

国际医学信息学领域研究现状与前沿分析^{*}

刘玉婷

(首都医科大学图书馆 北京 100069)

[摘要] **目的/意义** 讨论医学信息学发展趋势, 以期为医学信息学研究和发​​展提供参考。**方法/过程** 以科学引文索引扩展版收录的 31 种医学信息学领域期刊 2002—2021 年刊载论文为研究对象, 应用文献计量法、共词分析法分析该领域的发文总量、期刊、国家及地区、机构、作者、资助基金情况, 以及近 5 年研究方向、研究热点。**结果/结论** 医学信息学领域发文总量快速增长, 近 5 年研究方向集中在医疗保健科学与服务、计算机科学的跨学科应用、数学与计算生物学等方面, 研究热点集中在移动医疗、计算机技术、电子病历和自然语言处理技术等方面。该领域呈现多学科领域深度交叉融合、向健康信息学拓展、健康医疗数据开放共享、医疗信息化走向智慧医疗和智慧健康的趋势。

[关键词] 医学信息学; 文献计量法; 共词分析; 研究热点; 现状分析

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2023.07.006

Research Status and Frontier Analysis of International Medical Informatics Field

LIU Yuting

Library of Capital Medical University, Beijing 100069, China

[Abstract] **Purpose/Significance** To discuss the development trend of medical informatics, and to provide references for the research and development of medical informatics. **Method/Process** The research objects are the papers published in 31 medical informatics journals included in SCIE from 2002 to 2021. The bibliometric method and keyword co-occurrence analysis are used to analyze the number of published papers, journals, countries and regions, institutions, authors, funding status, as well as research directions, research hotspots in this field in recent five years. **Result/Conclusion** The total number of publications in the field of medical informatics has increased rapidly. In the past five years, the research directions have been focused on healthcare science and services, interdisciplinary applications of computer science, mathematics and computational biology, etc., research hotspots have been focused on mobile health, computer technology, electronic medical records (EMR), natural language processing (NLP) technology, etc. This field presents the trend of deep cross-integration of multi-disciplinary fields, expansion to health informatics, open sharing of health and medical data, and medical informatization moving towards smart medical treatment and smart health.

[Keywords] medical informatics; bibliometric method; co-word analysis; research hotspot; status analysis

[修回日期] 2023-04-13

[作者简介] 刘玉婷, 副研究馆员, 发表论文 10 余篇。

[基金项目] 教育部人文社会科学研究规划基金项目 (项目编号: 21YJA870008)。

1 引言

医学信息学是一门新兴交叉学科,以生物医学信息、数据、知识为研究对象,以信息科学与计算机技术为主要手段,融合多学科知识与方法,分析解决卫生管理、临床决策、生命健康等问题。20世纪50年代,医学信息学伴随计算机的发展而产生;70年代医学信息学作为一门学科被正式提出^[1];80年代以来,信息技术、生命科学的迅速发展推动医学信息学蓬勃发展。进入21世纪,医学信息学的研究范围不断扩展,近年来随着语义网、高性能计算及数据科学的发展,科学研究走向数据密集型研究,医学信息学也迎来更多的机遇与挑战。

当前有一些关于医学信息学领域的文献计量分析、现状分析,主要侧重从主题角度分析,如以Web of Science数据库检索主题为医学信息学的研究论文,应用CiteSpace绘制2010—2019年该领域的知识图谱^[2];以中国知网、中文社会科学引文索引、中国科学引文数据库为数据源,按照主题检索医学信息学领域论文,基于共词方法分析研究热点^[3];通过主题识别、聚类分析发现医学信息学分化出生物信息学和健康信息学两大分支^[4];还有研究从期刊角度,对2015—2019年国际医学信息学主题领域进行文献计量分析^[5]。本文基于Web of Science数据库医学信息学领域31种期刊所刊载论文,分析2002—2021年近20年发文总量、期刊、国家及地区、机构、作者、基金分布及近5年研究方向、研究热点等内容,以期对医学信息学的研究与发展提供参考。

2 数据来源与方法

2.1 数据来源

本文数据来源于科睿唯安的科学引文索引扩展版(science citation index expanded, SCIE)数据库、期刊引证报告(journal citation reports, JCR)数据库和InCites数据库。在JCR数据库中选择medical informatics类别,筛选出共31种期刊([https://](https://www.researchgate.net/publication/362277472_31_medical_informatics_journals)

www.researchgate.net/publication/362277472_31_medical_informatics_journals),在InCites中检索这31种期刊,文献类型选为“ARTICLE”OR“REVIEW”,出版年为2002—2021年,共得到54781条记录。检索日期为2022年7月1日。

2.2 指标说明

学科规范化引文影响力(category normalized citation impact, CNCI)是论文实际被引次数除以同文献类型、同出版年、同学科领域文献的期望被引次数^[6]。CNCI消除了文献类型、出版年、学科领域差异的影响,可以实现跨学科比较,与全球平均水平对比,如果CNCI>1,说明该论文的学术表现超过全球平均水平,反之则低于全球平均水平^[7]。引文影响力指论文篇均被引频次。h指数表示如果一位学者至少有n篇论文被引频次不低于n,则该学者h指数为n^[8]。h指数用于评估研究人员的学术产出数量与水平。

2.3 分析方法及工具

主要采用文献计量法、共词分析法分析。文献计量法是采用数学、统计学等方法对文献各方面属性、特征、规律等进行分析的一种方法^[9]。共词分析法是指分析文献中词语共同出现情况,利用聚类方法把复杂的共词网络简化为较少有代表性的词语类团^[10],被广泛用于对某一研究领域的热点分析^[3, 11-12]。使用BibExcel提取文献研究方向,应用可视化软件VOSviewer 1.6.18作为研究方向、关键词共现分析工具^[13-16]。

3 结果与分析

3.1 发文总量

2002—2021年医学信息学领域31种期刊发文总量快速增长,见图1。可分为3个阶段:第1阶段2002—2011年,平稳发展阶段;第2阶段2012—2018年,较第1阶段年均发文数量成倍增长;第3阶段2019—2021年,加速增长,年均发文量达到6000多篇。医学信息学领域期刊文献总量的增长趋势线指数模型拟合最优($y =$

$767.02e^{0.1036x}$, $R^2 = 0.9628$), 说明该领域研究呈指数快速增长。

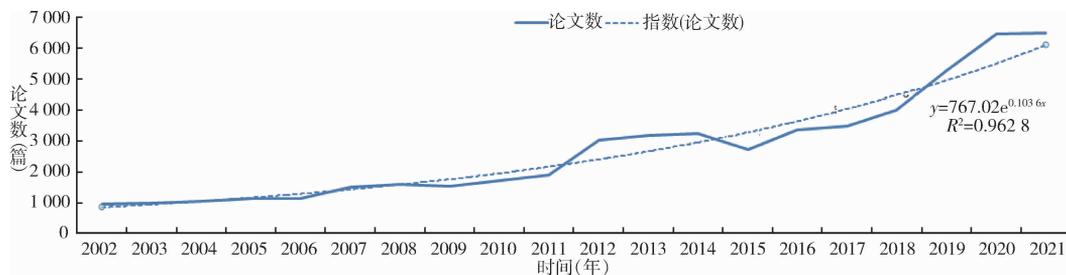


图 1 医学信息学领域发文情况

3.2 期刊情况

2002—2021 年 31 种医学信息领域期刊中发文量前 3 位是《医学统计》《医学互联网研究杂志》《生物医学中的计算机方法和程序》。被引频次前 3 位是《医学统计》《医学互联网研究杂志》《美国医学信息学协会杂志》。31 种期刊中 Q1 期刊 16 种,

Q2 期刊 5 种, Q3 期刊 7 种, Q4 期刊 3 种。CNCI 最高的为 5.16, 15 种期刊 CNCI 大于 1。期刊影响力因子《柳叶刀数字医疗》最高, 为 36.615。引文影响力前 3 位是《医学统计》《美国医学信息学协会杂志》《生物医学信息学杂志》。医学信息领域发文量前 10 位期刊情况, 见表 1。

表 1 医学信息学领域发文量前 10 位期刊情况

序号	刊名	ISSN	Web of Science 论文数 (篇)	被引 频次	期刊影响 因子分区	期刊 影响因子	CNCI	引文 影响力
1	《医学统计》(Statistics in Medicine)	0277-6715	6 127	195 068	Q1	2.497	1.09	31.84
2	《医学互联网研究杂志》 (Journal of Medical Internet Research)	1438-8871	5 689	130 380	Q1	7.076	1.70	22.92
3	《生物医学中的计算机方法和程序》 (Computer Methods and Programs in Biomedicine)	0169-2607	3 993	73 597	Q1	7.027	1.22	18.43
4	《美国医学信息学协会杂志》 (Journal of the American Medical Informatics Association)	1067-5027	3 071	97 398	Q1	7.942	1.65	31.72
5	《医学 & 生物工程 & 计算》 (Medical & Biological Engineering & Computing)	0140-0118	2 783	49 671	Q2	3.079	0.67	17.85
6	《临床实践评估杂志》 (Journal of Evaluation in Clinical Practice)	1356-1294	2 778	36 331	Q3	2.336	0.59	13.08
7	《医疗系统杂志》(Journal of Medical Systems)	0148-5598	2 635	42 169	Q1	4.920	1.01	16.00
8	《国际医学信息学杂志》 (International Journal of Medical Informatics)	1386-5056	2 485	55 390	Q1	4.730	1.09	22.29
9	《生物医学信息学杂志》(Journal of Biomedical Informatics)	1532-0464	2 482	76 572	Q1	8.000	1.26	30.85
10	《BMC 医学信息学与决策》 (BMC Medical Informatics and Decision Making)	0277-6715	2 336	33 462	Q3	3.298	0.87	14.32

3.3 国家及地区分布

文涉及国家或地区 167 个, 涉及论文数量前 10 位的国家及地区, 见表 2。

医学信息学 31 种期刊 2002—2021 年所刊载论

表 2 医学信息学领域论文国家或地区分布

序号	国家或地区	Web of Science 论文数 (篇)	被引频次	CNCI	引文影响力	被引次数排名前 1% 的论文百分比 (%)	所有开放获取论文百分比 (%)	占 31 种期刊所有论文百分比 (%)
1	美国	20 644	450 152	1.19	21.81	1.77	63.30	37.68
2	英国	6 169	187 579	1.49	30.41	3.02	63.62	11.26
3	英格兰	5 538	173 006	1.53	31.24	3.03	64.16	10.11
4	德国	4 597	56 147	0.72	12.21	0.87	39.35	8.39
5	中国	4 172	53 376	1.28	12.79	1.99	44.13	7.62
6	加拿大	3 750	86 307	1.23	23.02	2.08	48.99	6.85
7	澳大利亚	3 228	61 515	1.23	19.06	1.95	59.54	5.89
8	荷兰	2 676	65 125	1.24	24.34	1.68	64.42	4.88
9	西班牙	2 157	34 949	1.04	16.20	0.83	49.79	3.94
10	意大利	1 768	31 068	1.07	17.57	1.41	40.95	3.23

注：中国论文统计未包含港澳台地区。

3.4 机构分布

31 种期刊 2002—2021 年涉及发文机构 6 211 个，发文量前 10 位机构，见表 3。

3.5 高产作者

医学信息领域 31 种期刊 2002—2021 年涉及作者 26 万余人，论文产出前 10 位作者，见表 4。

表 3 医学信息领域前 10 位发文机构

序号	发文机构	Web of Science 论文数 (篇)	被引频次	CNCI	国家或地区	国际合作百分比 (%)	被引次数排名前 1% 的论文百分比 (%)	所有开放获取论文百分比 (%)
1	欧洲研究型大学联盟 (League of European Research Universities)	4 530	129 622	1.40	欧盟	46.64	2.30	60.15
2	哈佛大学 (Harvard University)	2 223	62 405	1.55	美国	33.56	3.42	75.03
3	加州大学系统 (University of California System)	1 613	34 681	1.40	美国	27.15	2.98	75.39
4	伦敦大学 (University of London)	1 265	37 199	1.69	英格兰	54.23	3.48	72.02
5	德克萨斯大学系统 (University of Texas System)	1 132	20 291	1.08	美国	24.56	1.06	72.00
6	多伦多大学 (University of Toronto)	1 005	33 367	1.69	加拿大	37.81	3.58	57.21
7	法国研究型大学联盟 (UDICE - French Research Universities)	878	14 764	1.00	法国	42.37	1.14	50.23
8	哥伦比亚大学 (Columbia University)	877	25 613	1.44	美国	26.23	1.94	78.22
9	美国国立卫生研究院 (National Institutes of Health (NIH) - USA)	825	21 804	1.30	美国	22.67	2.91	72.24
10	美国退伍军人事务部 (US Department of Veterans Affairs)	816	21 732	1.48	美国	13.36	2.08	72.92

表 4 医学信息领域前 10 位高产作者

序号	作者	所属机构	Web of Science 论文数 (篇)	被引频次	CNCI	国际合作 百分比 (%)	被引次数 排名前 1% 的 论文百分比 (%)	h 指数
1	Bates D W	哈佛大学 (Harvard University)	143	4 524	1.61	37.06	1.40	37
2	Hripcsak G	哥伦比亚大学 (Columbia University)	86	2 382	1.86	15.12	2.33	27
3	Weng C H	哥伦比亚大学 (Columbia University)	79	1 676	1.19	13.92	1.27	20
4	Wright A	哈佛大学 (Harvard University)	73	2 011	1.34	2.74	1.37	26
5	Sittig D F	德克萨斯大学休斯顿健康科学中心 (University of Texas Health Science Center Houston)	71	1 515	1.30	12.68	0.00	24
6	Xu H	德克萨斯大学休斯顿健康科学中心 (University of Texas Health Science Center Houston)	68	1 411	1.70	41.18	2.94	20
7	Liu H F	梅奥医学中心 (Mayo Clinic)	66	1 393	1.70	18.18	4.55	18
8	Denny J C	范德堡大学 (Vanderbilt University)	63	2 785	2.39	7.94	7.94	30
9	Del Fiol G	犹他大学 (University of Utah)	61	622	0.70	4.92	0.00	15
10	Acharya U R	亚洲大学 (Asia University)	57	3 664	3.00	96.49	12.28	24

3.6 基金分布

31 种期刊 2002—2021 年涉及基金资助机构 678 个, 基金资助论文 22 806 篇, 占论文总数的

41.63%。前 10 位资助基金, 见表 5。中国国家自然科学基金资助论文国际合作百分比相对较高, 开放论文占比相对较低。

表 5 医学信息学论文前 10 位资助基金

序号	名称	Web of Science 论文数 (篇)	被引频次	CNCI	国际合作 百分比 (%)	被引次数排名 前 1% 的论文 百分比 (%)	所有开放 获取论文 百分比 (%)
1	美国卫生和公众服务部 (United States Department of Health & Human Services)	9 033	233 899	1.31	19.01	1.97	81.18
2	美国国立卫生研究院 (National Institutes of Health (NIH) - USA)	8 217	214 793	1.32	19.42	2.02	82.63
3	欧盟委员会 (European Commission)	2 870	80 350	1.44	40.66	2.75	65.05
4	中国国家自然科学基金 (National Natural Science Foundation of China)	2 170	26 920	1.27	35.07	1.89	38.85
5	美国国家癌症研究所 (NIH National Cancer Institute)	1 896	45 089	1.13	19.88	1.95	80.80
6	美国国立医学图书馆 (NIH National Library of Medicine)	1 728	55 512	1.39	13.25	1.97	92.65
7	英国科研创新办公室 (UK Research & Innovation)	1 495	59 182	2.01	40.27	5.08	74.92
8	美国国立卫生研究院国家转化科学促进中心 (NIH National Center for Advancing Translational Sciences)	1 065	20 624	1.49	13.99	2.16	94.84
9	英国医学研究理事会 (Medical Research Council UK)	1 036	47 503	2.15	40.54	5.69	72.97
10	美国卫生保健和质量管理局 (Agency for Healthcare Research & Quality)	945	25 427	1.31	12.49	1.80	77.35

为大健康信息学, 狭义的健康信息学将其看作医学信息学的并列或者分支。国际健康信息学快速发展, 并不断扩展, 电子健康病历、卫生保健、健康信息技术、健康数据应用等是健康信息学的发展方向。中国健康信息学在初步发展阶段, 在健康中国、数字中国发展战略下, 在数字健康、慢病控制、疾病预防、健康促进等方面, 健康知识信息服务、大众健康素养的提升将成为医学信息知识服务的新内容。

4.3 健康医疗数据的开放共享

随着健康医疗数据的开放共享, 数据的可获得性提高, 真实世界研究将日益广泛, 多源异构数据的整合, 医学知识数据的处理、表达, 语义技术的发展, 数据的深度挖掘和广泛应用将进一步促进精准医学的发展。健康医疗大数据的开放共享在临床诊疗与决策、药物研发、公共卫生监测、公共健康管理、医疗卫生政策决策等方面具有重要意义。国际健康医疗开放数据已具规模, 形成相对完善的应用体系、政策法规以及平台, 中国也在积极推动健康医疗数据的开放和应用, 目前已成立多个健康医疗大数据研究院, 形成多个数据平台, 促进其应用发展。健康医疗数据开放共享的同时更加强调关注远程医疗与移动医疗、人工智能与深度学习等计算机技术应用、数据伦理与数据安全等。应积极推动健康医疗大数据的开放获取及应用, 从过程维度、支撑维度、管理维度等方面, 促进相关工作, 提高数据的共享和重用, 促进科学交流, 服务于科技创新、医疗卫生事业发展。

4.4 医疗信息化走向智慧医疗、智慧健康

5G 网络、云计算、大数据、人工智能、区块链和物联网等新一代信息技术的发展及其与医疗服务深度融合, 将进一步推进医疗信息化建设走向智慧医疗、智慧健康。2021 年国务院办公厅印发《关于推动公立医院高质量发展的意见》, 提出推进电子病历、智慧服务、智慧管理“三位一体”的智慧医院建设和医院信息标准化建设。大力发展远程医疗和互联网诊疗, 推动手术机器人等智能医疗设备和

智能辅助诊疗系统的研发与应用。智慧病房、智慧医院、移动医疗将实现以患者为中心的个性化、精准智慧医疗服务。可穿戴设备、基因测序技术、虚拟现实技术等将提供智能健康监测、评估与指导。数智环境下的智慧健康信息服务成为健康信息服务发展的新方向。

5 结语

医学信息学作为新兴交叉学科, 近年来快速发展, 受到学术界越来越多的关注, 对该领域 31 种高影响力期刊 2002—2021 年刊载论文情况进行研究, 有助于了解掌握医学信息学发展现状以及趋势。2017—2021 年医学信息学交叉关联比较强的方向有医疗保健科学与服务、计算机科学的跨学科应用、数学与计算生物学、计算机科学及信息系统、统计与概率等; 研究重点包括移动医疗、计算机技术、电子病历和自然语言处理技术等。医学信息学相关研究在公共卫生政策、临床决策支持、医学科学研究、医药研发、科研评价等方面占有重要地位。医学信息学发展还存在一些问题^[18], 如电子病历信息化数据整合与标准化^[19]、数据共享、数据安全与数据伦理等, 科研人员应以发展的眼光、包容的态度、合作的精神致力于医学信息学的发展, 并根据当前实际状况与社会发展需求深入研究, 以助力卫生健康信息事业的蓬勃发展, 提高公众健康水平。

参考文献

- 1 COLLEN M F. Origins of medical informatics [J]. The western journal of medicine, 1986, 145 (6): 778 - 785.
- 2 孙晶晶, 张帆, 张军亮, 等. 基于 WOS 的医学信息学领域研究现状分析 [J]. 中华医学图书情报杂志, 2021, 30 (1): 60 - 66.
- 3 钱巨敏, 徐清, 孟瑞君. 基于共词分析的医学信息学热点研究 [J]. 中国多媒体与网络教学学报: 下旬刊, 2021 (6): 7 - 9.
- 4 徐璐璐, 杜建, 叶鹰. 21 世纪以来医学信息学研究走向及其健康信息学转向 [J]. 情报学报, 2020, 39 (7): 777 - 786.

(下转第 67 页)

- 采集与分析系统构建研究 [J]. 湖南中医药大学学报, 2020, 40 (1): 70-74.
- 2 DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]. Minneapolis: The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019.
 - 3 王东波, 刘畅, 朱子赫, 等. SikuBERT 与 SikuRoBERTa: 面向数字人文的《四库全书》预训练模型构建及应用研究 [J]. 图书馆论坛, 2022, 42 (6): 31-43.
 - 4 LEE J, YOON W, KIM S, et al. BioBERT: a pretrained biomedical language representation model for biomedical text mining [J]. Bioinformatics, 2020, 36 (4): 1234-1240.
 - 5 ALSENTZER E, MURPHY J, BOAG W. Publicly available clinical BERT embeddings [C]. Minneapolis: The 2nd Clinical Natural Language Processing Workshop, 2019.
 - 6 IZ B, KYLE L, ARMAN C. SciBERT: a pretrained language model for scientific text [C]. Hong Kong: The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.
 - 7 LEE J S, HSIANG J. PatentBERT: patent classification with fine-tuning a pre-trained bert model [J]. World patent information, 2020, 61 (6): 409-418.
 - 8 王永炎. 中国现代名中医医案精粹 [M]. 北京: 人民卫生出版社, 2010.
 - 9 苏礼. 古今医案按 [M]. 北京: 人民卫生出版社, 2007.
 - 10 伊广谦, 李占永. 明清十八家名医医案 [M]. 北京: 中国中医药出版社, 2017.
 - 11 邹鹤瑜. 丁甘仁医案 [M]. 上海: 上海科学技术出版社, 2001.
 - 12 MIKOLOV T, SUTSKEVER L, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]. New York: The 26th International Conference on Neural Information Processing Systems, 2013.
 - 13 PENNINGTON J, SOCHER R, MANNING C D. Glove: global vectors for word representation [C]. Doha: The 2014 Conference on Empirical Methods in Natural Language Processing, 2014.

(上接第 44 页)

- 5 单治易, 安新颖, 关陟昊. 2015—2019 年国际医学信息学计量研究 [J]. 中国数字医学, 2021, 16 (1): 88-95.
- 6 MOED H F. Measuring contextual citation impact of scientific journals [J]. Journal of informetrics, 2010, 4 (3): 265-277.
- 7 陈斯斯, 刘春丽. 基于学科规范化引文影响力和相对引用率的论文学术影响力评价 [J]. 中华医学图书情报杂志, 2021, 30 (12): 25-30.
- 8 HIRSCH J E. An index to quantify an individual's scientific research output [J]. Proceedings of the national academy of sciences of the United States of America, 2005, 102 (46): 16569-16572.
- 9 季淑娟, 董月玲, 王晓丽. 基于文献计量方法的学科评价研究 [J]. 情报理论与实践, 2011, 34 (11): 21-25.
- 10 钟伟金, 李佳, 杨兴菊. 共词分析法研究 (三)——共词聚类分析法的原理与特点 [J]. 情报杂志, 2008 (7): 118-120.
- 11 虞秋雨, 徐跃权. 共词分析中高频词阈值确定方法的实证研究——以新冠肺炎文献高频词选取为例 [J]. 情报科学, 2020, 38 (9): 90-95.
- 12 丁敬达, 陈一帆, 刘超, 等. 基于共词和 Word2Vec 加权向量的文献——主题语义匹配分析方法 [J]. 图书情报工作, 2022, 66 (12): 108-116.
- 13 周晓分, 黄国彬, 白雅楠. 科学计量可视化软件的对比与数据预处理研究 [J]. 图书情报工作, 2013, 57 (23): 64-72.
- 14 VAN ECK N J, WALTMAN L. Text mining and visualization using VOSviewer [J]. ISSI newsletter, 2011, 7 (3): 50-54.
- 15 李艳, 张悦, 曾可, 等. 文献信息分析工具的比较 [J]. 中华医学图书情报杂志, 2015, 24 (11): 41-47.
- 16 高凯. 文献计量分析软件 VOSviewer 的应用研究 [J]. 科技情报开发与经济, 2015, 25 (12): 95-98.
- 17 周晓英, 裴俊良. 健康信息学的学科范畴与中国健康信息学的发展——兼述健康信息学学科建设与发展学术研讨会 [J]. 中国图书馆学报, 2022, 48 (2): 76-93.
- 18 钱庆. 医学信息学发展现状与展望 [J]. 中华医学信息导报, 2020, 35 (18): 10.
- 19 崔少国, 陈俊桦, 李晓虹. 融合语义及边界信息的中文电子病历命名实体识别 [J]. 电子科技大学学报, 2022, 51 (4): 565-571.