

我国近 10 年医学信息学研究计量分析*

林 鑫 郭进京 李梦园 任慧玲

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 基于 Web of Science 数据库, 收集 2008 - 2017 年我国作者发表的医学信息学领域的外文文献, 从年度发文量、合作国家、重要发文机构和来源期刊分布 4 方面进行定量分析。对高频关键词进行聚类, 借助 Ucinet 对高频关键词进行可视化分析, 归纳近 10 年我国医学信息学领域外文文献的 6 大研究热点。

〔关键词〕 医学信息学; 文献计量学; 研究热点; 共词分析; 聚类分析

〔中图分类号〕 R - 056 〔文献标识码〕 A 〔DOI〕 10. 3969/j. issn. 1673 - 6036. 2019. 02. 013

Bibliometrics Analysis of Study on Medical Informatics in China during the Last Decade LIN Xin, GUO Jinjing, LI Mengyuan, REN Huiling, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 The paper collects foreign literature on medical informatics published by Chinese authors during 2008 - 2017 from the database of Web of Science, and conducts bibliometrics analysis in four aspects such as annual amount of literature, cooperative country, important issuing institution and distribution of source journals. It conducts clustering of high - frequency keywords and then visual analysis of such keywords by Ucinet to generalize top 6 hotspots in foreign literature on medical informatics study over the past decade in China.

〔Keywords〕 medical informatics; bibliometrics; study hotspots; co - word analysis; cluster analysis

1 引言

医学信息学 (Medical Informatics) 是在卫生保健和医学方面应用计算机对数据进行分析 and 传播的信息科学领域^[1], 是一门发展较快的新兴交叉学科^[2]。近年来随着研究范围的不断扩大, 语义网、大数据处理以及高性能技术的飞速发展, 医学信息学发展迎来更多机遇, 研究热点不断涌现, 新的有

价值的研究成果也层出不穷^[3]。目前我国医学信息学的发展越来越走向专业化和国际化, 随着计算机技术的发展和新兴技术的出现, 不同阶段的研究热点也会发生变化。本文通过对 2008 - 2017 年我国医学信息学领域的外文文献进行计量学分析, 以期掌握该领域的发展情况和研究热点, 为我国医学信息学研究水平的提高提供理论基础。

2 数据来源及方法

2.1 数据来源

本研究以 Web of Science (WOS) 核心合集作为数据源, 检索我国大陆、台湾、澳门和香港地区发表的医学信息学研究文献。检索策略为 ((SU = Medical Informatics) AND CU = (CHINA OR HONG KONG OR MACAU OR PEOPLES R CHINA OR TAIWAN)),

〔收稿日期〕 2018 - 09 - 17

〔作者简介〕 林鑫, 硕士研究生, 发表论文 2 篇; 通讯作者: 任慧玲, 研究馆员, 硕士生导师。

〔基金项目〕 中国医学科学院医学与健康科技创新工程“中文临床医学术语系统构建研究”(项目编号: 2017 - I2M - 3 - 014)。

时间限定为 2008 – 2017 年，文献类型选择 “Article” 和 “Proceedings Paper” 进行精炼，检索时间为 2018 年 5 月 7 日，共检索到文献 4 726 篇。

2.2 研究方法

将全部文献以纯文本的格式导出后直接导入专业数据分析工具 TDA，清洗整合后即可进行分析。首先采用定量的方法对近 10 年我国医学信息学领域发表的外文文献基本特征（年度发文量、合作国家、重要发文机构和来源期刊）进行分析，然后利用 SPSS 19.0、Ucinet 和共词聚类分析方法归纳我国近 10 年医学信息学领域国际发文的研究热点。

3 结果与分析

3.1 文献基本特征

3.1.1 年度发文量 科学文献的数量是衡量科学知识量的重要尺度之一，因此某个领域年代论文量的变化能够直观反映出该领域学科发展态势^[4]。近 10 年我国医学信息学领域发表的 4 726 篇外文文献的年代分布，见图 1。可以发现 2008 – 2011 年的年度发文量变化幅度较大，之后进入快速发展时期，年文献量增长迅速，在 2016 年达到历史最高峰。2017 年的发文量虽然有所下降，但这并不意味着研究的停滞，一方面可能是由于 WOS 数据库文献收录尚不完善，另一方面可能是由于医学信息学领域正面临新的突破^[5]，这也符合学科发展的规律。整体来看 10 年间我国医学信息学领域外文文献发文总体呈逐年上升趋势，未来也将会受到越来越多学者关注。

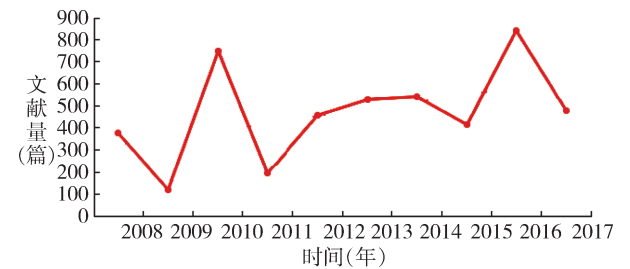


图 1 年度发文量趋势

3.1.2 合作国家 4 726 篇外文文献中与其他国家合作的发文量为 1 450 篇，占文献总量的 30.68%，作者来自 59 个国家，具有鲜明的国际合作化特点。与国内研究人员合作发文并且发文量在 10 篇以上的研究人员国籍信息，见表 1。可发现近 10 年与中国合作最多的是来自美国的研究人员，合作发文量为 667 篇，占与其他国家合作发文总数量的 46%，即接近一半的合作文献都有美国的参与；其次是澳大利亚、英国、新加坡和加拿大等国家。

表 1 合作国家（合作发文量≥10 篇）					
序号	国家	合作 发文量（篇）	序号	国家	合作 发文量（篇）
1	美国	667	10	芬兰	19
2	澳大利亚	126	11	瑞典	19
3	英国	106	12	荷兰	15
4	加拿大	81	13	印度	13
5	新加坡	71	14	印度尼西亚	11
6	日本	67	15	巴基斯坦	11
7	法国	28	16	瑞士	11
8	韩国	26	17	沙特阿拉伯	10
9	德国	25			

3.1.3 重要发文机构 在 WOS 数据库中对中国研究者的姓名进行区分时，由于存在中文同音不同字但英文简写相同等现象，因此具有很大难度。为降低研究误差不对发文作者进行统计，而是将发文机构作为研究对象，对机构近 10 年的总发文量进行分析。这些文献中共涉及国内外机构 2 191 所，其中发文量在 60 篇以上的 23 所机构，见表 2。可以发现这 23 所机构发文量共占总发文量的 51.03%。在这 23 所机构中中国大陆的研究机构有 12 所，发文量占 27.42%；台湾地区的研究机构 8 所，发文量占 16.93%；香港地区的研究机构 3 所，发文量占 6.69%。从发文机构的性质来看主要集中在大学。在此研究范围内可以看出 2008 – 2017 年间中国大陆的科研成果可以作为我国医学信息学国际发文的重要来源。

表 2 重要发文机构统计（发文量≥60 篇）

序号	机构	发文量（篇）	百分比（%）	序号	机构	发文量（篇）	百分比（%）
1	中国科学院大学	237	5.01	13	中国科学技术大学	97	2.05
2	国立台湾大学	172	3.64	14	北京航空航天大学	91	1.93
3	上海交通大学	145	3.07	15	香港理工大学	89	1.88
4	国立阳明大学	131	2.77	16	中国医科大学	80	1.69
5	台北医学大学	125	2.64	17	北京大学	80	1.69
6	浙江大学	119	2.52	18	国立台湾大学医学院	73	1.54
7	香港大学	116	2.45	19	长庚大学	63	1.33
8	国立成功大学	115	2.43	20	华中科技大学	63	1.33
9	清华大学	115	2.43	21	东北大学	61	1.29
10	香港中文大学	111	2.35	22	台北荣民总医院	61	1.29
11	哈尔滨工业大学	107	2.26	23	国立中正大学	60	1.27
12	北京理工大学	101	2.14				

3.1.4 来源期刊 统计近 10 年我国医学信息学领域外文文献的期刊收录和国际会议收录情况可以发现其分布广泛，共包括外文期刊和国际会议 381 种，涉及医学信息学、计算机和医学统计等多个领域。收录文献数量在 15 篇以上的 21 种外文期刊以及这些期刊的 2017 年影响因子，见表 3。在这 21 种期刊中有 18 种属于 2017 版《期刊引用报告》（*Journal Citation*

Reports，JCR）^[6] 中医学信息学类目下的 25 种国际高影响力期刊。一般情况下期刊影响因子、论文影响力与论文的被引频次呈正相关，即在高影响因子期刊上的论文其国际影响力相对较高^[7]。这 21 种期刊的 2017 年影响因子在医学信息学领域大部分相对较高，在一定程度上说明 2008 – 2017 年我国医学信息学领域的外文期刊文献普遍具有较高的国际影响力。

表 3 发文外文期刊（载文量≥15 篇）

序号	刊名	国家	文献量（篇）	2017 年影响因子
1	《医疗系统杂志》	美国	421	2.098
2	《生物医学的计算机方法与程序》	荷兰	297	2.674
3	《医学研究中的统计方法》	英国	193	2.284
4	《IEEE 生物医学与健康信息学杂志》	美国	160	3.850
5	《医学与生物工程与计算》	德国	160	1.971
6	《生物医学信息学杂志》	美国	116	2.882
7	《IEEE 生物医学信息技术汇报》	美国	96	—
8	《医学互联网研究杂志》	加拿大	93	4.671
9	《医学人工智能》	荷兰	82	2.879
10	《BMC 医学信息与决策》	英国	79	2.134
11	《国际医学信息学杂志》	爱尔兰	77	2.957
12	《临床实践评估杂志》	英国	71	1.483
13	《计算机、信息学、护理学》	美国	58	1.056
14	《医学统计学》	英国	45	1.932
15	《生物医学工程、生物医学技术》	德国	38	1.096
16	《美国医学信息学会杂志》	英国	35	4.270
17	《实验医学与生物学进展》	美国	25	1.760
18	《医疗创新与监管科学》	美国	19	0.882
19	《IEEE 医学工程与生物工程杂志》	美国	17	—
20	《JMIR 移动医疗与健康照护》	加拿大	15	4.541
21	《医学信息方法》	德国	15	1.531

注：“—”表示在 JCR 中没有查到 2017 年影响因子

3.2 研究热点

3.2.1 关键词词频统计 利用 TDA 软件对 4 726 篇文献进行统计,共得到 12 530 个关键词。使用 TDA 软件的 List Cleanup 功能对这 12 530 个关键词进行自动清理,合并同根词、同义词、近义词、单复数形式,结合人工进行筛选和处理。同时结合医学信息学主题删除没有实质意义的词以及一些区域性词汇。经过处理共得到关键词 11 916 个。关键词最高频次为 105,选择词频不少于 16 次的 55 个关键词作为聚类分析的基础,见表 4。

表 4 发文关键词统计 (词频≥16 次)

序号	关键词	词频	序号	关键词	词频
1	支持向量机	105	29	遗传算法	21
2	电子健康档案	94	30	信息系统	21
3	特征选择	64	31	仿真	21
4	分类	50	32	社交媒体	21
5	数据挖掘	48	33	文本挖掘	21
6	认证	41	34	卷积神经网络	20
7	心电图	40	35	脑电图扫描器	20
8	磁共振成像	38	36	医疗保健	20
9	中国传统医学	34	37	蛋白质相互作用	19
10	移动医疗	33	38	云计算	19
11	机器学习	31	39	医学图像	19
12	安全	30	40	神经网络	19
13	脑-机接口	29	41	粒子群优化	19
14	网络	29	42	糖尿病	18
15	聚类分析	28	43	图像处理	18
16	自然语言处理	28	44	预测	18
17	深度学习	27	45	随机对照试验	18
18	护理	27	46	远程医疗	18

续表 4

19	远程监护医疗	26	47	小波变换	18
	信息系统				
20	乳腺癌	25	48	有限元分析	17
21	图像分割	25	49	医学信息学	17
22	本体	25	50	隐私	17
23	病人安全	25	51	大数据	16
24	智能卡	25	52	决策树	16
25	无线射频识别	23	53	图像配准	16
26	匿名性	22	54	肺癌	16
27	生物信息学	22	55	模式识别	16
28	错误诊断	21			

3.2.2 关键词聚类 使用 TDA 软件生成关键词 55 ×55 共现矩阵,使用 Ochiai 系数^[8]将共现矩阵转换成相似矩阵,但由于相似矩阵中的 0 值过多,统计时容易造成误差,因此用 1 和全部相似矩阵数据相减得到关键词的相异矩阵,见表 5。将关键词的相异矩阵导入 SPSS 19.0 软件中,利用系统聚类、离差平方和法对关键词进行聚类分析,得到聚类树状图,见图 2。在高频关键词共现矩阵的基础上运用 Ucinet 软件绘制高频关键词共现网络,见图 3。图 3 中的节点表示关键词,节点之间的连线表示节点所表示的关键词在同一篇文章中共现。节点越大,关键词出现的频次越大;连线越粗,关键词共现次数越多。根据树状图以及关键词共现网络,结合相关文献,将 55 个高频主题词分成 6 类,分析各个类内主题词之间的语义关系,可得到 2008 – 2017 年我国医学信息学领域外文文献的 6 类研究热点。

表 5 关键词相异矩阵 (部分)

关键词	支持向量机	电子健康档案	分类	特征选择	数据挖掘	认证	中国传统医学	机器学习	安全	深度学习
支持向量机	0.000	1.000	0.931	0.878	0.958	1.000	1.000	0.965	1.000	0.981
电子健康档案	1.000	0.000	0.985	1.000	0.911	1.000	0.982	0.944	0.962	1.000
分类	0.931	0.985	0.000	0.876	0.980	1.000	0.976	0.924	1.000	0.946
特征选择	0.878	1.000	0.876	0.000	1.000	1.000	0.979	0.910	1.000	0.976
数据挖掘	0.958	0.911	0.980	1.000	0.000	1.000	0.975	0.922	1.000	1.000
认证	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	0.857	1.000
中国传统医学	1.000	0.982	0.976	0.979	0.975	1.000	0.000	0.969	1.000	1.000
机器学习	0.965	0.944	0.924	0.910	0.922	1.000	0.969	0.000	1.000	1.000
安全	1.000	0.962	1.000	1.000	1.000	0.857	1.000	1.000	0.000	1.000
深度学习	0.981	1.000	0.946	0.976	1.000	1.000	1.000	1.000	1.000	0.000

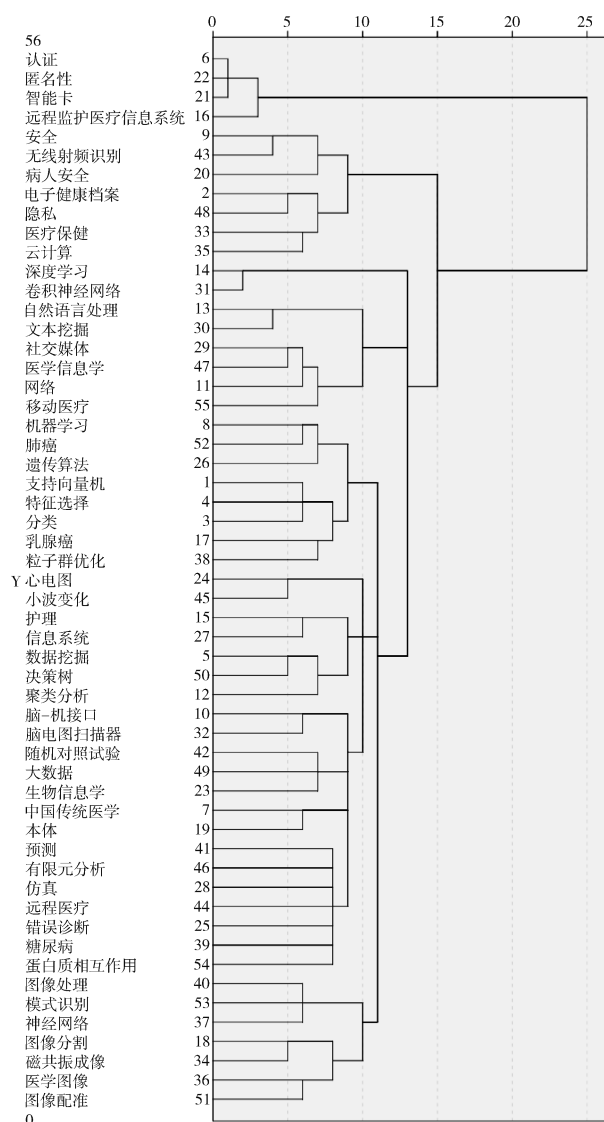


图2 发文研究热点聚类树

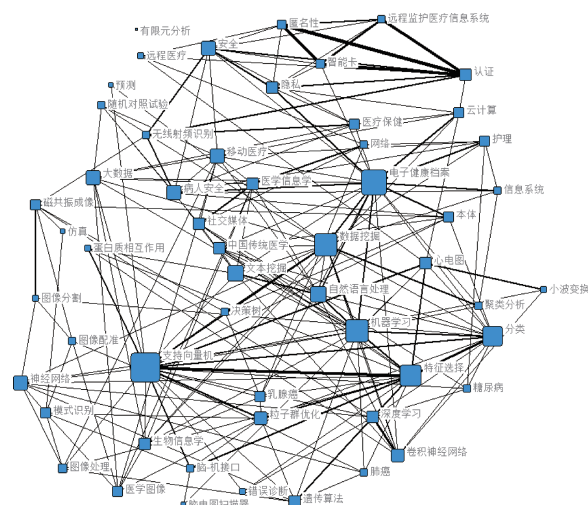


图3 发文关键词共现网络

3.2.3 6类研究热点 第1类：电子健康档案及远程医疗的相关隐私保护和安全管理（包括关键词6、22、21、16、9、43、20、2、48、33、35）。涉及的技术和方法主要有用户认证、智能卡技术和无限射频识别技术。如2012年Hsu CL等^[9]对基于智能卡的电子处方安全与隐私保护系统进行研究。Kuo KM等^[10]于2017年利用结构方程模型进行分析，提出能够潜在阻止护士违反电子健康档案隐私政策的策略。第2类：文本挖掘在生物医学领域中的应用（包括关键词14、31、13、30、29、47、11、55）。文本分类作为文本挖掘的基本任务之一，Yao L等^[11]于2017年对中医临床文本分类的不同类型的特征以及分类算法效果进行研究，并提出一种将深度学习文本表示与中医领域知识相结合的文本分类新方法，通过对中医临床记录进行文本分类的实证研究发现该方法比传统方法性能更优。第3类：机器学习方法在辅助临床疾病的诊断、治疗和临床决策中的应用（包括关键词8、52、26、1、4、3、17、38）。涉及的机器学习方法包括遗传算法、支持向量机、粒子群算法等。如Peng L等^[12]于2016年将最先进的生命科学研究与人工智能集成，并提出了一种半监督学习算法来对乳腺癌进行自动诊断。2017年Gu D等^[13]提出将遗传算法应用于基于案例的推理（Case Based Reasoning, CBR）分类法，该方法在乳腺癌诊断中具有很大的应用潜力。第4类：计算机技术在临床管理系统、医院信息系统和护理系统研发与管理中的应用（包括关键词24、45、15、27、5、50、12、10、32）。如2012年Xie M等^[14]介绍多维度心音数据库的开发；Qin Y等^[15]于2017年介绍重症监护信息系统在国内的开发和应用情况。第5类：生物信息学方法在蛋白质相互作用预测中的应用（包括关键词42、49、23、7、19、41、46、28、44、25、39、54）。如2016年Teng Z等^[16]开发一种利用基因本体术语预测蛋白质功能的新方法，实验结果证实该方法有效。第6类：医学图形图像的识别与处理研究（包括关键词40、53、37、18、34、36、51）。如Chen YW等^[17]考虑到以往的医学图像匹配方法仅具有理论意义且时间复杂度太高，于2011年提出一种用于匹配大

量 MR 图像的并行处理模型, 实验表明该模型不仅效果优于现有方法, 同时还具有更快的运行时间。Liu XB 等^[18]于 2015 年提出一种基于 Fisher 准则和遗传优化的新特征选择方法, 以解决肺部疾病 CT 征象的识别问题, 实验证明此方法具有较高识别效率。

4 结语

医学信息学作为一门新兴交叉学科, 目前正处于快速发展阶段。回顾医学信息学的内涵和发展历程有助于理清其发展面临的问题和趋势。本文采用文献计量学方法对 2008–2017 年我国医学信息学领域发表的外文文献从年度发文量、合作国家、重要发文期刊、来源期刊以及研究热点进行分析, 以期为我国医学信息学领域研究提供参考。

参考文献

- 1 National Library of Medicine. Medical Informatics [EB/OL]. [2018-06-23]. <https://meshb.nlm.nih.gov/record/ui?ui=D008490>.
- 2 代涛. 医学信息学的发展与思考 [J]. 医学信息学杂志, 2011, 32 (6): 2–16.
- 3 亢川博, 王伟, 张世玉, 等. 国内外医学信息学研究现状的可视化分析 [J]. 中华医学图书情报杂志, 2016, 25 (8): 24–30.
- 4 袁晓园, 华薇娜. 探索中国医学信息学国际发文的文献特征和研究主题 [J]. 现代情报, 2014, 34 (7): 125–129.
- 5 黄鹂, 曹东维. 国际医学信息学领域大数据研究热点分析 [J]. 医学信息学杂志, 2018, 39 (4): 2–7.
- 6 Clarivate Analytics. InCites Journal Citation Reports [EB/OL]. [2018-06-28]. <http://yxt.imicams.ac.cn/rwt/JCR/http://NJSYELUJN3SXT7DFPMYHI4DQNW3X85UTMW4YI3LTPMYGG55N/JCRJournalHomeAction.action?year=&editon=&journal=>.
- 7 李元, 王伟. 我国大陆体育科学研究国际化进展考察——基于 WOS 数据库的文献计量分析 [J]. 成都体育学院学报, 2013, 39 (1): 14–21.
- 8 陈慧茹, 肖相泽, 冯锋. 科技创新政策加权共词网络研

- 究——基于扎根理论与政策测量 [J]. 科学学研究, 2016, 34 (12): 1769–1776.
- 9 Hsu C L, Lu C F. A Security and Privacy Preserving E-prescription System Based on Smart Cards [J]. Journal of Medical Systems, 2012, 36 (6): 3637–3647.
- 10 Kuo K M, Talley P C, Hung M C, et al. A Deterrence Approach to Regulate Nurses' Compliance with Electronic Medical Records Privacy Policy [J]. Journal of Medical Systems, 2017, 41 (12): 198.
- 11 Yao L, Zhang Y, Wei B, et al. Traditional Chinese Medicine Clinical Records Classification Using Knowledge-powered Document Embedding [C]. Shenzhen: IEEE International Conference on Bioinformatics and Biomedicine. IEEE, 2017: 1926–1928.
- 12 Peng L, Chen W, Zhou W, et al. An Immune-inspired Semi-supervised Algorithm for Breast Cancer Diagnosis [J]. Computer Methods & Programs in Biomedicine, 2016, 134 (134): 259–265.
- 13 Gu D, Liang C, Zhao H. A Case-based Reasoning System Based on Weighted Heterogeneous Value Distance Metric for Breast Cancer Diagnosis [J]. Artificial Intelligence in Medicine, 2017 (77): 31.
- 14 Xie M, Xiao S, Liu T, et al. Multi-center, Multi-topic Heart Sound Databases and Their Applications [J]. Journal of Medical Systems, 2012, 36 (1): 33.
- 15 Qin Y, Zhou R, Wu Q, et al. The Effect of Nursing Participation in the Design of Acritical Care Information System: a case study in a Chinese hospital: [J]. BMC Medical Informatics & Decision Making, 2017, 17 (1): 165.
- 16 Teng Z, Guo M, Liu X, et al. Revealing Protein Functions Based on Relationships of Interacting Proteins and GO Terms [C]. Shenzhen: IEEE International Conference on Bioinformatics and Biomedicine. IEEE, 2017: 27.
- 17 Chen Y, Chen Y. Matching of a Huge Set of MR Images with a Parallel Processing Model [J]. Journal of Medical Systems, 2011, 35 (5): 795.
- 18 Liu X, Ma L, Li S, et al. Recognizing Common CT Imaging Signs of Lung Diseases Through a New Feature Selection Method Based on Fisher Criterion and Genetic Optimization [J]. Biomedical & Health Informatics IEEE Journal, 2015, 19 (2): 635–647.