

·热点分析·

2015年~2019年我国医疗大数据领域研究热点分析

苗豆¹,祝雅如¹,齐文娟²,吴雅琴¹

(内蒙古医科大学计算机信息学院¹,人文学院²,内蒙古呼和浩特 010110)

摘要:目的 了解我国国内医疗大数据发展及研究热点,为后续开展相关研究提供参考。方法 以中国知网(CNKI)为数据来源,运用 CiteSpace 及 VOSviewer 知识图谱分析工具,对 2015 年~2019 年发表在 CNKI 中的医疗大数据相关文献的研究热点进行图谱分析。结果 最终纳入文献 3123 篇,发文量自 2015 年呈上升趋势,来自若干机构的作者形成了一定的合作关系,其中核心作者共有 5 大类,分别以张政波、曹德森、黎檀实、薛万国及刘军为主;在医疗大数据领域,智慧医疗、人工智能、数据挖掘、云计算等受到广泛的关注。结论 有关医疗大数据的研究近年来受到研究者的重视,但研究者主要集中在一些科研水平较高的机构,研究者之间合作有待进一步加强,该领域相对缺乏应用广泛的成果。

关键词:医疗大数据;知识图谱;人工智能

中图分类号:R-1

文献标识码:A

DOI:10.3969/j.issn.1006-1959.2021.10.001

文章编号:1006-1959(2021)10-0001-04

Analysis of Research Hotspots in the Field of Medical Big Data in China from 2015 to 2019

MIAO Dou¹, ZHU Ya-ru¹, QI Wen-juan², WU Ya-qin¹

(School of Computer Information¹, School of Humanities², Inner Mongolia Medical University, Hohhot 010110, Inner Mongolia, China)

Abstract: **Objective** To understand the development and research hotspots of China's domestic medical big data, and to provide references for subsequent related research. **Methods** Using CNKI as the data source, using CiteSpace and VOSviewer knowledge graph analysis tools, a map analysis of the research hotspots of medical big data-related literature published in CNKI from 2015 to 2019. **Results** 3123 articles were finally included. The number of articles published had been on the rise since 2015. Authors from several institutions had formed a certain cooperative relationship. Among them, the core authors had 5 categories, mainly Zhang Zhengbo, Cao Deshen, Li Tanshi, Xue Wanguo and Liu Jun; In the field of medical big data, smart medical care, artificial intelligence, data mining, cloud computing, etc. had received extensive attention. **Conclusion** Research on medical big data has received attention from researchers in recent years, but researchers are mainly concentrated in some high-level scientific research institutions. The cooperation between researchers needs to be further strengthened, and there is a relatively lack of widely-used results in this field.

Key words: Medical big data; Knowledge graph; Artificial intelligence

随着医疗信息技术的发展,越来越多的医院使用各类医疗信息系统(如 HIS、EMR 等)开展医疗服务^[1]。由此在医疗领域产生了海量的医疗数据,并且健康医疗大数据正在从理论和概念逐步走向应用和实践^[2]。知识图谱是以科学知识为对象,显示科学知识的发展进程与结构关系的一种图形。借助知识图谱,人们可以洞察人类知识体系的各个领域和结构,构建复杂的知识网络,预测学科和知识前沿的发展趋势与研究热点。知识图谱作为符号主义发展的产物,是人工智能技术和系统中的重要组成部分,其在百科知识、生物信息、社交网络、社交网络以及网络安全等领域被广泛运用^[3,4]。本文使用 CiteSpace, VOSviewer 对医疗大数据领域的研究热点进行计量学分析,并利用软件绘制科学知识图谱,以期对医疗大数据领域的相关研究提供参考。

1 资料与方法

1.1 数据来源 以中国知网(CNKI)期刊数据库为数据来源,检索主题为“医疗大数据”和“medical big

data”,检索时间限定为 2015 年~2019 年,具体日期为 2019 年 12 月 29 日,共得到 3537 篇文献数据。纳入符合主题的期刊论文,剔除重复文献、会议论文、报纸、辑刊等与本文主题相关性较弱的文献,最终得到 3123 篇有效文献。

1.2 方法 将 CiteSpace 设置时区分割年份为 2015 年~2019 年,时间跨度分段长度为 1 年,术语选择为突显术语,节点类型可选择为关键词、主题,其中关键词共现数据选择 Create→Create a map based on bibliographic data,在 T type of analysis 选择 Co-occurrence,在 Unit of analysis 中选择 All keyword,加载自定义的词集,合并、替换或删除一些因写法不同(术语简称、全称的不同表达等)造成重复的关键词。VOSviewer 中,分析类型选择 Co-authorship→Authors,Co-occurrence→All keyword,计数方法选择全计数(Full counting),分析医疗大数据的研究热点。

1.3 数据统计分析 使用 CiteSpace.5.6.R3 和 VOSviewer 对医疗大数据相关文献的发文量及年代分布,学术研究核心学者,以及关键词频次、中心性、突现等进行可视化分析,全面展示医疗大数据研究的相关内容。

2 结果

2.1 文献发文量 2015 年~2018 年有关医疗大数据领域的发文量呈上升趋势,2018 年发文量高达 660

基金项目:1.内蒙古医科大学“三位一体”大学生创新创业培育项目(编号:SWYT2019014);2.内蒙古医科大学“青年领创”团队联盟项目(编号:QNLC-2020047)

作者简介:苗豆(1999.3-),女,内蒙古包头人,本科

通讯作者:吴雅琴(1978.11-),女,内蒙古乌兰察布人,硕士,副教授,主要从事医疗大数据分析与处理教学和科研工作

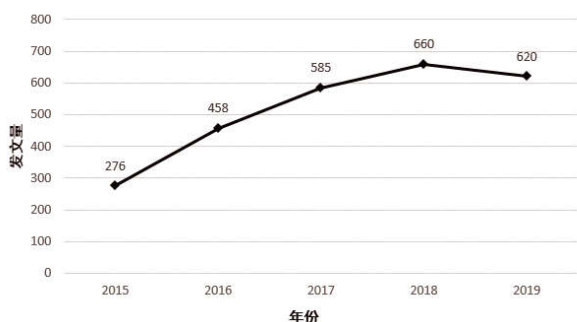


图1 2015年~2019年医疗大数据发文量趋势

篇,2019年发文量略有回落,发文量分布见图1。

2.2 作者合作网络 对纳入分析的3123篇有效文献进行核心作者合作网络制作,见图2。可视化结果展示所分析文献的核心作者共分为5大类,每类中发文量突出的分别是张政波、曹德森、黎檀实、薛万国、刘军等人。其中,作者张政波是北京市国际科技合作基地负责人,在创新型医疗器械和装备研发、电子

健康档案的二次分析利用等领域有较大贡献;作者曹德森曾与张政波、薛万国、黎檀实共同研究发表过急救大数据与Datathon活动等有关医疗大数据领域的论文;作者黎檀实在医药卫生领域有较大影响力,他主要在医疗大数据和医疗器械等领域展开研究;作者薛万国在医疗建设领域有独到的见解,从1988年开始从事医院信息化方面的研究开发与推广应用工作,他提出在做好病历共享的同时,还要注意病人隐私保护的问题;作者刘军在医学外科方面较为擅长,曾发表过肿瘤学以及外科学的相关学术论文。

2.3 关键词共现分析 利用VOSviewer进行关键词共现分析,所有关键词节点围绕“大数据”节点呈放射状展开,其中较为突出的节点有“医疗”“大数据时代”“智慧医疗”“人工智能”“数据挖掘”“云计算”等,见图3。

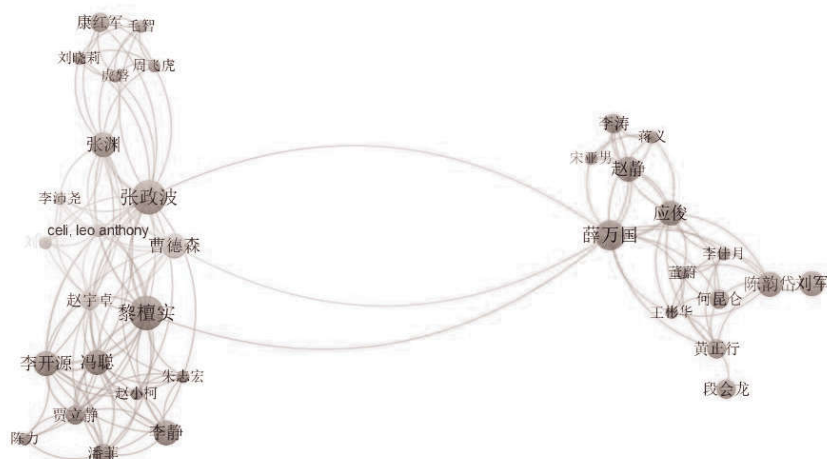


图2 核心作者合作网络

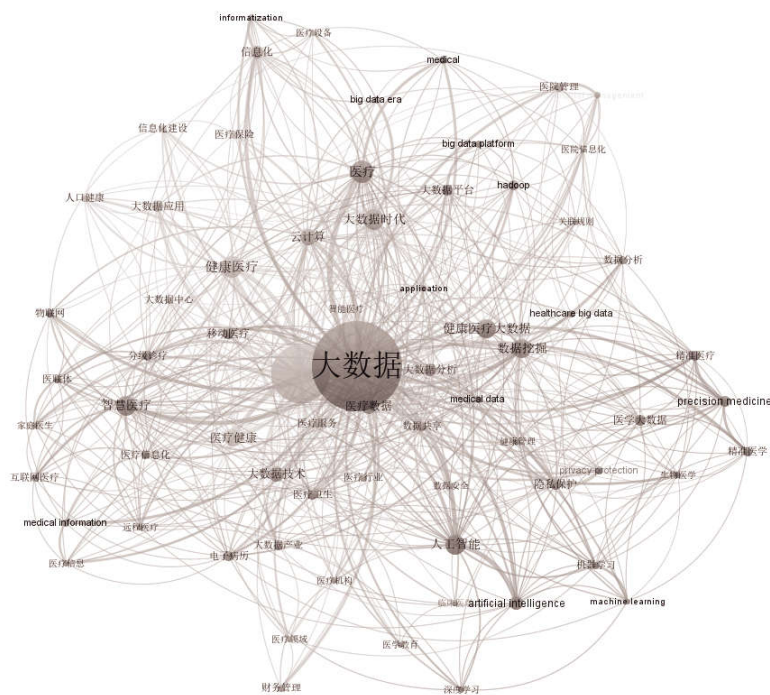


图3 关键词共现网络

2.4 高频关键词聚类 利用 CiteSpace 进行关键词聚类分析,医疗大数据研究领域关键词聚类共聚为 10 类,分别是“大数据”“医疗大数据”“精准医学”“大数据应用”“健康医疗大数据”“数据共享”“大数据时代”“数据挖掘”“移动医疗”和“信息资源共享”,见图

4;每一聚类所包括的聚类标签见表 1。
2.5 关键词突现分析 使用 CiteSpace 得到医疗大数据突现关键词,主要包括精准、数据资源、移动医疗等,见图 5。

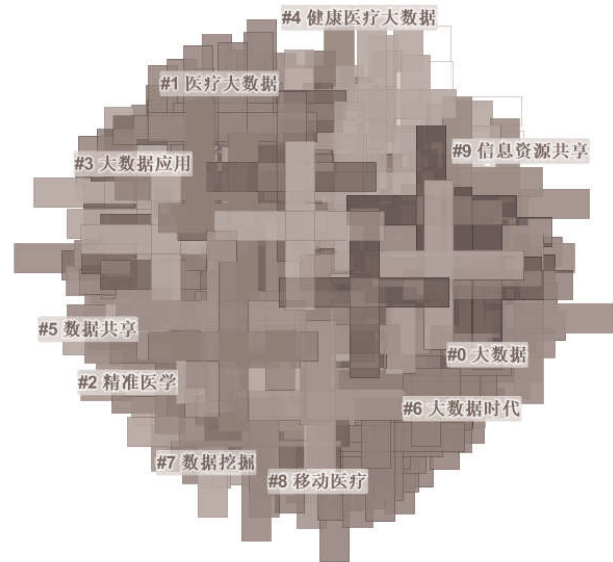


图 4 关键词聚类分析可视化网络

表 1 医疗大数据研究领域关键词聚类分布

ClusterID	Size	Silhouette	Mean(year)	Label(LLR)
0	81	0.513	2017	大数据;医疗大数据;医学信息等
1	62	0.648	2016	医疗大数据;全民健康;互联网医疗等
2	43	0.627	2016	精准医学;基因组学;医学模型等
3	40	0.689	2016	大数据应用;健康医疗;数据处理等
4	40	0.708	2017	健康医疗大数据;医疗保险等
5	39	0.687	2017	数据共享;数据质量;数据安全等
6	37	0.663	2016	大数据时代;医疗模式;个性化医疗等
7	36	0.53	2016	数据挖掘;Hadoop;云平台等
8	12	0.896	2015	移动医疗;机会网络;可植入传感器等
9	5	0.99	2015	信息资源共享;档案信息资源等

Top 11 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2015 - 2019
精准	2015	2.9362	2015	2016	■■■■■■■■
数据资源	2015	2.0959	2015	2016	■■■■■■■■
移动医疗	2015	3.5586	2015	2016	■■■■■■■■
医疗卫生	2015	2.9802	2015	2017	■■■■■■■■
领域	2015	3.2296	2015	2017	■■■■■■■■
挑战	2015	1.8528	2015	2016	■■■■■■■■
现代医药学	2015	2.0859	2015	2017	■■■■■■■■
数据采集	2015	1.9566	2016	2017	■■■■■■■■
数据处理	2015	2.5172	2016	2017	■■■■■■■■
大数据中心	2015	2.169	2017	2019	■■■■■■■■
分级诊疗	2015	2.5624	2017	2019	■■■■■■■■

图 5 2015 年~2019 年医疗大数据突现关键词

3 讨论

3.1 发文量分析 从刊载文献发文量来看,2015年~2018年发文量呈上升趋势,2019年略有回落,分析其中原因之一是本文所用数据的检索时间为2019年12月29日,很多期刊2019年所发表文献还未能CNKI检索,故分析数据中2019年文献数量有所下降。2016年,国家出台《国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见》,文件指出:健康医疗大数据应用发展将带来健康医疗模式的深刻变化,有利于激发深化医药卫生体制改革的动力和活力,提升健康医疗服务效率和质量,扩大资源供给,不断满足人民群众多层次、多样化的健康需求,有利于培育新的业态和经济增长点。相关文件的出台也在一定程度上推动了医疗大数据研究、应用的进一步增长。

3.2 作者分析 在作者合作网络中,每个节点代表一位作者,节点的大小代表作者发文量,节点的颜色代表按照默认的聚类方法得到的作者所属类群,网络中的连线代表了作者合作关系,线宽代表合作强度。作者合作网络表明,虽然形成了一些合作作者群,但深入分析会发现,合作作者大多来自同一个科研机构,覆盖面相对比较窄,国内还未形成较大范围的医疗大数据研究作者群。为进一步促进医疗大数据应用向纵深发展,今后亟待加强医疗大数据更广泛范围的合作。

3.3 研究热点分析 关键词是文献分析的重要组成部分,它高度精炼了文章的内容,通过分析关键词可以获知某领域的研究热点。根据关键词聚类结果,可将我国医疗大数据研究热点分为以下4个方面:①基于健康医疗大数据的数据应用及模型设计:在经过多年的医疗信息化建设后,各级各类医疗机构的医院信息系统存储了大量的数据,从规模上来讲形成了健康医疗大数据,但在应用层面上,这些规模庞大的数据很多处于“沉睡”状态。现代医学越来越依赖于具体数据的采集和判断^[9],如何借助于技术使这些沉睡的数据真正服务于医疗,服务于患者,提高医疗服务质量成为研究者关注的热点之一。②健康医疗大数据质量、数据资源共享及数据安全研究:大量的健康医疗数据分散在医疗机构不同的系统中,如何在保证数据安全的前提下,为研究者提供符合研究要求的高质量数据,并且高效共享健康医疗数据也是近些年研究者关注的热点。③健康医疗领域的技术探究:健康医疗大数据的应用离不开相关的大数据技术,如何将其它领域应用成熟的技术使之在医疗领域发挥更大作用,同样也受到研究者的关注。④移动医疗应用:2015年,移动医疗借助“互联网+”东风,在医疗行业大放异彩。从概念到产

品落地,再到如何切实有效的服务于患者,使患者在就医、问诊、复诊、转诊等各个环节真正体会到移动医疗的好处,也使得移动医疗成为近几年研究者关注的热点之一。

3.4 研究前沿分析 关键词突现分析可以了解研究领域的前沿热点。“突现”是指一个变量的值在短期内有非常大的波动。突现关键词分析有助于鉴定出某一段时期内最突出的文献知识,即文献信息的挖掘与文献知识扩散、文献知识吸收和文献知识扫描的发展进程^[6]。在本研究所设定的时间范围内,从时间序列上来看,2015年~2016年的突现词为“精准”“移动医疗”“数据资源”“卫生医疗”和“现代医学”,表明在这个时期,该领域的学者对与之相关的研究内容关注度较高。“大数据中心”和“分级诊疗”的有关内容在本研究时间段内,则集中出现在2018年和2019年。分级诊疗制度的建立是实现我国合理配置医疗资源、促进基本医疗卫生服务均等化的重要举措。推进分级诊疗制度建设是我国深化医药卫生体制改革的工作重点。国家计划到2020年完成分级诊疗的核心目标,即基本建立符合我国国情的分级诊疗制度^[7]。为实现此目标,加快“互联网+医疗健康”发展,完善国家级全民健康信息平台也就同样成为了工作重点,由此医疗机构大数据中心建设也在近几年成为学者关注和研究的重点之一。

综上所述,近几年医疗健康大数据在基础研究、临床应用等方面发挥了不可替代的作用。总体上看,我国已在医疗大数据的开发与利用上取得很大进展,在该领域研究中处于重要地位。健康医疗大数据成为当前该领域学者研究的一个重要热点,这为后续学者研究和实验提供了参考。

参考文献:

- [1]许培海,黄匡时.我国健康医疗大数据的现状、问题及对策[J].中国数字医学,2017,12(5):24-26.
- [2]Abramo G,D'Angelo CA,Reale E.Peer review vs bibliometrics:which method better predicts the scholarly impact of publications[J].Scientometrics,2019,121(1):537-554.
- [3]蔡立媛,李晓.中国广告学研究进展的知识图谱分析——基于Citespace的CSSCI数据库分析[J].全球传媒学刊,2020,7(3):139-154.
- [4]秦锦玉,翟洁,陈程,等.基于知识图谱的可视化技术研究[J].电子设计工程,2018,26(14):1-5.
- [5]埃里克·托普.颠覆医疗:大数据时代的个人健康革命[M].北京:电子工业出版社,2014.
- [6]程结晶,丁慢慢,朱彦君.国外信息管理领域知识流的新兴趋势及可视化分析[J].现代情报,2017,37(4):170-177.
- [7]施骞.上海市农村地区分级诊疗制度运行现状调查[D].沈阳农业大学,2020.

收稿日期:2020-12-23;修回日期:2021-01-04

编辑/成森