

机器学习在电子病历中应用的可视化分析

韦彩萍¹, 李翔¹, 翁梦晴², 吴旭生³, 胡庆元⁴, 刘建伟¹, 胡德华¹

1. 中南大学生命科学学院, 湖南 长沙 410013; 2. 广西医科大学信息与管理学院, 广西 南宁 530021; 3. 深圳市卫生健康发展研究和数据管理中心, 广东 深圳 518028; 4. 中南大学湘雅三医院, 湖南 长沙 410013

【摘要】为探讨机器学习方法在电子病历领域应用的研究现状、研究热点与前沿,以2000~2022年中国知网数据库和Web of Science核心合集数据库中关于机器学习在电子病历中应用的相关文献为数据来源,运用CiteSpace软件绘制国家/地区、作者、机构、关键词共现以及关键词突现5个方面科学知识图谱进行可视化对比分析,以便了解国内外研究的差异,为该领域的研究和发展提供参考。

【关键词】电子病历; 知识图谱; 机器学习; 深度学习; 可视化分析

【中图分类号】R318; R197.3

【文献标志码】A

【文章编号】1005-202X(2023)04-0521-08

Visualization analysis on machine learning application in electronic medical record

WEI Caiping¹, LI Xiang¹, WENG Mengqing², WU Xusheng³, HU Qingyuan⁴, LIU Jianwei¹, HU Dehua¹

1. School of Life Sciences, Central South University, Changsha 410013, China; 2. School of Information and Management, Guangxi Medical University, Nanning 530021, China; 3. Shenzhen Health Development Research and Data Management Center, Shenzhen 518028, China; 4. The Third Xiangya Hospital of Central South University, Changsha 410013, China

Abstract: The research status, hotspots and frontiers of machine learning application in electronic medical record are explored in the study. The research papers related to the application of machine learning in electronic medical record published from 2000 to 2022 in CNKI and Web of Science core collection are taken as data source. CiteSpace was used to draw the knowledge maps of countries/regions, authors, institutions, keywords co-occurrence and burst keywords for visualization analysis, so as to understand the research differences at home and abroad and provide reference for the related researches.

Keywords: electronic medical record; knowledge map; machine learning; deep learning; visualization analysis

前言

电子病历记录着患者住院全过程的所有诊疗信

息,包括病人姓名、性别等人口统计学信息,实验室检验结果、医学影像检查、医生诊断及护理记录等^[1-2]。随着机器学习和深度学习等人工智能技术的不断发展,医疗数据的大规模产生使得机器学习技术在电子病历中的应用迅速增加。对电子病历系统的利用,由传统的方便医院管理目的已转向利用机器学习技术对电子病历数据进行二次智能挖掘,如辅助临床诊断、药物警戒与新药研发、高风险疾病患病概率预测和个性化治疗方案制定等已成为研究热点^[3-5]。国内外学者对机器学习在电子病历中应用研究方面做出了诸多探索,并取得了重要进展。

目前有关于电子病历研究的文献梳理中,国内研究集中对电子病历某一子领域进行相关的综述和归纳;现有相关文献的研究方法上,以文本内容主观分析与总结居多,仅个别研究采用文献计量方法对研究结果进行可视化呈现并分析,极少研究者从宏

【收稿日期】2022-10-14

【基金项目】湖南省重点领域研发项目(2021WK2003);湖南省临床医疗技术创新引导项目(2021SK53703);深圳市卫生健康发展研究与大数据管理中心与中南大学合作项目(H202206140250001)

【作者简介】韦彩萍,硕士研究生,研究方向:健康信息传播、医药信息管理, E-mail: 212511068@csu.edu.cn; 李翔,硕士研究生,研究方向:医药信息管理、医学影像, E-mail: 1063049096@qq.com。韦彩萍和李翔系本文共同第一作者

【通信作者】刘建伟,硕士,副教授,研究方向:医学计算机技术、医院信息管理, E-mail: 802544@csu.edu.cn; 胡德华,博士,教授,博士生导师,研究方向:医学信息学、信息检索、医学大数据挖掘, E-mail: hudehua2000@163.com。刘建伟和胡德华系本文共同通信作者

观角度梳理机器学习方法在电子病历中的应用并做文献计量分析。为了更好地把握前沿动态,本文通过运用 CiteSpace 作为研究工具对国内外文献进行分析,绘制科学知识图谱进行可视化分析,整体把握国内外机器学习在电子病历中的应用研究热点以及发展趋势,以期对相关研究提供参考。

1 数据来源与研究方法

1.1 数据来源

中文文献主要选自“中国知网”(CNKI)数据库,采用高级检索方式,构建检索式为主题=(“机器学习”+“决策树算法”+“贝叶斯算法”+“聚类算法”+“支持向量机算法”+“关联规则”+“深度学习”+“人工神经网络”+“随机森林”+“循环神经网络”+“卷积神经网络”)AND 主题=(“电子病历”+“计算机化的病案系统”+“计算机的病人记录”),将检索语言类型限定为中文,选择期刊论文,进行同义词扩展检索,时间范围限制为 2000 年至 2022 年,检索时间为 2022 年 7 月 15 日,共检索到 214 篇文献。经过浏览文献标题及摘要部分,对于与研究主题不相关的文章进行排除,最后纳入 199 篇相关文献。通过以 Refworks 格式从知网中导出文献信息,以“download_xx”格式对文件命名,在 CiteSpace 将数据格式转化为 WOS 格式。

英文文献主要选自 Web of Science Core Collection 数据库,构建检索式为 Topic=(" machine learn* " OR " deep-learn* " OR " convolutional neural network* " OR " recurrent neural network* " OR " CNN* " OR " Decision tree learn* " OR " Association rule* ")AND Topic=("

electronic medical record* " OR " electronic patient record* " OR " EMR ")AND (Language=English)AND (Document Type=Article OR Review),时间范围限制为 2000 年至 2022 年,检索时间为 2022 年 7 月 15 日,共检索到 2361 篇文献。通过浏览标题及摘要部分等剔除不相关文献,最后筛选出 2 135 篇相关文献。最后,通过以纯文本文件格式从 WOS 中导出文献全记录与引用的参考文献,将所有文献导入 CiteSpace 进行可视化分析。

1.2 研究方法

本文采用陈超美博士及其团队研发的基于 Java 语言开发的计量学软件 CiteSpace,可对当前学科领域研究现状、研究热点和未来发展势等情况进行可视化分析^[6]。在 CiteSpace 软件中,对相关的参数进行设置。因中文文献和英文文献的发文年份不同,在时间跨度上的设置不同。在中文文献的分析中,时间跨度设置为 2007 年 1 月至 2022 年 7 月,时间分段选择为 1 年;在英文文献的分析中,时间跨度设置为 2006 年 1 月至 2022 年 7 月,时间分段设置为 1 年。文本运行框的设置选择为文章标题、摘要、作者关键词、附加关键词等;节点类型框的设置分别是国家、作者、机构、关键词等;在修剪框中选择关键路径和修剪切片网络,对文献进行可视化分析,以科学知识图谱形式展现相关研究结果。

2 国内外研究现状分析

2.1 年发文趋势分析

利用 WPS 对国内外年发文量数据进行制图并对比,如图 1 所示。

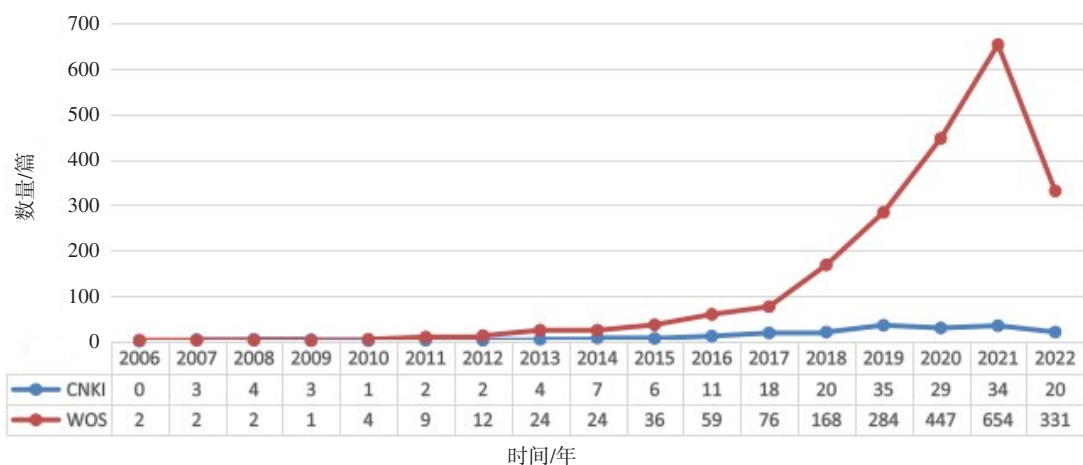


图 1 2006~2022 年国内外机器学习在电子病历中应用的年发文量变化趋势
Figure 1 Trends in annual number of published papers on machine learning application in electronic medical record at home and abroad from 2006 to 2022

2000~2005 年无机器学习在电子病历应用相关文献。从整体发文量来看,中文文献较少;2015 年之

前,每年发文量小于 10 篇;近 5 年每年发文数量少于 50 篇,整体呈现平稳的趋势,说明国内关于机器学习

在电子病历中应用的研究活跃度相对低。相比之下,2006~2022年机器学习在电子病历中应用的英文文献数量多于中文文献,尤其在2017年之后差距对比明显。2006~2016年间,英文文献数量呈缓慢增长趋势;2017年 AlphaGo 打败世界围棋冠军柯洁之后^[7],深度学习和神经网络算法成为学术界和业界关注焦点,促进机器学习的研究发文量快速增长,自2017年起机器学习在电子病历中应用的英文文献迅猛增长。

2.2 国家/地区合作分析

在 CiteSpace 中选择“country”节点对国内外文献进行国家/地区合作网络图谱分析,如图 2 所示。结果表明,世界各国之间合作发文较少;发文最多的是美国(1 149 篇),其次是中国(339 篇),第 3 是英国(145 篇),可见美国、中国和英国在机器学习应用于电子病历研究中发文量占据巨大优势;中心性较高的国家是葡萄牙(0.68)、西班牙(0.56)和法国(0.55),它们所在的合作团体内部成员各国之间合作密切,体现了这些国家在该领域的影响力,在该领域起到链接其他国家的重要作用。在国际上,虽然我国在该领域发文量排名第 2,但中心性为 0,不占据重要地位,说明我国英文文献在国际合作及内容深度上需进一步加强,未来应该寻求更多的国际合作机会,从而提升中国在国际上的学术影响力。

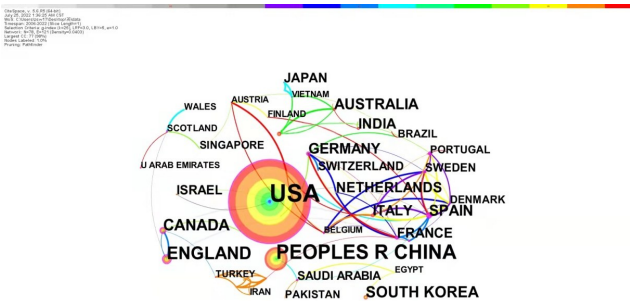


图 2 国内外机器学习在电子病历应用的国家/地区合作网络
Figure 2 Network on the cooperation of countries/regions on machine learning application in electronic medical record at home and abroad

2.3 国内外作者合作分析

在 CiteSpace 中选择“author”节点对国内外文献进行作者合作网络图谱分析,如图 3 和图 4 所示。

结果表明,国内发文量最多的是来自哈尔滨工业大学的关毅教授,其主要在利用机器学习方法在电子病历实体识别、实体关系抽取与智能诊断模型研究方面做了大量工作^[8-9];发文量排名第 2 的是南通大学的丁卫平教授,其主要研究利用支持向量机和聚类方法以提高电子病历挖掘性能和疾病诊断研究^[10-11];发文量排

名第 3 的是来自青岛科技大学的冯云霞副教授,其主要将关联规则和决策树算法利用在肺癌诊断及其致病因素的研究^[12-13]。其中合作关系较为紧密的科研团队为厦门市妇幼保健院吴谨淮为代表的研究团队、厦门大学附属第一医院的赵敏为代表的研究团队和无锡市人民政府的陈卫平为代表的研究团队,其他学者相对比较分散,相互之间协作关系较少,详见图 3。

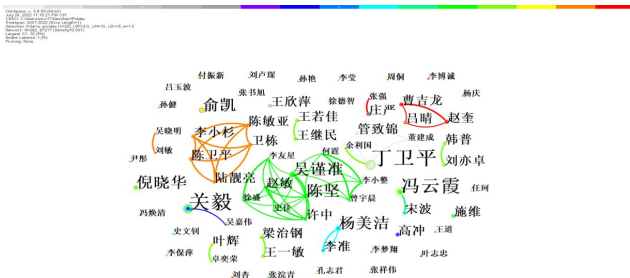


图 3 国内作者合作网络
Figure 3 Network on the cooperation of authors at home

同样地,国外学者之间合作网络关系也较为松散。发文量最多的是来自英国剑桥大学的 RITANKAR DAS,其利用多种神经网络、支持向量机和决策树等方法在疾病预测方面做了大量回顾性研究工作^[14-15];排名第 2 的是来自林肯大学的 JANA HOFFMAN, JANA HOFFMAN 和 RITANKAR DAS 存在多篇合作关系, JANA HOFFMAN 同样利用多种机器学习方法在疾病预测方面发表论文数篇,研究疾病较广,主要涉及急性肾损伤、脓毒症、胰腺炎和骨髓增生等多种疾病^[16-18]。排名第 3 的是来自美国梅奥诊所的 HONGFANG LIU,其主要利用自然语言处理技术结合机器学习方法在临床文本和影像资料应用中进行相关研究^[19-20]。从发文量较多的作者单位看,他们主要来自于西方发达国家。其中,以 RITANKAR DAS、JANA HOFFMAN、HONGFANG LIU、FEI WANG 和 NIGAM H SHAH 等为代表的团队内部合作紧密,与其他作者能看到较为清晰的连线,详见图 4。

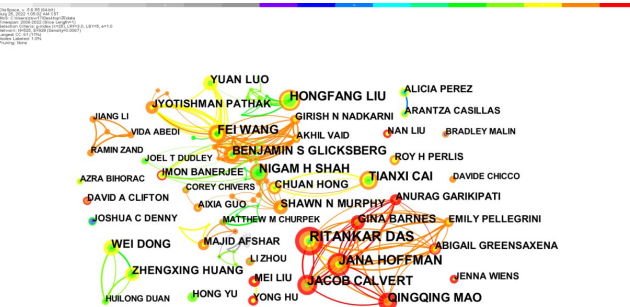


图 4 国外作者合作网络
Figure 4 Network on the cooperation of authors at abroad

中文文献中高频关键词包括“电子病历”“深度学习”“数据挖掘”“机器学习”“关联规则”“人工智能”等,说明利用机器学习方法实现在电子病历中进行数据挖掘一直是国内的研究热点。英文文献中最大的节点关键词是“machine learning(机器学习)”,其出现的频次为750次,此外高频关键词还包括“electronic medical record(电子病历)”、“risk(风险)”、“deep learning(深度学习)”、“classification(分类)”、“mortality(死亡率)”、“prediction(预测)”等,也是该领域关注的热点主题。

通过对国内外关键词对比分析,国内外研究热点既有相同之处也有不同之处,主要体现在以下几个方面。

3.1.1 国内外研究的相同之处 第一,国内外均关注辅助诊断研究。对于电子病历系统中的文本内容和医学影像资料进行多维度分析,提取相应的疾病标志特征,经过高复杂性的计算,帮助医生对疾病进行高效预判和诊断,从而给出相应治疗方案建议^[21]。对海量医学知识智能计算和分析,减小因医生经验及知识水平的受限和影响造成医生误诊和漏诊概率,提高日常诊疗效率,更能节约相关医疗费用和资源。国内外研究者对病历文本进行知识抽取、语义分析,对医学图像的分类、检测、分割和智能分析等方面做了大量相关工作^[22]。舒琛洁等^[23]依据“气虚”特征筛选出15种症状,按照7:3设定训练集和测试集比例,将随机森林、BP神经网络、CART决策树、支持向量机和K近邻5种模型应用在构建证候要素“气虚”的辅助诊断模型中,经过对比发现,基于随机森林算法构建模型的分类效果最好,AUC值达到0.923。Ning等^[24]提出了结合使用softmax分类器和XGBoost分类器对肺部结节的结构化数据和非结构化数据的融合特征来提高肺结节分类性能的问题。研究结果显示,在引入结构化数据的情况下,模型的性能均有显著提高;最佳准确率、灵敏度、特异性和ROC曲线下面积(AUC值)分别为0.936、0.919、0.956和0.971。融合结构化和非结构化数据可以发现更多的患者信息,为临床诊疗过程提供更好的决策支持。国内外共同关注的关键词包括:辅助诊断、临床路径、disease(疾病)、diagnosis(诊断)等。

第二,国内外均关注医疗活动中预测事件的研究。对于未发情况预测是实施医疗干预的前提,对降低医疗资源消耗、减少人民病痛和防止病情恶化具有重大意义。国内外现有研究中主要包括几大方面的预测:疾病预测、死亡率预测、再入院预测、预后预测、药物疗效及不良反应预测、产后出血预测等^[25-28]。研究对象主要是患有高风险疾病或重症疾

病的患者,例如:ICU重症患者、急性肾损伤患者、心梗患者、癌症患者等。随着人工智能技术的不断发展,借助机器学习和深度学习等方法在预测模型中的预测准确度和预测性能也越来越高。Liu等^[27]采用逻辑回归模型、随机森林、K近邻及Lightgbm和逻辑回归集成的模型对产后出血的临床危险因素指标进行预测,研究结果表明Lightgbm和逻辑回归集成的模型中预测产后出血效果最好,结合子宫收缩特征和2h内出血量的机器学习模型可以进一步提高阴道分娩中产后出血预测的准确性,为临床医生早期干预和减少不良妊娠结局提供了重要参考。国内外共同关注的关键词包括:疾病预测、risk(风险)、mortality(死亡率)、prediction(预测)、classification(分类)等。

第三,国内外均关注技术方法研究。最初应用在电子病历领域的机器学习模型有贝叶斯学习、关联规则、随机森林、决策树和人工神经网络等^[28]。但随着医疗健康领域对于人工智能的需求越来越迫切,深度学习技术被引入。深度学习是机器学习方法的一个分支领域,它专注于使用数据的深度层次模型进行学习。近年来计算机处理能力的巨大提高,深度学习在高维数据中的复杂结构挖掘任务中表现效果理想,在自然语言处理、语音识别、预测基因表达、药物治疗和图像分割等方面完胜于传统的机器学习算法^[29]。例如,在口腔诊疗中,相关研究证明对于高分辨率侧位片、全景片和电子计算机断层扫描等医学图像的智能定点测量中,深度神经网络比传统的随机森林等检测更准确^[30]。利用深度学习算法在人工智能、大数据领域等实现更深层的数据挖掘研究也引起国内外学者广泛关注。国内外共同关注的关键词包括:电子病历(electronic health record)、机器学习(machine learning)、深度学习(deep learning)、数据挖掘、人工智能(artificial intelligence)、大数据(big data)、神经网络(neural network)。

3.1.2 国内外研究的不同之处 国内研究主要关注热点体现在:第一,电子病历中的信息抽取的相关任务和方法研究。信息抽取的两大主要任务包括:生物医学命名实体识别(也叫实体抽取)和实体关系抽取。命名实体识别即对电子病历中非结构化的自由文本进行特定概念识别,如治疗手段、症状、药物、实验室检验、疾病等^[31]。研究中常用的命名实体识别方法包括基于规则与词典的方法,支持向量机、马尔可夫模型、长短期记忆网络和条件随机场的统计机器学习方法以及循环神经网络、双向长短期记忆网络和BERT模型的深度学习方法^[32]。命名实体识别

为后续的电子病历中不同专业概念之间存在的语义关系抽取起到重要作用,是关系抽取的前提。命名实体识别和实体关系抽取作为自然语言处理的基础任务,为知识图谱、语料库构建和数据挖掘等相关任务奠定基础。国内关注的关键词包括:信息抽取、实体抽取、知识图谱。第二,电子病历利用过程中患者信息的隐私保护问题。电子病历记载着患者个人姓名、出生年月、性别、家庭地址、联系方式、临床信息和治疗记录,这些数据非常敏感,这给机器学习和深度学习在电子病历领域的应用和发展带来了挑战。由于患者数据的私密性,它不仅涉及道德和法律问题,还涉及技术问题。如何在保护患者隐私数据的前提下,充分利用大数据的优势,促进机器学习和深度学习的发展,是人工智能在电子病历领域必须考虑的问题。程健一等^[33]提出支持向量机和条件随机场结合进行双分类处理的去隐私化方法,结果准确率达到0.94,召回率达到0.88,F值达到0.9110,卓具实效。国内关注的关键词包括:隐私保护。

国外研究主要关注热点体现在:第一,精准医疗与个性化医疗发展的研究。精准医疗(precision medicine)是充分利用个人疾病的基因组学、转录组学、代谢组学和蛋白质组学等信息,对疾病进行精细分类及精确诊断,以确定与疾病进展相关的机制,并将针对患者进行制定个性化治疗方案和用药建议,达到最佳治疗效果和医疗资源耗费最小化,核心在于疾病重新“分类”基础上的“对症下药”^[34-35]。美国政府的“精准医学倡议”和英国政府的“十万基因组计划”等,掀起全球精准医疗应用和研究热潮。电子病历作为医疗健康大数据组成之一,其中的患者基本信息、基本健康信息、门(急)诊病历记录、住院病历记录和健康体检记录等数据是实现精准医疗的有力支撑^[36]。国外研究人员专注于将基因组学、表观基因组学、代谢组学、生物信息学和医学图像进行融合,同时将这些数据与数据挖掘、深度学习和大数据方法等其他技术结合应用,研究疾病预测与诊断、个性化治疗方案制定、靶向治疗和智能可穿戴设备进行健康管理等^[37]。Sui等^[38]提出一种基于深度学习的肺癌放射基因组学分析图像生物标志物和基因表达数据相关框架,研究中构建了从肺部肿瘤图像到基因组数据的链接并依次实现生成过程,形成了一个双向框架映射多源医疗数据,它比传统的放射基因组学方法能获得更多相关的特征,研究使用生成对抗网络将基因组数据转化为肿瘤图像,证明该框架可以更有效地进行肺癌放射基因组研究。国外研究中主要关键词包括:care(照顾)、management(管理)和health(健康)。第二,临床应用系统或平台研

究。从关键词图谱信息可知,临床医生或研究者对于电子病历在实际应用中需求不尽相同。无论是疾病预测与诊断,还是药物警戒与新药研发,围绕临床实际需求搭建高效的系统工具或平台辅助工作将能达到事半功倍效果,是国外学者关注的重点之一。López-Martínez团队^[39]设计了基于机器学习和数据集成原理的智能大数据分析平台,该平台利用机器学习算法将电子健康记录、医院信息系统、放射学信息系统和实验室信息系统、公共卫生平台、社交媒体和临床网站门户网站生成的数据进行整合,集成大型标准化数据集,以提高公共卫生干预措施的有效性,改善诊断和临床决策支持。名为“MedAware”的分析平台被开发,用于检测医疗处方中的错误和临床错误,实时减少住院和再入院^[40]。国外研究中主要关键词包括:system(系统)。

3.2 关键词突现分析

关键词突现度可以反应在某段时间节点内的研究趋势和前沿动态的发展变化情况。应用CiteSpace软件在关键词共现的基础上绘制国内外文献关键词突现图谱,见图7和图8。

通过分析国内文献关键词突变图谱发现,国内研究大致分为3个阶段。第一阶段为2007~2012年。从影响周期来看,“关联规则”和“数据挖掘”研究持续最长,持续9年。早期阶段更侧重于理论基础研究,主要关注机器学习算法和数据挖掘方法的研究。第二阶段为2013~2019年。这一阶段,机器学习和深度学习技术得到发展,在电子病历中应用广泛。该阶段更加关注医疗信息中的实体抽取任务,应用于临床诊断和患者隐私保护研究,最受关注的疾病为糖尿病和高血压。第三阶段为2020~2022年。该阶段的突出特点是对人工智能需求更加迫切,主要关注信息提取和知识图谱构建工作,研究并发症和疾病预测。

通过分析国外文献关键词突变图谱发现,国外研究大致分为两个阶段。第一阶段为2010~2016年。该阶段中国外文文献中突现词主要分为3类,电子病历及文本类:medical record、electronic medical record、health record、architecture、text、information;技术类:support vector machine、data mining、extraction、text mining、big data;应用类:clinical decision support、system、Pharmacovigilance、phenotyping、regression。该阶段处于早期探索期,国外学者将相关理论方法和实践进行结合,在电子病历、技术和应用3方面研究产生了许多成果。第二阶段为2017~2022年。第一阶段的部分突现词的热度延续至第二阶段,如Pharmacovigilance、extraction等。这说明药物警戒和

文本的信息抽取依然是热点主题,另外的焦点还有关于词嵌入、循环神经网络和新型冠状病毒肺炎的研究。



图7 国内关键词突现分析图谱
Figure 7 Analysis on keywords with the strongest citation burst at home

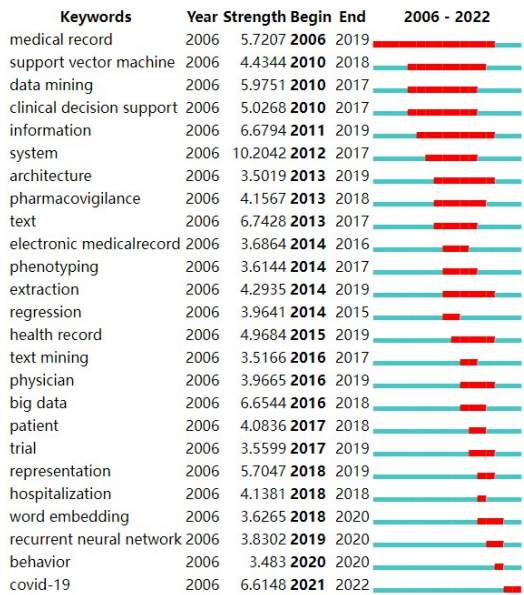


图8 国外关键词突现分析图谱
Figure 8 Analysis on keywords with the strongest citation burst at abroad

4 结 论

- (1)从发文量上看,中文文献相对较少,国外文献较多,在2017年之后国外文献迅猛增长,表明机器学习在电子病历中的应用一直是研究热点。
- (2)美国发文最多,然后是中国和英国;葡萄牙

在该领域中心性最高;我国在该领域影响力不足,我国学者应该主动寻求更多与其他国家学者合作的机会。

(3)国内外作者合作网络松散,仅存在少数团队合作关系。国内发文较多的是关毅、丁卫平和冯云霞;国外发文较多的是 RITANKAR DAS、JANA HOFFMAN和HONGFANG LIU。

(4)国内高校的信息管理学院和信息研究所、计算机研究院校及各高校附属医院为主要研究机构。但国内机构大部分处于各自独立研究状态,仅存在少数合作关系,应加强跨地区、跨学科、跨机构及跨国家的合作;国外机构之间合作更紧密,在跨机构和跨学科等方面表现更活跃。

(5)在研究热点上,国内外共同关注辅助诊断、预测事件和方法技术方面的研究。国内更关注信息抽取的相关任务和方法、患者信息的隐私保护问题。国外更关注精准医疗与个性化医疗发展的研究和临床应用系统或平台研究。

(6)人工智能、信息提取、知识图谱、并发症和疾病预测是国内研究的前沿。药物警戒、信息抽取、词嵌入、循环神经网络和新型冠状病毒肺炎等是国外近几年在该领域持续研究的重点。

本文仅选取WOS数据库及中国知网作为数据来源,文献涵盖面不够广。在未来研究中,应扩大研究的数据来源;另外应结合文献的引文、共被引分析等,才能更好地把握国内外机器学习在电子病历领域的研究进展与发展趋势。

【参考文献】

[1] 中华人民共和国中央人民政府. 卫生部关于印发《电子病历基本规范(试行)》的通知[EB/OL]. [2010-03-04/2022-07-18]. http://www.gov.cn/jzwgk/2010-03/04/content_1547432.htm.
The Central People's Government of the People's Republic of China. Notice of the Ministry of Health on the issuance of the Basic Specifications for Electronic Medical Records (Trial)[EB/OL]. [2010-03-04/2022-07-18]. http://www.gov.cn/jzwgk/2010-03/04/content_1547432.htm.

[2] Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health[J]. Ann Rev Public Health, 2015, 36: 345-359.

[3] 张景奇, 史文宝, 纪秀娟. 机器学习在医疗和公共卫生中应用[J]. 中国公共卫生, 2019, 35(10): 1449-1452.
Zhang JQ, Shi WH, Ji XJ. Application of machine learning in medical and public health researches: a review[J]. Chinese Journal of Public Health, 2019, 35(10): 1449-1452.

[4] Deo RC. Machine learning in medicine[J]. Circulation, 2015, 132(20): 1920-1930.

[5] Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics[J]. IEEE J Biomed Health Informatics, 2017, 21(1): 4-21.

[6] Chen CM, Chen Y, Horowitz M, et al. Towards an explanatory and computational theory of scientific discovery[J]. J Informetrics, 2009, 3(3): 191-209.

[7] 百度百科. 阿尔法围棋[EB/OL]. [2022-06-16/2022-9-18]. <https://baike.baidu.com/item/%E9%98%BF%E5%B0%94%E6%B3%95%>

- E5%9B%B4%E6%A3%8B/19319610?fr=kg_general.
Baidu Wikipedia. AlphaGo [EB/OL]. [2022-06-16/2022-9-18]. https://baike.baidu.com/item/%E9%98%BF%E5%B0%94%E6%B3%95%E5%9B%B4%E6%A3%8B/19319610?fr=kg_general.
- [8] 胡金鹏, 关毅. 面向呼吸内科智能诊断模型研究[J]. 智能计算机与应用, 2019, 9(5): 154-157.
Hu JP, Guan Y. Research on intelligent diagnosis model for respiratory medicine [J]. Intelligent Computer and Applications, 2019, 9(5): 154-157.
- [9] 王润奇, 关毅. 基于 Tri-Training 算法的中文电子病历实体识别研究[J]. 智能计算机与应用, 2017, 7(6): 132-134.
Wang RQ, Guan Y. Named entity recognition research in Chinese electronic medical records based on Tri-Training algorithm [J]. Intelligent Computer and Applications, 2017, 7(6): 132-134.
- [10] 沐燕舟, 丁卫平, 高峰, 等. 基于自适应 PSO 的改进 K-means 算法及其在电子病历聚类分析应用[J]. 计算机与数字工程, 2019, 47(8): 1861-1865.
Mu YZ, Ding WP, Gao F, et al. An improved K-means algorithm based on adaptive PSO and its application in cluster analysis of electronic medical records [J]. Computer & Digital Engineering, 2019, 47(8): 1861-1865.
- [11] 张琼, 丁卫平, 景炜, 等. 基于改进 PSO-SVM 算法的帕金森疾病诊断研究[J]. 计算机与数字工程, 2019, 47(8): 1851-1855.
Zhang Q, Ding WP, Jing W, et al. Research of Parkinson's disease diagnosis based on improved PSO-SVM algorithm [J]. Computer & Digital Engineering, 2019, 47(8): 1851-1855.
- [12] 冯云霞, 张润. 基于电子病历的肺癌诊断决策树算法[J]. 计算机系统应用, 2019, 28(10): 257-263.
Feng YX, Zhang R. Decision tree algorithms for lung cancer diagnosis based on electronic medical record [J]. Computer Systems & Applications, 2019, 28(10): 257-263.
- [13] 张润, 冯云霞. 基于改进 Apriori 算法的肺癌致病因素研究[J]. 计算机技术与发展, 2020, 30(2): 143-147.
Zhang R, Feng YX. Research on pathogenic factors of lung cancer based on improved Apriori algorithm [J]. Computer Technology and Development, 2020, 30(2): 143-147.
- [14] Sidney L, Emily P, Abigail G, et al. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS) [J]. J Crit Care, 2020, 60: 96-102.
- [15] Christopher B, Uli C, Yifan Z, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs [J]. Comput Biol Med, 2019, 109: 79-84.
- [16] Le S, Allen A, Calvert J, et al. Convolutional neural network model for intensive care unit acute kidney injury prediction [J]. Kidney Int Rep, 2021, 6(5): 1289-1298.
- [17] Ashwath R, Anurag G, Zohora I, et al. A machine learning approach to predicting risk of myelodysplastic syndrome [J]. Leuk Res, 2021, 109: 106639.
- [18] Rahul T, Zohora I, Anurag G, et al. Early prediction of severe acute pancreatitis using machine learning [J]. Pancreatolgy, 2022, 22(1): 43-50.
- [19] Wang Y, Zhao Y, Therneau TM, et al. Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records [J]. J Biomed Inform, 2020, 102: 103364.
- [20] Sunyang F, Bjoerg T, Xin Z, et al. A hybrid model to identify fall occurrence from electronic health records [J]. Int J Med Informatics, 2022, 162: 104736.
- [21] 刘新奎, 杨林朋, 管红英, 等. 基于疾病诊断相关分组的医学知识图谱构建方法[J]. 医学信息学杂志, 2020, 41(9): 27-30.
Liu XK, Yang LP, Jiu HY, et al. Building method of medical knowledge graph library based on diagnosis-related group (DRG-MKG) [J]. Journal of Medical Med Informatics, 2020, 41(9): 27-30.
- [22] 杨兆凯, 王龙, 陈金栋. 人工智能与深度学习在医学影像辅助诊断中的应用[J]. 电脑知识与技术, 2021, 17(35): 91-93.
Yang ZK, Wang L, Chen JD. Application of artificial intelligence and deep learning in medical image-assisted diagnosis [J]. Computer Knowledge and Technology, 2021, 17(35): 91-93.
- [23] 舒琛洁, 梁浩, 刘淑明, 等. 机器学习算法对证候要素“气虚”辅助诊断的性能评估[J]. 北京中医药大学学报, 2021, 44(10): 928-934.
Shu CJ, Liang H, Liu SM, et al. Performance evaluation of machine learning algorithms on the pattern elements of "qi deficiency" assisted diagnosis [J]. Journal of Beijing University of Traditional Chinese Medicine, 2021, 44(10): 928-934.
- [24] Ning TR, Zhang ZL, Wei XC, et al. Improving the performance of lung nodule classification by fusing structured and unstructured data [J]. Inform Fusion, 2022, 88: 161-174.
- [25] Lee Y, Kwon JM, Lee Y, et al. Deep learning in the medical domain: predicting cardiac arrest using deep learning [J]. Acute Crit Care, 2018, 33(3): 117-120.
- [26] Chi CY, Ao S, Winkler A, et al. Predicting the mortality and readmission of in-hospital cardiac arrest patients with electronic health records: a machine learning approach [J]. J Med Internet Res, 2021, 23(9): e27798.
- [27] Liu J, Wang C, Yan R, et al. Machine learning-based prediction of postpartum hemorrhage after vaginal delivery: combining bleeding high risk factors and uterine contraction curve [J]. Arch Gynecol Obstet, 2022, 306: 1015-1025.
- [28] Shamout F, Zhu T, Clifton DA. Machine learning for clinical outcome prediction [J]. IEEE Rev Biomed Engin, 2021, 14: 116-126.
- [29] Shickel B, Tighe PJ, Bihorac A, et al. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis [J]. IEEE J Biomed Health Inform, 2018, 22(5): 1589-1604.
- [30] Yang S, Zhu F, Ling X, et al. Intelligent health care: applications of deep learning in computational medicine [J]. Front Genet, 2021, 12: 607471.
- [31] 姜培, 方安, 赵琬清, 等. 电子病历信息抽取可视化分析[J]. 医学信息学杂志, 2021, 42(4): 35-40.
Lou P, Fang A, Zhao WQ, et al. Visualization analysis of electronic medical record information extraction [J]. Journal of Medical Informatics, 2021, 42(4): 35-40.
- [32] 崔博文, 金涛, 王建民. 自由文本电子病历信息抽取综述[J]. 计算机应用, 2021, 41(4): 1055-1063.
Cui BW, Jin T, Wang JM. Overview of information extraction of free-text electronic medical records [J]. Journal of Computer Applications, 2021, 41(4): 1055-1063.
- [33] 程健一, 关毅, 何彬. 基于 SVM 和 CRF 双层分类器的英文电子病历去隐私化[J]. 智能计算机与应用, 2016, 6(6): 17-19.
Cheng JY, Guan Y, He B. De-privacy of English electronic medical records based on SVM and CRF two-layer classifier [J]. Intelligent Computer and Applications, 2016, 6(6): 17-19.
- [34] National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease [M]. Washington (DC): National Academies Press (US), 2011.
- [35] Hulsen T, Jamuar S, Moody A, et al. From big data to precision medicine [J]. Front Med, 2019(6): 1-14.
- [36] 刘宁, 武琼, 陈敏. 个性化医疗服务类型及相关数据资源研究[J]. 中国卫生信息管理杂志, 2016, 13(1): 93-98.
Liu N, Wu Q, Chen M. Service types and data resources research on personalized medicine [J]. Chinese Journal of Health Informatics and Management, 2016, 13(1): 93-98.
- [37] Kim IW, Oh JM. Deep learning: from chemo informatics to precision medicine [J]. J Pharm Invest, 2017, 47: 317-323.
- [38] Sui D, Guo M, Ma X, et al. Imaging biomarkers and gene expression data correlation framework for lung cancer radio genomics analysis based on deep learning [J]. IEEE Access, 2021, 9: 125247-125257.
- [39] López-Martínez F, Núñez-Valdez ER., García-Díaz V, et al. A case study for a big data and machine learning platform to improve medical decision support in population health management [J]. Algorithms, 2020, 13: 102.
- [40] Digital Innovation and Transformation. MedAware-using AI to eliminate prescription errors [EB/OL]. [2018-04-09/2022-9-18]. <https://digital.hbs.edu/platform-digit/submission/medaware-using-ai-to-eliminate-prescription-errors/>.

(编辑:黄开颜)