DOI: 10.11918/j.issn.0367-6234.201806001

基于 EHR 的医疗知识图谱研究与应用综述

何 霆[1],吴雅婷[1],王华珍[1],熊英杰[1],孙 偲[1],徐汉川[2]

(1.华侨大学 计算机科学与技术学院,厦门 361021; 2.哈尔滨工业大学 计算机科学与技术学院,哈尔滨 150001)

摘 要: 电子健康记录(EHR)作为一种医疗信息化手段,在数十年的使用过程中储存和积累了越来越多的医疗过程和结果大数据. 知识图谱作为一种从海量数据中抽取结构化知识的手段,近年来在多个行业展示了广阔的应用前景. 知识图谱的优势在于对海量、异构的数据进行组织,完成知识推理. 知识图谱适用于自然语言的分析,有助于在以自然语言形式存在的海量 EHR 数据中获得宝贵的医疗知识和医疗经验. EHR 分析研究的价值主要集中在辅助诊断、辅助治疗和疾病预测. 利用大量的 EHR 数据构建医疗知识图谱,当新的患者数据来临之时,知识图谱可以发挥查询扩展、临床决策支持和疾病预测等作用. 本文首先简要介绍了 EHR 的发展现状,以及现有著名的 EHR 数据集及其应用成果. 其次,在概括介绍知识图谱发展总体现状基础上,分析了知识图谱在医疗领域的发展趋势和热点迁移. 然后,对基于 EHR 的医疗知识图谱研究与应用进展进行了比较全面的总结,包括 EHR 的信息抽取、数据整合、查询扩展、临床决策支持和疾病预测等. 最后,对该领域未来发展方向和面临的挑战作了展望.

关键词: EHR; 知识图谱; 医疗知识图谱; 研究和应用进展

中图分类号: TP182 **文献标志码:** A **文章编号:** 0367-6234(2018)11-0137-08

A survey of medical knowledge graph based on EHR

HE Ting[1], WU Yating[1], WANG Huazhen[1], XIONG Yingjie[1], SUN Cai[1], XU Hanchuan[2]

(1.School of Computer Science and Technology, Huaqiao University, Xiamen 361021, China;
2.School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Electronic health record (EHR) as a means of medical informatization has stored and accumulated big data of medical processes and results. Knowledge graph as a means to extract structured knowledge from massive data has recently shown broad application prospects in many industries. The application of knowledge graph technology to the analysis of EHR data can help obtain valuable medical knowledge and experience which could be used to improve the effectiveness and efficiency of medical industry significantly. This paper firstly introduced the development status of EHR, the existing famous EHR data sets and their application results. Secondly, on the basis of summarizing the current developments of knowledge graph, it analyzed the trend and hot-spot migration of knowledge graph in medical field as well. Then, it summarized the research and application progresses of EHR based medical knowledge graph which include the information extraction, data integration, query expansion, clinical diagnostic support and disease prediction of EHR, etc. Finally, this paper also gave the future development directions and challenges of EHR based medical knowledge graph.

Keywords: EHR; knowledge graph; medical knowledge graph; advances in research and application

一直以来,世界范围内的看病难、看病贵现象是社会关注的焦点. 医疗行业资源总量不足、分布不合理、优质资源匮乏、医师培养周期长,导致医疗服务质量不平衡、人才匮乏,不能满足患者的医疗需求,医患矛盾突出. 电子健康记录(Electronic Health Record,EHR)的使用,在一定程度上提高了医师的工作效率,使医师有更多的时间提高医疗技术水平,减少医患矛盾的发生. 作为临床医疗的信息载体,EHR 引起行业内研究者的广泛兴趣,取得了一定的成果. 然而,目前对 EHR 的研究还停留在传统的存储管理、整理与统计分析、以及部分较高层次的挖掘分析等方面,深层次应用方面的突破性进展较少. 知识图谱(Knowledge Graph,KG)技术作为一种从海量数据中抽取结构化知识的手段,近年来在多个行业展示了广阔的应用前景. 将知识图谱应用于 EHR 数据的分析,有助于获得存在于海量 EHR 数据中宝贵的医疗知识和医疗经验,并进行广泛的共享. 基于此,本文总结了基于 EHR 的医疗知识图谱研究与应用进展,希望能够对该领域的发展起到一定的推动作用.

收稿日期: 2018-06-01
基金项目: 国家自然科学基金(71571056); 华侨大学科研启动基金(16BS304)
作者简介: 何 霆(1972—),男,博士,教授,博士生导师
通信作者: 何 霆,xuantinghe@hit.edu.cn

1　电子健康记录(EHR)

美国卫生组织卫生标准 7(Health Level Seven,HL-7)对 EHR 归纳如下:"EHR 是向每个个人提供的、一份具有安全保密性的、记录其在卫生体系中关于健康历史与服务的终身档案[1]."以 20 世纪 60 年代 Lockheed 开发临床信息系统为起点,EHR 开始发展. 早期 EHR 是通过数码拍照或扫描方式将整本病案输入的计算机化病案,后来发展成为包含病人整个医疗过程的电子病历(electronic medical record,EMR),再后来发展成为内涵更为广泛的 EHR. 20 世纪 90 年代,随着对电子病历系统化研究的日益深入,EHR 的发展进入快车道[2]. 图 1 显示 Google Scholar 上 1960 年以来每十年间有关"EHR"研究和应用成果出版文献的数量.

如图 1,1960~1990 年间,EHR 并没有受到广泛的重视. 但是从 20 世纪 90 年代开始,EHR 的研究得到了突飞猛进的发展,说明 EHR 的研究越来越受到重视. 这些研究主要涉及电子健康记录的构建[3]、规范化[4]、存储[5-6]、管理[7-8]和挖掘[9-10],为改进医疗信息化、保险业和护理实践水平,提高效率,降低失误产生了贡献.

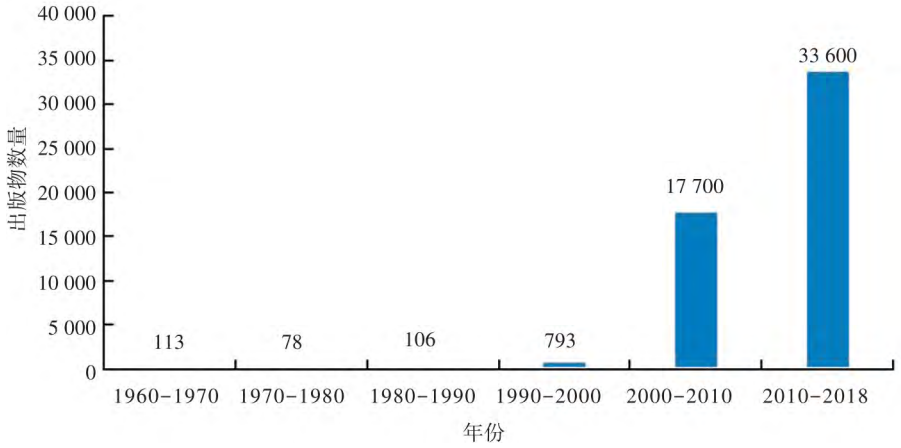

图 1　1960 年以来 EHR 相关出版文献数量

Fig.1　Publications of "EHR" since 1960

由于涉及到信息安全领域的法律限制以及公民隐私等方面的问题,EHR 数据本身的不可共享特性成为了其研究和应用发展的一大掣肘. 几十年来,虽然世界范围内的众多医疗机构积累了海量的 EHR 数据,但能够公开被人们研究和应用的数据集寥寥无几. 目前国内外广泛使用的 EHR 公开数据集如下:

1) PDD graph:该数据集包含了从大量 EHR 中提取出的医学知识,帮助医护人员准确掌握症状、诊断和治疗之间的关系. Wang 等利用该数据集研究 EHR 信息抽取[11].

2) MIMIC-III:记录 2001 和 2012 之间,Beth Israel Deaconess Medical Center 重症监护病房病历,包括诸如人口统计学、生命体征测量、影像报告和院外死亡率等信息[12]. Datla 等在该数据集基础上开发自动化临床诊断系统[13],Goodwin 等用此数据集进行 EHR 信息抽取中的知识嵌入问题的研究[14].

3) i2b2:是一个医疗领域综合数据集,包括吸烟、肥胖、药物治疗、临床叙事时间关系、心脏病数据集,供研究者使用. 通过此数据集,Harabagiu 研究 EHR 信息抽取和查询扩展问题[15,16],Ling Y 开发临床决策支持系统,进行诊断推断[17].

4) UMLS:是汇集健康与生物方面的专业词汇和标准的统一医学语言系统,解决了在不同医疗软件系统之间医疗术语消歧问题,用于确定 EHR 术语的概念映射. 利用此数据集,Hajihashemi[18] 进行早期疾病的判断,Kang BY[19] 等研究 EHR 术语的概念映射问题.

5) UCI:该网站存放糖尿病病人记录,可以从中得到糖尿病人一日三餐前后血糖含量和注射的胰岛素剂量信息. Overby C[20] 在此基础上进行药物不良反应的识别.

2　知识图谱

知识图谱是一种基于图的数据结构,由节点(实体)和边(实体间的关系)组成,实质是一个关系网络[21],2012 年由 Google 公司正式提出. 图 2 显示 Google Scholar 上 2012-2017 年度"知识图谱"公开出版文献数量.

从图 2 中可以看出,知识图谱出版文献的数量从 2012 年的 456 篇增长到 2017 年的2 630 篇,平均年增长率 41.98%. 知识图谱的研究历经了从语义网到知识库,再到知识图谱的发展脉络,主要包括以 Freebase、Wikidata、DBpedia 和 YAGO 为代表的横向

开放知识图谱和类似 IMDB、MusicBrainz 这样的垂直行业知识图谱，核心技术主要有知识抽取、知识表示、知识融合、知识推理 4 大方面[22]，应用领域包括但不局限于机器翻译、问答系统、情报分析等.

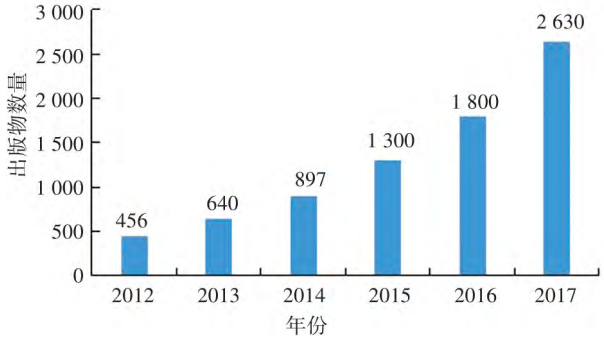

图 2　2012–2017 年间知识图谱出版文献数量

Fig.2　Publications of KG between 2012 and 2017

知识图谱适用于自然语言的分析，医疗领域中 EHR 信息作为自然语言的一种，非常适合采用知识图谱做分析. 近六年来，知识图谱在医疗健康领域的研究取得了一定进展. 以“knowledge graph AND medical”为关键词在 GoogleScholar 上搜索 2012～2017 年的研究报告(如图 3)，统计知识图谱在医疗健康领域研究的出版文献数量. 虽然这些年来相关文献的年均增长率较高(48.26%)，但从绝对数量上来看还较少，因此存在广阔的发展空间.

在医疗健康领域，国外著名且已经比较成熟的医疗知识库有 Gene Ontology、Drug-Bank、UMLS 等，而国内的研究都还处于起步阶段. 中文医疗知识图谱的构建研究显得十分迫切和必要，是我国实现智能医疗急需突破的瓶颈.

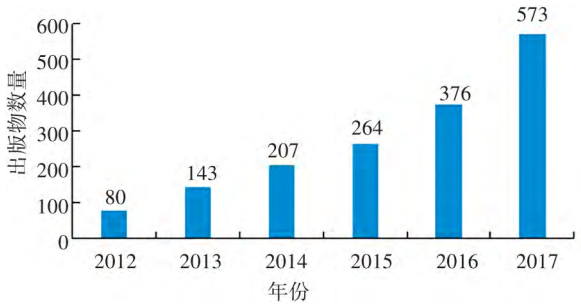

图 3　知识图谱在医疗健康领域研究的出版文献

Fig.3　Publications of knowledge graph in healthcare

知识图谱可以将各种医疗信息系统中琐碎、零散的知识相互连接，对信息进行分析，进行支持医疗信息获取[23]、医疗文本消歧[24]、综合性知识检索以及问答[25]、辅助决策支持[26-27]、疾病风险估计[28]等智能医疗应用. 对图 3 所示的 96 篇文献进行了可视化分析. 通过文献关键词进行了共现分析(见图 4)，知识图谱在医疗领域研究的高频词有查询、知识库、Unified medical language system 等，即为该领域的研究热点. 另外，这些高频词存在联系，例如存储、查询、知识库，可视化分析、文献计量、citespace，big data、human factors 等有较强的关联关系. 但是，从整体上来看，和传统研究领域相比，关键词的数目比较少. 由此可见，该领域的研究还存在较大的空白.

进一步考察 2004～2018 年间该领域研究热点的迁移情况，如图 5. 2007～2010 年间，文献侧重于讨论 UMLS 和机器学习；2010～2013 年间，研究转向预测、知识推理和分类；2013 年后，文献数量有了较大幅度的增长，关键词的变化速度加快，集中于大数据、医疗信息检索和医疗知识表现等.

图 4　文献关键词分析

Fig.4　Analysis of key words in literature

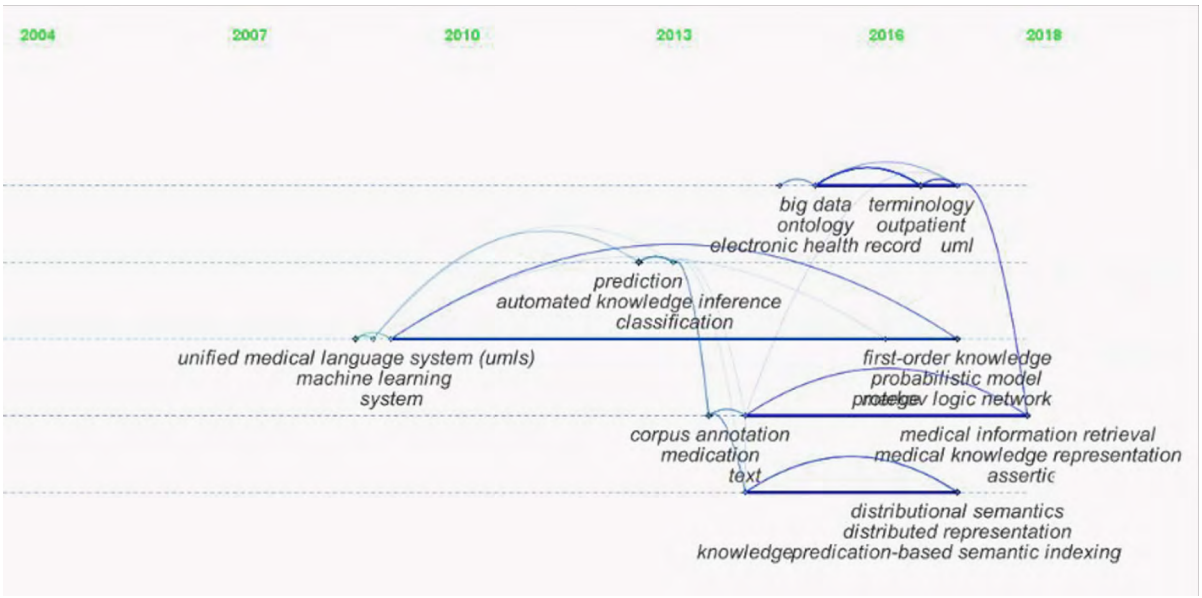

图 5 2004–2018 年间文献关键词的变化

Fig.5 Changes in the key words of the literature between 2004 and 2018

3 基于 EHR 的医疗知识图谱

以“电子健康记录 And 知识图谱”和“Electronic Health Records AND Knowledge Graph”中/英文为关键词，搜索了 Google Scholar 发布的 2012～2017 年的相关研究报告. 图 6 统计了 Google Scholar 上基于 EHR 的医疗知识图谱研究与应用的出版文献数量.

从图 6 可以看出，基于 EHR 的医疗知识图谱的研究目前还处于起步阶段.

表 1 总结了现有的几个比较典型的机构及其基于 EHR 的医疗知识图谱简介.

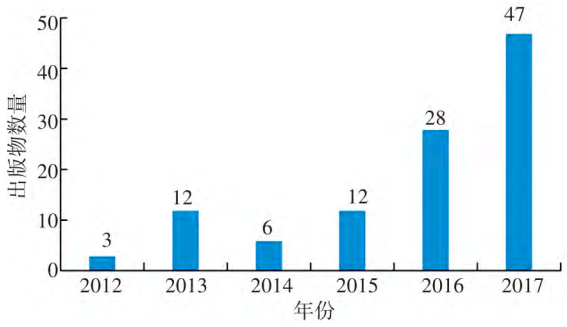

图 6 基于 EHR 的医疗知识图谱领域的研究文献

Fig.6 Publications of KG based on EHR

表 1 几个比较典型研究机构及其基于 EHR 的医疗知识图谱简介

Tab.1 Medicalknowledge graph based on EHR of some organizations

知识图谱名称	研究机构	主要贡献
PDD	西安交通大学	利用症状、诊断和相应治疗之间的关系，链接病人、疾病和药品[11]
口腔疾病和义齿组成知识图谱	北京大学口腔医学院	开发牙科中可摘取局部义齿的临床决策支持模型[29]
糖尿病知识图谱	浙江大学	为糖尿病患者提供计算机化的临床决策支持[30]
儿科混合知识图谱	中国科学院计算技术研究所	儿科疾病预测系统[31]
Partitioned knowledge Graphs	马里兰大学	在新的患者数据到来时，找出相应的主题，得到造成疾病的可能原因，给出相应建议[32]
QMKG	德克萨斯大学达拉斯分校	进行查询扩展，提高检索质量[15]
ADR-graph	伦敦国王学院	预测未知的药物不良反应[33]

表 2 给出了近期基于 EHR 的医疗知识图谱领域相关技术研究与应用成果.

3.1 EHR 信息抽取

EHR 信息抽取(Information Extraction，IE)是把 EHR 里包含的医疗过程和结果信息进行结构化处理. 输入信息抽取系统的是原始 EHR 数据，输出的是固定格式的信息点. 主要任务包括医疗实体提取、关系提取和主题提取，各方面的研究具体如下.

EHR 实体识别建立自然语言处理和知识表示的理论和方法的基础上，从文本数据中自动识别医学本体，包括疾病名称、症状、药物等专有名词. 例如：Lossioventura[34] 将知识图谱本体论用于自动化

数据处理，实现本体到数据库映射以及使用本体的术语在数据上添加元数据. Ritu Khare 等[34] 提出一种疾病命名实体识别方法，即使用主题词表 SNOMED-CT 和 MeSH 制备疾病词典，用 MetaMap 将生物医学文本映射到 UMLS，从而在给定药物标签的文字描述中识别所有疾病或适应症.

表 2　基于 EHR 的医疗知识图谱领域相关技术研究文献

Tab.2　Medical knowledge graph related literature based on EHR

相关研究和应用	下属研究和应用	相关人员(机构)
EHR 信息抽取	实体抽取	Khare R，et al.(U.S. National Institutes of Health)[34,35] Lossio-Ventura J A(University of Montpellier)[36]
	关系抽取	Harabagiu S M，et al.(University of Texas at Dallas)[16]
	主题抽取	Rose Yesha，et al.(University of Maryland)[32]
数据整合	数据整合	Bhuiyan M A R，et al.(East West University)[37] Cipière S，et al.(Clermont Université)[38] Lin L，et al.(Tsinghua University)[39] 阮彤等(华东理工大学)[40]
	知识嵌入	Kang B Y，et al.(Seoul National University)[19]
查询扩展	查询扩展	Harabagiu S M，et al.(University of Texas at Dallas)[15-16] Cipière S，et al.(Clermont Université)[38] Shen F，et al.(University of Missouri)[41]
	推理问答	Goodwin T R，et al.(University of Texas at Dallas)[42]
临床决策支持	慢性病	Zhang Y F，et al.(Zhejiang University)[30]
	专科诊断支持	Ling Y，et al.(Philips Research North America)[43] Sideris K(University of California)[26]
	药物不良反应	Jayaraman S，et al.(Pace University)[27] Iyer S V，et al.(Stanford University)[44]
疾病预测	疾病预测	Hajihashemi Z，et al.(University of Missouri-Columbia)[18] Esteban C，et al.(University of Munich)[45]

由于 EMR 的内聚性质，从 EMR 获得的临床知识可以体现医学概念之间的关系信息. 如：Ritu 等[34] 提出了一种通过捕捉医学概念相关的断言建立相关关系. Travis Goodwin 等[16] 集合 EMR 的共现信息生成关系并赋权，提出构建医学知识图谱的边和权重方法. Ritu[35] 利用 DailyMed 中的药物标签、PubMed 中的药物信息和 LabeledIn 中注释的药物-疾病-治疗关系，设计出预测新药物适应症的计算方法，利用 EMR 中的临床信息验证. Mihaela 等[18] 通过挖掘 UMLS，获得训练集中概念之间的关系路径，获得关系提取器，以更好地确定 EMR 数据中实体之间的语义关系.

EHR 包含了患者信息、病史、诊断、治疗方法和最终治疗结果，同一个主题下的同一个分区下，造成疾病的原因可能相同，由此可以给出相应建议，因此分析这些记录中的模式非常重要. Rose Yesha 等[32] 提出了一种自动发现患者数据中潜在主题和模式的方法，首先创建数据中的主要主题或模式，表示为主题知识图谱，根据 EHR 收集的术语对图进行分割，得到相应的主题.

3.2　数据整合

许多医院正汇总 EHR 数据建立数据仓库以提供各种监测和分析，帮助医生做出基于证据的临床决策. 由于 EHR 的格式不统一，所以在分析数据前首先要处理数据本身. Lin Liu 等[39] 收集了包括北京大学口腔医院、浙江大学口腔医院在内五家口腔医院的413 269 份电子医疗数据，建立基于语义网络的临床文档体系结构(CDA) 模板. 将 CDA 文档转化为知识图谱，实现特定单病种库文件的整合. 同时，医疗术语存在大量的同义词或上下位词，同一症状有多种文本表达形式，常常被不同的词语修饰. 医学诊断大量采用了国际疾病分类(ICD) 编码，但 ICD 编码结构并不包含完整的上下位关系. 通过 ICD 编码找到某一类患者基本不可能实现. 医生在编写一个疾病的 ICD 编码时，可粗可细，也会给 EHR 的自动处理带来困难. 阮彤等[40] 建立一个包含疾病、症状在内的医疗健康知识图谱，将结构化的文本与知识图谱相关联. 以此对 EHR 中可能包含

的家族史、时间事件、症状、检查、诊断、用药信息加以整合,形成结构化信息,从而得到结构化的病例.

3.3 查询扩展

医学是一个专业性极强的学科,非专业人士在搜索时可能会出现偏差,查询扩展尝试用同义词近义词替换、查询纠错、用同词根的词替换等解决这个问题.

Travis Goodwin 等[15]通过知识图谱中关系加权进行查询扩展,以提高检索质量. 给定查询后,提取关键字,通过与之相连的 20 个高权重的邻居来扩展关键字,根据加权查询进行评分,之后对文档进行重新排名,并对一些额外的约束进行评估. Guillaume 等[38]利用知识图谱以及 SPARQL 查询语言解决异构数据查询的问题. 将 SPARQL 查询语句分解重写,发送到不同医疗数据库,检索后将结果合并,以统一方式呈现. 谓词表示两个本体之间的关系,Feichen Shen[41]提出发现相邻关系谓词的层次模糊 C 均值聚类算法(HFCM),当某些属性比其他属性更紧密地相关,属性聚类和分区可以被用于高效地进行查询处理. 该算法以迭代的方式运行,直到集群中没有更多的邻居可以访问.

为找出关于复杂的医疗问题的答案,Travis Goodwin 等[42]从 EMR 系统提取的临床实践,从 PubMed Central 中的生物医学文章中取得的医学问题答案,通过知识推理,自动生成的医学知识的表示,最终实现医疗问答系统.

3.4 临床决策支持

临床诊断是病人治疗的关键,通常由专家的医学知识和直觉驱动. 临床决策支持(CDS)系统旨在为医生进行患者诊疗期间可能出现的复杂的临床决策提供建议[43]. 在知识图谱中检查每个疾病节点的症状,利用 EHR 记录的主诉症状和疾病节点症状的重叠对节点评分. 根据该评分对疾病重新排序,形成候选诊断集[13],是临床决策支持的一种重要方式. 这方面的成果包括:提供持续和个性化的慢病管理需要收集和分析 EHR 在内的患者数据,需要患者和医生在长期综合护理过程中密切合作. Yi-fan Zhang 等[30]建立了二型糖尿病知识图谱,为 544 个术语提供了丰富的定义和关系断言,为慢性病患者提供计算机化的临床决策支持. Vivek Datla 等描述了一个知识图谱为基础的临床诊断系统. Qingxiao Chen[29]用本体论设计患者口腔疾病和义齿组成部分的知识图谱,开发牙科中可摘局部义齿临床决策支持模型. 该模型利用余弦相似度算法计算患者与标准本体病例之间的相似度值,输出最相似的一组设计作为最终结果.

另外,涉及药品的临床决策需要考虑到药物不良反应(Adverse drug reaction,ADR). 尽管 ADR 在临床试验期间得到监测,但各种限制意味着并非所有 ADR 在药物被批准使用前都会被检测到. 因此,ADR 研究需要持续进行. 利用知识图谱进行 ADR 的研究,可以帮助医护人员避免药物不良反应的发生. Casey L[20]考虑 EHR 数据集,运用 DrugBank 和 PharmGKB 数据集,捕获语义知识,考虑药物之间和药物靶点之间的相似性等,得到患者药物使用-ADR-疾病结果,最终进行解释和预测. Saravanan[27]提出了一种利用知识图谱进行 ADR 知识表示的方法,为快速医生和患者推导药物不良反应,从而避免因人为失误造成高昂的代价.

3.5 疾病预测

每一种疾病的发生与发展都有各自的演变规律,如果能很好地了解和掌握这些规律,势必对各种疾病的预防起到干预和调控作用,对于预防疾病的大规模爆发至关重要.

例如:Zahra Hajihashemi[18]将传感器网络与 EHR 系统统一部署,以提供早期疾病识别. 利用统一医疗语言系统(Unified Medical Language System,UMLS)提供 Metamap 从每个护理笔记中提取一组概念. 然后将日常传感器序列与该患者的护理笔记相关的医学概念相关联. 推断未知时间的健康情况,计算其传感器序列与数据库中可用的序列之间的相似性.

4 讨 论

从 EHR 中获取医学信息,进行数据整合,构建医疗知识图谱,用医疗知识图谱协助医生进行查询扩展、临床决策支持、疾病预测,有利于提高医生的工作效率和诊断的准确率,减少医患矛盾的发生. 但是,由于该领域的发展还处于起步阶段,存在各种问题. 从 EHR 角度来说,各个机构的 EHR 系统独自发展,没有统一格式,不利于 EHR 的数据分析和二次利用. 从知识图谱角度来说,构建知识图谱这一环节大部分是由计算机学科的研究者完成的,领域专家的参与度较少,知识图谱的专业性并不一定能得到保证.

当前大多数 EHR 信息抽取研究是针对英文电子病历的,中文的医疗领域词典和知识库较少,词典和知识库的构建是接下来研究的必经之路. 数据整合主要是整合不同医疗机构的本地代码和不同数据字典的术语. 目前数据整合需要大量的人工干预,可以考虑使用合适的匹配算法进行自动匹配以提高效率. 最终形成 RDF/OWL 标准文件.

查询扩展是基于知识推理完成，结合语法和句法的搜索可能会增强查询扩展的效果. 为解决非医疗专业人士检索问题，在查询扩展的基础上，医疗问答系统研究兴起. 临床决策支持系统的准确性受到数据集的影响，未来的研究可以考虑结合基于模糊逻辑的算法，以减少对数据的依赖. 另外，临床决策尚不存在一个完整的评价体系，除了目前一般使用的完整性和准确性之外，还需要进行更全面的评价，包括效率、易用性和可扩展性等. 目前，疾病预测系统普遍精度较低，需要验证和人工干预，尝试使用预先训练分类器减少注释的成本和时间，提高精度. 具有相似逻辑和知识表示形式的科室可以考虑将目前已经存在的系统扩展到其他科室. 另外，将深度学习、表示学习等最新技术融入知识图谱，将有利于提高预诊辅诊的准确性.

5　总　结

EHR 存储了海量的医疗事实数据，研究者们希望从中提炼出有用的医学信息，帮助医生进行疾病诊断、治疗与预测. 目前对 EHR 的研究主要集中在构建、规范化、存储和基础挖掘方面. 要对 EHR 进行更加深入的分析，需要考虑如何更合理高效地处理海量 EHR 数据. 知识图谱提供了一种从海量文本抽取结构化知识的方法，可以对 EHR 进行有效的检索、比较、分析、整合和挖掘. 本文在研究 EHR 和知识图谱发展脉络的基础上，总结了基于 EHR 的医疗知识图谱在信息抽取、数据整合、查询扩展、临床决策支持和疾病预测的研究与应用进展，对未来发展方向和面临的挑战作了展望. 作为一种辅助工具，知识图谱与 EHR 的结合可以从现有的 EHR 中挖掘更多有用的医疗知识和医疗经验，实现广泛的共享，有利于实现医疗产业智能化，减少医患矛盾的发生，在一定程度上缓解看病难的问题.

参考文献

[1] 张涛，宗文红. 电子健康档案的发展与现状综述 [J]. 中国卫生信息管理杂志，2011，08(3)：83. DOI：10.3969/j.issn.1672-5166.2011.03.021
ZHANG Tao，ZONG Wenhong. Summary of development and current situation of electronic health records [J]. Journal of Chinese Health Information Management，2011，08(3)：83. DOI：10.3969/ j.issn.1672-5166.2011.03.021

[2] HAYRINEN K，SARANTO K. Definition，structure，content，use and impacts of electronic health records：a review of the research literature [J]. International Journal of Medical Informatics，2008，77(5)：291. Doi：10.1016/j.ijmedinf.2007.09.001

[3] EL-SAPPAGH S，ELMOGY M，RIAD A M，et al. EHR Datapreparation for case based reasoning construction [M]. Germany：Springer International Publishing，2014：483

[4] 谢冰洁，张旭峰，曾蔷，等. 门诊健康档案共享的探讨 [J]. 中国医疗设备，2010，25(3)：19. DOI：10.3969/j.issn.1674-1633.2010.03.005
XIE Bingjie，ZHANG Xunfeng，ZENG Qiang，et al. Discussion on the sharing of out-patient health records [J]. Chinese Medical Equipment Journal，2010，25(3)：19. DOI：10.3969/j.issn.1674-1633. 2010.03.005

[5] SUN Jinyuan，ZHU Xiaoyan，ZHANG Chi，et al. HCPP：Cryptography based secure EHR system for patient privacy and emergency healthcare [C] // International Conference on Distributed Computing Systems. USA：IEEE Press，2011：373. DOI：10.1109/ICDCS. 2011.83

[6] 刘愉，王立军. 基于 MongoDB 的 EHR 存储方案研究与设计 [J]. 中国数字医学，2013，6：20. DOI：10.3969/j.issn.1673-7571.2013.06.006
LIU Yu，WANG Lijun. Research anddesign of EHR storage solution based on MongoDB [J]. Journal of Chinese Digital Medicine，2013，6：20. DOI：10.3969/j.issn.1673-7571.2013. 06.006

[7] FOKOUE A，HASSANZADEH O. Tiresias：knowledge engineering and large-scale machine learning for interpretable drug-drug interaction prediction [C] // American Medical Informatics Association Symposium. [S.l.]：AMIA，2016：2116

[8] ZHANG Rui，LIU Ling，XUE Rui. Role-based and time-bound access and management of EHR data [M]. New York：John Wiley & Sons，Inc. 2014：994. DOI：10.1002/sec.817

[9] 朱寒阳. 面向转化医学的 EHR 数据接口与糖尿病数据挖掘研究 [D]. 浙江：浙江大学，2015
ZHU Hanyang. Research on EHRdata interface and diabetes data mining for translational medicine [D]. Zhejiang：Zhejiang University，2015

[10] HIRANO S，TSUMOTO S. Mining typical order sequences from EHR for building clinical pathways [C] // Workshops Held in Conjunction with the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Germany：Springer International Publishing，2014：39. DOI：10.1007/978-3-319- 13186-3_5

[11] WANG Meng，ZHANG Jiaheng，LIU Jun，et al. PDD Graph：bridging electronic medical records and biomedical knowledge graphs via entity linking [C] // International Semantic Web Conference. Germany：Springer International Publishing，2017：219. DOI：10.1007/978-3-319-68204-4_23

[12] JOHNSON A，POLLARD T，SHEN L，et al. MIMIC-III，a freely accessible critical care database [J]. Scientific Data，2016，3：160035

[13] DATLA V，HASAN S，QADIR A ，et al. Automated clinical diagnosis：The role of content in various sections of a clinical document [C] // IEEE International Conference on Bioinformatics and Biomedicine. USA：IEEE Press，2017：1004. DOI：10.1109/ BIBM. 2017.8217794

[14] GOODWIN T，HARABAGIU S. Embedding open-domain commonsense knowledge from text [C] // LREC Int Conf Lang Resour Eval，[S.l.]：LREC，2016：4621

[15] GOODWIN T，HARABAGIU S. Automatic generation of a qualified medical knowledge graph and its usage for retrieving patient cohorts from electronic medical records [C] // IEEE Seventh International Conference on Semantic Computing.USA：IEEE Press，2013：363. DOI：10.1109/ ICSC.2013.68

[16] GOODWIN T，HARABAGIU S. Graphical induction of qualified medical knowledge [J]. International Journal of Semantic Compu-

ting, 2013, 7(04): 377

[17] YUAN Ling. Methods and techniques for clinical text modeling and analytics [D]. Philadelphia: Drexel University, 2017

[18] HAJIHASHEMI Z, POPESCU M. Predicting health patterns using sensor sequence similarity and NLP [C] // IEEE International Conference on Bioinformatics and Biomedicine Workshops. USA: IEEE Press, 2013: 948. DOI: 10.1109/BIBMW.2012.6470278

[19] KANG B, KIM D, KIM H. Two-phase chief complaint mapping to the UMLS metathesaurus in Korean electronic medical records [J]. IEEE Transactions on Information Technology in Biomedicine, 2009, 13(1): 78

[20] OVERBY C, FLORES A, PALAM G, et al. Combining multiple knowledge sources: a case study of drug induced liver injury [C] // International Conference on Data Integration in the Life Sciences. Germany: Springer International Publishing, 2015: 3. DOI: 10.1007/978-3-319-21843-4_1

[21] 俞思伟, 范昊, 王菲, 等. 基于知识图谱的智能医疗研究 [J]. 医疗卫生装备, 2017, 38(3): 109. DOI: 10.7687/J.ISSN1003-8868.2017.03.109
YU Siwei, FAN Hao, WANG Fei, et al. Research on intelligent medicine based on knowledge graph [J]. Chinese Medical Equipment Journal, 2017, 38(3): 109. DOI: 10.7687/J.ISSN1003-8868.2017.03.109

[22] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述 [J]. 电子科技大学学报, 2016, 45(4): 589. DOI: 10.3969/j.issn.1001-0548.2016.04.012
XU Zenglin, SHENG Yongpan, HE Lirong, et al. Review onknowledge graph techniques [J]. Journal of University of Electronic Science and Technology, 2016, 45(4): 589. DOI: 10.3969/j.issn.1001.2016.04.012

[23] DESARKAR M, BHAUMIK S, SATHISH S, et al. Med-Tree: A user knowledge graph framework for medical applications [C] // International Conference on Bioinformatics and Bioengineering. USA: IEEE Press, 2013: 1. DOI: 10.1109/BIBE.2013.6701564

[24] EL-RAB W, EL-HAJJ M. Biomedical text disambiguation using UMLS [C] // IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. USA: IEEE Press, 2014: 943. DOI: 10.1145/2492517.2500251

[25] ZHANG Guangzhi, BIE Rongfang, SUN Yunchuan. Bring biomedical ontologies to personalized healthcare: a smart inquiry framework [C] // IEEE First International Conference on Connected Health. USA: IEEE Press, 2016: 84. DOI: 10.1109/CHASE.2016.32

[26] SIDERIS K. Mining electronic health records to improve remote health monitoring [D]. Los Angeles: University of California, 2016

[27] JAYARAMAN S, TAO L, GAI K, et al. Drug side effects data representation and full spectrum inferencing using knowledge graphs in intelligent telehealth [C] // International Conference on Cyber Security and Cloud Computing. USA: IEEE Press, 2016: 289. DOI: 10.1109/CSCloud.2016.49

[28] PATIL M, BHAUMIL S, PAUL S, et al. Estimating personalized risk ranking using laboratory test and medical knowledge (UMLS) [C] // Engineering in Medicine and Biology Society. USA: IEEE Press, 2013: 1274

[29] CHEN Qingxiao, WU Ji, LI Shusen, et al. An ontology-driven, case-based clinical decision support model for removable partial denture design [J]. Scientific Reports, 2016, 6: 27855. DOI: 10.1038/srep27855

[30] ZHANG Yifan, GOU Ling, ZHOU tianshu, et al. Anontology-based approach to patient follow-up assessment for continuous and personalized chronic disease management [J]. Journal of Biomedical Informatics, 2017, 72: 45. DOI: 10.1016/j.jbi.2017.06.021

[31] LIU Penghe. HKDP: a hybrid knowledge graph based pediatric disease prediction system [C] // International Conference on Smart Health. Germany: Springer International Publishing. 2016: 78

[32] YESHA R, et al. A graph-based method for analyzing electronic medical records [C] // IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. USA: ACM, 2015: 1036. DOI: 10.1145/2808797.2808806

[33] BEAN D, WU H, IQBAL E, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records [J]. Scientific Reports, 2017, 7(1): 16416

[34] KHARE R, LI J, LU Z. Automatic extraction of drug indications from FDA drug labels [J]. AMIA Symposium, 2014, 5: 787

[35] KHARE R, HSUAN C, LU Z. LabeledIn: Cataloging labeled indications for human drugs [J]. Journal of Biomedical Informatics, 2014, 52: 448. DOI: 10.1016/j.jbi.2014.08.004

[36] LOSSIOVENTURA J. Towards the french biomedical ontology enrichment [D]. Montpellier: University of Montpellier, 2015

[37] BHUIYAN A, ULLAH R. An interactive healthcare system of big data application with predictive model analysis [D]. Chicago: East West University, 2017

[38] CIPIERE S, ERETEO G, GAIGNARD A, et al. Global initiative for sentinel e-health network on grid (GINSENG): medical data integration and semantic developments for epidemiology [C] // IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. USA: IEEE Press, 2014: 755. DOI: 10.1109/CCGrid.2014.45

[39] LIU Liu, FENG Letong, CAO Zhanqiang. Requirements engineering for health data analytics: challenges and possible directions [C] // Requirements Engineering Conference. USA: IEEE Press, 2016: 266

[40] 阮彤, 高炬, 冯东雷, 等. 基于电子病历的临床医疗大数据挖掘流程与方法 [J]. 大数据, 2017, 3(5): 83. DOI: 10.11959/j.issn.2096-0271.2017054
RUAN Tong, GAO Ju, FENG DongLei, et al. Process and methods of clinical big data mining based on electronic medical records [J]. Big Data Research, 2017, 3(5): 83. DOI: 10.11959/j.issn.2096-0271.2017054

[41] SHEN Feichen. Predicate oriented pattern analysis for biomedical knowledge discovery [J]. Intell Inf Manag, 2016, 8(3): 66

[42] GOODWIN T, HARABAGIU S. Knowledgerepresentations and inference techniques for medical question answering [J]. Acm Transactions on Intelligent Systems & Technology, 2017, 9(2): 1. DOI: 10.1145/3106745

[43] YUAN Ling, HASAN S, DATLA V, et al. Learning to diagnose: Assimilating clinical narratives using deep reinforcement learning [C] // Proceedings of the Eighth International Joint Conference on Natural Language Processing. Taiwan: [s.n.], 2017: 895

[44] IYER S. Learningdrug-drug interactions from the unstructured text of electronic health records [D]. USA: Stanford University, 2013

[45] ESTEBAN C. Predicting sequences of clinical events by using a personalized temporal latent embedding model [C] // International Conference on Healthcare Informatics. USA: IEEE Press, 2015: 130. DOI: 10.1109/ICHI.2015.23

(编辑 苗秀芝)