

• 卫生管理 • doi:10.3969/j.issn.1671-8348.2021.01.037

# 支持临床决策的医学知识图谱的构建与应用\*

郑少宇<sup>1</sup>, 滕飞<sup>1△</sup>, 马征<sup>1</sup>, 陈泽君<sup>2</sup>, 马虹<sup>2</sup>, 吴洁<sup>2</sup>

(1. 西南交通大学信息科学与技术学院, 成都 611700; 2. 四川省成都市第六人民医院 610000)

**[摘要]** 知识图谱在语义检索、信息分类等领域扮演着重要的角色,但面向决策支持等强逻辑型任务的表现仍差强人意。该研究综合考虑了国内外主流的医学知识分类体系和知识图谱构建方法,针对临床医学主要知识源,利用实体识别、关系抽取等技术,探索了支持临床决策的医学图谱的本体结构和构建方法,并结合临床工作设计了与图谱结构相适应的临床决策支持系统,在重要临床环节提供有效的决策支持。

**[关键词]** 医学知识图谱; 临床决策支持; 知识检索

**[中图分类号]** R19

**[文献标识码]** B

**[文章编号]** 1671-8348(2021)01-0163-04

截至 2018 年 11 月,全国医疗卫生机构全年总诊疗人次达 75.4 亿人次,较 2017 年同比提高 3.2%<sup>[1]</sup>。在就医人数逐年增加的背后,是我国医疗卫生体制改革的不断深化和医疗行业体量的高速扩张。在这样的环境下,传统的医学知识存储模式和应用方法都受到了极大的挑战。

在存储模式方面,近年来兴起的知识图谱凭借其易拓展、易解释、易应用等诸多优势,在医学知识存储领域得到广泛应用。在中文医学知识图谱方面,近期较突出的研究成果有 CMeKG<sup>[2]</sup>、中医药知识图谱<sup>[3-4]</sup>等。这些知识图谱结构化地表示了常见疾病与其症状、治疗方案等之间的关系,对于临床工作有着重要意义。

随着医学知识存储模式一同改变的还有医学知识的应用方式。其中,临床决策支持作为最早提出的数字化医学知识应用方式之一,已从最初只能支持单病种诊疗辅助的专家系统,发展出了医学知识问答、临床环节质控、慢病管理等多种应用形式,业内亦不乏 Watson、惠美、嘉和美康等优秀成果。但由于临床工作的特殊性,使用知识图谱进行决策支持仍面临诸多困难。主要原因是:(1)临床知识结构复杂,应用级图谱的构建难度和成本较高;(2)临床工作专业化程度极高,致使临床辅助方法设计难度较大。

鉴于目前我国医疗行业对于高质量临床决策支持系统的迫切需求及临床医学知识图谱构建和应用所面临的诸多难点,本研究选择以知识图谱作为知识载体,从临床工作的实际流程和需求出发,探究医学知识图谱在临床决策支持领域的应用方法。设计了流程化的临床医学知识图谱构建方法,并根据临床工作的实际流程开发了支持在图谱上直接进行逻辑推

理的临床决策支持系统。

## 1 相关工作

### 1.1 临床医学知识图谱

与大多数知识图谱构建工作所涉及的流程类似,医学知识图谱的构建也需要经过本体建模和实例抽取两个步骤。且由于医学知识的特殊性,应用级医学图谱对于知识准确性的要求远高于通识领域,因此更需要合理的图谱构建流程。

在中文医学知识图谱方面,近年来较成熟的有 CMeKG<sup>[2]</sup>、中医药知识图谱<sup>[3-4]</sup>等面向多疾病的大规模图谱,以及乳腺肿瘤知识图谱<sup>[5]</sup>、高发性孕期疾病知识图谱<sup>[6]</sup>等面向单一病种的中小规模图谱。上述图谱多围绕疾病、药物和治疗手段这三类实体为核心进行构建,着重知识描述而非面向特定的知识应用,因此部分图谱的复杂度远不能达到临床决策支持的要求。以症状类实体为例,多数图谱中该类实体仅有“症状名”这一属性,而缺失了症状对于某一疾病而言的持续时间、发生条件等临床工作中重点关注的信息,这也是目前多数医学知识图谱的普遍缺陷。

### 1.2 医学知识图谱在临床决策支持工作中的应用

临床决策支持的概念最早可以追溯到 20 世纪 50 年代末,80 年代由 BUCHANAN 等研制出了首个临床决策支持系统(CDSS)。国内 CDSS 研究,大致可分为基于规则、基于案例和基于模型三类,陈全福等<sup>[7]</sup>使用案例推理(case-based reasoning, CBR),通过深度学习算法对案例进行匹配,进而给出诊疗建议。叶枫等<sup>[8]</sup>针对老年痴呆症的诊断,设计了一个老年痴呆症临床决策支持系统,用于辅助医生对老年痴呆症进行诊断。刘永斌等<sup>[9]</sup>提出基于知识库的临床决策支持系统技术框架,通过建立权威知识库来为临床工

\* 基金项目:四川省科技计划项目(2017SZYF0002)。 作者简介:郑少宇(1995—),工程师,硕士,主要从事医学数字化方面的研究。

△ 通信作者, E-mail: fteng@swjtu.edu.cn。

作提供建议。上述工作为中文 CDSS 构建提供了宝贵的实践经验,但也反映了目前 CDSS 构建中面临的诸如推理机可解释性欠佳、知识库构建困难等常见问题。

## 2 临床医学知识图谱构建

### 2.1 概述

知识图谱构建主要分为本体层建模和实例层构建两个阶段。目前本体层建模方法主要有 7 步法、骨架法、TOVE 法、SENSUS 法<sup>[10]</sup>等,这些抽取方法主要还是依托领域专家,根据需要覆盖的知识领域和面向的实际任务进行知识分类。本研究主要参考 7 步法和骨架法进行图谱本体层设计。

### 2.2 本体层建模流程

#### 2.2.1 提取领域知识中的类及属性

领域知识通过抽象以后可得到基本的类和属性。其中,类是指领域知识中的核心概念,如临床医学知识中的“疾病”“症状”等概念,类的实例称为知识元。属性可以分为对象属性和数据属性两种。对象属性用于定义类之间关系,如“(疾病)-表现为-(症状)”,而数据属性则用于定义类与描述该类某一特征的字符串之间的关系,如“(疾病)-名称-String”。根据临床工作中涉及的知识领域,可以将相关知识分为理论知识和临床经验两个大类,每个大类又包含若干主要知识源。通过对主要知识源中的重要术语进行标注和归类,提取出了临床知识领域的核心概念。

#### 2.2.2 临床医学知识图谱的本体层建模

将知识来源中抽象出的类及属性加以组织后得到了图 1 所示的本体层模型。该本体层中包含疾病、症状(症状下设一般症状、体征、异常检查结果 3 个子类)、检查、治疗等共计 9 个类和 ID、名称、别名、类型、描述、部位、条件等共计 26 种属性,其中 14 种为数据属性,12 种为对象属性。在这些属性中,为了使得图谱的知识维度足以支撑临床辅助工作,本研究围绕症状设计了多种特有数据属性,如附加词、条件、人群等,使得对于临床医学知识的描述更加完整、具体。部分属性和关系的释义见表 1、表 2。

表 1 数据属性释义

属性名称	概念
ID	知识元的唯一标识符
名称	知识元的名称
别名	知识元的别名
描述	对于知识元的描述
备注	知识元的生成日期、历史回溯、编写人员信息等
附加词	当多个知识元的名称同时出现并存在缺省,且缺省的内容不属于部位时,将缺省内容作为附加词存储
部位	症状或检查发生的部位

续表 1 数据属性释义

属性名称	概念
条件	症状发生的时间、动作条件等
程度	症状的严重程度和表现方式
人群	症状发生的特定人群
概率	症状在相关疾病中发生的概率
所在病程	症状在相关疾病中出现的时段
频率	症状发生的频率
单位	当涉及异常检查结果时,用于存储检查结果的单位
区间	当涉及异常检查结果时,用于存储检查结果的数值域

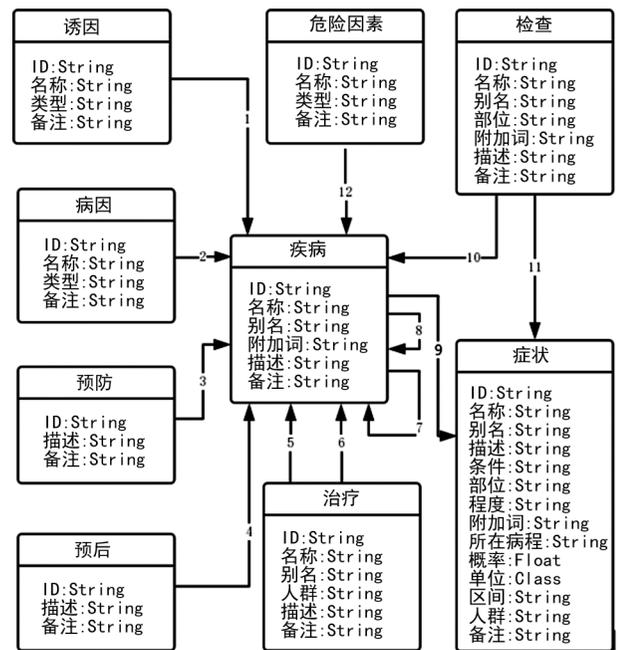


图 1 临床医学知识图谱的本体层

表 2 对象属性释义

属性编号	概念
1	(诱因)-诱发-(疾病)
2	(病因)-导致-(疾病)
3	(预防)-预防-(疾病)
4	(预后)-预后-(疾病)
5	(治疗)-改善-(疾病)
6	(治疗)-恶化-(疾病)
7	(疾病)-包含-(疾病)
8	(疾病)-伴随-(疾病)
9	(疾病)-表现为-(症状)
10	(检查)-发现-(症状)
11	(检查)-证实-(疾病)
12	(危险因素)-影响-(疾病)

### 2.3 实例层构建流程

由于目前无监督的方法在面向医学文本进行实体识别时效果并不理想,因此本研究选用了半监督的方法来进行知识抽取工作。工作主要集中在抽取各知识源中的疾病、症状实体及其相关属性。在面向医学教材和临床指南等核心知识源时,采用多角色并行

标注的方法开展知识抽取工作,具体流程如下,(1)前期准备:根据本体层的知识分类体系和语义标准制订了《中文医学文献命名实体和实体关系标注规范》,并根据多角色标注的任务需求开发了智能协同文本标注系统。(2)多角色标注:利用多名非医学专业人员并行标注,在抽取过程中,大部分症状属性缺乏专业词表,以症状的“条件”属性为例,常见的有“白天加重”“夜间加重”“躺卧时加重”等,这些词在医学系统中属于通用词汇,但缺少标准化的、统一的描述方法。因此需要将标注过程中发现的全部症状条件交由专家进行对齐,从而得到该属性的标准词表。对于文本中出现过的非标准词汇,则作为标准词的一种映射。映射表与标准词表共同构成了症状某一属性的专业词表。(3)知识评审:对于标注结果中出现的标注冲突、存疑结果和部分随机抽查结果,反馈给医院专家进行审核。(4)知识存储:通过审核的抽取结果转换成图结构进行存储,通过多角色+专家评审的方法,最大限度地保证了核心知识图谱的准确度。

在面向高质量的网络医学知识时,由于网页本身为半结构化数据,因此可以通过网络爬虫+模板对网络数据进行批量处理,所得结果同样需要经过上述(3)、(4)步骤。最终通过人工并行标注和自动标注两种手段在不同知识源中的合理应用,在保证知识质量的前提下,实现了图谱的快速扩充。

截止成稿时,构建的临床医学知识图谱已覆盖呼吸内科、消化内科、儿科、心血管内科等十余个主要科室的 800 余种常见疾病,以及相关症状、检查、治疗等数千条,初步具备了开展临床诊断辅助系统设计所需要的数据量,图谱中的实体数目和关系统计见表 3,部分关系三元组见图 2。

表 3 床医学知识图谱数据统计

数据来源	知识元	关系	知识元	关系
	总数	总数	类别总数	类别总数
医科教材	6 751	7 015	3	4
诊疗指南	1 957	2 031	3	4
网络数据	1 043	1 231	3	3
门诊病历	4 62	631	3	3
总计(去重后)	8 561	10 135	3	4

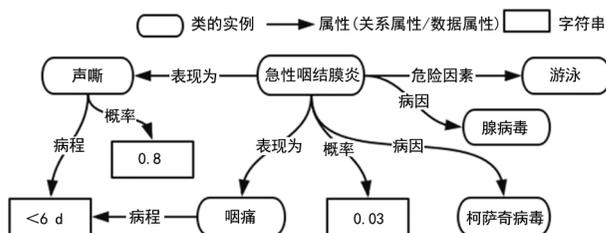


图 2 知识图谱中部分三元组

### 3 临床医学知识图谱的应用

目前基于医学知识图谱开展决策支持工作是医

疗信息化领域的研究热点之一。本研究在传统 CDSS 结构的基础上,设计了基于知识图谱的诊断推理模块,通过在知识图谱上模拟临床作业思维进行查询和计算,实现了在重要临床环节的同步决策支持。最终完成的原型系统由 CDSS 前端接口(A),推理诊断模块(B)和知识库管理模块(C)3 个部分组成,系统结构如图 4 所示。

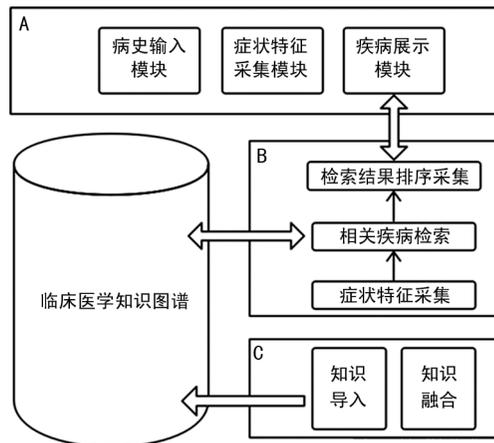


图 3 CDSS 结构图

为便于决策支持工作的介入,同时直接利用知识图谱进行临床推理,本研究将单病诊断辅助工作划分为下述 6 个步骤,(1)症状特征采集:通过监听现病史输入栏,症状输入后会先利用专业词表进行一次消歧,将同义词或描述不规范的症状名称规范化,并利用模板切分出症状的表现形式和持续时间。(2)相关疾病检索:在知识图谱中检索含有当前症状的疾病。(3)检索结果排序:利用病历中统计得到的疾病概率和症状权重进行排序,按疾病与现病史的相关度降序排列,排序结果以 JSON 文件的形式返回前端。(4)答案生成:通过解析 JSON 文件,重绘前端界面中疾病展示模块的内容,使用者可以直观地获取与现病史相关的所有疾病信息,包括相关疾病的伴随症状及确诊所需要的检查项目,为制订后续问诊计划提供思路。(5)生成流程记录:在下达诊断后,会评估操作流程的规范程度和诊断结果的可靠度并生成流程记录文件,对于不符合临床规范的诊断流程或诊断结果,将其流程记录模块等待专家核查。(6)下达诊断。其中步骤(2)、(3)、(4)一般会循环执行多次。总的来说,诊断辅助工作通过症状特征驱动,利用相关疾病排序、伴随症状提醒、相关检查提醒、诊断结果检验 4 种手段共同推动问诊工作的进行,帮助医务人员实时、全面地了解当前患者的相关疾病,为后续问诊提供思路。

考虑各医院硬件条件不同,CDSS 选择以 Web 端的形式进行开发,同时与传统 HIS 系统进行了整合,在不改变医务人员既有作业习惯的前提下,实现对于医务人员的决策支持。系统界面功能概览见图 4。



图 4 软件功能概览

通过学习成都市某三甲医院呼吸内科近 13 万份病历的统计特征,并用其现病史部分作为输入,第一诊断作为标准输出,通过识别现病史中的症状特征进行疾病推荐。由于本系统返回的症状是列表形式,故以实际诊断结果出现在推荐列表前 N 的比例作为实验评测结果,呼吸科门诊病历 TOP-1 命中率 61.4%, TOP-3 命中率 67.3%, TOP-5 命中率 68.8%。该实验证明了最终设计出的 CDSS 可以有效地在主要临床环节进行实时决策辅助,对于主要科室当中的常见疾病均有较好的预测结果。同时系统交互方式也符合门诊医生当前的工作习惯,得到了医院专家的认可。

#### 4 总 结

医学知识图谱应用范围极广,但由于临床医学知识的复杂性,使得流程化的知识图谱构建工作难以进行,同时,复杂的临床工作流程进一步加大了医学知识图谱在临床决策工作中的应用难度。本研究在充分分析临床知识结构和临床工作特点的基础上,设计了构建支持临床决策的医学知识图谱的可行流程,并开发了与图谱结构匹配的 CDSS 原型系统。在后续研究工作中,拟利用深度学习技术深入发掘医学领域的诊断规则和模式,并逐步扩大知识图谱规模,进一步提高 CDSS 的可靠性,为大型 CDSS 的开发打下基础。

#### 参考文献

[1] 卫计委统计信息中心. 2018 年 1-11 月全国医疗服务情况[R/OL]. [2019-1-23]. <http://www.nhc.gov.cn/mohwsbwstjxxzx/s7967/201901/57dec69d2c8c4e669864b067d2a1fb2e.shtml>.

- nhc.gov.cn/mohwsbwstjxxzx/s7967/201901/57dec69d2c8c4e669864b067d2a1fb2e.shtml.
- [2] 陆春吉,李军莲,郭进京. 医疗人工智能与临床医学术语标准[J]. 医学信息学杂志,2018,39(5): 155-158.
  - [3] 贾李蓉,于彤,崔蒙. 中医学语言系统研究进展[J]. 中国数字医学,2014,11(10):57-59.
  - [4] 贾李蓉,刘静,于彤. 中医药知识图谱构建[J]. 医学信息学杂志,2015,36(8):51-53.
  - [5] 崔洁,陈德华,乐嘉锦. 基于 EMR 的乳腺肿瘤知识图谱构建研究[J]. 计算机应用与软件,2017,34(12):122-126.
  - [6] 何斌杰,杨丹林,张栋,等. 面向高发性孕期疾病医学文献的知识图谱解析研究[J]. 福建医药杂志,2017,39(3):87-88,139.
  - [7] 陈全福,叶焕文,杨荣源,等. 基于案例库构建中医临床决策支持系统[J]. 广州中医药大学学报,2016,33(4):585-587.
  - [8] 叶枫,周根贵,南山. 老年痴呆症诊断临床决策支持系统设计与评估[J]. 中国生物医学工程学报,2009,28(6):872-877.
  - [9] 刘永斌,魏明月,于广军,等. 基于临床决策支持的智能化电子病历集成平台建设与实践[J]. 中国医院,2017,21(8):8-11.
  - [10] 王美琴,吴庆斌. 基于本体的医学知识库构建方法综述[J]. 医学信息学杂志,2017,38(3):73-76.

(收稿日期:2020-04-18 修回日期:2020-10-06)