

文章编号: 1003-0077(2019)10-0001-07

中文医学知识图谱 CMeKG 构建初探

奥德玛^{1,2}, 杨云飞^{1,2}, 穗志方^{1,2}, 代达勋^{1,2}, 常宝宝^{1,2}, 李素建^{1,2}, 咎红英^{2,3}

(1. 北京大学 计算语言学教育部重点实验室, 北京 100871;

2. 鹏城实验室, 广东 深圳 518055;

3. 郑州大学 信息工程学院, 河南 郑州 450001)

摘要: 医学知识图谱是智慧医疗应用的基石, 可以为机器阅读理解医学文本、智能咨询、智能诊断提供知识基础。现有的医学知识图谱从规模化、规范化、体系性、形式化等方面还不足以满足智慧医疗应用的需求。此外, 对复杂医学知识的精准描述更是构建医学知识图谱面临的重要挑战。针对上述问题, 该文利用自然语言处理与文本挖掘技术, 以人机结合的方式研发了中文医学知识图谱第一版 CMeKG 1.0 (Chinese Medical Knowledge Graph)。CMeKG 1.0 的构建参考了 ICD-10、ATC、MeSH 等权威的国际医学标准术语集以及规模庞大、多源异构的临床路径指南、临床实践、医学百科等资源, 覆盖了疾病、药物和诊疗技术, 包括 100 余万个医学概念关系的实例。该文综述了 CMeKG 1.0 构建过程中的描述体系、关键技术、构建流程以及医学知识描述等相关问题, 希望为医学领域知识图谱的构建与应用提供一些参考。

关键词: 知识图谱; 智慧医疗; 知识描述体系; 知识提取

中图分类号: TP391

文献标识码: A

Preliminary Study on the Construction of Chinese Medical Knowledge Graph

BYAMBASUREN Odmaa^{1,2}, YANG Yunfei^{1,2}, SUI Zhifang^{1,2}, DAI Damai^{1,2},
CHANG Baobao^{1,2}, LI Sujian^{1,2}, ZAN Hongying^{2,3}

(1. Key Laboratory of Computational Linguistics, Ministry of Education,
Peking University, Beijing 100871, China;

2. Peng Cheng Laboratory, Shenzhen, Guangdong 518055, China;

3. School of Information Engineering, Zhengzhou University, Zhengzhou, Henan 450001, China)

Abstract: The medical knowledge graph is the cornerstone of intelligent medical applications. The existing medical knowledge graphs are not enough from the perspectives of scale, specification, taxonomy, formalization as well as the precise description of the knowledge to meet the needs of intelligent medical applications. We apply natural language processing and text mining techniques with a semi-automated approach to develop the Chinese Medical Knowledge Graph (CMeKG 1.0). The construction of CMeKG 1.0 refers to the international medical coding systems such as ICD-10, ATC, and MeSH, as well as large-scale, multi-source heterogeneous clinical guidelines, medical standards, diagnostic protocols, and medical encyclopedia resources. CMeKG covers types such as diseases, drugs, and diagnosis/treatment technologies, with more than 1 million medical concept relationships. This paper presents the description system, key technologies, construction process and medical knowledge description of CMeKG 1.0, serving as a reference for the construction and application of knowledge graphs in the medical field.

Keywords: knowledge graph; intelligent medical treatment; knowledge description system; knowledge extraction

收稿日期: 2019-04-15 定稿日期: 2019-06-17

基金项目: 国家重点研发计划(2017YFB1002101); 国家自然科学基金(61772040, 61751201)

0 引言

2012 年,谷歌提出了知识图谱(Knowledge Graph)的概念,为世界知识和领域知识的构建提供了一个可资借鉴的手段^[1]。知识图谱的基本组成是由头实体、尾实体和两者之间的关系组成的三元组关系。目前,对知识图谱的研究应用主要包括通用知识图谱和垂直领域知识图谱。典型的通用知识图谱有 Google Knowledge Graph^[1]、YAGO^[2]、DBpedia^[3]、CN-DBpedia^[4]、XLORE^[5]等。虽然通用知识图谱收集了大量的领域知识,但是受到概念约束,无法完整描述比较复杂的领域知识。垂直领域知识图谱在领域知识的描述方面优于通用知识图谱,但常采用手工构建方法,因此其构建成本很高。

在知识图谱概念提出之前,众多研究机构已经利用知识库的存储方式构建了大量核心医学资源,如世界卫生组织维护的国际疾病分类代码 ICD-10^[6]、美国国立医学图书馆的一体化医学语言系统 UMLS^[7]及其收录的 100 多种词表和分类体系、国际医疗术语标准开发组织维护的系统化临床医学术语集 SNOMED-CT^[8]、Linked Open Data(LOD)收集的 1 000 多种生命科学知识库、BioPortal^[9]、Bio2RDP^[10]等。上述医学核心资源对医学领域的相关知识进行了专业的描述,这些资源可服务于医学文献检索或医学术语标准化等医学信息化应用,但还不能满足现代智慧医疗对医学知识描述的结构化、精细化和自动化的需求。

鉴于医学知识图谱在知识推理、智能问答、辅助诊断等智能医疗应用中的重要作用,已有一些单位陆续开展了医学知识图谱的构建工作。中国中医科学院中医药信息研究所基于已有的中医药学语言系统^[11]构建了中医药知识图谱^[12]、上海曙光医院和华东理工大学构建了中医药知识图谱^[13]、华东理工大学构建了中文症状库^[14]、中国医学科学院医学信息研究所构建了医药卫生知识服务系统知识图谱^[15]等。不过,现有医学知识图谱从规模化、规范化、形式化、体系性等方面都仍有很大的提升空间。如何基于高效的工程方法、权威的医学数据资源、精准的知识描述体系和先进的文本挖掘技术,构建大规模、高质量的医学知识图谱,仍是极具挑战性的课题。

针对上述挑战,本文利用自然语言处理与文

本挖掘技术,以人机结合的方式研发了中文医学知识图谱第一版 CMeKG 1.0。本文综述了 CMeKG 构建过程中的描述体系、关键技术、构建流程以及医学知识描述相关问题。本文组织结构如下:第 1 节介绍医学知识图谱相关研究工作,第 2 节介绍 CMeKG 构建的总体方案,第 3 节介绍医学知识描述体系,第 4 节介绍知识提取技术,第 5 节介绍医学知识图谱构建的初步结果,第 6 节为总结与展望。

1 相关工作

医学领域术语和知识资源的收集与构建由来已久,已取得显著的成果。比较典型的包括一些应用广泛的医学术语集,如 ICD-10、ATC^[16]和 MeSH^[17]等。

ICD-10 是世界卫生组织维护的国际疾病分类代码,它是包含疾病、症状、体征、异常发现、社会环境以及外部原因导致的损伤或疾病的代码。ICD-10 编码以树状层级结构描述了 22 大类的 10 000 多个概念,通常辅助于公共卫生组织跟踪疾病、保险公司理赔以及医院存储电子病历等任务。ATC 是世界卫生组织维护的解剖学、治疗学及化学分类系统,它的本体结构依据的是药物成分对人体解剖学的治疗、药理和化学特征效应,ATC 共包含消化道、血液形成器官、心血管等 14 大类解剖学概念对应的药物成分及其用法用量标准。MeSH 是美国国立医学图书馆编制的医学主题词表,它包含 15 大类所属的 18 000 多个医学主题词,主要辅助 PubMed 标引和检索医学文献。

中国中医科学院中医药信息研究所构建的中医药知识图谱包含中医药学语言系统的 100 余万语义关系,其主要目的是形象地表达概念之间的关系,通过浏览界面提高用户检索领域知识的体验。随后,上海曙光医院也构建了中医药知识图谱,实现了面向临床应用的半自动化知识图谱的构建和知识问答、辅助开药等应用。华东理工大学构建的中文症状库是结合中医和西医知识自动构建的知识图谱。中文症状库涉及 10 万多个实体及实体之间的 20 余种语义关系,共包含 60 多万个三元组关系。中国医学科学院医学信息研究所构建的医药卫生知识服务系统知识图谱基于医学百科的医学知识,以可视化界面展开医学百科知识。

综上所述,国外医学信息化领域存在结构化程度较高、规模较大的资源(例如,UMLS 等),而

中文医学知识目前以非结构化信息为主,缺少大规模结构化的开源资源。目前中文的医学知识图谱在覆盖的知识规模^[15]、知识的描述规范体系^[14]、知识的开源性^[12-13]和形式化等方面还存在诸多不足。因此,我们的研究目标是建立大规模、高质量的中文医学知识图谱,为智慧医疗奠定专业知识基础。

2 中文医学知识图谱 CMeKG 构建总体方案

目前关于现代医学方面的中文医学知识图谱还比较少。因此本文定位于现代医学中文知识图谱工程构建研究,旨在介绍我们开发的中文医学知识图谱,集中描述其总体构建流程和资源概貌。CMeKG 的最终目标是制订层次清晰、高度关联的结构化医学知识描述体系,研发高性能的医学知识图谱构建方法和关键技术;搭建基于自然语言处理技术的知

识图谱构建平台,形成知识图谱构建的自动化和规范化工程模式;采集加工多级医疗数据,建立高质量的中文医学知识图谱。作为 CMeKG 构建的第一阶段,我们搭建了中文医学知识图谱的基础框架,并建立知识图谱第一版。

在 CMeKG 的构建过程中,我们在语料处理和知识提取时充分利用了中文分词、命名实体识别以及关系提取技术,通过技术手段提升了医学知识图谱构建的自动化程度,扩大了可处理的医学文本的来源和规模,使得知识图谱的信息更充分。同时,作为一个医学知识图谱,不但要保证知识来源的充分性,而且要考虑知识组织和知识内容的权威性。针对后者我们充分利用了医学领域国际标准,如 ICD-10、ATC、MeSH 来组织和引导知识图谱构建过程,保证我们的医学知识的权威性和系统性。基于以上考虑,中文医学知识图谱 CMeKG 的构建流程如图 1 所示。

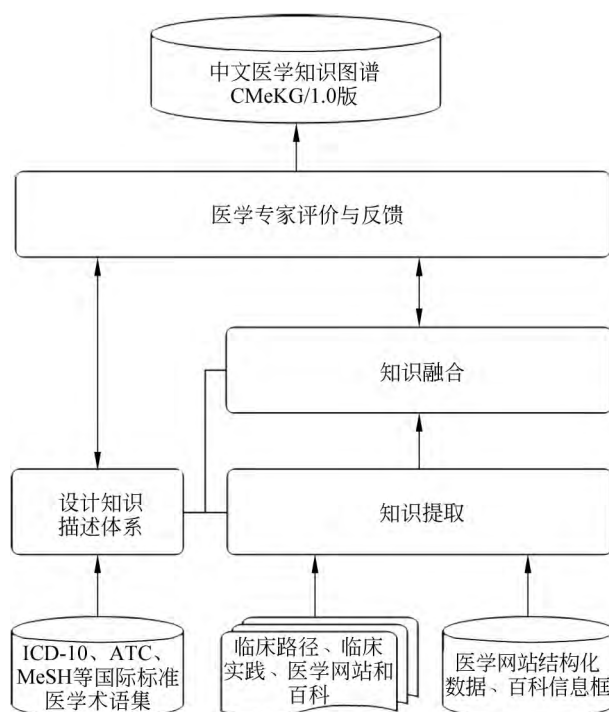


图 1 CMeKG 构建流程

首先,我们参考国际标准,在医学专家的指导下设计医学知识图谱模式层的规范体系。之后,在算法自动提取及人工标注、校对的基础上,整合提取医学概念关系实例,并进行实体对齐和归一化处理。在上述步骤完成后,根据医学专家的评价和反馈,迭代地修正医学知识图谱。最终,形成中文医学知识图谱 CMeKG 1.0,并在可视化平台上展示。

3 医学知识描述体系设计

CMeKG 主要由疾病、药物和诊疗技术类概念及其关系和属性组成。CMeKG 的医学知识抽象描述框架如图 2 所示,图 2 中以支气管肺癌为例,描述了对应的主要实体关系和属性。

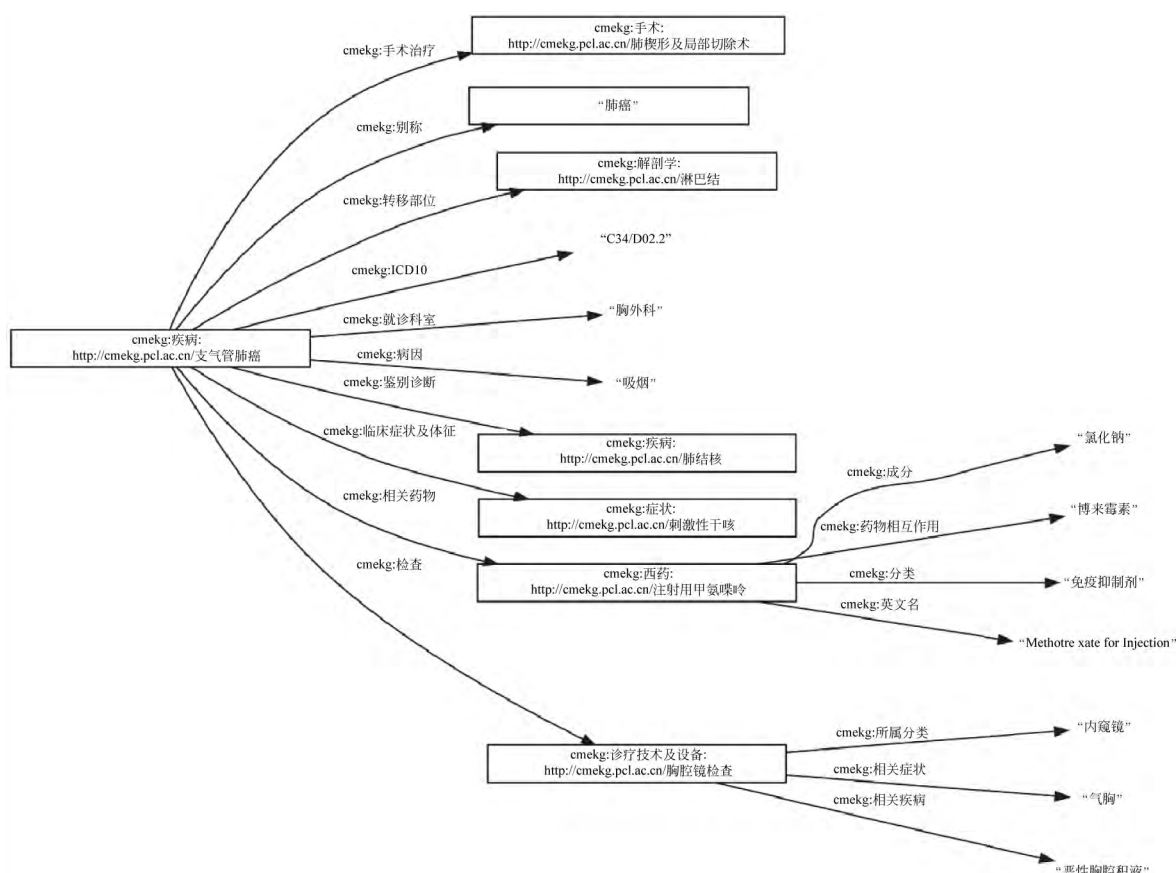


图 2 医学知识抽象描述框架

3.1 医学概念分类体系

概念分类体系的专业性和权威性对医学知识图谱的质量具有十分重要的作用,本文参考 ICD-10、ATC、MeSH 等国际标准医学术语集,将 CMeKG 的概念层设计为 15 大类,如表 1 所示。

表 1 概念分类体系规范

概念分类	参考标准	样例
A 解剖学	MeSH	半月板,胫骨
B 有机体	MeSH	革兰氏阴性菌
C 疾病	ICD-10	颅骨血管瘤
D 药物	ATC	氟化钠
E 诊疗技术及设备	MeSH	血管成形术,气囊,激光辅助
F 心理学和精神病学	MeSH	单亲
G 生物科学	MeSH	三核苷酸重复扩增
H 自然科学	MeSH	结肠镜检查,计算机断层摄影
I 社会科学	MeSH	殡葬仪式

续表

概念分类	参考标准	样例
J 工艺学,工业及农业	MeSH	酒精性饮料
K 人文科学	MeSH	生活质量
L 信息科学	MeSH	信息管理
M 人	MeSH	医院护理人员
N 医疗保健	MeSH	卫生保健费用
Z 地理名称	MeSH	加拿大

3.2 医学概念的关系描述框架

为了更丰富、更精准地描述不同种类的医学知识,基于上述医学概念分类体系,我们针对疾病、药物和诊疗技术及设备等各类医学概念进行细化描述,定义了各类概念的关系描述框架。每个概念的关系描述框架由概念间的关系(概念关系)和概念与属性之间的关系(属性关系)构成。我们共定义了 67 种概念关系,例如,〈药物类—症状类—适应症〉、〈疾病类—诊疗技术及设备类—检查〉,以及 194 种属性关系,用来描述某个概念实例的属性值(数字

或字符串),如同义词、规格、成分、发病年龄、住院时间等。表 2 以支气管肺癌为例,展示了疾病类概念的关系描述框架。

对于其他类概念的主要描述信息,简要列举如下:

- 药物类西药子类实体间常见关系: OTC 类型、不良反应、分类、商品名、性状、成分、英文名、药品监管分级、药品类型、规格、适

应症

- 药物类中草药子类实体间常见关系: 主治、入药部位、分布区域、别称、功效、属、性味、毒性、界、目、种、科、纲、贮藏、采集时间
- 诊疗技术及设备类实体间常见关系: 就诊科室、所属分类、相关疾病、相关症状、英文名

表 2 疾病类概念实体的关系描述框架

关系类型	关系子类型	关系取值	样例
语义	ICD 编码	字符串	C34、D02. 2
	同义词、别称	字符串	肺癌,支气管肺癌,支气管癌
疾病—部位	转移部位	解剖学类	淋巴结,锁骨上淋巴结,纵隔淋巴结
	外侵部位		心脏,交感神经
	发病部位		左肺,右肺
疾病—症状	临床症状	症状类	刺激性干咳,血痰/痰中带血,咯血
	侵及周围组织转移的症状		声音嘶哑,面部水肿,颈部水肿
疾病—检查	辅助检查	诊疗技术及设备类	胸部影像学检查,纤维支气管镜,肺穿刺活检
	影像学检查		X 线胸片,CT,磁共振成像
	内窥镜检查		支气管镜检查,经支气管针吸活检术
	筛查		低剂量 CT,常规胸片
	实验室检查		细胞学检查,剖胸探查术,ECT 检查
	其他检查		胸腔穿刺术,胸膜活检术,浅表淋巴结及皮下转移结节活检术
疾病—疾病	鉴别诊断	疾病类	肺结核,肺部感染,肺部良性肿瘤
	并发症		肺源性骨关节增生症,异位激素分泌综合症,异位促肾上腺皮质激素分泌综合症
	病理分型		小细胞肺癌,非小细胞肺癌,肺腺癌
疾病—其他治疗	放射治疗	诊疗技术及设备类	根治性放疗,姑息放疗,辅助放疗
疾病—手术治疗	手术治疗	诊疗技术及设备类	肺楔形及局部切除术,肺段切除术,肺叶切除术
疾病—药物治疗	药物治疗	药物类	血管内皮抑素,靶向治疗药物,多西紫杉醇
疾病—流行病学	多发群体	人类	吸烟者、慢性肺疾病者,接受过量放射线者
疾病—其他	就诊科室	数值、字符串	胸外科,肿瘤科
	预防		禁止吸烟,控制吸烟,保护环境
	转移方式		直接扩散,血行转移,淋巴道转移
疾病—社会学	病因	有机体类、字符串	吸烟,职业暴露,环境污染
	风险评估因素		吸烟史(现在和既往),氡暴露史,职业史,患癌史
	高危因素		吸烟指数>400,年龄>45 岁,环境与职业因素

4 医学知识提取技术

4.1 数据来源

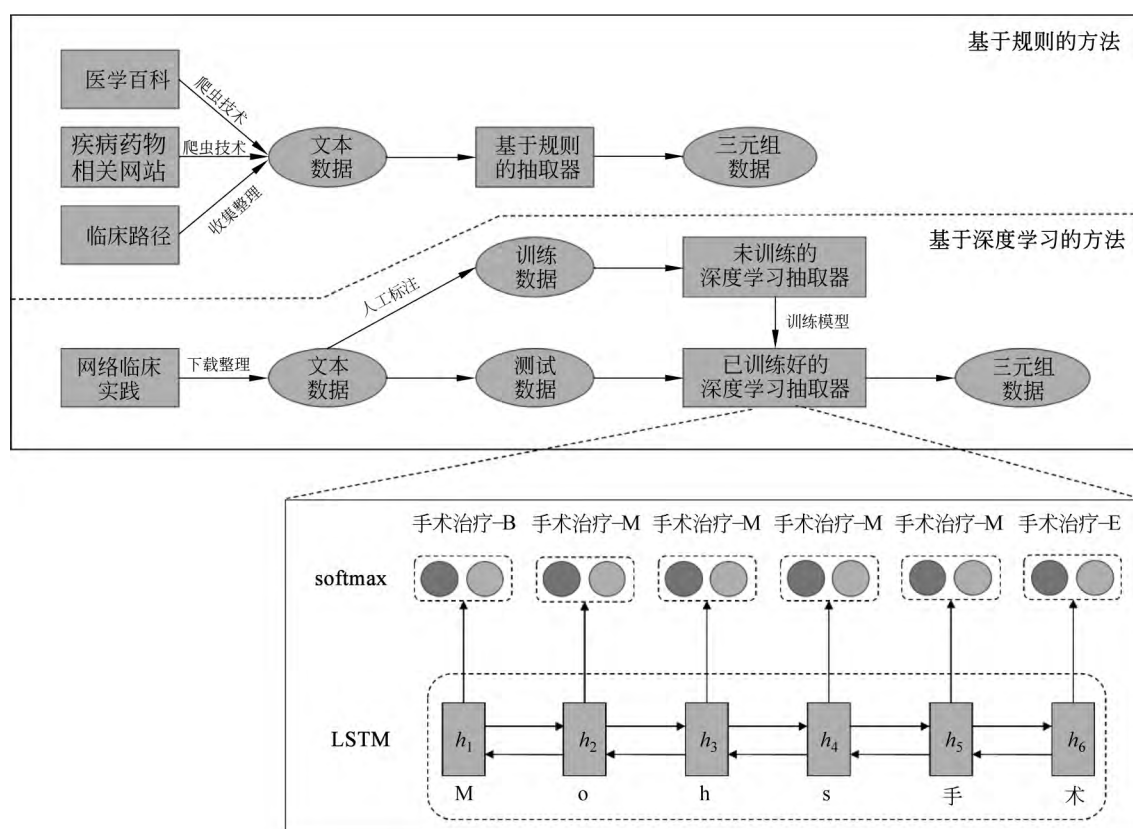
我们在知识提取的过程中根据知识来源的权威性排序,分为:国际/国家医学标准术语集、医学教材、临床路径指南、临床实践文件和医学百科。CMcKG 构建过程中使用了 200 余个权威的临床路径文档、900 余个临床实践文件以及百科类网络资源等多源异构的医学文本数据。

医学标准术语集和医学教材通常是国际或国家级机构统一规定的资源,其优点是权威性高、数据公开可靠。临床路径是针对某一种疾病建立的一套标准化治疗模式与治疗程序,是一个有关临床治疗的综合模式,以循证医学证据和指南为指导来促进治疗组织和疾病管理的方法。临床路径通常包含医院

处理某一疾病时的标准指导和流程,具有专业性和权威性。临床实践是面向临床医生和患者提供特定临床情况处理和指导的临床决策支持工具。它将最新的研究成果、诊断步骤、治疗步骤、指南、证据、专家意见整合在一起,为实际临床工作及疑难情况提供可靠的信息。

4.2 自动构建技术

不同的医学文本具有不同的形式特点。基于文本特点,分别使用了基于规则和基于深度学习的两种方法对多来源医学文本信息进行知识提取,本文主要涉及提取医学概念关系三元组信息。自动提取技术框架如图 3 所示,其中利用基于规则的方法提取临床路径、医学网站和医学百科中的医学知识,利用基于深度学习的方法来提取临床实践中的医学知识。



基于规则的知识提取 对于临床路径、医学网站和医学百科文本,我们缺乏标注数据,所以难以运用监督学习的方法来从中自动提取三元组。但是,这些数据并非是纯粹的文本,我们在从网络上爬取它们的过程中,同时也获取了它们的半结构化信息,

例如,段落层级、标题和小标题信息等。实践发现,在这些半结构化信息的帮助下可以构建一个基于规则的自动提取系统,针对不同来源文本的不同特征,构造出多样化的规则来对其进行概念关系的提取。我们对医学百科数据构造了 38 类规则,对医学网站

数据构造了 28 类规则,对临床路径数据构造了 17 类规则,每一类规则由数条更具体的规则所组成,最后在包含了 83 类、数百条规则的规则系统下,我们从临床路径、医学网站和医学百科文本中提取出了百万量级的概念关系三元组。通过抽样的人工评测,评估了基于规则的知识提取的精确率,对于从临床路径、医学网站和医学百科中提取出来的三元组,其精确率分别达到了 97%、96% 和 94%。

基于深度学习的知识提取 随着基于规则提取的三元组数据和人工标注的临床实践数据的积累,我们拥有了足够丰富的资源,已在临床实践数据上实施基于深度学习的医学知识提取。从文本中抽取三元组是一个被研究已久的任务,目前已经存在许

多三元组抽取的方法,主要分为两类:一类是先行实体识别再进行关系分类的 pipeline 方法;另一类是同时进行实体和关系抽取的联合抽取方法。但无论是哪种方法,在现有数据集上的实验结果都达不到令人满意的效果,作为医学知识图谱中使用的三元组数据,其精确度一定要有较高程度的保障,否则可能会引来许多问题。我们的任务是从一篇文档中抽取三元组,不同于以往的三元组抽取,其特殊性在于,数据中的一篇文档总是围绕着一个疾病实体来展开的,所以本质上只需要抽取另一个实体以及两个实体之间的关系。如图 4 所示,对于围绕“皮肤鳞状细胞癌”疾病的文档,我们只需要抽取“Mohs 手术”实体以及“手术治疗”的关系即可。

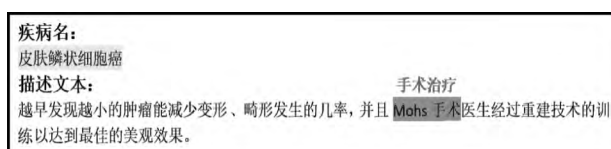


图 4 关系提取任务图示

考虑到这样的特殊性,我们提出了一套新的基于标注的三元组抽取方法。我们采取“关系名+BMES”的标注方式,如果某个实体与文档描述的疾病之间存在关系 R,则用“R-B”来标注该实体的第

一个字,“R-E”标注该实体的最后一个字,“R-M”标注该实体中间的字。如果该实体本身只有一个字,则用“R-S”来标注该实体。其他无关的字我们用“O”来标注。标注方式的示例如图 5 所示。

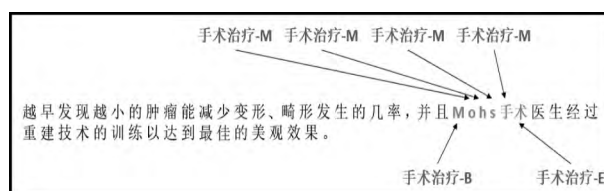


图 5 “关系名+BMES”标注方式图例

在这样的标注方式下,我们采用了 BiLSTM + softmax 分类模型来预测每个字的标签,先用双向 LSTM 得到每个字的隐向量,然后将该隐向量通过 MLP 和 Softmax 层,使其变为一个 c 维的概率向量,其中 c 为标签种类数。训练时,我们首先按照“关系名+BMES”的标注方法将训练集中的数据进行标注,然后将其送入模型进行学习。测试时,我们首先预测出每个字的标签,然后对标签进行匹配,即每找到一个配对同一关系 R 的完整的“BMES”集合,我们便取出这个集合对应的实体 E,形成一个(疾病名,R,实体 E)的三元组。我们的模型结构如图 3 下方展开的方框部分所示。

我们用该方法在临床实践数据中提取了 20 余

万条三元组数据,通过抽样的人工评测评估了基于深度学习的知识提取的精确率,其精确率达到了 89.2%。这样高精确率的自动提取方法在大幅提升知识库构建效率的同时,能够保证知识库内容的高度可靠性。

5 中文医学知识图谱 CMeKG 1.0

我们在医学专家的配合下,制订了医学知识描述体系,收集了多来源的大规模医学文本,通过人工标注和自动提取方法相结合,构建了中文医学知识图谱 CMeKG。CMeKG 1.0 版包括 6 310 种疾病、19 853 种药物、1 237 种诊疗技术及设备,涵盖疾病

的临床症状、发病部位、药物治疗、手术治疗、鉴别诊断、影像学检查、高危因素、传播途径、多发群体、就诊科室等以及药物的成分、适应症、用法用量、有效期、禁忌证等常见关系类型,关联到的医学实体达 20 余万,CMeKG 目前的概念关系实例及属性三元组达 100 余万。同时,基于以上医学实体和关系三元组,我们开发了 CMeKG 的展示平台,网址为:
<http://cmekg.pcl.ac.cn/>。

CMeKG 使用百度开源可视化库 Echarts^[16] 展示知识图谱,平台展示内容主要包括树状结构和网状图谱两部分,提供的功能包括实体的搜索、匹配、图谱的链接和遍历。对于每一个实体,选择以该实体为主语的三元组进行显示,连接同一节点的相同颜色节点代表相同的语义关系,整体效果呈现为以查询实体为中心,具有语义关系的相关实体发散至四周的网状结构,如图 6 所示。

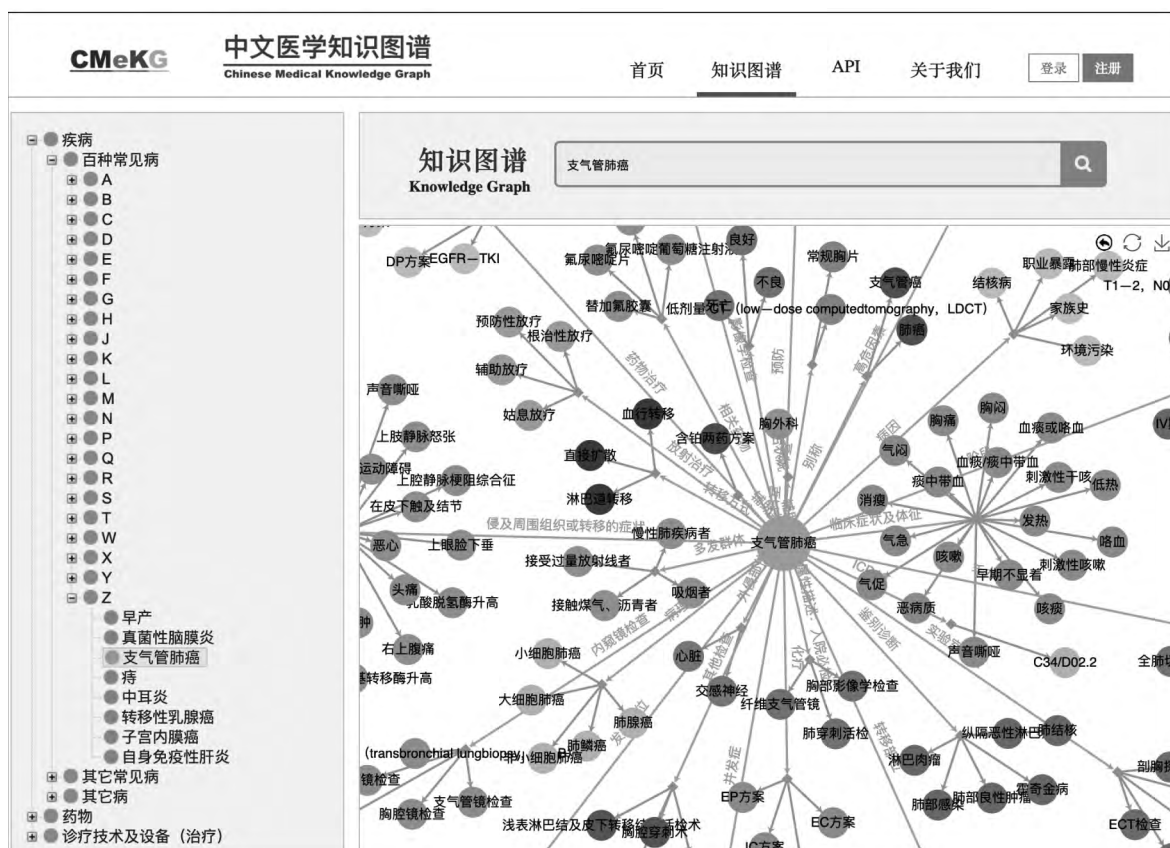


图 6 CMeKG 结果展示图

6 结束语

我们对医学知识图谱的构建研究才刚刚起步,本文综述了 CMeKG 1.0 构建过程中的描述体系、关键技术、构建流程以及初步的构建结构。我们认为,未来服务于智慧医疗的知识图谱应该满足以下方面的要求:知识来源具有权威性、知识组织具有系统性、知识内容具有丰富性、知识描述具有精准性、知识更新具有高效性。因此,医学知识图谱的构建任重而道远,在未来我们将朝着构建大规模、高质量、融合文本、图像与视频等多模态信息的医学知识图谱的方向继续努力。

致谢

感谢郑州大学第一附属医院牛承志老师、郑州大学第三附属医院赵悦淑主任及北京大学医学部詹思延教授给予的指导建议!

参考文献

- [1] Singhal A. Introducing the knowledge graph: things, not strings [EB/OL]. Official google blog, 2012. <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>.
- [2] Fabian M, Suchanek Gjergji, Kasneci Gerhard Wei-

- kum. YAGO: A large ontology from Wikipedia and WordNet[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2008(3): 203-217.
- [3] Christian Bizera, Jens Lehmannb, Georgi Kobilarova, et al. DBpedia-A crystallization point for the Web of data[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2009(07): 154-165.
- [4] Bo Xu, Yong Xu, Jiaqing Liang, et al. CN-DBpedia: A never-ending Chinese knowledge extraction system [J]. Lecture Notes in Computer Science, vol 10351. Springer, Cham. 2017(10351): 428-438.
- [5] Zhigang Wang, Juanzi Li, Zhichun Wang, et al. XLORE: A large-scale English-Chinese bilingual knowledge graph[C]//Proceedings of the 12th International Semantic Web Conference (ISWC 2013) on Posters & Demonstrations Track.
- [6] World Health Organization. International Statistical Classification of diseases and related health problems. (10th Revision)[OL] <https://icd.who.int/browse10/2016/en>.
- [7] Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases [J]. Journal of the American Medical Informatics Association, 2001, 8(1): 80-91.
- [8] Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth [J]. Studies in Health Technology and Informatics, 2006, 121(121): 279-90.
- [9] Natalya F Noy, Nigam H Shah, Patricia L Whetzel, et al. BioPortal: Ontologies and integrated data resources at the click of a mouse [J]. Nucleic Acids Research, 2009, 2009(37): 170-173.
- [10] FrancBois Belleau, Marc-Alexandre Nolin, Nicole Tourigny, et al. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems [J]. Journal of Biomedical Informatics, 2008, 2008 (41): 706-716.
- [11] 贾李蓉, 于彤, 崔蒙, 等. 中医药学语言系统研究进展 [J]. 中国数字医学, 2014, 9(10): 57-62.
- [12] 贾李蓉, 刘静, 于彤, 等. 中医药知识图谱构建 [J]. 医学信息学杂志, 2015, 36(8): 51-59.
- [13] 阮彤, 孙程琳, 王昊奋, 等. 中医药知识图谱构建与应用 [J]. 医学信息学杂志, 2016, 37(4): 8-13.
- [14] Tong Ruan, Mengjie Wang, JianSun, et al. An automatic approach for constructing a knowledge base of symptoms in Chinese [J]. Journal of Biomedical Semantics, 2017, 8(33): 71-79.
- [15] 刘燕, 傅智杰, 李姣, 等. 医学百科知识图谱构建 [J]. 中华医学图书情报杂志, 2018, 27(6): 28-34.
- [16] Deqing Li, Honghui Mei, Yi Shen, et al. ECharts: A declarative framework for rapid construction of web-based visualization[J]. Visual Informatics, 2018, 2(2): 136-146.
- [17] ATC/DDD index of 2019 [EB/OL]. WHO collaborating Centre of Drug Statistics Methodology, 2019. https://www.whocc.no/atc/structure_and_principles/.
- [18] Nelson S J, Johnston W D, Humphreys B L. Relationships in medical subject headings (MeSH) [M]. Relationships in the Organization of Knowledge. Springer Netherlands, 2001: 171-184.



奥德玛(1994—), 硕士研究生, 主要研究领域为知识图谱构建与融合。

E-mail: odmaa_b@pku.edu.cn



穗志方(1970—), 通信作者, 博士, 教授, 主要研究领域为计算语言学、文本知识工程。

E-mail: szf@pku.edu.cn



杨云飞(1995—), 硕士研究生, 主要研究领域为知识图谱与信息抽取。

E-mail: yfy@pku.edu.cn