

基于临床指南的知识图谱构建技术研究

尹梓名, 杜方芮, 赵紫彤, 贾玉娇, 田 甜, 徐 丹

(上海理工大学 医疗器械与食品学院, 上海 200093)

摘 要: 随着人工智能的迅猛发展以及智慧医疗的提出, 具备高关联性, 高结构化的医学知识图谱成为智慧医疗的研究热点。然而基于互联网大数据构建的知识图谱, 其知识质量难以保证, 难以适用于具有高专业、高要求、高准确度的临床辅助诊疗场景。本文提出一种基于临床指南的知识图谱构建方法, 着重在知识源头进行优化, 主要对其中包含的陈述性知识和流程性知识进行建模。最后以非小细胞肺癌与冠心病两类疾病为例, 分别构建了疾病知识图谱, 验证了该方法的可用性。

关键词: 知识图谱; 临床指南; 本体; Neo4j

中图分类号: TP391.5 **文献标识码:** A **DOI:** 10.3969/j.issn.1003-6970.2020.09.047

本文著录格式: 尹梓名, 杜方芮, 赵紫彤, 等. 基于临床指南的知识图谱构建技术研究[J]. 软件, 2020, 41 (09): 178-184+197

Research on Knowledge Graph Construction Technology Based on Clinical Guidelines

YIN Zi-ming, DU Fang-rui, ZHAO Zi-tong, JIA Yu-jiao, TIAN Tian, XU Dan

(School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200090, China)

[Abstract]: With the rapid development of artificial intelligence and the proposal of intelligent medicine, medical knowledge graph with high relevance and high structure has become the research hotspot. However, the knowledge graph based on Internet big data is difficult to ensure the quality of knowledge, and it is difficult to apply to the clinical auxiliary diagnosis and treatment scene with high professional, high requirements and high accuracy. This paper proposes a construction method of knowledge graph based on clinical guidelines, focusing on knowledge optimization at the source of knowledge, mainly modeling the declarative knowledge and process knowledge. Finally, the knowledge graph of non-small cell lung cancer and coronary heart disease was constructed to verify the availability of this method.

[Key words]: Knowledge graph; Clinical guidelines; Ontology; Neo4j

0 引言

随着人工智能的迅速发展以及智慧医疗的提出, 涉及到知识抽取、表示、融合、推理、问答等内容的知识图谱成为知识服务领域的一个研究热点。作为目前人工智能领域最前沿且最重要的一种知识表示形式, 知识图谱基于语义网络, 提供从“关系”的角度去分析问题的能力, 从而构建高关联性, 高结构化的知识模型。目前, 现有的面向互联网应用场景的通用知识图谱构建技术^[1-5], 虽然能够快速、大量地建立知识图谱, 但是基于网络数据建立的知识图谱, 其准确性和可靠性存疑, 知识质量难以保证, 形成的知识图谱未必完全正确, 难以适用于具有高准确度的临床辅助诊疗场景, 且这种方法生成的知识图谱冗余度较大^[6-7]。所以如何从各种专业的医学资料中构建知识图谱是医学知识图谱研究的重点内容。临床指南 (Clinical Guidelines) 是基于循证医学的观点系统开发的多组临床指导意见, 用于帮助医生和病人针对特定的临床问题做出恰当的处理, 减少医疗差错, 避免资源浪费^[8]。

因此, 基于临床指南构建用于辅助临床决策的医学知识图谱是一种值得研究的方法。本论文提出一种基于临床指南构建知识图谱技术, 围绕流程性知识与陈述性知识两种临床指南知识表达方式, 建立结构化本体库和知识图谱模型。该方法着重在知识源头对知识进行优化, 设计并调整概念间的关系结构, 对实例与实例间的联系进行细化完善, 保证数据的可用性与可靠性。依据此方法, 本文以美国 NCCN 公布的非小细胞肺癌临床诊疗指南和中华医学会公布的冠心病诊疗指南为例, 构建了这两种疾病的知识图谱, 并验证了其可行性, 为医疗智能辅助决策提供了一种高质量构建知识图谱的方法。

1 基于临床指南的知识图谱构建

1.1 概念层 (本体层)

临床指南由大量临床概念、术语及逻辑关系组成。构建基于临床指南的知识图谱, 首先应从指南中抽象出其概念层次, 将指南结构化。本体设计是知识表达的一种方式, 相当于知识图谱的结构层次构建, 突出

基金项目: 国家自然科学基金资助(批准号: 81801797), 国家重点研发计划资助(批准号: 2020YFC2005801)

作者简介: 尹梓名(1986-), 男, 博士, 讲师, 研究方向: 医学人工智能, 医疗大数据分析; 杜方芮(1999-), 女, 本科生, 专业: 医学信息工程; 赵紫彤 (2000-), 女, 本科生, 专业: 医学信息工程; 贾玉娇(1999-), 女, 本科生, 专业: 医学信息工程; 田甜(1998-), 女, 本科生, 专业: 医学信息工程; 徐丹(1999-), 女, 本科生, 专业: 医学信息工程。

强调概念及概念间的关系，通过本体形成的知识库不仅层次结构较强，而且冗余度较小，是知识图谱概念层构建必不可少的环节。因此，可通过对指南进行知识抽取，提取指南的概念层内容，进而进行本体设计。

1.1.1 概念结构表设计

文本指南中的知识主要由陈述性知识和流程性知识两大类构成^[9]。流程性知识关注的是陈述性知识的推理过程及推理结果。如何从不同格式的知识中提取概念层次内容，是概念层设计的核心部分。从临床指南的角度来看，我们关注概念模型的三个内容：概念内容，概念实例，以及概念间的逻辑关系。

(1) 概念内容：概念内容是对指南中属于同一类医学知识的归一化过程。

(2) 概念实例：是概念内容的子集，需要通过命名实体识别技术来完成。应注意，概念实例的提取可与数据层的实体抽取同步进行。

(3) 概念间的逻辑关系：描绘两个概念内容的关系，在流程性知识中通常以有向图的指向关系来表示，在陈述性知识中需要通过语句语义来确认概念间的逻辑关系。

概念层设计相当于知识图谱的架构设计，是知识图谱结构化、规范化的根基。因此，将概念内容按上下位关系，整理成疾病概念结构表，便于知识图谱构建的理解与运用，如表 1 所示。

表 1 疾病概念结构表
Tab.1 Disease conceptual structure table

一级概念层	二级概念层	三级概念层	实例
概念内容 1	概念内容 4	概念内容 7	概念实例 1, 概念实例 2, 概念实例 3
概念内容 2	概念内容 5	概念内容 8	概念实例 4, 概念实例 5, 概念实例 6

1.1.2 概念关系表设计

要实现指南结构化的概念整合，更重要的一步是使这些概念层在语义上形成关联，以构成一个完整的指南结构体。确定概念层之间的关系主要通过两种手段来实现：确定同类型医学概念关系，确定不同类型医学概念关系。同类型医学概念关系的确定主要体现在实体的上下位关系。概念结构表确定了概念层的上下位关系后，这一步主要关心不同类型医学概念关系。

不同类型医学概念关系的确定以类与类之间的映射关系为核心，是对概念间逻辑关系的归纳。由于疾病的治疗普遍包含检查、评估、治疗、用药、监测等步骤，我们设计了一张以疾病知识图谱为使用对象的通用概念关系表，不同疾病知识图谱的概念关系可在此表的基础上进行设计使用，也可对特定关系进行增补。

1.2 数据层（实体和关系）

构建指南数据层的目的是从不同格式的临床指南中提取医学知识，并将医学知识转换成可被计算机处理的数据格式。知识图谱数据应用的前提是，这些数据具有规范化的数据模型和物理模型。数据模型是对数据的有

表 2 实体关系表
Tab.2 Entity relationship table

实体关系名	起始实体	指向实体	关系解释
has_class_of	E_1	E_2	表示实体 E_2 是实体 E_1 的一个子类
has	E_1	E_2	表示实体 E_1 包含实体 E_2
has_treatment	E_1	E_2	表示实体 E_2 是实体 E_1 的治疗方法
has_assessment	E_1	E_2	表示实体 E_2 是实体 E_1 的评估方法
has_detection	E_1	E_2	表示实体 E_2 是实体 E_1 的检测方法
Detection_result	E_1	E_2	表示实体 E_2 是实体 E_1 的检测结果
has_monitor	E_1	E_2	表示实体 E_2 是实体 E_1 的监测方法
assess_result	E_1	E_2	表示实体 E_2 是实体 E_1 的评估结果
surgery_see	E_1	E_2	表示实体 E_1 手术可见实体 E_2
Suggest	E_1	E_2	表示实体 E_1 的指南建议是实体 E_2
drug_use	E_1	E_2	表示实体 E_1 的使用药品是实体 E_2
belong_to	E_1	E_2	表示实体 E_2 属于实体 E_1
refer_to	E_1	E_2	表示实体 E_2 是实体 E_1 的参考项

效表示，是从人的角度对知识图谱的数据进行描述，包含三元组模型和图模型两种模型。物理层面的存储是从计算机的角度对数据进行存储^[10]，包含关系表存储和图存储两者存储方式。数据模型和物理模型的合理化构建关系到知识图谱的有效性利用。对于数据层的构建，我们利用三元组数据模型来描述指南知识，通过知识抽取，知识存储两个内容，将指南知识有效地存储在图数据库管理系统 Neo4j 中。

1.2.1 知识抽取

医学临床指南中含有大量医学临床术语及指南流程，临床指南的知识抽取就是从临床指南中抽取与概念层设计相匹配的实体和关系，并将其整理成实体库和三元组关系库。知识抽取主要包括两方面内容：

(1) 医学实体抽取：从指南中识别命名实体，并将其分类到定义类别。

(2) 医学关系抽取：从指南中识别并抽取实体和实体之间的关系。关系必须是三元组形式，即实体-关系-实体。整理出的多元关系如图所示。

表 3 三元组关系表
Tab.3 Triad relation table

概念内容 1	关系	概念内容 2
实体 1	关系名称	实体 2
实体 1	关系名称	实体 3

在实体抽取和关系抽取时，亦可针对不同实体类设定相关属性。例如药品类别含有药品名称，药品英文名称，药品类别等属性。

1.2.2 知识存储

明确了所有实体以及实体的关系后，需要设计数据的物理模型，对知识进行存储，形成知识图谱。选择存储体系时需考虑知识可靠、可融合、可应用等问题。为表示多样化，复杂化的医学实体关系，我们选用基于 Neo4j 图数据库作为存储体系。Neo4j 是基于 Java 开发的开源原生数据库，是一个兼备高性能、嵌入式、轻量级等优势 NoSQL 图数据库。基于抽取

的三元组数据模型,设计导入的医学节点和关系节点,考虑到疾病临床指南所整理的数据为中等规模以及数据可实时插入的效果,采用导入Csv文件为主的方式,对数据进行存储。

2 知识图谱的应用案例

为验证上述方法的可行性,本文以非小细胞肺癌和冠心病的临床指南为例,依照本文所述方法分别建立两种疾病的知识图谱。

2.1 案例一:非小细胞肺癌知识图谱构建

近10多年来,恶性肿瘤发病人数成增长趋势。美国癌症协会(ACS)在《2020年癌症统计》上预估了2020年美国新发癌症病例数和死亡人数,统计显示,在所有癌症死亡中四分之一是由于肺癌所致,高于乳腺癌、前列腺癌和结直肠癌三者病例。基于上述知识图谱构建方法,我们构建了非小细胞肺癌临床本体以及可推理的知识图谱,从而实现非小细胞肺癌的临床辅助决策。

在当今中国的肿瘤防治领域,为实现肿瘤治疗的规范化、科学化,普遍倡导以美国国家综合癌症网络(NCCN)发布的临床实践指南为首选的循证学依据^[11]。该指南以有向流程图的形式,描述了非小细胞肺癌的临床诊治过程,例如根据各个疾病分期或临床表现给出对应的评估,治疗流程,如图1所示。



图1 NCCN非小细胞肺癌临床实践指南(部分)
Fig.1 NCCN clinical practice guidelines of non-small cell lung cancer (part)

2.1.1 概念层设计

首先,对图1所示的部分临床指南进行概念层提取。第一步,提取概念内容,分别为“临床表现”,“初始治疗”,“辅助治疗”和“治疗方法”四个内容。“治疗方法”在有向图中并未提示,但通过归纳可知,“手术”,“化疗”,“术前同步放化疗”均为一种“治疗方法”。第二步,提取概念实例,分别为“肺上沟瘤(T3侵犯, N0-1)”,“术前同步放化疗”,“手术”,“化疗”四个实例,此步骤要尽可能多的挖掘出临床指南中高质量的医学词汇。第三步,提取概念间的逻辑关系,分别为“临床表现的初始治疗是(某一治疗方法)”,“临床表现的辅助治疗是(某一治疗方法)”。通过对临床指南内容的整理归纳(例如由“初始治疗”,“辅助治疗”,“后续治疗”挖掘出上级概念层“治疗时期”),整理出非小细胞肺癌本体的三级概念结构,如表4所示。

其次,进行概念关系层设计,基于提取的概念间的逻辑关系,可将语义关系表设计如下表5所示。

2.1.2 数据层设计

数据层设计主要是从临床指南中挖掘出具有高质量的医学实体(概念实例)和关系,并将其整理成三元组的格式,存储在Neo4j数据库中。实体关系必须以概念层设计的关系表为框架,实现实体与实体之间

表4 非小细胞肺癌临床指南本体概念结构表
Tab.4 The ontology concept structure table of non-small cell lung cancer clinical guidelines

一级概念层	二级概念层	三级概念层	实例
临床表现			疑似多发性肺肿瘤,多发性肺癌(N0-1),偶发结节疑似肺癌等
临床分期			I期(中央型T1abcT2a, N0), IVB期 弥散性转移分散的肺结节(IIB,IIIA,IV期)等 无症状冠心病,心脏X综合征,心肌桥
治疗	治疗方法	药物治疗	{药品}
		手术	保留肺实质切除术, SVC 支架, 栓塞……
		放疗	完全根治性放疗, 立体定向消融放疗(SABR), 外照射放疗……
		化疗等	诱导化疗, 序贯化疗, 序贯化疗……
	治疗时期	初始治疗	{治疗方法}
		辅助治疗	{治疗方法}
		复发与转移的治疗	{治疗方法}
		一线治疗	{治疗方法}
		后续治疗	{治疗方法}
		继续维持治疗	{治疗方法}
评估	评估时期	风险评估	{评估方法}
		初始评估	{评估方法}
		治疗前评估	{评估方法}
		临床评估	{评估方法}
	评估方法	患者因素	年龄, 吸烟史, 既往肿瘤病史, 家族史, 职业暴露, 其它肺部疾病等
		放射性因素	肺结节的大小, 肺结节的形态, 肺结节的密度, 相关的实质异常等
监测	监测频率	其他	病史问诊, 体格检查, 肺功能检查, FDG PET/CT 扫描, 胸腔镜等
		监测频率	每6个月一次, 每年一次, 每3-6月一次等
		监测方法	戒烟建议, 胸部低剂量平扫CT, 增强CT等
	监测结果	监测结果	局部复发, 远处转移
		检测方法	分子检测, PD-L1 检测
	检测组织	检测组织	大细胞, 鳞状细胞癌, 腺癌, 非小细胞肺癌非特指型(NOS)等
药品	检测结果	检测结果	ALK 阳性, BRAF V600E 阳性, EGFR 敏感突变阳性等
			厄洛替尼, 贝伐珠单抗, 拉罗替尼等

表 5 本体设计中实体语义关系表
Tab.5 Entity semantic relation table in ontology design

实体关系名	起始实体类别	指向实体
has_class_of	疾病	临床分期/临床表现/评估/监测/检测/药品
has	临床分期	临床分期
	治疗	治疗时期, 治疗方法
	检测	检测方法, 检测结果, 检测组织
	评估	评估方法, 评估时期
	监测	监测方法, 监测频率, 监测结果
has_treatment	临床表现/临床分期/临床评估/检测结果	治疗方法
has_assessment	临床表现/临床分期/临床评估/检测结果	评估方法
has_detection	临床表现/临床评估	检测方法
detection_result	检测结果	临床表现/临床分期
has_monitor	临床分期	监测
assess_result	评估方法	评估结果
	临床分期	评估结果
surgery_see	评估结果	临床分期
Suggest	评估结果	治疗方法
	评估结果	(进一步) 评估结果
drug_use	临床分期/治疗方法	药品
refer_to	评估结果/检测结果/临床分期/临床表现	临床分期/临床表现
Belong_to	临床分期/临床表现/发现/评估/监测/检测/药品	其他可能类型

的关系连接。

如图 2 所示, 该部分指南表示的意思为“临床评

估为 IA 期 (周围型 T1abc, N0) 分期的病人, 需先进行治疗前评估, 评估方法有肺功能检查, 支气管镜, 纵隔淋巴结病理学评估, FDG PET/CT 扫描”, 故提取实体“IA 期 (周围型 T1abc, N0)”为临床分期, “肺功能检查”, “支气管镜”, “纵隔淋巴结病理学评估”和“FDG PET/CT 扫描”为评估方法, 概念内容“临床分期”与“评估方法”的关系为“has_assessment” (有…评估方法), 生成表 6 所示的三元组关系表。



图 2 《NCCN 肿瘤学临床实践指南 非小细胞肺癌》(部分)
Fig.2 NCCN clinical practice guidelines in non-small cell lung cancer (part)

表 6 非小细胞肺癌指南知识抽取 (部分)
Tab.6 Non-small cell cancer guideline knowledge extraction (part)

临床分期	评估方法	评估方法
IA 期 (周围型 T1abc, N0)	治疗前评估	肺功能检查
IA 期 (周围型 T1abc, N0)	治疗前评估	支气管镜
IA 期 (周围型 T1abc, N0)	治疗前评估	纵隔淋巴结病理学评估
IA 期 (周围型 T1abc, N0)	治疗前评估	FDG PET/CT 扫描

最后将实体与关系整理出 Neo4j Load csv 导入格式的实体与关系文件 (共导入 12 个实体文件与 18 个关系文件), 导入 Neo4j 数据库, 形成 284 个医学实体与 824 个医学关系, 如图 3、图 4 所示。

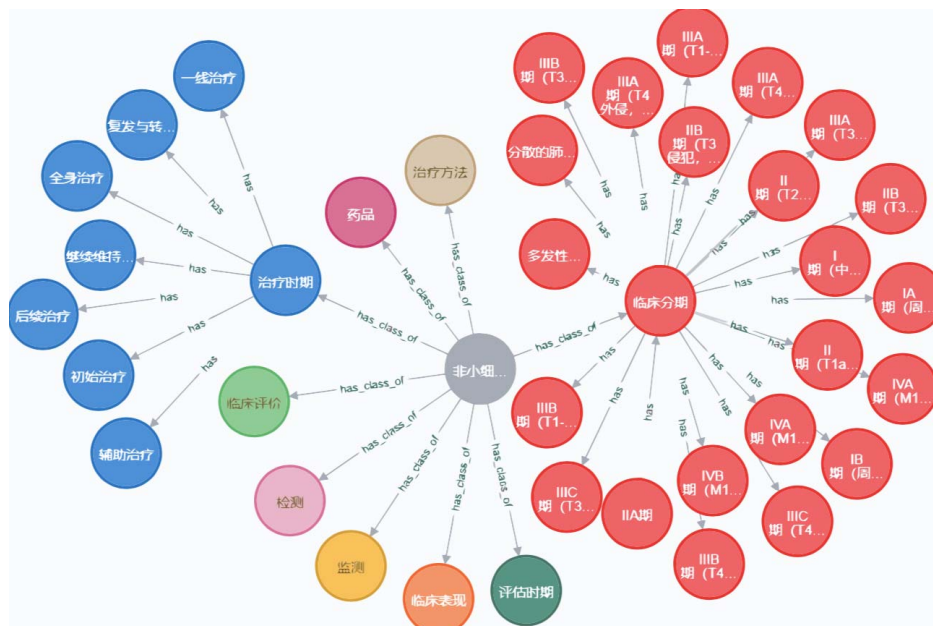


图 3 非小细胞肺癌知识图谱 (部分)
Fig.3 The knowledge graph of non-small cell lung cancer (part)

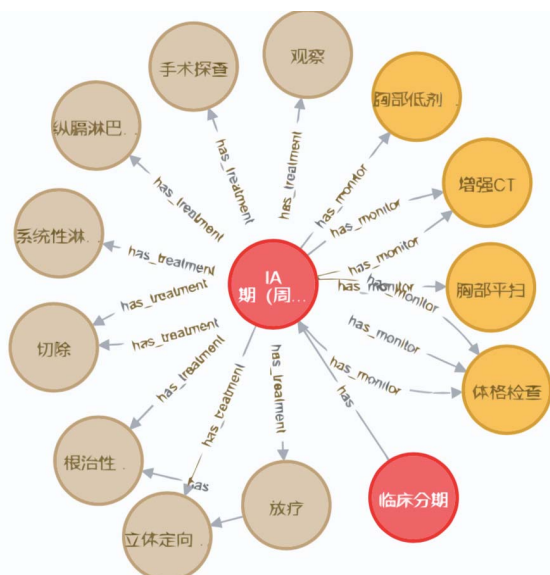


图4 非小细胞肺癌知识图谱 (IA 期 (周围型 T1abc, N0))
Fig.4 The knowledge graph of non-small cell lung cancer (Stage IA (peripheral T1abc, N0))

2.1.3 临床实例验证

非小细胞肺癌知识图谱可应用于临床辅助决策, 实现搜索与推荐功能。随着图数据规模的日益增大, 大图上的子图搜索问题变得极为重要^[12]。我们可以通过不同的方法来实现搜索功能, 例如基于规则的子图匹配, 基于节点与结构相似度的子图匹配等^[13]。本文通过规则匹配的方法, 实现了非小细胞肺癌系统的后台操作代码。

效果如下: 当用户点击“II 期 (T2b, N0)”临床分期和搜索条件“治疗方法”时, 系统会根据非小细胞肺癌知识图谱接收搜索条件内容, 通过规则匹配, 最后将其转换为 Cypher 语句可以在 Neo4j 中查询答案。结果如图 5 所示。

II 期 (T2b, N0) 的治疗方法有:
['纵膈淋巴结清扫', '根治性放疗', '根治性化疗', '手术探查', '系统性淋巴结采样']

图5 非小细胞肺癌验证案例
Fig.5 The validating test case of non-small cell lung cancer

非小细胞肺癌知识图谱的应用也可体现在病人的诊中和诊后的环节, 为医生提供诊断辅助决策。例如, 基于病人主诉, 医生评估病人的疾病分期为“A 期 (周围型 T1abc, N0)”, 临床决策支持系统可查询知识图谱, 知识图谱会针对该疾病分期给出不同的评估方法, 治疗方法与监测方法, 如图 6 所示。

2.2 案例二: 冠心病知识图谱构建

冠状动脉粥样硬化性心脏病是冠状动脉血管发生动脉粥样硬化病变而引起血管腔狭窄或阻塞, 造成心肌缺血、缺氧或坏死而导致的心脏病, 常常被称为“冠心病”。冠心病在美国和许多发达国家排在死亡原因的第一位。参照本文对流程性知识的构建方法, 以《临床诊疗指南·心血管分册 (2017 年修订版)》冠心病临床指南^[14]为基础, 对临床指南陈述性知识进行分析,

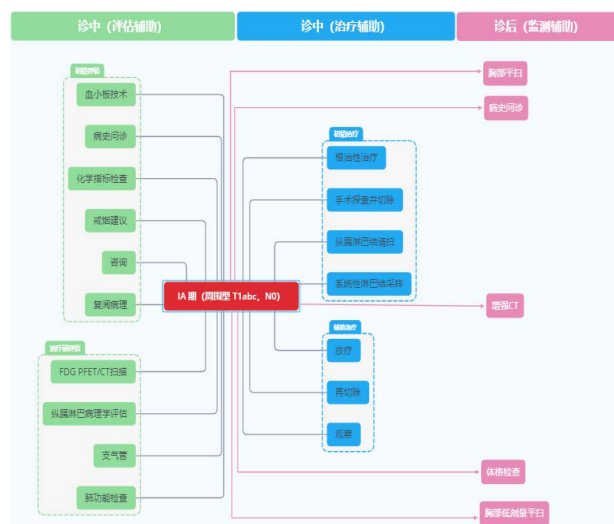


图6 非小细胞肺癌辅助决策 (IA 期 (周围型 T1abc, N0))
Fig.6 Assisted decision making for non-small cell lung cancer (stage IA (peripheral T1abc, N0))

构建冠心病临床本体以及知识图谱, 实现冠心病领域的临床辅助决策。

2.2.1 概念层设计

医学知识关系错综复杂并且医学知识抽取的精确度要求高, 为保证本体构建的正确性, 参考医学主题词表 (MeSH), 对冠心病领域内的基本概念和概念之间的相互关系进行分析, 通过这些关系和概念完成本体的构建。在理解 MeSH 词表和临床诊疗指南的基础上, 将冠心病知识本体分为患者病历、疾病类型、检测方法、检测结果、治疗方法、危险因素 6 大类, 知识层级结构如图 7 所示。

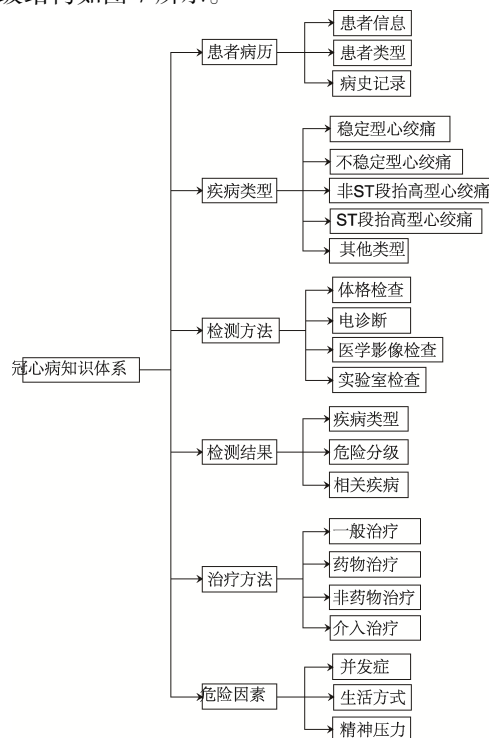


图7 冠心病知识层级结构
Fig.7 The knowledge hierarchy of coronary heart disease

在《临床诊疗指南·心血管分册（2017年修订）》的基础上，结合 MeSH 主题词表，对构建冠心病临床指南中概念与概念之间的关系进行详细的分析，整理出冠心病本体三级概念结构。具体概念结构如下表 7 所示。

参考非小细胞肺癌关系层设计方法，对相似的部分进行整合和复用，保证两类疾病构建知识图谱结构

上的统一性。确定冠心病同类型医学概念实体的上下位关系，并根据指南内容，以类与类之间的映射关系为核心提取不同类型的医学概念关系，以此整理出 8 种冠心病临床指南实体与实体之间的关系，分别为 has_class_of、has_treatment、has_assessment、has_detection、assess_result、belong_to、refer_to。冠心病临床指南的概念关系图下表 8 所示。

表 7 冠心病临床指南本概念结构表
Tab.7 Conceptual structure table of clinical guidelines for coronary heart disease

一级概念层	二级概念层	三级概念层	实例
患者病历	患者信息	姓名、性别、年龄、民族、婚否等	
	患者类型	一般人群、易患人群、冠心病患者	
	病史记录	既往史、个人史、家族史	
疾病类型	稳定型心绞痛		静息性心绞痛、初发心绞痛、恶化劳力型心绞痛
	不稳定型心绞痛		
	非 ST 段抬高型心肌梗死		
	ST 段抬高型心肌梗死		
	其他类型		无症状冠心病，心脏 X 综合征，心肌桥
检测方法	体格检查	视诊	
		听诊	心音、心率、心律、杂音、心包摩擦音
		触诊	心尖搏动、震颤、心包摩擦感
		问诊	寒热、出汗、胸部不适
	电诊断	心电图	
	医学影像检查	超声检查	负荷超声心动图、血管内超声（IVUS）
		CT 检查	CT 造影（CTA）、冠状动脉造影、核素心肌显像、多排螺旋 CT、电子束 CT
		核磁共振显像（MRI）检查	
	实验室检查		空腹血糖，血脂，糖耐量，血肌钙蛋白，肌酸激酶，肌红蛋白
检测结果	疾病名称	低危	
	危险分级	高危	
		中危	
	相关疾病		心脏病、心肌病、高血压、甲状腺亢进、严重贫血、冠状动脉痉挛、微血管病变、其他非心脏性疾病
治疗方法	一般治疗		持续心电监护、卧床休息、吸氧
	药物治疗	β 受体阻滞剂	普萘洛尔、美托洛尔、阿替洛尔、比索洛尔
		硝酸酯类药物	硝酸甘油
		镇痛剂	吗啡
		钙离子通道阻滞剂	
		抗血小板类药物	阿司匹林、血小板膜糖蛋白（GP）II b/IIIa 受体拮抗剂、氯吡格雷
		抗凝药物	肝素
		他汀类药物	
		抗焦虑剂	
		ACEI	
	介入治疗	经皮冠状动脉介入治疗（PCI）	溶栓、冠状动脉内支架术
	非药物治疗	饮食疗法	低盐低糖，合理膳食、戒烟戒酒
		运动疗法	控制体重、适时运动
		心理疗法	心理平衡、健康教育
		其他疗法	冠状动脉旁路移植术（CABG）、激光血运重建术、增强型体外反搏、脊髓电刺激
危险因素	并发症		高血压、高血脂
	生活方式		生活不规律
	精神压力		精神压力大

表 8 本体设计中实体语义关系表
Tab.8 Entity semantic relation table in ontology design

实体关系名	起始实体类别	指向实体
has_class_of	冠心病	患者病历/疾病类型/检测方法/检测结果/治疗方法/并发症
	患者病历	患者信息/患者类型/病史记录
	检测方法	体格检查/电诊断/医学影像检查/实验室检查
	治疗方法 危险因素	一般治疗/药物治疗/非药物治疗/介入治疗 并发症/生活方式/精神压力
has	疾病类型	疾病类型
	患者信息	姓名/年龄等
	患者类型	一般人群/易患人群/冠心病换患者
	病史记录	既往史/个人史/家族史
	体格检查	视诊/听诊/问诊/触诊
	医学影像检查	超声检查/CT/MRI
	非药物治疗	饮食疗法/运动疗法/心理疗法/其他疗法
has_treatment	疾病类型	治疗方法
has_detection	疾病类型	检测方法
detection_result	检测结果	疾病类型/危险分级/相关疾病
refer_to	治疗方法	检测结果/病史记录
Belong_to	无症状冠心病, 心脏其他类型 X 综合征, 心肌梗	

2.2.2 数据层构建

通过知识抽取, 对冠心病概念层和关系层进行详细设计, 构建冠心病本体, 明确了实体及实体关系后, 将关系数据和实体数据以三元组的形式导入到 Neo4j 图形数据库中, 具体知识存储方法参考非小细胞肺癌的知识图谱构建知识存储部分的详细说明, 冠心病知识图谱共包含 133 个医学实体, 137 条实体间关系, 可

为冠心病临床医学决策提供参考。图 8 为冠心病临床指南知识图谱可视化展示。

2.2.3 临床应用实例

冠心病知识图谱的构建为冠心病临床决策提供了便捷的查询方式。例如, 医学中药品种类繁多, 难以准确记忆, 在脱离临床指南的情况下, 如何在临床上帮助医生了解稳定型心绞痛患者进行药物治疗时应当使用何种药物。我们借助 Cypher 语言在冠心病知识图谱中查询这一信息, 输入查询语句“match data=(na:疾病类型 {name:'稳定型心绞痛'})-[r1]->(nb:相关信息)-[r2]->(nc:治疗方法)-[r3]->(nd:药物治疗) return data”对药品信息进行查询, 查询结果如图 9 所示。从查询结果可看出对稳定型心绞痛患者进行药物治疗时, 通常可采用硝酸酯类药物、抗凝药物等。药品信息通过冠心病知识图谱被及时获取, 减轻了医生记忆海量医学知识和错综复杂医学关系的难度, 在临床医学决策上起到至关重要的作用。

3 讨论

基于临床指南的知识图谱构建方法是一种专注单一疾病, 追求高质量医学知识表达的图谱构建方法, 可以根据患者症状以及指标数据, 为专业领域的医生提供诊前、诊中、诊后的辅助性临床决策方案。但此种方法仍有不足之处, 首先, 此种方法虽然可以保证录入知识图谱的数据质量, 但是需要由知识工程师和临床医生共同对临床指南进行整理与归纳, 需要消耗一定的人力和时间。其次, 文本指南中存在确定性知识和不确定性知识两种类型, 对不确定知识的取舍与分割亦是该方法的难点之一, 较好的办法是引入多种知识表达方式, 增强知识图谱的语义表达能力。综上, 本文所提出的方法只适用于专业领域的医学知识图谱构建, 对于通用医学知识图谱的构建还需要依靠大数据技术来获取实体, 或是由多个专业领域图谱融合来完成。

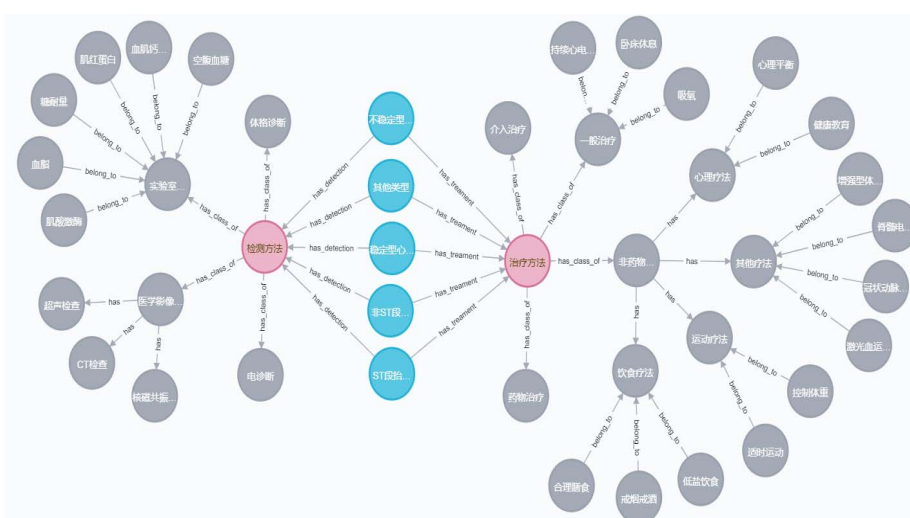


图 8 冠心病临床指南知识图谱可视化展示
Fig.8 The visualization of knowledge graph of clinical guidelines for coronary heart disease

(下转第 197 页)

- [4] 张铨伟, 王彪, 徐贵力. 摄像机标定方法研究[J]. 计算机技术与发展, 2010, 20(11): 174-179.
- [5] 马颂德, 张正友. 计算机视觉: 计算机理论与算法基础[M]. 北京: 科学出版社, 1998: 72-75.
- [6] 曲学军, 张璐, 张凌云. 基于新的成像模型的摄像机标定与三维重建[J]. 仪器仪表学报, 2011, 32(8): 1830-1836.
- [7] 夏茂盛, 孟祥磊, 宋占伟, 等. 基于双目视觉的嵌入式三维坐标提取系统[J]. 吉林大学学报: 信息科学版, 2011, 29(1): 61-66.
- [8] 张曙. 基于机器视觉的机械手抓取控制研究[D]. 上海: 上海理工大学, 2018.

- [9] Lu-qiao FAN, Wen-qiong ZHOU, Ban-xiang DUAN, etc. Obtain 3D coordinate of EOD robots based on computer vision[J]. Machine Tool & Hydraulics, 2019, 47(12): 143-150.
- [10] 韩博. 基于双目立体视觉的工件识别与定位关键技术研究[D]. 哈尔滨工业大学, 2018.
- [11] 张广军. 视觉测量[M]. 北京: 科学出版社, 2008: 49-53.
- [12] 曲学军, 张璐. 基于双目视觉的三维测量方法[J]. 计算机仿真, 2011, 28(2): 373-377.
- [13] Y Yabuta, H Mizumoto, S Arai. Binocular robot vision with active viewpoint for sphere measurement[J]. Key Engineering Materials, 2012, 516: 343-348.

(上接第 184 页)

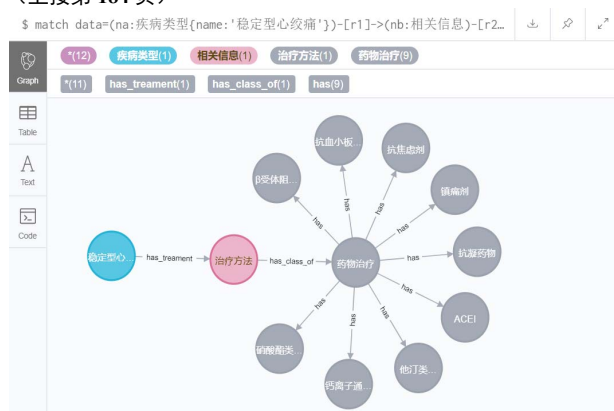


图9 药品信息查询结果
Fig.9 The query results of drug information

4 总结

本文提出了一种基于临床指南的知识图谱搭建方法, 并针对陈述性知识和流程性知识两种文本指南进行知识图谱搭建, 实现了非小细胞肺癌和冠心病两种疾病的知识图谱。疾病知识图谱的构建为医学临床决策中查找和整合医学知识, 进而进行临床辅助决策奠定了基础。下一步工作将基于知识图谱开发智能临床辅助决策系统, 实现知识图谱的语义搜索与智能问答, 帮助临床医生提高临床决策的效率和质量。

参考文献

- [1] 杨美洁, 熊相超. 基于爬虫技术和电子病历的糖尿病知识图谱的构建[J]. 中国数字医学, 2020, 2(2): 6-8.

- [2] 廖凌斌, 周欣, 何小海, 等. 面向涉恐领域的知识图谱构建方法[J]. 信息技术与网络安全, 2019, 38(9): 34-38.
- [3] 康准, 王德军. 基于知识图谱的生物学科知识问答系统[J]. 软件, 2018, 39(2): 07-11.
- [4] 王宁. 基于Web的领域知识图谱构建平台的研究与实现[D]. 北京市: 北京邮电大学, 2019.
- [5] Xin-Hong Jia, Wen-Ai song, Wei-Yan li, etc. Semi-automatic Construction Method Of Chronic Obstructive Pulmonary Disease Knowledge Graph[R]. Wisconsin, United States: IEEE 43rd Annual Computer Software and Applications Conference, 2019.
- [6] 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.
- [7] 孙郑煜, 鄂海红, 宋美娜, 等. 基于大数据技术的医学知识图谱构建方法[J]. 软件, 2020, 41(1): 13-17.
- [8] 曹红霞, 梁万年. 循证临床指南的开发程序[J]. 中国全科医学, 2003, 6(4): 342-344.
- [9] 吴彬飞. 临床指南知识表达与应用方法研究[D]. 浙江大学. 2010.
- [10] 肖仰化等. 知识图谱: 概念与技术[M]. 北京: 电子工业出版社, 2020. 1: [286].
- [11] 王书航, 王洁. 2018年V3版NCCN非小细胞肺癌指南更新要点解读[J]. 华西医学, 2018, 33(4): 388-392.
- [12] 胡一然, 宋中山, 孙翀, 等. NVSA: 一种具有可变节点值的查询图搜索算法[J]. 软件, 2018, 39(3): 16-21.
- [13] 张玲玉, 尹鸿峰. 基于OAN的知识图谱查询研究[J]. 软件, 2018, 39(1): 54-59.
- [14] 中华医学会. 临床诊疗指南. 心血管分册[M]. 人民卫生出版社, 2017.