

# 基于知识图谱的智能医疗研究

俞思伟, 范 昊, 王 菲, 徐 雷

**[摘要]** 目的: 利用知识图谱技术将各种医疗信息系统中琐碎、零散的知识相互连接, 以支持综合性知识检索以及问答、辅助决策支持等智能医疗应用。方法: 在构建医疗领域本体和对医疗知识库进行语义标注的基础上, 以慢病为例, 构建医疗知识图谱并将其应用于智能医疗中。结果: 提出了医疗知识图谱的构建方法和将其应用于基于医疗知识库的语义分析、推理以及疾病辅助诊断系统的方法。结论: 基于知识图谱的智能医疗应用将在解决国内优质医疗资源供给不足和医疗服务需求持续增加的矛盾中产生重要的作用。

**[关键词]** 知识图谱; 智能医疗; 本体; 语义分析; 辅助诊断

**[中国图书资料分类号]** R318; TP315 **[文献标志码]** A **[文章编号]** 1003-8868(2017)03-0109-04

**DOI:** 10.7687/J.ISSN1003-8868.2017.03.109

## Research on intelligent medicine based on knowledge graph

YU Si-wei<sup>1</sup>, FAN Hao<sup>2</sup>, WANG Fei<sup>2</sup>, XU Lei<sup>2</sup>

(1. School of Medicine, Wuhan University, Wuhan 430071, China;

2. School of Information Management, Wuhan University, Wuhan 430071, China)

**Abstract Objective** To discuss the use of knowledge graph technology to connect various trivial and fragmented knowledge in various medical information systems and support comprehensive knowledge retrieval and intelligent medical applications such as Q & A and clinical decision support. **Methods** Based on the construction of medical field ontology and semantic labeling of medical knowledge base, the medical knowledge graph was constructed and applied to intelligent medicine, with chronic disease taken as an example. **Results** The construction method of medical knowledge graph and its application in semantic analysis, reasoning and disease-assisted diagnosis system based on medical knowledge base were put forward. **Conclusion** The application of intelligent medicine based on knowledge graph will play an important role in contradiction between the supply of high-quality medical resources and increasing medical requirements. [Chinese Medical Equipment Journal, 2017, 38(3): 109-111, 126]

**Key words** knowledge graph; intelligent medicine; ontology; semantic analysis; assistant diagnosis

## 0 引言

自计算机技术应用于医疗领域以来, 人们就不满足于仅仅完成一些简单的医疗信息采集和记录, 而是希望能够帮助医生提升诊疗水平<sup>[1]</sup>。随着计算机科学技术的发展, 尤其是近年来云计算、物联网和大数据的快速发展, 使得人工智能的水平得到了很大的提高。医疗领域成为人工智能应用最活跃和最广泛的领域之一。

智能医疗是指运用现代信息、诊疗技术和设备等手段, 依托信息化技术平台建立以个人健康档案为核心的区域医疗卫生协作模式; 利用先进的信息技术改善疾病预防、诊断和研究, 继而实现对人群健康的科学管理, 最终让医疗生态圈的各个组成部分受益。智能医疗是一个以患者信息为本的协作体系,

该体系把患者基本信息、病历记录、各种实验室检查结果和财务信息等整合在一起, 以期达到诊疗精确化、成本集约化和就诊便捷化的工作目标, 最终实现疾病的有效预防, 有助于提高全民健康水平。

智能医疗的核心和关键是智能诊疗, 也就是让计算机成为具有医疗知识的大脑, 从而对医生诊疗提供辅助决策。让计算机成为具有医疗知识的大脑, 简单来说就是成为一种知识库, 加上基于此的逻辑推理和计算, 它可以泛指医疗从业人员的智慧总和。具体来说, 它具备诊断、治疗的能力, 可以独立提供用药辅助、分诊导诊、健康咨询等服务, 也可以高质量地辅助医疗从业人员完成某项工作。知识图谱是人工智能领域的一项重要技术, 是构建计算机医疗知识大脑的技术基础, 因此, 知识图谱成为智能医疗的关键技术之一。

## 1 知识图谱

Google 于 2012 年提出知识图谱的概念, 将其技术应用于 Google 搜索引擎上面, 用以提高 Google 查

基金项目: 国家自然科学基金(71661167007)

作者简介: 俞思伟(1966—), 男, 博士, 副主任技师, 主要从事医疗卫生信息化方面的研究工作, E-mail: 85289125@qq.com。

作者单位: 430071 武汉, 武汉大学医学部(俞思伟), 信息管理学院(范 昊, 王 菲, 徐 雷)

询返回结果的质量。Google 高级副总裁艾米特·辛格博士认为“构成这个世界的是实体,而非字符串。知识图谱能够理解真实世界中的实体以及实体间关系”<sup>[2]</sup>。

### 1.1 知识图谱的概念

知识图谱是语义网<sup>[3]</sup>(semantic web)技术之一,是一种基于图的数据结构,由节点(实体)和标注的边(实体间的关系)组成,它本质上是一种语义网络。知识图谱中的每个节点表达的是现实世界中存在的“实体”,每条边为实体与实体之间的“关系”。也就是说,知识图谱就是把所有不同种类的信息连接在一起而得到的一个关系网络。知识图谱提供了从“关系”的角度去分析问题的能力,也强调了数据驱动应用的概念。通过知识图谱,用户可以快速有效地获取相关知识以及知识之间的逻辑关系,并将不易理解的抽象信息以用户能够理解的可视化方式展示出来<sup>[4]</sup>,并且能够返回更全面、更丰富的信息。

### 1.2 知识图谱的发展现状

从 Google 知识图谱提出以来,美国的微软必应,中国的百度、搜狗等搜索引擎公司也纷纷宣布了各自的“知识图谱”产品,如微软“Satori”、百度“知心”、搜狗“知立方”等<sup>[5]</sup>。这些公司都希望利用知识图谱技术真正理解用户要查询的到底是什么,然后返回用户所关心的信息。

在国外,知识图谱技术已经比较成熟,大规模的知识图谱已经被构建好,包括 YAGO<sup>[6]</sup>、DBpedia<sup>[7]</sup>、NELL<sup>[8]</sup>、Freebase<sup>[9]</sup>和 Google Knowledge Graph<sup>[2]</sup>,这些知识图谱包括百万级别的实体以及十亿级别的关系。

我国针对知识图谱构建技术的研究还处于初步探索阶段,主要的有百度“知心”、搜狗“知立方”、复旦大学“CN-DBpedia”和“Zhishi.me”等。针对于生物医疗领域,国外著名且已经比较成熟的医疗知识库有 Gene Ontology、SNOMED—CT、Drug-Bank、Disease Ontology 等,而国内主要有区域本体知识库<sup>[10]</sup>和中医药知识图谱<sup>[11]</sup>,都还处于起步研究阶段。

综上所述,相对于英文,中文语言本身比较复杂,无法直接使用国外领先的技术进行语义分析,而是需要对技术进行研究改进。其次,中文医疗的专业语料库比较匮乏,也缺乏相关专业人士参与到知识共享。虽然可以利用机器翻译技术将医疗相关的英文知识库翻译成对应的中文,但是由于语言的差异性,语料知识的不完整性,容易产生语法错误、歧义问题以及专有名词的不一致,无法满足对医疗知识图谱构建高准确性的要求。因此,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。

## 2 知识图谱的构建

### 2.1 构建方案

知识图谱的构建方案一般分为 4 种<sup>[12]</sup>:

(1)“精确策划构建”方案:会用一组相关的专家通过手工的方法来创建三元组。该方法构建的知识图谱虽然具有较高的准确性,但是这种技术无法大规模应用,因为它依赖于具有一定背景知识的专家。

(2)“合作构建”方案:会以众包的方法,以一定的开放原则,招揽自愿者参与手工构建三元组的项目。这种方法一度用于 Wikipedia 和 Freebase 的构建,相比较于第一种方法具有一定的规模,但是依然具有一定的局限性。

(3)“半自动或自动化的半结构化文本构建”方案:凭借人工拟订的规则,学习到的规则或者正则表达式从半结构化的文本(例如:Wikipedia 中的信息盒)中自动地抽取到三元组(比如 YAGO 和 DBpedia)。通过这种方法所构建的知识图谱有着非常高的准确率。然而半结构化的文本中往往只是信息中的判断信息存储在 Web 中,不能完全覆盖所有知识,因此有着第 4 种构建方案。

(4)“半自动或自动化的非结构化文本构建”方案:通过机器学习和自然语言处理技术从非结构化文本中自动地挖掘三元组信息<sup>[13]</sup>。

### 2.2 构建方式

知识图谱的构建方式一般分为 3 种:自顶向下的方式、自底向上的方式以及这 2 种方法相结合的方式<sup>[12]</sup>。自顶向下的方式首先构建顶层关系本体,然后将抽取到的实体匹配更新到所构建的顶层本体中。自底向上的方式则直接将抽取数据中发现到的类别、实体、属性以及关系合并到知识图谱中。不管使用哪种方式构建知识图谱,其构建流程分为 4 个模块:知识获取、知识表示、知识存储以及知识可视化。

### 2.3 慢病知识图谱的构建

慢病知识图谱的构建需要整合多方资源,如关于慢病的医学研究文献、慢病保健常识以及临床诊疗案例库数据等。这些数据分布在互联网、科学文献数据库、专科诊疗数据集、纸质书籍等多个地方,并且这些资源类型多样,具有不同的侧重点和不同的质量。构建关于慢病的知识图谱,需要考虑该体系的全面性、高质量,因此需要对众多的资源进行过滤、抽取,使之结构化、系统化。同时,为了实现慢病知识的有效组织及扩展性,需要对慢病知识中涉及的医学专业术语进行规范化,以保证慢病相关的资料数据都能够被有效地收集和处理<sup>[14]</sup>。

慢病的治疗知识大部分来自医生的临床经验、已形成的关于慢病治疗的有效方法等,这种医疗经验及知识在具体的医疗诊治中的差异性给慢病知识

的规范、标准化带来了困难,目前尚没有规范化体系的慢病病理、发病规律、医疗控制等方面的资源。因此,为了实现慢病治疗相关知识的体系化,需要对慢病知识资源进行收集、归类与整理,过滤不相关的内容,构建关于慢病诊疗的知识图谱,从而实现慢病的智能医疗。

慢病资源抽取与清理流程如图1所示。首先全面收集关于慢病的各种资源,使用数字化技术对其进行电子化,对于复杂的专业性医学经验知识,其形式包括各种图片、音频和视频资源,由慢病专家指导进行分类归纳使之结构化、系统化,形成关于慢病诊疗的知识集合。对于文本型的慢病资源,在相关医学术语标准下,对其进行文本分析,如语义标注、实体及关系抽取、关联分析等,获取关于慢病的所有规范或不规范的概念词汇集合,并根据资源内容的差异以及概念之间的关系类型,组成构建关于慢病的顶层本体。最后融合慢病知识集合和慢病顶层本体构建关于慢病的知识库。

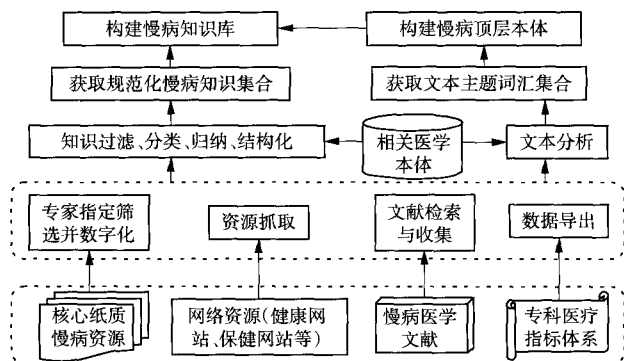


图1 慢病知识抽取与清理流程

慢病知识图谱的构建包括2个部分:慢病治疗经验知识库和慢病领域本体。慢病治疗经验知识库是关于慢病治疗的各种病因、有效药物、治疗过程、治疗周期、禁忌等知识,是慢病知识库的核心;慢病领域本体是关于慢病的医学概念及概念间关系的术语集合,是为了实现慢病知识库的共享与扩展性需要而设计的,其中大部分的内容可通过引用已有的临床术语集合或疾病本体来实现。

### 3 智能慢病辅助诊断系统

针对构筑的慢病知识图谱,设计慢病知识库的语义分析、推理以及慢病辅助诊断系统。该系统包含核心的慢病知识库,使用过程中不断地收集患者的临床诊疗数据、日常监测数据以及专科体检数据和各种输入的患者症状数据等,通过慢病知识库的语义匹配与推理功能辅助确认患者的慢病类型、治疗手段与计划、慢病保健常识与注意事项等。系统整体操作流程如图2所示。

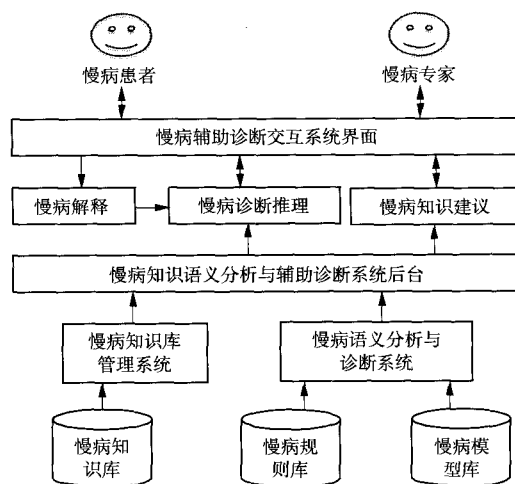


图2 智能慢病辅助诊断系统操作流程

如图2所示,慢病患者和医生都可使用该平台,通过提交慢病患者的各种疾病数据特征,知识库后台通过解释器对用户输入的各种数据、概念等进行解释、转换,进行慢病的语义分析和智能诊断,并提供给患者和专家关于慢病的知识建议。慢病知识语义分析与辅助诊断系统包括慢病知识库管理系统和慢病语义分析与诊断系统,该系统需要对慢病知识库、规则库、模型库进行查询、更新、调整等操作,以保证该系统的动态扩展性。其中,慢病规则库是关于慢病诊治的各种经验规则,可根据患者不同的病理特征数据推断出患者预防、用药等治疗措施,可动态扩展;慢病模型库是关于慢病诊治的个性化精确推断模型,可根据不同患者的病理特征,采用统计推断、数学模型等方法来设计有效的慢病诊断算法<sup>[15-16]</sup>。

### 4 结语

慢病知识图谱是智能慢病辅助诊断系统的基础,如何将不同来源、不同结构的慢病知识,例如临床路径、诊疗规范等按照一定的标准(如病情阶段等)和结构结合成一个可供利用的有机知识体系的方法,本文进行了深入的研究。在此基础上,探讨了慢病知识图谱的构建和智能慢病辅助诊断系统的系统框架,为系统开发和实现奠定了良好的基础。医疗知识图谱是智能医疗的重要基础,随着人工智能技术的快速发展,自然语言处理、图像识别、深度学习、机器学习等技术的日益成熟,为医疗知识图谱的发展奠定了良好的技术基础。基于知识图谱的智能医疗的应用将在解决我国优质医疗资源供给不足和医疗服务需求持续增加的矛盾中产生重要的作用。

### [参考文献]

- [1] 陈金雄,王海林. 迈向智能医疗:重构数字化医院理论

(▶▶下转第126页▶▶)

高端产品,已更新换代数次。目前,万元级别价格的医院便携式内窥镜系统已成功推向市场,国内外厂家均已开发出自己的产品。不过,即便该医疗设备已发展多年且国内外产品众多,其发展仍处于起步阶段。随着人工智能技术、计算机技术、大数据技术、云技术等新兴技术与医学领域的不断融合,医用便携式内窥镜将趋于小型化、多样化、智能化,同时能够支持物联网医疗。虽然医用便携式内窥镜无法取代现有的光纤内窥镜及电子内窥镜,但其将逐渐发展成为内窥镜领域一个不可分割的重要分支,其便携性的特点将在机动医疗中发挥无法替代的作用。

#### 4 结语

内窥镜的发展日新月异,便携式内窥镜的出现开启了一个新的发展方向。它凝聚了最尖端的电子技术,医生工作服的口袋即可容下整个内窥镜。已经上市的医用便携式内窥镜为其发展提供了很好的参考,LED、FPGA、DSP 和 SoC 等技术的发展及应用使医用便携式内窥镜在短短数年中就完成了跨越式的发展<sup>[12]</sup>。

从目前对医用便携式内窥镜研究的发展趋势来看,它依然倾向于微型化、集成化、智能化、低功耗化。此外,人体工学等设计理念的进步及其他最新技术在医学领域的应用,超声内镜、三维内镜、染色内镜等未来医用便携式内窥镜发展的必然趋势。

#### [参考文献]

- [1] 陈月存. 新型工业内窥镜的研究[D]. 长春: 长春理工大学, 2009.
- [2] 姜萍萍, 颜国正, 丁国清, 等. 一种便携式工业视频内窥镜的开发[J]. 光学精密工程, 2002, 10(4): 407-410.
- [3] 吴启海. 像素、CCD 尺寸与图像、色彩分辨率[J]. 照相机, 2010(2): 76-77.
- [4] 李定川. 高端 3CCD 数码摄像机的新机型综述[J]. 影像技术, 2007(4): 3-8.
- [5] 谢晴. 富士“Super CCD EXR”传感器[J]. 新电脑, 2008(10): 179.
- [6] 杨天领. 医学内窥镜高性能 LED 冷光源研究[D]. 杭州: 浙江大学, 2013.
- [7] 徐思燕. FPGA 器件设计技术发展综述[J]. 通讯世界, 2015(10): 223-223.
- [8] 王国章, 刘战, 须自明, 等. 一种新的硬件设计方法——结构化 ASIC 技术[J]. 微计算机信息, 2006, 22(2): 153-155.
- [9] 刘中汉. 基于 SOC 的医疗电子内窥镜摄像处理系统的设计[D]. 沈阳: 东北大学, 2014.
- [10] 陈子龙, 张红雨, 李俊斌. 蓝牙 4.0 无线传感网数据采集及以太网传输设计[J]. 电声技术, 2013, 37(10): 74-77.
- [11] 周文光, 王春飞, 许新建, 等. 便携式内镜的无线传输系统设计[J]. 生物医学工程与临床, 2015, 19(2): 182-185.
- [12] 叶斌. 高清晰医用电子内窥镜关键技术研究[D]. 杭州: 浙江大学, 2011.

(收稿: 2016-06-24 修回: 2016-11-01)

(◀◀上接第 111 页◀◀)

- 体系[D]. 北京: 电子工业出版社, 2014: 16-18.
- [2] SINGHAL A. Introducing the knowledge graph: things, not strings[EB/OL]. (2012-05-24)[2016-12-10]. <http://google-blog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>.
- [3] BEMERSLEE T, HENDLER J, LASSILA O. The semantic web[J]. Sci Am, 2001, 284(5): 34-43.
- [4] 陈悦, 刘则渊, 陈劲, 等. 科学知识图谱的发展历程[J]. 科学学研究, 2008, 26(3): 449-460.
- [5] 刘知远, 崔安硕. 大数据智能: 互联网时代的机器学习和自然语言处理技术[M]. 北京: 电子工业出版社, 2016: 23-27.
- [6] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: a core of semantic knowledge[C]//The 16th International Conference on World Wide Web, May 8-12, 2007, Banff, Alberta, Canada. New York: ACM, 2007: 697-706.
- [7] AUER S, BIZER C, KOBILAROV G, et al. Dbpedia: a nucleus for a web of open data[M]. Berlin Heidelberg: Springer, 2007.
- [8] CARLSON A, BETTERIDGE J, KISIEL B, et al. Toward an architecture for never-ending language learning[C]//The Twenty-Fourth AAAI Conference on Artificial Intelligence, July 11-15, 2010, Atlanta Georgia, USA. USA: AAAI, 2010: 3.
- [9] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//The 2008 ACM SIGMOD International Conference on Management of Data, June 9-12, 2008, Vancouver, Canada. New York: ACM, 2008: 1 247-1 250.
- [10] 王巍巍, 王志刚, 潘亮铭, 等. 双语影视知识图谱的构建研究[J]. 北京大学学报: 自然科学版, 2016, 52(1): 25-34.
- [11] 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): 51-53.
- [12] 王昊奋. 大规模知识图谱技术[EB/OL]. [2016-12-10]. [www.bi168.cn/thread-6745-1-1.html](http://www.bi168.cn/thread-6745-1-1.html).
- [13] 李劲松, 黄智生. 生物医学语义技术[M]. 杭州: 浙江大学出版社, 2012.
- [14] 俞思伟, 范昊, 王菲. 多维度临床知识组织方法及其知识库构建与平台开发[J]. 医学信息学杂志, 2016, 37(4): 2-7.
- [15] 俞思伟. 临床决策知识管理中语义检索技术的应用研究[J]. 中国数字医学, 2013, 8(4): 25-28.
- [16] 俞思伟, 姜赢, 董慧. 基于语义技术的临床路径工作流建模方法研究[J]. 中国数字医学, 2010, 5(8): 8-10, 37.

(收稿: 2016-09-10 修回: 2016-12-15)