

基于医学知识图谱的交互式可视化医学知识服务系统

沈明辉 吴结凤

琚生根

丁智刚 江顺权 付鹏 姚歆

(四川省卫生和计划生育
信息中心 成都 610041)(四川大学
成都 610000)(四川省卫生和计划生育
信息中心 成都 610041)

〔摘要〕 系统研究医学知识表示模型及医学知识图谱构建、语义检索和可视化分析方法, 基于医学知识图谱构建交互式可视化知识服务系统, 即虚拟患者, 阐述技术实现方式、应用及预期成效, 指出该系统有助于快速提高基层医生临床能力, 减少医患矛盾。

〔关键词〕 虚拟患者; 智能化; 模拟培训; 医学知识图谱

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.09.011

The Interactive Visualized Medical Knowledge Service System Based on the Medical Knowledge Map SHEN Minghui, WU Jiefeng, Health and Family Planning Information Center of Sichuan Province, Chengdu 610041, China; JU Shenggen, Sichuan University, Chengdu 610000, China; DING Zhigang, JIANG Shunquan, FU Peng, YAO Xin, Health and Family Planning Information Center of Sichuan Province, Chengdu 610041, China

〔Abstract〕 To study on medical knowledge representation model, the building of medical knowledge map, semantic retrieval and visualized analysis method systematically, the interactive visualized knowledge service system, namely, the virtual patient is built based on the medical knowledge map. The paper elaborates on the technology realization mode, application and anticipated effectiveness, as well as points out that the system is conducive to the rapid improvement of the clinical capability of primary doctors, so as to relief the tensions between doctors and patients.

〔Keywords〕 virtual patient; intelligence; simulation training; medical knowledge map

〔收稿日期〕 2018-06-08

〔作者简介〕 沈明辉, 硕士, 工程师, 发表论文 5 篇; 通讯作者: 吴结凤, 硕士, 工程师。

〔基金项目〕 四川省科学技术厅“基于医学知识图谱的智能化虚拟病人研究”(项目编号: 2018GZ0182)。

1 引言

2011 年国家发布《教育信息化十年发展规划(2011-2020 年)》, 其中着重强调加强医疗教育教学培训软件的研发。新一轮医学教育改革正在兴起, 其动因包括卫生工作相互依存性的提高、教学方法的进步、公众对健康关注度的增加以及意识到变革势在必行。目前国内外医疗教育培训主要采用

理论授课, 结合各种专业技能虚拟设备培训, 但这些设备少则几十万, 多则数百万, 培训人员每人每年仅能使用几次, 培训频率、力度均不足。总体来说现有的医疗培训机制方式单调、价格昂贵, 无法满足受训群体的培训需求。医学虚拟培训作为一种基于计算机技术的相对低廉的培训方式势必能够有效缓解以上问题^[1]。

2 研究目标

系统研究医学知识表示模型、医学知识图谱构建、检索和分析方法, 建立交互式可视化知识服务系统, 即虚拟患者, 支持模拟临床培训与辅助临床诊断, 主要研究目标为: 一是设计临床医学知识表示模型。能够对多类型医学实体、高维度实体属性以及不确定性实体关联进行准确表达, 取得医学知识表示模型的创新。二是提出基于医学知识表示模型的医学知识图谱自动构建方法。能够从多源异构的医疗大数据中抽取医学信息, 将医学知识进行融合与推理, 进而自动构建大规模医学知识图谱^[2]。三是提出医学知识图谱的个性化检索和可视化分析方法、基于知识子图匹配的语义检索算法以及基于知识图谱摘要的可视化分析方法, 以提供智能化的医学知识服务。四是建立交互式可视化知识服务系统, 即虚

拟患者。能够进行智能化和实时的临床模拟诊疗, 提供辅助诊断决策支持。设计系统测评方案, 能够全面测评系统的性能和有效性。

3 研究内容

3.1 概述

本研究针对医疗大数据多源异构、语义多样以及医学知识多类型、高维度、不确定关联的特点, 遵循表示模型 - 图谱构建 - 知识服务 - 系统研发的研究思路, 对项目展开深入研究。研究内容框架, 见图 1。首先从理论上研究医学知识表示模型, 进而基于该模型研究医学知识图谱的自动构建方法, 包括医学信息抽取、知识融合和知识推理方法。从语义层面统一抽取、融合和推理医疗大数据蕴涵的医学知识。在此基础上进一步研究大规模知识图谱中语义检索和可视化分析方法, 具体包括知识子图匹配算法和知识图谱摘要算法, 以提供高效的智能医学知识服务, 最后整合上述研究内容的成果研发交互式可视化知识服务原型系统, 包括交互式个性化检索和知识图谱可视化分析模块, 为医学诊疗培训与临床诊断辅助决策提供支持。4 项主要研究内容既相互关联又各具独立性, 形成合理的研究体系。

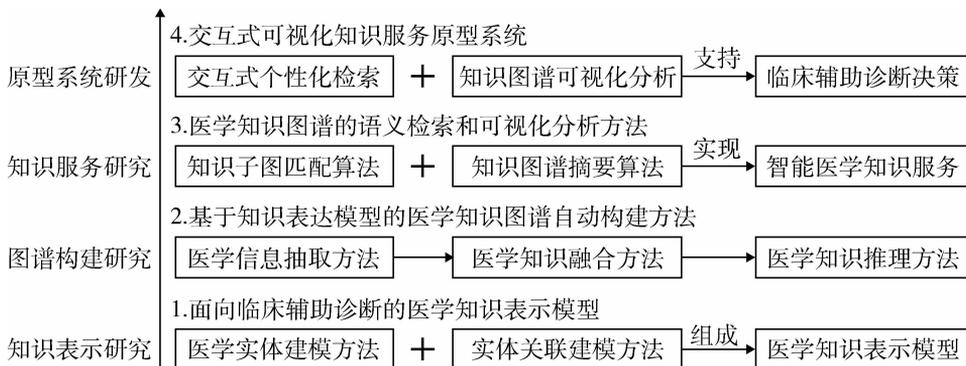


图 1 研究内容框架

3.2 医学知识表示模型

3.2.1 概述 医学知识的表示是大数据驱动医疗管理与决策的基础。医学领域的专业性和权威性决

定医学知识图谱的构建前提是对医学知识的正确表示。医学知识图谱本质上是具有语义网络结构的知识库。本研究将针对医疗知识不确定性关联的特点, 设计医学知识表示模型以破解医学知识表示的

难题。

3.2.2 医学实体分类体系建立方法 实体是知识图谱中的知识结点, 首先根据临床诊疗流程研究医学实体分类体系建立方法, 以自动地构建医学实体分类体系, 在领域专家的帮助下进一步优化该体系。

3.2.3 医学实体不确定性关联建模方法 临床诊断过程中状态的转换和知识的关联具有不确定性, 其中一些可以用定量的方式精确表达, 另外一些只能定性地表达, 本研究融合使用定量和定性的方式对实体不确定性关联进行表达并研究建模的方法。

3.2.4 医学知识表示模型设计 在以上研究的基础上分析总结现有知识表示模型的特点, 基于语义网络表示方法设计医学知识表示模型, 为医学知识表达与存储提供更加贴切的模型。

3.3 基于知识表示模型的医学知识图谱自动构建方法

3.3.1 多源异构医学大数据信息抽取方法 信息抽取包括实体、关系和属性抽取 3 个步骤。实体抽取也称为命名实体识别, 是信息抽取的第 1 步。研究内容第 1 部分所提出的医学知识表示模型给出医学知识的概念关系, 而现有的医学主题词表 (如 MeSH) 和医学术语库是从实例级别定义实体的取值, 均可以作为命名实体识别的先验知识。因此首先研究医学命名实体识别方法, 基于先验知识从海量医疗文本数据中自动抽取医学实体^[3]。实体抽取的结果是一系列离散的命名实体, 为得到语义信息, 还需要抽取实体之间的关联关系, 以形成知识图谱。因此研究医学实体关系抽取方法, 根据医学知识表示模型中的实体不确定性关联模型, 从文本数据中同时抽取确定性与不确定性的实体关系。针对医学实体属性高维度的特点, 研究医学实体属性抽取方法, 从多源异构的医学大数据中汇集属性信息, 实现对医学实体属性的完整刻画。

3.3.2 基于知识表示模型的医学知识融合方法 信息抽取的结果中可能包含大量的冗余和错误信息, 数据之间的关系也缺乏层次性和逻辑性。因此研究医学知识融合方法, 针对医学大数据语义多样

的特点, 消除概念的歧义, 根据知识表示模型合并已有结构化数据, 以保证知识质量, 提供医学知识的全景视图。

3.3.3 基于图模型的医学知识推理方法 知识推理是指从知识库中已有的实体关系推理出实体之间新的关联, 从而拓展和丰富知识图谱。知识图谱中实体为结点, 关系或属性为边, 研究基于图模型的医学知识推理方法, 根据实体在图结构上的特征推测实体之间的关系。

3.4 医学知识图谱的语义检索和可视化分析方法

3.4.1 基于知识子图匹配的语义检索 用户对医学知识图谱的个性化语义检索需求 (如患者的主诉和既往病史) 可以转化为实体和实体之间的关系, 构成查询子图, 其中结点为实体, 边为实体关系。语义检索实际上是从医学知识图谱的大图中获取和查询子图匹配的知识子图。考虑到医学知识图谱中的实体具有高维度属性, 检索结果的知识子图不仅在结构上与查询子图保持一致, 也需要满足匹配的结点 (即对应的实体) 属性相似。本研究基于知识子图匹配的语义检索算法, 根据用户的个性化语义检索需求, 从医学知识图谱大图中高效地获取同时满足对应的实体属性相似、实体关联一致的知识子图 (即诊断结论), 为临床辅助诊断决策提供支持。

3.4.2 基于知识图谱摘要的可视化分析 医学知识图谱规模巨大, 给用户进行可视化分析带来挑战, 用户难以在复杂的全局知识视图中找到所需的知识并进行实时分析。本研究知识图谱摘要算法对医学知识图谱进行自动缩放, 为用户展示不同粒度的知识图谱摘要, 帮助用户对医学知识进行可视化分析^[4]。

3.5 交互式可视化医学知识服务系统 (虚拟患者)

交互式可视化医学知识服务系统是医学大数据驱动管理与决策的典型应用。根据研究内容的理论与方法, 基于医学知识图谱研发支持医学临床诊疗培训和临床辅助诊断决策的交互式可视化知识服务系统, 通过虚拟患者模式进行呈现, 重点设计交互式的个性化检索模块和知识图谱可视化分析模块。该系统能够模拟医生临床诊断流程, 与用户进行交

互, 获取个性化检索需求, 返回满足条件的诊断结论 (医学知识子图); 同时能够对医学知识图谱进行摘要, 按照语义关联和缩放粒度进行展示, 帮助用户进行变粒度的可视化分析, 增强诊疗模拟和诊断结论的可解释性。

4 技术实现方式

采用理论与实证分析、应用研究相结合的方法, 在大量收集资料与数据、阅读文献的基础上梳理和总结经典的医学管理与决策理论以及大数据管理与分析方法; 结合数据实验的实证分析进行深入理论探索。在此基础上开展应用研究, 研发系统对理论成果进行验证, 根据评测标准对应用效果进行测评。技术路线, 见图 2。总体技术路线为建立知识表示模型、构建医学知识图谱、提供医学知识服务、研发知识服务系统, 具有较强的可行性和创新性。首先基于资源描述框架网络本体语言 (Re-

source Description Framework/Web Ontology Language, RDF/OWL) 建立医学知识表示模型, 包括医学实体分类体系以及建模实体不确定性关联; 然后从电子病历、临床指南和医学主题词表等多源异构医学大数据中抽取医学信息, 采用条件随机场模型抽取实体、朴素贝叶斯模型抽取实体关系, 关联规则挖掘方法抽取实体属性。提出实体链接方法和基于图的重启随机游走方法进行知识融合, 进一步提高知识质量, 构建医学知识图谱。提出概念和实体匹配方法, 合并第 3 方的关系数据库和知识库, 扩大医学知识图谱的规模。在此基础上提出基于图编码索引的知识子图匹配算法, 从医学知识图谱中获取满足用户个性化检索需求的诊断结论, 总结基于语义的结点聚类方法, 对知识图谱进行层次化摘要, 支持用户对大规模医学知识图谱进行可视化分析。最后整合以上方法和技术, 研发交互式可视化知识服务原型系统, 包括语义检索和可视化分析模块, 在医院进行示范应用, 设计测试基准对原型系统进行评测。

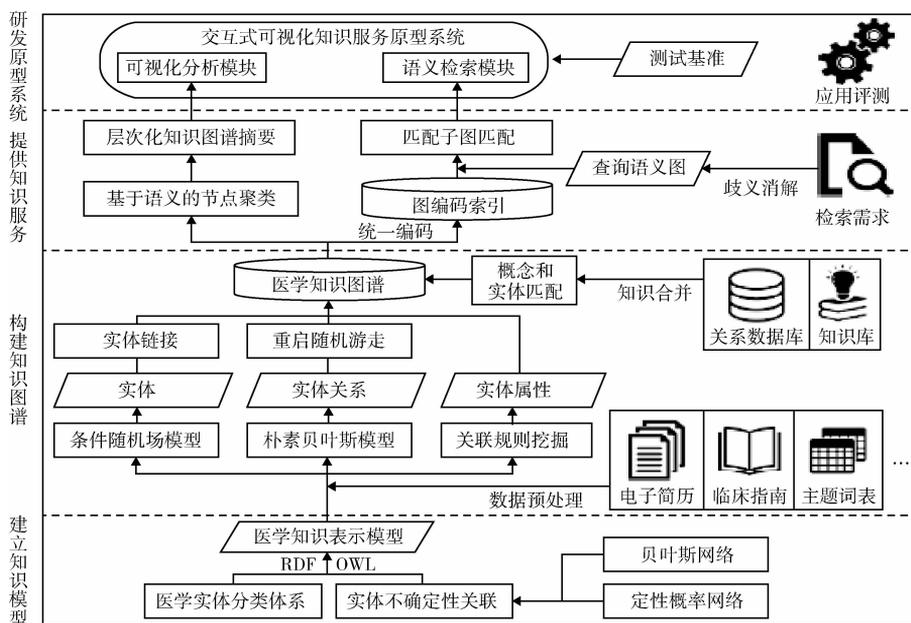


图 2 技术路线

5 应用与成效

5.1 主要应用场景

5.1.1 临床诊疗培训 以临床胜任力为导向、医

学知识图谱为基础, 构建智能化的虚拟患者, 模拟临床问诊、体检、辅检、诊断和治疗的全程。根据学员的干预, 虚拟患者的病情会向不同方向进行转归, 从而达到临床诊疗模拟的目的, 避免临床实习风险大、成本高等不足。该场景能够供学员和教师

随时随地教、学、练、考、评,从而帮助机构管理员实现机构资源的合理分配、导师即时掌握学员学习情况、学员随时随地学习临床知识。

5.1.2 临床辅助诊断 任何医生的知识面都是有限的,临床过程也容易由于医生经验因素造成遗漏或疏忽,知识查阅在临床过程必不可少,尤其是对于年轻的医生。而传统技术主要是通过关键字进行简单的概念定义查询,缺乏基于知识图谱的联想式查询以及推理。该场景主要是将传统的医生主动搜索转变为系统主动推荐,帮助医生做出临床决策,包括智能诊断、病史采集、治疗措施推荐等,从而大幅降低临床过程的差错率。

5.2 预期成效

5.2.1 有效解决目前临床教学实践成本高、风险高、案例少等困难 2014 年 11 月李克强总理提出医教协同深化临床医学人才培养改革的指导思想,提倡逐步建设具有中国特色的标准化、规范化临床医学人才培养体系,克服目前临床教学实践成本高、风险高、案例少等困难。然而在现代医学的发展中医生对患者进行有创的医疗操作已被法律与道德限制,直接学习无创性诊疗操作的机会也变得极为困难。因此医学教育必需建立更符合道德规范与法律规定、安全无风险、仿真医疗环境与流程的临床实践能力教学模式,以替代高风险的现实临床培训。本研究计算机虚拟现实技术构建虚拟患者,计算机人机交互模拟各种医学临床诊疗过程,采取所有医疗行为手段,实现模拟病房、问诊、体检、诊断、医嘱以及在线分享、交流等数 10 个特色功能,使基层医生能够完全基于浏览器或移动终端实现真实病例的模拟诊断和治疗,培训过程生动、逼真,且能力评估更为客观、科学,从而实现零风险、低成本、易普及、更科学的全新培训效果。

5.2.2 快速提高基层医生临床能力,减少医患矛盾 2015 年 9 月国务院办公厅发布《关于分级诊疗制度建设的指导意见》,从侧面提升基层医疗机构的地位。我国基本建立农村 3 级(县、乡、村)卫

生服务网和卫生队伍。其中处于卫生服务体系中心地位的基层医疗机构(乡镇卫生院和村卫生室)在人力、设备、资产、人才培养特别是服务能力和效率方面仍处于较低水平和相对落后地位。随着我国经济体制、医疗机构改革逐步深化以及分级诊疗制度的执行,大量患者将由三甲医院等大型医院分散到县级以下医疗机构,基层医院重要性尤为突显,基层医生能力培养迫在眉睫,否则医患矛盾依然会经常发生。在我国传统的医疗培训资源缺乏、成本高、效率低的情况下,应用本研究进行基层医护人员的诊疗能力培训能快速提高基层医生临床能力,减少医患矛盾。

6 结语

“21 世纪医学教育展望报告”指出目前我国医护人员每年的人均教育培训费用在 1 000 元左右,市场规模 60 亿/年,而这仅为欧美发达国家的 1/35。若要改善我国医疗环境,减少医患纠纷事件,医学教育和培训需要更大规模的投入,未来每年在该领域的市场规模必将达到千亿级别。而在线虚拟诊疗培训作为一种零风险、低成本、易普及、更科学的基于互联网的医学培训方式能够解决现有医疗培训存在的诸多问题,必将受到各医疗机构、医护人员的青睐,产生不可估量的经济效益和社会效益,潜力巨大。

参考文献

- 1 刘鸿燕,万艳丽,于凡,等. 临床指南知识表达模型研究现状分析 [J]. 中华医学图书情报杂志, 2017, 26 (5): 1-7.
- 2 吴彬飞. 临床指南知识表达和应用方法研究 [D]. 杭州: 浙江大学, 2010.
- 3 刘峤,李杨,段宏,等. 知识图谱构建技术综述 [J]. 计算机研究与发展, 2016, 53 (3): 582-600.
- 4 阮彤,王梦婕,王昊奋,等. 垂直知识图谱的构建与应用研究 [J]. 知识管理论坛, 2016 (3): 226-234.