

基于临床知识图谱的慢性肾病 辅助决策模型研究

Analysis of Chronic Kidney Disease Decision Support Model Based on Clinical Knowledge Graph

王琦¹, 康亮环², 刘国臻² (1. 北京大学, 北京 100871; 2. 北京聚人天下科技有限公司, 北京 100080)

Wang Qi¹, Kang Lianghuan², Liu Guozhen² (1. Peking University, Beijing 100871, China; 2. Beijing Jurentianxia Technology Ltd., Beijing 100080, China)

摘要:

提出了一种基于知识图谱的临床辅助决策系统构建方法。首先,通过建立标准数据模型,对多源数据进行汇聚,建立医学语料库;通过专家路径及医学教科书,参考UMLS的两层结构建立知识图谱的主干;然后通过医学自然语言处理及人工标注机制,对数据进行标准化处理,并通过数据挖掘建立概率模型,为知识图谱补充枝叶;基于知识图谱建立智能问诊路径,解决医生与辅助决策系统的交互问题。通过以上的路径建立起面向基层医生的肾病临床辅助决策系统,实现基层医生对慢性肾病标准化、规范化诊断和转诊,做到慢性肾病早发现、早治疗,起到了较好的效果。

Abstract:

A method of constructing clinical decision-making system based on knowledge graph is proposed. Firstly, medical corpus is built by building standard data model and aggregating multi-source data. Through expert path and medical textbooks, the backbone of knowledge graph is established by referring to the two-layer structure of UMLS. The standard medical language corpus is built by common data model, and at the same time gives the medical language processing tools and human-annotation rules to broaden and modify the knowledge map. Based on this knowledge map, the intelligent inquiry path is established to solve the interaction problem between doctors and assistant decision-making system. Through the above paths, the intelligent CKD decision support model for primary doctors is established, which can guide the primary physicians work standardly and regularly, and guarantee CKD early detection and treatment.

Keywords:

Chronic kidney disease (CKD); Knowledge map; Clinical decision support; Medical language processing

引用格式: 王琦, 康亮环, 刘国臻. 基于临床知识图谱的慢性肾病辅助决策模型研究[J]. 邮电设计技术, 2018(12): 68-71.

0 引言

近年来,慢性疾病发病率呈急剧上升趋势,已经成为影响居民健康水平提高、阻碍经济社会发展的重大公共卫生问题和社会问题。2012年柳叶刀文章指出,中国成人慢性肾病的患病率约10.8%,总数高达1.2亿人,而这一数字依然在持续增长。但实际检出率远远低于这个数字。实行分级诊疗、赋能基层医疗,能够解决医疗资源的供需矛盾,对实现慢性疾病控

制、患者健康管理等起到十分关键的作用。

在此背景下,利用医疗大数据和自然语言处理技术,基于临床数据和专家经验构建慢性肾病知识图谱,形成面向基层医疗业务的慢性肾病的临床辅助模型,实现基层医生对慢性肾病标准化、规范化诊断和转诊,做到慢性肾病早发现、早治疗。

1 研究现状分析

回顾决策支持系统理论与技术在临床应用与发展的30年,临床决策支持系统(CDSS)无论从其架构还是构建方法上,都发生了巨大的变化。在此领域里

收稿日期: 2018-11-19

的研究者和临床医生否定了原先构建专家系统的交互模式,基于专家经验的决策支持系统是不可能实现的。这存在2个方面的原因:一方面,专家经验并不是CDSS知识唯一的来源,对于不同的专家在同一问题上的表述存在差异,一个专家在不同时间对同一问题的看法同样也存在着不同;同时,个人的医学经验在不断变化之中,这使CDSS利用“生产式规则”表示专家经验,为非专家用户提供决策建议时,组合相关规则易出现冲突。另一方面,在专家系统中以“产生式规则作为知识块”来表示医学知识和人类的经验,无论就系统的复杂性还是人类认知来说,都是过于简单的模型,是致使专家系统低智能化的直接原因。

继医学专家系统之后,人工神经网络、遗传算法、模糊聚类算法等模式识别技术和基于数据仓库的数据挖掘技术在知识发现中的应用,不断提高CDSS的决策能力与决策范围。近年来,随着计算机技术的发展和临床信息系统的普及,基于人工智能技术的临床决策支持系统已成为各方研究的热点,并被认为是改善病人医护质量和提高医护人员工作效率的有效工具。根据决策系统的应用模式可以分为:

a) 提醒提示系统:实现化验危急值报警、临床疾病监测、药物使用及副作用的监测。

b) 评判式系统:对医生的诊疗行为进行自动评价,辅助医生修正诊疗行为。

c) 建议系统:基于临床指南或者预测模型等知识,针对患者状况进行推理,给出建议,供医生参考。

但是,CDSS在很多技术点上都面临挑战,尤其是基于中文医学信息的临床决策支持模型。生物系统是无比复杂的,临床决策可能需要利用庞大的潜在相关信息资源。例如,当向患者推荐治疗方案时,循证医学需要考虑到患者的症状和体征、既往疾病史、家族遗传史、疾病发生的历史和地理趋势、已发表的有效临床资料等。而且,最新发布的信息需要不断被整合到系统中去维持系统的实用价值,这一平台信息整合技术至今还不尽如人意。因此,有必要进一步完善现有的方法并开发新方法。在临床业务过程中,应加强临床辅助决策与实际业务环境的适用性研究,把注意焦点从提供正规知识转向改善医生的沟通与业务交流上,有利于沟通的系统无疑会极大地改善临床决策;另外,通过医学专家的知识专业性与临床数据的多元性需要有机结合,如临床知识的特性、个体数据的差异,如何构筑这类知识的模型及怎样将这些知识

应用到不同的环境中去等,能够有效增强临床辅助决策系统的适用范围和精确程度。本研究针对专家知识与临床数据挖掘的结合,业务流程中辅助决策模型与医生的交互2个问题设计慢性肾病辅助模型。

2 慢性肾病辅助模型研究

基于经典医学经验以及数据挖掘形成的知识图谱,通过添加复杂的、可定义的语义规则,用于进行健康预测、复杂疾病的推理等,构建慢性肾病辅助模型。在给定临床信息背景的前提下,通过各类医学实体间关系,研究类似规则、冲突规则、歧义规则的消除方法,定义规则的拟合度计算方法。基于知识图谱构建医疗知识搜索引擎和语义理解引擎,从“首诊症状”出发,构建了标准化智能辅助问诊流程,帮助医生对患者的病情进行判断。如图1所示,该流程可以智能规划问诊路径,并结合内置的高危疾病预警模型对医生进行诊断提示,并最终导向疑似诊断、治疗或进一步检查、转诊等路径。

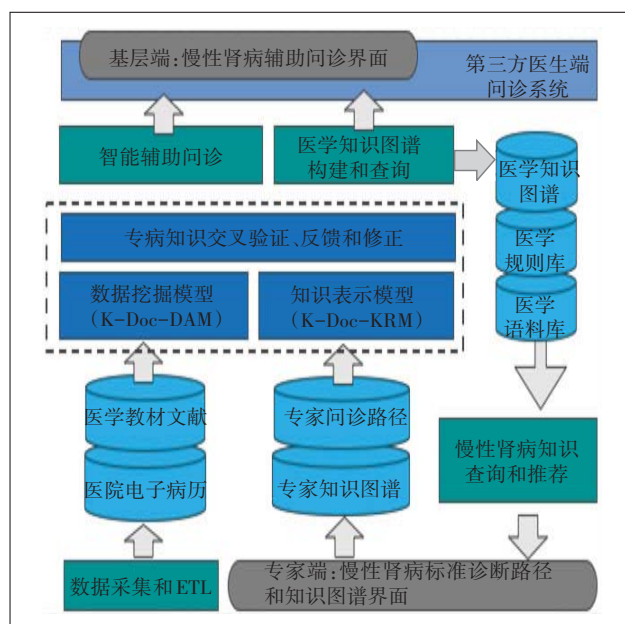


图1 慢性肾病辅助模型架构图

2.1 临床数据采集与整合

临床数据和知识信息是构建慢性肾病临床知识图谱的骨架,慢性肾病的诊治过程中,对于真实世界临床数据,患者体征、症状、诊断、药物、治疗、检验检查等数据最为关键。基于专家知识体系和临床指南标准建立知识图谱的骨架,形成通用标准数据模型,再整合真实世界临床数据,通过分析各部分数据的通

用结构,整合临床数据各个信息源,利用自然语言处理技术进行抽取整合,并利用通用标准数据模型形成统一的临床数据源和数据仓库。

2.2 医学知识图谱构建

慢性肾病临床知识图谱的构建是临床辅助决策系统的核心,通过收集整理临床大数据与专家知识体系,集成疾病、诊断、药物、治疗、健康管理等临床信息,以及专家经验、临床指南、经典病案等相关知识,进行自动抽取、深度学习和专家审编,形成证据分级的临床知识图谱,满足慢性肾病临床业务需求。

2.2.1 临床术语本体库构建

由于原始临床数据仓库的信息极为庞大,包含多种描述方式,其内容难免有重复和雷同,需要对临床术语进行归一化处理。其中包括:医疗信息统一化;同义词、近义词及衍生词的判别与合并,压缩信息数据库在词空间上的维度,减少文本匹配计算量,优化匹配效率;近似信息的归类和语义抽象,减少信息数量。最后,建立语义归一化的术语标准。在术语标准化的过程中,参考了国内外成熟的标准体系。疾病名称的标准术语字典使用国标2016版ICD-10疾病分类与编码系统,手术与操作标准术语字典使用ICD-9,检验标准字典使用LOINC中文版,药品标准术语字典使用NCCD系统。对归一化后的数据库进行自主学习式实体标注,先通过规则定义自动分析一部分术语,然后再启动人工标注和第三方审核这些步骤,迅速地累积语料以及标注结果。再利用LSTM神经网络算法来对这些实体标注模式进行学习,发现新的未入库语料和实体,最后再通过专家审核,把机器学习出的语料库和实体、实体关系标注入库,逐步适应临床业务数据习惯。

2.2.2 实体对齐与实体链接

通过定义患者五维视图(如图2所示)形成描述患者必须要抽取的一个最小网络,其他的节点都以任务为导向来进行添加,以明确实体对齐的目标和范围,

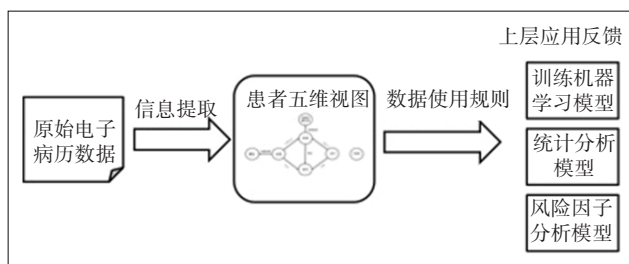


图2 患者五维视图构建与使用流程图

来优化实体对齐的效率,提高实体对齐的质量。在底层参考美国医学图书馆UMLS数据模型来定义数据结构,五维视图是作为这个数据结构中的子集;并构建底层统一任务器,以分解任务执行过程、统一数据标准和结果入库。所有的任务规则会基于标准数据模型进行,避免不同的任务需求所融合的知识不能够相互兼容。同时,参考美国医学图书馆UMLS采取建立超级叙词表的方式来定义概念、概念属性、上下层级。首先规定最为通用的词作为通用概念,对于同一概念的不同术语以及不同的变异形式,再采用三级结构模式,即概念(I级)-术语(II级)-词串(III级),基于此定义将原有术语库里已有概念进行关联和分级,由于之前已经统一了概念,把这些关联关系相互补充后,形成最终上下位关系结构。对于实际抽取到的临床实体,将知识图谱中的实体和关系,通过嵌入(embedding)方式投影到低维向量空间,并在向量空间中通过向量平移转换操作,计算头、尾实体及关系在向量空间中的损失函数值,实现头尾实体的关系链接。

2.2.3 知识图谱存储与更新

基于图数据库,患者五维视图中(如图3所示),每个临床术语本体用一个全局唯一确定的ID来标识,作为它们的标识符,每个属性-值对刻画实体的内在特性,而关系用来连接2个实体,刻画它们之间的关联。结合实际业务的知识表示模型与数据挖掘模型,根据标注开发标准流程(MATTER循环),拟采取宽进严出标注审核机制管理知识图谱的更新,宽进是指很多的知识都是通过数据挖掘而自动产生的,严出是指最初的本体框架、分类规则,以及最后的入库这些环节,还是基于行业专家判定,来保障本体库的质量和知识的准确度,也通过这个环节来增强机器的学习能力。

2.3 慢性肾病临床辅助决策模型构建

2.3.1 慢性肾病辅助模块

整个慢性肾病辅助模型分为以下3个模块。

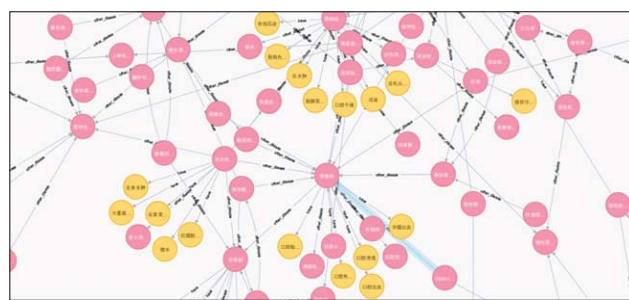


图3 慢性肾病知识图谱示意图

a) 智能辅助问诊模块:以“首诊症状”为入口,根据患者信息智能化设计问诊问题、规划问诊路径,辅助医生进行问诊和疾病判断。问诊路径的规划以知识图谱和专家路径为依托。以首诊症状为“水肿”为例,首先通过知识图谱挖掘“水肿”关联的疾病,以及这些疾病的关联症状、症状权重等内容,并智能计算出鉴别诊断的最优问诊路径。在此基础上,结合专家路径,形成规范化、标准化问诊路径,帮助医生收集信息。

b) 疾病风险评估模块:模型内置了基于大数据机器学习算法的疾病风险预警模型,当达到预警阈值时,会弹出提醒,并建议下一步的检验、检查或量表评估。疾病风险评估依托于海量电子病历资源,从中挖掘出疾病、症状、检验、检查之间的关联关系,动态计算疾病可能性。

c) 治疗与转诊建议模块:智能辅助问诊的最终结果会导向诊断、治疗、进一步检查或建议转诊。对于基层可以应对的常见疾病,如果可以确诊则提出诊疗建议;如果无法确诊则提供可能的疾病,并建议进一步的检查;如果涉及危重疾病、需要影像学检查,则给出相应的转诊建议。

2.3.2 知识图谱交互更新模型设计

为了实现医生经验与大数据规律的更新迭代,构建了两大标准化更新模型,保障知识图谱与临床业务能够相互结合。

a) K-Doc-KRM 知识表示模型:基于已有知识图谱,将新增专家经验转化为现有知识图谱可以理解的结构,并且可以自动化生成带有逻辑判断的结构化问诊流程,保障临床专家的知识体系作为知识图谱应用的骨架结构。

b) K-Doc-DAM 数据挖掘模型:基于医学自然语言处理算法,分析特定真实世界临床数据集,挖掘分析其中知识图谱未纳入的知识,基于专家通过标注开发标准流程(MATTER循环)审核后,补充至知识图谱,增强知识图谱对特定问题或特定场景的学习能力和适用性。

c) 专家经验的逻辑判断信息与数据规律的权重信息需要进行交叉验证,保证特定问题或特定场景下最终模型的准确性。

3 总结与展望

当医院信息系统发展到一定阶段,面对大数据时

代爆炸式增长的临床数据,完成针对业务功能上的应用后,临床决策支持系统是智慧医疗与医疗信息化建设的下一个目标,这将对提高医疗水平、促进医学科学的发展、充分发挥数字化医院的效能具有重要的作用,是先进计算机技术和现代医疗科研的完美结合。结合现有医疗信息化系统与业务需求,通过人机交互技术,实现智能辅助问诊、分诊、诊断和筛查等业务能力提升;有助于推进分级诊疗模式实行以及提升基层医疗水平。

目前,虽然逻辑推理算法很多,但是由于医学的复杂性和个体性因素较多,高智能、高集成的CDSS目前尚未实现,特别是医学知识库的建立更是一项复杂的系统工程,需要广大医学信息与医学大数据工作者继续共同奋斗实现。

参考文献:

- [1] ZHANG L, WANG F, WANG L, et al. Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet*, 2012, 379: 815-822.
- [2] YANO Y, FUJIMOTO S, ASAMI K, et al. Prevalence of chronic kidney disease in China[J]. *Lancet*, 2012, 380(9838): 214-214.
- [3] ZHANG L, LONG J, JIANG W, et al. Trends in Chronic Kidney Disease in China.[J]. *N Engl J Med*, 2016, 375(9): 905-906.
- [4] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. *电子科技大学学报*, 2016, 45(4): 589-606.
- [5] 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.
- [6] NICKEL M, MURPHY K, TRESP V, et al. A Review of Relational Machine Learning for Knowledge Graphs [J]. *Proceedings of the IEEE*, 2016, 104(1): 11-33.
- [7] DEMNERFUSHMAN D, MORK J G, SHOOSHAN S E, et al. UMLS content views appropriate for NLP processing of the biomedical literature vs. clinical text.[J]. *Journal of Biomedical Informatics*, 2010, 43(4): 587-594.
- [8] LEAMAN R, KHARE R, LU Z. Challenges in clinical natural language processing for automated disorder normalization[J]. *Journal of Biomedical Informatics*, 2015, 57(C): 28.
- [9] VOSS E A, MAKADIA R, MATCHO A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases [J]. *Journal of the American Medical Informatics Association* *Jamia*, 2015, 22(3): 553-564.

作者简介:

王琦,大数据高级经理,博士,主要从事医疗大数据分析数据挖掘工作;康亮环,工程师,博士,主要从事自然语言处理算法研究工作;刘国臻,工程师,主要从事临床知识图谱设计与医疗大数据产品设计。