

基于疾病诊断相关分组的医学知识图谱库构建方法^{*}

刘新奎 杨林朋

咎红英

牛承志

(郑州大学第一附属医院 郑州 450052) (郑州大学 郑州 450000) (郑州大学第一附属医院 郑州 450052)

〔摘要〕 详细阐述基于疾病诊断相关分组的医学知识图谱库构建方法, 对该知识图谱库应用前景进行展望, 包括健康宣教、辅助诊断决策系统、优化疾病诊断相关分组器等方面。

〔关键词〕 DRG - MKG; 辅助诊断; 健康宣教

〔中图分类号〕 R - 056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.09.005

Building Method of Medical Knowledge Graph Library Based on Diagnosis - related Group (DRG - MKG) LIU Xinkui, YANG Linpeng, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China; ZAN Hongying, Zhengzhou University, Zhengzhou 450000, China; NIU Chengzhi, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China

〔Abstract〕 The paper expounds the building method of Medical Knowledge Graph (MKG) library based on Diagnosis - related Group (DRG) in detail, prospects the application of this knowledge graph library, including health education, assistant diagnosis and decision - making system and optimization of DRG burster, etc.

〔Keywords〕 DRG - MKG; assistant diagnosis; health education

1 引言

1.1 研究背景

知识图谱最早于 2012 年 5 月 17 日被 Google 公司正式提出^[1], 随着人工智能技术发展和应用, 现已被广泛应用于智能搜索、个性化推荐等领域^[2]。在国外已有较多重要的知识图谱研究成果, 如

Google Knowledge Graph^[3]、DBpedia^[4]和 Freebase^[5]等。当前人工智能已广泛应用于医疗健康领域, 如在医学影像识别方面已达到顶级医生水平, 甚至在某些疾病诊断过程中智能决策支持系统对疾病的诊断正确率超过一般医生^[6], 人工智能应用于医学领域将会推动传统医学向智能医学转型, 为传统医学带来巨大变革, 逐渐走向智能化发展^[7]。与此同时, 医学人工智能发展可以更快、更准确地解决临床诊断问题, 将人工智能更多地应用到医学领域是大势所趋^[8]。

1.2 相关概念

随着计算机技术发展, 医学知识图谱 (Medical Knowledge Graphs, MKG) 应用越来越广泛, 能够直观地通过图谱形式表达出每个节点之间的

〔收稿日期〕 2020 - 02 - 18

〔作者简介〕 刘新奎, 博士, 副教授, 高级统计师, 硕士生导师, 发表论文 25 篇。

〔基金项目〕 河南省医学科技攻关计划省部共建项目“基于疾病诊断相关分组 (DRG) 的医学知识图谱构建研究” (项目编号: SB201901021)。

联系。疾病诊断相关分组 (Diagnosis - related Group, DRG) 是一种根据患者年龄、疾病诊断、并发症、合并症及治疗方式等因素将患者分成若干 DRG 组进行管理的体系^[9]。根据 DRG 分组器的分组原理,通过知识图谱形式将影响分组的相关信息以知识图形式展现出来,即以直观形式来展示 DRG 分组方式,对于 DRG 推广、应用具有重要意义。

2 DRG - MKG 存在的问题

目前大多数医学知识图谱主要是针对某个科室、某类疾病或药物进行构建,尚未有学者针对 DRG 建立医疗知识图谱库。建立较为客观精确的 DRG 医学知识图谱库需要不断融合不同医疗知识库,将不断发现的新的医学知识纳入到 DRG 医学知识图谱库中。基于 DRG 的医疗知识图谱构建是一个不断更新的过程。我国卫生发展水平以及费用结构与西方国家存在差异,应研究适合我国国情的病例组合方案,一些学者探讨 DRG 在中国的引进和应用,为建立我国病例组合积累资料和经验^[10]。由于各省市病种结构略有不同,DRG 在国内的发展过程中出现各种版本,如北京版、上海版,由此也出现很多问题^[11],MKG 可以辅助进行 DRG 分组器优化,从而进行反向的推进。

3 DRG - MKG 构建方法

3.1 概述

本文在以往研究基础上分析总结现有知识表示模型特点,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并构建 DRG - MKG 库。

3.2 分析整理文本信息及数据

为保证数据真实有效,采用随机抽样方式,利用 OntoManager、OOPS 以及 Core 等文本评估工具对收集到的病历首页进行检测,主要解决数据可能存

在的逻辑、编码标准化等问题。

3.3 标注语料数据

语料库是医学文本挖掘的重要资源,是实现语义标引、机器翻译、知识关联、数据挖掘、智能检索等功能的基础支撑。语料标注过程分为预标注和正式标注。预标注阶段主要完成对标注规范的修订,首先在分析现有医学语料库标注规范基础上形成标注规范初稿,明确标注实体类型范畴及边界等问题,随后采取多轮迭代模式进行规范修订和标注工作,主要利用迁移学习这一机器学习方法将通用领域的知识迁移到医学领域的命名实体识别。

3.4 DRG - MKG 描述体系

利用专家知识,融合 DRG,对现有知识图谱改进。现有医学知识图谱完备性不足,其体系某些部分有所缺失。疾病诊断系统在做判断时,很容易因为医学知识图谱的不完备性发生诊断或判断不明的情况,因此要融合 DRGs 的分组内容,对 MKG 描述体系进行细化。

3.5 实体识别和关系抽取

采用自下向上的构建方法,从非结构化或半结构化的医学文本中抽取医学实体和医学关系。针对医疗领域实体,拟用深度学习架构中的 NN - CRF 架构和滑动窗口分类思想,通过神经网络的学习来完成对实体的抽取;针对医疗领域关系,拟用 Lattice - LSTM 方法,对医学文本中的关系进行提取。

3.6 DRG - MKG 库

现今医学知识图谱库之间的融合技术有了一定发展,但仍需要大量人工干预,知识融合算法效率较低,采用众包方式来解决 DRG 知识图谱库构建中所遇到的知识融合难题。从非结构化或半结构化的医学文本中抽取医学实体和医学关系。针对医疗领域实体,拟用深度学习架构中的 NN - CRF 架构和滑动窗口分类思想,通过神经

网络的学习来完成对实体的抽取；针对医疗领域关系，拟用 Lattice - LSTM 方法，对医学文本中的关系进行提取，以此构建 DRG - MKG 库，见图 1。

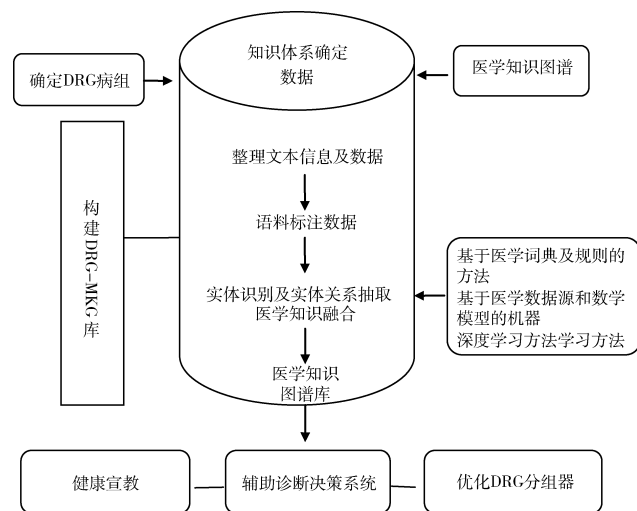


图 1 DRG - MKG 库构建框架

4 DRG - MKG 应用展望

4.1 健康宣教

MKG 为医疗信息系统中海量、异构、动态的医疗大数据表达、组织、管理及利用提供一种更为有效的方式，使系统智能化水平更高，更接近于人类认知思维。DRG - MKG 库能够将较为抽象的信息以可视化方式呈现，提高用户获取知识的效率和完整性。DRG - MKG 通过生动形象的图形信息帮助患者理解疾病相关注意事项。患者对健康宣教知识的掌握及遵守情况将对其康复结果及预防产生较大影响。此外由于患者文化水平参差不齐，往往难以在短时间内接受并掌握大量自护康复信息。通过知识图谱丰富图形及色彩效果吸引患者注意力，使患者在宣教过程中保持关注。利用机器学习算法以及医学数据特点对模型进行训练，然后识别实体。常用方法包括支持向量机、人工神经网络、隐马尔可夫模型、条件随机场等。

4.2 辅助诊断决策系统

使用知识图谱技术能够将结构化知识从大量文

本和图像中抽取出来，知识图谱、大数据技术以及深度学习技术 3 者的结合正在成为推动人工智能发展的核心驱动力。在医疗领域中知识图谱技术具有较好应用前景，针对 DRG - MKG 库的研究将有助于解决优质医疗资源供给不足和医疗服务需求持续增加的矛盾。随着知识图谱技术不断发展，DRG - MKG 的医疗辅助诊断成为可能。这是一种利用抽象建模、机器学习等现代信息化手段进行诊断的辅助系统，能够自动解析相关文本、图像等知识，提取相应特征，之后使用机器学习、深度学习等技术完成更高层次的特征抽象和多重非线性复杂计算，最终自动输出相应诊断结果。DRG - MKG 的医疗辅助诊断能够辅助医生诊断病情，显著提高诊断效率和准确度。特别是当医生接触到疑难杂症时，可能需要花费较多时间查找相关案例以及知识库。而借助智能辅助诊断，可利用已学习到的大量知识进行自动解析与处理，极大提高医生工作效率，降低不必要的医疗资源消耗。

4.3 优化 DRG 分组器

充分利用基于 DRG 的知识图谱库强大的数据处理能力，结合国内外现有 DRG 分组系统，对此次研究样本医院住院病案首页信息进行 DRG 分析，记录不同 DRG 系统分组情况，比较不同 DRG 分组系统分组效果，归纳不同 DRG 系统分组特点，追溯形成不同分组效果的影响因素，由此提出改善和优化现有 DRG 分组系统方案，形成本土化 DRG 分组系统。利用优化分组系统进行模拟测算，适时调整疾病分组模型和权重评价体系。

5 结语

实现 DRG - MKG 知识库的核心是结合病案首页、临床知识等多组学数据，利用基于医学数据源和数学模型的机器学习、基于医学词典及规则的自然语言处理以及深度学习、统计分析等方法，综合考虑病种诊疗数量、平均住院日、平均费用等资料的情况下，利用 OntoManager、OOPS 以及 Core 等文

本评估工具对收集到的病案首页进行检测, 解决数据可能存在的逻辑、编码标准化等问题, 筛选病种, 确定具有研究价值的 DRG 病种。研发面向疾病风险预测、早期诊断的临床决策支持系统, 为 DRG - MKG 库转化提供有效途径。通过构建并完善 DRG - MKG 库, 依托大型医院验证推广, 不仅可以将与 DRG 相关的医学知识抽取出来, 融合形成结构化知识, 还能在此基础上对健康宣教、辅助诊疗进行研究和应用, 推动智能化发展。另外可以促进患者对疾病的客观认知, 尽早配合治疗, 降低医疗费用, 节约医疗资源。

参考文献

- 1 Pujara J, Miao H, Getoor L, et al. Knowledge Graph Identification [M]. Berlin: Springer, 2013.
- 2 侯梦薇, 卫荣, 陆亮. 知识图谱研究综述及其在医疗领域的应用 [J]. 计算机研究与发展, 2018, 55 (12): 2587 - 2599.
- 3 Singhl A. Official Google Blog: introducing the knowledge

graph: things, no strings [J]. Official Google Blog, 2012 (5): 1 - 8.

- 4 Suchanek F M, Kanaseci G, WeiKum G. Yago: a core of semantic knowledge [C]. Homburg: International Conference on World Wide Web, 2007: 697 - 706.
- 5 刘知远, 崔安欣. 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.
- 6 袁冰, 范钢. 中医学如何走进人工智能时代 [J]. 中华中医药杂志, 2018, 33 (2): 698 - 703.
- 7 王锡山. 未来医学时代——人工智能诊疗 [J]. 中华结直肠疾病电子杂志, 2017, 6 (4): 349 - 352.
- 8 周路茜. 医学人工智能开始加速产业化 [J]. 新经济导刊, 2017 (7): 18 - 22.
- 9 刘新奎, 杨林朋. 疾病诊断相关分组 (DRG) 在我国的发展与应用 [J]. 中华医学图书情报杂志, 2018, 27 (8): 48 - 51.
- 10 刘小林. DRGs 在医疗管理中的抓手作用及存在的难题 [J]. 中国卫生产业, 2018, 15 (6): 40 - 41.
- 11 郎婧婧, 江芹, 王珊, 等. 典型国家 DRG 分组的比较研究与启示 [J]. 中国卫生经济, 2017, 36 (4): 50 - 53.

(上接第 19 页)

- 7 Bobak Moazzami, Niloofar Razavi - Khorasani, Arash Dooghaie Moghadam, et al. COVID - 19 and Telemedicine: immediate action required for maintaining healthcare providers well - being [EB/OL]. [2020 - 03 - 21]. https://www.researchgate.net/publication/340068485_Telehealth_for_global_emergencies_Implications_for_coronavirus_disease_2019_COVID-19.
- 8 Kimberly Lovett Rockwell, Alexis S Gilroy. Incorporating Telemedicine as Part of COVID - 19 Outbreak Response Systems [J]. The American Journal of Managed Care, 2020, 26 (4): 147 - 148.
- 9 Robin Ohannessian, Tu Anh Duong, Anna Odone. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID [EB/OL]. [2020 - 04 - 02]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7124951>.
- 10 Adam J Gadzinski, John L Gore. Implementing Telemedicine in Response to the COVID - 19 Pandemic [EB/OL]. [2020 - 04 - 03]. <https://doi.org/10.1097/JU.0000000000001033>.
- 11 Adel Elkbuli, Haley Ehrlich, Mark McKenney. The Effective Use of Telemedicine to Save Lives and Maintain Struc-

ture in a Healthcare System: current response to COVID - 19 [EB/OL]. [2020 - 04 - 07]. <https://doi.org/10.1016/j.ajem.2020.04.003>.

- 12 Ivy Lee, Carrie Kovarik, Trilokraj Tejasvi, et al. Telehealth: helping your patients and practice survive and thrive during the COVID - 19 crisis with rapid quality implementation [J]. J Am Acad Dermatol, 2020, 82 (5): 1213 - 1214.
- 13 Abdulmunir Azizy, Mujtaba Fayaz, Mehmet Agirbasli. Do Not Forget Afghanistan in Times of COVID - 19: telemedicine and the Internet of things to strengthen planetary health systems [EB/OL]. [2020 - 04 - 23]. https://www.researchgate.net/publication/340881470_Do_Not_Forget_Afghanistan_in_Times_of_COVID-19_Telemedicine_and_the_Internet_of_Things_to_Strengthen_Planetary_Health_Systems.
- 14 Jeydith Gutierrez, Ethan Kuperman, Peter J Kaboli. Using Telehealth as a Tool for Rural Hospitals in the COVID - 19 Pandemic Response [EB/OL]. [2020 - 04 - 11]. https://www.researchgate.net/publication/340589793_Using_Telehealth_as_a_Tool_for_Rural_Hospitals_in_the_COVID-19_Pandemic_Response.