

# 探讨 DRG-MKG 库构建方法\*

刘新奎<sup>1</sup>, 杨林朋<sup>1</sup>, 咎红英<sup>2</sup>, 牛承志<sup>1</sup>

1 (郑州大学第一附属医院, 郑州 450052)

2 (郑州大学, 郑州 450000)

**【摘要】** 疾病诊断相关分组 (Diagnosis-related Group, DRG) 和医学知识图谱 (Medical Knowledge Graphs, MKG) 是近年国内外研究的热点。本文对构建基于 DRG 医学知识图谱进行全面解析, 首次将 DRG 与医学知识图谱融合, 构建基于 DRG 的医学图谱知识库 (DRG-MKG), 丰富医学知识图谱内容, 扩展医学知识图谱应用。利用 DRG-MKG 库优化 DRG 分组器, 为基于 DRG 的绩效考核和医保支付改革提供技术支持。构建基于 DRG-MKG 的健康宣教和辅助诊断决策系统, 对于我国 DRG 的推广和医学知识图谱的应用具有重要意义。

**【关键词】** DRG-MKG; 健康宣教; 辅助诊断; 优化分组器

[中图分类号] R197.3 [文献标志码] A [文章编号]

## Discussion on the Construction Method of DRG-MKG Library

LIU Xin-kui, YANG Lin-peng, ZAN Hong-ying, NIU Cheng-zhi

(The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China)

**【Abstract】** The Diagnosis-linked Group (DRG) and Medical Knowledge Graphics (MKG) are hot topics in recent years. This paper comprehensively analyzes the construction of DRG-based medical knowledge maps, integrates DRG with medical knowledge maps for the first time, builds a DRG-based medical map knowledge base (DRG-MKG), enriches the content of medical knowledge maps, and expands the application of medical knowledge maps. The DRG-MKG library is used to optimize the DRG Packer and provide technical support for DRG-based performance assessment and medical insurance payment reform. Constructing health education and auxiliary diagnosis decision system based on DRG-MKG is of great significance to the application of DRG in China and the application of medical knowledge map.

**【Key word】** DRG-MKG; Health Education; Auxiliary Diagnosis; Optimize Group

知识图谱最早于 2012 年 5 月 17 日被 Google 公司正式提出<sup>[1]</sup>, 随着人工智能的技术发展和应用, 现已被广泛应用于智能搜索、个性化推荐等领域<sup>[2]</sup>。在国外已有较多重要的知识图谱研究成果, 如 Google Knowledge Graph<sup>[3]</sup>、DBpedia<sup>[4]</sup>和 Freebase<sup>[5]</sup>等。当前人工智能已广泛应用于医疗健康领域, 如在医学影像的识别方面已经达到顶级医生的水平, 甚至在某些疾病的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的水平<sup>[6]</sup>, 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<sup>[7]</sup>。与此同时, 医学人工智能的发展可以更快、更准确地解决临床诊断问题, 将人工智能更多的应用到医学领域亦是大势所趋<sup>[8]</sup>。

\* 本文系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) 的研究成果之一

随着计算机技术的发展，医学知识图谱（Medical Knowledge Graphs, MKG）的应用越来越广，它能够直观地通过图谱的形式表达出每个节点之间的联系。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<sup>[9]</sup>。如果能够根据 DRG 分组器的分组原理，通过知识图谱的形式将影响分组的相关信息以知识图的形式展现出来，即以直观的形式来展示 DRG 分组的方式，对于 DRG 推广应用具有重要意义。

## 1 DRG-MKG 问题

目前大多数的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。如要建立较为客观精确的 DRG 医学知识图谱库，就需要不断融合不同的医疗知识库，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 的医疗知识图谱的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，这就决定了我们须研究适合我国国情的病例组合方案，为此，一些学者探讨了 DRG 在中国的引进和应用，为建立我国的病例组合积累了资料和经验<sup>[10]</sup>。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了各种版本，比如说北京版与上海版，由此也出现很多的问题<sup>[11]</sup>，MKG 可以帮助我们进行 DRG 分组器优化，从而进行一个反向的推进。

## 2 DRG-MKG 构建方法

本文在以往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。构建 DRG-MKG 库。

### 2.1 分析整理文本信息及数据

为保证数据的真实有效，将采用随机抽样方式，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。

### 2.2 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，本研究首先在对现有医学语料库标注规范进行分析基础上，形成标注规范初稿，明确标注的实体类型的范畴及边界等问题，随后采取多轮迭代的模式进行规范的修订和标注工作，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。

### 2.3 DRG-MKG 描述体系

利用专家知识，融合 DRG，对现有知识图谱改进。现有医学知识图谱有很大的不完备性，其体系某些部分有所缺失。因此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，而发生诊断或者判断不明的情况，因此，要融合 DRGs 的分组内容，对 MKG 的描述体系进行细化。

### 2.4 实体识别和关系抽取

我们将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，抽取出来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的思想，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 的方法，对医学文本中的关系进行

提取。

### 2.5 DRG-MKG 库

现今的医学知识图谱库之间的融合技术有了一定的发展，但仍需要大量人工干预，知识融合算法的效率较低，因此，我们采用众包的方式来解决构建 DRG 知识图谱库中所遇到的知识融合难题。将从非结构化或半结构化的医学文本中，抽取出来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的思想，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 的方法，对医学文本中的关系进行提取，以此构建 DRG-MKG 库如图 1。

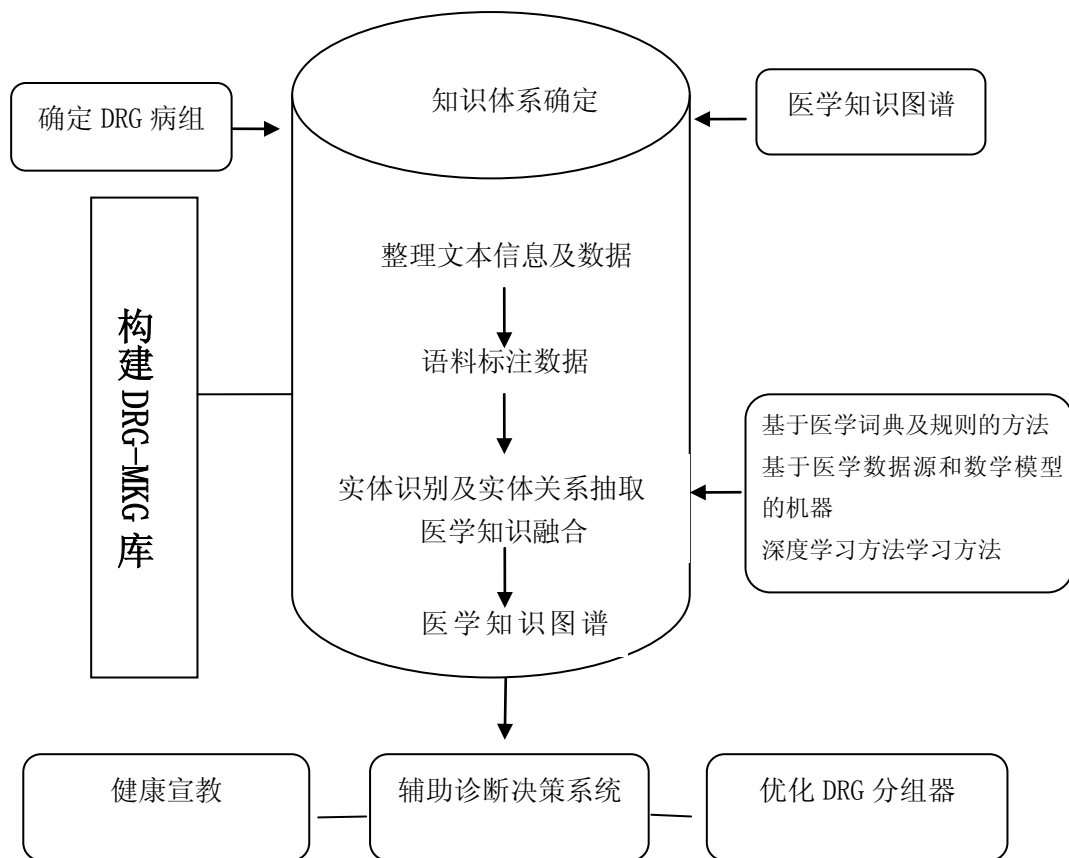


图 1 DRG-MKG 库构建框架

## 3 DRG-MKG 应用展望

### 3.1 健康宣教

MKG 为医疗信息系统中海量、异构、动态的医疗大数据的表达、组织、管理及利用提供了一种更为有效的方式，使系统的智能化水平更高，更加接近于人类的认知思维。其提供了从“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。DRG-MKG 库能够将较为抽象的信息以可视化的方式进行呈现，提高用户获取相关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并接受疾病的相关注意事项。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会产生较大的影响。同时，由于患者的文化层次参差不齐，往往难以在短时间内接受并掌握大量的自护康复信息。利用知识图谱的丰富图形及色彩效果，吸引患者的注意力，使患者在宣教过程中能够保持积极关注。利用

机器学习算法，利用医学数据的特点对模型进行训练，然后识别实体。常用的方法包括支持向量机、人工神经网络、隐 Markov 模型、条件随机场等。

### 3.2 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三者的结合，正在成为推动人工智能发展的核心驱动力。目前在医疗领域中，知识图谱技术也具有着不错的应用前景，针对 DRG-MKG 库的研究将有效帮助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。而随着知识图谱技术的不断发展，DRG-MKG 的医疗辅助诊断也成为了可能。

DRG-MKG 的医疗辅助诊断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。它能够自动化地解析与之相关的文本、图像等知识，并进行相应的特征提取，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，最终自动输出相应的诊断结果。

DRG-MKG 的医疗辅助诊断不但能够辅助医生诊断病人的病情，还能够显著提高诊断的效率以及诊断的准确度。特别是当医生接触到疑难杂症时，可能需要花费相当多的时间去找相关案例以及知识库。而借助智能辅助诊断，就可以利用已经学习到的大量知识，进行自动化解析与处理。由此，大大提高医生的工作效率，降低不必要的医疗资源消耗。

### 3.3 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，结合国内外现存的 DRG 分组系统，对此次研究的样本医院的住院病案首页信息进行 DRG 分析，记录不同 DRG 系统的分组情况，从而比较不同 DRG 分组系统的分组效果，归纳不同 DRG 系统的分组特点，追溯造成不同分组效果的影响因素，由此提出改善和优化现有 DRG 分组系统的方案，形成本土化的 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。

## 4 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、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、菲尔德法等方法，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、早期诊断的诊疗临床决策支持系统，为 DRG-MKG 库转化提供有效途径。通过 DRG-MKG 库的构建和完善，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，融合形成结构化的知识，还能在此基础上对健康宣教、辅助诊疗进行研究和应用，推动智能化的发展，另外，还可以促进患者对疾病的客观认知，尽早配合治疗，从而降低医疗费用，节省医疗资源，对住院治疗产生的总费用进行预测，将明显提高重大疾病的医疗水平和防治效益。

### 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.
- [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]周路菡. 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, 于丽华. 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.

(通讯作者: 刘新奎 E-mail:liuxk666@163.com)

### 作者贡献声明\* :

- 刘新奎: 提出研究思路, 设计研究方案;  
杨林朋, 咎红英: 研究过程的调查实施;  
刘新奎, 杨林朋: 采集、清洗和分析数据;  
杨林朋, 牛承志: 论文起草  
刘新奎: 论文最终版本修订。