

构建难度还是很高的。

传统的知识推理方法包括基于描述逻辑的推理、

基于规则的推理、基于分布式的知识推理等,各方法的比较见表4。

表4 推理方法的比较^[34]

推理方法	推理方式	方法优点	方法缺点
基于描述逻辑的推理	借助 TBox 和 ABox 工具将推理问题归结为 ABox 的一致性检验问题	达到了表达能力与推理复杂度的平衡	数据利用率不高
基于规则的推理	根据简单规则和统计特征的推理	可解释性强;规则正确时,准确率高	对规则的依赖度高,规则制定困难且适用性较低
基于分布式的推理	对知识图谱的低维向量进行推理	计算方便快捷	由于忽视深层次的组合语义信息,推理能力低
混合推理	结合多种方法的推理	结合了不同方法的优势	结合效率差,即:一种推理方法为主,另一种推理方法为辅,缺乏更深层次的合作

这些方式很难满足医学大数据下的快速推理和对于增量知识和规则的快速加载,所以现在应用更为广泛的是结合人工智能技术的知识推理模型,常见的有人工神经网络模型 (artificial neural network model)、遗传算法 (genetic algorithm) 和反向传播网络模型 (back propagation) 等. 文献^[35]中就提出了一种表示本体,以将文献抽象数据表征为4个知识元素(背景,目标,解决方案和发现). 案例研究表明,所提出的本体模型可以用来表示嵌入在文献摘要中的知识,并且可以通过 NLP 模型自动提取本体元素. 所提出的框架可以增强文献计量分析,以从文献中探索更多知识,实现知识推理的功能。

无论是传统的知识推理方法还是人工智能技术的推理方法都是以知识图谱作为数据源进行推理,而图挖掘计算则是基于图论的相关算法,把知识图谱看作图,把医学实体看作节点,实体间的关系看作边,实现对图谱的探索和挖掘,更有利于解决大规模的图数据分析问题^[36]. 基于此, Jagvaral 于 2019 年提出具有注意机制的 CNN-BiLSTM 方法用于知识图谱基于路径的推理^[37]. 论文中提到,他们研发的路径编码器从大型图形的路径中提取特征更有效,更是说明了应用多步推理在基于路径的推理中可能会有用. 此项研究只使用一种类型来表示实体,而大多数知识图谱中的实体具有多种类型,因此,多种类型合并到路径编码中的路径推理推理还有待研究。

以上为比较具体的领域知识图谱构建流程,虽然领域知识图谱应用比较广,但目前还尚未实现自动构建,而在 2018 年,清华大学知识工程实验室发表一篇

名为“一种准确而高效的领域知识图谱构建方法”的文章^[38],介绍了一种快速构建较高质量的领域知识图谱的方法,为领域知识图谱构建提供另一种思路,该方法称为“四步法”:① 领域本体构建;② 众包半自动语义标注;③ 外源数据补全;④ 信息抽取. 在领域知识图谱构建过程中,权衡效率和准确率,平衡自动化和人工构建,以高效地构建图谱,这是当前面临的一个很大问题。

3.5 知识图谱绘制工具

图5是以心律失常为关键词绘制的医学领域知识图谱,它展现了知识图谱力导向布局图的视图形式。

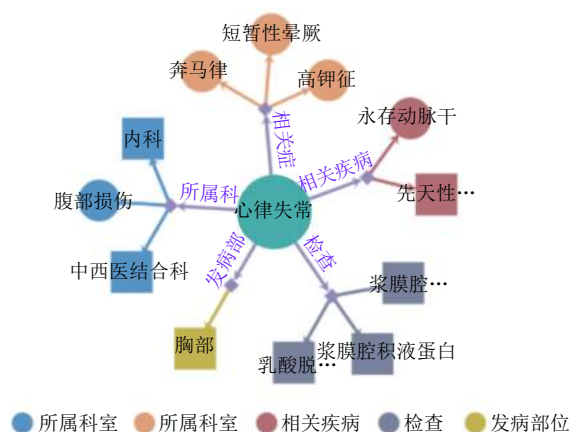


图5 医疗领域知识图谱举例

知识图谱的绘制工具可分为两大类:通用软件,如 SPSS、Ucinet、PajekWordsmithTools 和 GIS 等. 另一类是专门用于知识图谱绘制的软件,也有许多类型,有些是针对某些特定领域,有些是个人未公开的. 表5对知识图谱绘制工具做一个汇总。

表5 知识图谱绘制工具^[32]

绘制工具	说明	功能描述
SPSS ^[39]	美国 SPSS 公司 20 世纪 80 年代开发的大型统计分析软件, 商业用途.	具有完整的数据输入、编辑、统计分析、报表、图形绘制等功能. 常用于多元统计分析、数据挖掘和数据可视化.
Bibexcel ^[39]	瑞典科学计量学家 Persoon 开发的科学计量学软件, 用于科学研究免费软件.	具有文献计量分析、引文分析、共引分析、耦合分析、聚类分析和数据可视化等功能. 可用于分析 ISI 的 SCI、SSCI 和 A&HCI 文献数据库.
Histicite ^[40]	Eugene Garfield 等于 2001 年开发的科学文献引文链接分析和可视化系统, 免费软件.	可对 ISI 的 SCI、SSCI 和 SA&HCI 等文献数据库的引文数据进行计量分析, 生成文献、作者和期刊的引文矩阵和实时动态引文编年图. 直观的反映文献之间的引用关系、主题的宗谱关系、作者历史传承关系、科学知识发展演进等.
CiteSpace ^[41]	陈超美博士开发的专门用于科学知识图谱绘制的免费软件.	可用于追踪研究领域热点和发展趋势, 了解研究领域的研究前沿及演进关键路径, 重要的文献、作者及机构. 可用于对 ISI、CSSCI 和 CNKI 等多种文献数据库进行分析.
Pajek ^[42]	来自斯洛文尼亚的分析大型网络的社会网络分析免费软件.	Pajek 基于图论、网络分析和可视化技术, 主要用于大型网络分解, 网络关系展示, 科研作者合作网络图谱的绘制.

4 领域知识图谱的现状和应用

4.1 领域知识图谱的现状

随着近几年知识图谱技术的发展, 知识图谱研究与落地发生了一些转向. 其中一个重要变化就是领域

知识图谱的建设成为主流. 知识图谱技术与各行业的深度融合已经成为一个重要趋势^[4].

接下来, 本文对搜索、医疗、电商、社交、教育这几个热门领域规模比较大的知识图谱进行汇总, 见表 6.

表6 热门领域知识图谱汇总

领域	知识图谱
搜索领域	百度百科知识图谱、Google 知识图谱
医疗领域	中医药知识图谱 (http://www.tcmkb.cn/kg/index.php)、中医药知识地图 (http://www.tcmkb.cn/knowledge_maps/index.php)、中文医学知识图谱 (http://zstp.pcl.ac.cn:8002/)
电商领域	阿里知识图谱、京东商品知识图谱
社交领域	Facebook 推出的 Graph Search
教育领域	中国基础教育知识图谱 (edukg.org)、课程图谱 (http://coursegraph.com/)、百度教育大脑、

医疗领域是当前建设很火热的领域, 仅是对中文医学知识图谱的相关检索就达 200 多条, 大到中文疾病知识图谱, 小到甲状腺知识图谱, 医疗领域知识图谱的理论实践化是有原因的: (1) 医疗信息化浪潮. 步入信息化社会以来, 医疗信息化的发展从未停歇过, 从最初的医院信息系统开始, 电子病历、临床智慧医疗等技术层出不穷. (2) 庞大的医学数据. 除医院提供的病例信息, 基因组学, 蛋白组学也给医疗领域贡献了大量的数据. (3) 人工智能出现后, 为体量庞大的医学数据处理提供方向. 知识图谱正是作为大数据到人工智能的理想桥梁. 整合异构数据, 建立语义关系, 最重要的是知识推理, 医疗知识图谱在智慧医疗的建设中起到越来越重要的作用, 通过知识问答, 知识推理将更好的为社会服务. 所以医疗知识图谱发展迅速. 与之相似, 教育领域同样具有数据量大, 面临信息化建设等优点, 相信教育知识图谱也将会得到越来越多的关注.

4.2 领域知识图谱的应用

知识图谱作为近十年内新兴的概念, 其可以将各种信息和数据整合为知识, 为各研究领域提供可视化分析, 各类大规模知识图谱在智能搜索、智能问答、智能推荐、情报分析等方面发挥了重要作用.

4.2.1 智能搜索

基于知识图谱的智能搜索可以直接给出知识卡片而不是给出相关的链接序列. 在知识图谱的帮助下, 搜索引擎可以将搜索关键词映射到知识图谱中匹配度较高的一个或一组概念上, 最后以知识卡片的形式展现给用户. 知识卡片可以以 3 种形式展示知识^[3]: ① 对于单一关键词的搜索, 返还用户查询的实体的结构化摘要. 比如搜索姚明, 将给出姚明的身份介绍以及主要关系介绍; ② 对于问题类的搜索, 知识卡片直接给出答案. 比如搜索“姚明的身高是多少?”, 搜索结构将是显示 226.0 cm 的知识卡片; ③ 对于模糊类的查询, 将给

出相关网页列表. 例如搜索“姚明最近的活动有哪些?”, 搜索结果是包含姚明活动的新闻网页.

4.2.2 智能问答

Gowild 狗尾草的 AI 虚拟生命“琥珀虚颜”和苹果的智能语音助手 Siri 都是知识图谱应用于智能问答方面的实例. 智能问答是信息检索系统的一种高级形式, 能够用自然语言为用户提供问题的解答或者实现人机交流. 目前, 语音助手研发十分火热, 比如百度自然语言部开发的小度机器人, 阿里巴巴人工智能实验室研发的天猫精灵, 亚马逊 Alexa 语音服务等都是为智能问答更加智能、准确做出地探究.

4.2.3 智能推荐

电商、教育、社交等行业都需要借助大数据行为分析进行用户画像, 以指导广告投放和提高用户体验. 相较于原先对关联性较差的数据进行用户行为分析, 知识图谱一个天然的优势就是更突出数据之间的关系, 这样就能根据知识关联关系获得更加精确的用户画像, 有助于精准营销、精细化运营. 除了用户画像, 智能推荐还要依靠商品之间的关联提供使用建议、搭配等.

4.2.4 情报分析

江苏大学刘桂峰利用 CiteSpace 软件信息可视化方法, 对 1990–2010 年间来自 Web of Science (SCIE) 数据库的太赫兹技术领域研究的文献数据进行统计和可视化分析, 揭示出该领域的领军人物、知识基础和研究前沿等信息^[43]. 赵蓉英等^[44]利用 CiteSpace II 的爆发词探测方法绘制知识图谱, 并绘制爆发词随时间演化的学科前沿发展趋势图, 进而发现学科前沿. 胡泽文等在文献^[28]中借助通过 CiteSpace II 界定了改革开放以来情报学的 3 个发展阶段. CiteSpace 是一款应用于科学文献中识别并显示科学发展新趋势和新动态的软件, 通过它绘制知识图谱, 能够发现经典文献、研究热点和研究前沿. 可见知识图谱用于情报分析方面有很大的发展潜力.

除此之外, 知识图谱应用于医学、教育等领域, 对于建设智能医疗、智慧教育起着支撑作用.

5 结语

知识图谱从最初作为辅助 Google 搜索的技术被提出, 到现在很多行业都在建设自己的知识图谱, 它的价值正在被慢慢挖掘出来. 知识图谱不是知识的终点, 但是它确实能解决很多学科领域的瓶颈问题, 成为智

能化建设的基石.

结合医学知识图谱的构建和发展, 本文认为信息抽取技术仍是当前的研究热点, 最理想的信息抽取方式是结合实体抽取、关系抽取和属性抽取三者的联合抽取, 但该技术还没有典型代表. 而知识推理作为知识图谱最大的亮点和功能, 将其技术发展成熟还需要付出很大地努力. 在人工智能还有很大发展潜力的今天, 借助人工智能技术实现知识推理有很大的发展前景. 知识推理不仅是智能问答、智能推荐等应用的关键技术, 更是智能化建设的基石.

对于领域知识图谱的发展方向, 本文倾向于领域划分更精细, 领域交互更频繁的发展方向. 类比于医学领域中各种疾病的知识图谱, 也许教育领域会出现各种学科知识图谱, 因为越精细, 专业性越强, 知识越准确. 这也是越来越多的人主张建立企业知识图谱的原因. 此外, 各领域的知识图谱不该是独立存在的, 领域知识图谱之间有交互, 才能真正地构成知识网.

知识图谱仍在发展初期, 笔者仅希望通过本文的写作, 能抛砖引玉, 吸引更多人了解这门技术并投入到相关的研究中来.

参考文献

- 1 Huang HQ, Yu J, Liao X, *et al.* Review on knowledge graphs. *Computer Systems & Applications*, 2019, 28(6): 1–12.
- 2 Pan JZ, Vetere G, Gomez-Perez JM, *et al.* Exploiting Linked Data and Knowledge Graphs in Large Organisations. Cham: Springer, 2017.
- 3 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述. *电子科技大学学报*, 2016, 45(4): 589–606. [doi: 10.3969/j.issn.1001-0548.2016.04.012]
- 4 肖仰华. 领域知识图谱落地实践中的问题与对策. 复旦大学知识工厂实验室. https://www.sohu.com/a/280006592_100099320. (2018-12-06)[2019-10-05].
- 5 李涓子, 侯磊. 知识图谱研究综述. *山西大学学报(自然科学版)*, 2017, 40(3): 454–459.
- 6 侯梦薇, 卫荣, 陆亮, 等. 知识图谱研究综述及其在医疗领域的应用. *计算机研究与发展*, 2018, 55(12): 2585–2599.
- 7 袁凯琦, 邓扬, 陈道源, 等. 医学知识图谱构建技术与研究进展. *计算机应用研究*, 2018, 35(7): 1929–1936. [doi: 10.3969/j.issn.1001-3695.2018.07.002]
- 8 van der Heijden M, Lucas PJF. Extracting qualitative knowledge from medical guidelines for clinical decision-

- support systems. Proceedings of AIME 2009 Workshop KR4HC 2009 Knowledge Representation for Health Care. Verona, Italy. 2017. 100–112.
- 9 Merhej E, Schockaert S, Mckelvey TG, *et al.* Generating conflict-free treatments for patients with comorbidity using answer set programming. Proceedings of HEC 2016 International Joint Workshop on Knowledge Representation for Health Care. Munich, Germany. 2017. 111–119.
- 10 Bonacin R, Pruski C, Da Silveira M. Careflow personalization services: Concepts and tool for the evaluation of computer-interpretable guidelines. Proceedings of AIME 2011 Workshop KR4HC 2011 Knowledge Representation for Health-Care. Bled, Slovenia. 2012. 80–93.
- 11 López-Vallverdú JA, Riaño D, Collado A. Detecting dominant alternative interventions to reduce treatment costs. Proceedings of AIME 2011 Workshop KR4HC 2011 Knowledge Representation for Health-Care. Bled, Slovenia. 2012. 131–144.
- 12 Kaiser K, Seyfang A, Miksch S. Identifying treatment activities for modelling computer-interpretable clinical practice guidelines. Proceedings of ECAI 2010 Workshop KR4HC 2010 Knowledge Representation for Health-Care. Lisbon, Portugal. 2011. 114–125.
- 13 Henriksson A, Zhao J, Dalianis H, *et al.* Ensembles of randomized trees using diverse distributed representations of clinical events. BMC Medical Informatics and Decision Making, 2016, 16(Suppl 2): 69.
- 14 Friedman C, Alderson PO, Austin JHM, *et al.* A general natural-language text processor for clinical radiology. Journal of the American Medical Informatics Association, 1994, 1(2): 161–174. [doi: [10.1136/jamia.1994.95236146](https://doi.org/10.1136/jamia.1994.95236146)]
- 15 Mu XF, Wang W, Xu AP. Incorporating token-level dictionary feature into neural model for named entity recognition. Neurocomputing, 2019, 375: 43–50.
- 16 Azalia FY, Bijaksana MA, Huda AF. Name indexing in Indonesian translation of hadith using named entity recognition with Naïve Bayes classifier. Procedia Computer Science, 2019, 157: 142–149. [doi: [10.1016/j.procs.2019.08.151](https://doi.org/10.1016/j.procs.2019.08.151)]
- 17 Ghiasvand O, Kate RJ. Learning for clinical named entity recognition without manual annotations. Informatics in Medicine Unlocked, 2018, 13: 122–127. [doi: [10.1016/j.imu.2018.10.011](https://doi.org/10.1016/j.imu.2018.10.011)]
- 18 Zhang CW, Li YL, Du N, *et al.* On the generative discovery of structured medical knowledge. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018:). London, UK. 2018. 2720–2728. [doi: [10.1145/3219819.3220010](https://doi.org/10.1145/3219819.3220010)]
- 19 Lee HG, Park G, Kim H. Effective integration of morphological analysis and named entity recognition based on a recurrent neural network. Pattern Recognition Letters, 2018, 112: 361–365. [doi: [10.1016/j.patrec.2018.08.015](https://doi.org/10.1016/j.patrec.2018.08.015)]
- 20 Wang DS, Tiwari P, Garg S, *et al.* Structural block driven enhanced convolutional neural representation for relation extraction. Applied Soft Computing, 2020, 86: 105913. [doi: [10.1016/j.asoc.2019.105913](https://doi.org/10.1016/j.asoc.2019.105913)]
- 21 李智恒, 桂颖溢, 杨志豪, 等. 基于生物医学文献的化学物质致病关系抽取. 计算机研究与发展, 2018, 55(1): 198–206.
- 22 Zhang Y, Lin H, Yang Z, *et al.* A hybrid model based on neural networks for biomedical relation extraction. Journal of Biomedical Informatics, 2018: S1532046418300534.
- 23 Li Z, Yang JS, Gou X, *et al.* Recurrent neural networks with segment attention and entity description for relation extraction from clinical texts. Artificial Intelligence in Medicine, 2019, 97: 9–18. [doi: [10.1016/j.artmed.2019.04.003](https://doi.org/10.1016/j.artmed.2019.04.003)]
- 24 张元博. 医疗知识图谱构建与应用[硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2018.
- 25 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述. 计算机研究与发展, 2016, 53(3): 582–600. [doi: [10.7544/issn1000-1239.2016.20148228](https://doi.org/10.7544/issn1000-1239.2016.20148228)]
- 26 Cardillo E, Serafini L, Tamin A. A hybrid methodology for consumer-oriented healthcare knowledge acquisition. Proceedings of AIME 2009 Workshop KR4HC 2009 Knowledge Representation for Health-Care. Data, Processes and Guidelines. Verona, Italy. 2010. 38–49.
- 27 Zelaia A, Arregi O, Sierra B. Combining singular value decomposition and a multi-classifier: A new approach to support coreference resolution. Engineering Applications of Artificial Intelligence, 2015, 46: 279–286. [doi: [10.1016/j.engappai.2015.09.007](https://doi.org/10.1016/j.engappai.2015.09.007)]
- 28 Zuheros C, Tabik S, Valdivia A, *et al.* Deep recurrent neural network for geographical entities disambiguation on social media data. Knowledge-Based Systems, 2019, 173: 117–127. [doi: [10.1016/j.knosys.2019.02.030](https://doi.org/10.1016/j.knosys.2019.02.030)]
- 29 Delgado AD, Martínez R, Montalvo S, *et al.* Person name disambiguation on the web in a multilingual context. Information Sciences, 2018, 465: 373–387. [doi: [10.1016/j.ins.2018.07.024](https://doi.org/10.1016/j.ins.2018.07.024)]
- 30 Mendes PN, Hleisen H, Bizer C. Sieve: Linked data quality assessment and fusion. Proceedings of the 2012 Joint EDBT/ICDT Workshops. Berlin, Germany. 2012. 116–123.

- 31 解峥, 王盼卿, 彭成. 本体的自动构建方法. 电子设计工程, 2015, 23(15): 39–41. [doi: 10.3969/j.issn.1674-6236.2015.15.012]
- 32 胡泽文, 孙建军, 武夷山. 国内知识图谱应用研究综述. 图书情报工作, 2013, 57(3): 131–137, 84. [doi: 10.7536/j.jssn.0252-3116.2013.03.024]
- 33 McDaniel M, Storey VC, Sugumaran V. Assessing the quality of domain ontologies: Metrics and an automated ranking system. *Data & Knowledge Engineering*, 2018, 115: 32–47.
- 34 官赛萍, 靳小龙, 贾岩涛, 等. 面向知识图谱的知识推理研究进展. 软件学报, 2018, 29(10): 2966–2994. [doi: 10.13328/j.cnki.jos.005551]
- 35 Chen HN, Luo XW. An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing. *Advanced Engineering Informatics*, 2019, 42: 100959. [doi: 10.1016/j.aei.2019.100959]
- 36 Yan D, Tian YY, Cheng J. *Systems for Big Graph Analytics*. Cham: Springer, 2017.
- 37 Jagvaral B, Lee WK, Roh JS, *et al.* Path-based reasoning approach for knowledge graph completion using CNN-BiLSTM with attention mechanism. *Expert Systems with Applications*, 2020, 142: 112960. [doi: 10.1016/j.eswa.2019.112960]
- 38 杨玉基, 许斌, 胡家威, 等. 一种准确而高效的领域知识图谱构建方法. 软件学报, 2018, 29(10): 2931–2947. [doi: 10.13328/j.cnki.jos.005552]
- 39 岳晓旭, 袁军鹏, 高继平, 等. 常用科学知识图谱工具实例对比. 数字图书馆论坛, 2014, (5): 66–72. [doi: 10.3772/j.issn.1673-2286.2014.05.011]
- 40 董立平, 郭继军. 利用 Histcite 的人胚胎干细胞引文编年图主要路径分析. 医学信息学杂志, 2010, 31(11): 38–40, 49. [doi: 10.3969/j.issn.1673-6036.2010.11.011]
- 41 陈超美, 陈悦, 侯剑华, 等. CiteSpace II: 科学文献中新趋势与新动态的识别与可视化. 情报学报, 2009, 28(3): 401–421.
- 42 Batagelj V, Mrvar A. Analiza Sieci Społecznych Pajek. <http://mrvar.fdv.uni-lj.si/pajek/>. 2016.
- 43 刘桂锋, 杨国立. 基于 CiteSpace II 的国际太赫兹技术知识图谱研究. 图书情报研究, 2012, 5(3): 47–53.
- 44 赵蓉英, 许丽敏. 文献计量学发展演进与研究前沿的知识图谱探析. 中国图书馆学报, 2010, 36(5): 60–68.