基于知识增强的医学语言模型:现状、 技术与应用*

康砚澜 郭倩宇 张文强1,2 王昊奋3

(¹复旦大学工程与应用技术研究院,上海 200433 ²复旦大学计算机学院 上海 200433 ³同济大学 上海 200438)

[摘要] 目的/意义 介绍生成式语言模型在医学领域的应用现状和挑战,并提出一种基于知识增强的医学语言模型,以提高模型专业性、准确性和可信性,为医学、语言模型及知识图谱领域相关研究人员提供参考。方法/过程 回顾大语言模型的发展、现状及主要技术,分析其在数据安全、专业性、伦理规范和模型可解释性等方面面临的挑战。介绍医学生成式语言模型常见应用场景和技术要点,重点阐述基于知识图谱和多模态数据融合知识增强的医学语言模型,包括其优势、技术原理和具体案例。结果/结论知识增强的医学语言模型可提高语言模型对专业医学知识的理解、认知和应用能力,增强对自然语言的生成能力,拓展对多模态数据的处理能力,在医疗问答、智能辅助诊断、个性化医疗决策等方面具有广泛应用前景。

[关键词] 生成式语言模型; 医学问答; 知识图谱; 人工智能; 医疗

[中图分类号] R-058 [文献标识码] A [DOI] 10. 3969/j. issn. 1673-6036. 2023. 09. 002

Knowledge - enhanced Medical Language Models: Current Status, Techniques, and Applications

KANG Yanlan¹, GUO Qianyu², ZHANG Wengiang^{1,2}, WANG Haofen³

¹Academy for Engineering and Technology, Fudan University, Shanghai 200433, China; ²School of Computer Science, Fudan University, Shanghai 200433, China; ³Tongji University, Shanghai 200438, China

[Abstract] Purpose/Significance The paper introduces the application status and challenges of generative language model in the medical field, and proposes a knowledge – enhanced medical language model to improve the specialization, accuracy and credibility of the model, and provides references for researchers in the fields of medicine, language model and knowledge graph. Method/Process It reviews the development, current status, and major technologies of large language models, and analyzes the challenges in data security, professionalism, ethics, and model interpretability. It introduces the common application scenarios and technical points of medical generative language model, and focuses on the medical language model based on knowledge graph and multi – modal data fusion knowledge enhancement, including the advantages, technical principles and specific cases. Result/Conclusion The knowledge – enhanced medical language model can improve the understanding, cognition and application capability of language model to professional medical knowledge, enhance the generative capability of natural language, and expand the processing capability of multi – modal data, which has a wide application prospect in medical question answering, intelligent assisted diagnosis, personalized medical decision making and so on.

[Keywords] generative language model; medical question answering; knowledge graph; artificial intelligence (AI); medical treatment

[[]修回日期] 2023-09-19

[[]作者简介] 康砚澜,博士研究生;通信作者:张文强,研究员;王昊奋,特聘研究员。

[[]基金项目] 国家自然科学基金项目(项目编号: 62176185)。

1 引言

生成式语言模型是自然语言处理领域的核心技术之一,能够利用大规模无标注文本自动学习语言的语义、结构、语法等知识,从而提升对自然语言的理解能力及生成能力,根据不同输入条件和目标生成多种风格和形式的文本。随着预训练技术和变压器模块(Transformer)的发展,以生成式预训练转换器(generative pretrained transformer – 3.5, GPT)系列^[1-9]为代表的通用大语言模型在文本生成、对话、问答等多种任务中展现出强大能力,广泛应用于办公、推荐系统、服务问答等。

随着自动化辅助医疗和互联网医疗的发展,高 端医疗服务机器人、高质量临床问答诊断、专业健 康助手等均对语言模型与医学领域的结合提出更高 要求。然而, 医学领域包含大量专业医学知识及用 语、多模态复杂医疗影像等,通用语言模型难以通 过大量无监督自然语料训练习得。同时, 医疗领域 对语言模型的可解释性、可信度、效率等有更高要 求。因此,虽然生成式语言模型可应用于医学领域 的药品说明理解、病例报告生成、简单的医疗服务 对话等[10],但仍然不能满足临床问答等专业医学任 务[11]。为解决以上问题,研究人员基于知识引入的 方法提升通用语言模型对医疗专业知识的理解和应 用能力。构建更全面、专业的医学知识图谱、知识 库,探索在医疗问答中更准确的知识增强方法是提 升医学语言模型问答有效性的关键。基于此,本文 综述了生成式语言模型在医学领域的发展现状和面 临挑战、基于知识增强的技术原理及其在医疗领域 的主要应用前景等,以期为相关研究人员快速了解 语言模型在医疗领域中的研究现状和未来发展方 向、推动通用语言模型在医疗领域中更专业、广泛 和深入应用提供参考。

2 生成式人工智能前夕: 大语言模型的崛起

2.1 通用人工智能的发展

通用人工智能是指在各个领域和环境中, 机器

能够像人类一样具备广泛的认知能力、自主适应复杂环境,并实现自主感知、理解、学习、推理、决策与规划。近年来,人工智能在不同领域和场景取得重大进展,如围棋^[12-14]、图像生成^[15-17]和蛋白质结构预测^[18-22]等。作为通用人工智能的核心能力之一,自然语言理解和生成在自然语言处理任务中取得显著突破,以 ChatGPT 为代表的生成式人工智能(AI - generated content,AIGC)在文本生成、问答等任务上取得重大突破^[1-6]。在现有技术条件下,大语言模型成为实现通用人工智能的有效技术路径。

2.2 通用大语言模型

2.2.1 核心思想 通用大语言模型是一种能够理解和生成自然语言文本的深度学习模型。其核心思想是通过掩码预测下一个词汇,从而训练模型以实现对自然语言的理解和生成。为了实现这一目标,通用大语言模型需要大规模模型参数和预训练数据。通过大规模参数和海量预训练数据应用,通用大语言模型实现了泛化能力。在预训练阶段,其学习通用的语言表示、词法和句法信息以及世界知识。基于此,在文本分类、命名实体识别、机器翻译等传统自然语言处理任务中能够达到与单一领域最先进(state of the art, SOTA)模型性能相当的水平。此外,在多轮对话、复杂语言推理和自动编码等领域,通用大语言模型展现了出色的性能,并具备情景学习、指令遵循和思维链推理等能力。

2.2.2 情景学习(in - context learning, ICL) 是一种基于贝叶斯推断的学习方法,能够使模型根据任务示例或指令生成预期输出,无需额外训练。这种方法在多个自然语言处理基准测试中表现出有效性。利用概念蕴含的语义先验知识,模型可以学习输入、标签和输出之间的映射关系,从而提高在特定任务上的性能。

2.2.3 指令遵循 是一种提高模型泛化能力的方法,使模型能够根据任务指令执行新任务,而无需明确示例。如谷歌公司 T5 模型^[23],其使用文本到文本的转换框架,以统一各种自然语言处理任务,并且可以根据自然语言指令生成不同类型输出,如

摘要、翻译、对话等。

2.2.4 思维链推理(chain - of - thought, CoT) 是一种通过分解推理步骤处理复杂任务、增强模型推理能力和可解释性的方法。其体现了语言模型对推理过程的深度建模和控制能力。思维链推理采用提示策略,将问题分解为多个子问题,并将每个子问题的答案作为下一个子问题的输入,形成连续推理链条。语言模型可以逐步解决每个子问题并最终汇总得出完整答案。这种方法不但提高了模型的准确性,还增强了推理过程的可解释性,从而提高模型的可信度和用户体验。

3 医学生成式语言模型场景需求及技术要点

3.1 医学文本生成功能及应用

医学文本生成是自然语言处理在医疗行业中的重

要应用。其利用自然语言生成技术,根据医学知识图 谱、专业数据库、医学影像、医学信号等数据源,生 成专业化医学文本,以实现辅助阅读、创建病例报 告、编写摘要、提供问答建议等多种功能。近年来, 生成式语言模型在医学领域取得显著进展和突破。例 如, 生成式预训练转换器 4 (generative pretrained transformer 4, GPT-4) 在医学问答任务上展现出较 强的性能和准确性[24];专门针对生物医学文本预训 练的大语言模型 PubMed GPT[25] 在多个医学基准上达 到较先进水平。Galactica^[26]模型是集成科学知识图 谱进行预训练的大语言模型, 经测试发现在医学领 域具备一定零样本生成能力。谷歌的 Med - PaLM 2[27] 模型是一种针对医学问题进行优化的垂直领域 大语言模型, 能够针对患者问题提供专家级医师水 平的回复,并在多个国家的医学测试中达到及格水 平。医学生成式语言模型技术框架,见图1。



图 1 医学生成式语言模型技术框架

3.2 常见应用场景

- 3.2.1 医学报告自动生成 基于医学影像、检测结果和其他医疗数据,语言模型可以自动生成结构 化或非结构化的医学报告。这些报告可能涉及 X 光诊断、心电图分析、血液化验单等,其信息量大且要符合专业规范,通过生成报告可以极大减轻医生的文书负担。
- 3.2.2 个性化医疗建议生成 语言模型可以根据 患者的病史、症状等信息,自动生成针对个体情况 的医疗咨询、治疗方案或用药指导,提供个性化医 患服务。

- 3.2.3 医学知识精简总结 语言模型可以利用医学文献、数据库和知识图谱,对特定医学主题进行汇总、提炼和知识精简,自动生成知识概述或病例总结,帮助医生快速掌握医学前沿进展。
- 3.2.4 医患问答服务 语言模型可以理解患者和 医生的问题,提供准确回复,涵盖病因、治疗、预 后等信息。在问诊、健康咨询等场景中应用,可以 改善医患沟通效率。

3.3 技术要点

3.3.1 数据收集 从互联网获取大量结构化和非结构化医学文本数据,如医学图像、检验报告、病

例描述等。通过专业的数据清洗和标注工作去除无 关、有害或有偏见的信息,结合医学论文、专著等 文本提高预训练语料质量。

- 3.3.2 模型训练 采用无监督预训练方法,基于 Transformer 的单一解码器架构,使模型能够学习和 掌握医学文本统计规律。
- 3.3.3 模型微调 采用监督微调方法,通过多样 化的高质量指令数据,使模型具备泛化理解能力, 如对话问答和文本生成等,并提升模型在不同任务 上的性能和准确性。
- 3.3.4 意图对齐 在此过程中,通常由人类标注 师对模型输出结果进行评分,并基于评分训练一个自动打分的奖励模型,通过奖励模型与微调后的语言模型之间的反馈和迭代,使模型能够输出符合人类医师价值观的结果,避免产生违背医学伦理或带有偏见的内容。
- 3.3.5 低秩自适应 针对需要深入领域知识的任 务,可以采用低秩自适应方法,而无须重新训练整

个模型。该方法冻结预训练模型权重,并在每个 Transformer 层中引入秩分解矩阵,从而使模型适应 特定医学专科任务。

4 医学生成式语言模型的兴起

4.1 通用医疗人工智能的定义

医学生成式语言模型是近年来医学人工智能领域的一项重要发展。Moor M 等^[28]定义了通用医疗人工智能(generalist medical artificial intelligence,GMAI)模型,即能够使用少量或未标记数据执行多个任务的医疗人工智能模型。其通过在大规模、多样化的数据集上进行自我监督学习,能够灵活地解释不同医疗模式组合,并通过先进的医学推理能力呈现易于人类理解的结果。GMAI 模型具有高度灵活的可交互性、精准的医学语言表达专业性以及强大的领域泛化理解能力。医学生成式语言模型的代表^[27,29,31-38],见图 2。

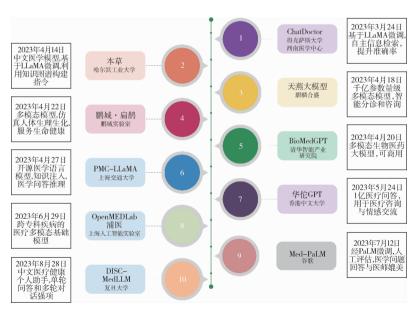


图 2 现有医学生成式语言模型脉络

4. 2 ChatDoctor

ChatDoctor^[29]是基于 LLaMA 微调的医学大语言模型,能适应医疗场景需求。其使用 10 万个在线医疗咨询网站的真实患者 - 医师对话数据进行微调,学习医学知识和医患对话语言风格。模

型采用独特的自主信息检索机制,能够访问和利用来自维基百科或疾病数据库等专业医疗数据库的信息,提升准确性。大量实验证实了该模型在准确率、召回率和 F1 值等方面的强大性能。研究结果表明, ChatDoctor 能够更好地理解患者需求并提供有根据的建议,具有广泛的医疗应用

前景。

4.3 Med - PaLM

Med - PaLM^[27]是基于谷歌公司 Pathways 语言模型(pathways language model, PaLM)基座的医学生成式语言模型。通过指令微调训练 Flan - PaLM模型,在 MedQA、MedMCQA、PubMedQA 和大规模多任务语言理解(massive multitask language understanding,MMLU)临床问题数据集上取得了最优结果。进一步通过与医学领域对齐的指令提示微调方法,建立 Med - PaLM模型以及基于事实性、理解性、推理能力、潜在伤害与偏见等多角度的人工评测框架。人工医师评估发现 Med - PaLM模型在回答医学问题方面具有与临床医师相当的效果。为进一步评估语言模型在医学问题回答方面的能力和局限性,还建立了新的医学问题回答基准测试数据集MultiMedQA。

4. 4 MIMIC - Diff - VQA

Hu X 等^[30]设计了一个多模态专家知识感知的多关系图网络模型,并构建了最大的医学视觉问答数据集 MIMIC - Diff - VQA。该数据集基于放射科胸片报告,构建具有逻辑递进和种类丰富的视觉问答对。同时还提出了图像对比视觉问答任务,即给定两张医学影像图片,回答关于这两张图片差异性的问题。这项研究进一步推动模型贴近临床实践,以提升医师诊疗效率。

4.5 华佗 GPT

ZHANG H 等^[31] 开发的华佗 GPT 旨在使语言模型具备类似于医师的中文医学诊断和医疗咨询建议能力。该模型结合 ChatGPT 生成的"蒸馏数据"和真实世界中医生回复患者问题的对话数据来构建微调语料,以保持与用户交流内容的知识丰富性。经过自动测评和医师评估,展现出较强性能。

5 挑战与风险

5.1 数据安全与隐私

医疗数据包含大量患者隐私信息,在生成式语·16·

言模型的训练和微调过程中,应严格遵循国内外相关法律法规,包括《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》,美国《健康保险携带和责任法案》(Health Insurance Portability and Accountability Act,HIPPA)及欧盟《通用数据保护条例》(General Data Protection Regulation,GDPR)等。为确保患者隐私和数据安全,必须采用加密、脱敏、匿名化和去标识化等技术手段,防范医疗数据泄露或滥用,并避免模型生成的输出暴露个人敏感信息。

5.2 专业性

生成式语言模型在输出医学文本时必须具备丰富的医学知识和经验,确保其内容符合医学规范和标准,包括正确使用医学术语、符号以及规范格式,以保证输出的准确性与可信度。模型应避免使用具有误导性和不专业的表达方式,并避免导致误诊、漏诊或不当治疗的情况。

5.3 安全性与伦理

生成式语言模型在输出内容时必须尊重和保护 人类的生命健康,不能提供有害、歧视性或违反医 学伦理和价值观的建议。模型内部应嵌入医学伦理 约束,并对输出进行风险评估,确保其安全性,不 产生对患者生命健康存在潜在影响的建议。

5.4 可解释性

为确保模型的可信度和透明性,生成的内容要 具有强烈的可解释性。模型应能够清晰地展示其推 理依据、循证支持和诊疗逻辑,及时检测并纠正可 能存在的偏见、错误或虚构等问题,确保其输出结 果既准确又可靠。

6 知识增强的医学语言模型优势及技术框架

6.1 基于知识图谱约束的医学语言模型

生成式语言模型在医学常识问答中具有广阔应 用前景。要应对上述挑战并确保生成文本的专业 性、准确性和可信性,针对医学数据的复杂性、稀 缺性和敏感性,降低偏见、事实性错误或虚构输出的风险,可利用领域知识图谱对生成模型的输出进行约束。迫切须要建立基于知识图谱约束的医学语言模型,知识图谱蕴含丰富的领域结构化知识,其中基于实体、关系和属性的图结构具有高度清晰的医学行业诊疗逻辑。通过将这些领域知识图谱的约束应用于语言模型的生成过程,可以输出专业、准确和可信赖的医学文本。

6.2 知识图谱增强的优势

6.2.1 医学生成式语言模型面临多重挑战 如数

据安全隐私保护、知识专业性不足和伦理规范性欠缺等问题,迫切需要有效的解决方案以确保模型在医疗领域的安全合规应用。结构化的医学领域知识图谱应用于生成式语言模型,可以有效提供外部约束和指导,帮助模型生成专业准确、符合规范的医学文本。知识图谱是一种将实体概念及其语义关联关系以网络化、结构化形式表示的知识框架。相较于文本语料,知识图谱的结构化表示形式更易被模型理解和应用。具体来说,知识图谱可以为生成式语言模型提供专业理解力等5方面支持,见图3。

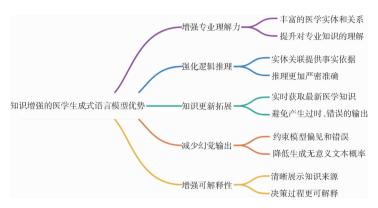


图 3 知识图谱增强的医学生成式语言模型优势框架

- 6.2.2 增强专业理解力 依托知识图谱中丰富的 医学实体及其关联关系,可以大幅提升生成式语言 模型对专业知识的理解能力。充足的专业知识引导 是避免模型产生幻觉、输出不准确或不规范医学文 本的重要前提。
- 6.2.3 强化逻辑推理 生成式语言模型可以进行 更准确的逻辑推理。实体之间的关联为模型提供可 靠的事实依据和严密的推理链,有助于生成条理清 晰、事实准确的医学文本。
- 6.2.4 知识更新拓展 借助持续扩展的知识图谱,模型可以实时获取最新医学知识,避免产生过时、错误的输出。
- 6.2.5 减少幻觉输出 知识图谱的约束作用可以减少模型偏见、事实性错误及幻觉输出,降低生成无意义或有害医疗文本的可能性。
- 6.2.6 增强可解释性 知识图谱可以清晰展示每 条生成语句所基于的知识来源,促进诊疗决策过程 有据可循。

6.3 知识增强的医学语言模型技术方法

6.3.1 知识结构引导的医学语言模型预训练 为 提升医学语言模型对医学知识的了解和认知,通常 将医学知识结构融入大语言模型的预训练阶段。通 用大语言模型通过对现有医学知识语料、知识库进 行大量的基于自监督学习的无监督训练,或通过相 似性任务(判断模型生成的医学回答是否与标准答 案相似)、选择任务(在给定的问题中选择最优答 案)等^[28]监督学习训练,将医学文本内容输入语 言模型,不断对语言模型输出的医学文本进行监督 和优化,提升语言模型对医学领域的基本了解和认 知,形成初步医学语言模型。

6.3.2 医学知识增强的语言模型微调训练 现有 医学知识库、知识图谱等,可以自动生成用于微调 的指令数据,从而将知识融入训练过程。基于知识 图谱生成大量高质量指令模板,医学语言模型可以 将大部分指令分类并快速学习^[39],在面对真实用户 的指令要求时,语言模型可以简单匹配用户指令与 模板指令,理解任务要求,从而完成对应任务。基 于知识图谱生成的医学指令模板可以高效提升语言 模型面对医学任务指令的响应能力。

6.3.3 医学外挂知识库检索增强推理 将医学知识库、知识图谱作为医学模型推理时的外挂知识库,类似使用检索增强的方式协助语言模型获得更多领域知识^[40]。通过对用户问题中的关键信息在相关医学知识库中进行自动检索,辅助语言模型理解和推理用户意图、生成更具有解释性的内容。

6.3.4 基于医学知识引导的提示学习 基于医学知识图谱生成合适的提示学习^[41],可以提升医学模型的推理能力,特别是在复杂场景和多步交互的情况下。现有工作^[42-43]表明情景学习对其中包含的例子顺序、质量等非常敏感,一个非常小的改变可能会引起模型效果的较大变化。因此,通过从医学知识图谱中抽取相关医学专业术语等知识,自动引导产生有利于医学模型挖掘相关的医学信息,提升在交互问答时的推理性能。

6.3.5 知识协同的智能中控路由 单独一个医疗语言模型往往难以独立完成复杂的任务,通常须要调用外部的应用程序接口、组合利用小模型、使用第三方工具等,这使语言模型类似于一个自适应的智能中控路由。医学语言模型在理解用户任务后进行任务规划,自适应地依据任务规划结果调用相关工具产生更专业的结果。

7 知识增强的医学语言模型应用案例

7.1 基于通用语言模型的医学知识图谱构建

7.1.1 通用语言模型应用价值 传统基于规则的 医学知识图谱构建方法需要花费大量时间和精力进行 模板设计;而依赖预训练语言模型的方法,则需要大 量医学文本的人工标注和模型调参。这两类方法在专 业医学领域泛化理解能力方面都存在局限,构建过程 耗时费力。通用语言模型的出现彻底改变了这一状况,通过大量多样化数据预训练,内化丰富的医学领域知识,无需额外的医学领域微调就可以直接应用 到下游任务。独特的语义理解能力使其可以准确解 析医学文本语义,准确识别实体并抽取实体关系,从而实现对医学知识图谱的高效自动化构建。具体来说,通用语言模型可以快速从海量非结构化医学文献中,准确抽取实体(如疾病、药物等)以及实体之间的关系(如"症状-疾病""药物-疾病"等),并持续更新知识图谱,确保知识图谱的及时性。构建于通用语言模型之上的医学知识图谱质量更高、覆盖面更广,为依托知识图谱进行医疗决策支持、信息检索等提供了坚实的知识基础。

7.1.2 实体与关系抽取 在处理医学文本时,语 言模型可以识别文本中的特定领域词汇, 并对疾 病、药物、患者、医生等实体进行分类。同时,还 可以根据参数化的领域知识,通过语义理解能力构 建这些实体的属性和关系[4]。语言模型由此可以实 现命名实体识别、实体分类、实体链接、关系抽取 和指代消解等多项任务。试将以下段落基于 GPT-4 进行命名实体识别、实体分类和关系抽取 分析:"李某某,60岁女性,在北京协和医院被某 医师诊断出患有糖尿病, 医师嘱咐用胰岛素治疗"。 在该过程中大语言模型成功完成以下任务。(1)命 名实体识别。识别出"李某某""糖尿病""某医 师""北京协和医院""胰岛素"等实体。(2)实 体分类。"李某某"被分类为患者(Patient 类), "糖尿病"分类为疾病(Disease 类),"某医师"分 类为医生(Doctor类),"北京协和医院"分类为机 构 (Organization 类), "胰岛素"分类为药物 (Drug 类)。(3) 实体链接。将新识别的实体与已 有的知识图谱进行关联,避免冗余或歧义。(4)关 系抽取。识别出"诊断" (< diagnosed with >)、 "医嘱"(rescribed>)等关系。将这些信息以 表格形式表达,见表1。

表 1 基于语言模型的实体识别与关系抽取示例

实体	类型	关联关系
李某某	Patient	被诊断与治疗对象
糖尿病	Disease	李某某被诊断的疾病
某医师	Doctor	诊断并治疗李某某的对象
北京协和医院	Organization	李某某的诊断和治疗机构
胰岛素	Drug	治疗李某某的药物

7.1.3 知识补全 通过模型参数化的医学知识,结合语言模型的逻辑推理能力,可以实现知识图谱的补全。假设已有知识图谱记录了胰岛素作为治疗糖尿病的药物,但对于其不良反应、副作用或替代药物等方面描述不完整。在这种情况下,大语言模型可以提供专业信息来填补这些缺失,见表 2。

表 2 基于大语言模型的药物不良反应知识补全示例

实体	类型	药物不良反应描述补全
胰岛素	药物	用于治疗糖尿病,可能导致低血糖,需配合合
		理饮食、适当运动,并定期监测血糖水平
二甲双胍	药物	用于治疗糖尿病,可能的副作用包括消化系统
		不适、疲劳、头痛、口腔有金属味、出现皮疹

7.2 知识增强的医学常识问答

医学常识问答是一项典型任务,基于语言模型 回答与医学相关的常识性问题,对促进患者健康教 育和基层医师培训等方面具有一定价值。然而由于 医学领域的专业性与复杂性,生成式语言模型往往 可能会生成片面、事实性错误甚至幻觉的回答。为 解决准确性问题, 医学常识问答往往通过医学知识 图谱进行提示学习,引导语言模型输出,使其生成 专业的医学文本[45-46]。具体而言可以将数据源中 的实体、关系和属性作为输入条件, 然后使用 GPT 模型生成相应文本描述或问答。例如,给定实体 "阿司匹林"和其属性"药理作用",模型可能生成 的文本描述为: "阿司匹林是一种常用的非甾体抗 炎药,具有抗炎、镇痛、退热和抗血小板聚集等药 理作用"。又如给定实体"阿司匹林"和问题"阿 司匹林有哪些不良反应",模型可能根据知识图谱 中相关实体和关系生成如下回答: "阿司匹林的不 良反应主要包括胃肠道出血、过敏反应、雷耶氏综 合征、耳鸣、头痛等"。通过这种方式不仅可以提 高医患沟通和教育的便利性和质量,还可以提升医 学研究和写作的效率和质量。为了展示这种方法的 效果,使用大规模中文开放知识图谱(OpenKG)[47] 中抽取的一些实体、关系和属性, 并利用 GPT - 4 的语言表征能力生成相应的文本描述或问答,见 表3。

表 3 基于知识图谱约束的医学常识问答示例

输入	GPT -4 输出文本
实体:糖尿病;属性:定义	糖尿病是一种以高血糖为特征的代谢性疾病,主要由胰岛素分泌不足或作用受损引起
实体:糖尿病;属性:症状	糖尿病的常见症状包括多饮、多尿、多食、体重下降、乏力、口干、皮肤感染等
实体:糖尿病;问题:糖尿病有哪些并发症?	糖尿病的常见并发症有心血管疾病、神经病变、肾脏损害、眼部损害、足部溃疡等

7.3 基于知识图谱增强的医学跨模态语言模型

7.3.1 医学视觉问答 是将自然语言处理和计算机视觉相结合的复杂任务,旨在对医学图像进行深度分析,并生成自然语言形式的问答。这一任务在医学领域具有重要的潜力和价值,但同时也面临挑战,包括医学图像的多样性、复杂性和噪声性,以及医学知识的动态性和不确定性。目前已有的视觉语言模型实践主要基于微调现有开源模型^[48-49]。例如 MedSAM^[50]项目利用深度学习技术构建医疗影像模型,从医学图像中提取有用信息。PMC - CLIP^[51]项目则是一个医疗多模态模型,基于科学文

献构建了高质量的医疗多模态数据集,称为 PMC - OA 数据集,并在医疗图文检索、图像分类及视觉问答等任务中取得最佳效果。此外,微软基于GPT - 4 构建的医疗病历生成多模态模型可以将床边对话的语音识别与语言模型的逻辑能力相结合,生成临床病历草稿,并经过医师审核,最终形成严谨的临床病历文书。

7.3.2 局限 上述医学跨模态语言模型在提供基础疾病诊断时存在一定局限性。例如无法准确理解和回答具有高度专业性和复杂性的医学问题,在处理医学罕见病理或特定类型影像时遇到困难。此外在面对医学知识的动态性和不确定性时,模型可能

无法提供最新、最准确的医学知识,从而给出过时 的医疗诊断和决策回复。

7.3.3 基于知识图谱和多模态数据融合技术的医学语言模型 此类模型可以有效解决上述问题。例如在解读一张 X 光片显示的"左肺上叶有一块圆形致密影"时,模型不仅从知识图谱中提取肺结核、肺癌、肺脓肿和肺水肿等可能的病因,还能结合患者的具体情况作出更全面的判断。

考虑以下案例:一位名为赵某的 45 岁女性患者。患有 II 型糖尿病(已持续 10 年)和原发性高血压 2 级(已持续 5 年),其父亲有冠心病史,母亲则患有 II 型糖尿病和高血压。近期为了缓解腰痛,自行服用了非甾体类抗炎药萘普生钠,并于服药 5 日后出现气喘和伴随的胸痛症状。考虑到萘普生钠在少数情况下可能引发肺水肿,结合对患者的病史事件图谱分析,模型可以进一步识别其药物使用记录中的潜在风险,并根据诊断建议待完善的检查和治疗,从而为医生、研究人员和患者提供更加准确、专业和可靠的医学视觉问答服务。

可见模型能有效解决多种内科慢性疾病和药物 干预带来的复杂诊断问题,融合最新医疗领域知识 图谱作为可靠知识来源,利用语言模型强大的逻辑 推理能力和丰富的领域知识图谱,综合处理医学图 像、语音和医患对话文本等多模态信息。协助医疗 专家快速获取初步诊断、治疗建议和决策支持,进 而提高医疗服务质量和效率。

8 结语

本文全面论述和分析了生成式语言模型在医学领域的应用前景及挑战。医学文本生成的复杂性和专业性对自然语言处理带来巨大考验,借助预训练模型能够生成一定程度具备专业性和准确性的医学文本。即使 GPT - 4 等模型在医学文本生成方面已经取得较良好表现,其性能的优化和提升还依赖更深层次医学知识的辅助增强,以确保答案的精确性和可靠性,防止对医疗决策和患者生命健康产生负面影响。知识图谱增强的语言模型提供了一种理想的解决方案,可以从无标注的大规模文本中挖掘和

学习丰富的医学知识,从而加深对医学文本知识的 理解和应用。然而,如何设计更加精确和高效的知识图谱构建方法,以更好地满足医学领域的应用需求,仍然面临挑战。此外,利用知识图谱可以进一步提高模型在处理医学视觉问答任务时的性能,实现更深层次的视觉理解和解释。但如何有效地整合知识图谱和视觉信息,以提高模型的解释能力和准确性,仍需更深入的研究。

生成式语言模型在医学领域具有巨大应用潜力,但在模型泛化能力、准确性、解释性等方面仍有待提升。未来可以借助因果机器学习增强语言模型的逻辑推理和医学证据支持,进而增强医学问答的完整性和可信度。

参考文献

- 1 LUND B D, WANG T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries [J]. Library hi tech news, 2023, 40 (3): 26-29.
- 2 JIAO W, WANG W, HUANG J, et al. Is ChatGPT a good translator? A preliminary study [EB/OL]. [2023 09 18]. https://ar5iv.org/abs/2301.08745.
- 3 FUCHS K. Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse [J]. Frontiers in education, 2023, 8 (1166682): 1-4.
- 4 DERGAA I, CHAMARI K, ZMIJEWSKI P, et al. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing [J]. Biology of sport, 2023, 40 (2): 615-622.
- 5 GUO B, DING Y. How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection [EB/OL]. [2023 09 18]. https://ar5iv.org/abs/2301.07597.
- 6 DALE R. Navigating the text generation revolution: traditional data to text NLG companies and the rise of ChatG-PT [J]. Natural language engineering, 2023, 29 (4): 1188 1197.
- 7 RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre training [EB/OL]. [2023 09 18]. https://s3 us west 2. amazonaws. com/openai assets/research covers/language unsupervised/language understanding paper. pdf.

- 8 RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners [J]. OpenAI blog, 2019, 1 (8): 9.
- 9 BROWN T, MANN B, RYDER N, et al. Language models are few - shot learners [J]. Advances in neural information processing systems, 2020 (33): 1877 - 1901.
- 10 THIRUNAVUKARASU A J, TING D S J, ELANGOVAN K, et al. Large language models in medicine [J]. Nature medicine, 2023 (29): 1-11.
- 11 HARRIS E. Large language models answer medical questions accurately, but can't match clinicians' knowledge [J]. Journal of the American medical association, 2023, 330 (9): 792-794.
- 12 SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search [J]. Nature, 2016, 529 (7587): 484-489.
- 13 SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of Go without human knowledge [J]. Nature, 2017, 550 (7676): 354-359.
- 14 SILVER D, HUBERT T, SCHRITTWIESER J, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self play [J]. Science, 2018, 362 (6419): 1140 1144.
- 15 DING M, ZHANG J, LI C, et al. CogView: mastering text to - image generation via transformers [J]. Advances in neural information processing systems, 2021 (34): 19822 - 19835.
- 16 RAMESH A, DHARIWAL P, NICHOL A, et al. Hierarchical text conditional image generation with CLIP latents [EB/OL]. [2023 09 18]. https://ar5iv.org/abs/2204.06125.
- 17 XUE Z, SONG G, GUO Q, et al. Raphael; text to image generation via large mixture of diffusion paths [EB/OL].
 [2023 09 18]. https://ar5iv.org/abs/2305.18295.
- 18 HUANG B, XU Y, HU X, et al. A backbone centred energy function of neural networks for protein design [J]. Nature, 2022, 602 (7897): 523 528.
- 19 CAO L, COVENTRY B, GORESHNIK I, et al. Design of protein – binding proteins from the target structure alone [J]. Nature, 2022, 605 (7910): 551 – 560.
- 20 DAS P, SERCU T, WADHAWAN K, et al. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations [J]. Nature biomedical engineering, 2021, 5 (6): 613-623.

- 21 TALLORIN L, WANG J, KIM W E, et al. Discovering de novo peptide substrates for enzymes using machine learning [J]. Nature communications, 2018, 9 (1): 5253.
- 22 JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021, 596 (7873): 583 – 589.
- 23 KALE M, RASTOGI A. Text to text pre training for data to text tasks [EB/OL]. [2023 09 18]. https://ar5iv.org/abs/2005.10433.
- 24 NORI H, KING N, MCKINNEY S M, et al. Capabilities of GPT 4 on medical challenge problems [EB/OL]. [2023 09 18]. https://ar5iv.org/abs/2303.13375.
- VENIGALLA A, FRANKLE J, CARBIN M. BioMedLM: a domain specific large language model for biomedical text [EB/OL]. [2022 12 15]. https://www.mosaicml.com/blog/introducing pubmed gpt.
- 26 TAYLOR R, KARDAS M, CUCURULL G, et al. Galactica: a large language model for science [EB/OL]. [2023 09 18]. https://ar5iv.org/abs/2211.09085.
- 27 SINGHAL K, AZIZI S, TU T, et al. Large language models encode clinical knowledge [J]. Nature, 2023 (620); 1-9.
- 28 MOOR M, BANERJEE O, SHAKERI Z, et al. Foundation models for generalist medical artificial intelligence [J]. Nature, 2023, 616 (7956); 259 – 265.
- 29 LI Y, LI Z, ZHANG K, et al. ChatDoctor: a medical chat model fine tuned on a large language model Meta AI (LLaMA) using medical domain knowledge [J]. Cureus, 2023, 15 (6): 2-5.
- 30 HU X, GU L, AN Q, et al. Expert knowledge aware image difference graph representation learning for difference aware medical visual question answering [C]. Long Beach: The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'23), 2023.
- ZHANG H, CHEN J, JIANG F, et al. HuatuoGPT, towards taming language model to be a doctor [EB/OL]. [2023 09 15]. https://github.com/SCIR HI/Huatuo Llama Med Chinese.
- 32 WANG H, LIU C, XI N, et al. HuaTuo: tuning LLaMA model with Chinese medical knowledge [EB/OL]. [2023 -09-15]. https://ar5iv.org/abs/2304.06975.
- 33 LUO Y, ZHANG J, FAN S, et al. BioMedGPT: open multimodal generative pre trained transformer for Biomedicine [EB/OL]. [2023 09 15]. https://github.com/taokz/

BiomedGPT.

- 34 BAO Z, CHEN W, XIAO S, et al. DISC MedLLM: bridging general large language models and real - world medical consultation [EB/OL]. [2023 - 09 - 15]. https:// github.com/FudanDISC/DISC - MedLLM.
- OpenMEDLab. Promoting large scale pre trained foundation models and adaptation in healthcare [EB/OL]. [2023 09 15]. https://github.com/OpenMEDLab.
- 36 WU C, ZHANG X, ZHANG Y, et al. PMC LLaMA: further finetuning llama on medical papers [EB/OL]. [2023 09 15]. https://github.com/chaoyi wu/PMC LLaMA.
- 37 CHEN Y, WANG Z, XING X, et al. BianQue 1.0: improving the 'Question' ability of medical chat model through finetuning with hybrid instructions and multi turn doctor QA datasets [EB/OL]. [2023 09 15]. https://github.com/scutcyr/BianQue.
- 38 阿帕斯大模型. 中国首个自研多模态大模型, 致力于用人工智能改善生活 [EB/OL]. [2023 09 15]. https://ailme.apusai.com.
- 39 RENZE L, ZHANG K, YIN W. Is prompt all you need? No. A comprehensive and broader view of instruction learning [EB/OL]. [2023 - 09 - 18]. https://arxiv.org/abs/2303.10475.
- 40 HE H, ZHANG H, ROTH D. Rethinking with retrieval: faithful large language model inference [EB/OL]. [2023 09 18]. https://arxiv.org/abs/2301.00303.
- 41 YAO F, PENG H, SABHARWAL A, et al. Complexity based prompting for multi step reasoning [EB/OL]. [2023 09 18]. https://arxiv.org/abs/2210.00720.
- 42 DENNY Z, SCHÄRLI N, HOU L, et al. Least to most prompting enables complex reasoning in large language models [EB/OL]. [2023 09 18]. https://arxiv.org/abs/

2205. 10625.

- 43 TUSHAR K, TRIVEDI H, FINLAYSON M, et al. Decomposed prompting: a modular approach for solving complex tasks [EB/OL]. [2023 09 18]. https://arxiv.org/abs/2210.02406.
- 44 ZHANG C, GAO S, WANG H, et al. Position aware joint entity and relation extraction with attention mechanism [C]. Vienna: The International Joint Conference on Artificial Intelligence (IJCAI), 2022.
- 45 OpenKG. Open Knowledge Graph Platform [EB/OL]. [2023 09 18]. http://www.openkg.cn/.
- 46 LI Z, WANG H, ZHANG W. Translational relation embeddings for multi hop knowledge base question answering [J]. Journal of web semantics, 2022, 74 (1): 100723.
- 47 CHEN H, HU N, QI G, et al. OpenKG Chain: a Block-chain infrastructure for open knowledge graphs [J]. Data intelligence, 2021, 3 (2): 205-227.
- 48 LIANG Y, WANG H, ZHANG W. A knowledge guided method for disease prediction based on attention mechanism [C]. Beijing: International Conference on Web Information Systems and Applications, 2022.
- 49 ZHANG A, TAO W, LI Z, et al. Type aware medical visual question answering [C]. Singapore: ICASSP 2022 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.
- 50 MA J, HE Y, LI F, et al. Segment anything in medical images [EB/OL]. [2023 09 18]. https://arxiv.org/abs/2304.12306.
- 51 LIN W, ZHAO Z, ZHANG X, et al. Pmc clip: contrastive language image pre training using biomedical documents [C]. Vancouver: The International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI), 2023.

欢迎订阅 欢迎赐稿