四川大学学报（医学版） 2023，54（5）：855－856
J Sichuan Univ（Med Sci）

在线全文

• 大数据与人工智能技术在生物医学多场景的应用 •

‖ 院士笔谈 ‖

医疗大数据结合大语言模型的应用展望

陈润生[1,2]

1. 四川大学华西医院 生物医学大数据中心（成都 610041）；
2. 中国科学院生物物理研究所 健康大数据研究中心 核酸生物学重点实验室（北京 100101）

【摘要】 在医学领域，大数据技术结合大语言模型的应用预计将具有巨大的影响。本文将从几个方面展望医疗大数据结合大语言模型的应用：首先是辅助医生诊断和鉴别诊断方面，其次是在循证医学领域，此外医疗大数据结合大语言模型也可以应用于辅助医生进行临床和医学研究方面。通过将医疗大数据与人工智能大语言模型相结合，可以实现更加精准、高效、智能化的医疗诊断和治疗，并将为人类的健康领域做出更大的贡献。

【关键词】 大数据技术 大语言模型 应用

Prospects for the Application of Healthcare Big Data Combined With Large Language Models

CHEN Runsheng[1,2].
1. *West China Biomedical Big Data Center*, *West China Hospital*, *Sichuan University*, Chengdu 610041, China; 2. *Key Laboratory of RNA Biology*, *Center for Big Data Research in Health*, *Institute of Biophysics*, *Chinese Academy of Sciences*, Beijing 100101, China

【**Abstract**】 The application of big data technology combined with large language models is expected to make an enormous impact in the field of medicine. Herein, the prospects for the application of healthcare big data combined with large language models were discussed in several aspects, including first in assisting doctors in making diagnosis and differential diagnosis and, then, in the field of evidence-based medicine. In addition, healthcare big data combined with large language models could also be applied in assisting doctors to conduct clinical and medical research. Through combining healthcare big data with large language models, medical diagnosis and treatment with improved precision, efficiency, and intelligence will be realized and greater contributions will be made to the field of human health.

【**Key words**】 Big data technology Large language model Applicaitons

在医学领域，大数据技术与大语言模型的集成应用预计将具有巨大的潜力与价值。医疗大数据是指医疗诊疗及科学研究过程中产生的大量数据，包括病历、检查报告、生理参数、影像、组学等多种形式的数据。这些数据蕴含着丰富的信息和知识，但是由于数据量太大、结构复杂、质量参差不齐等问题，很难直接被医生和研究人员利用。而人工智能大语言模型则可以通过自然语言处理技术，将这些数据转化为可读、可理解、可分析的形式，从而为医疗诊疗和医学研究提供更加高效、准确和智能化的支持。人工智能大语言模型通常是指利用大量文本数据做预训练得到的深度生成模型，它们可以理解和生成自然语言，实现多种语言任务，如对话、翻译、文本生成等。大语言模型通过对大规模语料库进行自监督的预训练，学习文本的语法、语义、逻辑和风格等特征，然后通过对特定任务的数据进行微调，适应不同的下游应用。其优势在于可以利用大量的无标注数据，提高模型的泛化能力和表达能力，缓解数据稀缺和标注成本高的问题，同时也可以实现跨领域和跨语言的迁移学习，提高模型的可复用性和可扩展性。把医疗大数据和人工智能大语言模型结合起来，可以为临床诊疗和医学研究带来巨大的变革和进步。

人工智能大语言模型通常采用Transformer架构。Transformer架构于2017年被谷歌团队提出[1]，核心是由编码器和解码器两部分组成，编码器将输入的文本序列编码成一个高维的向量表示，解码器根据编码器生成的高维向量生成输出的文本序列。编码器和解码器都由多个相同的层堆叠而成，每个层都包含一个多头自注意力子层和一个前馈神经网络子层，以及残差连接和层归一化操作。当前，人工智能大语言模型的代表作有令人瞩目的ChatGPT、Llama、ChatGLM、文心一言、通义千问等，它们在各种自然语言处理的基准测试中取得了惊人的成绩，也在各个领域产生了广泛的影响，例如搜索、医疗、教育、影视等。

人工智能大语言模型在临床诊疗和医学研究的应用是一个前沿且有前景的领域，它们可以帮助医生和患者提高诊断、治疗和预防的效率和质量，也可以帮助医学研究人员发现新的知识和方法。但目前的大语言模型在专业领域尚不成熟，突出的一点就是给出的答案往往深度不够，而且会出现答非所问、错误生成参考文献等问题，所以未来可以运用提示词工程、微调等技术，对通用的大语言模型进行改造，以适应在医学专业领域运用。谷歌医疗团队最近在*Nature*发表了最新版本的医疗大模型Med-PalM

v2.0[2], 这个工作中提出了全新的基准测试——MultiMedQA, 涵盖了医学考试、医学研究等领域的问题, 在这个基准测试中Med-PalM v2.0达到了所有模型的业内最好成绩; 同时在USMLE美国执业医师考试类似的问题上, Med-PalM v2.0达到了86.5%的准确率, 已经比肩经过系统训练的医学毕业生水平; 科学共识和安全性方面, 临床医生给出的答案与Med-PaLM的一致性为92.9%。这意味着在医学领域, Med-PaLM能够提供与临床医生相似的答案, 具有较高的科学共识和安全性。从中不难看出, 大语言模型在处理医学数据时展现出了显著优势, 作为一种先进的人工智能模型, 其具备自我学习和理解的能力, 可以快速消化并吸收海量的医学信息。医学语言模型不仅能理解并生成医学专业术语, 还能通过分析历史病例和患者信息, 辅助医生进行疾病诊断和治疗决策。结合医疗领域的具体需求, 人工智能大语言模型可以在以下方面发挥重要作用:

首先在辅助医生诊断和鉴别诊断方面, 可以利用人工智能大语言模型对患者的病历、检查报告、生理参数等进行自然语言处理和分析, 帮助医生快速准确地诊断疾病、制定治疗方案。例如, 可以通过对患者的病历进行自动摘要和分类, 提取关键信息和特征, 辅助医生做出正确的诊断和治疗决策。还可以通过对医学影像进行自动分析和识别, 帮助医生发现疾病的早期征兆和异常信号, 提高诊断准确率和效率。此外, 在药物治疗方面, 可以利用人工智能大语言模型对药物剂量、作用机制、副作用等进行分析和预测, 为医生制定个性化的治疗方案提供参考。在一些医生数量和医疗资源稀缺的地区, 通过人工智能大模型进行网络诊断的方法能够很好地缓解医疗的供需矛盾。

其次, 循证医学是一种基于证据的医学实践方法, 它通过收集、评估和综合现有的临床研究证据, 来指导医生和患者做出最佳的医疗决策。大语言模型可以通过文本挖掘技术, 自动从大量的临床研究文献中提取有用的信息和知识。传统的文献综述需要耗费大量的时间和精力来收集、筛选和整合文献, 而且可能会存在遗漏或者误解等问题。而大语言模型可以通过自然语言处理技术, 自动从海量的文献中提取出与特定疾病或治疗方案相关的信息和知识, 从而为循证医学提供更加全面、准确的支持。循证医学需要依靠大量的临床数据来支持决策, 但是这些数据往往存在着复杂的关联和规律。而大语言模型可以通过数据挖掘技术, 自动发现这些潜在的关联和规律, 从而为循证医学提供更加深入、准确的支持。

此外, 在辅助医生进行临床和医学研究方面, 可以利用人工智能大语言模型对医学文献、期刊论文等进行自然语言处理和分析, 从海量的医学数据、组学数据、药物数据中挖掘出潜在的规律、关联、趋势和见解, 帮助研究人员发现新的知识和规律。例如, 在癌症领域, 可以通过对肿瘤基因组数据进行分析和挖掘, 发现新的癌症标志物和治疗靶点; 从临床试验数据中探索新的药物对某种疾病的效果, 从电子健康记录数据中发现新的风险因素或预后指标; 在药物开发领域, 可以利用人工智能大语言模型对药物分子结构进行分析和预测, 加速新药开发的进程。在公共卫生方面, 可以利用人工智能大语言模型对流行病学数据进行分析和预测, 帮助政府和医疗机构制定应对措施。

综上所述, 人工智能大语言模型在医疗服务和医学研究的应用具有巨大的潜力和价值, 它们可以缓解医疗资源紧张的问题, 提高医疗质量和效率, 降低医疗成本, 促进医学创新和进步。然而, 将医疗大数据与人工智能大语言模型结合起来也存在一些挑战和难点。首先是数据安全和隐私保护问题。医疗数据属于敏感信息, 需要采取安全可靠的措施来保护数据的安全性和隐私性。其次是数据质量和标注问题。医疗数据往往存在噪声和缺失等问题, 需要采取有效的方法来清洗和预处理数据。同时, 由于医疗领域专业性强、术语复杂, 需要在人工智能大语言模型中引入专业知识和领域规则来提高模型的准确性和可解释性。大语言模型和大数据技术的结合也将为医疗行业带来新的商业机会。例如, 基于模型的智能医疗推荐系统, 可以为患者提供个性化的医疗服务; 而医疗大数据平台则可以为医疗机构和制药企业提供市场信息, 帮助他们做出更明智的商业决策。未来, 随着医疗大数据的不断积累和人工智能大语言模型技术的不断进步, 医疗领域将迎来更加广阔的发展空间和机遇。我们可以期待, 在不久的将来, 通过将医疗大数据与人工智能大语言模型相结合, 实现更加精准、高效、智能化的医疗诊断和治疗, 并将为医学研究、公共卫生政策制定、医疗服务提供和商业模式创新带来巨大的机遇, 为人类的健康事业做出更大的贡献。

参 考 文 献

[1] VASWANI A, SHAZEER N, PARMAR N, *et al*. Attention is all you need. arXiv, 2017. doi: 10.48550/arXiv.1706.03762.

[2] SINGHAL K, AZIZI S, TU T, *et al*. Large language models encode clinical knowledge. Nature, 2023, 620: 172–180. doi: 10.1038/s41586-023-06291-2.

(2023 – 09 – 07收稿, 2023 – 09 – 13修回)

编辑 姜 恬