



## (12) 发明专利申请

(10) 申请公布号 CN 117851558 A

(43) 申请公布日 2024. 04. 09

(21) 申请号 202311698917.9

(22) 申请日 2023.12.08

(71) 申请人 成都知识视觉科技有限公司

地址 610000 四川省成都市金牛高新技术  
产业园区信息园东路99号212号

(72) 发明人 王一哲 向飞 焦龙

(74) 专利代理机构 成都顶峰专利事务所(普通  
合伙) 51224

专利代理师 陈绍富

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

G16H 70/00 (2018.01)

G06F 21/62 (2013.01)

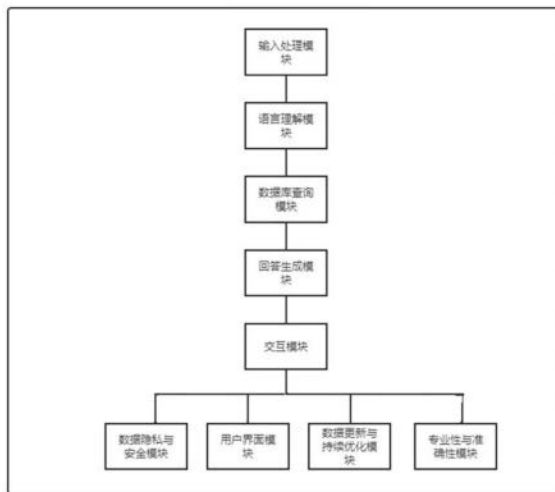
权利要求书3页 说明书8页 附图2页

## (54) 发明名称

一种基于大语言模型的医学问答系统

## (57) 摘要

一种基于大语言模型的医学问答系统,其特征在于,包括:输入处理模块,用于负责接收用户问题,并进行预处理;语言理解模块,用于理解用户输入的问题,识别问题的意图;数据库查询模块,用于连接医学知识库和医学文献数据库,查询相关的医学信息和答案;回答生成模块,用于生成系统的回答,根据查询到的医学信息,生成准确、专业的回答;交互模块,用于负责处理用户与系统的对话式交互,处理用户追问、澄清等情况;数据隐私与安全模块,用于保护患者信息安全;用户界面模块,用于构建用户友好的界面;数据更新与持续优化模块,用于定期更新数据库;专业性与准确性模块,用于系统提供专业知识的回答。本发明提供便捷的医学咨询与快速响应和帮助。



1. 一种基于大语言模型的医学问答系统,其特征在于,包括:

输入处理模块,用于负责接收用户输入的医学问题,并进行预处理;

语言理解模块,用于使用大语言模型来理解用户输入的问题,识别问题的主题、意图,并从中提取关键信息;

数据库查询模块,用于连接医学知识库和医学文献数据库,根据用户的问题,查询相关的医学信息和答案;

回答生成模块,用于使用大语言模型来生成系统的回答,根据查询到的医学信息,以及系统的训练和微调,生成准确、专业的回答;

交互模块,用于负责处理用户与系统的对话式交互,处理用户追问、澄清并保持对话的连贯;

数据隐私与安全模块,用于确保患者的个人健康信息得到充分的保护,遵守相关的隐私和数据安全法规;

用户界面模块,用于构建用户友好的界面,供患者输入问题并查看系统生成的回答;

数据更新与持续优化模块,用于负责定期更新医学知识库和医学文献数据库,以确保系统提供最新的和准确的信息,同时持续优化大语言模型,以提高系统的回答准确性;

专业性与准确性模块,用于确保系统提供基于医学专业知识的回答,对系统的回答进行评估和校对,避免出现误导性的信息。

2. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述负责接收用户输入的医学问题,并进行预处理,包括:

去除特殊字符、标点符号、多余空格、甚至拼写检查和纠正;对于分词,将输入的文本分割成单词或短语,分词考虑医学术语和缩写的特殊处理,以确保术语被正确理解;移除停用词,去除在文本中出现频率很高但没有实际信息量的词语,以减少噪音并提高系统的效率;考虑上下文信息,尤其是在对话式交互中,系统需要理解先前的问题和回答,以确保连贯性和更准确的回应;将文本转换为数学表示如词嵌入,有助于模型理解词语之间的语义关联,从而更好地回答问题。

3. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述使用大语言模型来理解用户输入的问题,识别问题的主题、意图,并从中提取关键信息,包括:

对大语言模型进行微调,使其更好地理解 and 回答医学相关问题,通过引入医学领域的文本数据进行模型训练来实现;将输入的问题归类到医学领域的特定主题,通过配置额外的分类器,识别问题所涉及的领域如药物、症状、疾病分类;确定用户提出问题的意图,如用户是在寻求解释、寻求建议,还是需要特定的信息,有助于更好地定位问题,并为用户提供更准确的答案;从问题中提取关键信息,包括实体如药物、症状、疾病、时间、事件,需要结合实体识别和上下文理解,以确保关键信息被正确地识别和提取;考虑对话中的上下文信息,确保问题的解释和答案与先前对话一致,从而提供连贯的对话体验;在需要时,能解释模型对问题的理解方式,有助于增强系统的可信度和用户对系统的信任感。

4. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述连接医学知识库和医学文献数据库,根据用户的问题,查询相关的医学信息和答案,包括:

整合多个医学知识库和文献数据库,数据库包括医学书籍、期刊、研究论文、临床指南以及专业组织的文档;确保不同数据库的数据格式和结构一致,以便系统能够统一查询和

提取信息,涉及到数据清洗、标准化和统一化处理;根据用户提出的问题,构建相应的查询以搜索相关信息,查询需要考虑用户问题的语义、实体识别的结果以及上下文信息,以提高查询的精准度;从数据库中提取相关的医学信息和答案,包括抽取文本、图表、统计数据,甚至是以结构化形式呈现的信息如数据库中的特定字段;将从数据库中检索到的信息与用户问题进行匹配,并对匹配度进行排序,涵盖文本相似度匹配、语义匹配方法,以确保返回最相关和准确的答案;确保数据库中的信息及时更新,以提供最新的医学知识,需要定期的数据同步和更新机制,以反映医学领域的最新发展;利用知识图谱构建数据库间关系,以帮助系统更好地理解医学信息的关联性和结构,从而提供更全面的答案;考虑用户问题的语境,以便更精准地过滤和提取相关信息,包括对于多义词、复杂问题或需要更详细上下文理解的情况。

5. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述使用大语言模型来生成系统的回答,根据查询到的医学信息,以及系统的训练和微调,生成准确、专业的回答,包括:

整合从医学知识库和文献数据库检索到的信息,包括从多个来源获取的文本、数据、图表各种形式的信息,将检索到的医学信息与用户问题和对话上下文相结合,有助于确保系统的回答与用户问题相关,并且能够维持对话的连贯性,将从数据库中检索到的信息进行摘要和概括,以确保回答的简洁性和易懂性,使用大语言模型进行微调,以适应医学领域的特定要求,通过在医学相关文本上进行微调,可以让模型更好地理解和生成医学内容;模型生成回答时可以通过控制多样性来提供更多样的答案,有助于覆盖多个方面并避免单一性的陈述;实施自动或人工的质量评估机制,以确保系统生成的回答是准确、可信的,并避免误导性信息;采用用户反馈机制,用于改进模型的回答质量,用户反馈可以用于优化和改进模型的训练,以提高系统的表现。

6. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述负责处理用户与系统的对话式交互,处理用户追问、澄清情况,包括:

系统需要理解和记住先前的对话内容,以确保连贯性,包括使用对话状态跟踪器来保持对话状态和上下文信息;对用户的追问或澄清进行分析,确定用户想要的是更多细节、解释还是确认,以适当地响应用户的请求,确定系统如何回应用户的不同类型追问或澄清,包括请求更多信息、提供更详细的解释、确认用户的意图;系统需要能够检测并纠正错误理解或处理错误,以避免提供错误答案并保持对话的准确性;在需要时,系统可以主动向用户提出问题以澄清或确认用户意图,确保系统理解准确,设计友好且直观的用户界面,以便用户清楚地理解系统的期望和功能,清晰的信息展示和用户指引是关键;若语音交互处理适用,则对于语音交互,系统需要适应语音识别的技术和语义理解,以实现对话式交互;当用户提出超出系统知识范围的问题时,系统需要适当地处理并引导用户,以确保用户理解系统的能力和局限性;系统需要实时或周期性地评估对话的质量,包括连贯性、准确性和用户满意度,以改进对话交互体验。

7. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述确保患者的个人健康信息得到充分的保护,遵守相关的隐私和数据安全法规,包括:

对于存储和传输的数据进行加密处理,确保即使数据被获取,无法被未授权的人或系统阅读;设定严格的访问权限和身份验证措施,只允许授权人员访问特定的患者信息,同时

记录和监控访问;对数据进行脱敏处理,特别是对于测试、开发和非生产环境,以减少对敏感信息的曝光;定期对系统进行安全审查和审计,以确保数据处理符合安全标准,发现潜在的风险并采取相应措施。

8. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述构建用户友好的界面,供患者输入问题并查看系统生成的回答,包括:

提供一个清晰、直观的文本框或语音输入框,让用户轻松地输入问题,在输入时可能会提供建议、自动完成或指导用户输入的格式;允许用户通过文本输入、语音输入甚至图像上传来提出问题,以适应不同用户的偏好和需求,当用户输入问题时,系统可以即时给予反馈,如自动完成建议、错误纠正或输入格式提示,以帮助用户更好地表达问题;以易于理解和浏览的方式呈现系统生成的回答,包括文本、图表、摘要、甚至相关文献链接,确保信息排版清晰,易于阅读;可展开内容、标签分类、答案的进一步深入查看交互元素,帮助用户深入了解问题的答案;确保界面在移动设备上的兼容性和友好性,包括响应式设计、触控友好的交互元素;提供一个渠道供用户提供反馈,以改进界面设计和系统功能,促进系统的不断优化。

9. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述负责定期更新医学知识库和医学文献数据库,同时持续优化大语言模型,包括:

建立定期的同步机制,以保证数据的及时更新,包括每日、每周或每月的自动同步;新数据获取后,进行清洗、去重、标准化和结构化处理,确保数据的一致性和可用性;建立数据库版本控制和备份机制,确保数据完整性和安全性,并且有能力回滚至之前的版本;建立自动化的数据更新流程,减少人工干预并保证数据更新的及时性和准确性;定期对大语言模型进行微调,使用最新的医学文献和数据,以适应医学领域的变化和需求;数据增利用新的数据和语料库对模型进行训练,扩大模型对医学知识的覆盖范围;定期评估模型的性能,包括准确性、效率和对新信息的适应能力;优化模型参数和架构,以提高回答的准确性和系统性能;及时应用模型更新,保持系统处于最新版本,以反映最新的模型改进和医学知识;利用用户反馈和系统日志,持续改进和优化模型,解决用户提出的问题和改进需求。

10. 如权利要求1所述的基于大语言模型的医学问答系统,其特征在于,所述确保系统提供基于医学专业知识的回答,对系统的回答进行评估和校对,包括:

整合来自可信赖的医学专业数据库和权威机构的信息,如医学期刊、学术论文、临床指南和专业学会的文档;引入医学领域的专家对系统回答进行审核、指导和补充,以确保回答符合医学标准和专业实践;使用知识图谱构建医学领域的信息关系和结构,帮助系统更好地理解信息间的相关性和逻辑关系;系统需要能够从文本中提取和验证医学知识,包括症状、诊断、治疗方法,确保回答的准确性;将来自多个权威来源的信息整合,以确保回答的全面性和准确性;建立自动化的质量评估机制,检查系统回答的准确性、一致性和合规性;让医学专家团队对系统回答进行人工审核和校对,确认回答的正确性和专业性;鼓励用户提供反馈,特别是对医学信息的准确性和实用性,以及根据用户反馈改进系统回答;基于错误分析和专家反馈,系统不断学习和更新,以改进和纠正回答中的错误或不准确信息;定期进行系统回答的审计和监管,确保系统遵循最新的医学标准和规定。

## 一种基于大语言模型的医学问答系统

### 技术领域

[0001] 本发明属于医学问答技术以及大语言模型的人工智能技术领域,尤其涉及一种基于大语言模型的医学问答系统。

### 背景技术

[0002] 传统模式下,患者通常依赖医生或医疗专业人员获取信息,这限制了信息的可及性和覆盖范围,患者需要预约和前往医疗机构,耗费时间和精力,而且仅能在特定的时间点获得帮助。在一些地区或特定领域,医疗资源可能稀缺,导致患者获取医学信息和建议困难。医生和患者之间的信息不对称,医生有专业知识,而患者可能对医学领域了解有限,导致交流和理解上的困难,在传统问诊模式下,患者可能需要长时间等待才能获得医生的回应或预约。患者在咨询医生后,可能由于记忆和理解差异,对医生建议的理解存在偏差或遗漏。有些患者可能因为主观难以启齿的问题或尴尬感而不愿直接面对医生进行问诊。传统问诊可能伴随着较高的医疗成本,尤其是在一些特定的医疗保健系统中。

### 发明内容

[0003] 本部分的目的在于概述本发明的实施例的一些方面以及简要介绍一些较佳实施例。在本部分以及本申请的说明书摘要和发明名称中可能会做些简化或省略以避免使本部分、说明书摘要和发明名称的目的模糊,而这种简化或省略不能用于限制本发明的范围。

[0004] 鉴于上述或现有的医学问答信息不对称、待时间长、时间地点限制的问题,提出了本发明。

[0005] 为解决上述技术问题,本发明提供如下技术方案:

[0006] 本发明实施例提供了一种基于大语言模型的医学问答系统,其特征在于,包括:输入处理模块,用于负责接收用户输入的医学问题,并进行预处理;语言理解模块,用于使用大语言模型来理解用户输入的问题,识别问题的主题、意图,并从中提取关键信息;数据库查询模块,用于连接医学知识库和医学文献数据库,根据用户的问题,查询相关的医学信息和答案;回答生成模块,用于使用大语言模型来生成系统的回答,根据查询到的医学信息,以及系统的训练和微调,生成准确、专业的回答;交互模块,用于负责处理用户与系统的对话式交互,处理用户追问、澄清并保持对话的连贯;数据隐私与安全模块,用于确保患者的个人健康信息得到充分的保护,遵守相关的隐私和数据安全法规;用户界面模块,用于构建用户友好的界面,供患者输入问题并查看系统生成的回答;数据更新与持续优化模块,用于负责定期更新医学知识库和医学文献数据库,以确保系统提供最新的和准确的信息,同时持续优化大语言模型,以提高系统的回答准确性;专业性与准确性模块,用于确保系统提供基于医学专业知识的回答,对系统的回答进行评估和校对,避免出现误导性的信息。

[0007] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述负责接收用户输入的医学问题,并进行预处理,包括:去除特殊字符、标点符号、多余空格、甚至拼写检查和纠正;对于分词,将输入的文本分割成单词或短语,分词考虑医学术语和缩

写的特殊处理,以确保术语被正确理解;移除停用词,去除在文本中出现频率很高但没有实际信息量的词语,以减少噪音并提高系统的效率;考虑上下文信息,尤其是在对话式交互中,系统需要理解先前的问题和回答,以确保连贯性和更准确的回应;将文本转换为数学表示如词嵌入,有助于模型理解词语之间的语义关联,从而更好地回答问题。

[0008] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述使用大语言模型来理解用户输入的问题,识别问题的主题、意图,并从中提取关键信息,包括:对大语言模型进行微调,使其更好地理解 and 回答医学相关问题,通过引入医学领域的文本数据进行模型训练来实现;将输入的问题归类到医学领域的特定主题,通过配置额外的分类器,识别问题所涉及的领域如药物、症状、疾病分类;确定用户提出问题的意图,如用户是在寻求解释、寻求建议,还是需要特定的信息,有助于更好地定位问题,并为用户提供更准确的答案;从问题中提取关键信息,包括实体如药物、症状、疾病、时间、事件,需要结合实体识别和上下文理解,以确保关键信息被正确地识别和提取;考虑对话中的上下文信息,确保问题的解释和答案与先前对话一致,从而提供连贯的对话体验;在需要时,能解释模型对问题的理解方式,有助于增强系统的可信度和用户对系统的信任感。

[0009] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述连接医学知识库和医学文献数据库,根据用户的问题,查询相关的医学信息和答案,包括:整合多个医学知识库和文献数据库,数据库包括医学书籍、期刊、研究论文、临床指南以及专业组织的文档;确保不同数据库的数据格式和结构一致,以便系统能够统一查询和提取信息,涉及到数据清洗、标准化和统一化处理;根据用户提出的问题,构建相应的查询以搜索相关信息,查询需要考虑用户问题的语义、实体识别的结果以及上下文信息,以提高查询的精准度;从数据库中提取相关的医学信息和答案,包括抽取文本、图表、统计数据,甚至是以结构化形式呈现的信息如数据库中的特定字段;将从数据库中检索到的信息与用户问题进行匹配,并对匹配度进行排序,涵盖文本相似度匹配、语义匹配方法,以确保返回最相关和准确的答案;确保数据库中的信息及时更新,以提供最新的医学知识,需要定期的数据同步和更新机制,以反映医学领域的最新发展;利用知识图谱构建数据库间关系,以帮助系统更好地理解医学信息的关联性和结构,从而提供更全面的答案;考虑用户问题的语境,以便更精准地过滤和提取相关信息,尤其是对于多义词、复杂问题或需要更详细上下文理解的情况。

[0010] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述使用大语言模型来生成系统的回答,根据查询到的医学信息,以及系统的训练和微调,生成准确、专业的回答,包括:整合从医学知识库和文献数据库检索到的信息,包括从多个来源获取的文本、数据、图表各种形式的信息,将检索到的医学信息与用户问题和对话上下文相结合,有助于确保系统的回答与用户问题相关,并且能够维持对话的连贯性,将从数据库中检索到的信息进行摘要和概括,以确保回答的简洁性和易懂性,使用大语言模型进行微调,以适应医学领域的特定要求,通过在医学相关文本上进行微调,可以让模型更好地理解和生成医学内容;模型生成回答时可以通过控制多样性来提供更多样的答案,有助于覆盖多个方面并避免单一性的陈述;实施自动或人工的质量评估机制,以确保系统生成的回答是准确、可信的,并避免误导性信息;采用用户反馈机制,用于改进模型的回答质量,用户反馈可以用于优化和改进模型的训练,以提高系统的表现。

[0011] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述负责处理用户与系统的对话式交互,处理用户追问、澄清情况,包括:系统需要理解和记住先前的对话内容,以确保连贯性,包括使用对话状态跟踪器来保持对话状态和上下文信息;对用户的追问或澄清进行分析,确定用户想要的是更多细节、解释还是确认,以适当地响应用户的请求,确定系统如何回应用户的不同类型追问或澄清,包括请求更多信息、提供更详细的解释、确认用户的意图;系统需要能够检测并纠正错误理解或处理错误,以避免提供错误答案并保持对话的准确性;在需要时,系统可以主动向用户提出问题以澄清或确认用户意图,确保系统理解准确,设计友好且直观的用户界面,以使用户清楚地理解系统的期望和功能,清晰的信息展示和用户指引是关键;若语音交互处理适用,则对于语音交互,系统需要适应语音识别的技术和语义理解,以实现对话式交互;当用户提出超出系统知识范围的问题时,系统需要适当地处理并引导用户,以确保用户理解系统的能力和局限性;系统需要实时或周期性地评估对话的质量,包括连贯性、准确性和用户满意度,以改进对话交互体验。

[0012] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述确保患者的个人健康信息得到充分的保护,遵守相关的隐私和数据安全法规,包括:对于存储和传输的数据进行加密处理,确保即使数据被获取,无法被未授权的人或系统阅读;设定严格的访问权限和身份验证措施,只允许授权人员访问特定的患者信息,同时记录和监控访问;对数据进行脱敏处理,特别是对于测试、开发和非生产环境,以减少对敏感信息的曝光;定期对系统进行安全审查和审计,以确保数据处理符合安全标准,发现潜在的风险并采取相应措施。

[0013] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述构建用户友好的界面,供患者输入问题并查看系统生成的回答,包括:

[0014] 提供一个清晰、直观的文本框或语音输入框,让用户轻松地输入问题,在输入时可能会提供建议、自动完成或指导用户输入的格式;允许用户通过文本输入、语音输入甚至图像上传来提出问题,以适应不同用户的偏好和需求,当用户输入问题时,系统可以即时给予反馈,如自动完成建议、错误纠正或输入格式提示,以帮助用户更好地表达问题;以易于理解和浏览的方式呈现系统生成的回答,包括文本、图表、摘要、甚至相关文献链接,确保信息排版清晰,易于阅读;可展开内容、标签分类、答案的进一步深入查看交互元素,帮助用户深入了解问题的答案;确保界面在移动设备上的兼容性和友好性,包括响应式设计、触控友好的交互元素;提供一个渠道供用户提供反馈,以改进界面设计和系统功能,促进系统的不断优化。

[0015] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述负责定期更新医学知识库和医学文献数据库,同时持续优化大语言模型,包括:建立定期的同步机制,以保证数据的及时更新,包括每日、每周或每月的自动同步;新数据获取后,进行清洗、去重、标准化和结构化处理,确保数据的一致性和可用性;建立数据库版本控制和备份机制,确保数据完整性和安全性,并且有能力回滚至之前的版本;建立自动化的数据更新流程,减少人工干预并保证数据更新的及时性和准确性;定期对大语言模型进行微调,使用最新的医学文献和数据,以适应医学领域的变化和 demand;数据增利用新的数据和语料库对模型进行训练,扩大模型对医学知识的覆盖范围;定期评估模型的性能,包括准确性、效率和对新信息的适应能力;优化模型参数和架构,以提高回答的准确性和系统性能;及时应用模



型更新,保持系统处于最新版本,以反映最新的模型改进和医学知识;利用用户反馈和系统日志,持续改进和优化模型,解决用户提出的问题和改进需求。

[0016] 作为本发明所述基于大语言模型的医学问答系统的一种优选方案,其中:所述确保系统提供基于医学专业知识的回答,对系统的回答进行评估和校对,包括:整合来自可信赖的医学专业数据库和权威机构的信息,如医学期刊、学术论文、临床指南和专业学会的文档;引入医学领域的专家对系统回答进行审核、指导和补充,以确保回答符合医学标准和专业实践;使用知识图谱构建医学领域的信息关系和结构,帮助系统更好地理解信息间的相关性和逻辑关系;系统需要能够从文本中提取和验证医学知识,包括症状、诊断、治疗方法,确保回答的准确性;将来自多个权威来源的信息整合,以确保回答的全面性和准确性;建立自动化的质量评估机制,检查系统回答的准确性、一致性和合规性;让医学专家团队对系统回答进行人工审核和校对,确认回答的正确性和专业性;鼓励用户提供反馈,特别是对医学信息的准确性和实用性,以及根据用户反馈改进系统回答;基于错误分析和专家反馈,系统不断学习和更新,以改进和纠正回答中的错误或不准确信息;定期进行系统回答的审计和监管,确保系统遵循最新的医学标准和规定。

[0017] 本发明的有益效果为:患者可以随时随地通过系统提出问题,无需待医生预约或前往医疗机构,提供了便捷的医学咨询途径,系统整合了多个医学知识库和文献数据库的信息,覆盖了丰富的医学领域知识,提供了更广泛的信息资源,患者可以匿名提问,保护了个人隐私,并允许他们就一些尴尬或私密的问题进行咨询,系统能够快速响应用户问题,提供即时的帮助和信息,减少了患者等待医疗咨询的时间,系统通过回答问题,帮助患者理解医学知识,提高医学素养和对疾病的认知,释放部分医生资源,让医生可以更专注于更复杂、紧急的病例,提高医疗资源的有效利用,通过持续更新和优化,系统可以不断学习和改进,提供更准确、及时的信息和建议,对于一些常见问题,患者无需花费额外医疗成本,同时也为医疗保健系统降低了压力。

## 附图说明

[0018] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其它的附图。其中:

[0019] 图1为本发明实施例提供的一种基于大语言模型的医学问答系统结构示意图。

[0020] 图2为本发明实施例提供的一种基于大语言模型的医学问答系统的流程图。

## 具体实施方式

[0021] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面结合说明书附图对本发明的具体实施方式做详细的说明。

[0022] 在下面的描述中阐述了很多具体细节以便于充分理解本发明,但是本发明还可以采用其他不同于在此描述的其它方式来实施,本领域技术人员可以在不违背本发明内涵的情况下做类似推广,因此本发明不受下面公开的具体实施例的限制。

[0023] 其次,此处所称的“一个实施例”或“实施例”是指可包含于本发明至少一个实现方



式中的特定特征、结构或特性。在本说明书中不同地方出现的“在一个实施例中”并非均指同一个实施例,也不是单独的或选择性的与其他实施例互相排斥的实施例。

[0024] 实施例

[0025] 参照图1~图2,为本发明第一个实施例,该实施例提供了一种基于大语言模型的医学问答系统,包括:

[0026] S1:负责接收用户输入的医学问题,并进行预处理。

[0027] 优选的,去除特殊字符、标点符号、多余空格、甚至拼写检查和纠正;对于分词,将输入的文本分割成单词或短语,分词考虑医学术语和缩写的特殊处理,以确保术语被正确理解;移除停用词,去除在文本中出现频率很高但没有实际信息量的词语,以减少噪音并提高系统的效率;考虑上下文信息,尤其是在对话式交互中,系统需要理解先前的问题和回答,以确保连贯性和更准确的回应;将文本转换为数学表示如词嵌入,有助于模型理解词语之间的语义关联,从而更好地回答问题。

[0028] S2:使用大语言模型来理解用户输入的问题,识别问题的主题、意图,并从中提取关键信息。

[0029] 优选的,对大语言模型进行微调,使其更好地理解和回答医学相关问题,通过引入医学领域的文本数据进行模型训练来实现;将输入的问题归类到医学领域的特定主题,通过配置额外的分类器,识别问题所涉及的领域如药物、症状、疾病分类等;确定用户提出问题的意图,如用户是在寻求解释、寻求建议,还是需要特定的信息等,有助于更好地定位问题,并为用户提供更准确的答案;从问题中提取关键信息,包括实体如药物、症状、疾病、时间、事件等,需要结合实体识别和上下文理解,以确保关键信息被正确地识别和提取;考虑对话中的上下文信息,确保问题的解释和答案与先前对话一致,从而提供连贯的对话体验;在需要时,能解释模型对问题的理解方式,有助于增强系统的可信度和用户对系统的信任感。

[0030] S3:连接医学知识库和医学文献数据库,根据用户的问题,查询相关的医学信息和答案。

[0031] 优选的,整合多个医学知识库和文献数据库,数据库包括医学书籍、期刊、研究论文、临床指南以及专业组织的文档;确保不同数据库的数据格式和结构一致,以便系统能够统一查询和提取信息,涉及到数据清洗、标准化和统一化处理;根据用户提出的问题,构建相应的查询以搜索相关信息,查询需要考虑用户问题的语义、实体识别的结果以及上下文信息,以提高查询的精准度;从数据库中提取相关的医学信息和答案,包括抽取文本、图表、统计数据,甚至是以结构化形式呈现的信息如数据库中的特定字段;将从数据库中检索到的信息与用户问题进行匹配,并对匹配度进行排序,涵盖文本相似度匹配、语义匹配等方法,以确保返回最相关和准确的答案;确保数据库中的信息及时更新,以提供最新的医学知识,需要定期的数据同步和更新机制,以反映医学领域的最新发展;利用知识图谱构建数据库间关系,以帮助系统更好地理解医学信息的关联性和结构,从而提供更全面的答案;考虑用户问题的语境,以便更精准地过滤和提取相关信息,尤其是对于多义词、复杂问题或需要更详细上下文理解的情况。

[0032] 进一步的,对从不同数据库获取的数据进行清洗,去除不一致、重复或无效的数据条目,这包括处理缺失值、修复错误或格式不一致的数据等;统一不同数据库中数据的格

式,确保数据在不同源之间保持一致性,这包括统一日期格式、单位标准、命名规范等;对于相似但命名不同的字段,建立映射关系,以便系统能够识别并提取相同类型的信息,例如,将不同数据库中的“症状”、“病历”字段进行映射;对于异常值检测和处理,基于统计学规则识别异常值,使Z-score方法处理,其表达式如下所示:

$$[0033] \quad Z = (x - \mu) / \sigma$$

[0034] 其中: $x$ 代表单个数据点的数值, $\mu$ 代表整个数据集的平均值, $\sigma$ 代表整个数据集的标准差。

[0035] 进一步的,医学知识库中的各种实体(疾病、症状、药物、治疗方法等)表示为图谱的节点,定义实体之间的关系,比如疾病与症状之间的关联、药物与治疗方法之间的应用关系等,为图谱中的边/关系添加属性,比如关联的强度、频率、权重等,以反映实体之间关系的特征;建立一个图数据库来存储知识图谱,使得系统可以高效地查询和分析图谱中的信息,利用本体(Ontology)或标准化语义模型来定义实体和关系,帮助系统理解和推断医学概念的含义,通过图数据库优化查询,实现对图谱的高效搜索和信息提取,以支持系统对用户问题的准确回答,利用知识图谱的结构和关联,系统可以进行推理,找到潜在的关联信息并提供更深层次的解答。

[0036] S4:使用大语言模型来生成系统的回答,根据查询到的医学信息,以及系统的训练和微调,生成准确、专业的回答。

[0037] 优选的,整合从医学知识库和文献数据库检索到的信息,包括从多个来源获取的文本、数据、图表等各种形式的信息,将检索到的医学信息与用户问题和对话上下文相结合,有助于确保系统的回答与用户问题相关,并且能够维持对话的连贯性,将从数据库中检索到的信息进行摘要和概括,以确保回答的简洁性和易懂性,使用大语言模型进行微调,以适应医学领域的特定要求,通过在医学相关文本上进行微调,可以让模型更好地理解和生成医学内容;模型生成回答时可以通过控制多样性来提供更多样的答案,有助于覆盖多个方面并避免单一性的陈述;实施自动或人工的质量评估机制,以确保系统生成的回答是准确、可信的,并避免误导性信息;采用用户反馈机制,用于改进模型的回答质量,用户反馈可以用于优化和改进模型的训练,以提高系统的表现。

[0038] 进一步的,在系统界面中嵌入反馈按钮或表单,让用户能够方便地提供反馈。提供多种反馈选项,比如“回答有帮助”、“回答不清楚”、“存在错误”等。系统记录用户的反馈,包括用户提供的具体信息,比如问题、系统回答,以及用户对回答的评价或修正。这些数据被用于模型训练的改进。分析用户反馈的情境和模式,找出用户对回答不满意的共同特点或模式。根据用户反馈,对模型进行更新和改进。例如,针对频繁出现的错误回答或不清晰的解释,对模型进行微调和修正。模型不断学习用户的反馈,并进行自我优化,以逐步提高回答的质量和准确性。

[0039] S5:负责处理用户与系统的对话式交互,处理用户追问、澄清等情况。

[0040] 优选的,系统需要理解和记住先前的对话内容,以确保连贯性,包括使用对话状态跟踪器来保持对话状态和上下文信息;对用户的追问或澄清进行分析,确定用户想要的是更多细节、解释还是确认,以适当地响应用户的请求,确定系统如何回应用户的不同类型追问或澄清,包括请求更多信息、提供更详细的解释、确认用户的意图等;系统需要能够检测并纠正错误理解或处理错误,以避免提供错误答案并保持对话的准确性;在需要时,系统可

以主动向用户提出问题以澄清或确认用户意图,确保系统理解准确,设计友好且直观的用户界面,以使用户清楚地理解系统的期望和功能,清晰的信息展示和用户指引是关键;若语音交互处理适用,则对于语音交互,系统需要适应语音识别的技术和语义理解,以实现对话式交互;当用户提出超出系统知识范围的问题时,系统需要适当地处理并引导用户,以确保用户理解系统的能力和局限性;系统需要实时或周期性地评估对话的质量,包括连贯性、准确性和用户满意度,以改进对话交互体验。

[0041] 进一步的,若用户询问一个非常专业或特定的医学问题,超出系统的知识范围,则系统以友善的语气回应用户,如:“非常感谢您的问题,我们的系统致力于提供医学领域的帮助。”系统明确表示自己的专业范围和限制,“我们主要涵盖常见疾病和一般性的医学知识。”建议用户向专业医生咨询,或是查阅特定的医学期刊或专业网站,“对于更专业或特定的问题,我们建议您咨询专业医生或查阅权威医学文献。”系统鼓励用户再次使用系统,“如果您有更多问题,随时欢迎再次向我们咨询。”

[0042] S6:确保患者的个人健康信息得到充分的保护,遵守相关的隐私和数据安全法规。

[0043] 优选的,对于存储和传输的数据进行加密处理,确保即使数据被获取,无法被未经授权的人或系统阅读;设定严格的访问权限和身份验证措施,只允许授权人员访问特定的患者信息,同时记录和监控访问;对数据进行脱敏处理,特别是对于测试、开发和非生产环境,以减少对敏感信息的曝光;定期对系统进行安全审查和审计,以确保数据处理符合安全标准,发现潜在的风险并采取相应措施。

[0044] 进一步的,要求授权人员进行双因素身份验证,比如结合密码和动态验证码,使用生物识别技术确认授权人员身份;设定不同角色的权限,根据员工职责授予对应权限,比如医生、护士、行政人员等。授权人员只拥有完成工作所需的最小权限;记录所有对患者信息的访问,包括时间、访问人员、访问内容等详细信息,定期审计访问日志,检查潜在的异常访问行为;对存储和传输的患者信息进行加密保护,确保数据安全性;设定访问控制列表,只允许特定人员或部门访问特定信息;限制访问患者信息的网络范围,确保只有内部受信任的网络可以访问敏感信息;限制只有特定的IP地址或区域可以访问敏感信息。

[0045] S7:构建用户友好的界面,供患者输入问题并查看系统生成的回答。

[0046] 优选的,提供一个清晰、直观的文本框或语音输入框,让用户轻松地输入问题,在输入时可能会提供建议、自动完成或指导用户输入的格式;允许用户通过文本输入、语音输入甚至图像上传来提出问题,以适应不同用户的偏好和需求,当用户输入问题时,系统可以即时给予反馈,如自动完成建议、错误纠正或输入格式提示,以帮助用户更好地表达问题;以易于理解和浏览的方式呈现系统生成的回答,包括文本、图表、摘要、甚至相关文献链接等,确保信息排版清晰,易于阅读;可展开内容、标签分类、答案的进一步深入查看等交互元素,帮助用户深入了解问题的答案;确保界面在移动设备上的兼容性和友好性,包括响应式设计、触控友好的交互元素等;提供一个渠道供用户提供反馈,以改进界面设计和系统功能,促进系统的不断优化。

[0047] 进一步的,当用户开始输入时,系统可以提供建议或自动补全,根据先前提问的常见模式或相似问题进行推荐。例如,用户开始输入“头痛”,系统可以自动建议“头痛的原因”、“头痛的治疗方法”等问题。系统能够即时检测和纠正输入中的拼写错误、语法错误或格式问题,并给予相应提示或建议。如果用户输入含糊不清或模糊的问题,系统可以提示用

户澄清或提供更多细节。系统在用户输入的同时,实时给予反馈。例如,当用户说出一句话或输入文字时,系统可以立即进行文本或语音回应,确认问题是否被正确理解。

[0048] 进一步的,若用户在输入框中键入“高血压是什么?”,系统立即给出类似问题的建议,如“高血压的症状”、“高血压的治疗方法”等;如果用户拼写错误,系统会即时提示并建议正确的拼写;用户语音说出“我了解高血压的症状。”,系统进行语音回应,“了解高血压症状,请稍等,正在为您查找相关信息。”;用户上传一张含有“高血压是什么?”问题的图片,系统进行光学字符识别,将图像中的文字转换为问题输入,并进行相关搜索。

[0049] S8:负责定期更新医学知识库和医学文献数据库,同时持续优化大语言模型。

[0050] 优选的,建立定期的同步机制,以保证数据的及时更新,包括每日、每周或每月的自动同步;新数据获取后,进行清洗、去重、标准化和结构化处理,确保数据的一致性和可用性;建立数据库版本控制和备份机制,确保数据完整性和安全性,并且有能力回滚至之前的版本;建立自动化的数据更新流程,减少人工干预并保证数据更新的及时性和准确性;定期对大语言模型进行微调,使用最新的医学文献和数据,以适应医学领域的变化和需求;数据增利用新的数据和语料库对模型进行训练,扩大模型对医学知识的覆盖范围;定期评估模型的性能,包括准确性、效率和对新信息的适应能力;优化模型参数和架构,以提高回答的准确性和系统性能;及时应用模型更新,保持系统处于最新版本,以反映最新的模型改进和医学知识;利用用户反馈和系统日志,持续改进和优化模型,解决用户提出的问题和改进需求。

[0051] S9:确保系统提供基于医学专业知识的回答,对系统的回答进行评估和校对。

[0052] 优选的,整合来自可信赖的医学专业数据库和权威机构的信息,如医学期刊、学术论文、临床指南和专业学会的文档;引入医学领域的专家对系统回答进行审核、指导和补充,以确保回答符合医学标准和专业实践;使用知识图谱构建医学领域的信息关系和结构,帮助系统更好地理解信息间的相关性和逻辑关系;系统需要能够从文本中提取和验证医学知识,包括症状、诊断、治疗方法等,确保回答的准确性;将来自多个权威来源的信息整合,以确保回答的全面性和准确性。

[0053] 优选的,建立自动化的质量评估机制,检查系统回答的准确性、一致性和合规性;让医学专家团队对系统回答进行人工审核和校对,确认回答的正确性和专业性;鼓励用户提供反馈,特别是对医学信息的准确性和实用性,以及根据用户反馈改进系统回答;基于错误分析和专家反馈,系统不断学习和更新,以改进和纠正回答中的错误或不准确信息;定期进行系统回答的审计和监管,确保系统遵循最新的医学标准和规定。

[0054] 应说明的是,以上实施例仅用于说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明技术方案的精神和范围,其均应涵盖在本发明的权利要求范围当中。

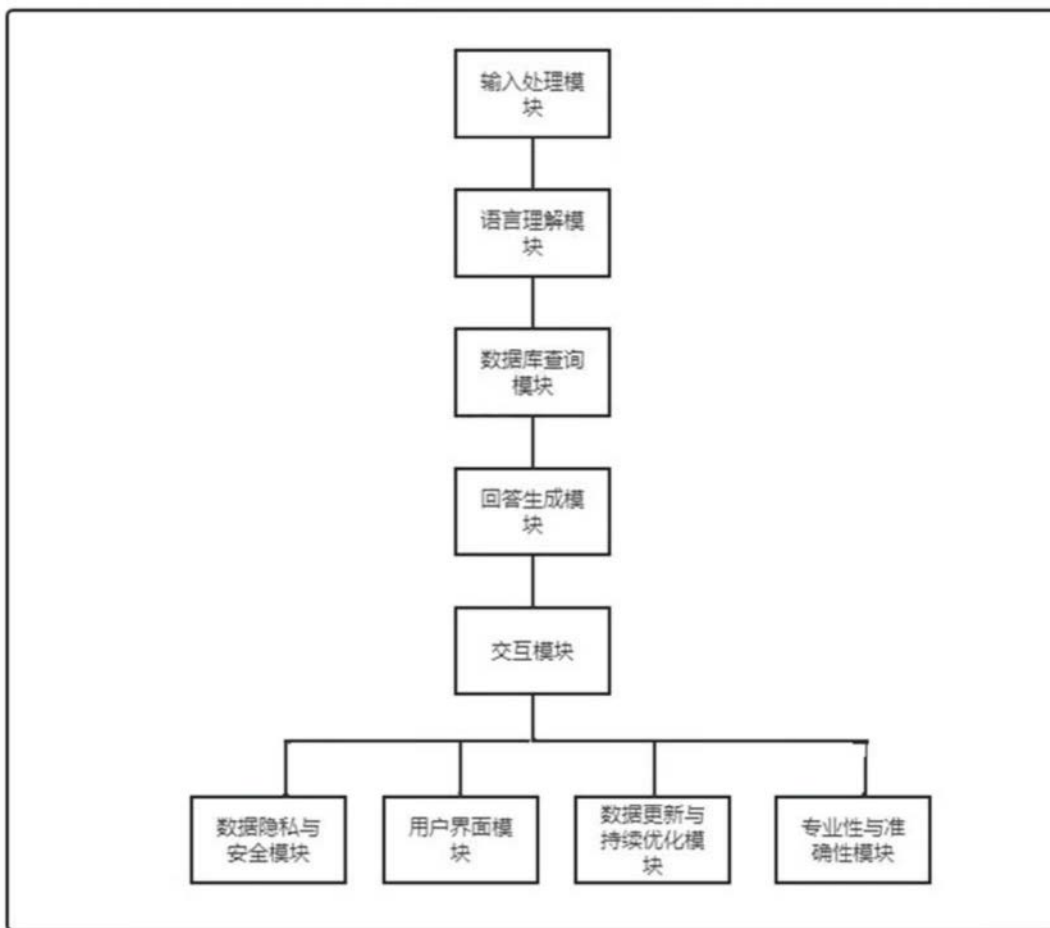


图1

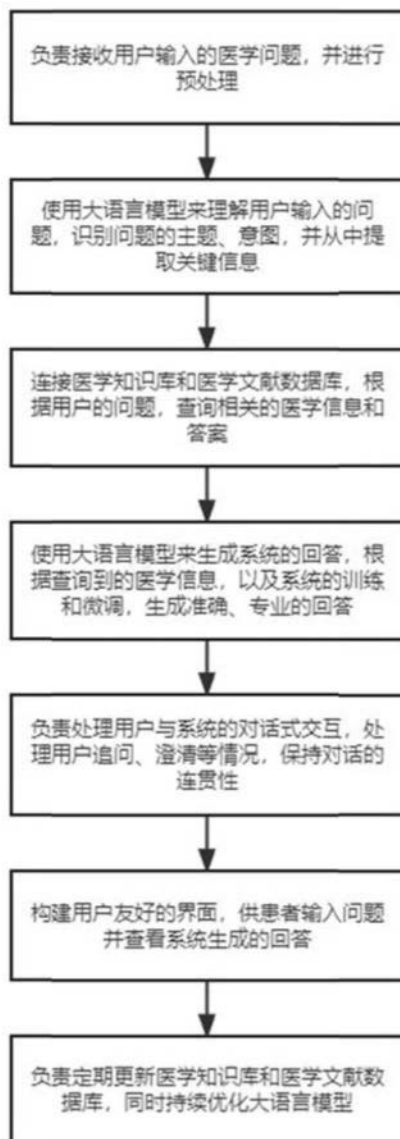


图2