



(12) 发明专利申请

(10) 申请公布号 CN 114090789 A

(43) 申请公布日 2022. 02. 25

(21) 申请号 202111369070.0

G06F 16/33 (2019.01)

(22) 申请日 2021.11.12

(71) 申请人 复旦大学

地址 200433 上海市杨浦区邯郸路220号

申请人 浙江智树网络科技有限公司

(72) 发明人 周峰 吕智慧 严绍根 陈宇

林榕健 徐杨川

(74) 专利代理机构 上海德昭知识产权代理有限公司

公司 31204

代理人 郁旦蓉

(51) Int. Cl.

G06F 16/36 (2019.01)

G06F 16/338 (2019.01)

G06F 16/335 (2019.01)

G06F 16/332 (2019.01)

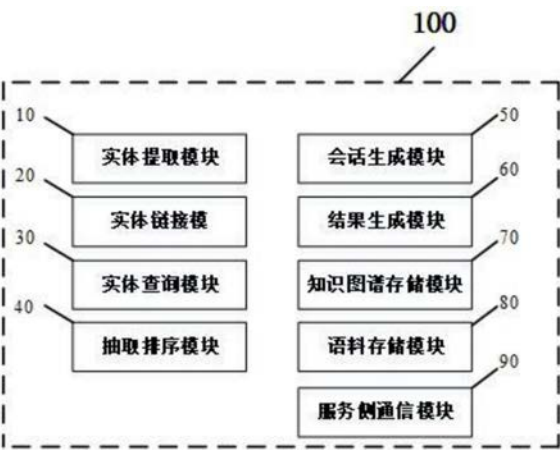
权利要求书2页 说明书6页 附图4页

(54) 发明名称

基于知识图谱的中医养生智能多轮交互系统

(57) 摘要

本发明属于人工智能知识图谱领域,提供了一种基于知识图谱的中医养生智能多轮交互系统,因为具备记忆能力可以主动关联与用户交互的内容,用户向系统输入临床表现后,系统会根据用户输入的临床表现的特征,主动向用户提问相关主诉表现,用户根据系统提问的问题输入反馈内容,智能交互系统使用这样的方式,经过与用户多轮交互后,最终给予用户有较高准确率的中医养生建议。本发明的交互系统采用多轮对话的交互形式,在终端用户的体验上和中医养生咨询问答的结论准确度上均高于普通单轮交互的问答系统。本发明针对前后文中识别到的实体进行实体链接,结合前后文的文本信息基于记忆序列的实体链接模型来解决多轮会话中语句的实体省略和指代消解。



1. 一种基于知识图谱的中医养生智能多轮交互系统,其特征在于,包括:
实体提取模块,用于提取每轮交互中用户输入的临床信息中的实体;
实体链接模块,用于对所述实体进行指代消解与实体消歧,得到处理后的所述实体;
实体查询模块,用于从预存储的中医养生知识图谱中获取与处理后的所述实体相关的多个三元组;

抽取排序模块,用于筛选需要所述用户再次输入的所述临床信息的若干个相关三元组,并将所述若干三元组按照预先设定的优先级排序,得到三元组优先级序列并确定最高优先级三元组;

会话生成模块,用于根据所述最高优先级三元组和以及预存储的所述中医养生知识图谱生成问题,以对所述用户进行下一轮提问;以及

结果生成模块,用于根据所述用户多次输入的所述临床信息以及预存储的所述中医养生知识图谱生成养生建议结果,

其中,所述实体链接模块包括前后文编码器,输入文本编码器,候选文本编码器,注意力机制层,输出结果层。

2. 根据权利要求1所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于:

其中,所述前后文编码器用于提取相邻的两轮交互中的所述实体的特征。

3. 根据权利要求1所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于:

其中,所述前后文编码器,所述输入文本编码器,所述候选文本编码器均包括双向长短期记忆神经网络。

4. 根据权利要求1所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于,还包括:

知识图谱存储模块,

其中,所述知识图谱存储模块用于存储所述中医养生知识图谱,

所述中医养生知识图谱包括中医养生本体知识与中医养生自定义词典。

5. 根据权利要求1所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于:

其中,所述中医养生知识图谱的源数据包括中医院病例报告表,

所述中医院病例报告表的内容包括:性别、年龄、职业、民族、婚姻、心率、血压、呼吸、首次发病时间、病程、主诉、临床表现、其他疾病史、简要病史、检查单、病理报告单、舌脉、西医诊断、中医辨证、中医诊断、方药、治疗方案、养护方案、饮食、运动锻炼、生活习惯、按摩推拿。

6. 根据权利要求5所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于:

其中,所述中医养生知识图谱的构建包括如下步骤:

步骤S1,对所述中医院病例报告表的内容进行实体抽取、关系抽取以及属性抽取并得到实体抽取结果、关系抽取结果以及属性抽取结果;

步骤S2,根据所述实体抽取结果、所述关系抽取结果以及所述属性抽取结果进行节点创建和关系创建,并得到所述中医养生知识图谱。

7. 根据权利要求1所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于,还包括:

语料存储模块,

其中,所述语料存储模块用于向所述会话生成模块和所述结果生成模块提供语料。

8. 根据权利要求7所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于:

其中,存储在所述语料存储模块中的语料的生成过程包括如下步骤:

步骤A1,分析所述中医养生知识图谱中的实体和实体关系,得到实体关系图;

步骤A2,将所述实体关系图转化为有向图;

步骤A3,以所述实体关系图的每个所述实体为起点,遍历所述有向图获取路径图;

步骤A4,对所述路径图中的路径进行分拆与组合获取子路径;

步骤A5,将所述中医养生知识图谱中的所述实体和所述实体关系对应到各个所述子路径中,得到所述语料。

9. 根据权利要求1所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于,还包括:

服务侧通信模块,

其中,所述服务侧通信模块用于接受所述用户发送的所述临床信息数据。

10. 根据权利要求1所述的基于知识图谱的中医养生智能多轮交互系统,其特征在于:

其中,所述服务侧通信模块包括Nginx代理集群服务器、负载均衡器以及Web应用集群服务器。

基于知识图谱的中医养生智能多轮交互系统

技术领域

[0001] 本发明属于人工智能知识图谱领域,具体涉及一种基于知识图谱的中医养生智能多轮交互系统。

背景技术

[0002] 国家已建立符合中医药特点的管理制度,让中医药在我国医药卫生领域充分发挥重要作用,而中医养生是我国医药卫生领域里中医药的重要组成部分。随着移动互联网的普及,人们通过移动终端进行中医养生相关知识咨询已经成为常态化应用场景。当前中医养生领域智能交互系统主要为问答系统,而这些问答系统多为单轮智能交互系统,即用户提问一个问题,系统根据问题返回答案,当用户再次提问问题时,之前的提问问题系统不会关联,即系统没有记忆能力。这会造成用户需要进行大量重复提问,才可能得到问题的答案,甚至无法得到准确答案。

[0003] 目前也有一些问答系统采用了多轮交互的方式,但是在多轮交互过程中,由于常规的会话生成模型是将交互的前后文本拼接,对拼接的文本进行建模,但是前后文本都是由多个文字语句组成,拼接后的文本较长,这种建模方式得到的前后文本向量信息会因为前文的信息被分散而变的不完善,用户和系统的会话会出现指代消解和实体省略的情况,而这种情况会对实体链接造成极大的困扰,进而导致用户无法得到其问题的准确答案。

发明内容

[0004] 本发明是为了解决上述问题而进行的,目的在于提供一种基于知识图谱的中医养生智能多轮交互系统。

[0005] 本发明提供了一种基于知识图谱的中医养生智能多轮交互系统,具有这样的特征,包括:实体提取模块,用于提取每轮交互中用户输入的临床信息中的实体;实体链接模块,对实体进行指代消解与实体消歧,得到处理后的实体;实体查询模块,用于从预存储的中医养生知识图谱中获取与处理后的实体相关的多个三元组;抽取排序模块,用于筛选需要用户再次输入的临床信息的若干个相关三元组,并将若干三元组按照预先设定的优先级排序,得到三元组优先级序列并确定最高优先级三元组;会话生成模块,用于根据最高优先级三元组和以及预存储的中医养生知识图谱生成问题,以对用户进行下一轮提问;以及结果生成模块,用于根据用户多次输入的临床信息以及预存储的中医养生知识图谱生成养生建议结果,其中,实体链接模块包括前后文编码器,输入文本编码器,候选文本编码器,注意力机制层,输出结果层。

[0006] 发明的作用与效果

[0007] 根据本发明所涉及的基于知识图谱的中医养生智能多轮交互系统,因为交互系统具备记忆能力,可以主动关联与用户交互的内容,用户向系统输入临床表现后,系统会根据用户输入的临床表现的特征,主动向用户提问相关主诉表现,用户根据系统提问的问题,向系统输入相关反馈内容,智能交互系统使用这样的方式,经过与用户多轮交互后,最终给予

用户有较高准确率的中医养生建议。而且,本发明的交互系统由于是采用多轮对话的交互形式,在终端用户的体验上和中医养生咨询问答的结论准确度上均高于普通单轮交互的问答系统。此外,本发明针对前后文中识别到的实体进行实体链接,提出了结合前后文的文本信息基于记忆序列的实体链接模型来解决多轮会话中语句的实体省略和指代消解。

附图说明

[0008] 图1是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的结构框图;

[0009] 图2是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的硬件架构图;

[0010] 图3是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的软件架构图;

[0011] 图4是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的实体链接模块的工作逻辑图;

[0012] 图5是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的中医养生知识图谱的构建流程图;

[0013] 图6是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的会话生成模块的工作逻辑图;

[0014] 图7是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的语料存储模块中的语料的生成流程图;

[0015] 图8是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的运行逻辑图。

具体实施方式

[0016] 为了使本发明实现的技术手段、创作特征、达成目的与功效易于明白了解,以下实施例结合附图对本发明基于知识图谱的中医养生智能多轮交互系统作具体阐述。

[0017] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征:其中,前后文编码器用于提取相邻的两轮交互中的实体的特征。

[0018] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征:其中,前后文编码器,输入文本编码器,候选文本编码器均包括双向长短期记忆神经网络。

[0019] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征,还包括:知识图谱存储模块,其中,知识图谱存储模块用于存储中医养生知识图谱,中医养生知识图谱包括中医养生本体知识与中医养生自定义词典。

[0020] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征:其中,中医养生知识图谱的源数据包括中医院病例报告表,中医院病例报告表的内容包括:性别、年龄、职业、民族、婚姻、心率、血压、呼吸、首次发病时间、病程、主诉、临床表现、其他疾病史、简要病史、检查单、病理报告单、舌脉、西医诊断、中医辨证、中医诊断、方药、治疗方案、养护方案、饮食、运动锻炼、生活习惯、按摩推拿。

[0021] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征:其中,中医养生知识图谱的构建包括如下步骤:步骤S1,对中医院病例报告表的内容进行实体抽取、关系抽取以及属性抽取并得到实体抽取结果、关系抽取结果以及属性抽取结果;步骤S2,根据实体抽取结果、关系抽取结果以及属性抽取结果进行节点创建和关系创建,并得到中医养生知识图谱。

[0022] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征:语料存储模块,其中,语料存储模块用于向会话生成模块和结果生成模块提供语料。

[0023] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征:其中,存储在语料存储模块中的语料的生成过程包括如下步骤:步骤A1,分析中医养生知识图谱中的实体和实体关系,得到实体关系图;步骤A2,将实体关系图转化为有向图;步骤A3,以实体关系图的每个实体为起点,遍历有向图获取路径图;步骤A4,对路径图中的路径进行分拆与组合获取子路径;步骤A5,将中医养生知识图谱中的实体和实体关系对应到各个子路径中,得到语料。

[0024] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征,还包括:服务侧通信模块,其中,服务侧通信模块用于接受用户发送的临床信息数据。

[0025] 在本发明提供的基于知识图谱的中医养生智能多轮交互系统中,还可以具有这样的特征:其中,服务侧通信模块包括Nginx代理集群服务器、负载均衡器以及Web应用集群服务器。

[0026] 本实施例中以Android手机端作为用户终端为例对本发明的基于知识图谱的中医养生智能多轮交互系统进行说明,其中,多个用户终端由用户持有。

[0027] 本领域技术人员可以理解,虽然本实施例中以Android手机端作为用户终端为例,但是用户终端也可以为IOS手机端、PC端等用户终端。

[0028] 图1是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的结构框图。

[0029] 如图1所示,基于知识图谱的中医养生智能多轮交互系统100可以包括:实体提取模块10、实体链接模块20、实体查询模块30、抽取排序模块40、会话生成模块50、结果生成模块60、知识图谱存储模块70、语料存储模块80以及服务侧通信模块90。

[0030] 服务侧通信模块90可以用于接受用户发送的临床信息数据。

[0031] 图2是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的硬件架构图。图3是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的软件架构图。

[0032] 如图2、图3所示,服务侧通信模块包括Nginx代理集群服务器91、负载均衡器92以及Web应用集群服务器93。

[0033] 服务侧通信模块通过网络层的API接口接收用户发送的数据包,用户的载体是Android客户端,Android客户端依据RESTFUL接口要求发送JSON字符串与系统进行数据交互,同时向用户提供交互UI前端界面;Spring Boot负责将Nginx等服务整合到服务端;Spring MVC负责将Model、View、Controller接口提供给服务端,以及提供RESTFUL接口服务。

[0034] 实体提取模块10可以用于提取每轮交互中用户输入的临床信息中的实体。

[0035] 实体链接模块20,可以用于对实体进行指代消解与实体消歧,得到处理后的实体。

[0036] 图4是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的实体链接模块的工作逻辑图。

[0037] 如图4所示,实体链接模块20包括前后文编码器,输入文本编码器,候选文本编码器,注意力机制层以及输出结果层。其中,前后文编码器负责提取前后文的特征。前后文编码器,输入文本编码器,候选文本编码器均是采用双向LSTM,以输入文本编码器为例,输入文本 $\{w_1, w_2, \dots, w_n\}$ (w_n 即w中的第n个词),即在f时刻,正向LSTM得出 $\overrightarrow{w_f}$,反向LSTM得出 $\overleftarrow{w_f}$, f时刻的输出 w_f 可以通过拼接 $\overrightarrow{w_f}$ 和 $\overleftarrow{w_f}$ 得到。公式如下:

$$[0038] \quad w_f = [\overrightarrow{w_f}; \overleftarrow{w_f}] = [LSTM(w_f, \overrightarrow{w_{f-1}}); LSTM(w_f, \overleftarrow{w_{f+1}})]$$

[0039] 最后,再使用注意力机制计算过的候选向量c、前后文向量s、输入文本向量w传递到reply层中,得出实体链接结果。计算公式如下:

$$[0040] \quad s = \text{Attention}(c, w)$$

$$[0041] \quad c = \text{Attention}(s, w)$$

[0042] 本实施例通过使用注意力机制层提取多轮交互会话中有价值的特征,以及提取双向LSTM中因随时间传递靠前减弱或消失的内容特征,进而过滤掉无效特征,最终得出有效实体链接。

[0043] 基于知识图谱的中医养生智能多轮交互系统是使用知识图谱节点实体发现和链接等方式将相关信息融合到一起,实现多轮交互。在中医养生咨询的会话过程中,用户咨询的问题都是有关联的、有连续性的,为确保系统回复内容的准确率和丰富性,需要结合交互的历史信息和以及与交互相关的知识图谱进行回复。

[0044] 基于知识图谱的中医养生智能多轮交互系统的建模采用点对点的生成模型完成,先对知识图谱、交互历史、用户当前输入进行建模,再依据建模信息进行解码回复。交互历史与知识图谱和生成的回复并不是一一映射的关系,因此,进行知识图谱与交互历史正确的选择和建模方式的正确选择会对交互系统生成回复的效果产生较大影响。

[0045] 实体查询模块30,可以用于从预存储的中医养生知识图谱中获取与处理后的实体相关的多个三元组。

[0046] 知识图谱存储模块70可以用于存储中医养生知识图谱,中医养生知识图谱包括中医养生本体知识与中医养生自定义词典。

[0047] 中医养生知识图谱的源数据包括中医院病例报告表,中医院病例报告表的内容包括:性别、年龄、职业、民族、婚姻、心率、血压、呼吸、首次发病时间、病程、主诉、临床表现、其他疾病史、简要病史、检查单、病理报告单、舌脉、西医诊断、中医辨证、中医诊断、方药、治疗方案、养护方案、饮食、运动锻炼、生活习惯、按摩推拿。

[0048] 图5是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的中医养生知识图谱的构建流程图。

[0049] 如图5所示,中医养生知识图谱的构建包括如下步骤:

[0050] 步骤S1,对中医院病例报告表的内容进行实体抽取、关系抽取以及属性抽取并得

到实体抽取结果、关系抽取结果以及属性抽取结果；

[0051] 步骤S2,根据实体抽取结果、关系抽取结果以及属性抽取结果进行节点创建和关系创建,并得到中医养生知识图谱。

[0052] 抽取排序模块40,用于筛选需要用户再次输入的临床信息的若干个相关三元组,并将若干三元组按照预先设定的优先级排序,得到三元组优先级序列并确定最高优先级三元组。

[0053] 在智能多轮交互的过程中,会话中会出现指代消解和实体省略的情况,而这种情况会对实体链接造成极大的困扰,因此,本发明针对前后文中识别到的实体进行实体链接,提出了结合前后文的文本信息基于实体链接模块20来解决多轮会话中语句的实体省略和指代消解。

[0054] 先将与前后文中识别到的实体相关的三元组拼接成候选文本,再使用一个全连接层对候选文本编码器输出的候选向量c、前后文编码器输出的前后文向量s、输入文本编码器输出的文本向量w拼接后进行线性变化,最后通过softmax函数计算出输出结果reply,公式如下:

[0055] $\text{reply} = \text{softmax}(F \cdot [c; s; w])$

[0056] 公式中的F为全连接层参数。

[0057] 会话生成模块50可以用于根据最高优先级三元组和以及预存储的中医养生知识图谱生成问题,以对用户进行下一轮提问。

[0058] 图6是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的会话生成模块的工作逻辑图。

[0059] 如图6所示,本实施例提出了句子的文本信息使用一个编码器建模,句子组成的前后文本信息使用另外一个编码器建模,这种基于实体链接模块20的记忆能力生成模型。基于实体链接模块20的会话生成模块50使用的编码器是由一个单向的GRU和一个双向的GRU组成,句子组成的前后文本信息使用的编码器是由一个单向的GRU,句子的文本信息使用的编码器是一个双向的GRU。

[0060] 一个前后文由 $\{s_1, s_2, \dots, s_n\}$ 组成, s_n 是前后文中第n句话, $s_n = \{w_{n1}, w_{n2}, \dots, w_{nn}\}$, w_{nn} 是 s_n 中第n个词, Word Encoder使用的编码器是一个双向的GRU,它将前后文中的每个句子 $\{s_1, s_2, \dots, s_n\}$ 编码成句子向量 $\{\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n\}$, Context Encoder使用的编码器是一个单向的GRU,它将句子向量 $\{\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n\}$ 编码成前后文向量t,计算步骤如下:

[0061] $s_n = [\vec{s}_{nn}; \vec{s}_{1n}] = [GRU(w_{nn}, \vec{s}_{nn-1}); GRU(w_{1n}, \vec{s}_{2n})]$

[0062] $t = s_n = GRU(s_n, s_{n-1})$

[0063] 结果生成模块60可以用于根据用户多次输入的临床信息以及预存储的中医养生知识图谱生成养生建议结果。

[0064] 语料存储模块80可以用于向会话生成模块和结果生成模块提供语料。

[0065] 图7是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的语料存储模块中的语料的生成流程图。

[0066] 如图7存储在语料存储模块80中的语料的生成过程包括如下步骤:

[0067] 步骤A1,分析中医养生知识图谱中的实体和实体关系,得到实体关系图;

- [0068] 步骤A2,将实体关系图转化为有向图;
- [0069] 步骤A3,以实体关系图的每个实体为起点,遍历有向图获取路径图;
- [0070] 步骤A4,对路径图中的路径进行分拆与组合获取子路径;
- [0071] 步骤A5,将中医养生知识图谱中的实体和实体关系对应到各个子路径中,得到语料。
- [0072] 图8是本发明的实施例中基于知识图谱的中医养生智能多轮交互系统的运行逻辑图。
- [0073] 如图8所示,基于知识图谱的中医养生智能交互系统对用户咨询的中医养生问题的当前会话语句和历史会话语句进行命名实体识别后,得到与用户问题相关的实体列表,使用基于记忆序列实体链接模型进行指代消解与实体消歧,并从中医养生知识图谱中提取与实体相关的三元组信息,并将三元组信息传递到会话系统中,使用基于记忆序列建模的生成模型生成传递给用户的回复。
- [0074] 实施例的作用与效果
- [0075] 根据本实施例所涉及的基于知识图谱的中医养生智能多轮交互系统,因为交互系统具备记忆能力,可以主动关联与用户交互的内容已经成为必不可少的要求。用户向系统输入临床表现后,系统会根据用户输入的临床表现的特征,主动向用户提问相关主诉表现,用户根据系统提问的问题,向系统输入相关反馈内容,智能交互系统使用这样的方式,经过与用户多轮交互后,最终给予用户有较高准确率的中医养生建议。而且,本实施例的交互系统由于是采用多轮对话的交互形式,在终端用户的体验上和中医养生咨询问答的结论准确度上均高于普通单轮交互的问答系统。此外,本实施例针对前后文中识别到的实体进行实体链接,提出了结合前后文的文本信息基于记忆序列的实体链接模型来解决多轮会话中语句的实体省略和指代消解。
- [0076] 上述实施方式为本发明的优选案例,并不用来限制本发明的保护范围。

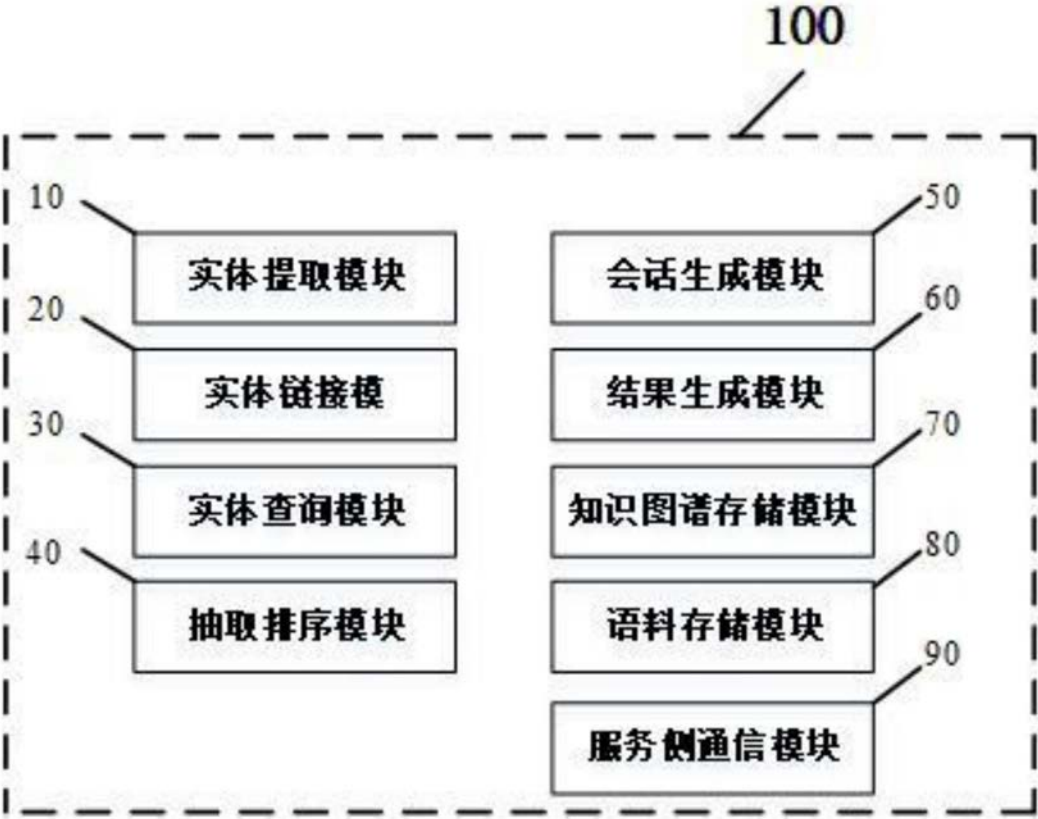


图1

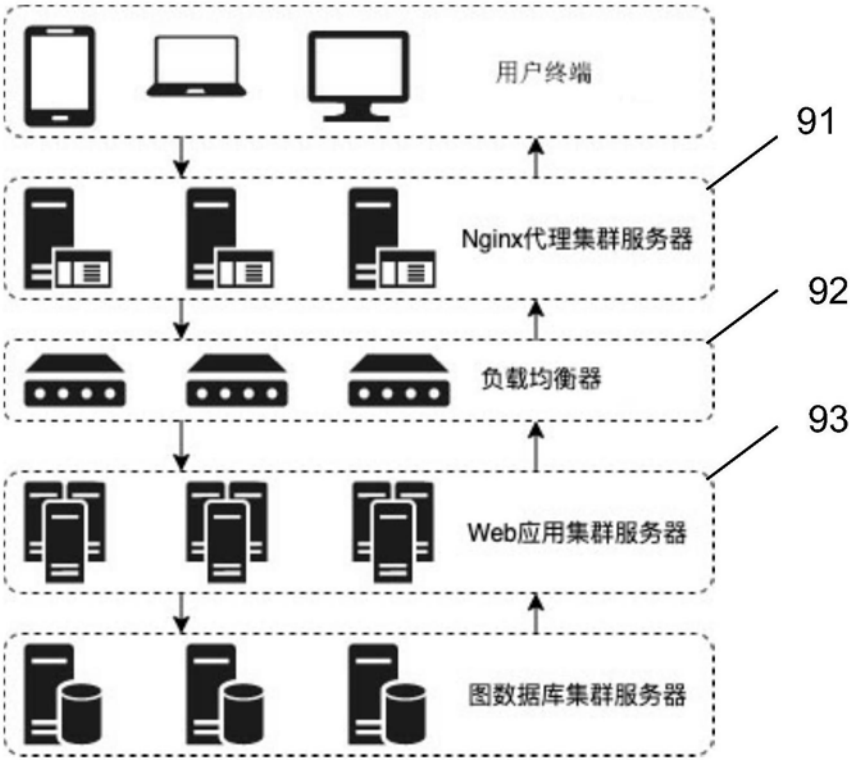


图2

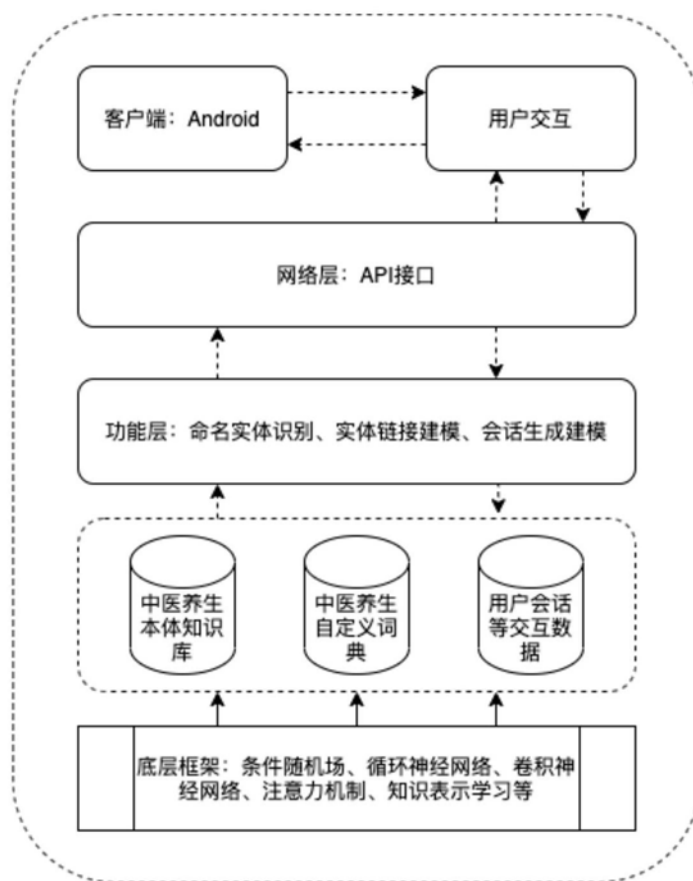


图3

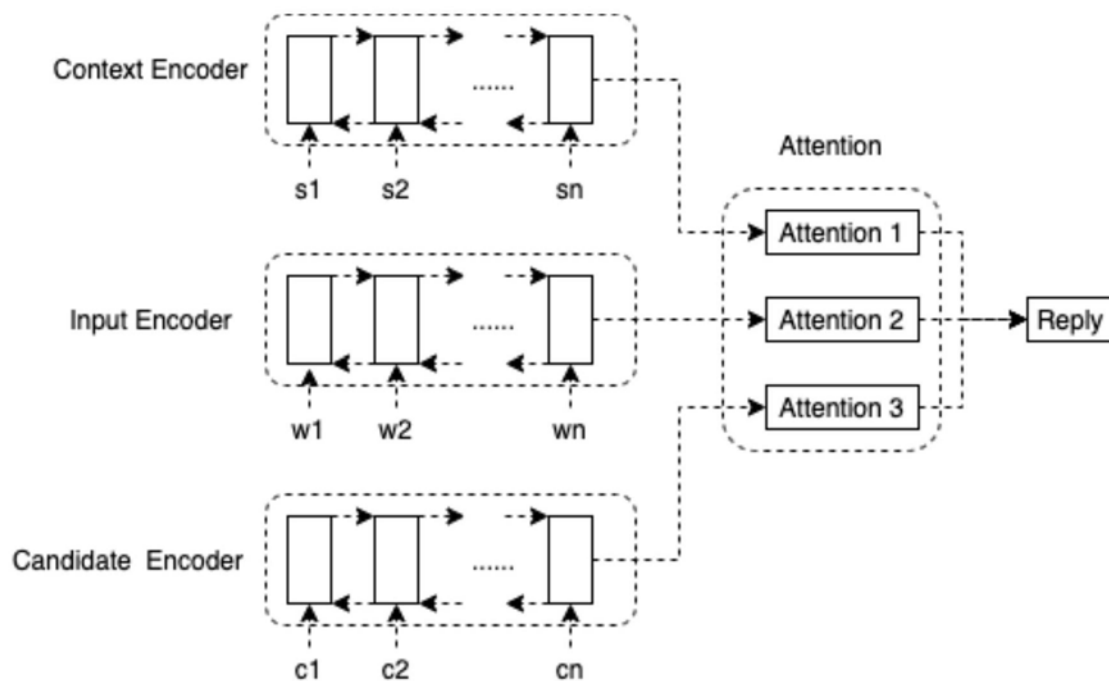


图4

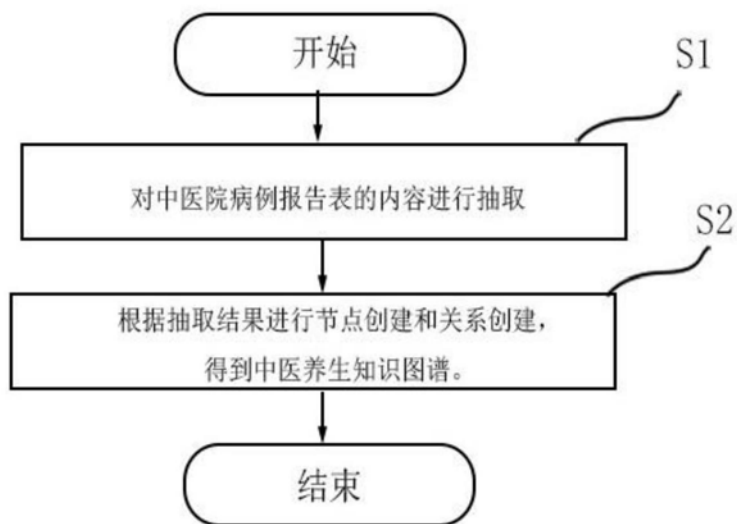


图5

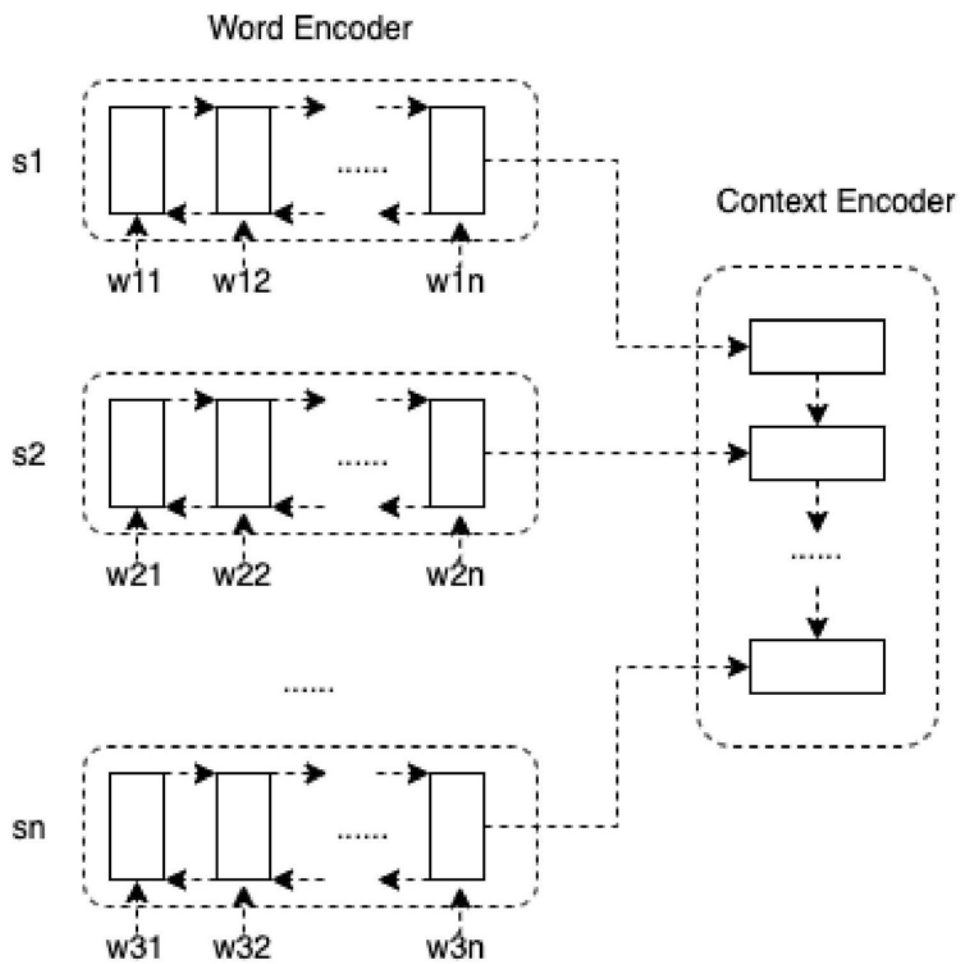


图6

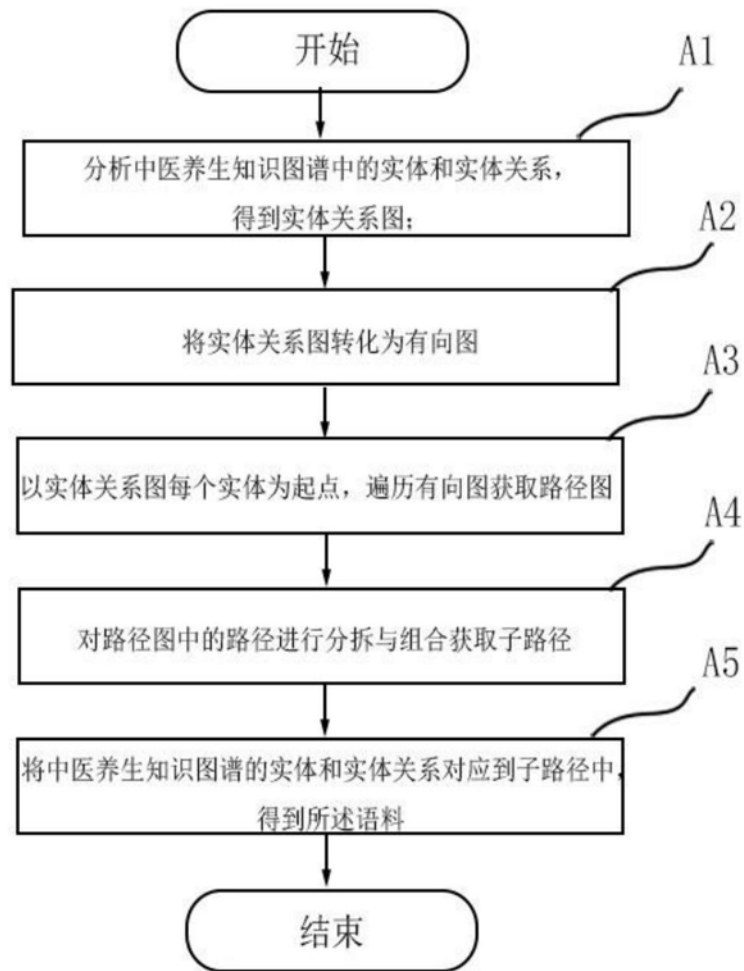


图7

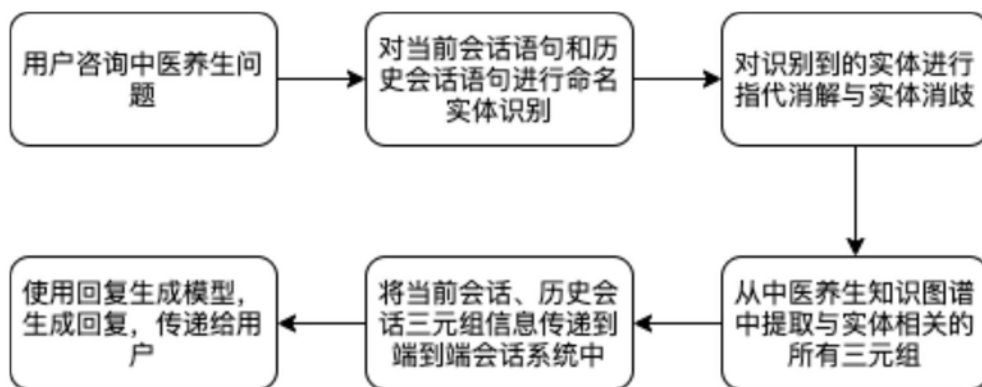


图8