



(12) 发明专利申请

(10) 申请公布号 CN 117633189 A

(43) 申请公布日 2024. 03. 01

(21) 申请号 202311688184.0

(22) 申请日 2023.12.08

(71) 申请人 北京城市网邻信息技术有限公司

地址 100015 北京市朝阳区酒仙桥北路甲  
10号院101号楼1-7层内1层103室

(72) 发明人 请求不公布姓名 请求不公布姓名  
请求不公布姓名

(74) 专利代理机构 北京太合九思知识产权代理  
有限公司 11610

专利代理师 孙明子 刘戈

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/335 (2019.01)

G06N 5/022 (2023.01)

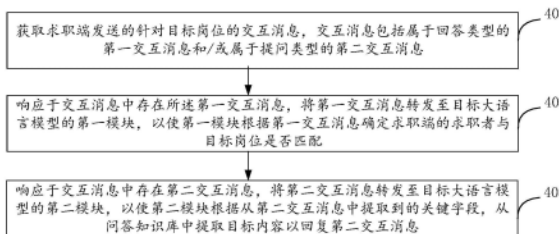
权利要求书2页 说明书11页 附图5页

(54) 发明名称

问答会话处理方法、装置、设备和存储介质

(57) 摘要

本发明实施例提供一种问答会话处理方法、装置、设备和存储介质,应用于服务端,该方法包括:获取求职端发送的针对目标岗位的交互消息,交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息;响应于交互消息中存在第一交互消息,将第一交互消息转发至目标大语言模型的第一模块,以使第一模块根据第一交互消息确定求职端的求职者与目标岗位是否匹配;响应于交互消息中存在第二交互消息,将第二交互消息转发至目标大语言模型的第二模块,以使第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复第二交互消息。本方案通过目标大语言模型实现对求职端的准确回复的和流畅问答。



1. 一种问答会话处理方法,应用于服务端,其特征在于,包括:

获取求职端发送的针对目标岗位的交互消息,所述交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息;

响应于所述交互消息中存在所述第一交互消息,将所述第一交互消息转发至目标大语言模型的第一模块,以使所述第一模块根据所述第一交互消息确定所述求职端的求职者与所述目标岗位是否匹配;

响应于所述交互消息中存在所述第二交互消息,将所述第二交互消息转发至目标大语言模型的第二模块,以使所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。

2. 根据权利要求1所述的方法,其特征在于,所述第一模块根据所述第一交互消息确定所述求职端的求职者与所述目标岗位是否匹配,包括:

所述第一模块从所述第一交互信息中提取与所述目标岗位的至少一个岗位招聘条件相关的求职者信息;

根据所述求职者信息与所述至少一个岗位招聘条件的匹配情况,确定所述求职端的求职者与所述目标岗位是否匹配。

3. 根据权利要求2所述的方法,其特征在于,所述根据所述求职者信息与所述至少一个岗位招聘条件的匹配情况,确定所述求职端的求职者与所述目标岗位是否匹配,包括:

若所述求职者信息与所述至少一个岗位招聘条件均匹配,则确定所述求职端的求职者与所述目标岗位匹配;

响应于确定所述求职端的求职者与所述目标岗位匹配,向所述求职端发送针对所述目标岗位的约面消息,所述约面消息用于确定所述求职者是否参加所述目标岗位的面试;

响应于所述求职端对所述约面消息的确认操作,向所述目标岗位对应的招聘端发送面试通知。

4. 根据权利要求1所述的方法,其特征在于,所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取目标内容,以回复所述第二交互消息,包括:

所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取至少一个目标问答对;其中,每个目标问答对包括目标问题和目标答案,所述至少一个目标问答对中的目标问题与所述关键字段的相似度大于预设相似度阈值;

根据所述至少一个目标问答对中的目标答案和所述目标岗位对应的岗位描述信息,生成所述第二交互消息对应的回复消息。

5. 根据权利要求1至4中任一项所述的方法,其特征在于,所述方法还包括:

确定所述目标岗位的岗位招聘条件中未获取到对应的求职者信息的目标岗位招聘条件,所述求职者信息从所述第一交互信息中获取;

通过所述目标大语言模型的第三模块根据所述目标岗位招聘条件生成属于提问类型的目标交互消息,以提示所述求职端反馈与所述目标岗位招聘条件相关的求职者信息。

6. 根据权利要求1所述的方法,其特征在于,所述服务端配置有分别适用于不同交互领域的多个大语言模型,所述获取求职端发送的针对目标岗位的交互消息之后,所述方法还包括:

根据所述目标岗位和/或所述交互信息,确定当前的目标交互领域;

从所述多个大语言模型中确定出适用交互领域与所述目标交互领域相同的所述目标大语言模型。

7. 根据权利要求6所述的方法,其特征在於,所述方法还包括:

获取所述目标大语言模型在模型测试阶段对应的多个问答测试消息;

根据所述多个问答测试消息分别对应的多个问答准确性评价结果,确定所述目标大语言模型的适用交互领域;和/或,

获取所述目标大语言模型在与所述求职端进行会话的过程中对应的多个问答消息;

根据所述多个问答消息分别对应的多个问答准确性评价结果,更新所述目标大语言模型的适用交互领域。

8. 一种问答会话处理装置,其特征在於,应用于服务端,包括:

获取模块,用于获取求职端发送的针对目标岗位的交互消息,所述交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息;

第一处理模块,用于响应于所述交互消息中存在所述第一交互消息,将所述第一交互消息转发至目标大语言模型的第一模块,以使所述第一模块根据所述第一交互消息确定所述求职端的求职者与所述目标岗位是否匹配;

第二处理模块,用于响应于所述交互消息中存在所述第二交互消息,将所述第二交互消息转发至目标大语言模型的第二模块,以使所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。

9. 一种电子设备,其特征在於,包括:存储器、处理器、通信接口;其中,所述存储器上存储有可执行代码,当所述可执行代码被所述处理器执行时,使所述处理器执行如权利要求1至7中任一项所述的问答会话处理方法。

10. 一种非暂时性机器可读存储介质,其特征在於,所述非暂时性机器可读存储介质上存储有可执行代码,当所述可执行代码被电子设备的处理器执行时,使所述处理器执行如权利要求1至7中任一项所述的问答会话处理方法。

## 问答会话处理方法、装置、设备和存储介质

### 技术领域

[0001] 本发明涉及人工智能技术领域,尤其涉及一种问答会话处理方法、装置、设备和存储介质。

### 背景技术

[0002] 大型语言模型(Large Language Model,LLM)是指使用大量文本数据训练的深度学习模型。大语言模型可以处理多种自然语言任务,比如:理解语言文本的含义,生成自然语言文本,实现问答、对话等,是实现人工智能的关键途径之一。

[0003] 在招聘场景中,为了提高求职端与招聘端之间的会话效率,招聘端通常会配置大语言模型,以通过大语言模型快速响应求职端发送的会话信息。然而,在实际应用中,大语言模型虽然提升了交互时的响应速度,但是回复的准确率确无法得到保障,容易出现答非所问的情况,影响用户体验。

### 发明内容

[0004] 本发明实施例提供一种问答会话处理方法、装置、设备和存储介质,用以提升服务端与求职端人机交互时,问答会话的准确性和流畅性。

[0005] 第一方面,本发明实施例提供一种问答会话处理方法,应用于服务端,所述方法包括:

[0006] 获取求职端发送的针对目标岗位的交互消息,所述交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息;

[0007] 响应于所述交互消息中存在所述第一交互消息,将所述第一交互消息转发至目标大语言模型的第一模块,以使所述第一模块根据所述第一交互消息确定所述求职端的求职者与所述目标岗位是否匹配;

[0008] 响应于所述交互消息中存在所述第二交互消息,将所述第二交互消息转发至目标大语言模型的第二模块,以使所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。

[0009] 第二方面,本发明实施例提供一种问答会话处理装置,应用于服务端,所述装置包括:

[0010] 获取模块,用于获取求职端发送的针对目标岗位的交互消息,所述交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息;

[0011] 第一处理模块,用于响应于所述交互消息中存在所述第一交互消息,将所述第一交互消息转发至目标大语言模型的第一模块,以使所述第一模块根据所述第一交互消息确定所述求职端的求职者与所述目标岗位是否匹配;

[0012] 第二处理模块,用于响应于所述交互消息中存在所述第二交互消息,将所述第二交互消息转发至目标大语言模型的第二模块,以使所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。

[0013] 第三方面,本发明实施例提供一种电子设备,包括:存储器、处理器、通信接口;其中,所述存储器上存储有可执行代码,当所述可执行代码被所述处理器执行时,使所述处理器至少可以实现如第一方面所述的问答会话处理方法。

[0014] 第四方面,本发明实施例提供了一种非暂时性机器可读存储介质,所述非暂时性机器可读存储介质上存储有可执行代码,当所述可执行代码被电子设备的处理器执行时,使所述处理器至少可以实现如第一方面所述的问答会话处理方法。

[0015] 在本发明实施例提供的方案中,首先,获取求职端发送的针对目标岗位的交互消息。其中按照信息类型的不同,交互消息可以包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息。之后,响应于交互消息中存在第一交互消息,将第一交互消息转发至目标大语言模型的第一模块,以通过第一模块根据第一交互消息确定求职端的求职者与目标岗位是否匹配;响应于交互消息中存在第二交互消息,将第二交互消息转发至目标大语言模型的第二模块,以通过第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。本方案通过在目标大语言模型中配置分别用于处理不同类型的交互消息的不同功能模块(第一模块和第二模块),实现对求职端的准确回复的和流畅问答。

## 附图说明

[0016] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图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的前提下,还可以根据这些附图获得其他的附图。

[0017] 图1为本发明实施例提供的一种问答会话处理方法的硬件执行环境的示意图;

[0018] 图2为本发明实施例提供的一种问答会话处理方法的云计算环境的示意图;

[0019] 图3为本发明实施例提供的一种问答会话处理方法的应用示意图;

[0020] 图4为本发明实施例提供的一种问答会话处理方法的流程图;

[0021] 图5为本发明实施例提供的另一种问答会话处理方法的流程图;

[0022] 图6为本发明实施例提供的又一种问答会话处理方法的流程图;

[0023] 图7为本发明实施例提供的一种问答会话处理装置的结构示意图;

[0024] 图8为与图7所示实施例提供的问答会话处理装置对应的电子设备的结构示意图。

## 具体实施方式

[0025] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0026] 需要说明的是,本发明实施例中所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据,并且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准,并提供有相应的操作入口,供用户选择授

权或者拒绝。

[0027] 下面结合附图对本发明的一些实施方式作详细说明。在各实施例之间不冲突的情况下,下述的实施例及实施例中的特征可以相互组合。另外,下述各方法实施例中的步骤时序仅为一种举例,而非严格限定。

[0028] 大型语言模型(Large Language Model,LLM)是指使用大量文本数据训练的深度学习模型。大语言模型可以处理多种自然语言任务,比如:理解语言文本的含义、生成自然语言文本、实现问答、对话等,是实现人工智能的关键途径之一。

[0029] 在购物、招聘等应用场景中,为提高会话效率,商家或招聘方通常在服务端配置有大语言模型,也即常说的客服机器人,通过大语言模型理解客户端输入的需求,可以快速回答客户端的问题。但是,在实际的使用过程中,存在大语言模型对问题回复不准确、答非所问的情况,影响用户体验。

[0030] 本实施例中,为了便于理解,以招聘场景为例,提出了一种问答会话处理方法,该方法应用于服务端,包括:首先,获取求职端发送的针对目标岗位的交互消息。其中按照信息类型的不同,交互消息可以包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息。之后,响应于交互消息中存在第一交互消息,将第一交互消息转发至目标大语言模型的第一模块,以通过第一模块根据第一交互消息确定求职端的求职者与目标岗位是否匹配;响应于交互消息中存在第二交互消息,将第二交互消息转发至目标大语言模型的第二模块,以通过第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。本方案中,将目标大语言模型的能力进行模块化,通过在目标大语言模型中分别配置用于处理会回答类型交互消息的第一模块,以及用于处理会问题类型交互消息的第二模块,实现对求职端的准确回复的和流畅问答。

[0031] 下面对本发明实施例提供的问答会话处理方法进行介绍说明。

[0032] 图1为本发明实施例提供的一种问答会话处理方法的硬件执行环境的示意图,如图1所示,该问答会话处理方法的硬件执行环境可以由求职端设备101和服务端设备102构成,求职端设备101与服务端设备102通信连接。其中,求职端设备101可以是某一种终端设备,比如智能手机、平板电脑、PC机等。服务端设备102可以是内容提供商的服务器,也可以是云服务商的云服务器。

[0033] 在实际的招聘场景中,服务端设备102还与招聘端设备(图1中未示意出)通信连接。服务端作为招聘服务提供方,服务端设备102上部署有招聘相关的服务,求职端设备101和招聘端设备通过部署与服务端设备102上部署的招聘相关的服务对应的应用程序,实现对该招聘相关的服务的使用。例如:求职端设备101通过应用程序向招聘端投递求职简历、招聘端设备通过应用程序向求职端发送岗位邀请等。

[0034] 本实施例中,服务端中部署有大语言模型,用于代替招聘端与求职端进行问答会话。

[0035] 实际应用中,上述服务端设备102可以是内容提供商维护的一个独立的物理服务器或物理服务器集群,也可以是云服务商维护的云服务器一称为计算节点。图2为本发明实施例提供的一种问答会话处理方法的云计算环境的示意图,在如图2所示的云计算环境中,可以包括分布式部署的若干(图2中示意的201-1,201-2,...)计算节点(云服务器),每个计算节点中都具有计算、存储等处理资源。在云计算环境中,可以组织由多个计算节点来提供

某种服务,当然,一个计算节点也可以提供一种或多种服务,比如图2中示意的服务A、服务B、服务C、服务D。云计算环境中提供该服务的方式可以是对外提供服务接口,求职端设备调用该服务接口以使用相应的服务。服务接口包括软件开发工具包(Software Development Kit,简称SDK)、应用程序接口(Application Programming Interface,简称API)等形式。

[0036] 上述服务是根据云计算环境所支持的各种虚拟化技术来部署的,比如基于虚拟机、容器的虚拟化技术。以基于容器的虚拟化技术为例,一个服务对应的若干容器可以被组装成一个容器组(pod)。比如图2中示意的服务B可以配置有一个或多个pod,每个pod内可以包括代理以及一个或多个容器。pod中的一个或多个容器用于处理与服务的一个或多个相应功能相关的请求,pod中的代理用于控制与服务相关的网络功能,比如路由、负载均衡等。

[0037] 在操作过程中,执行来自求职端设备的请求,可能需要调用云计算环境中的一个或多个服务,执行一个服务的一个或多个功能可能需要调用另一个服务的一个或多个功能。如图2所示,服务A接收到求职端设备发出的请求后,可以调用服务B,服务B可以请求服务D执行一个或多个功能。

[0038] 在上述云计算环境下,本发明实施例提供了如图3所示的一种问答会话处理方法的应用示意图。

[0039] 在图3中,云计算环境中提供有问答会话处理服务以及相应的服务接口,求职端设备调用该服务接口,以向该问答会话处理服务触发请求,该请求中包括求职端发送的针对目标岗位的交互消息,云计算环境响应于该请求,确定响应该请求的提供问答会话处理服务的计算节点,利用该计算节点中的处理资源来确定用于响应该交互消息的目标大语言模型,并通过该目标大语言模型中的对应模块处理该交互消息,以将处理结果反馈给求职端设备进行显示。

[0040] 下面结合图4详细介绍本发明实施例提供的问答会话处理方法的执行过程。该问答会话处理方法由上述服务端设备来执行,比如:由上述云计算环境中的计算节点来执行。

[0041] 图4为本发明实施例提供的一种问答会话处理方法的流程图,该方法应用于服务端,如图4所示,可以包括如下步骤:

[0042] 401、获取求职端发送的针对目标岗位的交互消息,交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息。

[0043] 402、响应于交互消息中存在所述第一交互消息,将第一交互消息转发至目标大语言模型的第一模块,以使第一模块根据第一交互消息确定求职端的求职者与目标岗位是否匹配。

[0044] 403、响应于交互消息中存在第二交互消息,将第二交互消息转发至目标大语言模型的第二模块,以使第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复第二交互消息。

[0045] 如图3所示,本实施例中的问答会话处理方法,将求职端发送的针对目标岗位的交互消息按照消息类型划分为了两种类型,分别为回答类型和问题类型。为便于区分,将属于回答类型的交互消息称为第一交互消息,将属于问题类型的交互消息称为第二交互消息。

[0046] 服务端对应的目标大语言模型中设置有第一模块和第二模块,第一模块被配置为处理第一交互消息,即根据第一交互消息确定求职端的求职者与目标岗位是否匹配;第二模块被配置为处理第二交互消息,即根据从第二交互消息中提取到的关键字段,从问答知

识库中提取目标内容以回复第二交互消息。

[0047] 本实施例中,可选地,目标大语言模型可以直接部署于服务端,也可以部署于其他独立于服务端但服务端能与之通信的位置,本实施例不对目标大语言模型的部署位置进行限制。

[0048] 在获取求职端发送的针对目标岗位的交互消息后,确定交互消息包含的消息类型。具体地,若该交互消息为回答消息,则仅包含属于回答类型第一交互消息;若该交互消息为问题消息,则仅包含属于问题类型第二交互消息;若该交互消息为问答消息,则该问答消息中的回答部分属于回答类型,该问答消息中的问题部分属于问题类型,该交互消息包含属于回答类型第一交互消息和属于问题类型第二交互消息。

[0049] 若交互消息中存在第一交互消息,则将第一交互消息转发至目标大语言模型的第一模块,以通过第一模块根据第一交互消息确定求职端的求职者与目标岗位是否匹配。

[0050] 作为一种可选地确定求职端的求职者与目标岗位是否匹配的方式,首先,通过第一模块从第一交互信息中提取与目标岗位的至少一个岗位招聘条件相关的求职者信息;之后,根据求职者信息与至少一个岗位招聘条件的匹配情况,确定求职端的求职者与目标岗位是否匹配。

[0051] 其中,目标岗位的至少一个岗位招聘条件,也即目标岗位的招聘要求。例如,目标岗位的招聘要求包括:年龄45岁以下、至少5年从业经验、具有高级技工证书等等。求职者信息,也即与目标岗位的招聘要求相对应的求职者个人信息,比如:求职者的年龄为35岁、从业6年、具有高级技工证书等等。

[0052] 具体实施过程中,若求职者信息与至少一个岗位招聘条件均匹配,则确定求职端的求职者与目标岗位匹配。举例来说,假设某一岗位的至少一个岗位招聘条件为:求职者的年龄在45岁以下、工龄为5年,若求职者信息对应表明求职者为38岁,工龄为6年,则认为求职者信息与至少一个岗位招聘条件均匹配,确定求职端的求职者与该岗位匹配。响应于确定求职端的求职者与目标岗位匹配,向求职端发送针对目标岗位的约面消息。其中,该约面消息用于确定求职者是否参加目标岗位的面试,该面试由求职端的求职者和招聘端的招聘者通过线上或线下的方式完成。响应于求职端对约面消息的确认操作,向目标岗位对应的招聘端发送面试通知,以使招聘端的招聘者对求职者在约定时间、约定地点进行面试。

[0053] 若求职者信息与至少一个岗位招聘条件存在一项不匹配,则确定求职端的求职者与目标岗位不匹配。响应于确定求职端的求职者与目标岗位不匹配,向求职端发送针对目标岗位的结束消息。其中,该结束消息用于告知求职者其自身条件与目标岗位的招聘要求不匹配,不能进行进一步的面试。

[0054] 若交互消息中存在第二交互消息,则将第二交互消息转发至目标大语言模型的第二模块,以通过第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复第二交互消息。

[0055] 其中,问答知识库中包括目标大语言模型在求职招聘场景中对应的多个问答对。该多个问答对包括:目标大语言模型在训练过程中使用的训练样本对应的问答对,以及目标大语言模型在使用过程中新产生的问答会话对应的问答对。其中,训练样本对应的问答对又包括:在对目标大语言模型进行与目标岗位无关的问答训练时使用的的第一训练样本对应的的第一问答对,以及在目标对大语言模型进行与目标岗位相关的问答训练时使用的第二

训练样本对应的第二问答对。

[0056] 作为一种可选地生成第二交互消息对应的回复消息的方式,可以通过第二模块先从第二交互消息中提取到的关键字段;然后,根据提取到的关键字段,从问答知识库中提取至少一个目标问答对;其中,每个目标问答对包括目标问题和目标答案,至少一个目标问答对中的目标问题与关键字段的相似度大于预设相似度阈值;之后,根据至少一个目标问答对中的目标答案和目标岗位对应的岗位描述信息,生成第二交互消息对应的回复消息。

[0057] 由于第二问答对的数量小于第一问答对、目标大语言模型在使用过程中新产生的问答会话对应的问答对对应于目标大语言模型的近期应用场景,因此,可选地,从问答知识库中提取至少一个目标问答对包括:从问答知识库的第二问答对或目标大语言模型在使用过程中新产生的问答会话对应的问答对中查询与提取到的关键字段匹配的至少一个目标问答对;若第二问答对和目标大语言模型在使用过程中新产生的问答会话对应的问答对中不存在与提取到的关键字段匹配的至少一个目标问答对,则从问答知识库的第一问答对中查询与提取到的关键字段匹配的至少一个目标问答对。

[0058] 目标岗位对应的岗位描述信息,可以理解为目标岗位对应的先验知识,比如:目标岗位涉及概念的解释信息等。可选地,可在目标大语言模型中预先配置不同岗位分别对应的岗位描述信息,以便在问答会话时使用。岗位描述信息用于辅助目标大语言模型的第二模块更好的理解第二交互信息,生成更准确的第二交互消息对应的回复消息。

[0059] 本实施例中,服务端在于求职端进行问答会话时,首先,服务端获取求职端发送的针对目标岗位的交互消息。其中按照信息类型的不同,交互消息可以包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息。之后,响应于交互消息中存在第一交互消息,将第一交互消息转发至目标大语言模型的第一模块,以通过第一模块根据第一交互消息确定求职端的求职者与目标岗位是否匹配;响应于交互消息中存在第二交互消息,将第二交互消息转发至目标大语言模型的第二模块,以通过第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。通过将目标大语言模型的能力进行模块化,即在目标大语言模型中分别配置用于处理会回答类型交互消息的第一模块,以及用于处理会问题类型交互消息的第二模块,实现对求职端发送的交互信息进行按照消息类型的区分处理,以提高问答会话的准确回复的和流畅问答。

[0060] 在另一实施例中,服务端还具备提问能力,该提问能力可通过目标大语言模型的第三模块实现。以下结合具体实施例进行说明。

[0061] 图5为本发明实施例提供的另一种问答会话处理方法的流程图,该方法应用于服务端,如图5所示,可以包括如下步骤:

[0062] 501、获取求职端发送的针对目标岗位的交互消息,交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息。

[0063] 502、响应于交互消息中存在所述第一交互消息,将第一交互消息转发至目标大语言模型的第一模块,以使第一模块从第一交互信息中提取与目标岗位的至少一个岗位招聘条件相关的求职者信息。

[0064] 503、确定目标岗位的岗位招聘条件中未获取到对应的求职者信息的目标岗位招聘条件。

[0065] 504、通过目标大语言模型的第三模块根据目标岗位招聘条件生成属于提问类型

的目标交互消息,以提示求职端反馈与目标岗位招聘条件相关的求职者信息,以使第一模块根据求职者信息和至少一个岗位招聘条件确定求职端的求职者与目标岗位是否匹配。

[0066] 505、响应于交互消息中存在第二交互消息,将第二交互消息转发至目标大语言模型的第二模块,以使第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复第二交互消息。

[0067] 其中,步骤501、步骤502和步骤503的具体实施过程可参考前述实施例,本实施例中不再进行赘述。

[0068] 可以理解的是,在与求职端进行交互时,目标大语言模型的第一模块需要获取从第一交互信息中提取与目标岗位的至少一个岗位招聘条件相关的求职者信息,以进一步确定求职端的求职者与目标岗位是否匹配。如果能够主动生成相应的与目标岗位的至少一个岗位招聘条件相关的提问问题并反馈给求职端,有利于提高会话效率。

[0069] 因此,本实施例中,服务端可以在会话过程中,不断地确定目标岗位的岗位招聘条件中是否存在未获取到对应的求职者信息的目标岗位招聘条件,比如:每隔一段时间确认一次,或者每获取到求职端发送的针对目标岗位的交互消息后确认一次等。

[0070] 若目标岗位的岗位招聘条件中存在未获取到对应的求职者信息的目标岗位招聘条件,则通过目标大语言模型的第三模块根据目标岗位招聘条件生成属于提问类型的目标交互消息,并发送给求职端,以提示求职端反馈与目标岗位招聘条件相关的求职者信息。

[0071] 之后,目标大语言模型的第一模块根据该目标交互消息对应的第一交互消息确定出新的求职者信息,并将求职者对应的全部求职者信息和至少一个岗位招聘条件确定求职端的求职者与所述目标岗位是否匹配。其中,全部求职者信息可能从不同的第一交互消息中提取得到。

[0072] 本实施例中,通过目标大语言模型的第三模块主动生成与目标岗位的至少一个岗位招聘条件相关的目标交互消息的方式,提升了目标大语言模型的灵活性,使得服务端在招聘场景中的问答会话效率更高。

[0073] 图6为本发明实施例提供的又一种问答会话处理方法的流程图,该方法应用于服务端,如图6所示,可以包括如下步骤:

[0074] 601、获取求职端发送的针对目标岗位的交互消息,交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息。

[0075] 602、根据目标岗位和/或交互信息,确定当前的目标交互领域。

[0076] 603、从服务端配置的分别适用于不同交互领域的多个大语言模型中确定出适用交互领域与目标交互领域相同的目标大语言模型。

[0077] 604、响应于交互消息中存在所述第一交互消息,将第一交互消息转发至目标大语言模型的第一模块,以使第一模块根据第一交互消息确定求职端的求职者与目标岗位是否匹配。

[0078] 605、响应于交互消息中存在第二交互消息,将第二交互消息转发至目标大语言模型的第二模块,以使第二模块根据从第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复第二交互消息。

[0079] 其中,步骤601、步骤604和步骤605的具体实施过程可参考前述实施例,本实施例中不再进行赘述。

[0080] 本实施例中,为了进一步提升服务端与求职端问答会话的准确度和流畅度,本实施例中服务端所使用的目标大语言模型适用的交互领域与当前会话对应的目标交互领域相同。

[0081] 其中,交互领域,可以理解为,与当前求职招聘场景中的目标岗位关联的领域等。不同的交互领域划分规则,对应的交互领域可能存在差别。例如,当按照岗位所属行业划分交互领域时,交互领域可以被划分为:家政领域、货车司机领域、租房领域、搬家领域等。当按照会话中交互信息涉及的具体语义内容所属领域划分交互领域时,交互领域可以被划分为:岗位内容咨询领域、岗位相关政策咨询领域、岗位薪酬待遇咨询领域等。

[0082] 可以理解的是,不同的交互领域有其领域的特异性,同一大语言模型在不同的交互领域的表现能力也有所不同,采用单个大语言模型响应所有交互领域的交互信息,无法保证会话消息的准确度。

[0083] 为解决这一问题,本实施例中,在获取求职端发送的针对目标岗位的交互消息之后,先根据目标岗位和/或交互信息,确定当前的目标交互领域;然后,从服务端配置的分别适用于不同交互领域的多个大语言模型中确定出适用交互领域与目标交互领域相同的目标大语言模型。

[0084] 作为一种可选地确定目标交互领域的方式,可以预先建立多种不同交互领域与其对应的描述词之间的对应关系。在获取到求职端发送的针对目标岗位的交互消息之后,对该目标岗位和/或交互消息进行关键词提取;然后,通过计算提取到的关键词与各不同交互领域对应的描述词之间的相似度,确定目标岗位和/或交互消息对应的目标交互领域。

[0085] 例如,可以将多种不同提问领域对应的描述词以及从目标岗位和/或交互消息中提取的关键词进行向量化,然后计算描述词对应向量与关键词对应向量之间的余弦相似度,若关键词对应向量与某一交互领域x的描述词对应向量之间的余弦相似度大于设定相似度阈值,则确定目标岗位和/或交互消息对应的目标交互领域为交互领域x。

[0086] 作为另一种可选地确定目标交互领域的方式,还可以直接根据目标岗位所属领域,确定当前对应的目标交互领域。例如,若目标岗位为保姆,则可确定当前对应的目标交互领域为家政领域。

[0087] 之后,确定服务端配置的分别适用于不同交互领域的多个大语言模型中是否存在目标交互领域对应的目标大语言模型。如果不存在目标交互领域对应的目标大语言模型,则通过非大语言模型的方式进行会话,比如输出相应的提示信息,以提示通过人工的方式进行会话等。如果存在目标交互领域对应的目标大语言模型,则确定通过目标大语言模型进行会话。

[0088] 本实施例中,在服务端配置分别适用于不同交互领域的多个大语言模型,且多个大语言模型与交互领域之间存在对应关系。实际应用中,无论是按照岗位所属行业划分交互领域,还是按照会话涉及的具体语义内容所属领域划分交互领域,通过与当前会话对应的目标交互领域匹配的目标大语言模型进行会话交互,都有利于提高服务端在实际过程中进行问答会话的准确性。

[0089] 实际应用中,为了让大语言模型在招聘这一应用场景中具有较优的模型性能,在大语言模型使用之前,会先获取岗位关联领域的先验知识信息,比如:不同领域的岗位的工作内容、招聘要求等;然后,根据先验知识信息对大语言模型进行微调训练,使大语言模型

具备在与岗位关联的多个不同领域进行会话的能力。但是,实际应用中,大语言模型具备在某一领域进行会话能力,并不意味着其擅长处理该领域的会话,因此,确认大语言模型适用的会话领域(也即交互领域)对于提高大语言模型会话的准确度尤为重要。

[0090] 可选地,本实施例中,通过增加“前置目标大语言模型适用交互领域确认”和“后置目标大语言模型适用交互领域更新”的方式,进一步提升大语言模型在问答会话过程中,会话的准确度和流畅度。

[0091] 前置大语言模型适用交互领域确认阶段,可选地,可以先获取目标大语言模型在模型测试阶段对应的多个问答测试消息;然后,根据多个问答测试消息分别对应的多个问答准确性评价结果,确定目标大语言模型的适用交互领域。例如,若多个问答测试消息对应的多个问答准确性评价结果均表明目标大语言模型的问答准确,则确定该多个问答测试消息对应的交互领域为目标大语言模型适用的交互领域。

[0092] 可选地,可以通过例如人工标注等方式确定问答准确性评价结果。

[0093] 在后置大语言模型适用交互领域更新阶段,可选地,可以先获取目标大语言模型在与求职端进行会话的过程中对应的多个问答消息;然后,根据多个问答消息分别对应的多个问答准确性评价结果,更新目标大语言模型的适用交互领域。

[0094] 例如,当多个问答消息分别对应的多个问答准确性评价结果中存在问答不准确的评价结果时,表明目标大语言模型在本次问答会话中的表现能力一般。可选地,可以从目标大语言模型的适用交互领域中剔除目标交互领域。当多个问答消息分别对应的多个问答准确性评价结果均表明目标大语言模型问答准确时,表明大语言模型在本次问答会话中的表现能力正常,仍将目标交互领域作为大语言模型的适用交互领域。由此,通过及时更新目标大语言模型适用交互领域的方式,能够保证目标大语言模型在会话交互过程中的准确性和流畅性。

[0095] 以下将详细描述本发明的一个或多个实施例的问答会话处理装置。本领域技术人员可以理解,这些装置均可使用市售的硬件组件通过本方案所教导的步骤进行配置来构成。

[0096] 图7为本发明实施例提供的一种问答会话处理装置的结构示意图,应用于服务端,如图7所示,该装置包括:获取模块11、第一处理模块12、第二处理模块13。

[0097] 获取模块11,用于获取求职端发送的针对目标岗位的交互消息,所述交互消息包括属于回答类型的第一交互消息和/或属于提问类型的第二交互消息。

[0098] 第一处理模块12,用于响应于所述交互消息中存在所述第一交互消息,将所述第一交互消息转发至目标大语言模型的第一模块,以使所述第一模块根据所述第一交互消息确定所述求职端的求职者与所述目标岗位是否匹配。

[0099] 第二处理模块13,用于响应于所述交互消息中存在所述第二交互消息,将所述第二交互消息转发至目标大语言模型的第二模块,以使所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取目标内容以回复所述第二交互消息。

[0100] 可选地,所述第一处理模块12,具体用于所述第一模块从所述第一交互信息中提取与所述目标岗位的至少一个岗位招聘条件相关的求职者信息;根据所述求职者信息与所述至少一个岗位招聘条件的匹配情况,确定所述求职端的求职者与所述目标岗位是否匹配。

[0101] 可选地,所述第一处理模块12,具体用于若所述求职者信息与所述至少一个岗位招聘条件均匹配,则确定所述求职端的求职者与所述目标岗位匹配;响应于确定所述求职端的求职者与所述目标岗位匹配,向所述求职端发送针对所述目标岗位的约面消息,所述约面消息用于确定所述求职者是否参加所述目标岗位的面试;响应于所述求职端对所述约面消息的确认操作,向所述目标岗位对应的招聘端发送面试通知。

[0102] 可选地,所述第二处理模块13,具体用于所述第二模块根据从所述第二交互消息中提取到的关键字段,从问答知识库中提取至少一个目标问答对;其中,每个目标问答对包括目标问题和目标答案,所述至少一个目标问答对中的目标问题与所述关键字段的相似度大于预设相似度阈值;根据所述至少一个目标问答对中的目标答案和所述目标岗位对应的岗位描述信息,生成所述第二交互消息对应的回复消息。

[0103] 可选地,所述第一处理模块12,还用于确定所述目标岗位的岗位招聘条件中未获取到对应的求职者信息的目标岗位招聘条件,所述求职者信息从所述第一交互信息中获取;通过所述目标大语言模型的第三模块根据所述目标岗位招聘条件生成属于提问类型的目标交互消息,以提示所述求职端反馈与所述目标岗位招聘条件相关的求职者信息。

[0104] 可选地,所述服务端配置有分别适用于不同交互领域的多个大语言模型,所述问答会话处理装置还包括:确定模块。所述确定模块,用于根据所述目标岗位和/或所述交互信息,确定当前的目标交互领域;从所述多个大语言模型中确定出适用交互领域与所述目标交互领域相同的所述目标大语言模型

[0105] 可选地,所述确定模块还用于获取所述目标大语言模型在模型测试阶段对应的多个问答测试消息;根据所述多个问答测试消息分别对应的多个问答准确性评价结果,确定所述目标大语言模型的适用交互领域;和/或,获取所述目标大语言模型在与所述求职端进行会话的过程中对应的多个问答消息;根据所述多个问答消息分别对应的多个问答准确性评价结果,更新所述目标大语言模型的适用交互领域。

[0106] 图7所示装置可以执行前述实施例介绍的步骤,详细的执行过程和技术效果参见前述实施例中的描述,在此不再赘述。

[0107] 在一个可能的设计中,上述图7所示问答会话处理装置的结构可实现为一电子设备,如图8所示,该电子设备可以包括:存储器21、处理器22、通信接口23。其中,存储器21上存储有可执行代码,当所述可执行代码被处理器22执行时,使处理器22至少可以实现如前述实施例中提供的问答会话处理方法。

[0108] 另外,本发明实施例提供了一种非暂时性机器可读存储介质,所述非暂时性机器可读存储介质上存储有可执行代码,当所述可执行代码被电子设备的处理器执行时,使所述处理器至少可以实现如前述实施例中提供的问答会话处理方法。

[0109] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0110] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助加必需的通用硬件平台的方式来实现,当然也可以通过硬件和软件结合的方式来实现。基于这样的理解,上述技术方案本质上或者说对现有技术做出贡献的部分可以以计算

机产品的形式体现出来,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0111] 最后应说明的是:以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

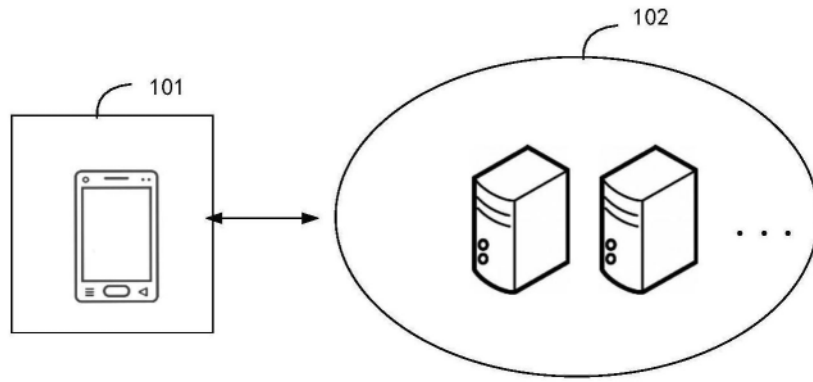


图1

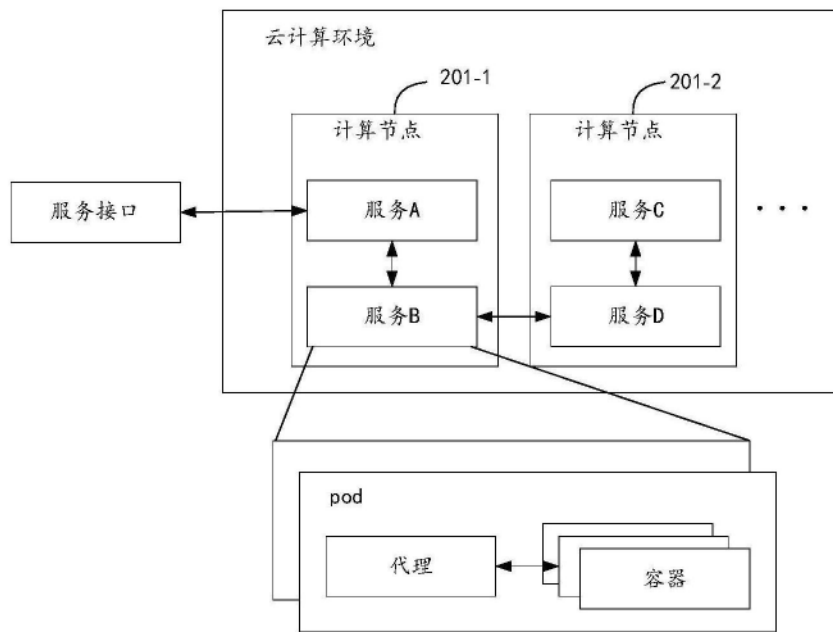


图2

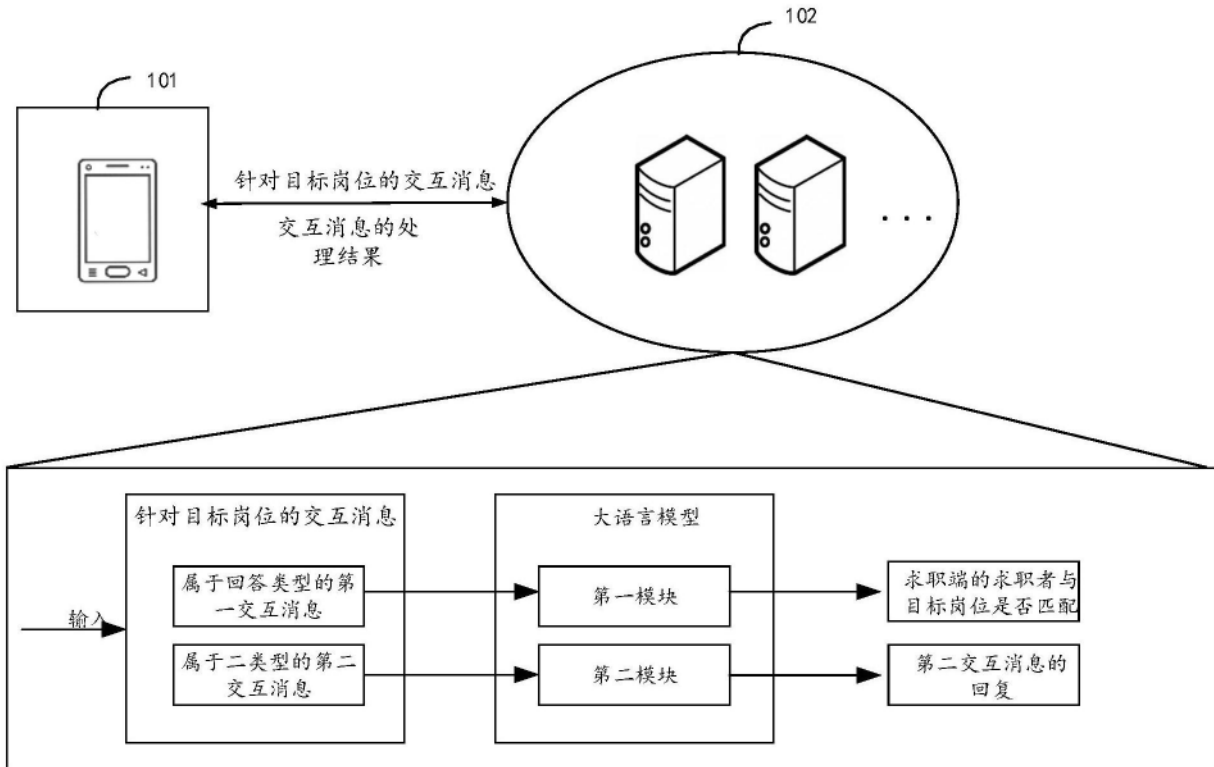


图3

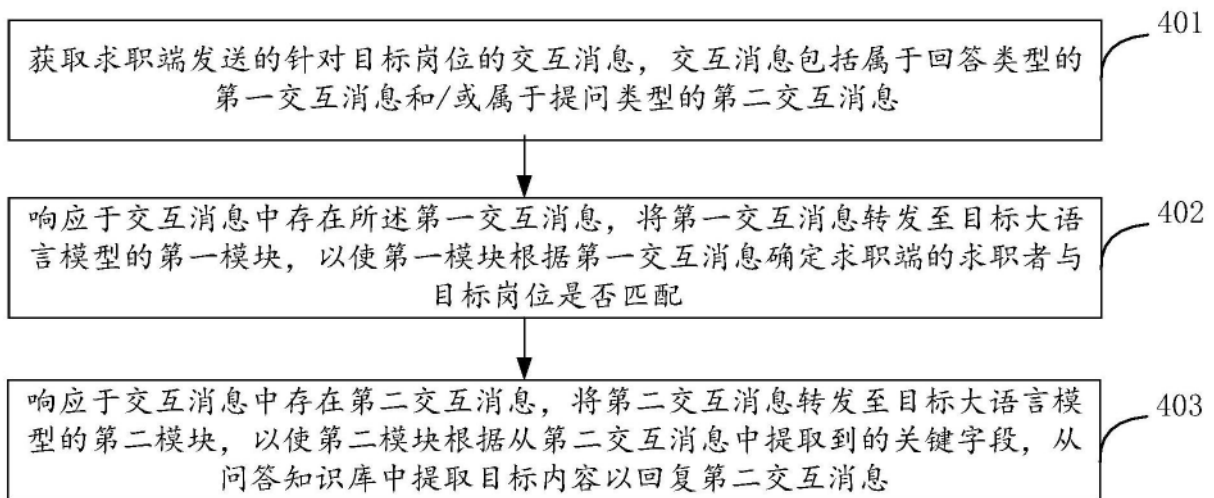


图4

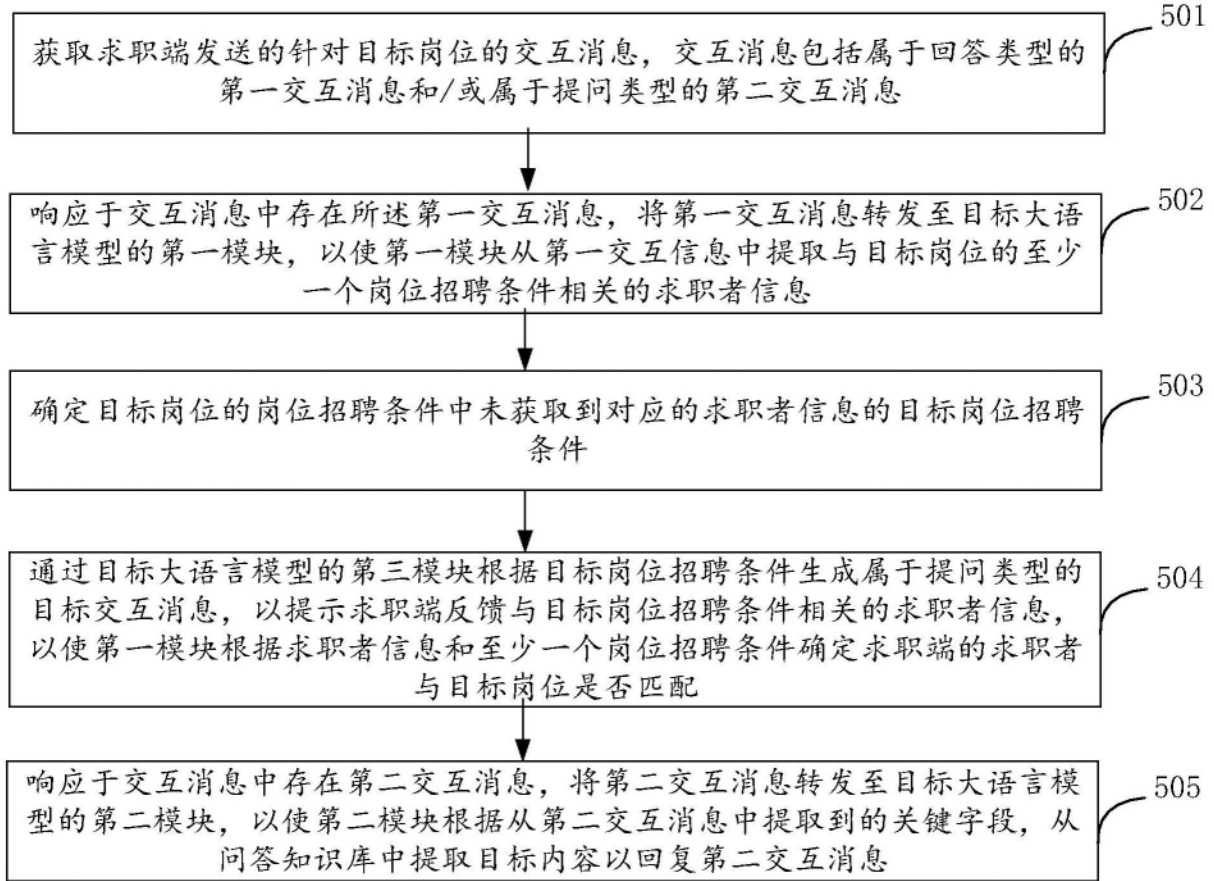


图5

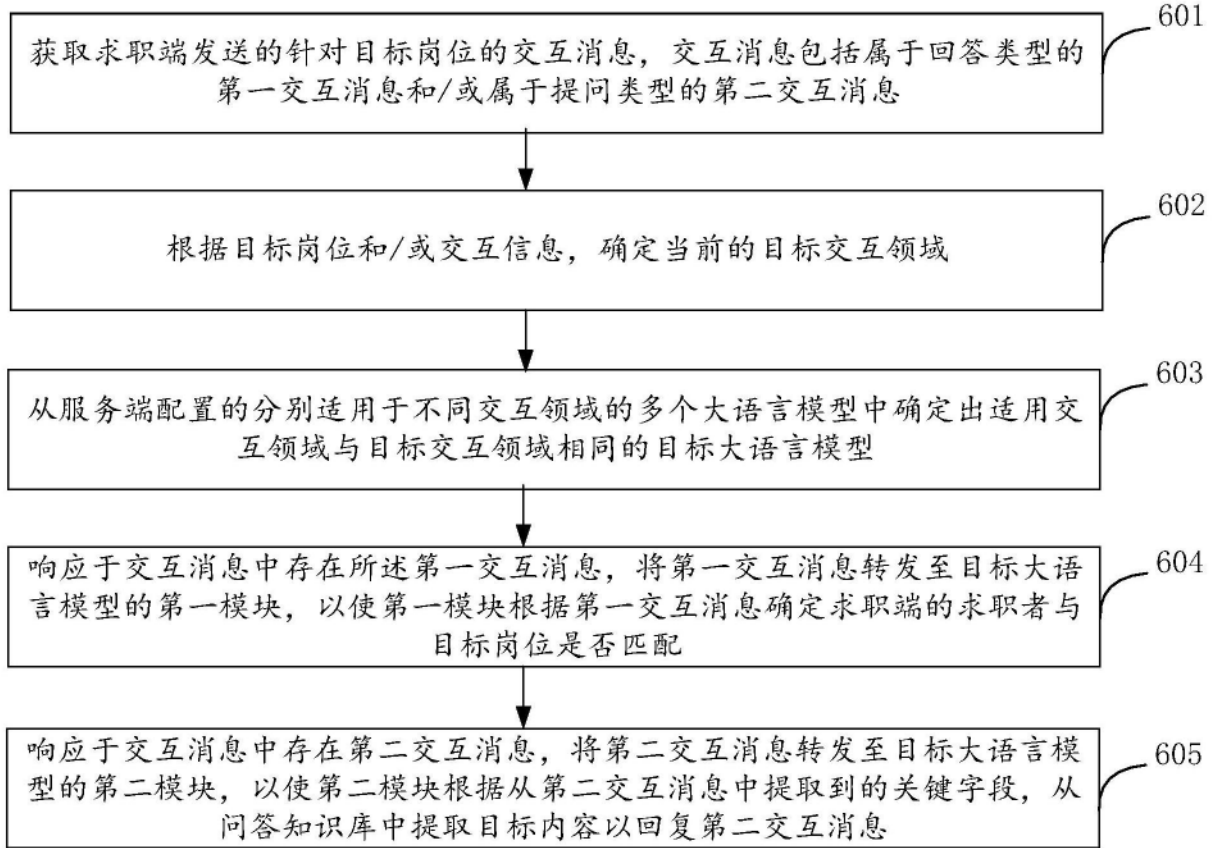


图6

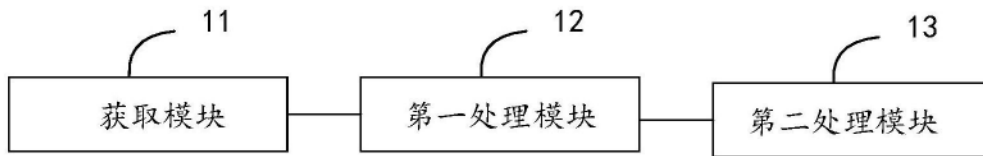


图7

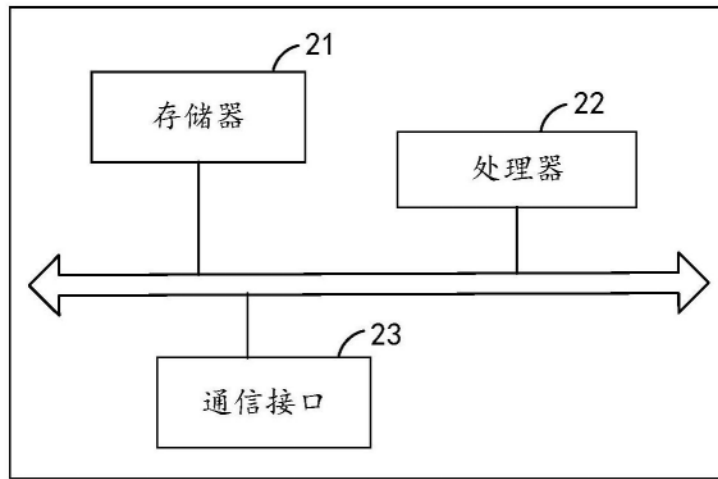


图8