



(12) 发明专利申请

(10) 申请公布号 CN 112100349 A

(43) 申请公布日 2020.12.18

(21) 申请号 202010915762.X

(22) 申请日 2020.09.03

(71) 申请人 深圳数联天下智能科技有限公司
地址 518000 广东省深圳市南山区粤海街道高新区社区科技南十二路011号方大大厦1706

(72) 发明人 王文

(74) 专利代理机构 深圳市六加知识产权代理有限公司 44372

代理人 孟丽平

(51) Int.Cl.
G06F 16/332 (2019.01)

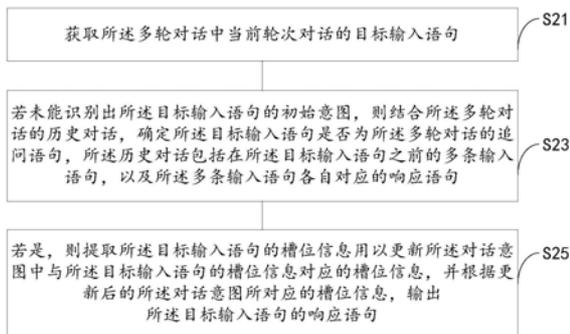
权利要求书2页 说明书11页 附图4页

(54) 发明名称

一种多轮对话方法、装置、电子设备及存储介质

(57) 摘要

本发明涉及人工智能技术领域,公开一种多轮对话方法、装置、电子设备及存储介质,通过在未能识别出目标输入语句的初始意图时,通过结合多轮对话的历史对话,确定目标输入语句是否为多轮对话的追问语句,若是,则提取目标输入语句的槽位信息用以更新对话意图中与目标输入语句的槽位信息对应的槽位信息,并根据更新后的对话意图所对应的槽位信息,输出目标输入语句的响应语句,从而,多轮对话能够根据历史对话进行追问回复,提高了追问回复的灵活性和准确性。



1. 一种多轮对话方法,其特征在于,所述方法包括:

获取多轮对话中当前轮次对话对应的目标输入语句,其中,所述多轮对话的对话意图的所有槽位已填充完成;

若未能识别出所述目标输入语句的初始意图,则结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,所述历史对话包括在所述目标输入语句之前的多条输入语句,以及所述多条输入语句各自对应的响应语句;

若是,则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息,并根据更新后的所述对话意图所对应的槽位信息,输出所述目标输入语句的响应语句。

2. 根据权利要求1所述的方法,其特征在于,所述结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句之前,还包括:

将所述目标输入语句输入至意图分类器,以获取多个意图分类结果及所述多个意图分类结果对应的概率值;

若各所述意图分类结果对应的概率值均小于预设概率阈值,则未能识别出所述目标输入语句的初始意图;

否则,将多个意图分类结果对应的概率值中最大概率值所对应的意图作为所述目标输入语句的初始意图。

3. 根据权利要求1所述的方法,其特征在于,所述结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,包括:

将所述历史对话与所述目标输入语句进行拼接,以得到拼接样本;

将所述拼接样本输入追问模型,以确定所述目标输入语句是否为所述多轮对话的追问语句。

4. 根据权利要求3所述的方法,其特征在于,所述将所述历史对话与所述目标输入语句进行拼接,以得到拼接样本,包括:

依次将所述历史对话中的多条输入语句及所述多条输入语句各自对应的响应语句进行拼接,得到历史对话文本;

将所述目标输入语句拼接于所述历史对话文本之后,以得到拼接样本。

5. 根据权利要求3或4所述的方法,其特征在于,所述将所述拼接样本输入追问模型,以确定所述目标输入语句是否为所述多轮对话的追问语句,包括:

将所述拼接样本转化为数字向量,并基于所述数字向量提取出M维特征向量,M为大于或等于1的正整数;

将所述M维特征向量转化为二维输出向量,其中,所述二维输出向量分别对应两种判定结果,所述两种判定结果分别为所述目标条输入语句为所述多轮对话的追问语句,和所述目标条输入语句不为所述多轮对话的追问语句;

将所述二维输出向量转换为概率分布,并将所述概率分布中最大概率值所对应的判定结果作为所述追问模型的输出结果。

6. 根据权利要求1-5任一项所述的方法,其特征在于,所述获取多轮对话中当前轮次对话对应的目标输入语句之前,所述方法还包括:

获取所述多轮对话的第1条输入语句,并根据所述第1条输入语句识别出所述对话意

图；

根据所述对话意图的槽位创建数据结构，所述数据结构用于储存所述对话意图的槽位信息；

提取所述第1条输入语句的槽位信息，并将所述第1条输入语句的槽位信息填充至所述数据结构中与所述第1条输入语句的槽位信息对应的槽位位置；

当所述数据结构的槽位未填满时，根据未填充槽位对应的关键词输出所述第1条输入语句的响应语句；

当所述数据结构的槽位填满时，则根据所述数据结构输出的响应语句。

7. 根据权利要求1所述的方法，其特征在于，所述方法还包括：

若识别出所述目标输入语句的初始意图，则将所述初始意图作为新一轮多轮对话的对话意图，开启新一轮多轮对话。

8. 一种多轮对话装置，其特征在于，所述装置包括：

输入获取模块，用于获取所述多轮对话中当前轮次对话对应的目标输入语句，其中，所述多轮对话的对话意图的所有槽位已填充完成；

追问确定模块，用于若未能识别出所述目标输入语句的初始意图，则结合所述多轮对话的历史对话，确定所述目标输入语句是否为所述多轮对话的追问语句，所述历史对话包括在所述目标输入语句之前的多条输入语句，以及所述多条输入语句各自对应的响应语句；

输出模块，用于若是所述当前轮次对话的追问语句，则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息，并根据更新后的所述对话意图所对应的槽位信息，输出所述目标输入语句的响应语句。

9. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机可执行程序，当所述计算机可执行程序被处理器执行时，使所述计算机执行如权利要求1-7中任一项所述的多轮对话方法。

10. 一种电子设备，其特征在于，包括：

处理器；以及

存储器，用于存储所述处理器的可执行指令；

其中，所述处理器配置为经由执行所述可执行指令来执行如权利要求1-7任一项所述的多轮对话方法。

一种多轮对话方法、装置、电子设备及存储介质

技术领域

[0001] 本发明涉及人工智能技术领域,特别是涉及一种多轮对话方法、装置、电子设备及存储介质。

背景技术

[0002] 随着人工智能技术的发展,人机对话技术中的多轮对话技术是实现智能人机交互系统的核心技术之一,具有广泛的应用场景,并且可以直接使用在具体的业务处理中,例如,询问天气、酒店、机票等。

[0003] 在现有的人机对话技术中,当用户通过多轮对话已经得到相应的回复之后,再进一步的基于历史对话进行追问时,由于历史对话已经形成一轮完整的多轮对话,而使得基于历史对话进行的追问成为新一轮多轮对话。然而,由于追问是基于历史对话提出,若脱离历史对话,则无法有效的识别出追问的对话意图并进行用户意图回复。

发明内容

[0004] 本发明实施例的一个目的旨在提供一种多轮对话方法、装置、电子设备及存储介质,其能够解决在用户追问时有效识别出对话意图并进行灵活回复,提高用户体验。

[0005] 为了解决上述技术问题,本发明提供以下技术方案:

[0006] 第一方面,本发明实施例提供多轮对话方法,所述方法包括:

[0007] 获取所述多轮对话中当前轮次对话对应的目标输入语句,其中,所述多轮对话的对话意图的所有槽位已填充完成;

[0008] 若未能识别出所述目标输入语句的初始意图,则结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,所述历史对话包括在所述目标输入语句之前的多条输入语句,以及所述多条输入语句各自对应的响应语句;

[0009] 若是,则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息,并根据更新后的所述对话意图所对应的槽位信息,输出所述目标输入语句的响应语句。

[0010] 可选地,所述结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句之前,还包括:

[0011] 将所述目标输入语句输入至意图分类器,以获取多个意图分类结果及所述多个意图分类结果对应的概率值;

[0012] 若各所述意图分类结果对应的概率值均小于预设概率阈值,则未能识别出所述目标输入语句的初始意图;

[0013] 否则,将多个意图分类结果对应的概率值中最大概率值所对应的意图作为所述目标输入语句的初始意图。

[0014] 可选地,所述结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,包括:

- [0015] 将所述历史对话与所述目标输入语句进行拼接,以得到拼接样本;
- [0016] 将所述拼接样本输入追问模型,以确定所述目标输入语句是否为所述多轮对话的追问语句。
- [0017] 可选地,所述将所述历史对话与所述目标输入语句进行拼接,以得到拼接样本,包括:
- [0018] 依次将所述历史对话中的多条输入语句及所述多条输入语句各自对应的响应语句进行拼接,得到历史对话文本;
- [0019] 将所述目标输入语句拼接于所述历史对话文本之后,以得到拼接样本。
- [0020] 可选地,所述将所述拼接样本输入追问模型,以确定所述目标输入语句是否为所述多轮对话的追问语句,包括:
- [0021] 将所述拼接样本转化为数字向量,并基于所述数字向量提取出M维特征向量;
- [0022] 将所述M维特征向量转化为二维输出向量,其中,所述二维输出向量分别对应两种判定结果,所述两种判定结果分别为所述目标条输入语句为所述多轮对话的追问语句,和所述目标条输入语句不为所述多轮对话的追问语句;
- [0023] 将所述二维输出向量转换为概率分布,并将所述概率分布中最大概率值所对应的判定结果作为所述追问模型的输出结果。
- [0024] 可选地,所述获取所述多轮对话中当前轮次对话对应的目标输入语句之前,所述方法还包括:
- [0025] 获取所述多轮对话的第1条输入语句,并根据所述第1条输入语句识别出所述对话意图;
- [0026] 根据所述对话意图的槽位创建数据结构,所述数据结构用于储存所述对话意图的槽位信息;
- [0027] 提取所述第1条输入语句的槽位信息,并将所述第1条输入语句的槽位信息填充至所述数据结构中与所述第1条输入语句的槽位信息对应的槽位位置;
- [0028] 当所述数据结构的槽位未填满时,根据未填充槽位对应的关键词输出所述第1条输入语句的响应语句;
- [0029] 当所述数据结构的槽位填满时,则根据所述数据结构输出的响应语句。
- [0030] 可选地,所述方法还包括:
- [0031] 若识别出所述目标输入语句的初始意图,则将所述初始意图作为新一轮所述多轮对话的对话意图,开启新一轮所述多轮对话。
- [0032] 第二方面,本发明实施例提供一种多轮对话装置,所述装置包括:
- [0033] 输入获取模块,用于获取所述多轮对话中当前轮次对话对应的目标输入语句,其中,所述多轮对话的对话意图的所有槽位已填充完成;
- [0034] 追问确定模块,用于若未能识别出所述目标输入语句的初始意图,则结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,所述历史对话包括在所述目标输入语句之前的多条输入语句,以及所述多条输入语句各自对应的响应语句;
- [0035] 输出模块,用于若是所述当前轮次对话的追问语句,则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息,并根

据更新后的所述对话意图所对应的槽位信息,输出所述目标输入语句的响应语句。

[0036] 第三方面,本发明实施例提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机可执行程序,当所述计算机可执行程序被处理器执行时,使所述计算机执行上述任一项所述的多轮对话方法。

[0037] 第四方面,本发明实施例提供一种电子设备,包括:

[0038] 处理器;以及

[0039] 存储器,用于存储所述处理器的可执行指令;

[0040] 其中,所述处理器配置为经由执行所述可执行指令来执行如上述任一项所述的多轮对话方法。

[0041] 相对于传统技术,在本发明各实施例提供的多轮对话方法、装置、电子设备及存储介质,在未能识别出所述目标输入语句的初始意图时,通过结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,若是,则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息,并根据更新后的所述对话意图所对应的槽位信息,输出所述目标输入语句的响应语句,从而,在用户追问时能够有效识别出对话意图,并能够根据该对话意图进行灵活回复,提高用户体验。

附图说明

[0042] 一个或多个实施例通过与之对应的附图中的图片进行示例性说明,这些示例性说明并不构成对实施例的限定,附图中具有相同参考数字标号的元件表示为类似的元件,除非有特别申明,附图中的图不构成比例限制。

[0043] 图1为本发明实施例提供的一种人机对话系统的结构示意图;

[0044] 图2为本发明实施例提供的一种多轮对话方法的流程示意图;

[0045] 图3为本发明实施例提供的又一人机对话方法的流程示意图;

[0046] 图4为本发明实施例提供的一种意图识别方法的流程示意图;

[0047] 图5为本发明实施例提供的多轮对话方法中追问语句的确定方法的流程示意图;

[0048] 图6为本发明实施例提供的一种追问模型的结构示意图;

[0049] 图7为本发明实施例提供的一种多轮对话方法的流程示意图;

[0050] 图8为本发明实施例提供的一种多人对话装置的结构示意图。

具体实施方式

[0051] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0052] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互任意组合。在附图的流程图中示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行。并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0053] 自然语言处理是人工智能的一个子领域,致力于使计算机能够理解和处理人类语

言,使计算机更接近于人类对语言的理解,旨在文本数据中提取信息,在文本上运行模型来提取实体。其中,自然语言理解是用于完成对文本的语义解析,提取关键信息,比如实体、意图等;自然语言生成则针对用户的输入产生响应的自然语言文本。从而,基于上述人工智能技术实现人机之间进行人机对话。

[0054] 请参阅图1,为本发明实施例提供的一种人机对话系统的结构框图,其中,人机对话系统100包括用户终端10和电子设备20。

[0055] 具体的,用户终端10为用户使用的终端设备,其具体可以是桌面电脑、笔记本电脑等个人计算机,也可以是手机、平板电脑、多媒体设备、智能音箱、智能可穿戴设备等,但不限于此。

[0056] 电子设备20可以提供人机对话服务,终端设备10可以经由互联网与电子设备20建立连接,从而使得用户可以经由终端设备10与电子设备20进行人机对话。用户在终端设备10上打开浏览器或者人机对话类应用(APP),例如人工智能助手,并通过文字输入用户语句(文本),由终端设备10将用户语句发送到电子设备20。电子设备20接收到用户输入语句后,对用户输入语句进行语义识别,并根据语义识别结果向终端设备10返回响应语句,从而实现人机对话。

[0057] 在一些实施例中,终端设备10还可以采集用户的语音数据,并对语音数据进行语音识别处理,得到用户语句;或者,终端设备也可以将语音数据发送到电子设备20,由电子设备20对语音数据进行语音识别处理,得到用户输入语句。

[0058] 电子设备20用于向用户终端10提供服务,其可以实现为服务器,例如应用服务器、Web服务器等;也可以实现为桌面电脑、笔记本电脑、处理器芯片、平板电脑等,但不限于此。具体的,电子设备20包括一个或多个处理器以及存储器,在本实施例中,以一个处理器为例进行说明,其中,处理器和存储器可以通过总线或者其他方式连接。

[0059] 存储器作为一种非易失性计算机可读存储介质,可用于存储非易失性软件程序、非易失性计算机可执行程序以及模块,如本发明下述实施例中的多轮对话方法的程序指令/模块。处理器通过运行存储在存储器中的非易失性软件程序、指令以及模块,从而执行服务器的各种功能应用以及数据处理,即实现下述方法实施例中多轮对话方法以及装置实施例的各个模块的功能。

[0060] 存储器可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实施例中,存储器可选包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至处理器。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。所述程序指令/模块存储在所述存储器中,当被所述一个或者多个处理器执行时,执行下述任意方法实施例中的多轮对话方法。

[0061] 应该说明的是,人机对话过程通常包括多轮对话,电子设备20需要对人机对话的对话状态进行跟踪,以便根据当前的对话状态,对用户输入语句进行准确的响应,从而使得人机对话能够流畅进行,其中,对话状态主要是指识别的对话意图所对应的槽位填充是否完成。

[0062] 在又一些实施例中,人机对话系统100还包括数据存储装置30。数据存储装置30可以是关系型数据库例如MySQL、ACCESS等,也可以是非关系型数据库例如NoSQL等;可以是驻

留于电子设备20中的本地数据库,也可以作为分布式数据库,例如HBase等设置于多个地理位置。总之,数据存储装置30用于存储数据,本发明实施例对数据存储装置30的具体部署、配置情况不做限制。电子设备20可以与数据存储装置30连接,并获取数据存储装置30中所存储的数据。例如,电子设备20可以直接读取数据存储装置30中的数据(当数据存储装置30为电子设备20的本地数据库时),也可以通过有线或无线的方式接入互联网,并通过数据接口来获取数据存储装置30中的数据。

[0063] 在一些实施例中,数据存储装置30还可以临时或者持久存储人机对话系统的对话语句,对话语句包括多个用户输入语句以及人机对话系统针对每一用户输入语句的响应语句。对于多轮对话而言,在本发明实施例中,当前轮次对话语句包括当前轮次的用户输入语句和针对该用户输入语句的响应语句。

[0064] 应该说明的是,人机对话时涉及的自然语言理解技术中,自然语言理解的目的是要获得一个语义表示,其中,语义表示方式主要包括分布语义、框架语义及模型论语义,在本实施例中,通过在框架语义表示方式的基础上进行变形,亦即,采用领域(Domain)、对话意图(Intent)和槽位(Slots)来表示语义结果。

[0065] 其中,领域是指同一类型的数据或者资源,以及围绕这些数据或资源提供的服务,比如,飞机票、火车票、天气等。对话意图是指对于领域数据的操作,具体是指在该人机对话过程中用户希望获取到的信息,比如,在飞机票领域中,有购票和/或退票等对话意图。槽位用于存放领域的属性,其预先定义于人机对话系统中,当人机对话系统识别出对话意图时,该对话意图所需的槽位即已确定,应该说明的是,槽位的数量可以为一个或多个,具体数量与对话意图相关。例如,询问天气的人机对话,则需要获取地点(location)和时间(time)这两个槽位信息。

[0066] 可以理解的,在人机对话过程中,一条用户的输入语句无法满足该输入语句所对应的对话意图的全部槽位填充,因此,人机对话系统需通过问答的形式提醒用户进行必要信息的输入,以进一步的获取空缺的槽位信息并完成填充,于是,产生了人机对话的多轮对话方法。具体的,多轮对话是指在人机对话过程中,根据用户的输入语句识别出对话意图后,进一步获取槽位信息以最终得到明确回复的对话。其中,多轮对话包括一个或多个轮次的问答形式的对话,每个轮次的对话包括1条用户的输入语句和/或针对该输入语句的响应语句。当对话意图的全部槽位信息均已填充完成,则人机对话系统则能根据槽位信息针对用户的输入语句进行回复,于是,该轮多轮对话结束。

[0067] 在实际使用时,用户在该多轮对话结束后,仍然希望基于结束的该轮多轮对话的基础上继续提问,于是,在本实施例中,提供一种多轮对话方法,该方法基于一轮多轮对话结束时,若继续接收到用户的目标输入语句,则请参阅图2,所述方法包括:

[0068] S21、获取所述多轮对话中当前轮次对话的目标输入语句,其中,所述多轮对话的对话意图的所有槽位已填充完成;

[0069] 当前轮次对话是指该轮多轮对话中最近一次时间接收到的用户的输入语句和/或针对该用户的输入语句输出的响应语句,在本实施例中,当前轮次对话仅包含最近一次时间接收到的用户的输入语句,即为目标输入语句。目标输入语句是用户得到针对对话意图的响应语句之后,继续输入的提问语句,以以下例子来说明本方法的应用场景。

[0070] 用户:明天的天气怎样?(可识别出领域“天气”、用户意图“询问天气”及槽位“date

=明天”)

[0071] 系统:哪个城市?(以问询的方式获取槽位信息)

[0072] 用户:深圳。(进一步填充槽位“location=深圳”)

[0073] 系统:小雨转阴(已完成全部槽位填充“date=明天,location=深圳”)。

[0074] 用户:后天呢?

[0075] 其中,人机对话系统已根据用户的输入语句“明天的天气怎样”及“深圳”识别出领域和用户意图,并完成全部槽位的填充,从而,系统输出了针对用户的输入语句的响应语句“小雨转阴”,到此为止,已完成了一轮多轮对话过程。在此基础上,若继续接收到用户的目标输入语句“后天呢”,于是,针对该目标输入语句,在一些实施例中,请参阅图3,所述方法还包括:

[0076] S22、根据所述目标输入语句识别初始意图;

[0077] 其中,所述初始意图是指目标输入语句所指示的意图,其可以与上述的用户意图相同或相异。可以理解的,准确识别所述目标输入语句的意图是人机对话能够流畅进行的基础,具体的,识别所述目标输入语句的初始意图时,请参阅图4,所述方法包括:

[0078] S221、将所述目标输入语句输入至意图分类器,以获取多个意图分类结果及所述多个意图分类结果对应的概率值;

[0079] 在一些实施例中,针对所述目标输入语句,首先对其进行预处理,包括定义用户词典、分词、去除特殊符号、去除停用词等,其中,用户词典是指在分词时手动添加的一些不可分的词,以防止分词算法时将一些固定搭配的短语拆分,进而影响了人机对话系统对初始意图的识别。

[0080] 针对经过预处理后的所述目标输入语句,采用意图分类器进行意图分类,其中,意图分类器可以采用业界较为成熟的及其学习方法,例如,逻辑回归算法(Logistic Regression,LR)、支持向量机算法(Support Vector Machine,SVM)。或者深度学习方法,例如,文本卷积神经网络算法(Text Convolutional Neural Network,Text CNN)等,具体采用哪一种方法可根据不同方法在数据上的表现效果来做进一步选择,本实施例不对意图分类器的方法进行限定。

[0081] 所述意图分类器根据所述目标输入语句输出多个分类结果及所述多个意图分类结果对应的概率值,例如,当多个意图分类结果所对应的概率值分别为0.2、0.35、0.5、0.75,于是,进一步根据上述多个意图分类结果所对应的概率值判断所述目标输入语句的初始意图。

[0082] S222、若各所述意图分类结果对应的概率值均小于预设概率阈值,则未能识别出所述目标输入语句的初始意图;

[0083] 在人机对话系统中,存储有预设概率阈值用以协助进行意图分类结果的概率判断,其中,所述预设概率阈值为根据实际需要进行设定的值。具体的,将多个意图分类结果所对应的概率值分别与所述预设概率阈值进行比较,判断是否存在大于所述预设概率阈值的概率值。若各所述意图分类结果对应的概率值均小于预设概率阈值,则未能识别出所述目标输入语句的初始意图,亦即,所述意图分类器未能识别出所述目标输入语句的初始意图。

[0084] S223、否则,将多个意图分类结果对应的概率值中最大概率值所对应的意图作为

所述目标输入语句的初始意图。

[0085] 若存在大于或等于所述预设概率阈值的概率值,则表示所述意图分类器识别出所述目标输入语句的初始意图,并将多个意图分类结果对应的概率值中最大概率值所对应的意图作为所述目标输入语句的初始意图。

[0086] 在本实施例中,将所述目标输入语句输入至意图分类器,用以识别出所述目标输入语句所对应的初始意图,提高了意图识别的准确程度。

[0087] S23、若未能识别出所述目标输入语句的初始意图,则结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,所述历史对话包括在所述目标输入语句之前的多条输入语句,以及所述多条输入语句各自对应的响应语句;

[0088] 其中,历史对话是指当前多轮对话中对应的产生于该目标输入语句之前的所有对话,包括在所述目标输入语句之前的多条输入语句,以及所述多条输入语句各自对应的响应语句,可以理解的,该轮多轮对话已根据历史对话完成了用户意图的响应。

[0089] 当人机对话系统未能识别出所述目标输入语句的初始意图,则进一步的结合所述多轮对话的历史对话,确定所述目标输入语句是否为历史对话的追问语句,其中,追问语句是指基于历史对话的输入语句,与历史对话紧密相关,只有当结合历史对话后,才能得到用户输入追问语句所希望达到的目的。具体的,在一些实施例中,采用如下方法用以确定所述目标输入语句是否为历史对话的追问语句,请参阅图5,所述方法包括:

[0090] S231、将所述历史对话与所述目标输入语句进行拼接,以得到拼接样本;

[0091] 具体的,依次将所述历史对话中的多条输入语句及所述多条输入语句各自对应的响应语句进行拼接,得到历史对话文本;将所述目标输入语句拼接于所述历史对话文本之后,以得到拼接样本。举例说明:

[0092] 用户:明天的天气怎样?

[0093] 系统:哪个城市?

[0094] 用户:深圳。

[0095] 系统:小雨转阴。

[0096] 用户:后天呢?

[0097] 那么,依次将所述历史对话中的多条输入语句及所述多条输入语句各自对应的响应语句进行拼接后得到历史对话文本为“明天的天气怎样?哪个城市?深圳。小雨转阴。”,将所述目标输入语句拼接于所述历史对话文本之后得到的拼接样本为“明天的天气怎样?哪个城市?深圳。小雨转阴。后天呢?”,于是得到所述拼接样本。

[0098] S232、将所述拼接样本输入追问模型,以确定所述目标输入语句是否为所述多轮对话的追问语句。

[0099] 所述追问模型为预设的神经网络语言模型,所述神经网络语言模型包括BERT层(Bidirectional Encoder Representations from Transformers,基于Transformer的双向编码器表征)、全连接层及Softmax层。其中,BERT层用于将输入的拼接样本转化为数字向量并提取出M维特征向量,例如,通过BERT层将输入的拼接样本转化为一个768维数字向量并提取出768维特征向量。在一些实施例中,在输入所述追问模型之前,先将拼接样本转化为数字向量,于是,BERT层则用于从输入的所述数字向量提取出M维特征向量。全连接层为该追问模型的“分类器”,用于将BERT层提取出的M维特征向量转换为二维输出向量,在本实

施例中,由于只需要得到两种判定结果,于是,在全连接层中将所述M维特征向量转化为二维输出向量,其中,所述二维输出向量分别对应两种判定结果,所述两种判定结果分别为所述目标条输入语句为所述多轮对话的追问语句,和所述目标条输入语句不为所述多轮对话的追问语句。应该说明的是,在其他实施例中,全连接层还可以根据需求将BERT层提取出的M维特征向量转换为任意维度的输出向量,在本实施例中不对全连接层转换的维度做限制。Softmax层位于该追问模型的最后一层,用于将全连接层转换出的二维输出向量进一步转换为概率分布,将任意时间发生的概率分布于0-1之间,并将所述概率分布中最大概率值所对应的判定结果作为所述追问模型的输出结果。

[0100] 请参阅图6,图6为本发明实施例提供的追问模型的结构示意图,可以理解的,追问模型还可以根据实际需求加入其他层,均属于本发明实施例的保护范围。具体的,将所述拼接样本输入所述追问模型后,所述追问模型将所述拼接样本转化为数字向量,并基于所述数字向量提取出M维特征向量;将所述M维特征向量转化为二维输出向量,其中,所述二维输出向量分别对应两种判定结果,所述两种判定结果分别为所述目标条输入语句为所述多轮对话的追问语句,和所述目标条输入语句不为所述多轮对话的追问语句;将所述二维输出向量转换为概率分布,并将所述概率分布中最大概率值所对应的判定结果作为所述追问模型的输出结果。如图6所示,假设历史对话文本为“a=深圳明天天气怎么样?小雨转阴。”追问语句为“b=后天呢?”,那么,将a和b拼接后的拼接样本为“a+b=深圳明天天气怎么样?小雨转阴。后天呢?”,在本实施例中,在输入所述追问模型之前,先将拼接样本转化为数字向量[64,32,5,11,11,87,98,23,72,8,90,11,12,17,2,99,11,71,8]。将该数字向量输入BERT层,BERT层从输入的所述数字向量提取出M维特征向量[0.3625,1.492,-1.5828,...,9.234,0.1111],其中,维度M可调,并将提取的M维特征向量输出至全连接层,全连接层将所述M维特征向量转化为二维输出向量[2.382,3.182],并将二维输出向量输出至Softmax层,于是,Softmax层将二维输出向量进一步转换为概率分布[0.36,0.64],并将概率较高的事件标签为1,概率较低的事件标签为0,最后,将概率较高的事件标签1作为该追问模型的输出结果进行输出。

[0101] 可以理解的,若所述概率分布中最大概率值所对应的判定结果为是所述历史对话的追问,则执行步骤S25;若所述概率分布中最大概率值所对应的判定结果为不是所述历史对话的追问时,在又一些实施例中,所述方法还包括:

[0102] S24、若识别出所述目标输入语句的初始意图,则将所述初始意图作为新一轮多轮对话的对话意图,开启新一轮多轮对话。

[0103] 其中,新一轮多轮对话是用于区别上一轮多轮对话而言,新一轮所述多轮对话具有与上一轮多轮对话相异的对话意图,即该初始意图与上一轮多轮对话的对话意图相异,且,该初始意图具有与其对应的槽位。然后,人机对话系统将根据所述初始意图进入新一轮问答式的多轮对话。

[0104] S25、若是,则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息,并根据更新后的所述对话意图所对应的槽位信息,输出所述目标输入语句的响应语句。

[0105] 若所述目标输入语句被判断为是所述历史对话的追问语句,则说明该目标输入语句与历史对话具有相同的领域和用户意图,因此,只需提取出所述目标输入语句的槽位信

息,用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位,即可根据更新后的所述对话意图所对应的槽位信息,输出所述目标输入语句的响应语句,进而得到了目标输入语句的响应。继续以上述例子为例来说明槽位更新:

[0106] 用户:明天的天气怎样?(可识别出领域“天气”、用户意图“询问天气”及槽位“date=明天”)

[0107] 系统:哪个城市?(以问询的方式获取槽位信息)

[0108] 用户:深圳。(进一步填充槽位“location=深圳”)

[0109] 系统:小雨转阴(已完成全部槽位填充“date=明天,location=深圳”)。

[0110] 用户:后天呢?

[0111] 其中,“后天呢?”为所述多轮对话的目标输入语句,当判断所述目标输入语句为所述多轮对话的追问语句时,则说明“后天呢?”具有与历史对话相同的领域和用户意图,那么,只需要提取出目标输入语句中“后天”的槽位信息并更新到对应槽位“date=后天”中,那么,语义识别后得到的语义表达为:领域“天气”,用户意图“询问天气”,槽位“date=后天,location=深圳”,于是,所述人机对话系统根据该语义表达进行响应。

[0112] 在本发明实施例中,通过在未能识别出所述目标输入语句的初始意图时,通过结合所述多轮对话的历史对话,确定所述目标输入语句是否为所述多轮对话的追问语句,若是,则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息,并根据更新后的所述对话意图所对应的槽位信息,输出所述目标输入语句的响应语句,从而,在用户追问时能够有效识别出对话意图,并能够根据该对话意图进行灵活回复,提高用户体验。

[0113] 应该说明的是,上述多轮对话方法的实施是基于一轮多轮对话完成的基础上,因此,在获取所述多轮对话中当前轮次对话对应的目标输入语句之前,请参阅图7,所述方法还包括:

[0114] S31、获取所述多轮对话的第1条输入语句,并根据所述第1条输入语句识别出所述对话意图;

[0115] S32、提取所述第1条输入语句的槽位信息,并将所述第1条输入语句的槽位信息填充至所述数据结构中与所述第1条输入语句的槽位信息对应的槽位位置;

[0116] S33、当所述数据结构的槽位未填满时,根据未填充槽位对应的关键词输出所述第1条输入语句的响应语句;

[0117] S34、当所述数据结构的槽位填满时,则根据所述数据结构输出的响应语句。

[0118] 可以理解的,当识别出对话意图时,则该对话意图所对应的槽位数量也是确定的,例如:

[0119] 用户:明天的天气怎样?(可识别出领域“天气”、用户意图“询问天气”及槽位“date=明天”)

[0120] 系统:哪个城市?(以问询的方式获取槽位信息)

[0121] 用户:深圳。(进一步填充槽位“location=深圳”)

[0122] 系统:小雨转阴(已完成全部槽位填充“date=明天,location=深圳”)。

[0123] 那么,根据用户输入的第1条输入语句“明天的天气怎样”即可判别出该轮多轮对话的领域及用户意图,其中,领域为“天气”,用户意图为“询问天气”,那么该用户意图对应

的槽位包括“date=,location=”。

[0124] 紧接着提取出第1条输入语句“明天的天气怎样”的槽位信息“明天”，并将槽位信息“明天”填充至“date=”的槽位，于是，得到填充后的槽位为“date=明天,location=”。然而，此时的槽位并未得到完全的填充，于是，人机对话系统根据未得到填充的槽位“location=”输出问询语句“哪个城市？”，以提醒用户输入槽位相关信息。

[0125] 若用户输入的第1条输入语句“深圳明天的天气怎样？”，则根据该输入语句“深圳明天的天气怎样？”可识别出该轮多轮对话的领域及用户意图，并能根据输入语句“深圳明天的天气怎样？”提取出槽位信息用以填充“date=明天,location=深圳”的槽位，从而，完成了槽位的填充。

[0126] 当所述对话意图所对应的槽位均已填充完成，则人机对话系统根据填充完成的槽位信息进行用户输入语句的响应。

[0127] 在本发明实施例中，通过识别用户的输入语句所对应的对话意图，并在人机对话过程中，以问答的形式提醒用户进行必要信息的输入，以完成所有槽位的填充，从而保证了人机对话的顺利进行。

[0128] 请参阅图8，本发明实施例提供一种多轮对话装置，所述装置400包括：

[0129] 输入获取模块41，用于获取所述多轮对话中当前轮次对话对应的目标输入语句，其中，所述多轮对话的对话意图的所有槽位已填充完成；

[0130] 追问确定模块42，用于若未能识别出所述目标输入语句的初始意图，则结合所述多轮对话的历史对话，确定所述目标输入语句是否为所述多轮对话的追问语句，所述历史对话包括在所述目标输入语句之前的多条输入语句，以及所述多条输入语句各自对应的响应语句；

[0131] 输出模块43，用于若是所述当前轮次对话的追问语句，则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息，并根据更新后的所述对话意图所对应的槽位信息，输出所述目标输入语句的响应语句。

[0132] 应该说明的是，所述装置400的具体实施方式可参照上述方式实施例，在此不再一一赘述。

[0133] 在本发明实施例中，通过在未能识别出所述目标输入语句的初始意图时，通过结合所述多轮对话的历史对话，确定所述目标输入语句是否为所述多轮对话的追问语句，若是，则提取所述目标输入语句的槽位信息用以更新所述对话意图中与所述目标输入语句的槽位信息对应的槽位信息，并根据更新后的所述对话意图所对应的槽位信息，输出所述目标输入语句的响应语句，从而，在用户追问时能够有效识别出对话意图，并能够根据该对话意图进行灵活回复，提高用户体验。

[0134] 以上所描述的装置或设备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模块可以是或者也可以不是物理上分开的，作为模块单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络模块单元上。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。

[0135] 本发明实施例提供了一种非易失性计算机可读存储介质，所述计算机可读存储介质存储有计算机可执行指令，该计算机可执行指令被一个或多个处理器执行，例如，执行以上描述的图2至图7的方法步骤。

[0136] 本发明实施例提供了一种计算机程序产品,包括存储在非易失性计算机可读存储介质上的计算程序,所述计算机程序包括程序指令,当所述程序指令被计算机执行时,使所述计算机执行上述任意方法实施例中的本体构建方法,例如,执行以上描述的图2至图7的方法步骤。

[0137] 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;在本发明的思路下,以上实施例或者不同实施例中的技术特征之间也可以进行组合,步骤可以以任意顺序实现,并存在如上所述的本发明的不同方面的许多其它变化,为了简明,它们没有在细节中提供;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范围。

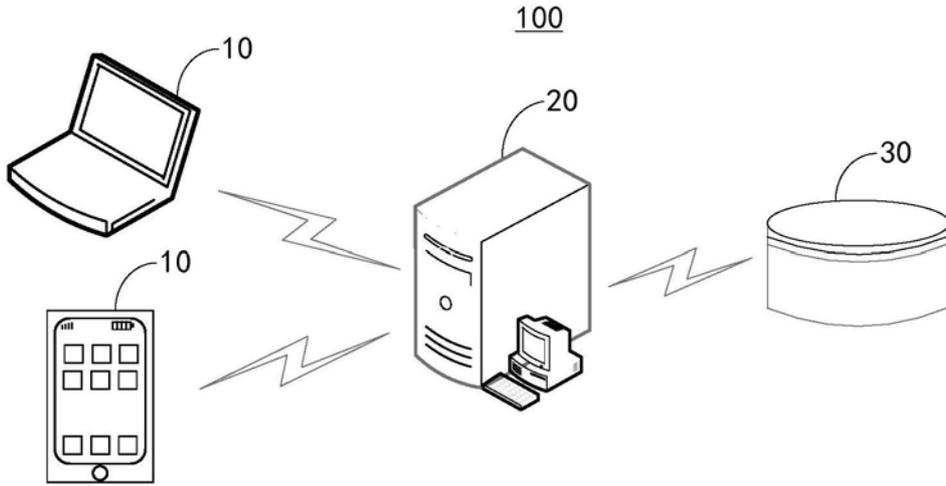


图1

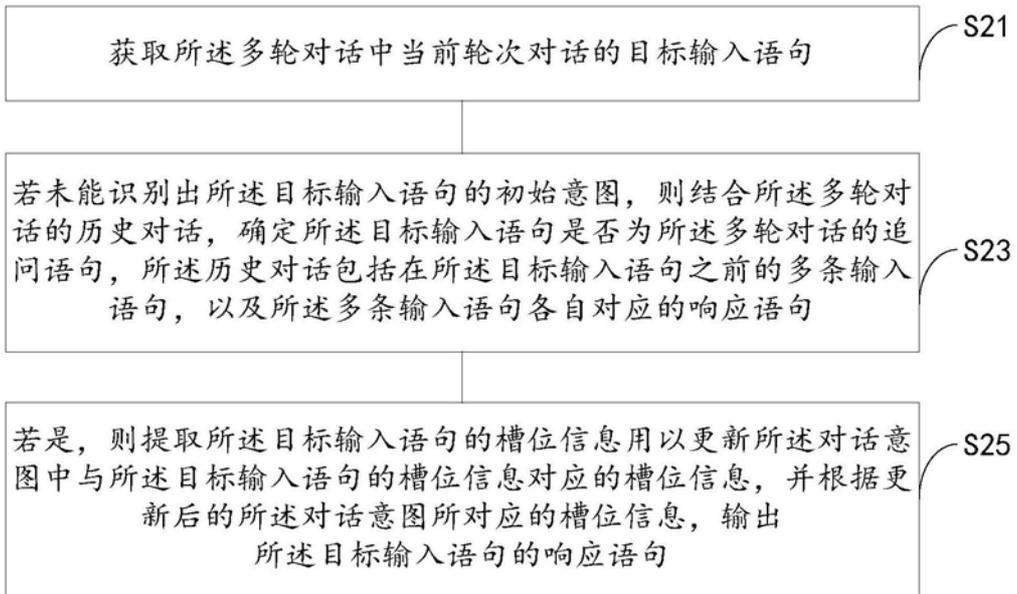
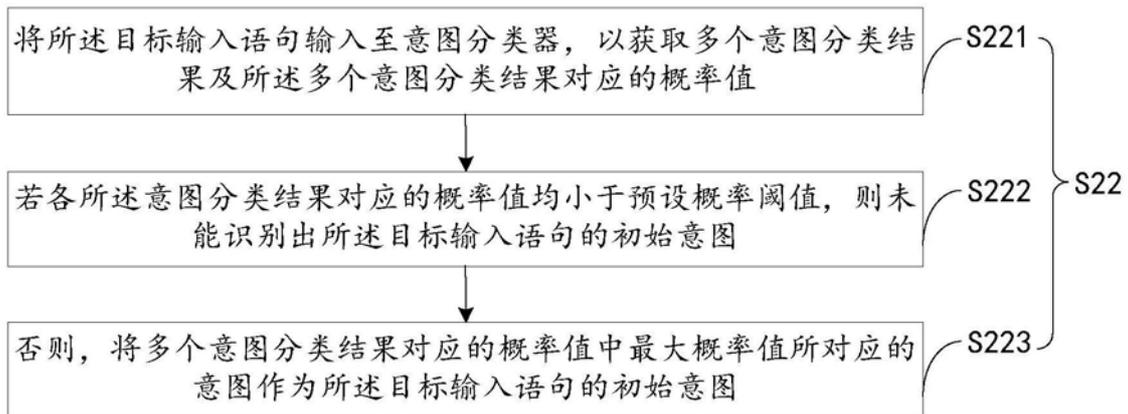
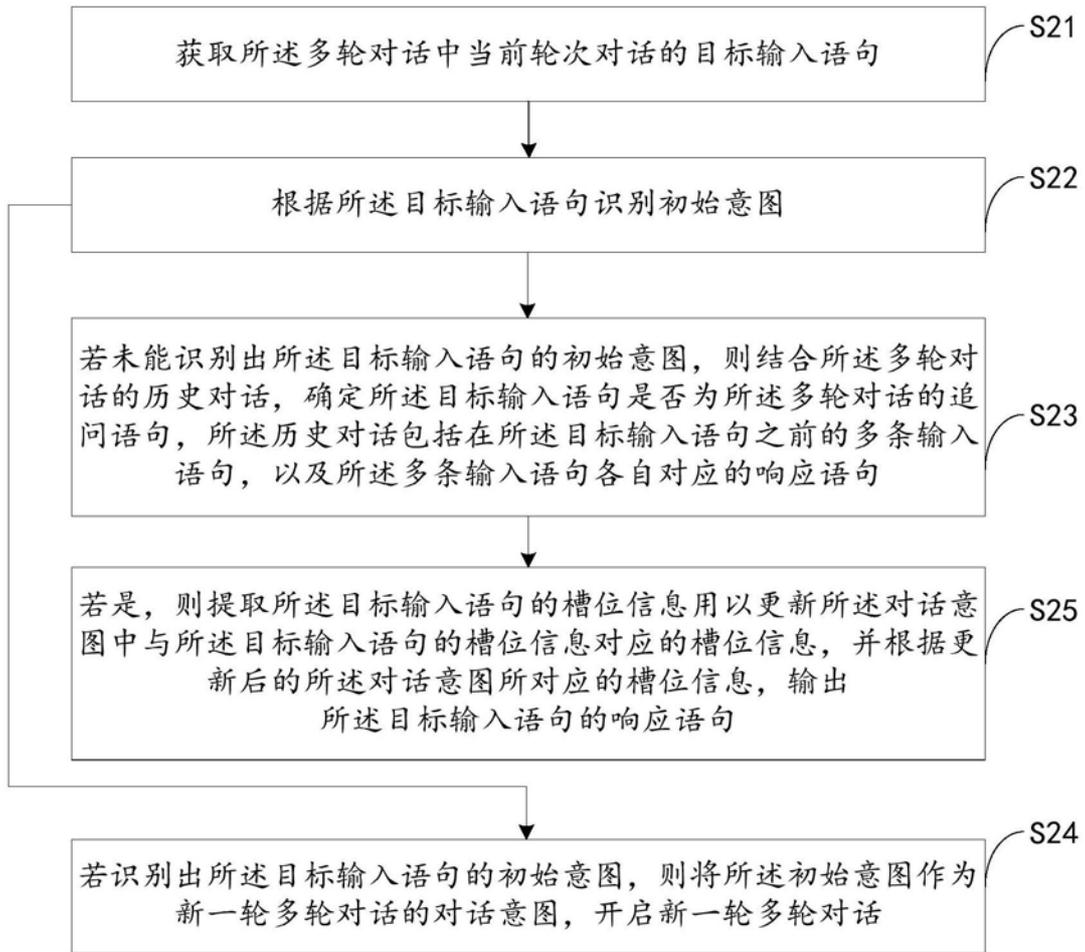


图2



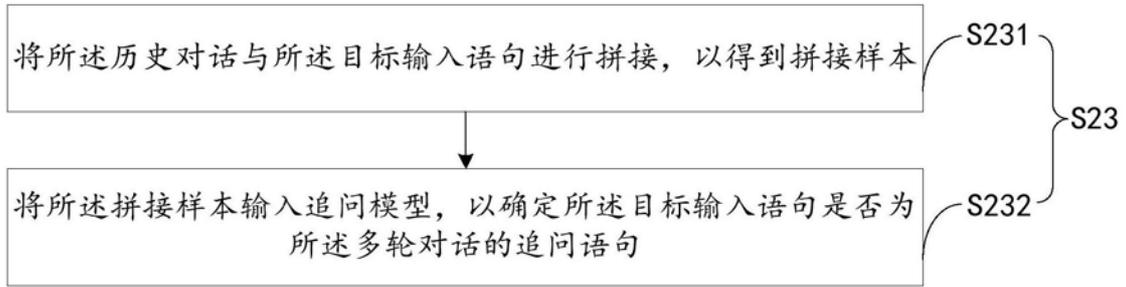


图5

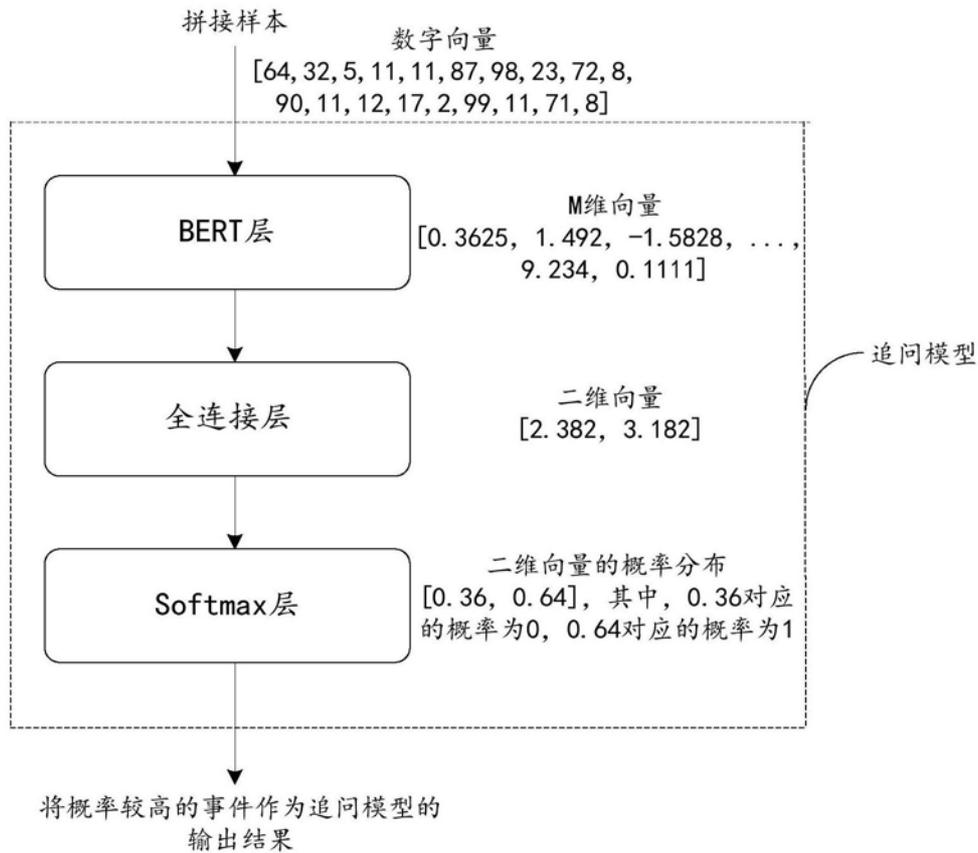


图6

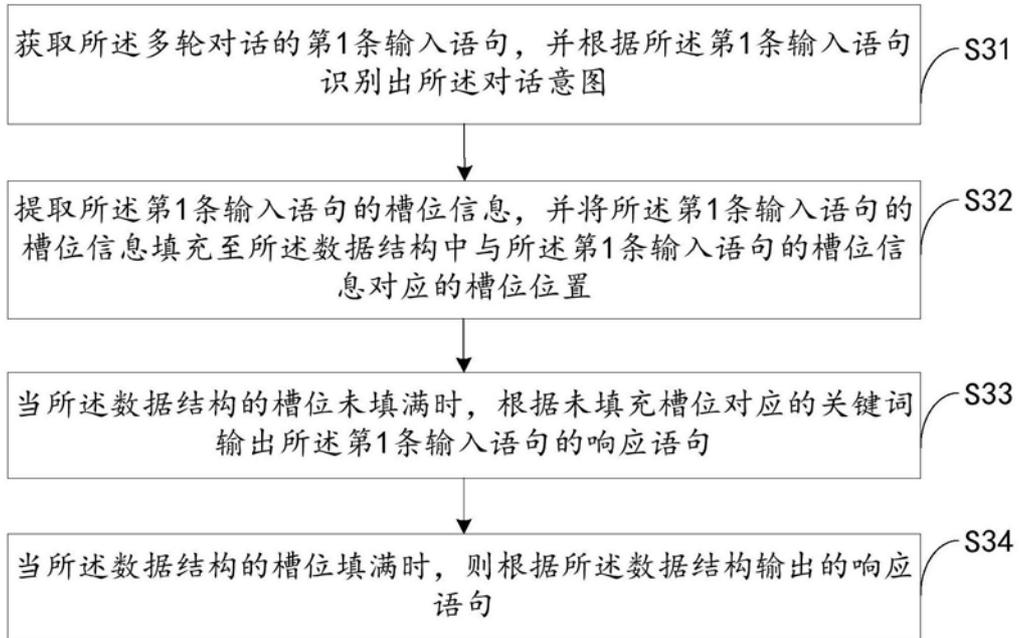


图7

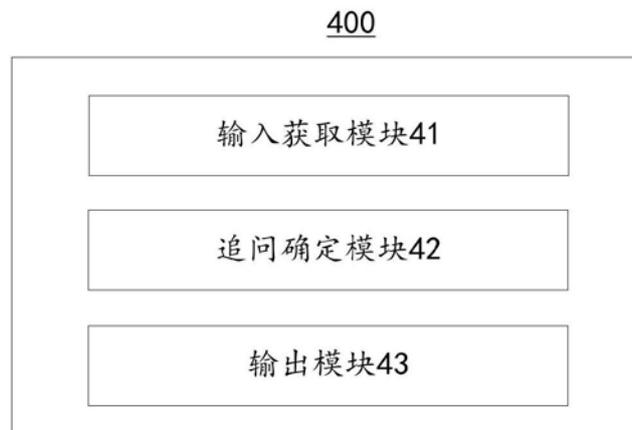


图8