



(12) 发明专利申请

(10) 申请公布号 CN 114626387 A

(43) 申请公布日 2022. 06. 14

(21) 申请号 202210246347.9

(22) 申请日 2022.03.14

(71) 申请人 京东科技信息技术有限公司

地址 100176 北京市北京经济技术开发区
科创十一街18号院2号楼6层601

(72) 发明人 刘航 陈蒙

(74) 专利代理机构 北京清亦华知识产权代理事
务所(普通合伙) 11201

专利代理师 单冠飞

(51) Int. Cl.

G06F 40/35 (2020.01)

G06F 16/332 (2019.01)

G06F 16/335 (2019.01)

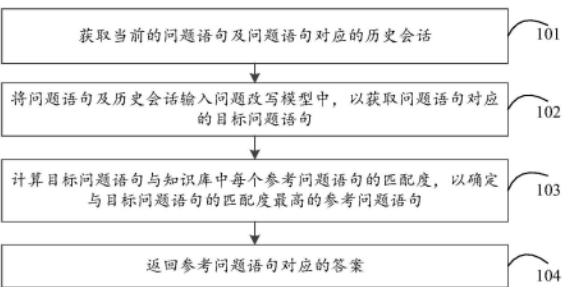
权利要求书3页 说明书12页 附图4页

(54) 发明名称

问题语句的处理方法、装置、计算机设备及
存储介质

(57) 摘要

本公开提出一种问题语句的处理方法、装置、计算机设备及存储介质,涉及计算机技术领域,尤其涉及智能对话领域。该方法包括:获取当前的问题语句及所述问题语句对应的历史会话;将所述问题语句及所述历史会话输入问题改写模型中,以获取所述问题语句对应的目标问题语句;计算所述目标问题语句与知识库中每个参考问题语句的匹配度,以确定与所述目标问题语句的匹配度最高的参考问题语句;返回所述参考问题语句对应的答案。由此,保证了问题语句输入信息的完整度,提高了系统应答结果的准确性。



1. 一种问题语句的处理方法,其特征在于,包括:
获取当前的问题语句及所述问题语句对应的历史会话;
将所述问题语句及所述历史会话输入问题改写模型中,以获取所述问题语句对应的目标问题语句;
计算所述目标问题语句与知识库中每个参考问题语句的匹配度,以确定与所述目标问题语句的匹配度最高的参考问题语句;
返回所述参考问题语句对应的答案。
2. 如权利要求1所述的方法,其特征在于,所述历史会话包括前N个问题语句至当前的所述问题语句之间的对话,其中,N为正整数。
3. 如权利要求1所述的方法,其特征在于,在所述将所述问题语句及所述历史会话输入问题改写模型中之前,还包括:
获取多个训练数据对,其中,每个所述训练数据对中包括:候选问题语句、所述候选问题语句对应的历史会话及对应的目标问题语句样本;
将所述候选问题语句及对应的历史会话,输入初始模型,以获取所述初始模型输出的预测语句;
根据所述预测语句与所述目标问题语句样本间的差异,对所述初始模型进行修订,以生成所述问题改写模型。
4. 如权利要求3所述的方法,其特征在于,所述获取多个训练数据对,包括:
从对话日志中获取多个目标问题语句样本及所述目标问题语句样本对应的历史会话;
对所述目标问题语句样本及对应的历史会话进行解析,以确定参考词语;
删除所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句。
5. 如权利要求4所述方法,其特征在于,在所述对所述目标问题语句样本及对应的历史会话进行解析,以确定参考词语之后,还包括:
使用预设的词语替换所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句。
6. 如权利要求4或5所述的方法,其特征在于,在所述对所述目标问题语句样本及对应的历史会话进行解析,以确定参考词语之后,还包括:
复制所述目标问题语句样本,以获取与所述目标问题语句样本对应的候选问题语句。
7. 如权利要求3所述的方法,其特征在于,所述获取多个训练数据对,包括:
根据所述对话日志中每个问题语句的长度,从所述对话日志中获取多个目标问题语句样本及所述目标问题语句样本对应的历史会话;
对所述目标问题语句样本及对应的历史会话进行解析,以确定所述参考词语;
为每个所述目标问题语句样本分别生成一个对应的随机概率;
在所述随机概率在第一数值范围内的情况下,删除所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句;
在所述随机概率在第二数值范围内的情况下,使用预设的词语替换所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句;
在所述随机概率在第三数值范围内的情况下,复制所述目标问题语句样本,以获取与

所述目标问题语句样本对应的候选问题语句。

8. 一种问题语句的处理装置,其特征在於,包括:

第一获取模块,用于获取当前的问题语句及所述问题语句对应的历史会话;

第二获取模块,用于将所述问题语句及所述历史会话输入问题改写模型中,以获取所述问题语句对应的目标问题语句;

确定模块,用于计算所述目标问题语句与知识库中每个参考问题语句的匹配度,以确定与所述目标问题语句的匹配度最高的参考问题语句;

返回模块,用于返回所述参考问题语句对应的答案。

9. 如权利要求8所述的装置,其特征在於,所述历史会话包括前N个问题语句至当前的所述问题语句之间的对话,其中,N为正整数。

10. 如权利要求8所述的装置,其特征在於,还包括训练模块,所述训练模块包括:

第一获取单元,用于获取多个训练数据对,其中,每个所述训练数据对中包括:候选问题语句、所述候选问题语句对应的历史会话及对应的目标问题语句样本;

第二获取单元,用于将所述候选问题语句及对应的历史会话,输入初始模型,以获取所述初始模型输出的预测语句;

生成单元,用于根据所述预测语句与所述目标问题语句样本间的差异,对所述初始模型进行修订,以生成所述问题改写模型。

11. 如权利要求10所述的装置,其特征在於,所述第一获取单元用于:

从对话日志中获取多个目标问题语句样本及所述目标问题语句样本对应的历史会话;

对所述目标问题语句样本及对应的历史会话进行解析,以确定参考词语;

删除所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句。

12. 如权利要求11所述装置,其特征在於,所述第一获取单元还用于:

使用预设的词语替换所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句。

13. 如权利要求11或12所述的装置,其特征在於,所述第一获取单元还用于:

复制所述目标问题语句样本,以获取与所述目标问题语句样本对应的候选问题语句。

14. 如权利要求10所述的装置,其特征在於,所述第一获取单元用于:

根据所述对话日志中每个问题语句的长度,从所述对话日志中获取多个目标问题语句样本及所述目标问题语句样本对应的历史会话;

对所述目标问题语句样本及对应的历史会话进行解析,以确定所述参考词语;

为每个所述目标问题语句样本分别生成一个对应的随机概率;

在所述随机概率在第一数值范围内的情况下,删除所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句;

在所述随机概率在第二数值范围内的情况下,使用预设的词语替换所述目标问题语句样本中的参考词语,以获取与所述目标问题语句样本对应的候选问题语句;

在所述随机概率在第三数值范围内的情况下,复制所述目标问题语句样本,以获取与所述目标问题语句样本对应的候选问题语句。

15. 一种计算机设备,其特征在於,包括存储器、处理器及存储在存储器上并可在处理

器上运行的计算机程序,所述处理器执行所述程序时,实现如权利要求1-7中任一所述的问题语句的处理方法。

16.一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1-7中任一所述的问题语句的处理方法。

17.一种计算机程序产品,其特征在于,包括计算机程序,所述计算机程序在被处理器执行时实现如权利要求1-7中任一所述的问题语句的处理方法。

问题语句的处理方法、装置、计算机设备及存储介质

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及一种问题语句的处理方法、装置、计算机设备及存储介质。

背景技术

[0002] 在智能对话系统中,用户咨询问题时往往会进行多次提问以获取更加准确的答案。然而,随着对话轮次的增加,用户在提出新问题时参考之前的对话内容进行表述。这种表述简短且口语化,通常以省略或者指代的方式代替对话内容中已经提及的信息。这些现象导致当前问题语句信息缺失,增加了系统理解用户意图的难度,严重影响了系统应答的准确性。因此,研究如何完善问题语句信息,具有重要意义。

发明内容

[0003] 本公开旨在至少在一定程度上解决相关技术中的技术问题之一。

[0004] 本公开第一方面实施例提出了一种问题语句的处理方法,包括:

[0005] 获取当前的问题语句及所述问题语句对应的历史会话;

[0006] 将所述问题语句及所述历史会话输入问题改写模型中,以获取所述问题语句对应的目标问题语句;

[0007] 计算所述目标问题语句与知识库中每个参考问题语句的匹配度,以确定与所述目标问题语句的匹配度最高的参考问题语句;

[0008] 返回所述参考问题语句对应的答案。

[0009] 本公开第二方面实施例提出了一种问题语句的处理装置,包括:

[0010] 第一获取模块,用于获取当前的问题语句及所述问题语句对应的历史会话;

[0011] 第二获取模块,用于将所述问题语句及所述历史会话输入问题改写模型中,以获取所述问题语句对应的目标问题语句;

[0012] 确定模块,用于计算所述目标问题语句与知识库中每个参考问题语句的匹配度,以确定与所述目标问题语句的匹配度最高的参考问题语句;

[0013] 返回模块,用于返回所述参考问题语句对应的答案。

[0014] 本公开第三方面实施例提出了一种计算机设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时,实现如本公开第一方面实施例提出的问题语句的处理方法。

[0015] 本公开第四方面实施例提出了一种非临时性计算机可读存储介质,存储有计算机程序,所述计算机程序被处理器执行时实现如本公开第一方面实施例提出的问题语句的处理方法。

[0016] 本公开第五方面实施例提出了一种计算机程序产品,当所述计算机程序产品中的指令处理器执行时,执行如本公开第一方面实施例提出的问题语句的处理方法。

[0017] 本公开提供的问题语句的处理方法、装置、计算机设备及存储介质,存在如下有益

效果：

[0018] 首先获取当前的问题语句及问题语句对应的历史会话,然后将问题语句及历史会话输入问题改写模型中,获取问题语句对应的目标问题语句,之后计算目标问题语句与知识库中每个参考问题语句的匹配度,最后返回与目标问题语句的匹配度最高的参考问题语句对应的答案。由此,完善了问题语句的省略信息,进而为后续的应答提供了语义完整的输入,保证了系统应答结果的准确性。

[0019] 本公开附加的方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本公开的实践了解到。

附图说明

[0020] 本公开上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0021] 图1为本公开一实施例所提供的问题语句的处理方法的流程示意图;

[0022] 图2为本公开另一实施例所提供的问题语句的处理方法的流程示意图;

[0023] 图3为本公开一实施例所提供的获取多个训练数据对的流程示意图;

[0024] 图4为本公开另一实施例所提供的获取多个训练数据对的流程示意图;

[0025] 图5为本公开一实施例所提供的问题语句的处理装置的结构示意图;

[0026] 图6为本公开另一实施例所提供的问题语句的处理装置的结构示意图;

[0027] 图7示出了适于用来实现本公开实施方式的示例性计算机设备的框图。

具体实施方式

[0028] 下面详细描述本公开的实施例,实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,旨在用于解释本公开,而不能理解为对本公开的限制。

[0029] 下面参考附图描述本公开实施例的问题语句的处理方法、装置、计算机设备和存储介质。

[0030] 图1为本公开实施例所提供的问题语句的处理方法的流程示意图。

[0031] 本公开实施例以该问题语句的处理方法被配置于问题语句的处理装置中来举例说明,该问题语句的处理装置可以应用于任一计算机设备中,以使该计算机设备可以执行问题语句的处理功能。

[0032] 其中,计算机设备可以为个人电脑(Personal Computer,简称PC)、云端设备、移动设备等,移动设备例如可以为手机、平板电脑、个人数字助理、穿戴式设备、车载设备等具有各种操作系统、触摸屏和/或显示屏的硬件设备。

[0033] 如图1所示,该问题语句的处理方法可以包括以下步骤:

[0034] 步骤101,获取当前的问题语句及问题语句对应的历史会话。

[0035] 其中,问题语句,可以为用户提出的任意问题,比如可以为用户咨询优惠活动的提问,或者用户咨询商品属性的提问,或者用户寻求情感帮助的提问等,本公开对此不做限定。

[0036] 需要说明的是,用户采用省略或指代的方式进行提问,往往是在和对话系统进行

多轮交互后。因此,当前的问题语句对应的历史会话,可以为前N个问题语句至当前的问题语句之间的对话,其中,N为正整数。

[0037] 可以理解的是,用户每提出一个问题,对话系统会做出相应的回答,用户的提问和系统的回答可以称为一轮对话。因此,前N个问题语句至当前的问题语句之间的对话,可以为在当前的问题语句提出前,用户与对话系统间发生的N轮对话。

[0038] 其中,N值可以根据需要设定。比如,历史会话可以为在当前的问题语句提出前,用户与对话系统间发生的2轮对话、4轮对话或6轮对话等,本公开对此不做限定。

[0039] 举例来说,用户咨询商品属性,用户与对话系统间发生的对话如下:

[0040] A:http://item.XX.com/27XX59.html

[0041] B:请问需要咨询该商品什么问题呢?

[0042] A:温度和无线标志中间有一个标志,是什么意思呀?

[0043] B:这个是电辅热的标志。

[0044] A:什么意思。

[0045] 此时,当前的问题语句可以为“什么意思”,若设定历史会话为在当前的问题语句提出前,用户与对话系统间发生的2轮对话,则对应的历史会话可以为“A:http://item.XX.com/27XX59.html.B:请问需要咨询该商品什么问题呢?A:温度和无线标志中间有一个标志,是什么意思呀?B:这个是电辅热的标志。”

[0046] 需要说明的是,在当前的问题语句提出前,若用户与对话系统间发生的对话轮次达不到设定的N值,则可以将用户与对话系统间发生的全部对话作为历史会话。

[0047] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中问题语句及历史会话等的限定。

[0048] 步骤102,将问题语句及历史会话输入问题改写模型中,以获取问题语句对应的目标问题语句。

[0049] 可以理解的是,用户在提问的问题语句里省略或指代的信息,通常已在历史会话中出现过。因此,通过将问题语句及对应的历史会话输入问题改写模型,可以使得问题改写模型根据历史会话对问题语句进行改写,以使改写后的目标问题语句包含完整信息。

[0050] 具体的,问题改写模型可以首先获取问题语句及历史会话的语义表示,然后根据语义表示从问题语句或历史会话中分别复制词语,形成一个语义完整且流畅的目标问题语句。

[0051] 比如,当用户提出的问题语句存在省略或指代的情况时,问题改写模型可以从历史会话中复制省略或指代的内容到问题语句中,以生成目标问题语句。

[0052] 举例来说,当前的问题语句可以为“什么意思”,对应的历史会话可以为“A:http://item.XX.com/27XX59.html.B:请问需要咨询该商品什么问题呢?A:温度和无线标志中间有一个标志,是什么意思呀?B:这个是电辅热的标志。”问题改写模型生成的目标问题语句可以为“电辅热是什么意思”,进而对话系统可以根据目标问题语句准确理解用户意图。

[0053] 可以理解的是,当用户提出的问题语句不存在省略或指代的情况时,问题改写模型可以将当前的问题语句作为目标问题语句输出。

[0054] 其中,问题改写模型的类型可以根据需要设定,比如可以为深度生成模型

Transformer等,本公开对此不做限定。

[0055] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中问题改写模型等的限定。

[0056] 步骤103,计算目标问题语句与知识库中每个参考问题语句的匹配度,以确定与目标问题语句的匹配度最高的参考问题语句。

[0057] 其中,知识库可以为智能对话系统预先设置的问答题库,包括多个参考问题语句及每个参考问题语句分别对应的答案。在用户提出问题后,智能对话系统可以根据用户的问题从知识库中检索匹配相似的问法,然后采用相似问法对应的答案应答用户。

[0058] 比如,可以根据目标问题语句的文本内容,首先计算目标问题语句与知识库中每个参考问题语句间的文本重复度;然后根据文本重复度,筛选与目标问题语句相似的多个参考问题语句;之后计算目标问题语句与每个相似的参考问题语句间的语义相似度;最后将语义相似度最高的参考问题语句,作为与目标问题语句的匹配度最高的参考问题语句。

[0059] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中匹配度等的限定。

[0060] 步骤104,返回参考问题语句对应的答案。

[0061] 其中,参考问题语句对应的答案,可以作为对话系统对用户提问问题的应答。

[0062] 本公开实施例中,首先获取当前的问题语句及问题语句对应的历史会话,然后将问题语句及历史会话输入问题改写模型中,获取问题语句对应的目标问题语句,之后计算目标问题语句与知识库中每个参考问题语句的匹配度,最后返回与目标问题语句的匹配度最高的参考问题语句对应的答案。由此,完善了问题语句的省略信息,进而为后续的应答提供了语义完整的输入,保证了系统应答结果的准确性。

[0063] 上述实施例,通过历史会话中的词语完善用户问题语句中省略和指代的内容,为系统应答提供了内容更加完整流畅的问句,以此提高了系统的应答能力。在一种可能的实现方式中,可以构建训练数据训练初始模型,从而生成问题改写模型,以提升改写结果的准确性。下面结合图2对上述过程进行进一步说明。

[0064] 步骤201,获取当前的问题语句及问题语句对应的历史会话。

[0065] 其中,步骤201的具体实现方式,可以参照本公开其他实施例的详细描述,在此不再赘述。

[0066] 步骤202,获取多个训练数据对,其中,每个训练数据对中包括:候选问题语句、候选问题语句对应的历史会话及对应的目标问题语句样本。

[0067] 其中,每个训练数据对,可以包括三种相互关联的训练数据,即候选问题语句、候选问题语句对应的历史会话及对应的目标问题语句样本。

[0068] 其中,候选问题语句,可以为用户在使用智能对话系统的过程中,可能提出的任意类型的问题语句。

[0069] 比如,候选问题语句,可以为省略部分信息的问题语句,或者为使用代词指代部分信息的问题语句,还可以为语义完整的问题语句,本公开对此不做限定。

[0070] 相应的,候选问题语句对应的历史会话,可以为对话日志中,前N个问题语句至候选问题语句之间的对话,也就是在候选问题语句提出前,用户与对话系统间发生的N轮对话。

[0071] 比如,历史会话可以为在候选问题语句提出前,用户与对话系统间发生的4轮对话,或者用户与对话系统间发生的6轮对话等,本公开对此不做限定。

[0072] 相应的,候选问题语句对应的目标问题语句样本,可以为与候选问题语句对应的语义完整的问题语句。

[0073] 比如,候选问题语句为省略部分信息的问题语句,则目标问题语句样本可以为对候选问题语句中的省略信息进行完善后的问题语句。

[0074] 或者,候选问题语句为使用代词指代部分信息的问题语句,则目标问题语句样本可以为使用具体词语对候选问题语句中的代词进行替换后的问题语句。

[0075] 或者,候选问题语句为语义完整的问题语句,则目标问题语句样本可以为与候选问题语句相同的问题语句。

[0076] 举例来说,候选问题语句对应的历史会话为:

[0077] A:怎么改订单?

[0078] B:只能取消订单,然后重新下单。

[0079] A:那么怎么取消订单?

[0080] B:在我的订单里面有取消入口。

[0081] A:我找不到。

[0082] 候选问题语句可以为“我找不到”,或者为“我找不到它”,或者为“我找不到取消入口”,对应的历史会话为“A:怎么改订单?B:只能取消订单,然后重新下单。A:那么怎么取消订单?B:在我的订单里面有取消入口。”目标问题语句样本可以为“我找不到取消入口”。

[0083] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中候选问题语句、历史会话、目标问题语句样本等的限定。

[0084] 步骤203,将候选问题语句及对应的历史会话,输入初始模型,以获取初始模型输出的预测语句。

[0085] 其中,初始模型,可以为任一预先构建的模型,比如Transformer模型等,本公开对此不做限定。

[0086] 具体的,初始模型可以对候选问题语句及对应的历史会话中的词语进行语义解析,以从历史会话中复制相关词语,对候选问题语句进行改写,形成预测语句。

[0087] 步骤204,根据预测语句与目标问题语句样本间的差异,对初始模型进行修订,以生成问题改写模型。

[0088] 比如,可以计算预测语句与目标问题语句样本间的误差,然后基于误差通过梯度下降的方法进行迭代求解,得到最小化的损失函数和模型参数值,生成问题改写模型。

[0089] 步骤205,将问题语句及历史会话输入问题改写模型中,以获取问题语句对应的目标问题语句。

[0090] 步骤206,计算目标问题语句与知识库中每个参考问题语句的匹配度,以确定与目标问题语句的匹配度最高的参考问题语句。

[0091] 步骤207,返回参考问题语句对应的答案。

[0092] 其中,步骤205-步骤207的具体实现方式,可以参照本公开其他实施例的详细描述,在此不再赘述。

[0093] 本公开实施例中,首先获取当前的问题语句及问题语句对应的历史会话,然后获

取多个训练数据对训练初始模型,从而生成问题改写模型;随后将问题语句及历史会话输入问题改写模型中,获取问题语句对应的目标问题语句,之后计算目标问题语句与知识库中每个参考问题语句的匹配度,最后返回与目标问题语句的匹配度最高的参考问题语句对应的答案。由此,提升了问题语句改写结果的准确性,为系统应答提供了内容更加完整流畅的问句,保证了系统应答结果的准确性。

[0094] 可以理解的是,训练数据是影响生成模型的准确性和可靠性的重要因素。因此,获取大量真实的对话数据对于模型的训练是十分重要的。

[0095] 在本公开实施例一种可能的实现方式中,如图3所示,获取多个训练数据对,可以包括以下步骤:

[0096] 步骤301,从对话日志中获取多个目标问题语句样本及目标问题语句样本对应的历史会话。

[0097] 其中,对话日志可以为智能对话系统中保存的历史对话记录等,本公开对此不做限定。

[0098] 可以理解的是,用户提出的问题语句的长度越长,问题语句的语义越完整。当长度超过一定范围后,问题语句会有冗余的情况。

[0099] 本公开实施例中,可以按照问题语句的长度,选取对话日志中语义完整的语句作为目标问题语句。

[0100] 比如,可以选取文本长度在15~30之间的问题语句和该问题语句对应的历史会话,作为模型训练的候选数据。

[0101] 此外,可以对候选数据做进一步过滤,仅保留问题语句和历史会话中存在相同词语的数据。其中,相同词语的词性可以为名词、动词或形容词,以保证相同词语具有业务含义。过滤后得到的数据可以作为目标问题语句样本及目标问题语句样本对应的历史会话。

[0102] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中目标问题语句样本及对应的历史会话的限定。

[0103] 步骤302,对目标问题语句样本及对应的历史会话进行解析,以确定参考词语。

[0104] 可以理解的是,用户在提问的问题语句里缺失的信息,通常已在历史会话中出现过。对问题语句进行改写的目的,是根据历史会话对用户提出的问题语句进行语义完善,以使对话系统能够准确理解用户提问的意图。

[0105] 因此,可以对目标问题语句样本及对应的历史会话进行解析,以确定目标问题语句样本和历史会话中存在的相同的词语,即在目标问题语句样本和历史会话中都出现过的词语,最终作为参考词语。

[0106] 步骤303,删除目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句。

[0107] 比如,目标问题语句样本可以为“我找不到取消入口”。对应的历史会话为“A:怎么改订单?B:只能取消订单,然后重新下单。A:那么怎么取消订单?B:在我的订单里面有取消入口。”

[0108] 此时,目标问题语句样本和历史会话中出现的相同的词语为“取消入口”,因此可以删除目标问题语句样本中的参考词语“取消入口”,得到候选问题语句可以为“我找不到”。

[0109] 步骤304,使用预设的词语替换目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句。

[0110] 需要说明的是,用户在提问的问题语句里,可能使用代词指代部分信息,比如它、这个、那个等。因此,为了使问题改写模型能够对使用指代的问题语句进行改写,可以将指代的词语作为预设的词语,替换目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句。

[0111] 比如,目标问题语句样本可以为“我找不到取消入口”。对应的历史会话为“A:怎么改订单?B:只能取消订单,然后重新下单。A:那么怎么取消订单?B:在我的订单里面有取消入口。”

[0112] 此时,可以使用预测的词语“它”替换目标问题语句样本中的参考词语“取消入口”,得到候选问题语句可以为“我找不到它”。

[0113] 步骤305,复制目标问题语句样本,以获取与目标问题语句样本对应的候选问题语句。

[0114] 需要说明的是,用户提出的问题语句本身可以具有完整的语义,此时,问题改写模型可以将用户原始的问题语句作为目标问题语句输出。因此,为了使问题改写模型能够对具有完整语义的问题语句进行处理,可以复制目标问题语句样本,作为目标问题语句样本对应的候选问题语句。

[0115] 比如,目标问题语句样本可以为“我找不到取消入口”。对应的历史会话为“A:怎么改订单?B:只能取消订单,然后重新下单。A:那么怎么取消订单?B:在我的订单里面有取消入口。”

[0116] 此时,可以复制目标问题语句样本,得到候选问题语句可以为“我找不到取消入口”。

[0117] 本公开实施例中,首先从对话日志中获取多个目标问题语句样本及目标问题语句样本对应的历史会话;然后对目标问题语句样本及对应的历史会话进行解析,以确定参考词语;最后删除目标问题语句样本中的参考词语,使用预设的词语替换目标问题语句样本中的参考词语,以及复制目标问题语句样本,以获取与目标问题语句样本对应的候选问题语句。由此,在无需任何标注成本和人力成本的情况下,可以有效获取大规模的训练数据,充分训练模型的语义表示能力和改写能力,大大提高了模型的准确性。

[0118] 需要说明的是,在智能对话系统的对话日志中,用户采用省略的问题语句进行提问,采用指代的问题语句进行提问,或者采用语义完整的问题语句进行提问的数据占比可能不同。

[0119] 因此,可以根据不同类型的问题语句的占比对训练数据进行分配,以使问题改写模型对不同类型的问题语句的改写结果更加准确。

[0120] 在本公开实施例一种可能的实现方式中,如图4所示,获取多个训练数据对,可以包括以下步骤:

[0121] 步骤401,根据对话日志中每个问题语句的长度,从对话日志中获取多个目标问题语句样本及目标问题语句样本对应的历史会话。

[0122] 步骤402,对目标问题语句样本及对应的历史会话进行解析,以确定参考词语。

[0123] 步骤401-步骤402的具体实现方式,可以参照本公开其他实施例的详细描述,在此

不再赘述。

[0124] 步骤403,为每个目标问题语句样本分别生成一个对应的随机概率。

[0125] 步骤404,在随机概率在第一数值范围内的情况下,删除目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句。

[0126] 步骤405,在随机概率在第二数值范围内的情况下,使用预设的词语替换目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句。

[0127] 步骤406,在随机概率在第三数值范围内的情况下,复制目标问题语句样本,以获取与目标问题语句样本对应的候选问题语句。

[0128] 其中,第一数值范围、第二数值范围及第三数值范围可以根据不同类型的问题语句的占比确定。

[0129] 比如,用户采用省略的问题语句进行提问的占比为25%,采用指代的问题语句进行提问的占比为25%,采用语义完整的问题语句进行提问的占比为50%。则第一数值范围可以为大于0小于等于25%,第二数值范围可以为大于25%小于等于50%,第三数值范围可以为大于50%小于等于100%。

[0130] 本公开实施例中,通过为每个目标问题语句样本分别生成一个对应的随机概率,然后根据随机概率的数值确定对目标问题语句样本进行删除、替换或复制的操作,最后使得不同类型的训练数据的分布,与实际的问题语句的占比保持一致,从而保证了问题改写模型,对不同类型的问题语句的改写结果的准确性。

[0131] 为了实现上述实施例,本公开还提出一种问题语句的处理装置。

[0132] 图5为本公开实施例所提供的问题语句的处理装置的结构示意图。

[0133] 如图5所示,该问题语句的处理装置100可以包括:第一获取模块110、第二获取模块120、确定模块130、返回模块140。

[0134] 其中,第一获取模块110,用于获取当前的问题语句及问题语句对应的历史会话。

[0135] 第二获取模块120,用于将问题语句及历史会话输入问题改写模型中,以获取问题语句对应的目标问题语句。

[0136] 确定模块130,用于计算目标问题语句与知识库中每个参考问题语句的匹配度,以确定与目标问题语句的匹配度最高的参考问题语句。

[0137] 返回模块140,用于返回参考问题语句对应的答案。

[0138] 在一种可能的实现方式中,历史会话包括前N个问题语句至当前的问题语句之间的对话,其中,N为正整数。

[0139] 本公开实施例中的上述各模块的功能及具体实现原理,可参照上述各方法实施例,此处不再赘述。

[0140] 本公开实施例的问题语句的处理装置,首先获取当前的问题语句及问题语句对应的历史会话,然后将问题语句及历史会话输入问题改写模型中,获取问题语句对应的目标问题语句,之后计算目标问题语句与知识库中每个参考问题语句的匹配度,最后返回与目标问题语句的匹配度最高的参考问题语句对应的答案。由此,完善了问题语句的省略信息,进而为后续的应答提供了语义完整的输入,保证了系统应答结果的准确性。

[0141] 进一步地,在本公开实施例的一种可能的实现方式中,参见图6,在图5所示实施例的基础上,还包括训练模块150。训练模块150包括:第一获取单元151、第二获取单元152及

生成单元153。

[0142] 其中,第一获取单元151,用于获取多个训练数据对,其中,每个训练数据对中包括:候选问题语句、候选问题语句对应的历史会话及对应的目标问题语句样本。

[0143] 第二获取单元152,用于将候选问题语句及对应的历史会话,输入初始模型,以获取初始模型输出的预测语句。

[0144] 生成单元153,用于根据预测语句与目标问题语句样本间的差异,对初始模型进行修订,以生成问题改写模型。

[0145] 在一种可能的实现方式中,第一获取单元151用于:

[0146] 从对话日志中获取多个目标问题语句样本及目标问题语句样本对应的历史会话;

[0147] 对目标问题语句样本及对应的历史会话进行解析,以确定参考词语;

[0148] 删除目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句。

[0149] 在一种可能的实现方式中,第一获取单元151还用于:使用预设的词语替换目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句。

[0150] 在一种可能的实现方式中,第一获取单元151还用于:复制目标问题语句样本,以获取与目标问题语句样本对应的候选问题语句。

[0151] 在一种可能的实现方式中,第一获取单元151用于:

[0152] 根据对话日志中每个问题语句的长度,从对话日志中获取多个目标问题语句样本及目标问题语句样本对应的历史会话;

[0153] 对目标问题语句样本及对应的历史会话进行解析,以确定参考词语;

[0154] 为每个目标问题语句样本分别生成一个对应的随机概率;

[0155] 在随机概率在第一数值范围内的情况下,删除目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句;

[0156] 在随机概率在第二数值范围内的情况下,使用预设的词语替换目标问题语句样本中的参考词语,以获取与目标问题语句样本对应的候选问题语句;

[0157] 在随机概率在第三数值范围内的情况下,复制目标问题语句样本,以获取与目标问题语句样本对应的候选问题语句。

[0158] 本公开实施例中的上述各模块的功能及具体实现原理,可参照上述各方法实施例,此处不再赘述。

[0159] 本公开实施例的问题语句的处理装置,首先获取当前的问题语句及问题语句对应的历史会话,然后获取多个训练数据对训练初始模型,从而生成问题改写模型;随后将问题语句及历史会话输入问题改写模型中,获取问题语句对应的目标问题语句,之后计算目标问题语句与知识库中每个参考问题语句的匹配度,最后返回与目标问题语句的匹配度最高的参考问题语句对应的答案。由此,提升了问题语句改写结果的准确性,为系统应答提供了内容更加完整流畅的问句,保证了系统应答结果的准确性。

[0160] 为了实现上述实施例,本公开还提出一种计算机设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,处理器执行程序时,实现如本公开前述实施例提出的问题语句的处理方法。

[0161] 为了实现上述实施例,本公开还提出一种非临时性计算机可读存储介质,存储有

计算机程序,计算机程序被处理器执行时实现如本公开前述实施例提出的问题语句的处理方法。

[0162] 为了实现上述实施例,本公开还提出一种计算机程序产品,当计算机程序产品中的指令处理器执行时,执行如本公开前述实施例提出的问题语句的处理方法。

[0163] 图7示出了适于用来实现本公开实施方式的示例性计算机设备的框图。图7显示的计算机设备12仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0164] 如图7所示,计算机设备12以通用计算设备的形式表现。计算机设备12的组件可以包括但不限于:一个或者多个处理器或者处理单元16,系统存储器28,连接不同系统组件(包括系统存储器28和处理单元16)的总线18。

[0165] 总线18表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构(Industry Standard Architecture;以下简称:ISA)总线,微通道体系结构(Micro Channel Architecture;以下简称:MAC)总线,增强型ISA总线、视频电子标准协会(Video Electronics Standards Association;以下简称:VESA)局域总线以及外围组件互连(Peripheral Component Interconnection;以下简称:PCI)总线。

[0166] 计算机设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机设备12访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0167] 存储器28可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器(Random Access Memory;以下简称:RAM) 30和/或高速缓存存储器32。计算机设备12可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统34可以用于读写不可移动的、非易失性磁介质(图7未显示,通常称为“硬盘驱动器”)。尽管图7中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如:光盘只读存储器(Compact Disc Read Only Memory;以下简称:CD-ROM)、数字多功能只读光盘(Digital Video Disc Read Only Memory;以下简称:DVD-ROM)或其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线18相连。存储器28可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本公开各实施例的功能。

[0168] 具有一组(至少一个)程序模块42的程序/实用工具40,可以存储在例如存储器28中,这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块42通常执行本公开所描述的实施例中的功能和/或方法。

[0169] 计算机设备12也可以与一个或多个外部设备14(例如键盘、指向设备、显示器24等)通信,还可与一个或者多个使得用户能与该计算机设备12交互的设备通信,和/或与使得该计算机设备12能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口22进行。并且,计算机设备12还可以通过网络适配器20与一个或者多个网络(例如局域网(Local Area Network;以下简称:LAN),广域网(Wide Area Network;以下简称:WAN)和/或公共网络,例如因特网)通信。如图

所示,网络适配器20通过总线18与计算机设备12的其它模块通信。应当明白,尽管图中未示出,可以结合计算机设备12使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0170] 处理单元16通过运行存储在系统存储器28中的程序,从而执行各种功能应用以及数据处理,例如实现前述实施例中提及的方法。

[0171] 本公开的技术方案,首先获取当前的问题语句及问题语句对应的历史会话,然后将问题语句及历史会话输入问题改写模型中,获取问题语句对应的目标问题语句,之后计算目标问题语句与知识库中每个参考问题语句的匹配度,最后返回与目标问题语句的匹配度最高的参考问题语句对应的答案。由此,完善了问题语句的省略信息,进而为后续的应答提供了语义完整的输入,保证了系统应答结果的准确性。

[0172] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本公开的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0173] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本公开的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0174] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现定制逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本公开的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本公开的实施例所属技术领域的技术人员所理解。

[0175] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤,例如,可以被认为用于实现逻辑功能的可执行指令的定序列表,可以具体实现在任何计算机可读介质中,以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用,或结合这些指令执行系统、装置或设备而使用。就本说明书而言,“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例(非穷尽性列表)包括以下:具有一个或多个布线的电连接部(电子装置),便携式计算机盘盒(磁装置),随机存取存储器(RAM),只读存储器(ROM),可擦除可编程只读存储器(EPROM或闪速存储器),光纤装置,以及便携式光盘只读存储器(CDROM)。另外,计算机可读介质甚至可以是可在其上打印程序的纸或其他合适的介质,因为可以例如通过对纸或其他介质进行光学扫描,接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得程序,然后将其存储在计算机存储器中。

[0176] 应当理解,本公开的各部分可以用硬件、软件、固件或它们的组合来实现。在上述

实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。如,如果用硬件来实现和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或他们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列(PGA),现场可编程门阵列(FPGA)等。

[0177] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成,的程序可以存储于一种计算机可读存储介质中,该程序在执行时,包括方法实施例的步骤之一或其组合。

[0178] 此外,在本公开各个实施例中的各功能单元可以集成在一个处理模块中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机可读存储介质中。

[0179] 上述提到的存储介质可以是只读存储器,磁盘或光盘等。尽管上面已经示出和描述了本公开的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本公开的限制,本领域的普通技术人员在本公开的范围内可以对上述实施例进行变化、修改、替换和变型。

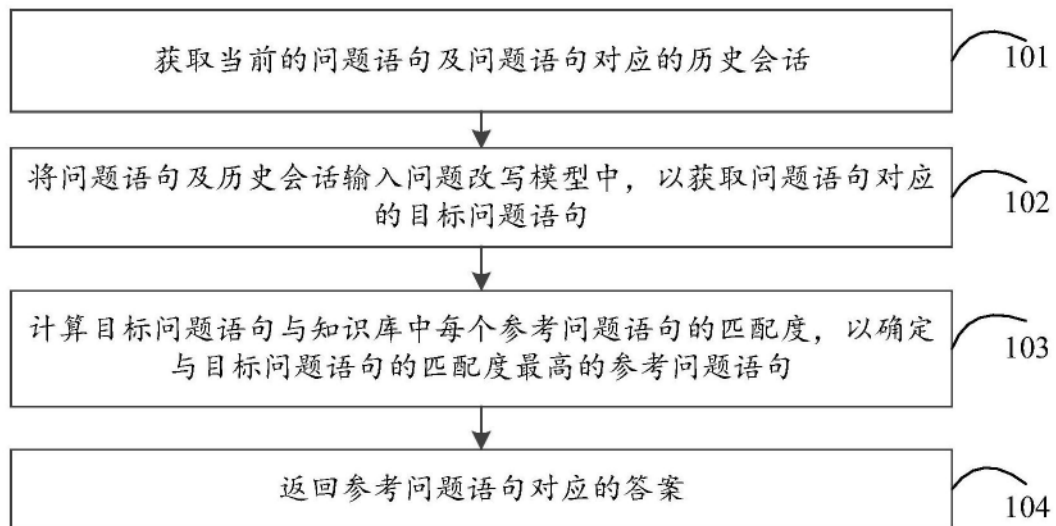


图1

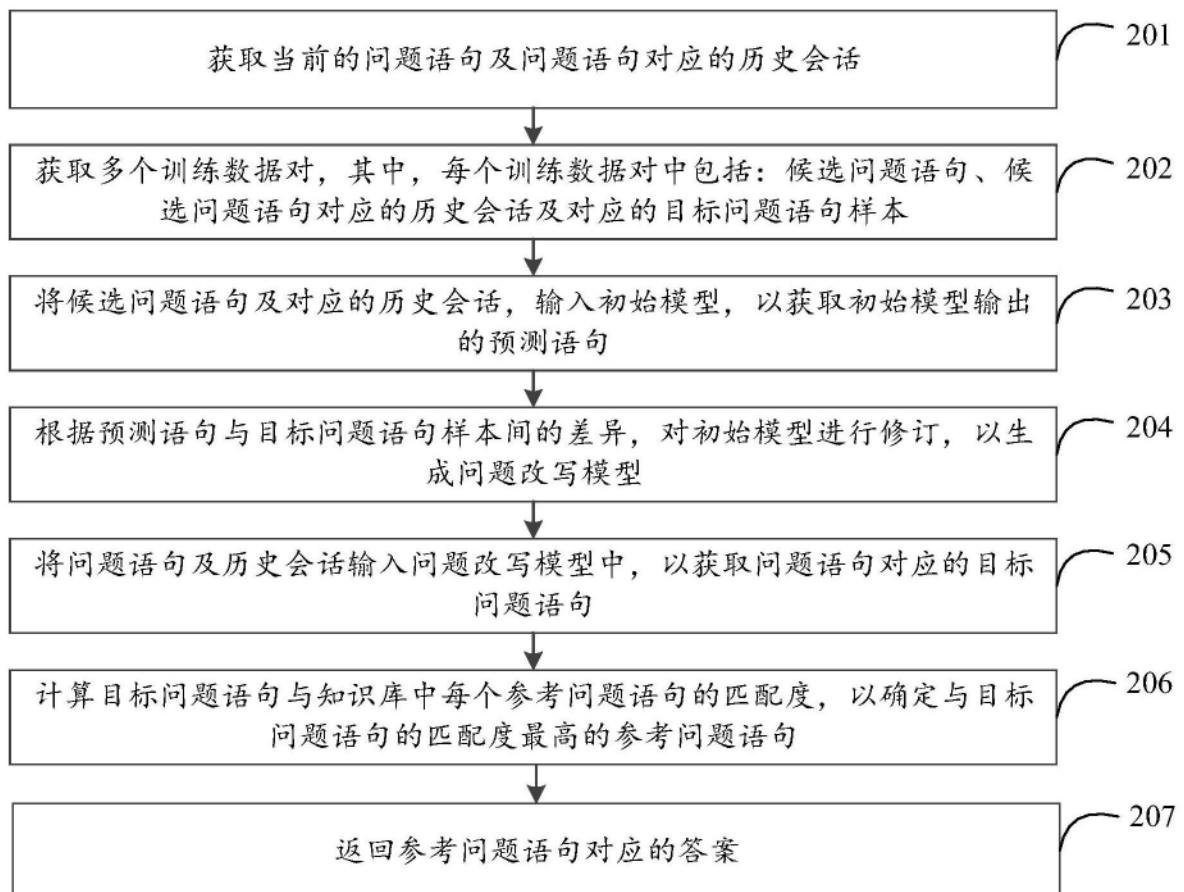


图2

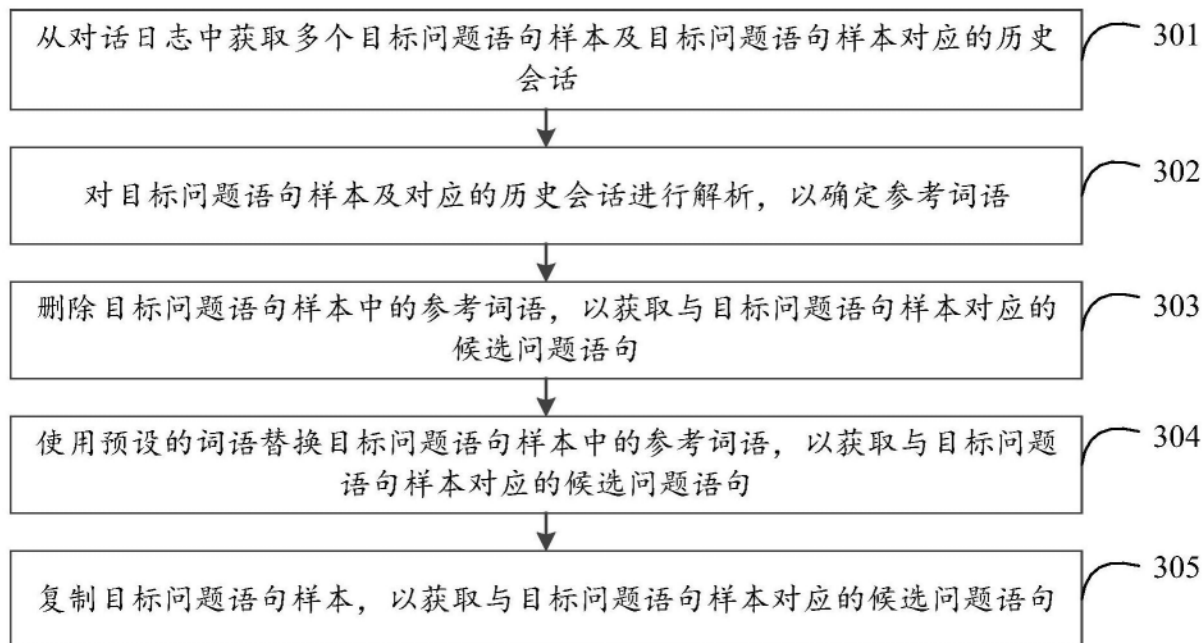


图3

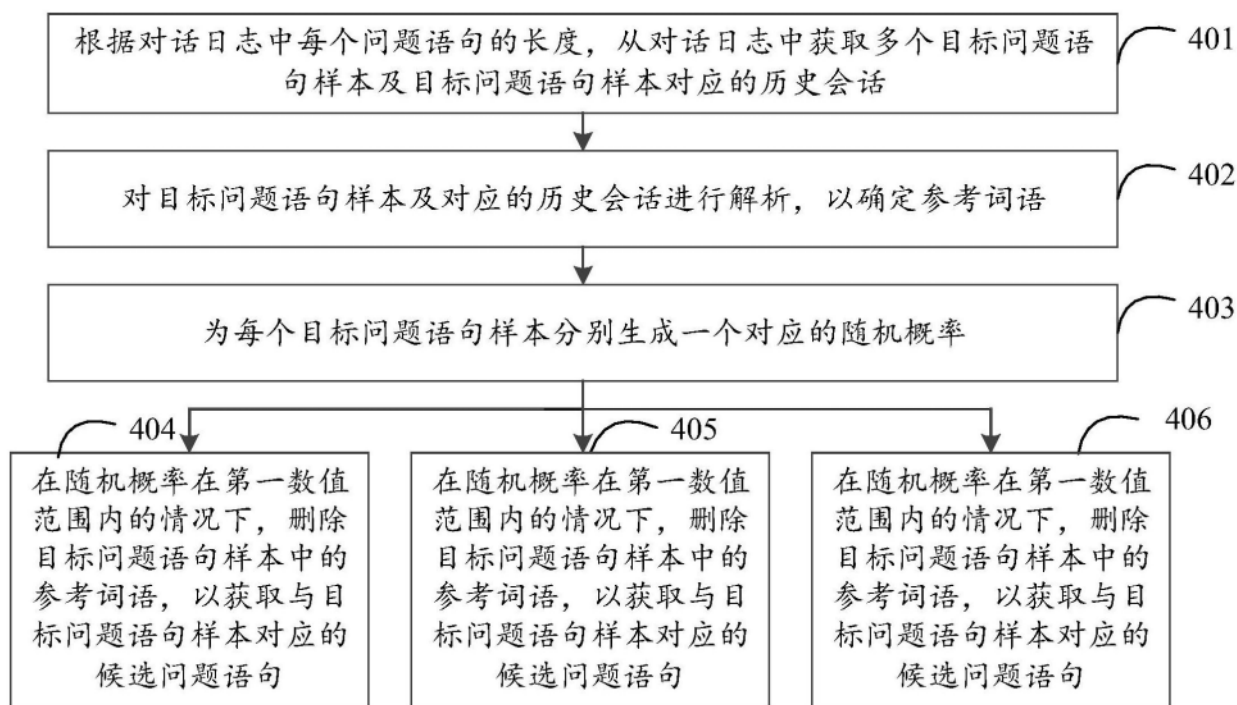


图4

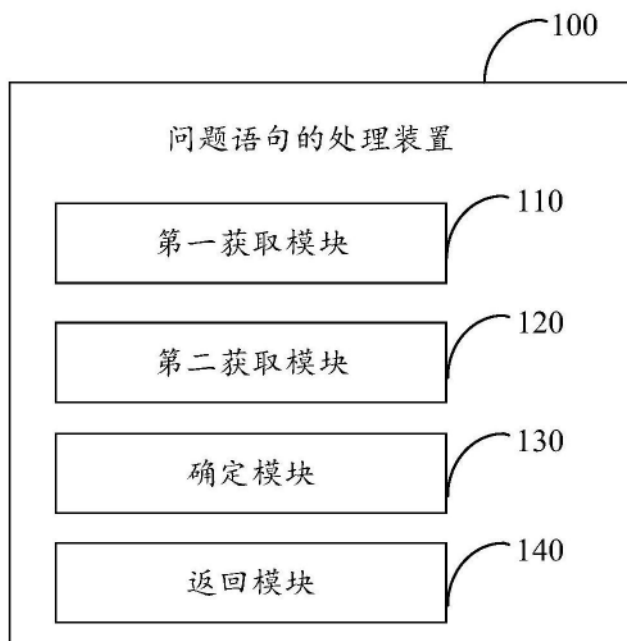


图5

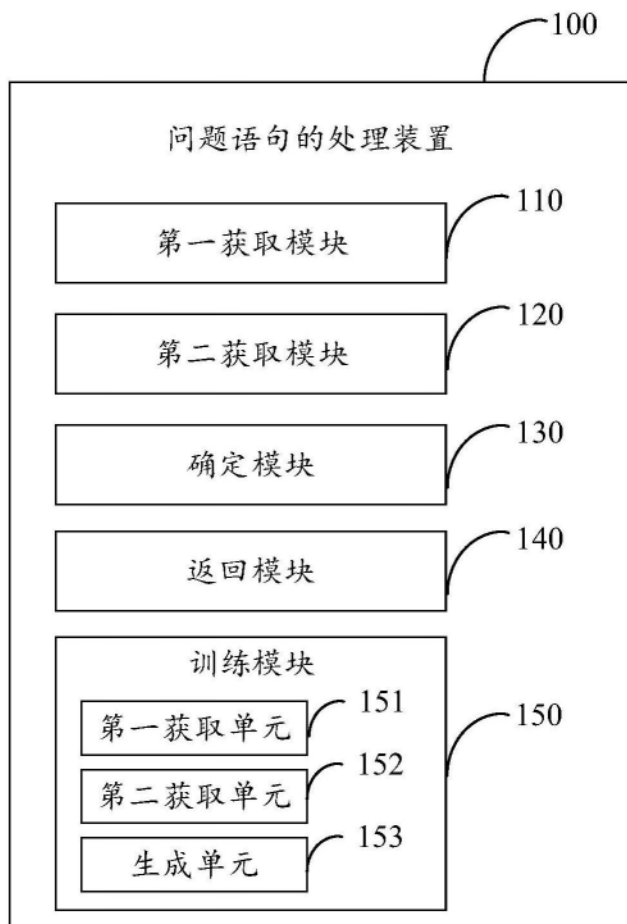


图6

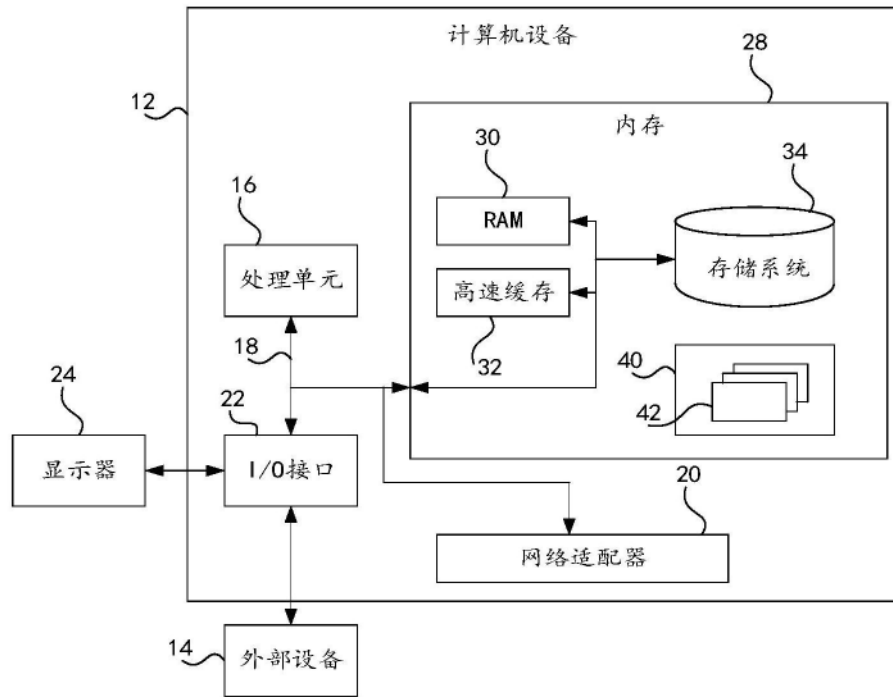


图7