



(12) 发明专利

(10) 授权公告号 CN 110647627 B

(45) 授权公告日 2022.05.27

(21) 申请号 201910720156.X

G06F 16/332 (2019.01)

(22) 申请日 2019.08.06

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 110647627 A

CN 107209861 A, 2017.09.26

CN 104268134 A, 2015.01.07

CN 108681749 A, 2018.10.19

(43) 申请公布日 2020.01.03

CN 102439595 A, 2012.05.02

CN 109815341 A, 2019.05.28

(73) 专利权人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号
百度大厦2层

CN 103810218 A, 2014.05.21

CN 106776532 A, 2017.05.31

CN 109657127 A, 2019.04.19

(72) 发明人 焦振宇 古桂元 孙叔琦 李婷婷
孙珂

US 2019087728 A1, 2019.03.21

US 2016071127 A1, 2016.03.10

(74) 专利代理机构 北京鸿德海业知识产权代理
有限公司 11412
专利代理师 田宏宾

罗宇翔等. 一种基于邮件列表的软件问答信息抽取方法.《计算机科学》.2015,第23-25页,第35页.

(51) Int. Cl.

审查员 唐文俊

G06F 16/35 (2019.01)

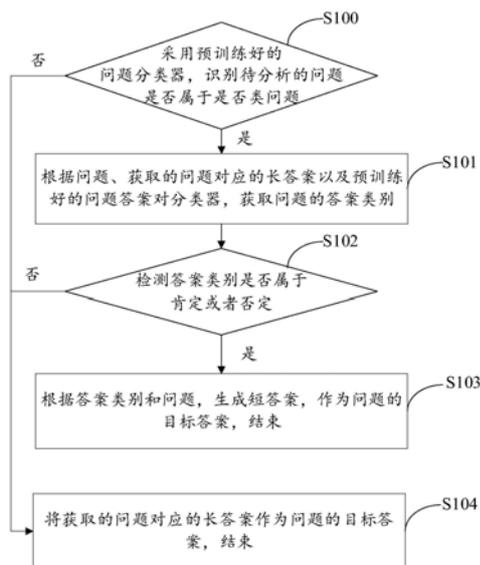
权利要求书2页 说明书9页 附图3页

(54) 发明名称

答案生成方法及装置、计算机设备与可读介质

(57) 摘要

本发明提供一种答案生成方法及装置、计算机设备与可读介质。其方法包括：采用预训练好的问题分类器，识别待分析的问题是否属于是否类问题，若属于，根据问题、获取的问题对应的长答案以及预训练好的问题答案对分类器，获取问题的答案类别；若答案类别属于肯定或者否定，根据答案类别和问题，生成短答案，作为问题的目标答案。本发明的技术方案，可以对是否类问题生成简洁、明了的短答案，与现有的人机对话系统提供的长答案相比，可以减少答案中的冗余信息，增强答案的精准度；能够进一步缩短用户阅读答案的时间，减少用户的阅读成本，从而能够有效地增强用户的使用体验度。



1. 一种答案生成方法,其特征在于,所述方法包括:
 - 采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题;
 - 若属于,根据所述问题、获取的所述问题对应的长答案以及预训练好的问题答案对分类器,获取所述问题的答案类别;
 - 检测所述答案类别是否属于肯定或者否定;
 - 若是,根据所述答案类别和所述问题,生成短答案,作为所述问题的目标答案;
 - 根据所述答案类别和所述问题,生成短答案,包括:
 - 对所述问题进行分析,获取生成所述短答案所需的目标信息内容;
 - 根据所述答案类别和所述目标信息内容,生成所述短答案。
2. 根据权利要求1所述的方法,其特征在于,若所述问题不属于是否类问题,所述方法包括:
 - 将获取的所述问题对应的长答案作为所述问题的目标答案。
3. 根据权利要求1所述的方法,其特征在于,若所述答案类别不是肯定或者否定,所述方法还包括:
 - 将获取的所述问题对应的长答案作为所述问题的目标答案。
4. 根据权利要求1所述的方法,其特征在于,对所述问题进行分析,获取生成所述短答案所需的目标信息内容,包括:
 - 对所述问题分别进行分词,得到对应的词序列;
 - 根据预设的多条规则和所述词序列,获取生成所述短答案所需的目标信息内容;
 - 或者,对所述问题进行分析,获取生成所述短答案所需的信息内容,包括:
 - 对所述问题分别进行分词和词性标注,得到对应的词序列和词性序列;
 - 根据预设的多条规则、所述词序列和所述词性序列,获取生成所述短答案所需的目标信息内容。
5. 根据权利要求4所述的方法,其特征在于,根据预设的多条规则和所述词序列,获取生成所述短答案所需的目标信息内容,包括:
 - 分别采用各条所述规则对所述词序列进行匹配处理,获取能够匹配到的数条原始信息内容和各条所述原始信息内容对应的所述规则;
 - 根据各条所述规则的优先级,从数条所述原始信息内容中获取对应的所述规则优先级最高的所述原始信息内容,作为所述目标信息内容;
 - 根据预设的多条规则、所述词序列和所述词性序列,获取生成所述短答案所需的目标信息内容,包括:
 - 分别采用各条所述规则对所述词序列和所述词性序列进行匹配处理,获取能够匹配到的数条原始信息内容和各条所述原始信息内容对应的所述规则;
 - 根据各条所述规则的优先级,从数条所述原始信息内容中获取对应的所述规则优先级最高的所述原始信息内容,作为所述目标信息内容。
6. 一种答案生成装置,其特征在于,所述装置包括:
 - 识别模块,用于采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题;
 - 获取模块,用于若所述识别模块识别所述问题属于是否类问题,根据所述问题、获取的所述问题对应的长答案以及预训练好的问题答案对分类器,获取所述问题的答案类别;

检测模块,用于检测所述答案类别是否属于肯定或者否定;

生成模块,用于若所述检测模块确定所述答案类别属于肯定或者否定,根据所述答案类别和所述问题,生成短答案,作为所述问题的目标答案;

所述生成模块,还包括:

获取单元,用于对所述问题进行分析,获取生成所述短答案所需的目标信息内容;

生成单元,用于根据所述答案类别和所述目标信息内容,生成所述短答案。

7. 根据权利要求6所述的装置,其特征在于,所述装置还包括:

设置模块,用于若所述识别模块识别所述问题不属于是否类问题,将获取的所述问题对应的长答案作为所述问题的目标答案。

8. 根据权利要求7所述的装置,其特征在于,所述设置模块,还用于若所述检测模块确定所述答案类别不属于肯定或者否定,将获取的所述问题对应的长答案作为所述问题的目标答案。

9. 根据权利要求6所述的装置,其特征在于,所述获取单元,用于:

对所述问题分别进行分词,得到对应的词序列;

根据预设的多条规则和所述词序列,获取生成所述短答案所需的目标信息内容;

或者,所述获取单元,用于:

对所述问题分别进行分词和词性标注,得到对应的词序列和词性序列;

根据预设的多条规则、所述词序列和所述词性序列,获取生成所述短答案所需的目标信息内容。

10. 根据权利要求9所述的装置,其特征在于,所述获取单元,具体用于:

分别采用各条所述规则对所述词序列进行匹配处理,获取能够匹配到的数条原始信息内容和各条所述原始信息内容对应的所述规则;

根据各条所述规则的优先级,从数条所述原始信息内容中获取对应的所述规则优先级最高的所述原始信息内容,作为所述目标信息内容;

或者所述获取单元,具体用于:

分别采用各条所述规则对所述词序列和所述词性序列进行匹配处理,获取能够匹配到的数条原始信息内容和各条所述原始信息内容对应的所述规则;

根据各条所述规则的优先级,从数条所述原始信息内容中获取对应的所述规则优先级最高的所述原始信息内容,作为所述目标信息内容。

11. 一种人机对话系统,所述系统中包括如上权利要求6-10任一所述的答案生成装置。

12. 一种计算机设备,其特征在于,所述设备包括:

一个或多个处理器;

存储器,用于存储一个或多个程序;

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-5中任一所述的方法。

13. 一种计算机可读介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现如权利要求1-5中任一所述的方法。

答案生成方法及装置、计算机设备与可读介质

【技术领域】

[0001] 本发明涉及计算机应用技术领域,尤其涉及一种答案生成方法及装置、计算机设备与可读介质。

【背景技术】

[0002] 在人机对话系统中,为满足用户的不同需求,对话机器人往往需要回答用户提出的各种各样的问题,例如“珠穆朗玛峰有多高”、“姚明拿过NBA冠军吗”、“怎么把手机中的照片传到电脑上”等,用户提出的问题中包含大量的的是否类问题,即可以通过具有明确肯定或否定语义的简短答案回答的问题。

[0003] 提出是否类问题的用户所期待的答案往往是一个简单且直接的肯定或否定应答。然而,目前的人机对话系统中,并不会对这类是否类问题进行特殊处理,而是按照统一的问题处理方式,像对待其他类型的问题一样,生成并返回一个包含若干句话的长答案的答复。这样,对于提出是否类问题的而用户,需要用户自己从长答案中归纳出肯定或否定的结果。因此,现有的人机对话系统生成是否类问题的答案包含过多的冗余信息,精准度较差。

【发明内容】

[0004] 本发明提供了一种答案生成方法及装置、计算机设备与可读介质,用于提高是否类为题对应的答案的精准度。

[0005] 本发明提供一种答案生成方法,所述方法包括:

[0006] 采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题;

[0007] 若属于,根据所述问题、获取的所述问题对应的长答案以及预训练好的问题答案对分类器,获取所述问题的答案类别;

[0008] 检测所述答案类别是否属于肯定或者否定;

[0009] 若是,根据所述答案类别和所述问题,生成短答案,作为所述问题的目标答案。

[0010] 本发明提供一种答案生成装置,所述装置包括:

[0011] 识别模块,用于采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题;

[0012] 获取模块,用于若所述识别模块识别所述问题属于是否类问题,根据所述问题、获取的所述问题对应的长答案以及预训练好的问题答案对分类器,获取所述问题的答案类别;

[0013] 检测模块,用于检测所述答案类别是否属于肯定或者否定;

[0014] 生成模块,用于若所述检测模块确定所述答案类别属于肯定或者否定,根据所述答案类别和所述问题,生成短答案,作为所述问题的目标答案。

[0015] 本发明还提供一种人机对话系统,所述系统中包括如所述的答案生成装置。

[0016] 本发明还提供一种计算机设备,所述设备包括:

[0017] 一个或多个处理器;

[0018] 存储器,用于存储一个或多个程序;

[0019] 当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如上所述的答案生成方法。

[0020] 本发明还提供一种计算机可读介质,其上存储有计算机程序,该程序被处理器执行时实现如上所述的答案生成方法。

[0021] 本发明的答案生成方法及装置、计算机设备与可读介质,通过采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题,若属于,根据问题、获取的问题对应的长答案以及预训练好的问题答案对分类器,获取问题的答案类别;若答案类别属于肯定或者否定,根据答案类别和问题,生成短答案,作为问题的目标答案。本发明的技术方案,可以对是否类问题生成简洁、明了的短答案,与现有的人机对话系统提供的长答案相比,可以减少答案中的冗余信息,增强答案的精准度;能够进一步缩短用户阅读答案的时间,减少用户的阅读成本,从而能够有效地增强用户的使用体验度。

【附图说明】

[0022] 图1为本发明的答案生成方法实施例的流程图。

[0023] 图2为本发明的答案生成装置实施例一的结构图。

[0024] 图3为本发明的答案生成装置实施例二的结构图。

[0025] 图4为本发明的计算机设备实施例的结构图。

[0026] 图5为本发明提供的一种计算机设备的示例图。

【具体实施方式】

[0027] 为了使本发明的目的、技术方案和优点更加清楚,下面结合附图和具体实施例对本发明进行详细描述。

[0028] 图1为本发明的答案生成方法实施例的流程图。如图1所示,本实施例的答案生成方法,具体可以包括如下步骤:

[0029] S100、采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题;若属于,执行步骤S101;若不属于,执行步骤S104;

[0030] 本实施例的答案生成方法的执行主体为答案生成装置,该答案生成装置可以应用于人机对话系统中,用于对人机对话系统中的是否类问题的答案进行进一步地处理,使得是否类问题的答案更加精准、简要。

[0031] 首先,本实施例中,需要预先训练一个问题分类器,该问题分类器可以采用神经网络模型来实现,例如可以基于Bidirectional Encoder Representations from Transformers (BERT) 模型来实现。具体地,该问题分类器用于来实现问题是否属于是否类问题。本实施例的是否类问题是一种较为宽泛的问题,不仅仅指问题的语句中包括是否。例如,这朵花漂亮吗?今天好玩吗?今天开不开心?等等之类的问题中虽然没有是否这两个字,但是回答时,根据问题中的句意,都可以归纳为是否类问题。本实施例的问题分类器,可以对任何是否类问题进行准确识别。

[0032] 本实施例的问题分类器,在训练时,可以采集数条各种语句格式的是否类问题和不属于是否类的问题,作为训练问题,并标注每一个训练问题是否属于是否类问题。本实施

例中,为了便于描述,仅将问题的类型分为两种类型,是否类问题和不是是否类问题。然后采用数条训练问题和标注的训练问题的类型,训练问题分类器。训练时,将每一个训练问题输入到问题分类器中,该问题分类器可以预测出该训练问题的类型,然后比对预测的训练问题的类型和标注的训练问题的类型是否一致,若不一致,调整问题分类器的参数,使得预测的训练问题的类型和标注的训练问题的类型一致。采用数条训练问题和对应的标注的训练问题的类型,采用上述方式,不断地对问题分类器进行训练,直至训练次数达到预设次数阈值,或者在连续的训练轮数中,预测的训练问题的类型和标注的训练问题的类型始终一致,此时确定问题分类器的参数,进而确定问题分类器,问题分类器的训练结束。

[0033] 本实施例的训练问题的条数根据需求,可以达到百万级别,训练问题的条数越多,训练的问题分类器的准确性越高。

[0034] 使用时,直接将问题输入至训练好的问题分类器中,该问题分类器可以预测并输出该问题的类型,本实施例中可以仅包括两种类型,一种为是否类问题,另一类为其他问题。

[0035] S101、根据问题、获取的问题对应的长答案以及预训练好的问题答案对分类器,获取问题的答案类别;

[0036] 本实施例的问题的长答案,可以从现有的人机对话系统中获取,或者基于现有的人机对话系统获取长答案的方式来获取。也就是说,现有的人机对话系统获取的答案都是包括较多信息的长答案,对于是否类问题而言,包括较多的冗余信息,不够精简。

[0037] 本实施例中预训练的问题答案对分类器也采用神经网络模型来实现,例如也可以基于BERT模型来实现。该预训练好的问题答案对分类器,用于来实现对问题的答案的类别进行识别。例如,本实施例中,问题的答案类别可以分为Yes(肯定)、No(否定)、Depends(视情况而定)、No_Opinion(无观点)。其中Yes和No类别明确包含肯定和否定语义的答案,而Depends和No_Opinion被认为不包含明确的肯定和否定语义。当然,实际应用中,也可以直接将问题的答案类别分为肯定、否定和其他三大类。

[0038] 本实施例的问题答案对分类器,在训练时,可以采集数条训练数据,每条训练数据中包括训练问题、训练问题对应的训练长答案以及已知的答案类别。训练时,将每一条训练数据中的训练问题和对应的训练长答案输入至问题答案对分类器中,该问题答案对分类器可以预测出该训练问题的答案类别,然后比对预测的训练问题的答案类别和已知的答案类别是否一致,若不一致,调整问题答案对分类器的参数,使得预测的训练问题的答案类别和已知的训练问题的答案类别一致。采用数条训练数据中的训练问题、训练问题对应的训练长答案以及已知的答案类别,按照上述方式,不断地对问题答案对分类器进行训练,直至训练次数达到预设次数阈值,或者在连续的训练轮数中,预测的训练问题的答案类别和已知的答案类别始终一致,此时确定问题答案对分类器的参数,进而确定问题答案对分类器,问题答案对分类器的训练结束。

[0039] 同理,本实施例的训练数据的条数根据需求,可以达到百万级别,训练数据的条数越多,训练的问题答案对分类器的准确性越高。

[0040] 使用时,将问题、获取的问题对应的长答案输入至预训练好的问题答案对分类器中,问题答案对分类器便可以预测并输出该问题的答案类别。且该答案类别中至少包括肯定、否定。

[0041] S102、检测答案类别是否属于肯定或者否定；若是，执行步骤S103；若不属于，执行步骤S104；

[0042] S103、根据答案类别和问题，生成短答案，作为问题的目标答案，结束。

[0043] S104、将获取的问题对应的长答案作为问题的目标答案，结束。

[0044] 本实施例中，主要用于对是否问题生成短答案，该短答案简洁、明了，用户一眼就能够看明白。所以，当答案类别不是肯定或者否定时，此时还是按照现有的人机对话系统获取长答案的方式，获取问题对应的长答案，作为问题的目标答案，并输出。而对于本实施例关注的是否类问题的答案也是肯定或者否定时，此时确定，可以生成简要的短答案。具体地，可以根据答案类别和问题，生成短答案，作为问题的目标答案。

[0045] 例如，进一步可选地，本实施例的该步骤S103“根据答案类别和问题，生成短答案，作为问题的目标答案”，具体可以包括如下步骤：

[0046] A、对问题进行分析，获取生成短答案所需的目标信息内容；

[0047] B、根据答案类别和目标信息内容，生成短答案。

[0048] 例如，对于问题“今天好玩吗”，人机对话系统中的长答案可能是“今天很好玩”或者“今天不好玩”。而在本实施例中，为了避免答案中包括过多的冗余信息，直接生成短答案。而为了与问题的描述一致，本实施例中还需要获取短答案所需的目标信息内容。例如，若答案类别是肯定的，则生成答案所需的目标信息内容可以是“好玩吗”。此时结合目标信息内容和肯定答案，则生成的短答案可以是“好玩”。若否定答案，对应地生成的短答案可以为“不好玩”。再例如，对于问题“可不可以借你的笔用一下”，人机对话系统中的长答案可能是“当然可以了，随使用吧”，或者“不好意思，我正在用，您稍等一会吧”等等之类的长答案。而在本实施例中，若答案类别是肯定的，生成短答案所需的目标信息内容可以是“可不可以”，而若答案类别是肯定的，生成的短答案直接就是“可以”；而若答案类别是否定，生成的短答案直接就是“不可以”。与长答案相比，本实施例的短答案简洁、明了，用户不用花太多时间阅读和理解，缩短了用户的阅读时间，减少了用户的阅读成本，能够有效地增强用户的使用体验度。

[0049] 基于以上描述，可以知道，本实施例中，生成短答案时，为了与问题的场景描述一致，不能出现问题问的是“可不可以”，而答案回答的是“是或者不是”之类的不对应的情况，本实施例中，还需要对问题进行分析，获取生成短答案所需的目标信息内容；然后才能根据答案类别和目标信息内容，生成短答案。

[0050] 其中步骤A对问题进行分析，获取生成短答案所需的目标信息内容，具体可以包括如下两种实现方式：

[0051] 第一种实现方式包括如下步骤：

[0052] (a1) 对问题分别进行分词，得到对应的词序列；

[0053] (b1) 根据预设的多条规则和词序列，获取生成短答案所需的目标信息内容。

[0054] 本实施例中，可以按照现有相关的分词技术对问题进行分词，得到对应的词序列。然后采用预先设置的多条规则分别匹配词序列，从中获取生成短答案所需的目标信息内容。例如，可以分别采用各条规则对词序列进行匹配处理，获取能够匹配到的数条原始信息内容和各条原始信息内容对应的规则；根据各条规则的优先级，从数条原始信息内容中获取对应的规则优先级最高的原始信息内容，作为目标信息内容。也就是说，预先设置规则的

时候,同时需要为每条规则配置对应的优先级。本实施例的规则可以为基于词的规则,例如可以采用正则表达式、关键字的方式定义,例如某条规则可以为“(.)不\1”,可以匹配到需要不需要、可以不可以,能不能等模式。这样,后续若是肯定的答案类别,结合答案类别,可以生成需要、可以、能作为肯定回答;而若是否定的答案类别,可以生成不需要、可以、能作为否定回答。当然,实际应用中,若预先设置的规则没有优先级,可以随机获取一条规则匹配的信息内容作为目标信息内容。还可以按照其他方式过滤出一条信息内容作为目标信息内容,在此不再一一举例赘述。

[0055] 第二种实现方式包括如下步骤:

[0056] (a2) 对问题分别进行分词和词性标注,得到对应的词序列和词性序列;

[0057] (b2) 根据预设的多条规则、词序列和词性序列,获取生成短答案所需的目标信息内容。

[0058] 本实现方式与上述第一种实现方式的区别在于:本实现方式中,在对问题分别进行分词后,还进一步进行词性标注,即对词序列中的每一个词进行词性标注,得到词性序列。此时对应的预设的多条规则中不仅包括有第一种实现方式中的关于词的规则,还可以增加有关于词性的规则。由于不同语言表达中,不同词性的词可能具备相同的功能,使用基于词性的规则可以提升短答案生成的泛化能力。例如某个词性的规则“word:好:0|pos:v:1|word:吗:2”,可以用来从词序列中抽取出好看吗、好用吗、好玩吗等作为目标信息内容,然后后续结合答案的类别,可以生成短答案。例如,若是答案类别是肯定,可以生成好看、好用、好玩,作为肯定回答;若答案类别是否定,可以生成不好看、不好用、不好玩,作为否定回答。

[0059] 同理,具体实现时,可以分别采用各条规则对词序列和词性序列进行匹配处理,获取能够匹配到的数条原始信息内容和各条原始信息内容对应的规则;根据各条规则的优先级,从数条原始信息内容中获取对应的规则优先级最高的原始信息内容,作为目标信息内容。同理,若预先设置的规则没有优先级,可以随机获取一条规则匹配的信息内容作为目标信息内容。还可以按照其他方式过滤出一条信息内容作为目标信息内容,在此不再一一举例赘述。

[0060] 本实施例的答案生成方法,通过采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题,若属于,根据问题、获取的问题对应的长答案以及预训练好的问题答案对分类器,获取问题的答案类别;若答案类别属于肯定或者否定,根据答案类别和问题,生成短答案,作为问题的目标答案。本实施例的技术方案,可以对是否类问题生成简洁、明了的短答案,与现有的人机对话系统提供的长答案相比,可以减少答案中的冗余信息,增强答案的精准度;能够进一步缩短用户阅读答案的时间,减少用户的阅读成本,从而能够有效地增强用户的使用体验度。

[0061] 图2为本发明的答案生成装置实施例一的结构图。如图2所示,本实施例的答案生成装置,具体可以包括:

[0062] 识别模块10用于采用预训练好的问题分类器,识别待分析的问题是否属于是否类问题;

[0063] 获取模块11用于若识别模块10识别问题属于是否类问题,根据问题、获取的问题对应的长答案以及预训练好的问题答案对分类器,获取问题的答案类别;

[0064] 检测模块12用于检测获取模块11获取的答案类别是否属于肯定或者否定；

[0065] 生成模块13用于若检测模块12确定答案类别属于肯定或者否定时，受检测模块12触发，根据获取模块11获取的答案类别和问题，生成短答案，作为问题的目标答案。

[0066] 本实施例的答案生成装置，通过采用上述模块实现答案生成的实现原理以及技术效果与上述相关方法实施例的实现相同，详细可以参考上述相关方法实施例的记载，在此不再赘述。

[0067] 图3为本发明的答案生成装置实施例二的结构图。如图3所示，本实施例的答案生成装置，在上述图2所示实施例的技术方案的基础上，进一步更加详细地描述本发明的技术方案。

[0068] 如图3所示，本实施例的答案生成装置，还包括：

[0069] 设置模块14用于若识别模块10识别问题不属于是否类问题，将获取的问题对应的长答案作为问题的目标答案。

[0070] 或者，设置模块14还用于若检测模块12确定答案类别不属于肯定或者否定，将获取的问题对应的长答案作为问题的目标答案。

[0071] 进一步可选地，如图3所示，本实施例的答案生成装置中，生成模块13还包括：

[0072] 获取单元131用于若检测模块12确定答案类别属于肯定或者否定时，受检测模块12触发，对问题进行分析，获取生成短答案所需的目标信息内容；

[0073] 生成单元132用于根据获取模块11获取的答案类别和获取单元131获取的目标信息内容，生成短答案。

[0074] 进一步可选地，获取单元131用于：

[0075] 对问题分别进行分词，得到对应的词序列；

[0076] 根据预设的多条规则和词序列，获取生成短答案所需的目标信息内容。

[0077] 此时对应地，获取单元131具体用于：

[0078] 分别采用各条规则对词序列进行匹配处理，获取能够匹配到的数条原始信息内容和各条原始信息内容对应的规则；

[0079] 根据各条规则的优先级，从数条原始信息内容中获取对应的规则优先级最高的原始信息内容，作为目标信息内容。

[0080] 或者，获取单元131用于：

[0081] 对问题分别进行分词和词性标注，得到对应的词序列和词性序列；

[0082] 根据预设的多条规则、词序列和词性序列，获取生成短答案所需的目标信息内容。

[0083] 此时对应地，获取单元131具体用于：

[0084] 分别采用各条规则对词序列和词性序列进行匹配处理，获取能够匹配到的数条原始信息内容和各条原始信息内容对应的规则；

[0085] 根据各条规则的优先级，从数条原始信息内容中获取对应的规则优先级最高的原始信息内容，作为目标信息内容。

[0086] 本实施例的答案生成装置，通过采用上述模块实现答案生成的实现原理以及技术效果与上述相关方法实施例的实现相同，详细可以参考上述相关方法实施例的记载，在此不再赘述。

[0087] 本发明还提供了一种人机对话系统，该系统中包括如上图2或者图3所示的答案生

成装置,并具体可以采用如上图1所示的答案生成方法,实现答案生成,详细可以参考上述实施例的记载,在此不再赘述。

[0088] 图4为本发明的计算机设备实施例的结构图。如图4所示,本实施例的计算机设备,包括:一个或多个处理器30,以及存储器40,存储器40用于存储一个或多个程序,当存储器40中存储的一个或多个程序被一个或多个处理器30执行,使得一个或多个处理器30实现如上图1所示实施例的答案生成方法。图4所示实施例中以包括多个处理器30为例。例如该计算机设备可以为一种人机对话系统。

[0089] 例如,图5为本发明提供的一种计算机设备的示例图。图5示出了适于用来实现本发明实施方式的示例性计算机设备12a的框图。图5显示的计算机设备12a仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0090] 如图5所示,计算机设备12a以通用计算设备的形式表现。计算机设备12a的组件可以包括但不限于:一个或者多个处理器16a,系统存储器28a,连接不同系统组件(包括系统存储器28a和处理器16a)的总线18a。

[0091] 总线18a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(ISA)总线,微通道体系结构(MAC)总线,增强型ISA总线、视频电子标准协会(VESA)局域总线以及外围组件互连(PCI)总线。

[0092] 计算机设备12a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12a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。

[0093] 系统存储器28a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(RAM)30a和/或高速缓存存储器32a。计算机设备12a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34a可以用于读写不可移动的、非易失性磁介质(图5未显示,通常称为“硬盘驱动器”)。尽管图5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CD-ROM, DVD-ROM或者其它光介质)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18a相连。系统存储器28a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发明上述图1-图3各实施例的功能。

[0094] 具有一组(至少一个)程序模块42a的程序/实用工具40a,可以存储在例如系统存储器28a中,这样的程序模块42a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42a通常执行本发明所描述的上述图1-图3各实施例中的功能和/或方法。

[0095] 计算机设备12a也可以与一个或多个外部设备14a(例如键盘、指向设备、显示器24a等)通信,还可与一个或者多个使得用户能与该计算机设备12a交互的设备通信,和/或与使得该计算机设备12a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22a进行。并且,计算机设备12a还可以通过网络适配器20a与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器20a通过总线18a与计算机设备12a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12a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、冗余处理器、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。

[0096] 处理器16a通过运行存储在系统存储器28a中的程序,从而执行各种功能应用以及数据处理,例如实现上述实施例所示的答案生成方法。

[0097] 本发明还提供一种计算机可读介质,其上存储有计算机程序,该程序被处理器执行时实现如上述实施例所示的答案生成方法。

[0098] 本实施例的计算机可读介质可以包括上述图5所示实施例中的系统存储器28a中的RAM30a、和/或高速缓存存储器32a、和/或存储系统34a。

[0099] 随着科技的发展,计算机程序的传播途径不再受限于有形介质,还可以直接从网络下载,或者采用其他方式获取。因此,本实施例中的计算机可读介质不仅可以包括有形的介质,还可以包括无形的介质。

[0100] 本实施例的计算机可读介质可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是一—但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0101] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括——但不限于——电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0102] 计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括——但不限于——无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0103] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码,所述程序设计语言包括面向对象的程序设计语言——诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言——诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0104] 在本发明所提供的几个实施例中,应该理解到,所揭露的系统,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0105] 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,作为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单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0106] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,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单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实现。

[0107] 上述以软件功能单元的形式实现的集成的单元,可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)或处理器(processor)执行本发明各个实施例所述方法的部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0108] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明保护的范围之内。

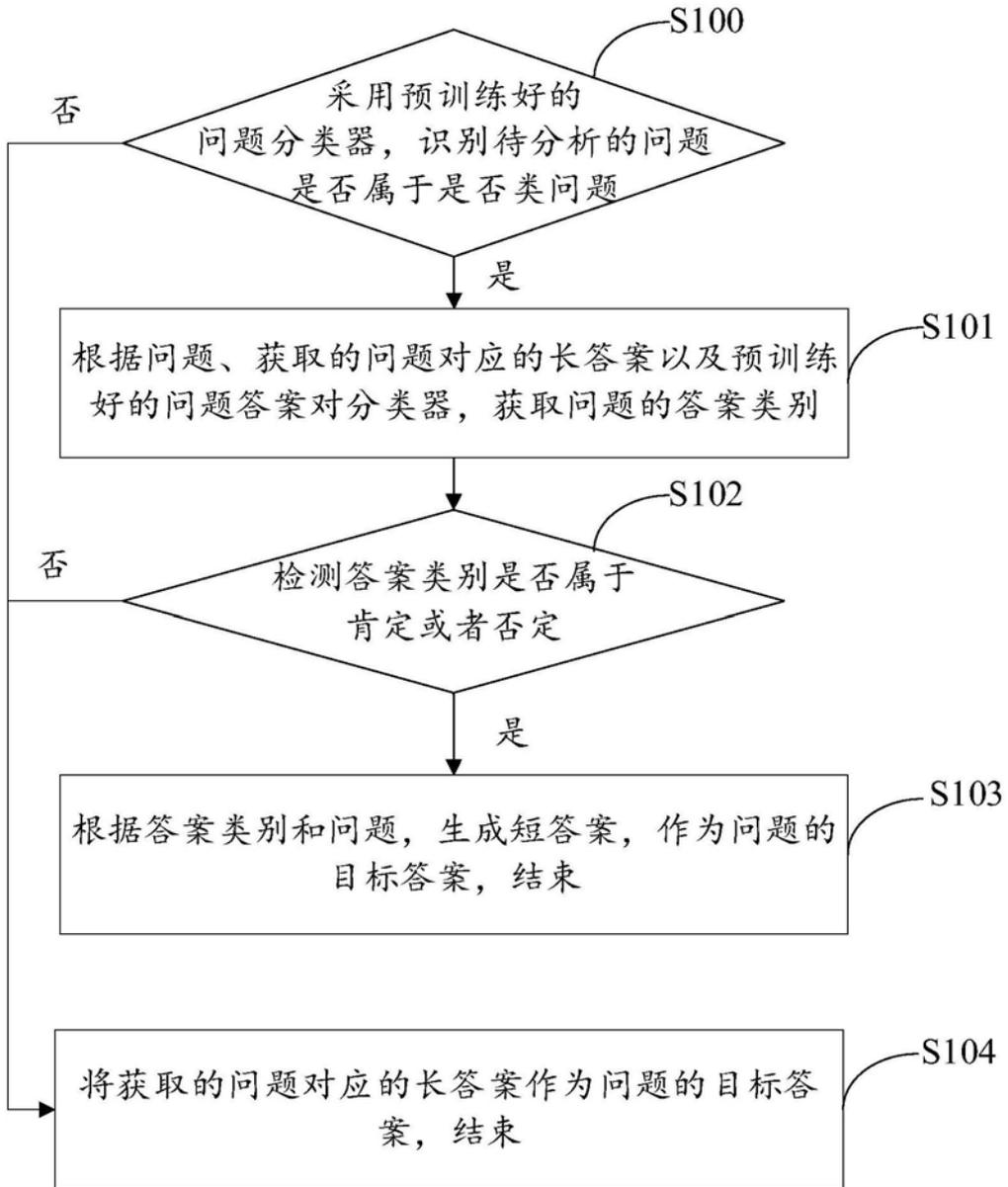


图1

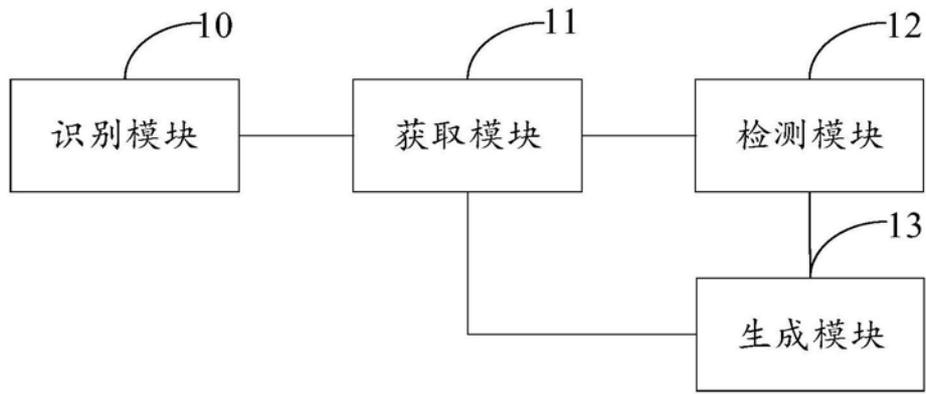


图2

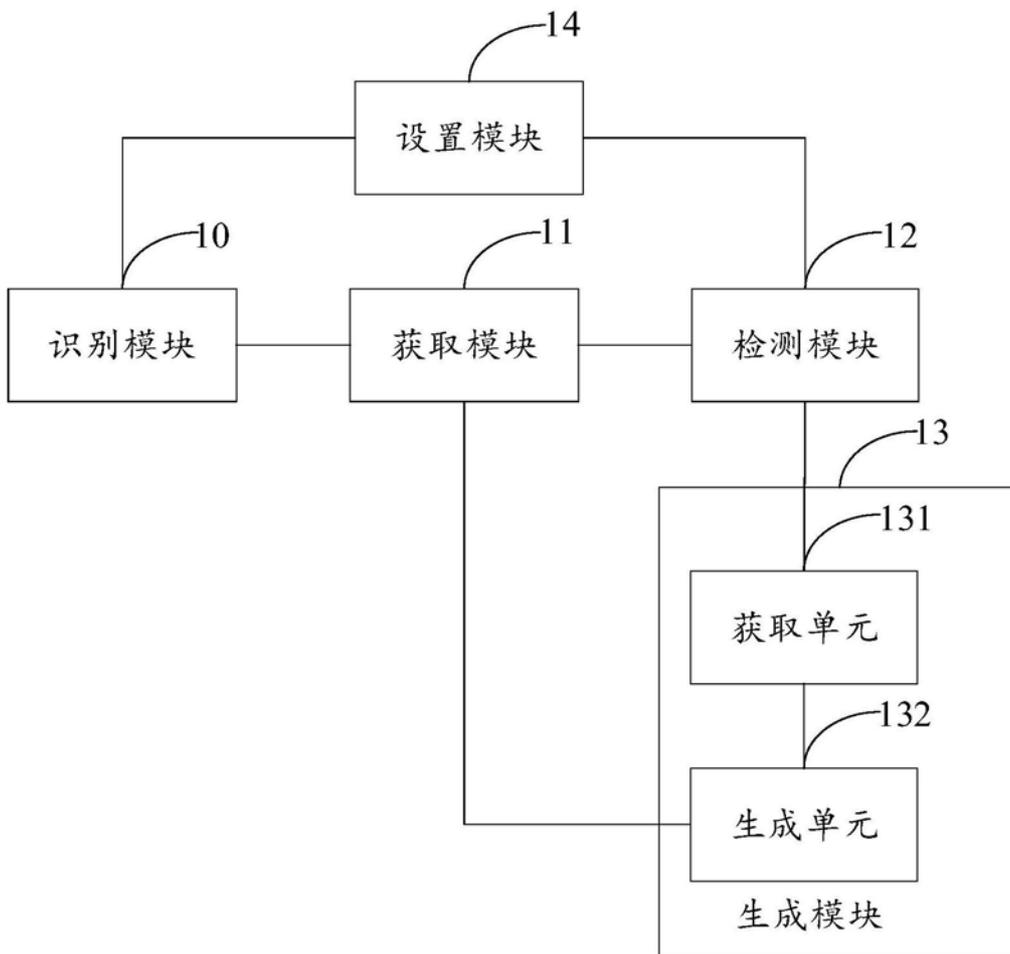


图3

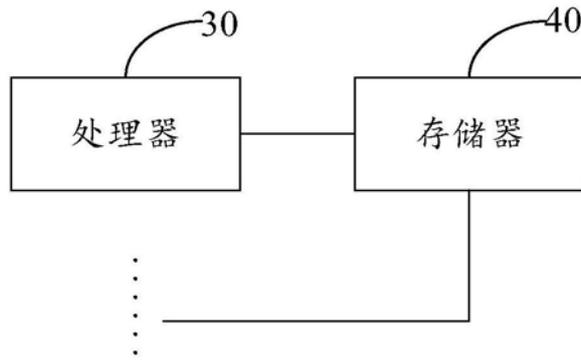


图4

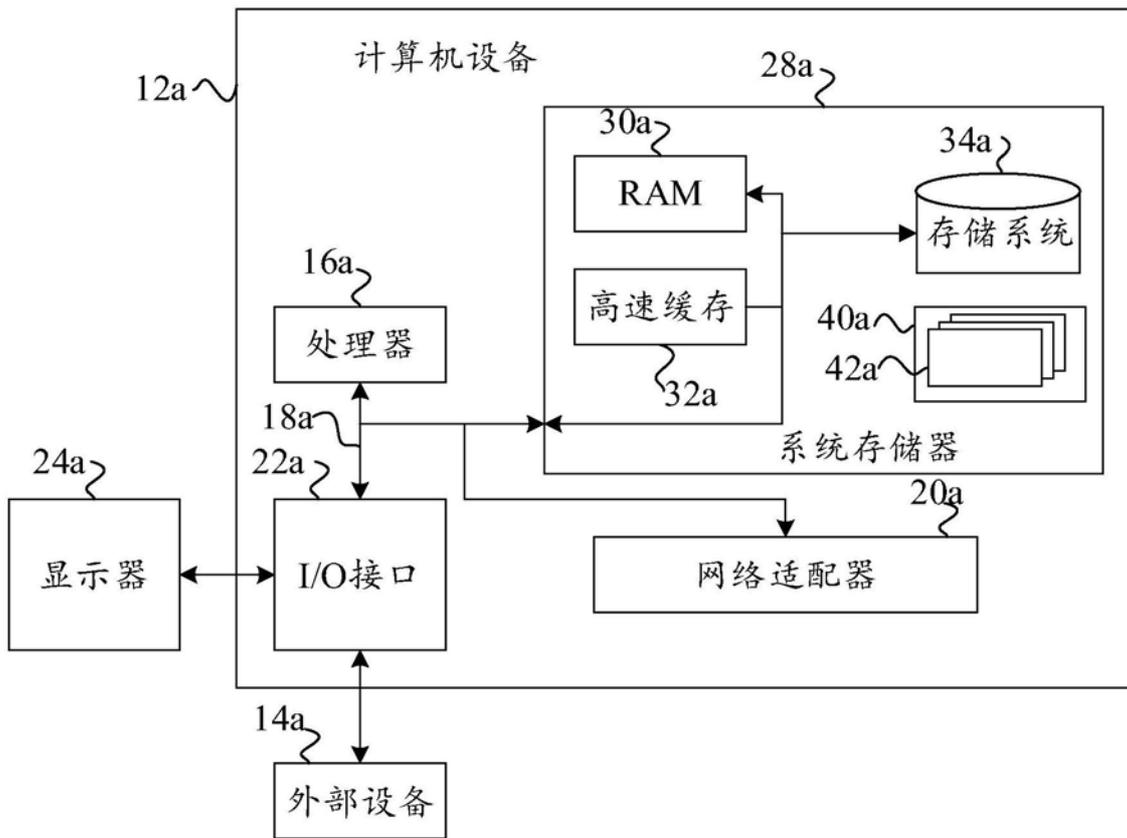


图5