



(12) 发明专利

(10) 授权公告号 CN 109284357 B

(45) 授权公告日 2022. 07. 19

(21) 申请号 201810995868.8

G06F 40/289 (2020.01)

(22) 申请日 2018.08.29

G06F 40/30 (2020.01)

(65) 同一申请的已公布的文献号
申请公布号 CN 109284357 A

(56) 对比文件

CN 106446109 A, 2017.02.22

CN 106446109 A, 2017.02.22

(43) 申请公布日 2019.01.29

CN 105243152 A, 2016.01.13

(73) 专利权人 腾讯科技(深圳)有限公司
地址 518000 广东省深圳市南山区高新区
科技中一路腾讯大厦35层

CN 107562774 A, 2018.01.09

CN 105653704 A, 2016.06.08

审查员 邓丽婉

(72) 发明人 冉邱 周霄 牛成 周杰

(74) 专利代理机构 深圳市联鼎知识产权代理有
限公司 44232
专利代理师 刘抗美

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/953 (2019.01)

权利要求书3页 说明书14页 附图8页

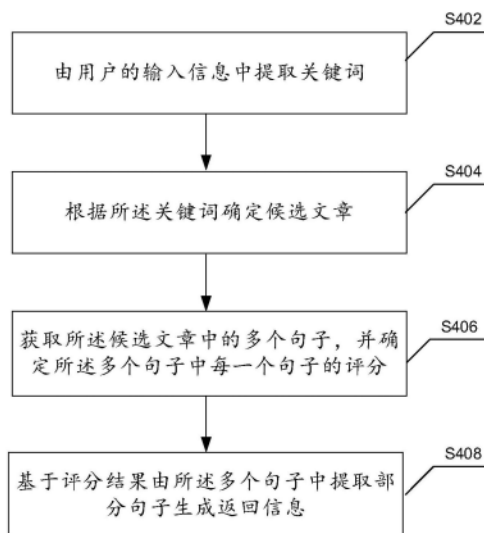
(54) 发明名称

人机对话方法、装置、电子设备及计算机可读介质

(57) 摘要

本公开涉及一种人机对话方法、装置、电子设备及计算机可读介质。该方法包括：由用户的输入信息中提取关键词；根据所述关键词确定候选文章；获取所述候选文章中的多个句子，并确定所述多个句子中每一个句子的评分；以及基于评分结果由所述多个句子中提取部分句子生成返回信息。本公开涉及的人机对话方法、装置、电子设备及计算机可读介质，能够使得机器人在人机对话过程中产生具有信息量的回答，辅助用户与外界信息进行观点交流。

40



1. 一种人机对话方法,其特征在于,包括:

将用户的输入信息进行分词处理,以获取多个词汇;

确定多个词汇的词汇向量;

确定所述多个词汇的词汇向量之间的余弦相似度;

基于词汇向量之间的余弦相似度构建相似度矩阵;

将词汇向量作为节点,采用共现关系构造任意两个节点之间的边,将所述相似度矩阵中的矩阵值作为对应节点之间边的权重,形成候选关键词图;所述共现关系表示两个节点之间存在边仅当它们对应的词汇在设定长度的窗口中共现;

给所述候选关键词图中的节点指定任意初值,并递归计算,直至所述候选关键词图中任意一节点的误差率小于预设值时,达到收敛,根据收敛后的候选关键词图确定词汇向量的重要度;

依据所述重要度排序确定关键词,所述关键词包括根据重要度由高至低选取的一个或多个词汇,以及由形成相邻词组的单词组合而成的多词关键词;

根据所述关键词确定候选文章;

获取所述候选文章中的多个句子,并根据句子间的语义相似度确定所述多个句子中每一个句子在所述候选文章中的重要度;将所述重要度作为每一个句子的第一评分;

根据所述多个句子中每一个句子与所述输入信息的相似性确定每一个句子的第二评分;

根据所述第一评分与所述第二评分的加权和确定所述多个句子中每一个句子的评分;

基于评分结果由所述多个句子中提取评分高的部分句子生成返回信息。

2. 如权利要求1所述的方法,其特征在于,所述用户的输入信息为用户的语音输入信息或用户的文字输入信息。

3. 如权利要求1所述的方法,其特征在于,根据所述关键词确定候选文章包括:

根据所述关键词在预定平台中进行检索以确定所述候选文章。

4. 如权利要求1所述的方法,其特征在于,根据句子间的语义相似度确定所述多个句子中每一个句子在所述文章中的重要度包括:

确定所述多个句子中每一个句子的句子向量;

确定所述多个句子的句子向量之间的语义相似度;以及

基于所述语义相似度获取每一个句子在所述文章中的重要度以确定所述第一评分。

5. 如权利要求4所述的方法,其特征在于,确定所述多个句子中每一个句子的句子向量包括:

将句子进行分词处理获取多个词汇;

确定多个词汇的词汇向量;

确定多个词汇的词频逆文档频率;以及

根据所述词频逆文档频率与所述词汇向量确定句子的句子向量。

6. 如权利要求4所述的方法,其特征在于,基于所述语义相似度获取每一个句子在所述文章中的重要度以确定第一评分包括:

基于所述语义相似度构建相似度矩阵;

将句子作为节点,将语义相似度作为边的权重,基于所述相似度矩阵构建图模型;

基于所述图模型通过迭代获取句子在所述文章中的重要度；
通过所述多个句子中每一个句子的重要度确定所述第一评分。

7. 如权利要求1所述的方法,其特征在於,根据所述多个句子中每一个句子与所述输入信息的相似性确定第二评分包括:

确定用户输入信息的信息向量;

确定所述多个句子中每一个句子的句子向量;

通过余弦相似度确定信息向量与每一个句子的句子向量之间的相似性;以及

通过所述多个句子中每一个句子的相似性确定所述第二评分。

8. 如权利要求1所述的方法,其特征在於,基于评分结果由所述多个句子中提取部分句子生成返回信息包括:

将所述多个句子按照评分结果进行排序;以及

依序由所述多个句子中提取部分句子生成所述返回信息。

9. 如权利要求8所述的方法,其特征在於,依序由所述多个句子中提取部分句子生成所述返回信息包括:

对所述部分句子进行自然语言处理以生成所述返回信息。

10. 如权利要求1或4所述的方法,其特征在於,还包括:

通过词向量训练算法训练文字数据获取所述词汇向量。

11. 一种人机对话装置,其特征在於,包括:

关键词模块,用于将用户的输入信息进行分词处理,以获取多个词汇;确定多个词汇的词汇向量;确定所述多个词汇的词汇向量之间的相似度;基于词汇向量之间的余弦相似度构建相似度矩阵;将词汇向量作为节点,采用共现关系构造任意两个节点之间的边,将所述相似度矩阵中的矩阵值作为对应节点之间边的权重,形成候选关键词图;所述共现关系表示两个节点之间存在边仅当它们对应的词汇在设定长度的窗口中共现;给所述候选关键词图中的节点指定任意初值,并递归计算,直至所述候选关键词图中任一节点的误差率小于预设值时,达到收敛,根据收敛后的候选关键词图确定词汇向量的重要度;依据所述重要度排序确定关键词,所述关键词包括根据重要度由高至低选取的一个或多个词汇,以及由形成相邻词组的单词组合而成的多词关键词;

文章模块,用于根据所述关键词确定候选文章;

评分模块,用于获取所述候选文章中的多个句子,并根据句子间的语义相似度确定所述多个句子中每一个句子在所述候选文章中的重要度;将所述重要度作为每一个句子的第一评分;根据所述多个句子中每一个句子与所述输入信息的相似性确定每一个句子的第二评分;根据所述第一评分与所述第二评分的加权和确定所述多个句子中每一个句子的评分;

返回信息模块,用于基于评分结果由所述多个句子中提取评分高的部分句子生成返回信息。

12. 一种电子设备,其特征在於,包括:

一个或多个处理器;

存储装置,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1-10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,所述程序被处理器执行时实现如权利要求1-10中任一所述的方法。

人机对话方法、装置、电子设备及计算机可读介质

技术领域

[0001] 本公开涉及计算机信息处理领域,具体而言,涉及一种人机对话方法、装置、电子设备及计算机可读介质。

背景技术

[0002] 聊天机器人(chatterbot)是一个用来模拟人类对话或聊天的程序。目前聊天机器人广泛运用于各种即时通讯平台,学习教育平台,以及各种信息服务平台。聊天机器人可基于数据库中知识与用户进行聊天,通常是研发者把自己感兴趣的回答放到数据库中,当一个问题被抛给聊天机器人时,它从数据库中找到最贴切的答案,回复给它的聊伴。这种简单的对话型聊天方式,用户获取信息范围较窄,逐渐退出市场。

[0003] 随着,序列数据的神经网络在人工智能与翻译交流领域的应用,聊天机器人可基于用户所说的话作为模型的输入,将模型的输出的语句作为回复来进行人机交互。但是这种方式的机器人回复内容较为空洞,不具有信息量,并没有实际内容的扩展,用户实际体验一般,在实际应用中受到限制。

发明内容

[0004] 有鉴于此,本公开提供一种人机对话方法、装置、电子设备及计算机可读介质,能够使得机器人在人机对话过程中产生具有信息量的回答,辅助用户与外界信息进行观点交流。

[0005] 本公开的其他特性和优点将通过下面的详细描述变得显然,或部分地通过本公开的实践而习得。

[0006] 根据本公开的一方面,提出一种人机对话方法,该方法包括:由用户的输入信息中提取关键词;根据所述关键词确定候选文章;获取所述候选文章中的多个句子,并确定所述多个句子中每一个句子的评分;以及基于评分结果由所述多个句子中提取部分句子生成返回信息。

[0007] 根据本公开的一方面,提出一种人机对话装置,该装置包括:关键词模块,用于由用户的输入信息中提取关键词;文章模块,用于根据所述关键词确定候选文章;评分模块,用于获取所述候选文章中的多个句子,并确定所述多个句子中每一个句子的评分;以及返回信息模块,用于基于评分结果由所述多个句子中提取部分句子生成返回信息。

[0008] 根据本公开的一方面,提出一种电子设备,该电子设备包括:一个或多个处理器;存储装置,用于存储一个或多个程序;当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如上文的方法。

[0009] 根据本公开的一方面,提出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文中的方法。

[0010] 根据本公开的人机对话方法、装置、电子设备及计算机可读介质,能够使得机器人在人机对话过程中产生具有信息量的回答,辅助用户与外界信息进行观点交流。

[0011] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性的,并不能限制本公开。

附图说明

[0012] 通过参照附图详细描述其示例实施例,本公开的上述和其它目标、特征及优点将变得更加显而易见。下面描述的附图仅仅是本公开的一些实施例,对于本领域的普通技术人员来讲,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。

[0013] 图1是根据一示例性实施例示出的一种人机对话方法及装置的系统框图。

[0014] 图2是根据一示例性实施例示出的一种人机对话方法的应用场景图。

[0015] 图3是根据一示例性实施例示出的一种人机对话方法的应用场景图。

[0016] 图4是根据一示例性实施例示出的一种人机对话方法的流程图。

[0017] 图5是根据另一示例性实施例示出的一种人机对话方法的流程图。

[0018] 图6是根据另一示例性实施例示出的一种人机对话方法的流程图。

[0019] 图7是根据另一示例性实施例示出的一种人机对话方法的流程图。

[0020] 图8是根据一示例性实施例示出的一种人机对话装置的框图。

[0021] 图9是根据另一示例性实施例示出的一种人机对话装置的框图。

[0022] 图10是根据一示例性实施例示出的一种电子设备的框图。

[0023] 图11是根据一示例性实施例示出的一种计算机可读存储介质示意图。

具体实施方式

[0024] 现在将参考附图更全面地描述示例实施例。然而,示例实施例能够以多种形式实施,且不应被理解为限于在此阐述的实施例;相反,提供这些实施例使得本公开将全面和完整,并将示例实施例的构思全面地传达给本领域的技术人员。在图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。

[0025] 此外,所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施例中。在下面的描述中,提供许多具体细节从而给出对本公开的实施例的充分理解。然而,本领域技术人员将意识到,可以实践本公开的技术方案而没有特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知方法、装置、实现或者操作以避免模糊本公开的各方面。

[0026] 附图中所示的方框图仅仅是功能实体,不一定必须与物理上独立的实体相对应。即,可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0027] 附图中所示的流程图仅是示例性说明,不是必须包括所有的内容和操作/步骤,也不是必须按所描述的顺序执行。例如,有的操作/步骤还可以分解,而有的操作/步骤可以合并或部分合并,因此实际执行的顺序有可能根据实际情况改变。

[0028] 应理解,虽然本文中可能使用术语第一、第二、第三等来描述各种组件,但这些组件不应受这些术语限制。这些术语乃用以区分一组件与另一组件。因此,下文论述的第一组件可称为第二组件而不偏离本公开概念的教导。如本文中所使用,术语“及/或”包括相关联的列出项目中的任一个及一或多者的所有组合。

[0029] 本领域技术人员可以理解,附图只是示例实施例的示意图,附图中的模块或流程并不一定是实施本公开所必须的,因此不能用于限制本公开的保护范围。

[0030] 目前常用于聊天机器人闲聊的技术有基于序列到序列神经网络模型(seq2seq)生成式方法。seq2seq神经网络模型是一种端到端(end-to-end)模型,能够对序列进行处理,该模型以一个序列作为输入,以另一个序列作为输出,常用于机器翻译和对话领域。而基于seq2seq神经网络模型的生成式方法:首先在大量对话训练数据上训练seq2seq模型,然后在用户与聊天机器人闲聊时,直接将用户所说的话作为该模型的输入,然后将该模型的输出作为对于用户的回复。

[0031] 采用基于seq2seq模型生成式方法的缺点是训练seq2seq模型时需要大量的对话训练数据,而这些数据由于范围广数据量大,使得seq2seq神经网络模型的训练数据在实际生活中较难获取。

[0032] 采用基于seq2seq模型生成式方法的聊天机器人在闲聊时,对用户的回答空洞,不具有信息量,没有实际内容,在实际应用中受到限制。例如,当用户想了解外界资讯或者和聊天机器人交流对一些事件观点或看法的时候,该方法只能简单地回复或回避用户问题,并不能正面回答用户问题,更不能给出其他有信息量的建设性的观点。

[0033] 有鉴于此,本公开提出了一种人机对话方法,以期实现根据用户的提问,在相关的文章中抽取关键句子作为自己的回复,实现资讯、观点或看法等的交流的目的。

[0034] 下面将对本公开的内容进行详细描述。

[0035] 图1是根据一示例性实施例示出的一种人机对话方法及装置的系统框图。

[0036] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0037] 用户可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用,例如购物类应用、网页浏览器应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等。

[0038] 终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、智能音箱、平板电脑、膝上型便携计算机和台式计算机等等。

[0039] 服务器105可以是提供各种服务的服务器,例如对用户利用终端设备101、102、103所调用的聊天机器人提供后台数据支撑的服务器。后台数据支撑的服务器可以对接收到的用户提问问题进行分析等处理,并将处理结果(例如返回信息、返回语音等)反馈给终端设备。

[0040] 终端设备101、102、103可由用户的语音输入信息中提取关键词;终端设备101、102、103还可由用户的文字输入信息中提取关键词。

[0041] 终端设备101、102、103可由用户的输入信息中提取关键词;终端设备101、102、103可根据所述关键词确定候选文章;终端设备101、102、103可获取所述候选文章中的多个句子,并确定所述多个句子中每一个句子的评分;终端设备101、102、103可基于评分结果由所述多个句子中提取部分句子生成返回信息。

[0042] 终端设备101、102、103可将用户的输入信息转发至服务器105中进行处理;服务器105可由用户的输入信息中提取关键词;服务器105可根据所述关键词确定候选文章;服务

器105可获取所述候选文章中的多个句子,并确定所述多个句子中每一个句子的评分;服务器105可基于评分结果由所述多个句子中提取部分句子生成返回信息;服务器105可将所述返回信息发送至终端设备101、102、103以便响应所述用户的输入信息。

[0043] 需要说明的是,本公开实施例所提供的人机对话方法可以由服务器105和/或终端设备101、102、103执行,相应地,人机对话装置可以设置于服务器105和/或终端设备101、102、103中。而提供给用户进行信息输入的请求端一般位于终端设备101、102、103中。

[0044] 根据本公开的人机对话方法及装置可以应用于聊天机器人相关产品中,如智能音箱、及时通讯软件中的聊天机器人等,主要应用在用户与聊天机器人闲聊的场景下。本公开的人机对话方法及装置可以实现新闻概要播报、观点交流等功能。

[0045] 在一个应用场景中,了解新闻是用户的常用需求,一般情况下播报完整的新闻会占用较长时间,会使得用户失去耐心,降低用户体验,这时播报新闻概要就显得十分必要了。例如,当用户想了解近期的热点新闻时,可以对聊天机器人说“我想听下最近的热点新闻”,或“最近有什么热点新闻吗”等其他表述,本公开的人机对话方法及装置会从近期的公众号文章中检索出几篇热点新闻文章,并抽取文章中的几个关键句子作为概要播报给用户。或者,当用户想了解具体方面如中美贸易战的新闻时,可以对聊天机器人说“现在中美贸易战局势怎样了”,本公开的人机对话方法及装置会从所检索出的该方面最相关的几篇新闻中抽取关键句子,组成概要回复给用户。

[0046] 图2是根据一示例性实施例示出的一种人机对话方法的应用场景图。如图2所述的人机对话场景中,用户可在即时聊天窗口中,或信息交互平台上与聊天机器人进行对话,用户可通过文字输入问题,聊天机器人对用户问题进行数据处理后,返回文字信息和/或语音信息作为用户提问的响应。

[0047] 闲聊时,用户可能会向聊天机器人提出对事情的看法等观点交流的问题,如“你对这次高考改革怎么看”,本公开的人机对话方法及装置根据用户提问的信息,搜索相关文章进而组成返回信息以答复用户,用户可从返回信息中获取到最近的高考政策及高考分数等。本公开的人机对话方法及装置可以使聊天机器人像真人一样进行直接回答,而不是简单回复“很好啊”等类似语句,或者回避用户提问。针对用户提出的问题,本公开的人机对话方法及装置可通过在大量热点文章中提取出高考改革评论文章,并从中提取出关键的评论句子,然后将关键评论句子组织成回复信息以作为机器人要表达的看法。

[0048] 图3是根据一示例性实施例示出的一种人机对话方法的应用场景图。如图3所述的人机对话场景中,用户可通过语音输入终端与聊天机器人进行对话,语音输入终端可为智能音箱或者智能机器人等设备。用户可通过语音提出问题,聊天机器人先将用户的语音问题转化为文字信息,然后再进行数据处理后,返回文字和/或语音信息作为用户提问的响应。

[0049] 再如当用户问“XXXX这部电影怎样”,本公开的人机对话方法及装置可在电影信息平台中提取相关文章中的关键评论句子,然后根据这些句子组成返回信息以答复用户。用户可以从返回信息中获取电影简介或评价内,而不会像基于seq2seq模型生成式方法那样简单地回复“很好看啊”等类似回答。

[0050] 根据本公开的人机对话方法及装置,能够解决现有技术中存在的技术缺陷,本公开的人机对话方法及装置可以根据用户的提问,在相关的文章中抽取关键句子作为自己的

回复,进而实现用户观点与外界资讯、观点或看法等的交流。

[0051] 图4是根据一示例性实施例示出的一种人机对话方法的流程图。人机对话方法40至少包括步骤S402至S408。

[0052] 如图4所示,在S402中,由用户的输入信息中提取关键词。可由用户的语音输入信息中提取关键词;或由用户的文字输入信息中提取关键词。

[0053] 在一个实施例中,由用户的文字输入信息中提取关键词,具体可包括:将所述输入信息进行分词处理,以获取多个词汇;确定多个词汇的词汇向量;通过余弦相似度与词汇向量确定多个词汇之间的相似度;根据所述相似度确定多个词汇的重要度;以及依据所述重要度排序确定所述关键词。其中,关键词的数量可为一个或多个,本公开不以此为限。关键词提取的相关内容将在后文图5的实施例中进行详细说明。

[0054] 在一个实施例中,由用户的语音输入信息中提取关键词,具体可通过智能音箱、智能机器人或者其他带有声音接收功能的音频设备接收用户语音输入的音频信息。在获取音频数据后对用户的音频信息音频识别,进而获取用户输入信息中的关键词。还可将用户的音频信息通过音频识别转化为文字信息,进而根据上一实施例中的方法,提取文字信息中的关键词。

[0055] 如图4所示,在S404中,根据所述关键词确定候选文章。可根据所述关键词在预定平台中进行检索以确定所述候选文章。例如可由近期的公众号文章中检索出几篇与关键词相关的热点新闻文章;还可在主流搜索引擎中对关键词进行搜索,根据检索出的结果进行排序,根据排序由高至低由检索结果中确定所述候选文章。其中,候选文章数量可为一个或多个,本公开不以此为限。

[0056] 如图4所示,在S406中,获取所述候选文章中的多个句子,并确定所述多个句子中每一个句子的评分。可包括:根据所述多个句子中每一个句子在所述文章中的重要度确定第一评分;根据所述多个句子中每一个句子与所述输入信息的相似性确定第二评分;根据所述第一评分与所述第二评分确定评分。

[0057] 在一个实施例中,根据所述多个句子中每一个句子在所述文章中的重要度确定所述第一评分,句子的重要度可通过文本排序算法获得,具体包括:确定所述多个句子中每一个句子的句子向量;确定所述多个句子的句子向量之间的相似度;基于所述相似度获取每一个句子在所述文章中的重要度以确定所述第一评分。

[0058] 在一个实施例中,文本排序算法可为Automatic Summarization(自动文摘算法)中的TextRank(文本抽取法)。TextRank通过把候选文章分割成若干组成单元(单词、句子)并建立图模型,利用投票机制对候选文章中的重要成分进行排序,从而能够仅利用候选文章本身的信息对文档中的每个句子的重要程度进行评分。通过TextRank算法确定句子在文章中的重要度的相关内容将在后文图7的实施例中进行详细说明。

[0059] 因为TextRank评分只体现对应句子在文章中的语义重要度,没要考虑与用户提问的相关性,所以不能直接根据该重要度抽取句子。为此,可用提问中关键词的词向量加权平均作为用户提问的向量表示,用各句子向量与该向量之间的余弦相似度作为句子与用户提问的相似性的第二评分。

[0060] 在一个实施例中,根据所述多个句子中每一个句子与所述输入信息的相似性确定所述第二评分包括:确定用户输入信息的信息向量;确定所述多个句子中每一个句子的句

子向量;通过余弦相似度确定信息向量与每一个句子的句子向量之间的相似性;以及通过所述多个句子中每一个句子的相似性确定所述第二评分。候选文章中的句子与所述输入信息之间的相似度计算具体步骤与图7中的过程相似;将用户输入信息也作为句子,在将图7中将候选文章句子与句子之间的相似性比较,替换为用户输入信息与句子之间的相似性比较即可,本公开在此不再赘述。

[0061] 在一个实施例中,根据所述第一评分与所述第二评分确定评分。可通过候选文章中,每个句子的重要度与相似性的加权和作为句子的最终评分并进行排序。例如,第一评分为A,第二评分为B,最终的评分C可表示为:

[0062] $C=qA+(1-q)B$;

[0063] 其中,q为权重系数。值得一提的是,q的取值范围为 $q \in [0,1]$;在q取值为1,或0时,总的评分C可由第一评分为A或第二评分为B确定。

[0064] 如图4所示,在S408中,基于评分结果由所述多个句子中提取部分句子生成返回信息。可将所述多个句子按照评分结果进行排序;以及依序由所述多个句子中提取部分句子生成所述返回信息。

[0065] 在一个实施例中,依序由所述多个句子中提取部分句子生成所述返回信息包括:对所述部分句子进行后处理以生成所述返回信息。具体可将候选文章的句子进行评分排序(第一评分与第二评分),优选出评分最高的几个句子组织成回复,并进行自然语言处理如指代消解、摘要首句首部去连词等,生成便于用户阅读理解的文字结果,然后将文字结果返回给用户。

[0066] 指代消解是自然语言处理的重要内容,就是在篇章中确定代词指向哪个名词短语的问题。在公开的一个实施例中,指代消解首先要构造先行候选录,然后再从候选中作多选一选择。具体而言指代消解的方法包括:基于句法的指代消解,以及基于语料库的方法,还有统计方法,统计机器学习方法等等,本公开中的指代消解过程可通过以上方法中的一种或几种实现。

[0067] 根据本公开的人机对话方法,通过检索用户问题相关文章,并在相关文章中提取部分语句,生成具有实质内容,包含信息量的返回信息的方式,能够使得机器人在人机对话过程中产生具有信息量的回答,辅助用户与外界信息进行观点交流。

[0068] 在非任务型对话即闲聊场景下,用户向聊天机器人提问时,本公开的人机对话方法,首先通过检索提取与用户提问相关的几篇候选文章,采用文本排序算法计算候选文章中各句子的重要度,再综合考虑各句子与用户提问的相关性生成候选文章中个句子的最终评分和排序,最后根据评分在候选文章中抽取几个关键句子组织成摘要返回给用户作为回答。根据以上内容的描述,本公开的人机对话方法,能够提取相关文章中的关键句子使得聊天机器人的回复内容包含一定信息而不空洞。本公开的人机对话方法,在生成返回信息时,综合考虑句子在候选文章中的重要度和句子与用户提问的相关性,能够获得既贴近用户问题,又具有信息量的返回信息。

[0069] 应清楚地理解,本公开描述了如何形成和使用特定示例,但本公开的原理不限于这些示例的任何细节。相反,基于本公开公开的内容的教导,这些原理能够应用于许多其它实施例。

[0070] 在本公开的一种示例性实施例中,还包括:通过词向量训练算法训练文字数据获

取所述词汇向量。其中,词向量(Word embedding),是自然语言处理(NLP)中的一组语言建模和特征学习技术的统称,在本公开实施例中,词向量作用就是将自然语言中的字词转为计算机可以理解的稠密向量。构建词向量常用的模型为word2vec模型和GloVe(Global Vectors for word representation)模型。

[0071] 其中,word2vec模型能训练大量的语料,并且可以使单词向量维度较低,能够捕捉单词之间的相似性,对单词使用代数运算就能计算相似的单词。GloVe模型是word2vec模型之后提出的词向量训练模型,GloVe模型word2vec模型的基础上提出了根据语料中的一些统计信息辅助进行词向量构建的想法。也就是说,如果一个词汇和哪个上下文词汇在一起的多,那么这个词汇与这个上下文词汇在一起要比与其他词在一起意义要大,那么在最终词向量的确定过程中,会优先考虑这两个词汇之间的关联关系。

[0072] 基于以上的词向量模型的特点分析,在本公开的一个实施例中,可通过GloVe算法对公共网络中的新闻数据或者文章数据进行训练以获取本公开中所用的词汇的词向量。在本公开的一个实施例中,还可通过现有技术中的其他词向量构建算法获取本公开中所用的词汇的词向量,本公开不以此为限。

[0073] 图5是根据另一示例性实施例示出的一种人机对话方法的流程图。如图5所示的人机对话方法50是对图4所示的人机对话方法40中“由用户的输入信息中提取关键词”的详细描述。

[0074] 如图5所述,在S502中,将所述输入信息进行分词处理,以获取多个词汇。如上文所述,用户输入的信息可为音频或文字形式,当用户输入的信息为音频时,可先将音频信息转化为文字,然后再进行本实施例中的分词处理。

[0075] 在一个实施例中,分词处理可为对汉字进行中文分词处理,中文分词(Chinese Word Segmentation)指的是将一个汉字序列切分成一个一个单独的词。分词就是将连续的字序列按照一定的规范重新组合成词序列的过程。现有的分词算法可分为三大类:基于字符串匹配的分词方法、基于理解的分词方法和基于统计的分词方法。按照是否与词性标注过程相结合,又可以分为单纯分词方法和分词与标注相结合的一体化方法。

[0076] 在一个实施例中,基于字符串匹配的分词方法,是按照一定的策略将待分析的汉字串与一个“充分大的”机器词典中的词条进行匹配,若在词典中找到某个字符串,则匹配成功(识别出一个词)。基于理解的分词方法,是通过让计算机模拟人对句子的理解,达到识别词的效果。其基本思想就是在分词的同时进行句法、语义分析,利用句法信息和语义信息来处理歧义现象。基于统计的分词方法,对语料中相邻共现的各个字的组合的频度进行统计,计算它们的互现信息。

[0077] 当然,在用户输入的信息为英文或其他语言信息时,也可通过其他语言对应的分词处理方法对用户输入信息进行分词处理,具体分词处理的方法不影响本公开后续实施例的处理步骤。在本公开中,可通过以上多种分词方法中的一种或几种来对本公开中的输入信息进行分词处理,本公开不以此为限。

[0078] 在S504中,确定多个词汇的词汇向量。可通过GloVe算法对公共网络中的新闻数据或者文章数据进行训练以获取本公开中所用的词汇的词向量。

[0079] 在一个实施例中,将通过训练得到的词向量为上文分词处理之后得到的词汇进行词向量标注,生成词汇向量集合。

[0080] 在S506中,通过余弦相似度与词汇向量确定多个词汇之间的相似度。可用词向量间的余弦相似度作为对应词之间的相似度。余弦相似度,是通过计算两个向量的夹角余弦值来评估他们的相似度。

[0081] 在一个实施例中,可按照文章中词汇出现的顺序,将多个词汇进行排序并编号,可由1号词汇向量作为起始词汇向量,分别比较1号词汇向量与其他词汇向量之间的相似度。在多次遍历所有的词汇向量,并将词汇向量两两比较之后,确定他们之间的相似度。

[0082] 在一个实施例中,将两个词汇向量根据其坐标值,分别绘制到向量空间中,如最常见的二维空间。计算求得他们的夹角,并得出夹角对应的余弦值,此余弦值就可以用来表征这两个向量的相似性。例如两个词汇向量之间的夹角越小,他们之间的余弦值越接近于1,则两个词汇向量越相似。

[0083] 在S508中,根据所述相似度确定多个词汇的重要度。可利用TextRank算法计算各词的重要度。TextRank算法是利用局部词汇之间关系(共现窗口)对后续关键词进行排序,直接从文本本身抽取。其主要步骤可如下:

[0084] (1) 构建候选关键词图 $G=(V,E)$,其中 V 为节点集(词汇向量集),然后采用共现关系(co-occurrence)构造任两点之间的边,即为两个节点之间存在边仅当它们对应的词汇在长度为 K 的窗口中共现, K 表示窗口大小,即最多共现 K 个单词。

[0085] (2) 根据如图7所述的步骤和公式,将节点之间边的权重设置为词汇向量之间的相似度,通过迭代传播各节点的权重,直至收敛。具体步骤与方法请参考图7以及图7对应的实施例,本申请在此不再赘述。

[0086] (3) 对节点权重进行倒序排序,从而得到最重要的 T 个词汇向量,节点的权重即为词汇向量的重要度。

[0087] 在S510中,依据所述重要度排序确定所述关键词。可按照重要度由高至低的选取一个或多个词汇向量,这些词汇向量对应的词汇作为本实施例中的关键词。

[0088] 在一个实施例中,还可由得到最重要的 N 个单词,在原始文本中进行标记,若形成相邻词组,则组合成多词关键词。例如,用户输入的信息文本中有句子“我想知道今天的天气预报”,如果“天气”和“预报”均属于候选关键词,则可将“天气预报”这一组合作为本公开中的关键词。

[0089] 本公开的人机对话方法,通过TextRank算法对用户输入的信息进行处理,获取用户输入信息的关键词的方式,能够准确快速的提取出用户输入信息的关键点,有利于后文中的信息检索处理。

[0090] 图6是根据另一示例性实施例示出的一种人机对话方法的流程图。如图6所示的人机对话方法60是对图4所示的人机对话方法40中“确定所述多个句子中每一个句子的评分”的详细描述。

[0091] 如图6所示,在S602中,获取用户输入信息与多个句子。在本实施例中,用户输入信息为文字信息,或者是将用户的语音进行转化得到的文字信息。多个句子为上文中在待选文章中提取出的句子。

[0092] 在S604中,确定所述多个句子中每一个句子的句子向量。可通过句子中所包含词的词向量来计算句子向量。句子向量可例通过由句子中包含的词向量的不同方式的组合得到。句子向量计算方法可包括:

[0093] 神经词袋模型,简单对文本序列中每个词嵌入进行平均,将所有的词进行加总,并将加总之后的结果作为整个序列的表示。

[0094] 递归神经网络模型(Recursive Neural Network),按照一个给定的外部拓扑结构(比如成分句法树),不断递归得到整个序列的表示。

[0095] 循环神经网络(Recurrent Neural Network),将文本序列看作时间序列,不断更新,最后得到整个序列的表示。

[0096] 卷积神经网络(Convolutional Neural Network),通过多个卷积层和子采样层,最终得到一个固定长度的向量。

[0097] 在一个实施例中,本公开通过加权表示的神经词袋模型,利用词向量构建句子向量,本部分内容将在图7的实施例中进行详细描述。

[0098] 在S606中,确定所述多个句子的句子向量之间的相似度。可根据上文所述的方法确定句子向量之间的余弦相似度,进而通过句子向量间的余弦相似度作为句子间的语义相似度。

[0099] 在S608中,基于所述相似度获取每一个句子在所述文章中的重要度。可根据句子之间的相似度构建相似度矩阵,根据该相似度矩阵,以句子为点,对应的相似度作为边的权重,构建图模型。在图模型的基础上,PageRank的迭代方法,计算出每个句子的最终重要度。本部分内容将在图7的实施例中进行详细描述。

[0100] 在一个实施例中,可将重要度作为第一评分。

[0101] 在S610中,确定用户输入信息的信息向量。可参照上文中句子向量的生成方式,将用户输入信息作为句子,通过句子中所包含词的词向量来计算用户输入信息的信息向量。

[0102] 在S612中,确定所述多个句子中每一个句子的句子向量。可参照上文中句子向量的生成方式,通过句子中所包含词的词向量来计算句子向量。

[0103] 在S614中,通过余弦相似度确定信息向量与每一个句子的句子向量之间的相似性。可用词向量间的余弦相似度作为对应词之间的相似度。余弦相似度,是通过计算两个向量的夹角余弦值来评估他们的相似度的方法。

[0104] 在一个实施例中,可将相似度作为第二评分。

[0105] 在S616中,确定评分。综合考虑第一评分与第二评分,例如第一评分为A,第二评分为B,最终的评分C可表示为:

[0106] $C=qA+(1-q)B$;

[0107] 其中,q为权重系数。

[0108] 本公开的人机对话方法,综合考虑了句子在候选文章中的语义重要度以及用户提问的内容与候选文章中句子的相关性,进而确定候选文章中句子评分的方式,能够从多方面、多角度、综合的对候选文章中的句子进行评分。

[0109] 图7是根据另一示例性实施例示出的一种人机对话方法的流程图。如图7所示的人机对话方法70是对图6所示的人机对话方法60中“确定所述多个句子中每一个句子的句子向量”与“基于所述相似度获取每一个句子在所述文章中的重要度”的详细描述。

[0110] 如图7所示,在S702中,将句子进行分词处理获取多个词汇。可通过上文所述的多种分词方法中的一种或几种来对本公开中的输入信息进行分词处理,本公开不以此为限。

[0111] 在S704中,确定多个词汇的词汇向量。可通过GloVe算法对公共网络中的新闻数据

或者文章数据进行训练以获取本公开中所用的词汇的词向量。

[0112] 在S706中,确定多个词汇的词频逆文档频率。可通过TF-IDF (term frequency-inverse document frequency) 技术确定多个词汇的词频逆文档频率。其中,TF-IDF是一种用于信息检索与数据挖掘的常用加权技术。TF是词频 (Term Frequency), IDF是逆文本频率指数 (Inverse Document Frequency)。

[0113] 本公开考虑到在句子中每个词的语义对句子的语义的影响程度不同,而每个词向量的TF-IDF值能反应该词在句子中的重要程度,所以在本实施例中,采用TF-IDF对词汇向量进行加权,用词汇向量的加权平均作为句子向量。

[0114] 在一个实施例中,词向量的TF-IDF计算方式如下:

[0115] $TF_w =$ 词汇向量w在某篇文档中出现的次数/该文档的总词数;

[0116] $IDF_w = \log(\text{文档总数}/(\text{含词w的文档总数}+1))$;

[0117] $TF-IDF = TF \times IDF$ 。

[0118] 通过以上的公式定义确定每一个词汇向量在待选文章中的词频逆文档频率。

[0119] 在S708中,根据所述词频逆文档频率与所述词汇向量确定句子的句子向量。由于待选文章中句子中每个词汇的语义对句子的语义的影响程度不同,而每个词的TF-IDF值能反应该词的重要程度,在本公开实施例中用TF-IDF对词进行加权,用词向量的加权平均作为句子向量,具体公式如下:

$$[0120] \quad V_s = \frac{\sum_{w \in s} TF-IDF_w V_w}{\sum_{w \in s} TF-IDF_w}$$

[0121] 其中, V_s 表示句子s的句子向量,s表示待选文章中的某一个句子,w表示词, V_w 表示词w的词向量,TF-IDF_w表示词w的TF-IDF值。

[0122] 在S710中,确定所述多个句子的句子向量之间的相似度。可用句子向量的余弦相似度作为句子间的相似度。

[0123] 在S712中,基于所述相似度构建相似度矩阵。若待选文章含n个句子,则构建一个n×n的相似度矩阵,矩阵中元素 a_{ij} (第i行第j列元素)的值为第i个句子和第j个句子间的相似度。

[0124] 在S714中,将句子作为节点,将相似度作为边的权重,基于所述相似度矩阵构建图模型。

[0125] 其中,所述相似度矩阵可例如为4*4的矩阵,点集合为 (a_1, a_2, a_3, a_4) ,对应的边集合为 $(a_{12}, a_{13}, \dots, a_{34})$:

$$[0126] \quad \begin{vmatrix} 1 & 0.3 & 0.5 & 0.1 \\ 0.3 & 1 & 0.7 & 0.5 \\ 0.5 & 0.7 & 1 & 0.4 \\ 0.1 & 0.5 & 0.4 & 1 \end{vmatrix};$$

[0127] 基于此相似度矩阵所构建的图模型具有4个节点,可分别为 a_1, a_2, a_3, a_4 , a_1, a_2, a_3, a_4 之间对应的矩阵值作为节点间的边的权重。由此,图模型中, a_1, a_2 之间的边 a_{12} 可记为0.3; a_1, a_3 之间的边 a_{13} 可记为0.5; a_1, a_4 之间的边 a_{14} 可记为0.1;以此类推可为图模型中的每一个

节点与节点之间的边进行赋值,以构建图模型。

[0128] 在S716中,基于所述图模型通过迭代获取句子在所述文章中的重要度。可将句子向量之间的余弦相似度作为句子间的相似度,进而迭代计算确定句子在文章中的重要度。

[0129] 例如,可以通过下述公式计算句子之间的相似度,本公开不以此为限。

$$[0130] \quad \text{Similarity}(S_i, S_j) = \frac{\left| \left\{ w_k \mid w_k \in S_i \ \& \ w_k \in S_j \right\} \right|}{\log(|S_i|) + \log(|S_j|)}$$

[0131] 其中, S_i 代表的是第*i*个句子; S_j 代表的是第*j*个句子; w_k 代表的是句子中第*k*个词汇;

[0132] $|S_i|$ 与 $|S_j|$ 代表的是句子中词汇的个数;

[0133] $w_k \in S_i \ \& \ w_k \in S_j$ 代表着同时在 S_i 和 S_j 中出现词汇。

[0134] 根据上述公式,对候选文章中的句子进行遍历迭代计算,以确定句子在所述文章中的相似度。

[0135] 对于本公开中的图模型,即有向有权图 $G = (V, E)$,由点集合 V (句子)和边集合 E (相似度)组成, E 是 $V \times V$ 的子集。图中任两点 V_i, V_j 之间边的权重为 w_{ji} ,对于一个给定的点 V_i , $In(V_i)$ 为指向该点的点集合, $Out(V_i)$ 为点 V_i 指向的点集合。点 V_i (句子)的重要度得分定义如下:

$$[0136] \quad ws(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} ws(V_j)$$

[0137] 其中, $ws(V_i)$ 代表的是 V_i 这个句子的重要度得分, $ws(V_j)$ 代表的是 V_j 句子的重要度得分;

[0138] d 为阻尼系数,取值范围为0到1,代表从图中某一特定点指向其他任意点的概率,可根据经验值进行设置,可取值为0.85;

[0139] $V_j \in In(V_i)$ 代表所有指向 V_i (句子)点的点(句子)集合, $V_k \in Out(V_j)$ 代表所有 V_j 点(句子)指向的点(句子)集合;

[0140] w_{ji} 与 w_{jk} 为 V_i, V_j 与 V_j, V_k 的边的权重。

[0141] 使用TextRank算法计算图中各点的得分时,需要给图中的点指定任意的初值,并递归计算直到收敛,在图中任意一点的误差率小于给定的预设值时就可以达到收敛。

[0142] 在一个实施例中,在迭代计算的过程中,将每个句子的重要度初始化为1。在一个实施例中,判断收敛的预设值取为0.0001。

[0143] 本公开的人机对话方法,在抽取用户输入信息的关键句子时考虑句子的语义。在传统TextRank算法中,句子的相似度计算没有考虑句子的语义,而本公开的人机对话方法用词频-逆文档频率(TF-IDF)作为词的权重,用句子中所包含词的词向量的加权平均作为句子的向量表示,用该向量来计算句子的相似度。

[0144] 值得一提的是,本公开中训练词向量时采用GloVe模型,此处也可以使用其他词向量模型如连续词袋模型CBOW、Skip-Gram等。

[0145] 本公开中用词向量的加权平均作为句子的向量表示,本公开的人机对话方法也可

以采用其他的句子向量表示方法,如Skip-Thought等方法。

[0146] 本公开中使用TextRank来计算文章中句子的重要度时,用到了句子向量间的余弦相似度作为句子间的语义相似度。范数是一种强化了的距离概念,在定义上比距离多了一条数乘的运算法则。在数学上,范数包括向量范数和矩阵范数,向量范数表征向量空间中向量的大小,矩阵范数表征矩阵引起变化的大小。在实际使用中,也可以考虑采用向量间的L1或L2范数等其他距离度量来表示该相似度。

[0147] 本领域技术人员可以理解实现上述实施例的全部或部分步骤被实现为由CPU执行的计算机程序。在该计算机程序被CPU执行时,执行本公开提供的上述方法所限定的上述功能。所述的程序可以存储于一种计算机可读存储介质中,该存储介质可以是只读存储器,磁盘或光盘等。

[0148] 此外,需要注意的是,上述附图仅是根据本公开示例性实施例的方法所包括的处理的示意性说明,而不是限制目的。易于理解,上述附图所示的处理并不表明或限制这些处理的时间顺序。另外,也易于理解,这些处理可以是例如在多个模块中同步或异步执行的。

[0149] 下述为本公开装置实施例,可以用于执行本公开方法实施例。对于本公开装置实施例中未披露的细节,请参照本公开方法实施例。

[0150] 图8是根据一示例性实施例示出的一种人机对话装置的框图。人机对话装置80包括:关键词模块802,文章模块804,评分模块806,以及返回信息模块808。

[0151] 关键词模块802用于由用户的输入信息中提取关键词;可由用户的语音输入信息中提取关键词;或由用户的文字输入信息中提取关键词。

[0152] 文章模块804用于根据所述关键词确定候选文章;可根据所述关键词在预定平台中进行检索以确定所述候选文章。具体可由近期的公众号文章中检索出几篇于关键词相关的热点新闻文章;还可在主流搜索引擎中对关键词进行搜索,根据检索出的结果进行排序,根据排序由高至低由检索结果中确定所述候选文章。其中,候选文章数量可为一个或多个,本公开不以此为限。

[0153] 评分模块806用于获取所述候选文章中的多个句子,并确定所述多个句子中每一个句子的评分;可包括:根据所述多个句子中每一个句子在所述文章中的重要度确定第一评分;根据所述多个句子中每一个句子与所述输入信息的相似性确定第二评分;根据所述第一评分与所述第二评分确定评分。

[0154] 返回信息模块808用于基于评分结果由所述多个句子中提取部分句子生成返回信息。可将所述多个句子按照评分结果进行排序;以及依序由所述多个句子中提取部分句子生成所述返回信息。

[0155] 根据本公开的人机对话装置,通过检索用户问题相关文章,并在相关文章中提取部分语句,生成具有实质内容,包含信息量的返回信息的方式,能够使得机器人在人机对话过程中产生具有信息量的回答,辅助用户与外界信息进行观点交流。

[0156] 图9是根据另一示例性实施例示出的一种人机对话装置的框图。如图9所示的人机对话装置90,在如图8所示的人机对话装置80的基础上还包括:词向量训练模块902。

[0157] 词向量训练模块902用于通过词向量训练算法训练文字数据获取所述词汇向量。在本公开的一个实施例中,可通过GloVe算法对公共网络中的新闻数据或者文章数据进行训练以获取本公开中所用的词汇的词向量。在本公开的一个实施例中,还可通过现有技术

中的其他词向量构建算法获取本公开中所用的词汇的词向量,本公开不以此为限。

[0158] 图10是根据一示例性实施例示出的一种电子设备的框图。

[0159] 下面参照图10来描述根据本公开的这种实施方式的电子设备1000。图10显示的电子设备1000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。

[0160] 如图10所示,电子设备1000以通用计算设备的形式表现。电子设备1000的组件可以包括但不限于:至少一个处理单元1010、至少一个存储单元1020、连接不同系统组件(包括存储单元1020和处理单元1010)的总线1030、显示单元1040等。

[0161] 其中,所述存储单元存储有程序代码,所述程序代码可以被所述处理单元1010执行,使得所述处理单元1010执行本说明书上述电子处方流转处理方法部分中描述的根据本公开各种示例性实施方式的步骤。例如,所述处理单元1010可以执行如图4,图5,图6,以及图7中所示的步骤。

[0162] 所述存储单元1020可以包括易失性存储单元形式的可读介质,例如随机存取存储单元(RAM) 10201和/或高速缓存存储单元10202,还可以进一步包括只读存储单元(ROM) 10203。

[0163] 所述存储单元1020还可以包括具有一组(至少一个)程序模块10205的程序/实用工具10204,这样的程序模块10205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。

[0164] 总线1030可以为表示几类总线结构中的一种或多种,包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

[0165] 电子设备1000也可以与一个或多个外部设备1000' (例如键盘、指向设备、蓝牙设备等)通信,还可与一个或者多个使得用户能与该电子设备1000交互的设备通信,和/或与使得该电子设备1000能与一个或多个其它计算设备进行通信的任何设备(例如路由器、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1050进行。并且,电子设备1000还可以通过网络适配器1060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,如因特网)通信。网络适配器1060可以通过总线1030与电子设备1000的其它模块通信。应当明白,尽管图中未示出,可以结合电子设备1000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。

[0166] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本公开实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、或者网络设备等)执行根据本公开实施方式的上述方法。

[0167] 图11示意性示出本公开示例性实施例中一种计算机可读存储介质示意图。

[0168] 参考图11所示,描述了根据本公开的实施方式的用于实现上述方法的程序产品1100,其可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码,并可以在终端设备,例如个人电脑上运行。然而,本公开的程序产品不限于此,在本文件中,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。

[0169] 上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该计算机可读介质实现如下功能:由用户的输入信息中提取关键词;根据所述关键词确定候选文章;获取所述候选文章中的多个句子,并确定所述多个句子中每一个句子的评分;以及基于评分结果由所述多个句子中提取部分句子生成返回信息。

[0170] 本领域技术人员可以理解上述各模块可以按照实施例的描述分布于装置中,也可以进行相应变化唯一不同于本实施例的一个或多个装置中。上述实施例的模块可以合并为一个模块,也可以进一步拆分成多个子模块。

[0171] 通过以上的实施例的描述,本领域的技术人员易于理解,这里描述的示例实施例可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本公开实施例的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、移动终端、或者网络设备等)执行根据本公开实施例的方法。

[0172] 以上具体地示出和描述了本公开的示例性实施例。应可理解的是,本公开不限于这里描述的详细结构、设置方式或实现方法;相反,本公开意图涵盖包含在所附权利要求的精神和范围内的各种修改和等效设置。

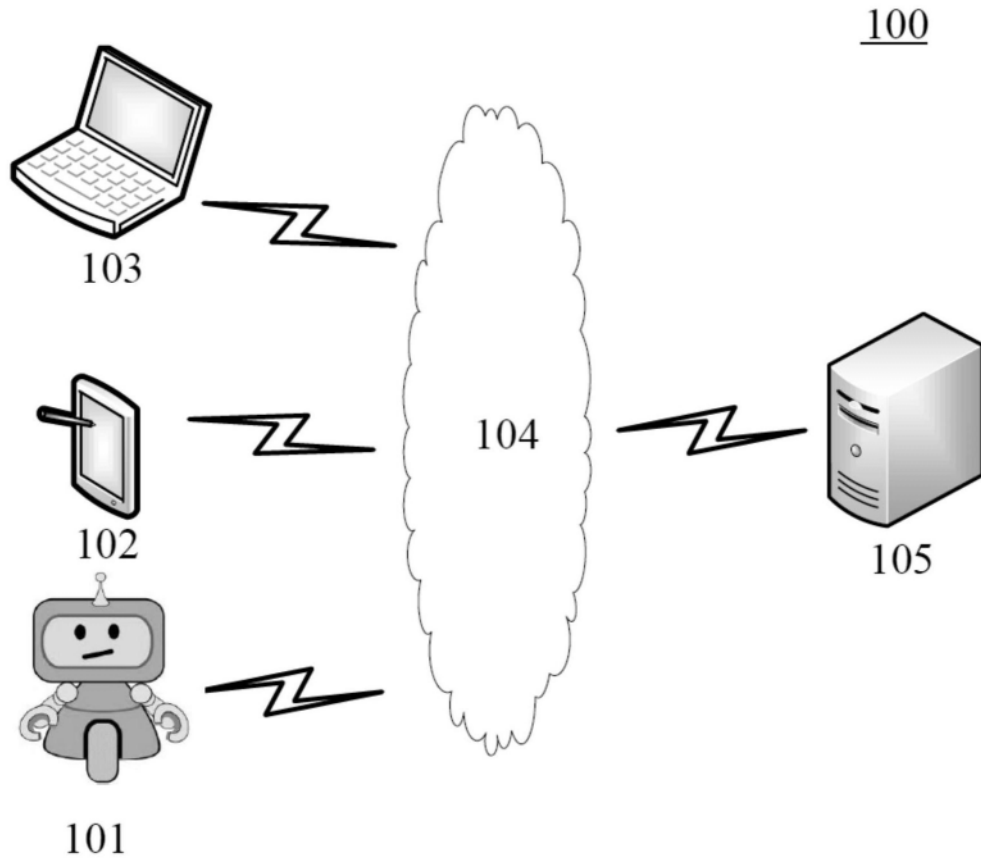


图1

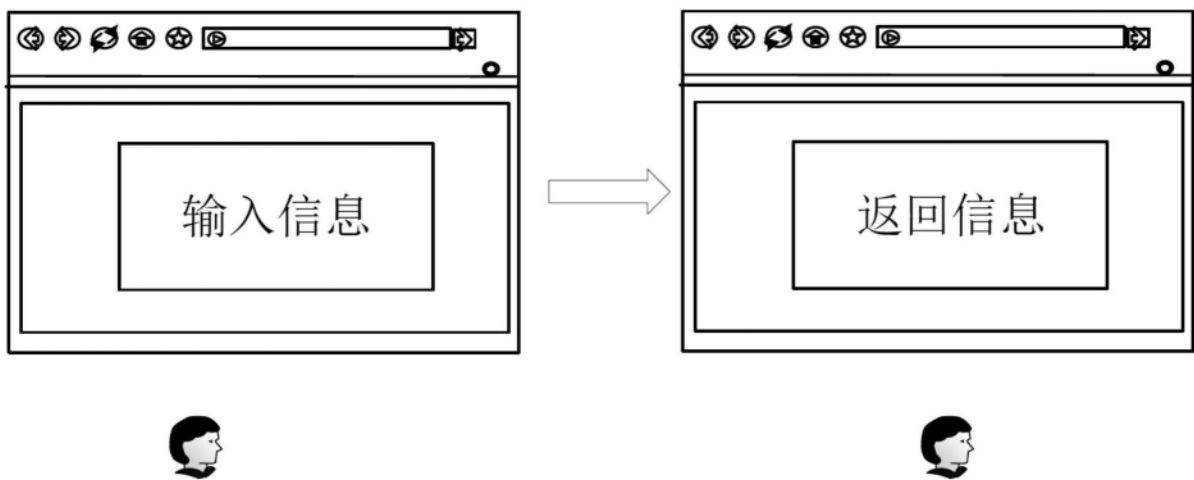


图2

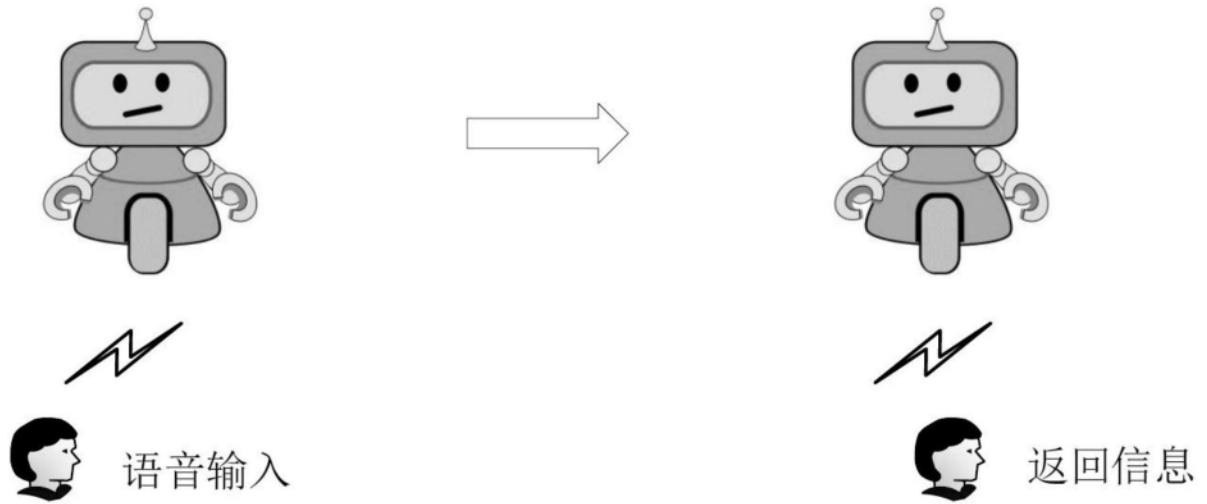


图3

40

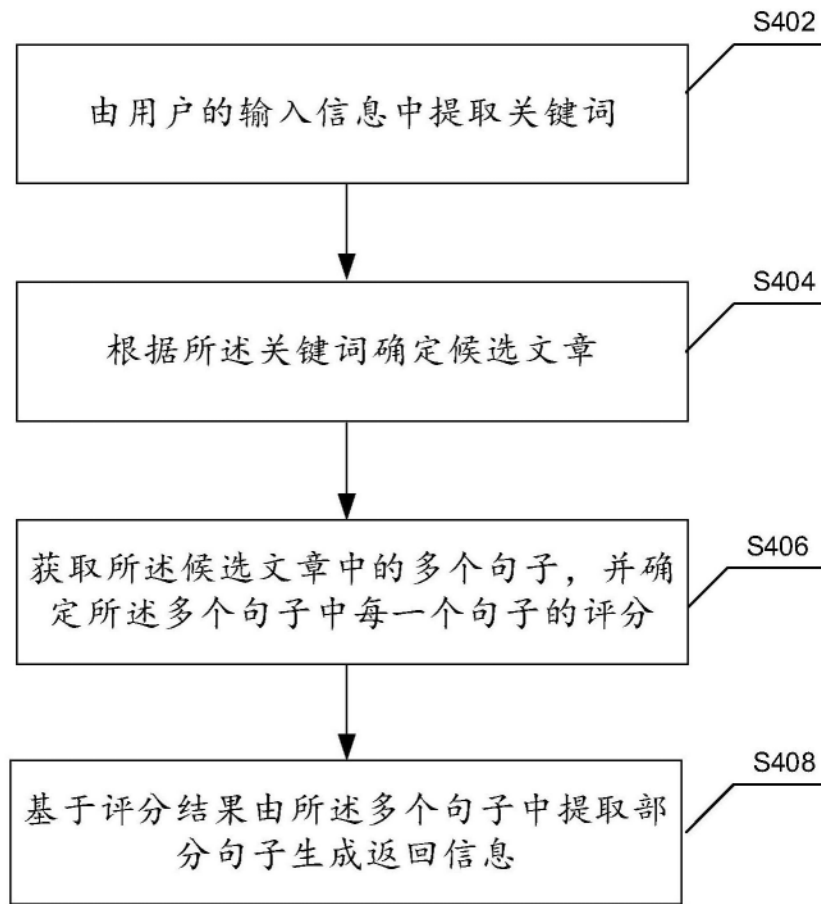


图4

50

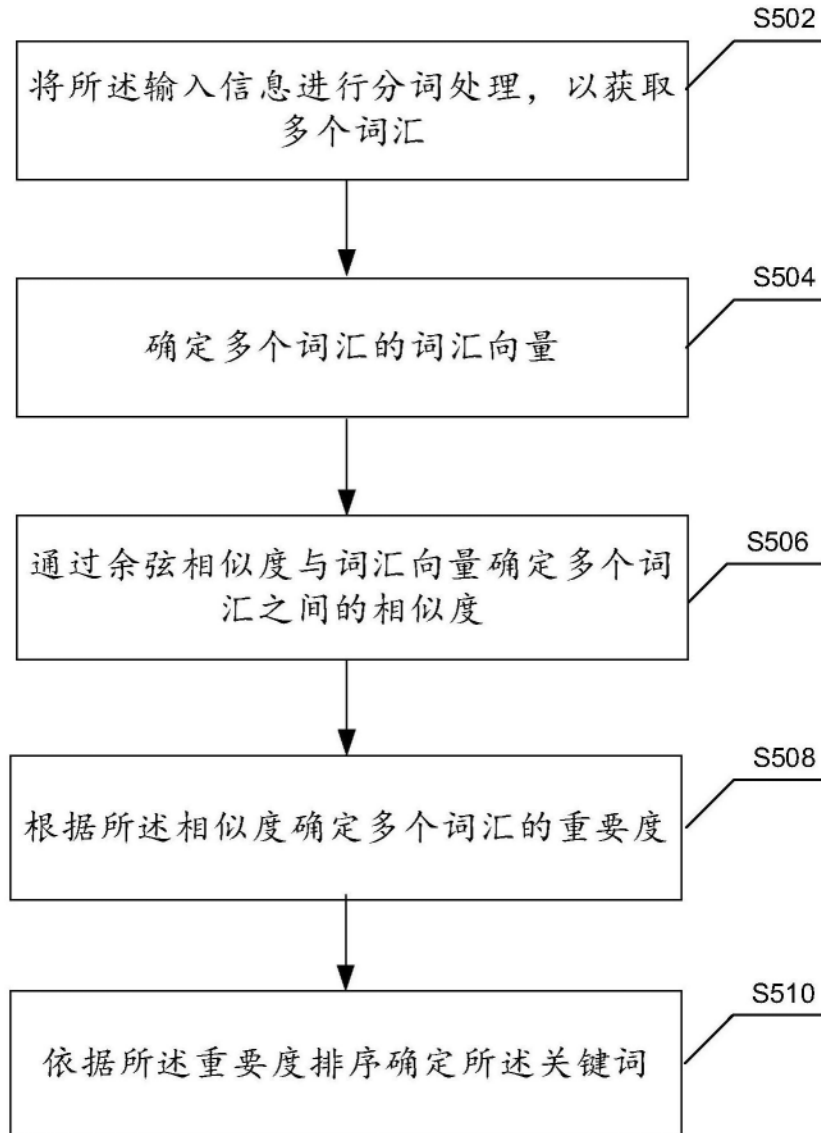


图5

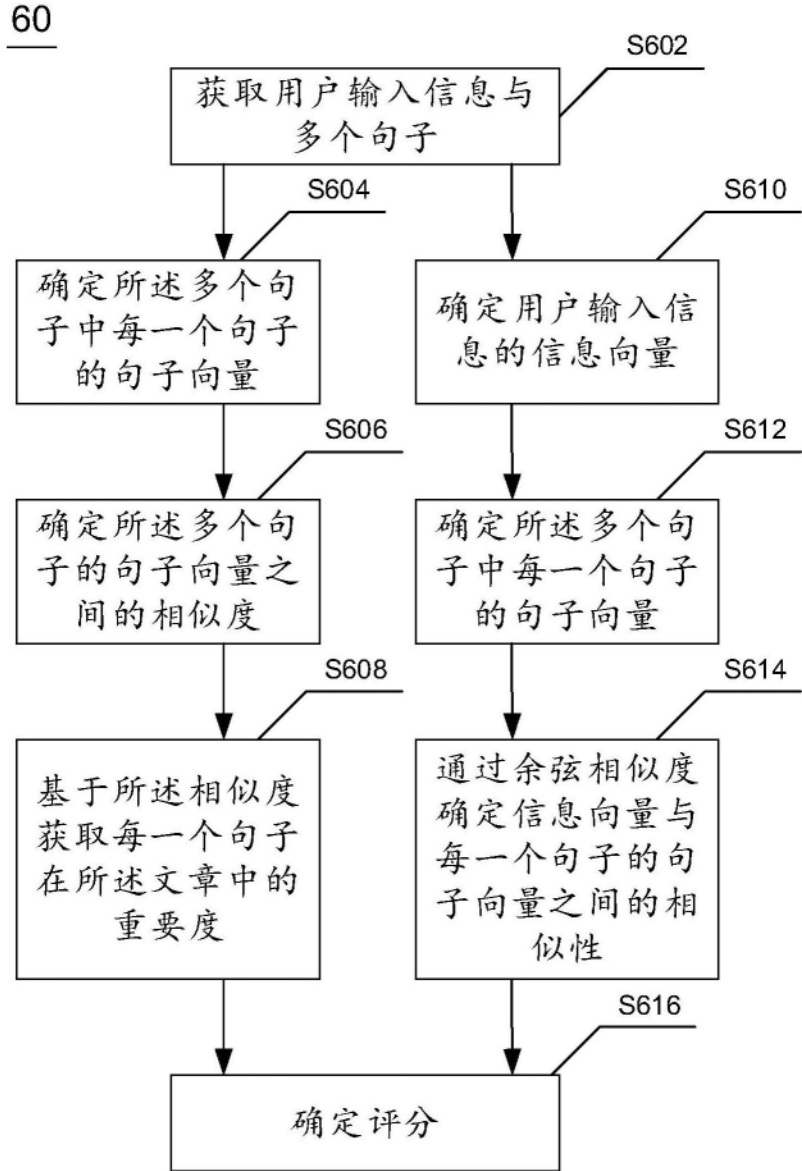


图6

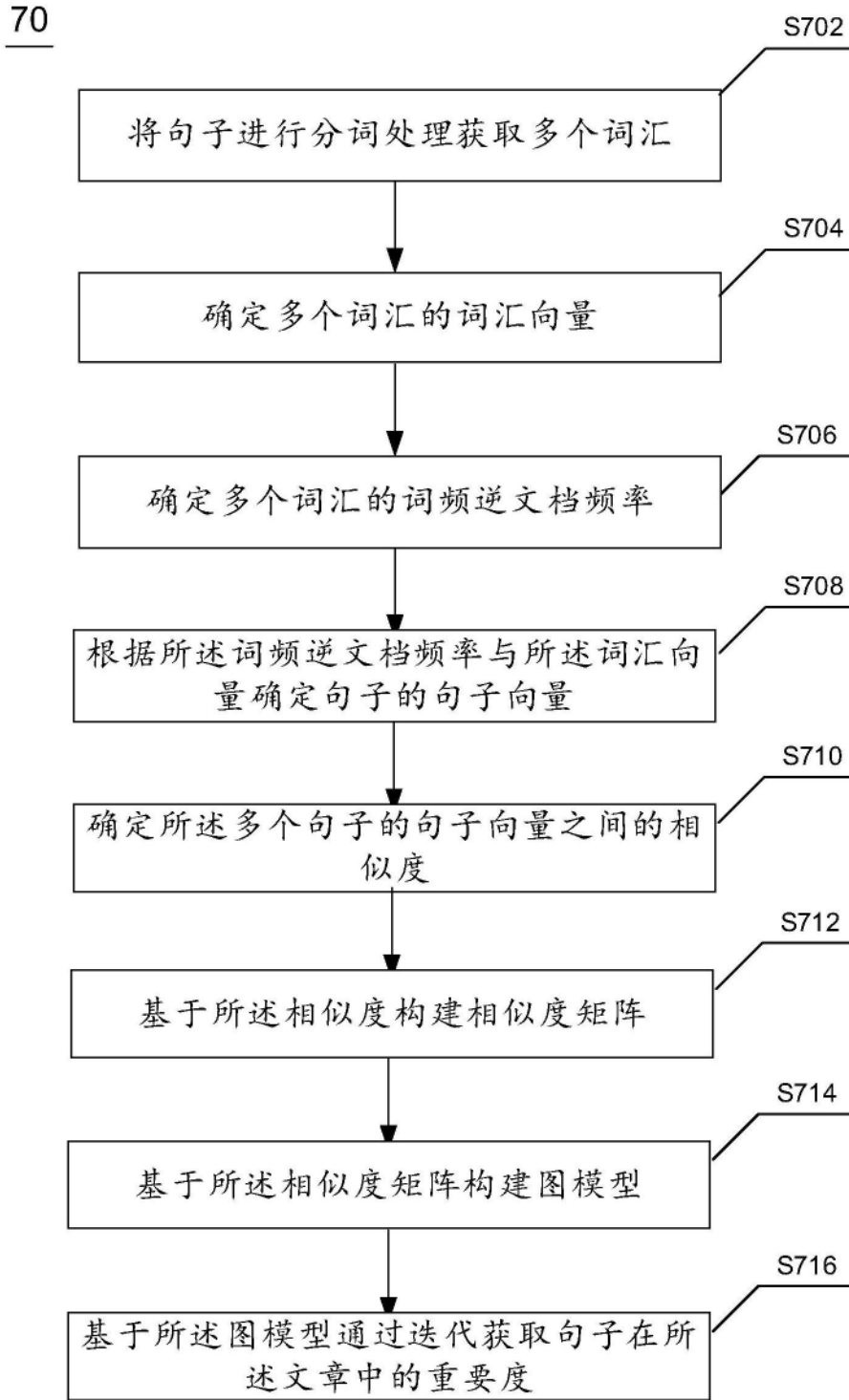


图7

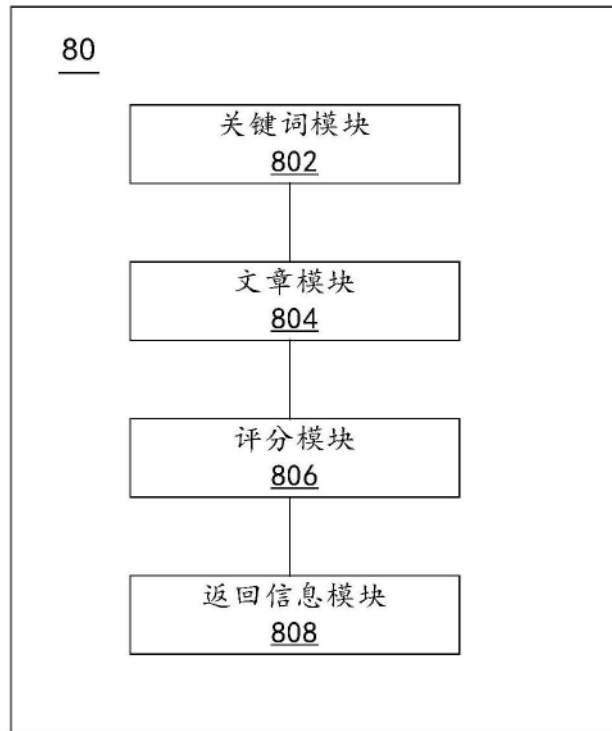


图8

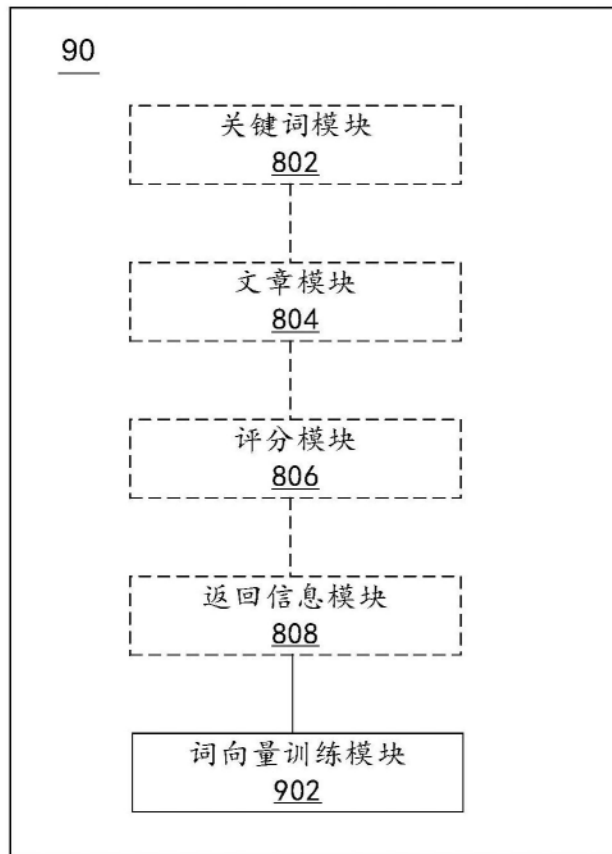


图9

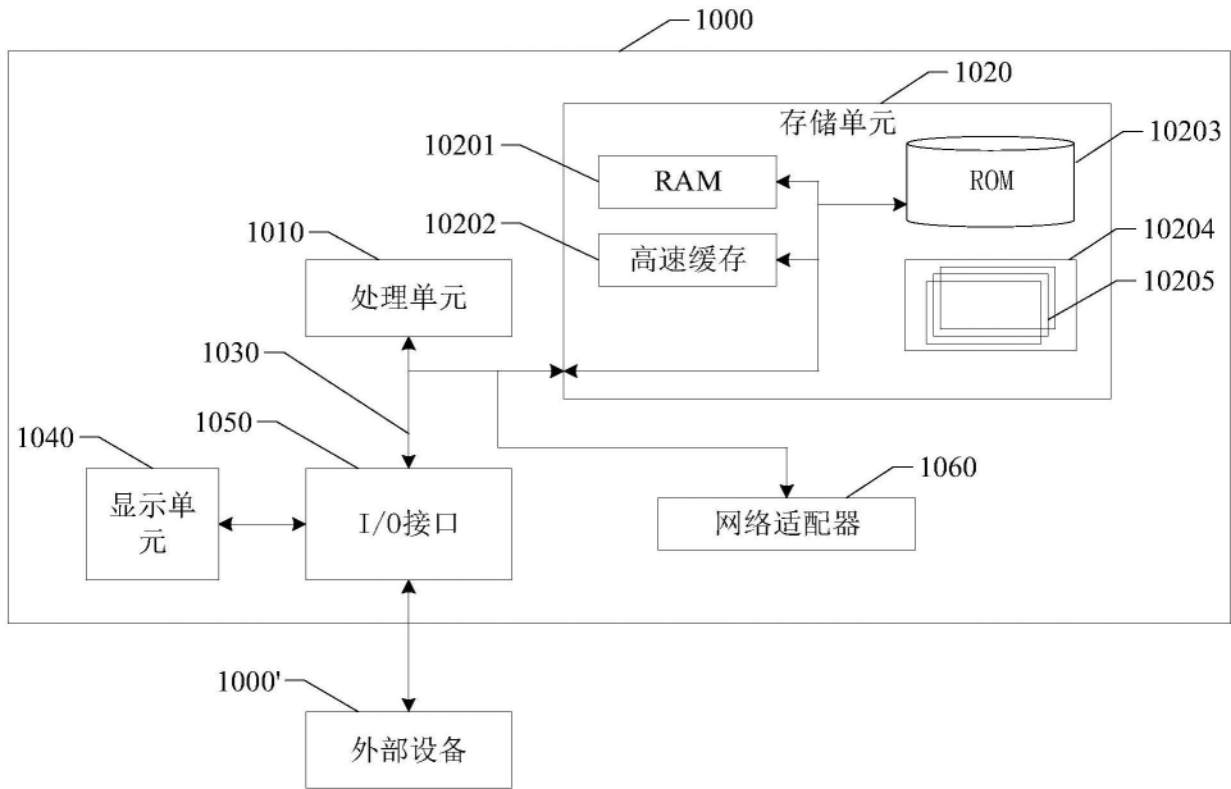


图10

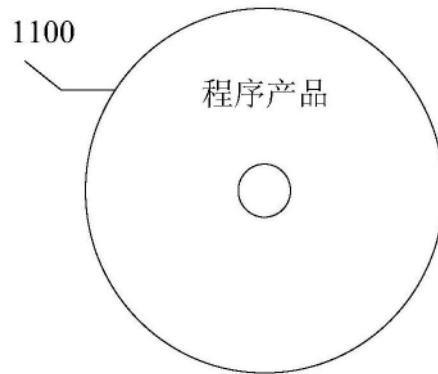


图11