



(12) 发明专利申请

(10) 申请公布号 CN 113761143 A

(43) 申请公布日 2021. 12. 07

(21) 申请号 202011275591.5

G06Q 30/02 (2012.01)

(22) 申请日 2020.11.13

(71) 申请人 北京京东尚科信息技术有限公司
地址 100086 北京市海淀区知春路76号8层
申请人 北京京东世纪贸易有限公司

(72) 发明人 刘航 陈蒙

(74) 专利代理机构 中原信达知识产权代理有限
责任公司 11219
代理人 郭晗 赵迪

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 40/289 (2020.01)

G06F 40/30 (2020.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

权利要求书2页 说明书11页 附图6页

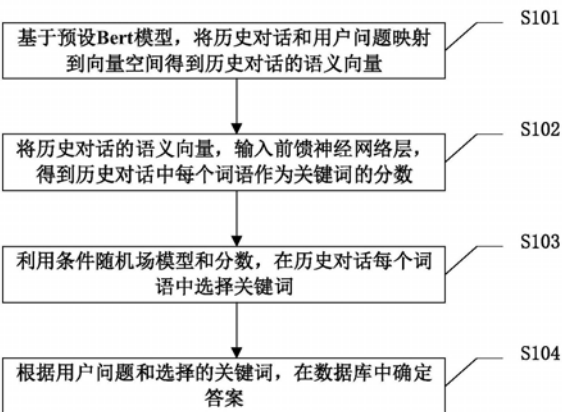
(54) 发明名称

为用户问题确定答案的方法、装置、设备和
介质

(57) 摘要

本发明公开了为用户问题确定答案的方法、装置、设备和计算机可读介质,涉及计算机技术领域。该方法的一具体实施方式包括:基于预设Bert模型,将历史对话和用户问题映射到向量空间得到历史对话的语义向量;将所述历史对话的语义向量,输入前馈神经网络层,得到所述历史对话中每个词语作为关键词的分数;利用条件随机场模型和所述分数,在所述历史对话每个词语中选择关键词;根据所述用户问题和选择的关键词,在数据库中确定答案。该实施方式能够提高根据用户问题确定答案的准确性。

100



1. 一种为用户问题确定答案的方法,其特征在于,包括:

基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量;

将所述历史对话的语义向量,输入前馈神经网络层,得到所述历史对话中每个词语作为关键词的分数;

利用条件随机场模型和所述分数,在所述历史对话每个词语中选择关键词;

根据所述用户问题和选择的关键词,在数据库中确定答案。

2. 根据权利要求1所述为用户问题确定答案的方法,其特征在于,所述根据所述用户问题和选择的关键词,在数据库中确定答案,包括:

将所述选择的关键词插入所述用户问题中,以补全所述用户问题,得到用户完整问题;

基于所述用户完整问题,在数据库中确定答案。

3. 根据权利要求2所述为用户问题确定答案的方法,其特征在于,将所述选择的关键词插入所述用户问题中,以补全所述用户问题,得到用户完整问题,包括:

将一个所述选择的关键词分别插入所述用户问题的多个词语中,得到多个候选用户问题;

采用N-gram语言模型计算所述多个候选用户问题的流畅度,并以最高流畅度的候选用户问题更新所述用户问题;

将以一个或多个所述选择的关键词更新后的用户问题,作为所述用户完整问题。

4. 根据权利要求3所述为用户问题确定答案的方法,其特征在于,所述将一个所述选择的关键词分别插入所述用户问题的多个词语中,包括:

将一个所述选择的关键词分别插入所述用户问题的所有词语中。

5. 根据权利要求3所述为用户问题确定答案的方法,其特征在于,将以一个或多个所述选择的关键词更新后的用户问题,作为所述用户完整问题,包括:

将以所有所述选择的关键词更新后的用户问题,作为所述用户完整问题。

6. 根据权利要求1所述为用户问题确定答案的方法,其特征在于,所述基于预设Bert模型,将历史对话的文本和用户问题的文本映射到向量空间得到所述历史对话的语义向量之前,还包括:

在原始历史对话中进行格式处理和/或文字处理,得到所述历史对话,所述文字处理包括删除预设模板词语和/或预设场景词语。

7. 根据权利要求1所述为用户问题确定答案的方法,其特征在于,所述基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量,包括:

以所述历史对话中的词语和所述用户问题中的词语,构建输入词语;

拼接所述输入词语的词向量、所述输入词语的位置向量和所述输入词语的片段向量得到输入向量,所述位置向量是所述词向量中每个词语的位置向量的集合,所述片段向量是所述词向量中每个词语的片段向量的集合;

将所示输入向量输入所述预设Bert模型,得到所述历史对话的语义向量。

8. 一种为用户问题确定答案的装置,其特征在于,包括:

映射模块,用于基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量;

评分模块,用于将所述历史对话的语义向量,输入前馈神经网络层,得到所述历史对话中每个词语作为关键词的分数;

选择模块,用于利用条件随机场模型和所述分数,在所述历史对话每个词语中选择关键词;

确定模块,用于根据所述用户问题和选择的关键词,在数据库中确定答案。

9.一种为用户问题确定答案的电子设备,其特征在于,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1-7中任一所述的方法。

10.一种计算机可读介质,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1-7中任一所述的方法。

为用户问题确定答案的方法、装置、设备和介质

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种为用户问题确定答案的方法、装置、设备和计算机可读介质。

背景技术

[0002] 对话系统作为人类与机器交流的一个重要入口,在个人生活和企业服务中发挥着越来越重要的作用。

[0003] 随着深度学习技术的发展,对话系统逐渐成熟并普及到各个行业领域。在电商领域,对话系统可以快速解答用户问题并提供咨询服务,极大地提升了用户体验,同时也为电商企业节省了大量的成本。

[0004] 在实现本发明过程中,发明人发现现有技术中至少存在如下问题:由于根据用户问题确定答案的准确性较低,导致对话系统的准确率较低。

发明内容

[0005] 有鉴于此,本发明实施例提供一种为用户问题确定答案的方法、装置、设备和计算机可读介质,能够提高根据用户问题确定答案的准确性。

[0006] 为实现上述目的,根据本发明实施例的一个方面,提供了一种为用户问题确定答案的方法,包括:

[0007] 基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量;

[0008] 将所述历史对话的语义向量,输入前馈神经网络层,得到所述历史对话中每个词语作为关键词的分数;

[0009] 利用条件随机场模型和所述分数,在所述历史对话每个词语中选择关键词;

[0010] 根据所述用户问题和选择的关键词,在数据库中确定答案。

[0011] 所述根据所述用户问题和选择的关键词,在数据库中确定答案,包括:

[0012] 将所述选择的关键词插入所述用户问题中,以补全所述用户问题,得到用户完整问题;

[0013] 基于所述用户完整问题,在数据库中确定答案。

[0014] 将所述选择的关键词插入所述用户问题中,以补全所述用户问题,得到用户完整问题,包括:

[0015] 将一个所述选择的关键词分别插入所述用户问题的多个词语中,得到多个候选用户问题;

[0016] 采用N-gram语言模型计算所述多个候选用户问题的流畅度,并以最高流畅度的候选用户问题更新所述用户问题;

[0017] 将以一个或多个所述选择的关键词更新后的用户问题,作为所述用户完整问题。

[0018] 所述将一个所述选择的关键词分别插入所述用户问题的多个词语中,包括:

- [0019] 将一个所述选择的关键词分别插入所述用户问题的所有词语中。
- [0020] 将以一个或多个所述选择的关键词更新后的用户问题,作为所述用户完整问题,包括:
- [0021] 将以所有所述选择的关键词更新后的用户问题,作为所述用户完整问题。
- [0022] 所述基于预设Bert模型,将历史对话的文本和用户问题的文本映射到向量空间得到所述历史对话的语义向量之前,还包括:
- [0023] 在原始历史对话中进行格式处理和/或文字处理,得到所述历史对话,所述文字处理包括删除预设模板词语和/或预设场景词语。
- [0024] 所述基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量,包括:
- [0025] 以所述历史对话中的词语和所述用户问题中的词语,构建输入词语;
- [0026] 拼接所述输入词语的词向量、所述输入词语的位置向量和所述输入词语的片段向量得到输入向量,所述位置向量是所述词向量中每个词语的位置向量的集合,所述片段向量是所述词向量中每个词语的片段向量的集合;
- [0027] 将所示输入向量输入所述预设Bert模型,得到所述历史对话的语义向量。
- [0028] 根据本发明实施例的第二方面,提供了一种为用户问题确定答案的装置,包括:
- [0029] 映射模块,用于基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量;
- [0030] 评分模块,用于将所述历史对话的语义向量,输入前馈神经网络层,得到所述历史对话中每个词语作为关键词的分数;
- [0031] 选择模块,用于利用条件随机场模型和所述分数,在所述历史对话每个词语中选择关键词;
- [0032] 确定模块,用于根据所述用户问题和选择的关键词,在数据库中确定答案。
- [0033] 根据本发明实施例的第三方面,提供了一种为用户问题确定答案的电子设备,包括:
- [0034] 一个或多个处理器;
- [0035] 存储装置,用于存储一个或多个程序,
- [0036] 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述的方法。
- [0037] 根据本发明实施例的第四方面,提供了一种计算机可读介质,其上存储有计算机程序,所述程序被处理器执行时实现如上述的方法。
- [0038] 上述发明中的一个实施例具有如下优点或有益效果:基于预设Bert模型,将历史对话和用户问题映射到向量空间得到历史对话的语义向量;将历史对话的语义向量,输入前馈神经网络层,得到历史对话中每个词语作为关键词的分数;利用条件随机场模型和分数,在历史对话每个词语中选择关键词;根据用户问题和选择的关键词,在数据库中确定答案。所确定的答案,不仅依据用户问题还考虑到所选择的关键词,所选择的关键词能够体现用户问题涉及的应用场景。结合应用场景所确定的答案,能够提高答案的准确性。
- [0039] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

附图说明

[0040] 附图用于更好地理解本发明,不构成对本发明的不当限定。其中:

[0041] 图1是根据本发明实施例的为用户问题确定答案的方法的主要流程的示意图;

[0042] 图2是根据本发明实施例的将历史对话和用户问题映射到向量空间得到历史对话的语义向量的流程示意图;

[0043] 图3是根据本发明实施例的补全用户问题以确定答案的流程示意图;

[0044] 图4是根据本发明实施例的将关键词插入用户问题中以补全用户问题的流程示意图;

[0045] 图5是根据本发明实施例的为用户问题确定答案的方法的应用流程的示意图;

[0046] 图6是根据本发明实施例的用户与机器人对话内容的示意图;

[0047] 图7是根据本发明实施例的为用户问题确定答案的装置的主要结构的示意图;

[0048] 图8是本发明实施例可以应用于其中的示例性系统架构图;

[0049] 图9是适于用来实现本发明实施例的终端设备或服务器的计算机系统的结构示意图。

具体实施方式

[0050] 以下结合附图对本发明的示范性实施例做出说明,其中包括本发明实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本发明的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0051] 对话系统在应答用户问题(query)之前,需要准确理解用户问题的语义和意图。将用户问题与数据库中的知识点进行语义匹配,选取语义上最接近的知识点作为答案以解答用户问题。其中,用户问题与知识点的语义匹配是重要环节。

[0052] 采用语义匹配模型,分别将用户问题与数据库中的知识点映射到相同的向量空间,然后通过计算向量之间的分数,筛选出最合适的答案。即,通过获取向量之间的关联性,实现系统的高效应答。

[0053] 在多轮对话的系统中,用户往往会根据历史对话内容进行提问,用户问题中普遍存在句子成分缺失的情况,例如“还有吗?”、“什么意思?”等。句子成分缺失直接导致用户问题语义模糊和意图不明,使得无法匹配到正确的知识点,严重影响根据用户问题确定答案的准确性,导致对话系统的准确率较低。

[0054] 为了解决根据用户问题确定答案的准确性较低,可以采用以下本发明实施例中的技术方案。

[0055] 参见图1,图1是根据本发明实施例的为用户问题确定答案的方法主要流程的示意图,将历史对话和用户问题映射到向量空间,以在历史对话中的词语中选择关键词;然后根据用户问题和选择的关键词确定答案。如图1所示,具体包括以下步骤:

[0056] S101、基于预设Bert模型,将历史对话和用户问题映射到向量空间得到历史对话的语义向量。

[0057] 为了解决根据用户问题确定答案的准确性较低,在本发明实施例中结合用户问题所涉及的应用场景补充问题。在实际应用过程中,发现历史对话可以表征用户问题所涉及

的应用场景。那么,具体可以结合历史对话补充用户问题。

[0058] 历史对话是用户发送用户问题之前的对话文本。作为一个示例,在对话系统以语音方式实现人机交流的情况下,可以将历史对话的语音转换为历史对话。作为另一个示例,在对话系统以文字方式实现人机交流的情况下,可以将历史对话中涉及的所有文本作为历史对话。

[0059] 在历史对话中包括多个词语,将历史对话和用户问题,进行语义匹配,以达到弥补用户问题的信息缺失。但是在真实对话场景中,历史对话中的文本信息复杂多样,历史对话中包含大量的无效信息。无效信息干扰语义匹配,严重影响语义匹配的准确性,进而难以精准确定答案。

[0060] 那么,需要从历史对话中的所有词语中挑选出关键词,再以关键词和用户问题确定答案。

[0061] 为了确定关键词,需要将历史对话和用户问题映射到向量空间。在本发明实施例中,基于预设Bert模型实现,将历史对话和用户问题映射到向量空间。

[0062] 变换器的双向编码器表征量(Bidirectional Encoder Representations from Transformers,Bert)模型是一种语言模型。Bert模型的目标是利用大规模无标注语料训练、获得包含丰富语义信息的文本的语义。Bert模型应用于问答、命名实体识别、自然语言推理和文本分类等。

[0063] 在本发明实施例中,预设Bert模型属于已完成训练的模型。为了应用于不同的领域,可以利用相关领域的的数据训练。作为一个示例,本发明实施例中的方案应用于电商领域,则可以利用电商领域的的数据训练Bert模型得到预设Bert模型。作为另一个示例,本发明实施例中的方案应用于通信领域,则可以利用通信领域的的数据训练Bert模型得到预设Bert模型。

[0064] 考虑到历史对话信息冗长,为了提高处理历史对话的速度,可以提前处理原始历史对话。具体包括格式处理和文字处理。在实际应用中,提前处理原始历史对话,可以采用格式处理和/或文字处理,进而得到历史对话。可以理解的是,原始历史对话是未经处理的历史对话的文本。历史对话是经处理的历史对话的文本。

[0065] 格式处理是按照预设方式处理文本中的字符。格式处理可以包括以下一种或多种:删除标点符号,删除表情符号,替换历史对话中的各种链接,删除网页标签和将繁体汉字转换为简体汉字。

[0066] 文字处理是在文本中删除词语。作为一个示例,文字处理包括删除预设模板词语和/或预设场景词语。预设模板词语是预先设置的词语。预设场景词语是在预设使用场景中涉及的词语。

[0067] 预设模板词语是以字符串的方式匹配模板词语。如:预设模板词语包括闲聊术语和转人工术语。闲聊术语包括:“你好”、“谢谢”和“欢迎光临,有什么可以帮您的”等。转人工术语包括:“有人吗”和“人工客服”等。

[0068] 预设场景词语可以包括商品描述语。商品描述语通常是由客服发送,而且出现情况比较固定的词语。在用户发送一款商品的链接后,客服会推送相应的商品描述语。另外,商品描述语的文本长度较长,一般会超过50个字符,而且为了突出商品的卖点,商品描述语中包含特殊标签。特殊标签是html网页标签。通过发送者的角色,文本出现的场景、文本的

长度和特殊标签等方面,可以确定历史对话内容是否为商品描述语。

[0069] 参见图2,图2是根据本发明实施例的将历史对话和用户问题映射到向量空间得到历史对话的语义向量的流程示意图,具体包括以下步骤:

[0070] S201、以历史对话中的词语和用户问题中的词语,构建输入词语。

[0071] 在文本到向量空间的映射过程中,历史对话中词语 $w_{context}^i$ 、用户问题中词语 w_{query}^j 和文本分隔符<SEP>,将上述三者拼接得到输入词向量。

[0072] 其中,在历史对话中包括n个词语,i代表第i个词语,n是大于等于1的整数;在用户问题中包括m个词语,j代表第j个词语,m是大于等于1的整数。

[0073] 输入词语中包括三部分,文本分隔符位于其他两部分中间。 $w_{context}^i$ 位于 w_{query}^j 之前,或 w_{query}^j 位于 $w_{context}^i$ 之前。参见公式(1),公式(1)中 w_{input} 是输入词语。

[0074] $w_{input} = w_{context}^1, \dots, w_{context}^n, \langle SEP \rangle, w_{query}^1, \dots, w_{query}^m$ 公式(1)

[0075] S202、拼接输入词语的词向量、位置向量和片段向量得到输入向量,位置向量是上述词向量中每个词语位置向量的集合,片段向量是上述词向量中每个词语片段向量的集合。

[0076] 根据语料预先建立的词表,将输入词语中每个词语映射为词向量,得到输入词语的词向量 e_{input}^{i+j} 。

[0077] 除此之外,为了区分每个词语在输入词语中的位置,如:词汇2在输入词语中位置是从左至右的第二位。输入词语中的每个词语被映射为位置向量 p_{input}^{i+j} 。也就是说,位置向量是输入词语的词向量中每个词语位置向量的集合。

[0078] 还可以区分每个词语在输入词语中的片段。片段包括历史对话中的词语和用户问题中的词语。输入词语中的每个词语被映射为片段向量 s_{input}^{i+j} 。即,片段向量是输入词语的词向量中每个词语片段向量的集合。

[0079] 拼接输入词语的词向量、位置向量和片段向量得到输入向量 $Input^i$,

[0080] $Input^{i+j} = concat(e_{input}^{i+j}; p_{input}^{i+j}; s_{input}^{i+j})$ 公式(2)

[0081] 需要说明的是,输入向量中输入词语的词向量、位置向量和片段向量可以按照任意顺序拼接。这是因为三个向量相加,对于顺序没有要求。

[0082] S203、将输入向量输入预设Bert模型,得到历史对话的语义向量。

[0083] 将输入向量输入预设Bert模型,通过Bert模型对历史对话的n个词语进行语义编码,得到历史对话的语义向量。可以理解的是,历史对话的语义向量是包含全局信息的语义向量。也就是说,历史对话的语义向量是包括历史对话和用户问题的语义向量。

[0084] 在图2的实施例中,通过预设Bert模型,得到历史对应的语义向量。

[0085] S102、将历史对话的语义向量,输入前馈神经网络层,得到历史对话中每个词语作为关键词的分数。

[0086] 历史对话的语义向量包含全局信息,将历史对话的语义向量,输入前馈神经网络层,得到历史对话中每个词语作为关键词的分数 $score^i$ 。

[0087] $\text{score}^i = \text{NN}(\text{BERT}(\text{Input}^i))$ 公式 (3)

[0088] 其中, $\text{NN}(\cdot)$ 表示前馈神经网络层, Input^i 表示输入中的第 i 个词语。

[0089] 前馈神经网络层是在 Bert 模型上添加的网络结构。前馈神经网络层是人工神经网络层的一种。前馈神经网络层采用一种单向多层结构。其中每一层包含若干个神经元。

[0090] 对于历史对话中每个词语来说, 均得到作为关键词的分数。一个词语作为关键词的分数和作为非关键词的分数的总和是 1。那么, 进而可知历史对话中每个词语作为非关键词的分数。

[0091] S103、利用条件随机场模型和分数, 在历史对话每个词语中选择关键词。

[0092] 条件随机场 (conditional random field, CRF) 模型, 是一种鉴别式机率模型, 是随机场的一种, 常用于标注或分析序列资料。

[0093] 在本发明实施例中, 利用 CRF 模型在历史对话的词语中做出全局决策。即, 从所有关键词路径中寻找一条分数最高的关键词路径进行输出, 并在上述分数最高的关键词路径中标记出历史对话中的每个词语是否为关键词。上述分数是指历史对话中每个词语作为关键词的分数。

[0094] 其中, 关键词路径可以采用 viterbi 算法得到。关键词路径可理解为候选的关键词路径, 并非最终的关键词路径。关键词路径中每个词语都有一个标签, 标识该词语是否为关键词。分数最高的关键词路径所包含的关键词为最终的关键词。

[0095] S104、根据用户问题和选择的关键词, 在数据库中确定答案。

[0096] 在历史对话中选择关键词后, 就可以根据用户问题和选择的关键词, 在数据库中确定答案。与仅依据用户问题确定答案相比而言, 由于关键词充分呈现用户问题的应用场景, 用户问题结合关键词所确定的答案, 准确性较高。

[0097] 在本发明的一个实施例中, 可以将所选择的关键词插入到用户问题中, 以便在数据库中确定答案。

[0098] 参见图 3, 图 3 是根据本发明实施例的补全用户问题以确定答案的流程示意图, 具体包括以下步骤:

[0099] S301、将选择的关键词插入用户问题中, 以补全用户问题, 得到用户完整问题。

[0100] 从历史对话中获取关键词后, 需要将这些关键词补全到用户问题中, 并且插入到合适位置, 使得补全后的用户问题在信息完善的同时, 仍然能够保持语言上的连贯流畅性。

[0101] 在本发明的一个实施例中, 可以基于 N-gram 语言模型, 补全用户问题。

[0102] 参见图 4, 图 4 是根据本发明实施例的将关键词插入用户问题中以补全用户问题的流程示意图, 具体包括以下步骤:

[0103] S401、将一个选择的关键词分别插入用户问题的多个词语中, 得到多个候选用户问题。

[0104] 所选择的关键词的数量是一个或多个。可以将一个选择的关键词分别插入用户问题的多个词语中。也就是说, 将一个关键词插入用户问题的一个词语中, 就能够得到一个候选用户问题。

[0105] 作为一个示例, 用户问题有 m 个词语, 可以得到 $m+1$ 个候选用户问题。即, 将关键词插入到用户问题 $(w_{query}^1, \dots, w_{query}^m)$ 中所有可能的位置上, 得到 $m+1$ 个候选用户问题。

[0106] S402、采用 N-gram 语言模型计算多个候选用户问题的流畅度, 并以最高流畅度的

候选用户问题更新用户问题。

[0107] 为了从多个候选用户问题选择出最佳的候选用户问题,可以采用N-gram语言模型计算多个候选用户问题的流畅度。然后,以最高流畅度的候选用户问题更新用户问题。也就是说,以最高流畅度的候选用户问题作为用户问题。

[0108] N-gram模型是一种语言模型,语言模型是一个基于概率的判别模型,它的输入是一句话,输出是这句话的概率,即这些单词的联合概率。

[0109] 具体来说,N-gram语言模型对m+1个候选用户问题分别评分,计算候选用户问题的流畅度。候选用户问题的评分越高,则流畅度越高。也就是说,将一个选择的关键词分别插入用户问题的所有词语中。

[0110] 按照流畅度的高低排序,使用最高流畅度的候选用户问题更新用户问题,即视为关键词插入到用户问题中,用户问题的词语数量变成了m+1。

[0111] S403、将以一个或多个选择的关键词更新后的用户问题,作为用户完整问题。

[0112] 对于多个关键词而言,可以按照S401和S402,在更新后的用户问题的基础上,继续更新用户问题。将以一个或多个选择的关键词更新后的用户问题,作为用户完整问题。

[0113] 作为一个示例,将以所有选择的关键词更新后的用户问题,作为用户完整问题。可以理解的是,将所有选择的关键词均插入用户问题,进而得到用户完整问题。

[0114] 在补全用户问题的过程中,可以将所有选择的关键词作为关键词集合。从关键词集合中选择一个关键词按照S401和S402插入到用户问题中。然后,再次从剩余的关键词集合中再次选择一个关键词按照S401和S402插入到用户问题中。类似地,直到关键词集合为空,则将以最后一个关键词更新后的用户问题,作为用户完整问题。

[0115] 在图4实施例中,将所选择的关键词插入用户问题中,通过计算候选用户问题的流畅度更新用户问题,进而得到用户完整问题。实现所选择的关键词与用户问题的充分结合。

[0116] S302、基于用户完整问题,在数据库中确定答案。

[0117] 基于用户完整问题,能够在数据中准确确定答案。

[0118] 在上述实施例中,基于预设Bert模型,将历史对话和用户问题映射到向量空间得到历史对话的语义向量;将历史对话的语义向量,输入前馈神经网络层,得到历史对话中每个词语作为关键词的分数;利用条件随机场模型和分数,在历史对话中每个词语中选择关键词;根据用户问题和选择的关键词,在数据库中确定答案。所确定的答案,不仅依据用户问题还考虑到所选择的关键词,所选择的关键词能够体现用户问题涉及的应用场景。结合应用场所确定的答案,能够提高答案的准确性。

[0119] 参见图5,图5是根据本发明实施例的为用户问题确定答案的方法的应用流程的示意图,具体包括以下步骤:

[0120] S501、用户问题。

[0121] 用户通过对话系统与机器客服沟通。参见图6,图6是根据本发明实施例的用户与机器人对话内容的示意图。

[0122] 在用户与机器客户的沟通中,用户提出用户问题。作为一个示例,图6中用户问题包括:什么意思。

[0123] S502、处理原始历史对话。

[0124] 图6中的用户问题:什么意思,在不结合用户问题涉及的应用场景,则无法确定该

用户问题的准确答案。

[0125] 那么,需要处理原始对话。在图6中原始对话是指用户问题上方的5句对话。具体地,可以采用格式处理和/或文字处理,进而得到历史对话。

[0126] S503、以历史对话补全用户问题。

[0127] 然后采用S101至S104的方案,以历史对话补全用户问题。图6中虚线框中的用户问题,即补全后的用户问题。补全后的用户问题是:电辅热是什么意思。

[0128] S504、语义匹配。

[0129] 基于补全后的用户问题,在数据库中确定答案。

[0130] S505、答案。

[0131] 由于补全后的用户问题是结合具体的应用场景,因而所确定的答案的准确性更高。

[0132] 参见图7,图7是根据本发明实施例的为用户问题确定答案的装置的主要结构的示意图,为用户问题确定答案的装置可以实现为用户问题确定答案的方法,如图7所示,为用户问题确定答案的装置具体包括:

[0133] 映射模块701,用于基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量;

[0134] 评分模块702,用于将所述历史对话的语义向量,输入前馈神经网络层,得到所述历史对话中每个词语作为关键词的分数;

[0135] 选择模块703,用于利用条件随机场模型和所述分数,在所述历史对话每个词语中选择关键词;

[0136] 确定模块704,用于根据所述用户问题和选择的关键词,在数据库中确定答案。

[0137] 在本发明的一个实施例中,确定模块704,具体用于将所述选择的关键词插入所述用户问题中,以补全所述用户问题,得到用户完整问题;

[0138] 基于所述用户完整问题,在数据库中确定答案。

[0139] 在本发明的一个实施例中,确定模块704,具体用于将一个所述选择的关键词分别插入所述用户问题的多个词语中,得到多个候选用户问题;

[0140] 采用N-gram语言模型计算所述多个候选用户问题的流畅度,并以最高流畅度的候选用户问题更新所述用户问题;

[0141] 将以一个或多个所述选择的关键词更新后的用户问题,作为所述用户完整问题。

[0142] 在本发明的一个实施例中,确定模块704,具体用于将一个所述选择的关键词分别插入所述用户问题的所有词语中。

[0143] 在本发明的一个实施例中,确定模块704,具体用于将以所有所述选择的关键词更新后的用户问题,作为所述用户完整问题。

[0144] 在本发明的一个实施例中,映射模块701,还用于在原始历史对话中进行格式处理和/或文字处理,得到所述历史对话,所述文字处理包括删除预设模板词语和/或预设场景词语。

[0145] 在本发明的一个实施例中,映射模块701,具体用于以所述历史对话中的词语和所述用户问题中的词语,构建输入词语;

[0146] 拼接所述输入词语的词向量、所述输入词语的位置向量和所述输入词语的片段向

量得到输入向量,所述位置向量是所述词向量中每个词语的位置向量的集合,所述片段向量是所述词向量中每个词语的片段向量的集合;

[0147] 将所示输入向量输入所述预设Bert模型,得到所述历史对话的语义向量。

[0148] 图8示出了可以应用本发明实施例的为用户问题确定答案的方法或为用户问题确定答案的装置的示例性系统架构800。

[0149] 如图8所示,系统架构800可以包括终端设备801、802、803,网络804和服务器805。网络804用以在终端设备801、802、803和服务器805之间提供通信链路的介质。网络804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。

[0150] 用户可以使用终端设备801、802、803通过网络804与服务器805交互,以接收或发送消息等。终端设备801、802、803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(仅为示例)。

[0151] 终端设备801、802、803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、平板电脑、膝上型便携计算机和台式计算机等等。

[0152] 服务器805可以是提供各种服务的服务器,例如对用户利用终端设备801、802、803所浏览的购物类网站提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的产品信息查询请求等数据进行分析等处理,并将处理结果(例如目标推送信息、产品信息--仅为示例)反馈给终端设备。

[0153] 需要说明的是,本发明实施例所提供的为用户问题确定答案的方法一般由服务器805执行,相应地,为用户问题确定答案的装置一般设置于服务器805中。

[0154] 应该理解,图8中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0155] 下面参考图9,其示出了适于用来实现本发明实施例的终端设备的计算机系统900的结构示意图。图9示出的终端设备仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0156] 如图9所示,计算机系统900包括中央处理单元(CPU)901,其可以根据存储在只读存储器(ROM)902中的程序或者从存储部分908加载到随机访问存储器(RAM)903中的程序而执行各种适当的动作和处理。在RAM 903中,还存储有系统900操作所需的各种程序和数据。CPU 901、ROM 902以及RAM 903通过总线904彼此相连。输入/输出(I/O)接口905也连接至总线904。

[0157] 以下部件连接至I/O接口905:包括键盘、鼠标等的输入部分906;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分907;包括硬盘等的存储部分908;以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分909。通信部分909经由诸如因特网的网络执行通信处理。驱动器910也根据需要连接至I/O接口905。可拆卸介质911,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器910上,以便于从其上读出的计算机程序根据需要被安装入存储部分908。

[0158] 特别地,根据本发明公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本发明公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分909从网络上被下载和安装,和/或从可

拆卸介质911被安装。在该计算机程序被中央处理单元(CPU) 901执行时,执行本发明的系统中限定的上述功能。

[0159] 需要说明的是,本发明所示的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或者是上述两者的任意组合。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可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(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。在本发明中,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。而在本发明中,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。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。

[0160] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0161] 描述于本发明实施例中涉及到的模块可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块也可以设置在处理器中,例如,可以描述为:一种处理器包括映射模块、评分模块、选择模块和确定模块。其中,这些模块的名称在某种情况下并不构成对该模块本身的限定,例如,映射模块还可以被描述为“用于基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量”。

[0162] 作为另一方面,本发明还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。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,使得该设备包括:

[0163] 基于预设Bert模型,将历史对话和用户问题映射到向量空间得到所述历史对话的语义向量;

[0164] 将所述历史对话的语义向量,输入前馈神经网络层,得到所述历史对话中每个词语作为关键词的分数;

[0165] 利用条件随机场模型和所述分数,在所述历史对话每个词语中选择关键词;

[0166] 根据所述用户问题和选择的关键词,在数据库中确定答案。

[0167] 根据本发明实施例的技术方案,基于预设Bert模型,将历史对话和用户问题映射到向量空间得到历史对话的语义向量;将历史对话的语义向量,输入前馈神经网络层,得到历史对话中每个词语作为关键词的分数;利用条件随机场模型和分数,在历史对话每个词语中选择关键词;根据用户问题和选择的关键词,在数据库中确定答案。所确定的答案,不仅依据用户问题还考虑到所选择的关键词,所选择的关键词能够体现用户问题涉及的应用场景。结合应用场景所确定的答案,能够提高答案的准确性。

[0168] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,取决于设计要求和因素,可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

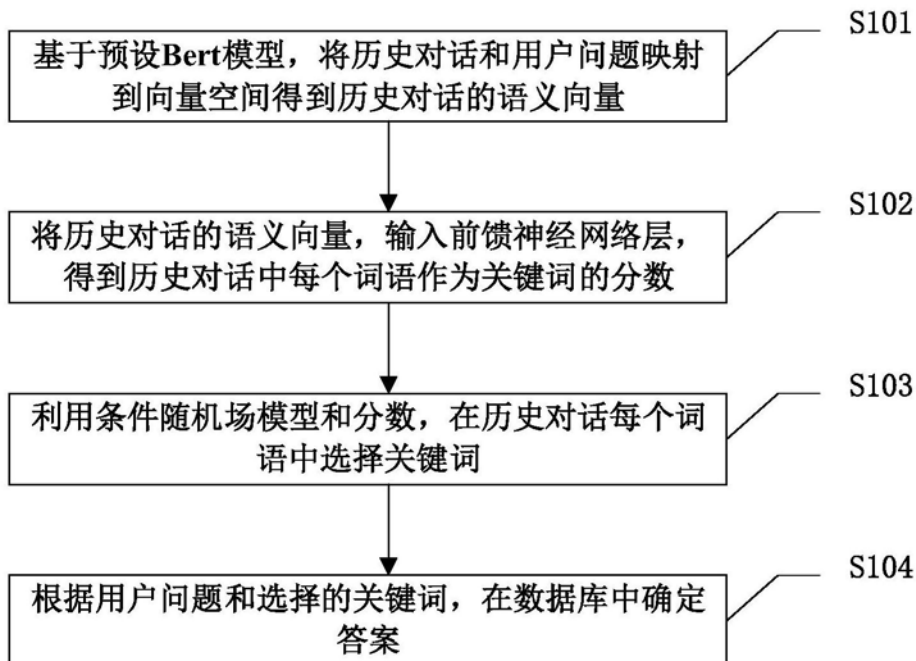
100

图1

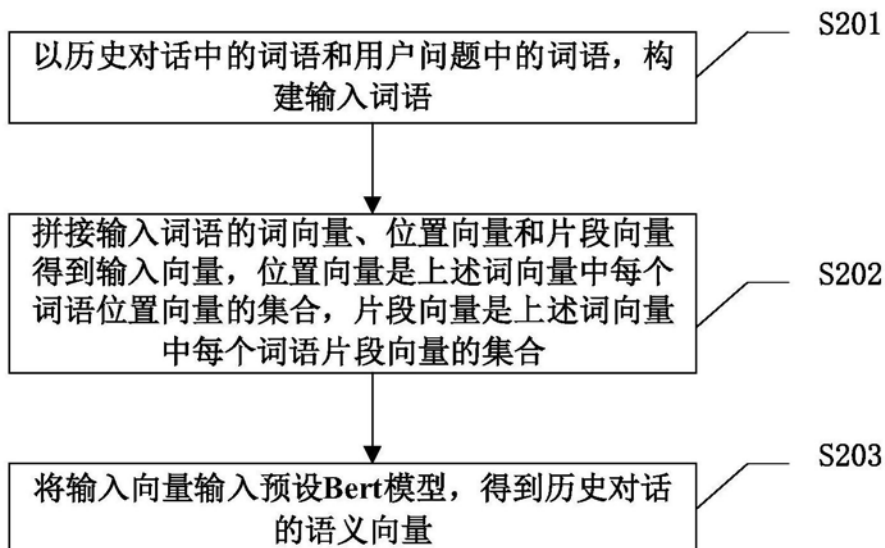
200

图2

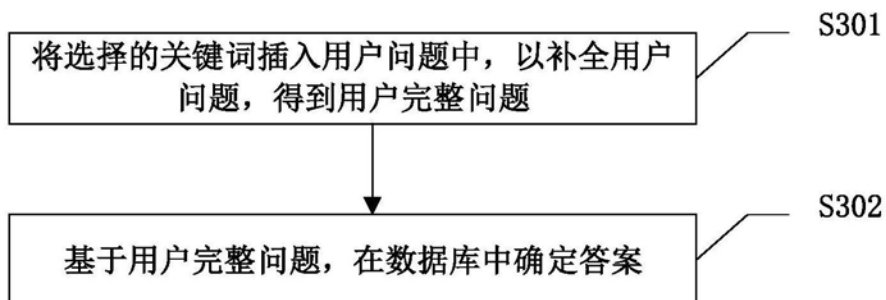
300

图3

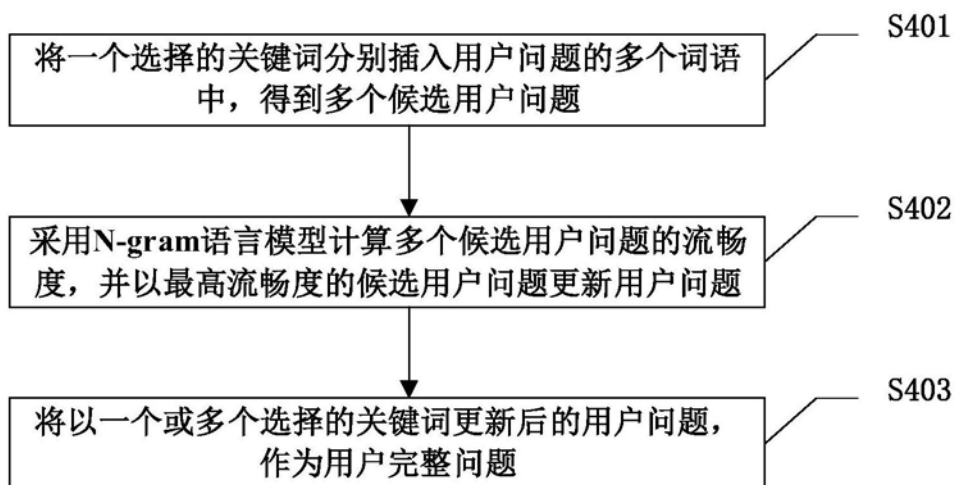
400

图4

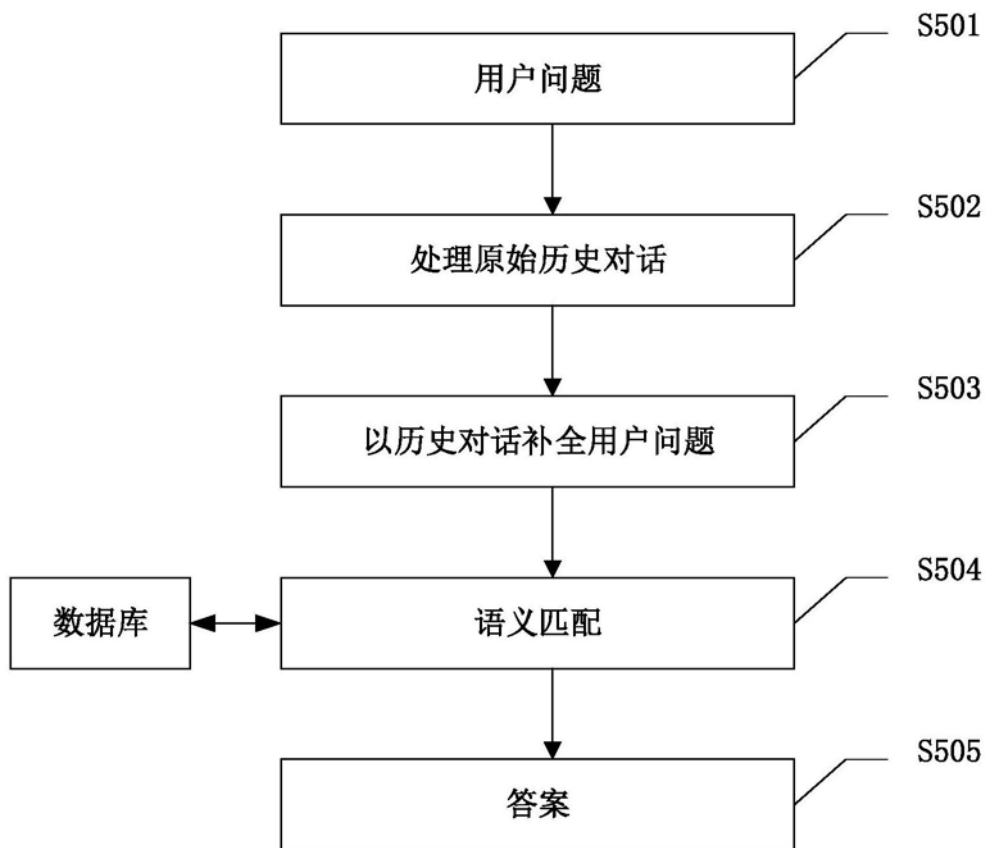
500

图5

600

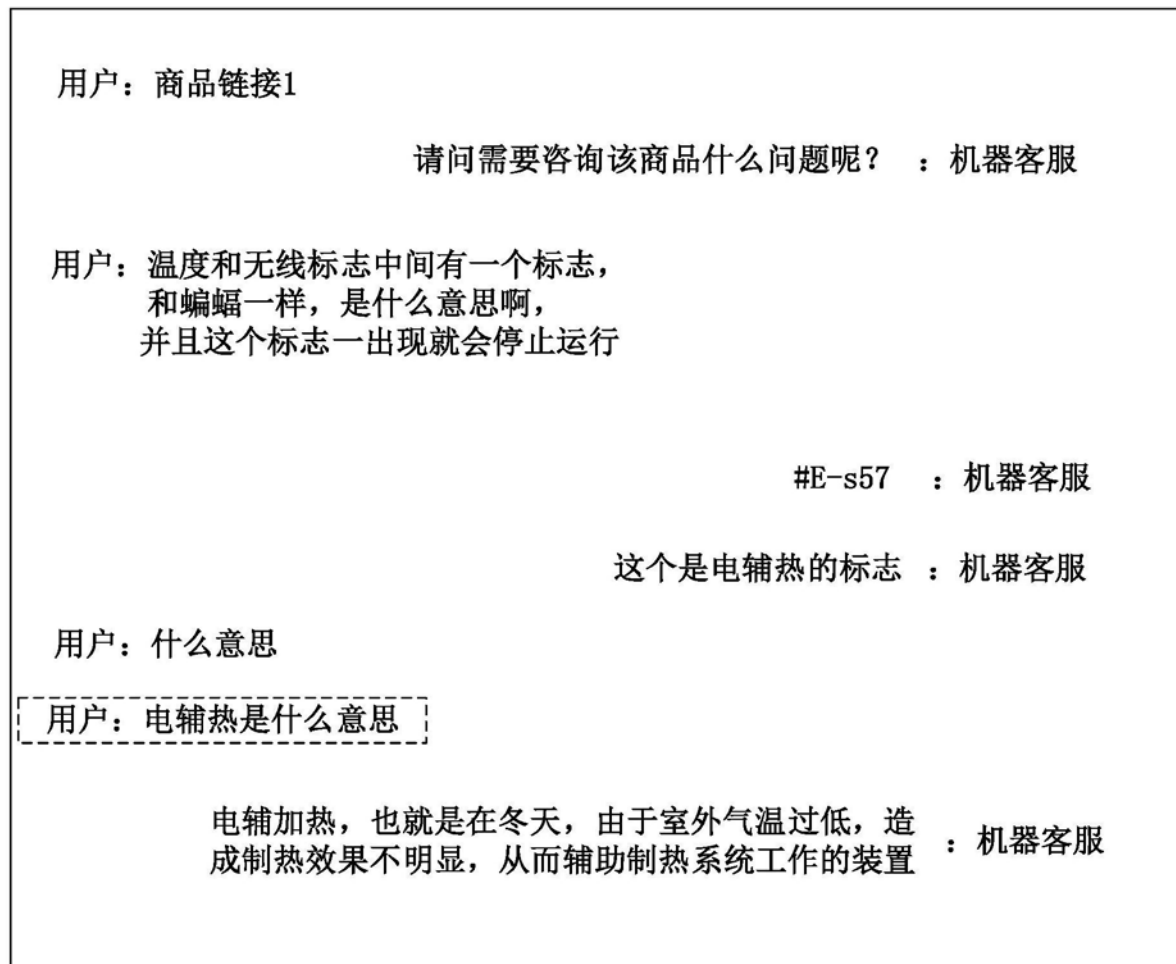


图6

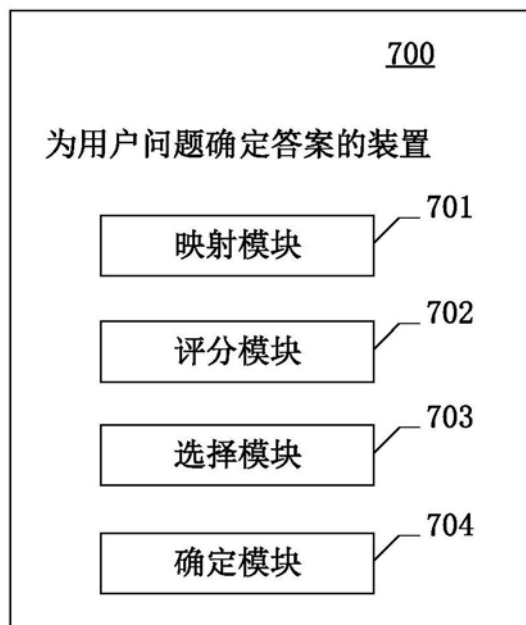


图7

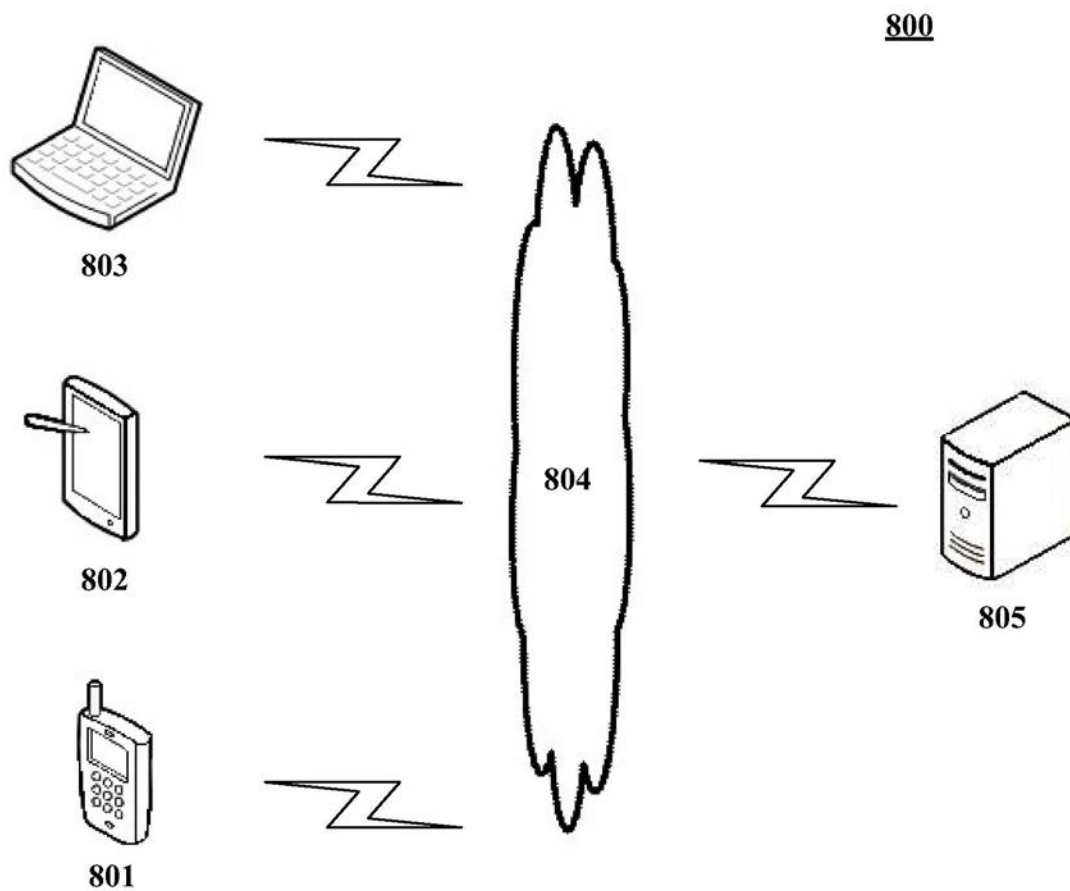


图8

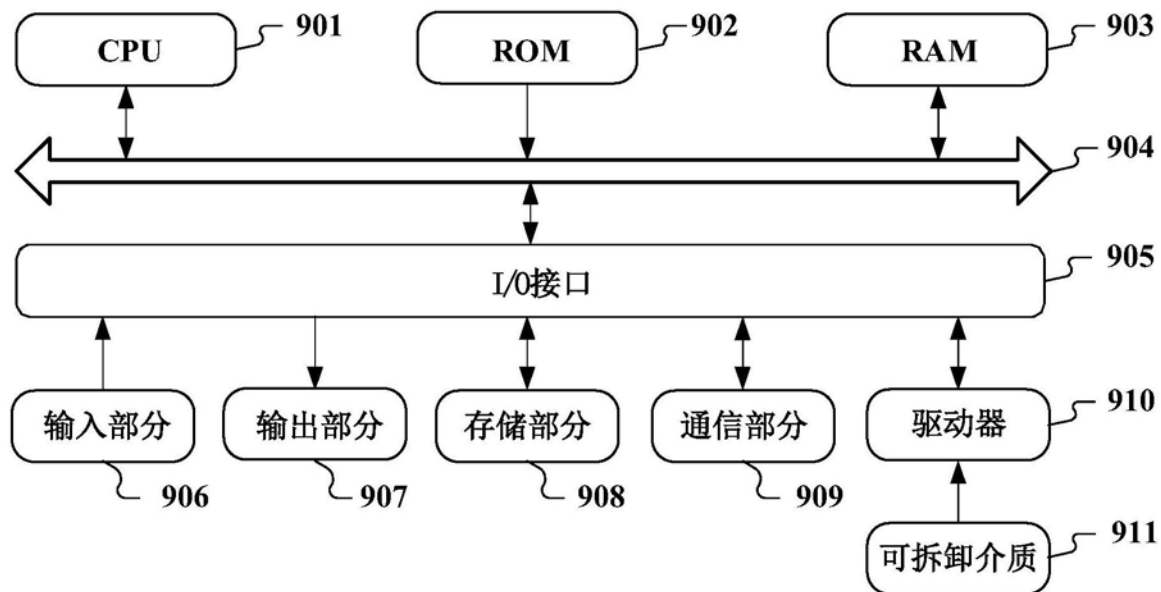
900

图9