



(12) 发明专利申请

(10) 申请公布号 CN 112347770 A

(43) 申请公布日 2021. 02. 09

(21) 申请号 202011282527.X

(22) 申请日 2020.11.16

(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 刘瑞雪 刘航 陈蒙

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021

代理人 赵婷

(51) Int. Cl.

G06F 40/279 (2020.01)

G06F 16/33 (2019.01)

G06F 16/332 (2019.01)

G06K 9/62 (2006.01)

权利要求书2页 说明书12页 附图4页

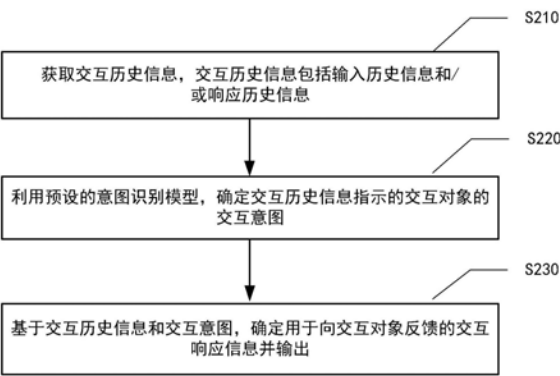
(54) 发明名称

交互处理方法及装置

(57) 摘要

本公开提供了一种交互处理方法,本方法包括:获取交互历史信息,所述交互历史信息包括输入历史信息和/或响应历史信息;利用预设的意图识别模型,确定所述交互历史信息指示的交互对象的交互意图;基于所述交互历史信息和所述交互意图,确定用于向所述交互对象反馈的交互响应信息并输出。本公开还提供了一种交互处理装置、一种电子设备及一种计算机可读存储介质。

200



1. 一种交互处理方法, 包括:

获取交互历史信息, 所述交互历史信息包括输入历史信息和/或响应历史信息;

利用预设的意图识别模型, 确定所述交互历史信息指示的交互对象的交互意图;

基于所述交互历史信息和所述交互意图, 确定用于向所述交互对象反馈的交互响应信息并输出。

2. 根据权利要求1所述的方法, 其中, 所述基于所述交互历史信息和所述交互意图, 确定用于向所述交互对象反馈的交互响应信息, 包括:

根据交互意图与意图描述词之间的预设关联关系, 确定与所述交互历史信息关联的至少一个意图描述词, 其中, 所述关联关系是利用意图识别模型基于交互样本数据训练得到的;

基于所述交互历史信息和所述至少一个意图描述词, 确定所述交互响应信息。

3. 根据权利要求2所述的方法, 其中, 所述基于所述交互历史信息和所述至少一个意图描述词, 确定所述交互响应信息, 包括:

确定与所述交互意图关联的交互候选信息库, 所述交互候选信息库包括至少一个输入候选信息和至少一个响应候选信息;

基于所述交互历史信息和所述至少一个意图描述词, 在所述交互候选信息库中确定与所述交互历史信息匹配度最高的目标输入候选信息;

将与所述目标输入候选信息关联的响应候选信息, 作为所述交互响应信息。

4. 根据权利要求3所述的方法, 其中, 所述基于所述交互历史信息和所述至少一个意图描述词, 在所述交互候选信息库中确定与所述交互历史信息匹配度最高的目标输入候选信息, 包括:

针对任一输入候选信息, 确定所述交互历史信息与所述输入候选信息间的第一匹配度;

确定所述至少一个意图描述词与所述输入候选信息间的第二匹配度;

基于所述第一匹配度和所述第二匹配度, 确定针对所述输入候选信息的第三匹配度;

重复以上步骤, 直至确定出针对各所述输入候选信息的第三匹配度;

至少基于针对各所述输入候选信息的第三匹配度, 确定所述目标输入候选信息。

5. 根据权利要求4所述的方法, 其中, 所述至少基于针对各所述输入候选信息的第三匹配度, 确定所述目标输入候选信息, 包括:

针对任一所述输入候选信息, 计算所述至少一个意图描述词的平均词向量与所述输入候选信息的平均词向量间的相似度, 以作为针对所述输入候选信息的第四匹配度;

计算所述输入候选信息中包含的意图描述词个数, 以作为针对所述输入候选信息的第五匹配度;

计算所述输入候选信息中的所有组成词与所述至少一个意图描述词间的平均相似度, 以作为针对所述输入候选信息的第六匹配度;

基于所述第三匹配度、第四匹配度、第五匹配度和第六匹配度, 确定所述目标输入候选信息。

6. 根据权利要求2所述的方法, 其中, 所述基于所述交互历史信息和所述至少一个意图描述词, 确定所述交互响应信息, 包括:

利用预设的交互信息生成模型,进行基于所述交互历史信息的编码操作,得到中间特征向量;以及

进行基于所述中间特征向量和所述至少一个意图描述词的解码操作,以生成所述交互响应信息。

7.根据权利要求6所述的方法,在所述交互信息生成模型的训练过程中,包括:

利用所述交互信息生成模型,进行基于交互样本信息的编码操作,得到样本中间特征向量,所述交互样本信息包括样本输入信息和样本响应信息,所述交互样本信息具有指示交互意图的样本标签;

进行基于所述样本中间特征向量的解码操作,得到预测响应信息;

基于所述预测响应信息和所述样本响应信息,计算所述交互信息生成模型的生成损失;

基于所述样本标签和所述预测响应信息指示的交互意图,计算所述交互信息生成模型的分类损失;

基于所述生成损失和所述分类损失,确定所述交互信息生成模型的总损失,以进行基于所述总损失的模型训练。

8.一种交互处理装置,包括:

获取模块,用于获取交互历史信息,所述交互历史信息包括输入历史信息和/或响应历史信息;

第一处理模块,用于利用预设的意图识别模型,确定所述交互历史信息指示的交互对象的交互意图;

第二处理模块,用于基于所述交互历史信息和所述交互意图,确定用于向所述交互对象反馈的交互响应信息并输出。

9.一种电子设备,包括:

一个或多个处理器;以及

存储器,用于存储一个或多个程序,

其中,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述一个或多个处理器实现权利要求1至7中任一项所述的方法。

10.一种计算机可读存储介质,其上存储有可执行指令,该指令被处理器执行时使处理器实现权利要求1至7中任一项所述的方法。

交互处理方法及装置

技术领域

[0001] 本公开涉及人工智能技术领域,更具体地,涉及一种交互处理方法、一种交互处理装置、一种电子设备以及一种计算机可读存储介质。

背景技术

[0002] 随者人工智能技术的迅速发展,智能交互在不同应用领域得以广泛渗透,智能交互能够为用户提供高效的交互服务,这不仅有利于提升基于智能交互的业务运行效率,还有利于节省平台用于与用户交互的人力成本。

[0003] 在实现本公开技术构思过程中,发明人发现相关技术的智能交互中,多基于与用户进行交互的交互历史数据,确定用于向用户反馈的交互响应信息,这存在确定出的交互响应信息可能与用户交互意图不匹配的问题,其不仅影响智能交互效率,还会影响智能交互效果。

发明内容

[0004] 有鉴于此,本公开提供了一种交互匹配性高、交互效率得以有效改善的交互处理方法及装置。

[0005] 本公开的一个方面提供了一种交互处理方法,包括:获取交互历史信息,上述交互历史信息包括输入历史信息 and/or 响应历史信息;利用预设的意图识别模型,确定上述交互历史信息指示的交互对象的交互意图;基于上述交互历史信息和上述交互意图,确定用于向上述交互对象反馈的交互响应信息并输出。

[0006] 可选地,上述基于上述交互历史信息和上述交互意图,确定用于向上述交互对象反馈的交互响应信息,包括:根据交互意图与意图描述词之间的预设关联关系,确定与上述交互历史信息关联的至少一个意图描述词,其中,上述关联关系是利用意图识别模型基于交互样本数据训练得到的;基于上述交互历史信息和上述至少一个意图描述词,确定上述交互响应信息。

[0007] 可选地,上述基于上述交互历史信息和上述至少一个意图描述词,确定上述交互响应信息,包括:确定与上述交互意图关联的交互候选信息库,上述交互候选信息库包括至少一个输入候选信息和至少一个响应候选信息;基于上述交互历史信息和上述至少一个意图描述词,在上述交互候选信息库中确定与上述交互历史信息匹配度最高的目标输入候选信息;将与上述目标输入候选信息关联的响应候选信息,作为上述交互响应信息。

[0008] 可选地,上述基于上述交互历史信息和上述至少一个意图描述词,在上述交互候选信息库中确定与上述交互历史信息匹配度最高的目标输入候选信息,包括:针对任一输入候选信息,确定上述交互历史信息与上述输入候选信息间的第一匹配度;确定上述至少一个意图描述词与上述输入候选信息间的第二匹配度;基于上述第一匹配度和上述第二匹配度,确定针对上述输入候选信息的第三匹配度;重复以上步骤,直至确定出针对各上述输入候选信息的第三匹配度;至少基于针对各上述输入候选信息的第三匹配度,确定上述目

标输入候选信息。

[0009] 可选地,上述至少基于针对各上述输入候选信息的第三匹配度,确定上述目标输入候选信息,包括:针对任一上述输入候选信息,计算上述至少一个意图描述词的平均词向量与上述输入候选信息的平均词向量间的相似度,以作为针对上述输入候选信息的第四匹配度;计算上述输入候选信息中包含的意图描述词个数,以作为针对上述输入候选信息的第五匹配度;计算上述输入候选信息中的所有组成词与上述至少一个意图描述词间的平均相似度,以作为针对上述输入候选信息的第六匹配度;基于上述第三匹配度、第四匹配度、第五匹配度和第六匹配度,确定上述目标输入候选信息。

[0010] 可选地,上述基于上述交互历史信息 and 上述至少一个意图描述词,确定上述交互响应信息,包括:利用预设的交互信息生成模型,进行基于上述交互历史信息的编码操作,得到中间特征向量;以及进行基于上述中间特征向量和上述至少一个意图描述词的解码操作,以生成上述交互响应信息。

[0011] 可选地,在上述交互信息生成模型的训练过程中,包括:利用上述交互信息生成模型,进行基于交互样本信息的编码操作,得到样本中间特征向量,上述交互样本信息包括样本输入信息和样本响应信息,上述交互样本信息具有指示交互意图的样本标签;进行基于上述样本中间特征向量的解码操作,得到预测响应信息;基于上述预测响应信息和上述样本响应信息,计算上述交互信息生成模型的生成损失;基于上述样本标签和上述预测响应信息指示的交互意图,计算上述交互信息生成模型的分​​类损失;基于上述生成损失和上述分类损失,确定上述交互信息生成模型的总损失,以进行基于上述总损失的模型训练。

[0012] 本公开的另一个方面提供了一种交互处理装置,包括:获取模块,用于获取交互历史信息,上述交互历史信息包括输入历史信息和/或响应历史信息;第一处理模块,用于利用预设的意图识别模型,确定上述交互历史信息指示的交互对象的交互意图;第二处理模块,用于基于上述交互历史信息和上述交互意图,确定用于向上述交互对象反馈的交互响应信息并输出。

[0013] 可选地,上述第二处理模块包括:第一处理子模块,用于根据交互意图与意图描述词之间的预设关联关系,确定与上述交互历史信息关联的至少一个意图描述词,其中,上述关联关系是利用意图识别模型基于交互样本数据训练得到的;第二处理子模块,用于基于上述交互历史信息和上述至少一个意图描述词,确定上述交互响应信息。

[0014] 可选地,上述第二处理子模块包括:第一处理单元,用于确定与上述交互意图关联的交互候选信息库,上述交互候选信息库包括至少一个输入候选信息和至少一个响应候选信息;第二处理单元,用于基于上述交互历史信息和上述至少一个意图描述词,在上述交互候选信息库中确定与上述交互历史信息匹配度最高的目标输入候选信息;第三处理单元,用于将与上述目标输入候选信息关联的响应候选信息,作为上述交互响应信息。

[0015] 可选地,上述第二处理单元包括:第一处理子单元,用于针对任一输入候选信息,确定上述交互历史信息与上述输入候选信息间的第一匹配度;第二处理子单元,用于确定上述至少一个意图描述词与上述输入候选信息间的第二匹配度;第三处理子单元,用于基于上述第一匹配度和上述第二匹配度,确定针对上述输入候选信息的第三匹配度;重复以上步骤,直至确定出针对各上述输入候选信息的第三匹配度;第四处理子单元,用于至少基于针对各上述输入候选信息的第三匹配度,确定上述目标输入候选信息。

[0016] 可选地,上述第四处理子单元具体用于:针对任一上述输入候选信息,计算上述至少一个意图描述词的平均词向量与上述输入候选信息的平均词向量间的相似度,以作为针对上述输入候选信息的第四匹配度;计算上述输入候选信息中包含的意图描述词个数,以作为针对上述输入候选信息的第五匹配度;计算上述输入候选信息中的所有组成词与上述至少一个意图描述词间的平均相似度,以作为针对上述输入候选信息的第六匹配度;基于上述第三匹配度、第四匹配度、第五匹配度和第六匹配度,确定上述目标输入候选信息。

[0017] 可选地,上述第二处理子模块包括:第四处理单元,用于利用预设的交互信息生成模型,进行基于上述交互历史信息的编码操作,得到中间特征向量;第五处理单元,用于以及进行基于上述中间特征向量和上述至少一个意图描述词的解码操作,以生成上述交互响应信息。

[0018] 可选地,在上述交互信息生成模型的训练过程中,包括:利用上述交互信息生成模型,进行基于交互样本信息的编码操作,得到样本中间特征向量,上述交互样本信息包括样本输入信息和样本响应信息,上述交互样本信息具有指示交互意图的样本标签;进行基于上述样本中间特征向量的解码操作,得到预测响应信息;基于上述预测响应信息和上述样本响应信息,计算上述交互信息生成模型的生成损失;基于上述样本标签和上述预测响应信息指示的交互意图,计算上述交互信息生成模型的分损失;基于上述生成损失和上述分损失,确定上述交互信息生成模型的总损失,以进行基于上述总损失的模型训练。

[0019] 本公开的另一个方面提供了一种电子设备。上述电子设备包括至少一个处理器,以及与上述至少一个处理器通信连接的存储器。其中,上述存储器存储有可被上述至少一个处理器执行的指令,上述指令被上述至少一个处理器执行,以使上述至少一个处理器实现本公开实施例的方法。

[0020] 本公开的另一方面提供了一种计算机可读存储介质,存储有计算机可执行指令,上述指令在被执行时用于实现本公开实施例的方法。

[0021] 本公开的另一方面提供了一种计算机程序,上述计算机程序包括计算机可执行指令,上述指令在被执行时用于实现本公开实施例的方法。

[0022] 通过本公开实施例,因为采用了获取交互历史信息,交互历史信息包括输入历史信息和/或响应历史信息;利用预设的意图识别模型,确定交互历史信息指示的交互对象的交互意图;基于交互历史信息和交互意图,确定用于向交互对象反馈的交互响应信息并输出的技术方案,所以至少部分地克服了相关技术中存在的交互匹配性差、交互效率不佳的技术问题,进而达到了有效提高交互匹配性和交互效率的技术效果。

附图说明

[0023] 为了更完整地理解本公开及其优势,现在将参考结合附图的以下描述,其中:

[0024] 图1示意性示出了根据本公开实施例的交互处理系统架构;

[0025] 图2示意性示出了根据本公开实施例的一种交互处理方法的流程图;

[0026] 图3示意性示出了根据本公开实施例的另一交互处理方法的流程图;

[0027] 图4示意性示出了根据本公开实施例的确定针对某一候选输入信息的第三匹配度方法的示意图;

[0028] 图5示意性示出了根据本公开实施例的再一交互处理方法的流程图;

[0029] 图6示意性示出了根据本公开实施例的交互处理装置的框图；

[0030] 图7示意性示出了根据本公开实施例的适于实现交互处理方法和装置的电子设备的框图。

具体实施方式

[0031] 以下，将参照附图来描述本公开的实施例。但是应该理解，这些描述只是示例性地，而并非要限制本公开的范围。在下面的详细描述中，为便于解释，阐述了许多具体的细节以提供对本公开实施例的全面理解。然而，明显地，一个或多个实施例在没有这些具体细节的情况下也可以被实施。此外，在以下说明中，省略了对公知结构和技术的描述，以避免不必要地混淆本公开的概念。

[0032] 在此使用的术语仅仅是为了描述具体实施例，而并非意在限制本公开。在此使用的术语“包括”、“包含”等表明了所述特征、步骤、操作和/或部件的存在，但是并不排除存在或添加一个或多个其他特征、操作步骤、操作或部件。

[0033] 在此使用的所有术语(包括技术和科学术语)具有本领域技术人员通常所理解的含义，除非另外定义。应注意，这里使用的术语应解释为具有与本说明书的上下文相一致的含义，而不应以理想化或过于刻板的方式来解释。

[0034] 在使用类似于“A、B和C等中至少一个”这样的表述的情况下，一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释(例如，“具有A、B和C中至少一个的系统”应包括但不限于单独具有A、单独具有B、单独具有C、具有A和B、具有A和C、具有B和C、和/或具有A、B、C的系统等)。

[0035] 本公开的各个实施例提供了一种交互处理方法以及能够应用该方法的交互处理装置。其中，该方法可以包括获取交互历史信息，交互历史信息包括输入历史信息和/或响应历史信息，和利用预设的意图识别模型，确定交互历史信息指示的交互对象的交互意图，然后基于交互历史信息和交互意图，确定用于向交互对象反馈的交互响应信息并输出。

[0036] 如图1所示，该系统架构100包括至少一个终端(图中示出了多个，如终端101、102、103)和服务器104。在该系统架构100中，服务器104获取与交互对象进行交互产生的交互历史信息，交互历史信息包括输入历史信息和/或响应历史信息，然后利用预设的意图识别模型，确定交互历史信息指示的交互对象的交互意图，最后基于交互历史信息和交互意图，确定用于向交互对象反馈的交互响应信息，和向终端(如终端101、102、103)发送交互响应信息。

[0037] 以下将结合附图和具体实施例详细阐述本公开。

[0038] 图2示意性示出了根据本公开实施例的一种交互处理方法的流程图。

[0039] 如图2所示，该方法例如可以包括操作S210～S230。

[0040] 在操作S210，获取交互历史信息，交互历史信息包括输入历史信息和/或响应历史信息。

[0041] 在本公开实施例中，具体地，用户通过终端与业务平台交互产生交互数据，例如，用户通过终端与业务平台提供的智能助理交互产生交互数据。终端例如可以包括智能手机、平板电脑、便携式计算机、台式计算机、智能手表、智能音箱、智能家居等。业务平台例如可以包括购物平台、外卖平台、打车平台、机器对话平台等。业务平台提供的智能助理例如

可以包括虚拟客服、智能助手、对话机器人等。

[0042] 进行交互处理的服务器获取用户与业务平台间的交互历史信息,交互历史信息包括来自用户的输入历史信息,以及来自业务平台的响应历史信息。前述服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或分布式系统。交互历史信息的形式可以是文本形式信息,也可以是语音形式信息,还可以是其他交互形式的信息。

[0043] 在获取用户与业务平台间的交互历史信息时,可以按照时间戳顺序,以用户的待响应的交互输入信息为起点,获取预设时间段内的交互历史信息,或者,还可以获取预设数量的交互历史信息。交互历史信息既包括来自用户的输入历史信息,也包括来自业务平台的响应历史信息,基于交互历史信息确定用于向用户反馈的交互响应信息,有利于改善确定出的交互响应信息的匹配度,有利于提升智能交互体验和提高智能交互效率。

[0044] 可选地,在基于交互历史信息确定用于向用户反馈的交互响应信息前,可以对获取的交互历史信息进行去重、筛除、文本替换、文本转换等形式的预处理,以实现将非标准信息转换为标准信息。示例性地,筛除会话文本信息中不包括文本内容的信息,例如筛除只包括标点符号或表情包的会话信息。通过对获取的交互历史信息进行预处理,一方面有利于减少后续数据处理的计算量,另一方面有利于提升数据处理结果的准确度。

[0045] 接下来,在操作S220,利用预设的意图识别模型,确定交互历史信息指示的交互对象的交互意图。

[0046] 在本公开实施例中,具体地,将获取的交互历史信息输入训练好的意图识别模型中,利用意图识别模型对交互历史信息进行识别处理,得到交互历史信息指示的交互对象的交互意图,即得到交互历史信息指示的用户的交互意图。针对不同应用场景和实际需要,交互意图具有细粒度划分。示例性地,在电商应用中,对于支付场景,交互意图可能包括付款方式确认、优惠信息咨询、邮费查询等;对于售后场景,交互意图可能包括物流信息查询、退换货咨询、商品问题反馈等。

[0047] 在确定交互历史信息指示的交互对象的交互意图时,利用意图识别模型对各交互历史信息进行编码,得到与各交互历史信息关联的特征向量。编码方法例如可以包括基于Transformer模型(翻译模型)进行编码,或者基于Bi-GRU(Bidirectional Gated Recurrent Unit,双向门循环单元)模型进行编码。

[0048] 针对大量预设的细粒度划分的交互意图,各交互意图具有对应的意图标签。在确定交互历史信息指示的交互意图时,对预设的交互意图的意图标签进行向量化,得到不同意图标签的特征向量。然后,确定各交互历史信息特征向量与意图标签特征向量间的相似度,并将相似度高于预设阈值的预设交互意图,确定为交互历史信息指示的交互意图。其中,相似度可以包括余弦相似度、欧氏距离倒数等,本实施例对相似度的具体类型不做限定。

[0049] 意图识别模型是利用大量具有分类结果的交互样本数据训练得到的,交互样本数据的分类结果指示与交互样本数据关联的交互意图,其中,意图识别模型例如可以是HAN(Hierarchical Attention Network,分层注意力网络)模型。

[0050] 在利用大量的具有分类结果的交互样本数据进行模型训练时,针对任一交互意图,利用意图识别模型标注每条交互样本数据中注意力分数最高的n个词汇,然后对各交互样本数据中注意力分数最高的n个词汇进行汇总,并把其中出现频次最高的预设个数的词

汇,确定为与该交互意图关联的意图描述词。出现频次最高的预设个数的词汇构成与该交互意图关联的意图描述词库。

[0051] 示例性地,针对“快递信息咨询”的交互意图,与该交互意图关联的交互样本数据共有20万条。利用意图识别模型标注每条交互样本数据中注意力分数最高的5个词汇,然后对20万条交互样本数据中的各交互样本数据的注意力分数最高的5个词汇进行汇总,并把其中出现频次最高的100个词汇,确定为与该交互意图关联的意图描述词。出现频次最高的100个词汇构成与“快递信息咨询”这一交互意图关联的意图描述词库。

[0052] 接下来,在操作S230,基于交互历史信息 and 交互意图,确定用于向交互对象反馈的交互响应信息并输出。

[0053] 在本公开实施例中,具体地,首先根据交互意图与意图描述词之间的预设关联关系,确定与交互历史信息关联的至少一个意图描述词,其中,关联关系是利用意图识别模型基于交互样本数据训练得到的,然后,基于交互历史信息 and 至少一个意图描述词,确定交互响应信息。

[0054] 在利用意图识别模型确定出交互历史信息指示的交互意图后,根据与该交互意图关联的意图描述词库,将其中预设个数的意图描述词确定为与交互历史信息关联的意图描述词。示例性地,在与交互意图关联的意图描述词库中,确定其中50个意图描述词为与交互历史信息关联的意图描述词。与交互意图关联的意图描述词库是在意图识别模型训练过程中得到的,具体已在操作S220中阐述,在此不做赘述。

[0055] 基于交互历史信息 and 交互历史信息指示的交互意图,具体来说,基于交互历史信息和与交互历史信息关联的至少一个意图描述词,确定用于向交互对象反馈的交互响应信息并输出。这有利于保证输出的交互响应信息与用户的交互意图相匹配,这不仅有利于改善自助交互效果,还有利于提升自助交互效率。

[0056] 在本公开实施例中,通过获取交互历史信息,其中,交互历史信息包括输入历史信息和/或响应历史信息,和利用预设的意图识别模型,确定交互历史信息指示的交互对象的交互意图,然后基于交互历史信息 and 交互意图,确定用于向交互对象反馈的交互响应信息并输出。基于获取的交互历史信息 and 交互历史信息指示的交互意图,确定用于向交互对象反馈的交互响应信息,这有利于提升输出的交互响应信息在交互意图层面的匹配程度;通过精准匹配交互输入信息,和提供准确度高的交互响应信息,不仅能够有效提升智能交互效率,还有利于改善智能交互效果。

[0057] 图3示意性示出了根据本公开实施例的另一交互处理方法的流程图。

[0058] 如图3所示,操作S230具体可以包括操作S310~S330。

[0059] 在操作S310,确定与交互意图关联的交互候选信息库,交互候选信息库包括至少一个输入候选信息和至少一个响应候选信息。

[0060] 在本公开实施例中,具体地,在确定出交互历史信息指示的交互意图后,确定与交互意图关联的交互候选信息库,交互候选信息库包括至少一个输入候选信息和至少一个响应候选信息。交互候选信息库例如可以是标准问答库,标准问答库中包括至少一个标准问题文本和至少一个标准答案文本,标准问题文本例如可以包括“请问xx快递的运费是多少”。

[0061] 在上述过程中,针对每一交互意图,可以以键值的方式存储与各交互意图关联的

交互候选信息。具体地,以交互意图的意图标签为键名,以交互候选信息为键值存储与各交互意图关联的交互候选信息。与交互意图关联的多个交互候选信息,构成与交互意图关联的交互候选信息库。

[0062] 接下来,在操作S320,基于交互历史信息 and 至少一个意图描述词,在交互候选信息库中确定与交互历史信息匹配度最高的目标输入候选信息。

[0063] 在本公开实施例中,具体地,针对任一输入候选信息,确定交互历史信息与输入候选信息间的第一匹配度;确定至少一个意图描述词与输入候选信息间的第二匹配度;基于第一匹配度和第二匹配度,确定针对输入候选信息的第三匹配度;重复以上步骤,直至确定出针对各输入候选信息的第三匹配度;至少基于针对各输入候选信息的第三匹配度,确定目标输入候选信息。

[0064] 交互历史信息包括至少一个输入历史信息和/或至少一个响应历史信息。针对交互候选信息库中的任一输入候选信息,在确定交互历史信息与该输入候选信息间的第一匹配度时,确定每一条交互历史信息与该输入候选信息间的第一匹配度。示例性地,交互历史信息包括来自用户的A1、A2、A3共3个历史问题文本,和来自虚拟客服的Q1、Q2共2个历史问题文本,对于输入候选信息“请问xx快递的运费是多少”,分别确定A1、A2、A3、Q1、Q2与输入候选信息间的第一匹配度。确定第一匹配度的方法,例如可以包括确定交互历史信息的特征向量与输入候选信息的特征向量间的相似度。

[0065] 此外,还需确定与交互历史信息关联的至少一个意图描述词,与交互候选信息库中的各输入候选信息间的第二匹配度。针对任一输入候选信息,可以将至少一个意图描述词作为额外的交互输入信息,确定由至少一个意图描述词构成的交互输入信息的特征向量。然后,确定此“额外的”交互输入信息的特征向量与输入候选信息的特征向量间的相似度,以得到至少一个意图描述词与输入候选信息间的第二匹配度。

[0066] 针对任一输入候选信息,在确定出各交互历史信息与该输入候选信息间的第一匹配度,和确定出至少一个意图描述词与该输入候选信息间的第二匹配度后,利用匹配矩阵对各第一匹配度和第二匹配度进行聚合,得到针对该输入候选信息的第三匹配度。重复上述操作,确定针对各输入候选信息的第三匹配度,然后基于针对各输入候选信息的第三匹配度,将其中第三匹配度最高的输入候选信息确定为目标输入候选信息。

[0067] 图4示意性示出了根据本公开实施例的确定针对某一候选输入信息的第三匹配度方法的示意图,如图4所示,获取的交互历史信息包括A1、Q1、A2、Q2、A3。针对任一输入候选信息,分别确定A1、Q1、A2、Q2、A3与该输入候选信息间的第一匹配度 p_1 、 p_2 、 p_3 、 p_4 、 p_5 。同时,将至少一个意图描述词作为额外的交互输入信息,基于与该交互输入信息关联的特征向量,确定至少一个描述词与输入候选信息间的第二匹配度 p_6 。然后,基于第一匹配度 p_1 、 p_2 、 p_3 、 p_4 、 p_5 和第二匹配度 p_6 ,利用匹配矩阵确定针对该候选输入信息的第三匹配度。在确定出针对各候选输入信息的第三匹配度后,将第三匹配度最高的候选输入信息确定为目标候选输入信息。

[0068] 除基于针对各输入候选信息的第三匹配度确定目标输入候选信息外,还可通过集成方法集成多种匹配度,然后基于集成后匹配度确定目标输入候选信息。具体地,针对任一输入候选信息,计算至少一个意图描述词的平均词向量与输入候选信息的平均词向量间的相似度,以作为针对输入候选信息的第四匹配度;计算输入候选信息中包含的意图描述词

个数,以作为针对输入候选信息的第五匹配度;计算输入候选信息中的所有组成词与至少一个意图描述词间的平均相似度,以作为针对输入候选信息的第六匹配度;基于第三匹配度、第四匹配度、第五匹配度和第六匹配度,确定目标输入候选信息。

[0069] 输入候选信息中所有组成词与至少一个意图描述词间的平均相似度,为所有组成词的平均相似度,具体是对所有组成词的相似度求均数得到。在计算某个组成词的相似度时,计算该组成词与各意图描述词间的相似度,并把其中的最大相似度确定为该组成词的相似度。在得到第四、五、六匹配度后,可以利用Xgboost等集成方法对包含第三匹配度在内的各匹配度进行集成,得到针对各输入候选信息的集成匹配度,然后将集成匹配度最高的输入候选信息确定为目标输入候选信息。

[0070] 接下来,在操作S330,将与目标输入候选信息关联的响应候选信息,作为交互响应信息并输出。

[0071] 在本公开实施例中,具体地,在确定出与历史交互信息和至少一个意图关键词匹配度最高的目标输入候选信息,将与目标输入候选信息关联的响应候选信息,作为用于向交互对象反馈的交互响应信息并输出。

[0072] 通过结合确定至少一个意图描述词与各输入候选信息间的第二匹配度,实现在交互候选信息库中检索出匹配度最高的目标输入候选信息。与目标输入候选信息关联的响应候选信息自然与交互历史信息及交互历史信息指示的交互意图间的匹配度最高,这有效提高了输出的交互响应信息与交互对象的交互意图的匹配准确率,有利于提高与交互对象间的智能交互效率和改善智能交互效果。

[0073] 图5示意性示出了根据本公开实施例的再一交互处理方法的流程图。

[0074] 如图5所示,操作S230具体可以包括操作S510~S520。

[0075] 在操作S510,利用预设的交互信息生成模型,进行基于交互历史信息的编码操作,得到中间特征向量。

[0076] 在本公开实施例中,具体地,预设的交互信息生成模型例如可以是sequence-to-sequence的序列模型,利用交互信息生成模型进行基于交互历史信息的编码操作,具体可以包括:首先遍历交互历史信息中的每一个字符,当交互历史信息中包括多个输入历史信息 and 多个响应历史信息时,遍历每一输入历史信息和每一响应历史信息种的字符。将交互历史信息中的各字符依次作为对应的编码时刻的输入。此外,每一编码时刻的输入还包括相邻的上一编码时刻输出的隐状态。同样,本编码时刻也会输出一个隐状态,本编码时刻输出的隐状态同样也会作为相邻的下一编码时刻的输入。对于seq2seq模型来说,通常保留最后一个编码时刻输出的隐状态,以作为基于交互历史信息的编码操作所得到的中间特征向量,中间特征向量可被认为编码了交互历史信息的语义信息。

[0077] 接下来,在操作S520,进行基于中间特征向量和至少一个意图描述词的解码操作,以生成交互响应信息并输出。

[0078] 在本公开实施例中,具体地,将中间特征向量作为第一个解码时刻的初始隐状态,即将最后一个编码时刻输出的隐状态作为第一个解码时刻的初始隐状态。利用神经网络模型的每一解码时刻输出字符和隐状态,同时将上一解码时刻输出的字符和隐状态作为输入,输入至相邻的下一解码时刻。此外,为提高生成的交互响应信息与交互对象的交互意图间的匹配度,确定至少一个意图描述词的平均词向量,并将平均词向量作为至少一个意图

描述词的表示向量,输入每一解码时刻。重复前述操作,直至解码输出EOS token,各个解码时刻输出的字符构成交互响应信息。

[0079] 可选地,可以引入注意力机制解决输入语句较长编码效果较差的问题,具体地,在进行t时刻解码计算时,除将上一解码时刻输出的字符和隐状态,以及将至少一个意图描述词的表示向量作为输入外,还可以将编码计算中所有时刻的输入加权得到上下文向量,然后将上下文向量同时作为每个解码时刻的输入,以此生成交互意图匹配良好的交互响应信息。

[0080] 交互信息生成模型为预先训练好的模型,在交互信息生成模型的训练过程中,可以利用交互信息生成模型,进行基于交互样本信息的编码操作,得到样本中间特征向量,交互样本信息包括样本输入信息和样本响应信息,交互样本信息具有指示交互意图的样本标签;进行基于样本中间特征向量的解码操作,得到预测响应信息;基于预测响应信息和样本响应信息,计算交互信息生成模型的生成损失;基于样本标签和预测响应信息指示的交互意图,计算交互信息生成模型的分损失;基于生成损失和分损失,确定交互信息生成模型的总损失,以进行基于总损失的模型训练。

[0081] 具体地,在利用交互信息生成模型生成交互响应信息时, $p(y_t)$ 表示t时刻生成某字的概率, $Emb(y_t)$ 表示该字的词向量, $Ewe(t;C,z)$ 表示t时刻生成字的期待向量值:

$$[0082] \quad Ewe(t;C,z)=\sum_{y_t \in V} p(y_t) Emb(y_t)$$

[0083] 其中, $p(y_t)$ 为生成词典中的某个字的概率,其可由交互信息生成模型的解码器训练得出,V表示预设的常用词典。根据样本响应信息和各时刻生成字的期待向量值 $Ewe(t;C,z)$,可确定出交互信息生成模型的生成损失 L_{CE} 。

[0084] 将所有时刻生成字的期待向量值累加平均,然后使用softmax激活函数得到 $q(z|Y)$, $q(z|Y)$ 表示生成语句的意图分别概率:

$$[0085] \quad q(z|Y)=f(W \cdot \frac{1}{T} \sum_{t=1}^T Ewe(t;C,z))$$

[0086] $p(z)$ 表示期待生成语句的意图(可由交互样本信息的样本标签确定得到), $q(z|Y)$ 表示实际生成语句的意图分别概率,交互信息生成模型的分损失可表示为:

$$[0087] \quad L_{CLS}=-p(z) \log q(z|Y)$$

[0088] 基于生成损失 L_{CE} 和分损失 L_{CLS} ,可确定出交互信息生成模型的总损失 $L=L_{CE}+\mu L_{CLS}$, μ 为可定义变量,生成损失 L_{CE} 和分损失 L_{CLS} 将共同作用于交互响应信息的生成。

[0089] 图6示意性示出了根据本公开实施例的交互处理装置的框图。

[0090] 如图6所示,该装置可以包括获取模块601、第一处理模块602和第二处理模块603。

[0091] 具体地,获取模块601,用于获取交互历史信息,所述交互历史信息包括输入历史信息和/或响应历史信息;第一处理模块602,用于利用预设的意图识别模型,确定所述交互历史信息指示的交互对象的交互意图;第二处理模块603,用于基于所述交互历史信息和所述交互意图,确定用于向所述交互对象反馈的交互响应信息并输出。

[0092] 在本公开实施例中,通过获取交互历史信息,其中,交互历史信息包括输入历史信息和/或响应历史信息,和利用预设的意图识别模型,确定交互历史信息指示的交互对象的交互意图,然后基于交互历史信息和交互意图,确定用于向交互对象反馈的交互响应信息并输出。基于获取的交互历史信息和交互历史信息指示的交互意图,确定用于向交互对象

反馈的交互响应信息,这有利于提升输出的交互响应信息在交互意图层面的匹配程度;通过精准匹配交互输入信息,和提供准确度高的交互响应信息,不仅能够有效提升智能交互效率,还有利于改善智能交互效果。

[0093] 作为一种可选的实施例,第二处理模块包括:第一处理子模块,用于根据交互意图与意图描述词之间的预设关联关系,确定与交互历史信息关联的至少一个意图描述词,其中,关联关系是利用意图识别模型基于交互样本数据训练得到的;第二处理子模块,用于基于交互历史信息和至少一个意图描述词,确定交互响应信息。

[0094] 作为一种可选的实施例,第二处理子模块包括:第一处理单元,用于确定与交互意图关联的交互候选信息库,交互候选信息库包括至少一个输入候选信息和至少一个响应候选信息;第二处理单元,用于基于交互历史信息和至少一个意图描述词,在交互候选信息库中确定与交互历史信息匹配度最高的目标输入候选信息;第三处理单元,用于将与目标输入候选信息关联的响应候选信息,作为交互响应信息。

[0095] 作为一种可选的实施例,第二处理单元包括:第一处理子单元,用于针对任一输入候选信息,确定交互历史信息与输入候选信息间的第一匹配度;第二处理子单元,用于确定至少一个意图描述词与输入候选信息间的第二匹配度;第三处理子单元,用于基于第一匹配度和第二匹配度,确定针对输入候选信息的第三匹配度;重复以上步骤,直至确定出针对各输入候选信息的第三匹配度;第四处理子单元,用于至少基于针对各输入候选信息的第三匹配度,确定目标输入候选信息。

[0096] 作为一种可选的实施例,第四处理子单元具体用于:针对任一输入候选信息,计算至少一个意图描述词的平均词向量与输入候选信息的平均词向量间的相似度,以作为针对输入候选信息的第四匹配度;计算输入候选信息中包含的意图描述词个数,以作为针对输入候选信息的第五匹配度;计算输入候选信息中的所有组成词与至少一个意图描述词间的平均相似度,以作为针对输入候选信息的第六匹配度;基于第三匹配度、第四匹配度、第五匹配度和第六匹配度,确定目标输入候选信息。

[0097] 作为一种可选的实施例,第二处理子模块包括:第四处理单元,用于利用预设的交互信息生成模型,进行基于交互历史信息的编码操作,得到中间特征向量;第五处理单元,用于以及进行基于中间特征向量和至少一个意图描述词的解码操作,以生成交互响应信息。

[0098] 作为一种可选的实施例,在交互信息生成模型的训练过程中,包括:利用交互信息生成模型,进行基于交互样本信息的编码操作,得到样本中间特征向量,交互样本信息包括样本输入信息和样本响应信息,交互样本信息具有指示交互意图的样本标签;进行基于样本中间特征向量的解码操作,得到预测响应信息;基于预测响应信息和样本响应信息,计算交互信息生成模型的生成损失;基于样本标签和预测响应信息指示的交互意图,计算交互信息生成模型的分损失;基于生成损失和分损失,确定交互信息生成模型的总损失,以进行基于总损失的模型训练。

[0099] 可选地,获取模块601、第一处理模块602和第二处理模块603中的模块、子模块中的任意多个、或其中任意多个的至少部分功能可以在一个模块中实现。根据本公开实施例的模块中的任意一个或多个可以被拆分成多个模块来实现。根据本公开实施例的模块中的任意一个或多个可以至少被部分地实现为硬件电路,例如现场可编程门阵列(FPGA)、可编

程逻辑阵列 (PLA)、片上系统、基板上的系统、封装上的系统、专用集成电路 (ASIC), 或可以通过对电路进行集成或封装的任何其他的合理方式的硬件或固件来实现, 或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者根据本公开实施例的模块中的一个或多个可以至少被部分地实现为计算机程序模块, 当该计算机程序模块被运行时, 可以执行相应的功能。

[0100] 图7示意性示出了根据本公开实施例的适于实现交互处理方法和装置的电子设备的框图。图7示出的计算机系统仅仅是一个示例, 不应对本公开实施例的功能和使用范围带来任何限制。

[0101] 如图7所示, 根据本公开实施例的计算机系统700包括处理器701, 其可以根据存储在只读存储器 (ROM) 702中的程序或者从存储部分708加载到随机访问存储器 (RAM) 703中的程序而执行各种适当的动作和处理。处理器701例如可以包括通用微处理器 (例如CPU)、指令集处理器和/或相关芯片组和/或专用微处理器 (例如, 专用集成电路 (ASIC)), 等等。处理器701还可以包括用于缓存用途的板载存储器。处理器701可以包括用于执行根据本公开实施例的方法流程的不同动作的单一处理单元或者是多个处理单元。

[0102] 在RAM 703中, 存储有系统700操作所需的各种程序和数据。处理器701、ROM 702以及RAM 703通过总线704彼此相连。处理器701通过执行ROM 702和/或RAM 703中的程序来执行根据本公开实施例的方法流程的各种操作。需要注意, 所述程序也可以存储在除ROM 702和RAM 703以外的一个或多个存储器中。处理器701也可以通过执行存储在所述一个或多个存储器中的程序来执行根据本公开实施例的方法流程的各种操作。

[0103] 可选地, 系统700还可以包括输入/输出 (I/O) 接口705, 输入/输出 (I/O) 接口705也连接至总线704。系统700还可以包括连接至I/O接口705的以下部件中的一项或多项: 包括键盘、鼠标等的输入部分706; 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707; 包括硬盘等的存储部分708; 以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分709。通信部分709经由诸如因特网的网络执行通信处理。驱动器710也根据需要连接至I/O接口706。可拆卸介质711, 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710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708。

[0104] 可选地, 根据本公开实施例的方法流程可以被实现为计算机软件程序。例如, 本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存储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709从网络上被下载和安装, 和/或从可拆卸介质711被安装。在该计算机程序被处理器701执行时, 执行本公开实施例的系统中限定的上述功能。可选地, 上文描述的系统、设备、装置、模块、单元等可以通过计算机程序模块来实现。

[0105] 本公开还提供了一种计算机可读存储介质, 该计算机可读存储介质可以是上述实施例中描述的设备/装置/系统中所包含的; 也可以是单独存在, 而未装配入该设备/装置/系统中。上述计算机可读存储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执行时, 实现根据本公开实施例的方法。

[0106] 可选地, 计算机可读存储介质可以是非易失性的计算机可读存储介质, 例如可以包括但不限于: 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)、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。在本公开中,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。例如,可选地,计算机可读存储介质可以包括上文描述的ROM 702和/或RAM 703和/或ROM 702和RAM 703以外的一个或多个存储器。

[0107] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、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。

[0108] 本领域技术人员可以理解,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合,即使这样的组合或结合没有明确记载于本公开中。特别地,在不脱离本公开精神和教导的情况下,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合。所有这些组合和/或结合均落入本公开的范围。

[0109] 以上对本公开的实施例进行了描述。但是,这些实施例仅仅是为了说明的目的,而并非为了限制本公开的范围。尽管在以上分别描述了各实施例,但是这并不意味着各个实施例中的措施不能有利地结合使用。本公开的范围由所附权利要求及其等同物限定。不脱离本公开的范围,本领域技术人员可以做出多种替代和修改,这些替代和修改都应落在本公开的范围之内。

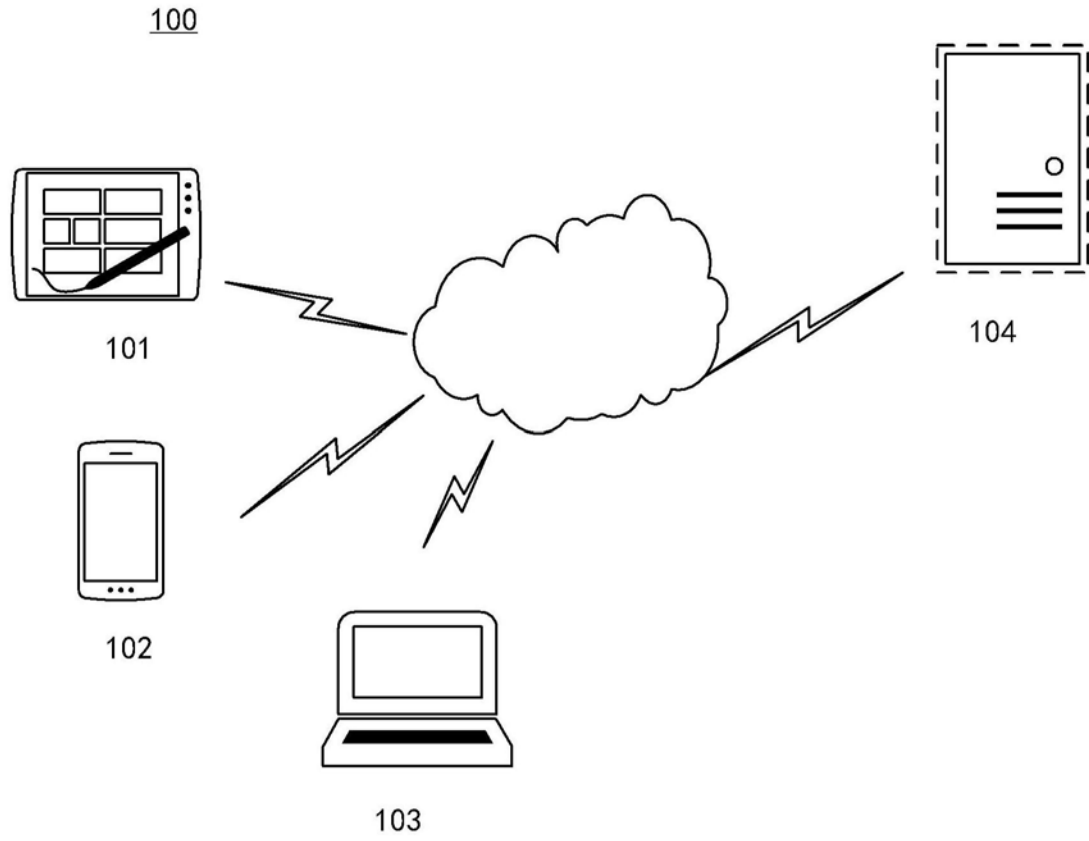


图1

200

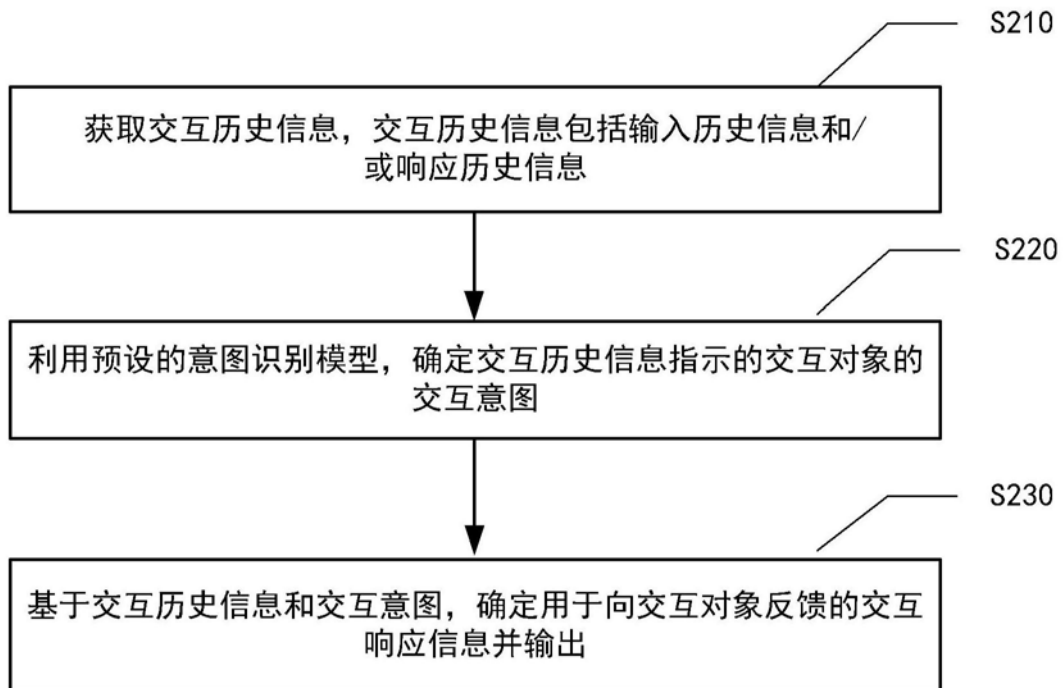


图2

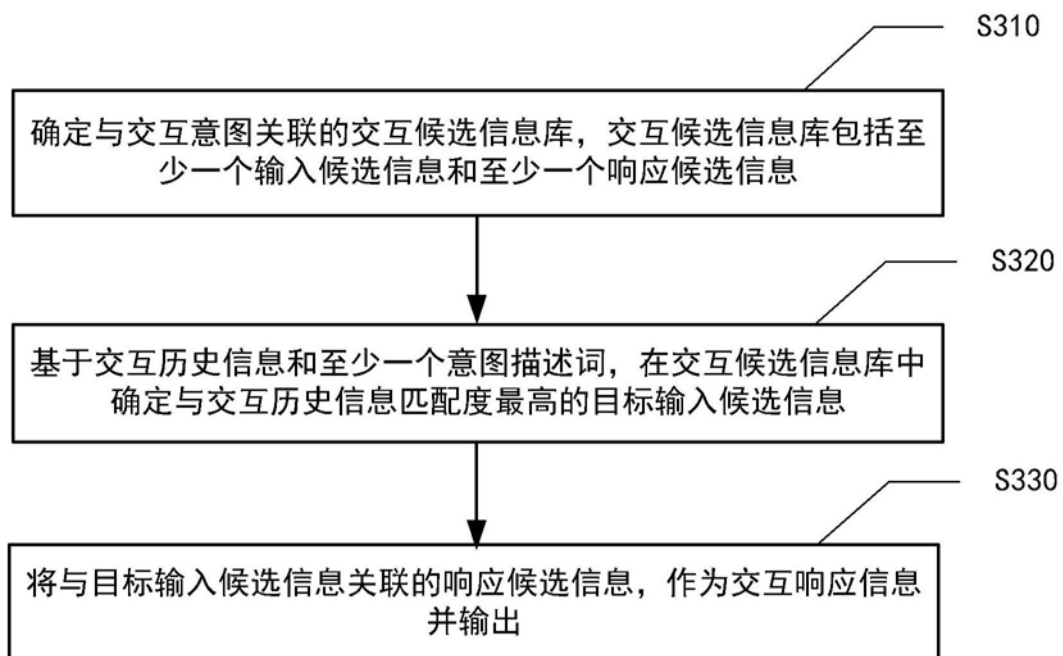


图3

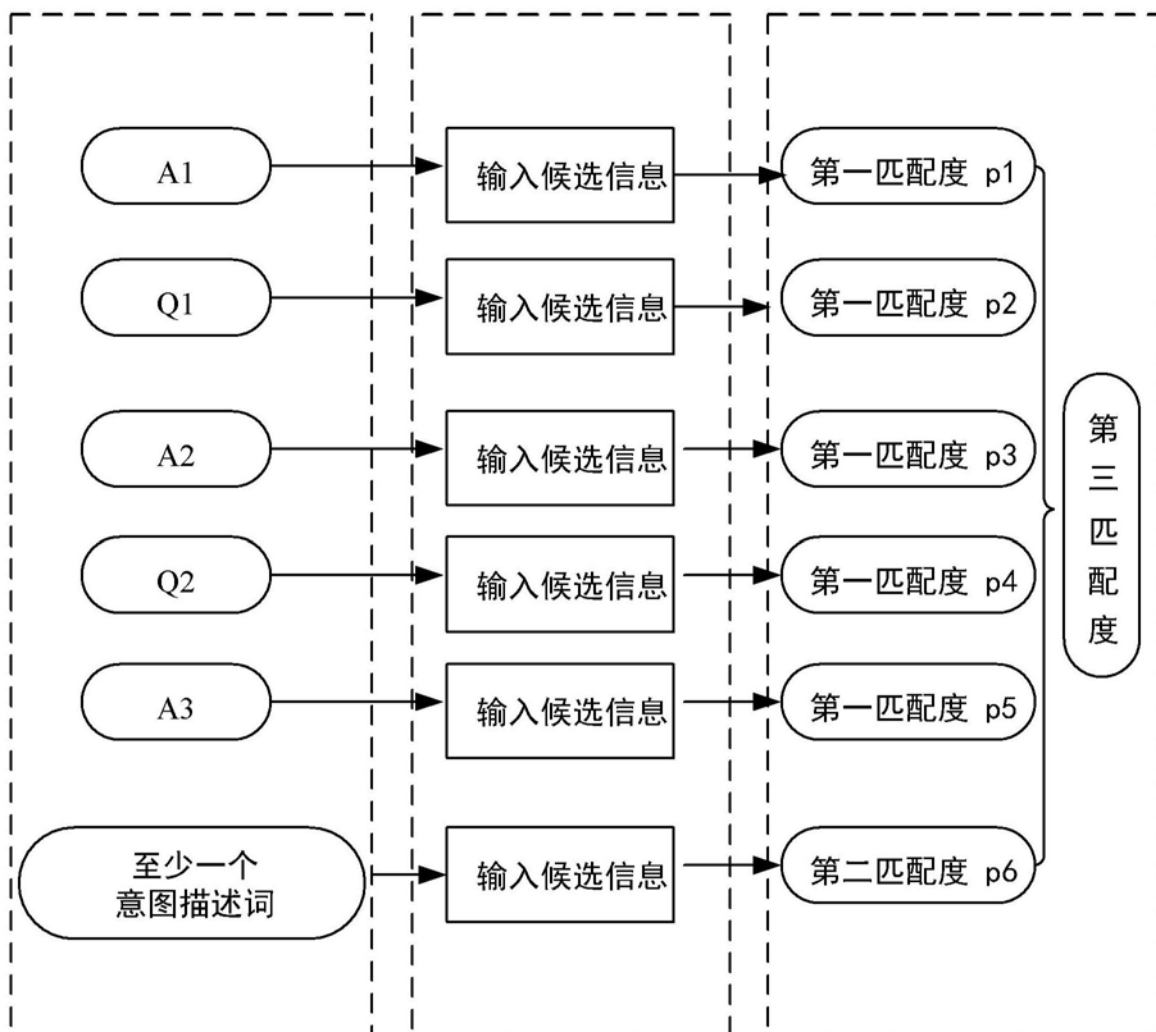


图4

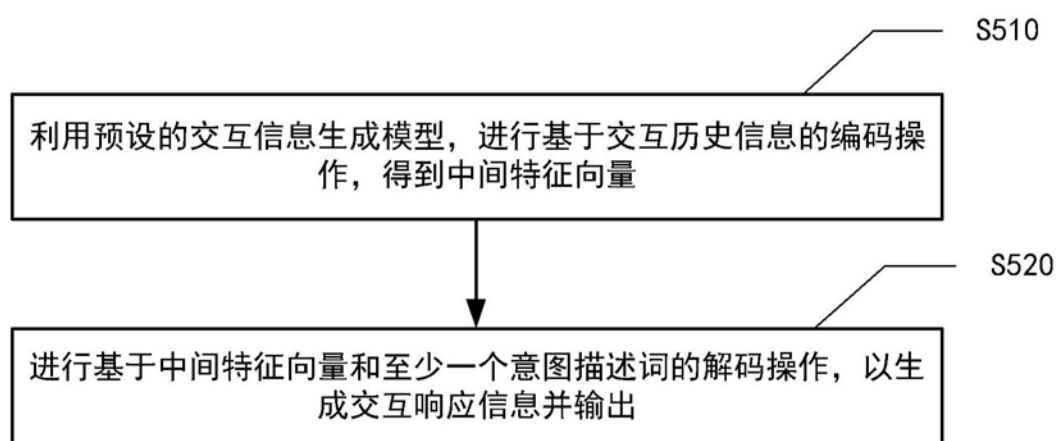


图5

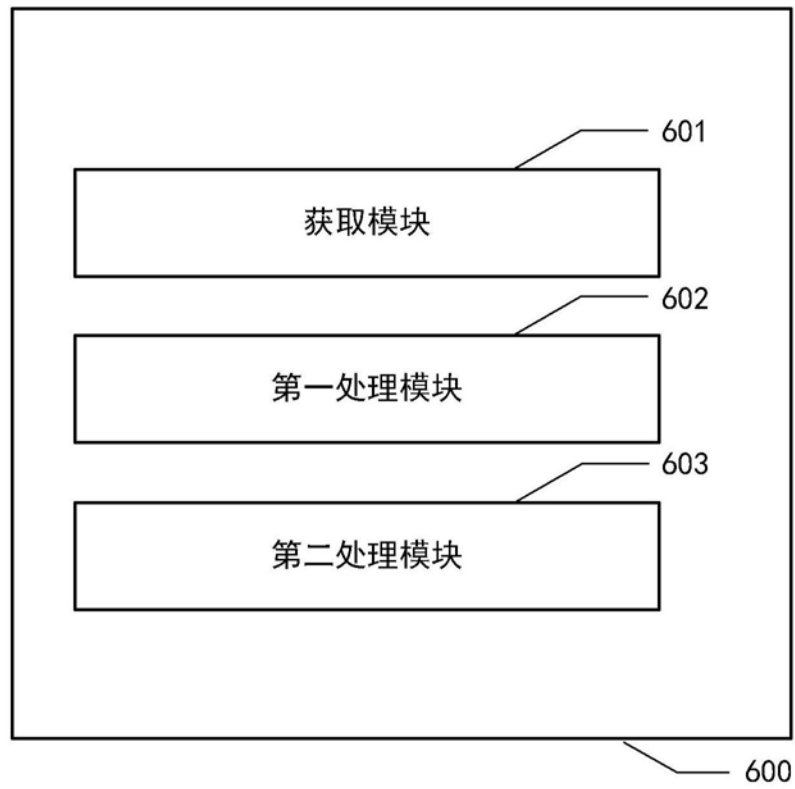


图6

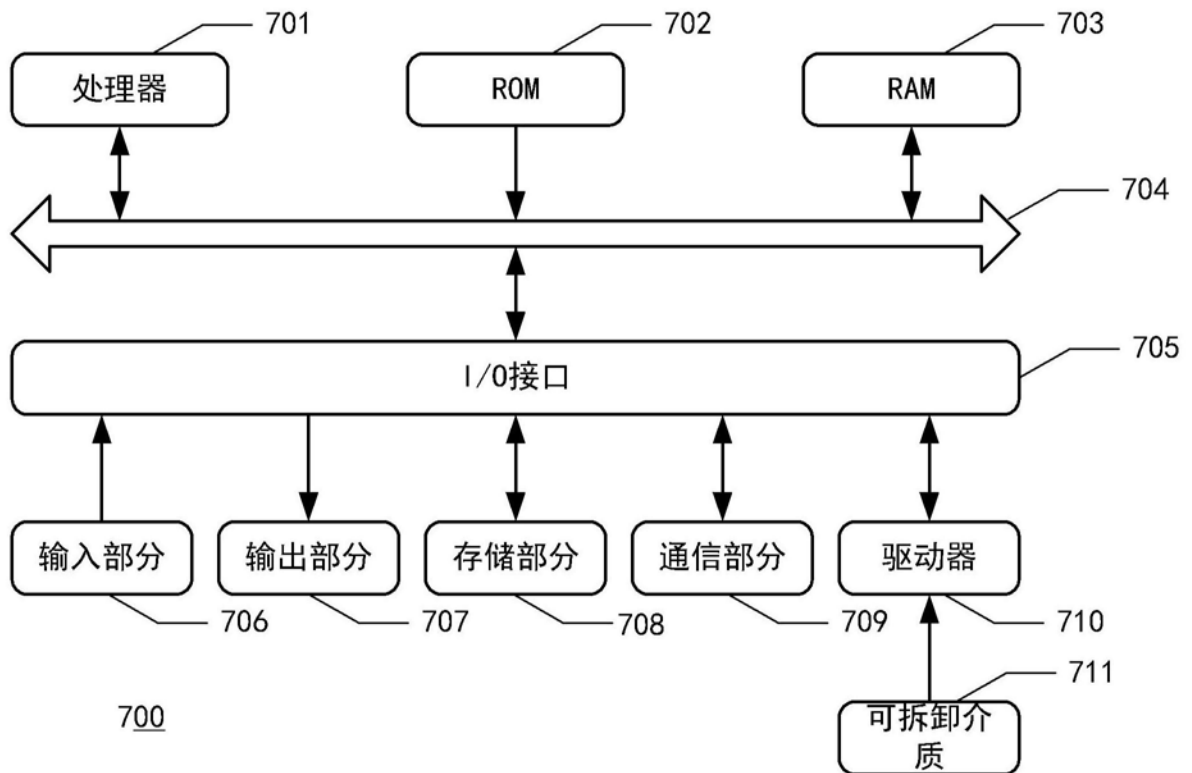


图7