(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117909585 A
(43)申请公布日 2024.04.19

(21)申请号 202410038789.3

(22)申请日 2024.01.10

(71)申请人 京东城市(北京)数字科技有限公司
地址 100086 北京市海淀区知春路76号(写字楼)1号楼9层1-7-5号
申请人 京东科技信息技术有限公司

(72)发明人 孔浩然 申磊 祝天刚 陈蒙

(74)专利代理机构 北京世辉律师事务所 16093
专利代理师 罗利娜

(51)Int.Cl.
G06F 16/9535(2019.01)
G06Q 30/0251(2023.01)
G06Q 30/0241(2023.01)

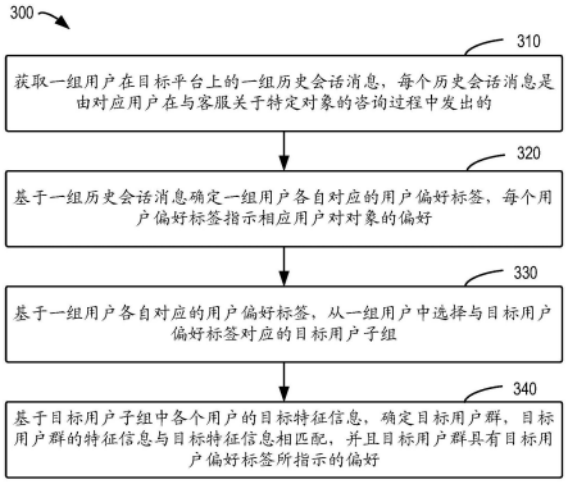
权利要求书2页 说明书12页 附图4页

(54)发明名称

用于信息处理的方法、装置、设备和可读介质

(57)摘要

本公开的实施例提供用于信息处理的方法、装置、设备和可读介质。该方法包括:获取一组用户在目标平台上的一组历史会话消息,每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的;基于一组历史会话消息确定一组用户各自对应的用户偏好标签,每个用户偏好标签指示相应用户对对象的偏好;基于一组用户各自对应的用户偏好标签,从一组用户中选择与目标用户偏好标签对应的目标用户子组;以及基于目标用户子组中各个用户的目标特征信息,确定目标用户群,目标用户群的特征信息与目标特征信息相匹配,并且目标用户群具有目标用户偏好标签所指示的偏好。以此方式,可以提升确定的标签的准确性和提升确定的目标用户群的准确性。



CN 117909585 A

1.一种信息处理方法,包括:

获取一组用户在目标平台上的一组历史会话消息,每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的;

基于所述一组历史会话消息确定所述一组用户各自对应的用户偏好标签,每个用户偏好标签指示相应用户对对象的偏好;

基于所述一组用户各自对应的用户偏好标签,从所述一组用户中选择与目标用户偏好标签对应的目标用户子组;以及

基于所述目标用户子组中各个用户的目标特征信息,确定目标用户群,所述目标用户群的特征信息与所述目标特征信息相匹配,并且所述目标用户群具有所述目标用户偏好标签所指示的偏好。

2.根据权利要求1所述的方法,其中基于所述一组历史会话消息确定所述一组用户各自对应的用户偏好标签包括:

基于所述一组历史会话消息各自所包括的消息内容,确定所述一组历史会话消息各自对应的意图标签,每个意图标签指示在历史会话消息中相应用户的意图;以及

基于意图标签与用户偏好标签之间的映射关系以及所述一组历史会话消息各自对应的意图标签,确定所述一组用户各自对应的用户偏好标签。

3.根据权利要求2所述的方法,其中确定所述一组历史会话消息各自对应的意图标签包括:对于一组历史会话消息中的每个历史会话消息,

确定该历史会话消息所包括的文本是否包括意图词;以及

如果该历史会话消息所包括的文本包括所述意图词,基于意图词与意图标签之间的映射关系,确定该历史会话消息对应的意图标签。

4.根据权利要求2所述的方法,其中确定所述一组历史会话消息各自对应的意图标签包括:

利用机器学习模型处理所述一组历史会话消息各自所包括的消息内容,以确定所述一组历史会话消息各自对应的意图标签。

5.根据权利要求1所述的方法,还包括:

基于所述一组用户各自对应的用户偏好标签,将所述一组用户划分为多个用户子组,每个用户子组中的用户对应相同的用户偏好标签;以及

基于所述多个用户子组各自对应的历史会话消息,确定用于描述每个用户子组对应的用户偏好标签的描述文本。

6.根据权利要求1所述的方法,其中确定目标用户群包括:

利用相似人群扩展算法,基于所述目标用户子组中各个用户的目标特征信息来确定所述目标用户群。

7.根据权利要求6所述的方法,其中利用相似人群扩展算法确定所述目标用户群包括:

利用所述相似人群扩展算法,确定与所述目标特征信息相匹配第一用户群,所述第一用户群包括所述目标用户子组和另外的多个用户;以及

从所述第一用户群中删除所述目标用户子组以确定所述目标用户群。

8.根据权利要求1所述的方法,还包括:

基于所述目标用户偏好标签所指示的偏好,向所述目标用户群提供对象推荐信息。

9.一种用于信息处理的装置,包括:

消息获取模块,被配置为获取一组用户在目标平台上的一组历史会话消息,每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的;

标签确定模块,被配置为基于所述一组历史会话消息确定所述一组用户各自对应的用户偏好标签,每个用户偏好标签指示相应用户对对象的偏好;

用户选择模块,被配置为基于所述一组用户各自对应的用户偏好标签,从所述一组用户中选择与目标用户偏好标签对应的目标用户子组;以及

用户确定模块,被配置为基于所述目标用户子组中各个用户的目标特征信息,确定目标用户群,所述目标用户群的特征信息与所述目标特征信息相匹配,并且所述目标用户群具有所述目标用户偏好标签所指示的偏好。

10.一种电子设备,包括:

至少一个处理单元;以及

至少一个存储器,所述至少一个存储器被耦合到所述至少一个处理单元并且存储用于由所述至少一个处理单元执行的指令,所述指令在由所述至少一个处理单元执行时使所述电子设备执行根据权利要求1至8任一项所述的方法。

11.一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序可由处理器执行以实现根据权利要求1至8任一项所述的方法。

用于信息处理的方法、装置、设备和可读介质

技术领域

[0001] 本公开的示例实施例总体涉及计算机技术领域，并且更具体地，涉及用于信息处理的方法、装置、设备和计算机可读存储介质。

背景技术

[0002] 随着互联网技术的发展，通过互联网获取信息，生活、娱乐和工作成为人们生活的一部分。推广方(例如商家)为了推广特定对象(例如商品)，往往通过互联网投放对象推荐信息(例如广告)。对于推广方，推广方期望将对象推荐信息针对性的推荐给目标用户群以提升推广的准确性。对于用户，用户期望能够针对性的获取对象推荐信息，这样用户可以直接获取符合自身偏好的对象的对象推荐信息，这有助于降低大量推广对用户造成的骚扰问题。因此，期望能够提升确定目标用户群的准确性。

发明内容

[0003] 在本公开的第一方面，提供一种用于信息处理的方法。该方法包括：获取一组用户在目标平台上的一组历史会话消息，每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的；基于一组历史会话消息确定一组用户各自对应的用户偏好标签，每个用户偏好标签指示相应用户对对象的偏好；基于一组用户各自对应的用户偏好标签，从一组用户中选择与目标用户偏好标签对应的目标用户子组；以及基于目标用户子组中各个用户的目标特征信息，确定目标用户群，目标用户群的特征信息与目标特征信息相匹配，并且目标用户群具有目标用户偏好标签所指示的偏好。

[0004] 在本公开的第二方面，提供一种用于信息处理的装置。该装置包括：消息获取模块，被配置为获取一组用户在目标平台上的一组历史会话消息，每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的；标签确定模块，被配置为基于一组历史会话消息确定一组用户各自对应的用户偏好标签，每个用户偏好标签指示相应用户对对象的偏好；用户选择模块，被配置为基于一组用户各自对应的用户偏好标签，从一组用户中选择与目标用户偏好标签对应的目标用户子组；以及用户确定模块，被配置为基于目标用户子组中各个用户的目标特征信息，确定目标用户群，目标用户群的特征信息与目标特征信息相匹配，并且目标用户群具有目标用户偏好标签所指示的偏好。

[0005] 在本公开的第三方面，提供了一种电子设备。该电子设备包括至少一个处理单元；以及至少一个存储器，至少一个存储器被耦合到至少一个处理单元并且存储用于由至少一个处理单元执行的指令，指令在由至少一个处理单元执行时使电子设备执行本公开第一方面的方法。

[0006] 在本公开的第四方面，提供了一种计算机可读存储介质。该计算机可读存储介质上存储有计算机程序，其可由处理器执行以执行根据本公开的第一方面的方法。

[0007] 应当理解，发明内容部分中所描述的内容并非旨在限定本公开的实施例的关键特征或重要特征，也不用于限制本公开的范围。本公开的其它特征将通过以下的描述而变得

容易理解。

附图说明

[0008] 在下文中,结合附图并参考以下详细说明,本公开各实现方式的上述和其他特征、优点及方面将变得更加明显。在附图中,相同或相似的附图标记表示相同或相似的元素,其中:

[0009] 图1示出了本公开的实施例能够在其中实现的示例环境的示意图;

[0010] 图2示出了根据本公开的一些实施例的用于信息处理的示例架构的示意图;

[0011] 图3示出了根据本公开的一些实施例的用于信息处理的过程的流程图;

[0012] 图4示出了根据本公开的一些实施例的用于信息处理的装置的示意性结构框图;以及

[0013] 图5示出了其中可以实施本公开的一个或多个实施例的计算设备的框图。

具体实施方式

[0014] 下面将参照附图更详细地描述本公开的实施例。虽然附图中示出了本公开的一些实施例,然而应当理解的是,本公开可以通过各种形式来实现,而且不应该被解释为限于这里阐述的实施例,相反,提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是,本公开的附图及实施例仅用于示例性作用,并非用于限制本公开的保护范围。

[0015] 在本公开的实施例的描述中,术语“包括”及其类似用语应当理解为开放性包含,即“包括但不限于”。术语“基于”应当理解为“至少部分地基于”。术语“一个实施例”或“该实施例”应当理解为“至少一个实施例”。术语“第一”、“第二”等等可以指代不同的或相同的对象。下文还可能包括其他明确的和隐含的定义。

[0016] 术语“响应于”表示相应的事件发生或者条件得以满足。将会理解,响应于该事件或者条件而被执行的后续动作的执行时机,与事件发生或者条件满足的时间,这二者不一定是强关联的。在某些情况下,后续动作可在事件发生或者条件成立时立即被执行;在另一些情况下,后续动作也可在事件发生或者条件成立后经过一段时间才被执行。

[0017] 需要说明的是,本公开的技术方案中,所涉及的用户个人信息的获取,存储和应用等,均符合相关法律法规的规定,且不违背公序良俗。

[0018] 可以理解的是,在使用本公开各实施例公开的技术方案之前,均应当根据相关法律法规通过适当的方式对本公开所涉及个人信息的类型、使用范围、使用场景等告知用户并获得用户的授权。

[0019] 例如,在响应于接收到用户的主动请求时,向用户发送提示信息,以明确地提示用户,其请求执行的操作将需要获得和使用到用户的个人信息,从而使得用户可以根据提示信息来自主地选择是否向执行本公开技术方案的操作的电子设备、应用程序、服务器或存储介质等软件或硬件提供个人信息。

[0020] 作为一种可选的但非限制性的实现方式,响应于接收到用户的主动请求,向用户发送提示信息的方式,例如可以是弹出窗口的方式,弹出窗口中可以以文字的方式呈现提示信息。此外,弹出窗口中还可以承载供用户选择“同意”或“不同意”向电子设备提供个人信息的选择控件。

[0021] 如上文所讨论的,推广方(例如商家)为了推广特定对象(例如实体\虚拟商品、服务等),往往通过互联网投放对象推荐信息(例如广告)。对于推广方,推广方期望将对象推荐信息针对性的推荐给目标用户群以提升推广的准确性。对于用户,用户期望能够针对性的获取对象推荐信息,这样用户可以直接获取符合自身偏好的对象的对象推荐信息,这有助于降低大量推广对用户造成的骚扰问题。传统上,大多是获取用户在目标平台(例如应用、网页等)上针对对象的评论文本,通过自然语言处理方法获取评论文本中的对象的属性信息(例如商品的尺寸),再结合“标签-属性”之间的映射关系,确定用户对应的标签。可以基于目标标签,对与目标标签对应的用户群进行对象推广。

[0022] 然而,针对对象的评论文本往往较为稀疏,且单个评论文本中出现的属性词缺少文本中的“上下文信息”后,不一定可以表示出用户的偏好(例如:“我看到对象A有5000毫安的电池,而对象B是4000毫安的”这句话中,并不表示用户偏好某一种电池类型,而是用户偏好“商品比较”意图)。此外,评论文本往往是用户对已经获取到的对象的评论文本(例如对已购买的商品的评论文本),这可能会导致当前再对发表评论文本的用户进行对象推广的效果不佳。

[0023] 有鉴于此,本公开的实施例提供了一种用于信息处理的方法。该方法包括:获取一组用户在目标平台上的一组历史会话消息,每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的。基于一组历史会话消息确定一组用户各自对应的用户偏好标签,每个用户偏好标签指示相应用户对对象的偏好。基于一组用户各自对应的用户偏好标签,从一组用户中选择与目标用户偏好标签对应的目标用户子组。基于目标用户子组中各个用户的目标特征信息,确定目标用户群,目标用户群的特征信息与目标特征信息相匹配,并且目标用户群具有目标用户偏好标签所指示的偏好。

[0024] 以此方式,可以基于用户在目标平台上的历史会话消息确定用户对应的用户偏好标签,可以保证确定的标签的准确性。可以基于与目标用户偏好标签相对应的用户组,确定特征信息与这个用户组所包含的用户的特征信息相匹配的目标用户群,有助于提升确定的目标用户群的准确性。

[0025] 图1示出了本公开的实施例能够在其中实现的示例环境100的示意图。如图1所示,环境100可以包括电子设备110。

[0026] 电子设备110可以获取一组用户101(例如可以包括用户101-1、101-2、……、101-M,其中M为正整数)在目标平台上的一组历史会话消息102(例如可以包括历史会话消息102-1、102-2、……、102-M)。每个历史会话消息102是由对应用户101在与客服关于特定对象的咨询过程中发出的,因此,用户历史会话消息与用户之间一一对应,一组历史会话消息102的数目M与一组用户101的M为相同数目。也即电子设备110可以获取一组用户101中每个用户101在目标平台上发出的历史会话消息102。目标平台可以是任意适当类型的平台,其例如可以包括但不限于网页、应用、链接、小程序等等。示例性地,目标平台可以是购物类应用。电子设备110可以获取一组用户101在该购物类应用上的一组历史会话消息102。

[0027] 电子设备110可以基于获取到的一组历史会话消息102确定一组用户101各自对应的用户偏好标签,每个用户偏好标签指示相应用户对对象的偏好。由于存在多个用户对应的用户偏好标签相同的情况,电子设备110可以基于一组用户101各自对应的用户偏好标签,从一组用户101中选择与目标用户偏好标签对应的目标用户子组112。可以理解,目标用

户子组112中所包括的至少一个用户为一组用户101中的用户。例如,目标用户子组112可以包括一组用户101中的用户101-i,101-j,101-m等等(这里的i,j,m仅用于表示不同的用户,不具备其他含义)。

[0028] 电子设备110可以基于确定的目标用户子组112来确定目标用户群114。目标用户群114例如可以包括一组用户101之外的多个用户103(例如可以包括用户103-1、103-2、……、103-K,其中K为与M无关的正整数)。

[0029] 电子设备110可以包括具有计算能力的任何计算系统,例如各种计算设备/系统、终端设备、服务端设备等。终端设备可以是任意类型的移动终端、固定终端或便携式终端,包括移动手机、台式计算机、膝上型计算机、笔记本计算机、上网本计算机、平板计算机、媒体计算机、多媒体平板、掌上电脑、便携式游戏终端、VR/AR设备、个人通信系统(Personal Communication System,PCS)设备、个人导航设备、个人数字助理(Personal Digital Assistant,PDA)、音频/视频播放器、数码相机/摄像机、定位设备、电视接收器、无线电广播接收器、电子书设备、游戏设备或者前述各项的任意组合,包括这些设备的配件和外设或者其任意组合。

[0030] 服务端设备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者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络、以及大数据和人工智能平台等基础云计算服务的云服务器。服务端设备例如可以包括计算系统/服务器,诸如大型机、边缘计算节点、云环境中的计算设备,等等。

[0031] 应当理解,仅出于示例性的目的描述环境100中各个元素的结构和功能,而不暗示对于本公开的范围的任何限制。

[0032] 以下将继续参考附图描述本公开的一些示例实施例。

[0033] 图2示出了根据本公开的一些实施例的用于信息处理的示例架构200的示意图。架构200可以在电子设备110处实现。下面参考图1描述架构200。

[0034] 如图2所示,架构200包括消息获取单元210。消息获取单元210被配置为获取一组用户101在目标平台上的一组历史会话消息102。在一些实施例中,消息获取单元210可以基于一组用户101的一组用户标识来获取这一组用户101在目标平台上的一组历史会话消息102。用户标识例如可以包括用户账号、用户ID、用户名称等等。

[0035] 这里的一组历史会话消息102中的每个历史会话消息至少包括对应用户101在与客服关于特定对象的咨询过程中所发出的消息。这里的客服例如可以为目标平台的客服、与特定对象相关联的客服(以特定对象为某一商品为例,与特定对象相关联的客服例如可以为出售该商品的商家的客服)等。在一些实施例中,消息获取单元210可以仅获取一组用户101在与客服关于特定对象的咨询过程中所发出的消息,也即消息获取单元210可以直接获取到一组历史会话消息。

[0036] 在一些实施例中,消息获取单元210还可以获取一组用户101中的每个用户101在目标平台上与客服之间的所有消息(包括用户所发出的消息和客服所发出的消息)。消息获取单元210进而可以对获取到的消息进行筛选以仅保留符合要求(也即符合为用户101在与客服关于特定对象的咨询过程中所发出的消息这一要求)的消息,这些被保留的消息被确定为一组历史会话消息102。

[0037] 如图2所示,架构200还包括标签确定单元220。一组历史会话消息102被提供给标签确定单元220以便标签确定单元220基于一组历史会话消息102,确定一组用户101各自对应的用户偏好标签225。在一些实施例中,为了提升标签确定的准确性,架构200还可以包括消息预处理单元215。一组历史会话消息102可以先被提供给消息预处理单元215。消息预处理单元215对这一组历史会话消息102执行预处理操作,并将经过预处理的一组历史会话消息提供给标签确定单元220。这里的预处理操作例如可以包括过滤操作。消息预处理单元215可以对一组历史会话消息102中的每个历史会话消息102执行过滤操作以过滤掉其中与对象无关的文本内容。示例性地,消息预处理单元215可以对每个历史会话消息102执行过滤操作以过滤掉每个历史会话消息102中的语气词(例如“呀、啊、哦”等),并将执行过滤操作后的一组历史会话消息102提供给标签确定单元220。

[0038] 标签确定单元220可以基于一组历史会话消息102(或者经过预处理的一组历史会话消息),确定一组用户101各自对应的用户偏好标签225。每个用户偏好标签225指示相应用户101对对象的偏好。标签确定单元220可以通过任意适当的方式确定一组用户101各自对应的用户偏好标签225。在一些实施例中,标签确定单元220可以包括意图标签确定子单元222和偏好标签确定子单元224。

[0039] 标签确定单元220获取到一组历史会话消息102后,将一组历史会话消息102提供给意图标签确定子单元222。意图标签确定子单元222可以基于一组历史会话消息102各自所包括的消息内容(包括但不限于文本、图像、音频等等),确定一组历史会话消息102各自对应的意图标签。每个意图标签指示在相应的历史会话消息102中相应用户101的意图。偏好标签确定子单元224可以预先获取到意图标签与用户偏好标签之间的映射关系。偏好标签确定子单元224可以基于意图标签与用户偏好标签之间的映射关系以及一组历史会话消息102各自对应的意图标签,确定一组用户101各自对应的用户偏好标签225。需要注意的是,在一些实施例中,一个意图标签仅可以对应与一个用户偏好标签,但一个用户偏好标签可以对应与多个意图标签,也即意图标签和用户偏好标签可以是多对一的关系。

[0040] 关于意图标签确定子单元222确定意图标签的具体方式,在一些实施例中,对于一组历史会话消息102中的每个历史会话消息102,意图标签确定子单元222可以确定该历史会话消息102所包括的文本是否包括意图词。这里的意图词可以是预先被设置好的意图词,其例如可以是用户预先确定的,也可以是电子设备110自行确定的。如果该历史会话消息所包括的文本包括意图词,意图标签确定子单元222可以基于预先获取到的意图词与意图标签之间的映射关系以及文本所包括的意图词,确定该历史会话消息对应的意图标签。示例性地,若历史会话消息102包括“比较XX和YY之间的区别”这样的文本,意图标签确定子单元222确定该文本中包括“比较”这个意图词后,可以基于意图词与意图标签之间的映射关系确定“比较”这个意图词对应的意图标签为“比较不同对象”,也即可以确定包括该文本的历史会话消息102对应的意图标签为“比较不同对象”。

[0041] 在一些实施例中,意图标签确定子单元222还可以利用机器学习模型处理一组历史会话消息102各自所包括的消息内容,以确定一组历史会话消息102各自对应的意图标签。这里的机器学习模型也称为意图识别模型,其可以是任意适当的模型,例如,BERT模型。意图标签确定子单元222可以依次将每个历史会话消息102输入该机器学习模型。机器学习模型可以预先获取到大量的意图标签。机器学习模型可以确定获取的历史会话消息102对

应于哪一个意图标签,并输出该历史会话消息102对应的意图标签。

[0042] 具体地,机器学习模型中的编码器基于如下公式可以将输入的历史会话消息102映射为输入编码向量:

[0043] $E_c=T_c+S_c+P_c$ (1)

[0044] $V_c=E_cX_c$ (2)

[0045] 其中,T_c为词嵌入,S_c为段落嵌入,P_c为位置嵌入,X_c为输入的历史会话消息102的消息内容,E_c为消息内容侧嵌入矩阵。

[0046] 机器学习模型进而可以基于如下公式确定输入编码向量对应的意图标签:

[0047] $y_{preference}=BERT(V_c)$ (3)

[0048] $$Attention(Q,K,V)=softmax(\frac{QK^T}{\sqrt{dk}})V \quad (4)$$

[0049] 其中,$y_{preference}$为意图判断结果,Q,K和V分别是查询(query)、键(key)和值(value)对应的矩阵,dk是键的维度。

[0050] 需要注意的是,一个历史会话消息102可以包括多个消息内容(例如可以包括多个文本、多个图像、多个音频等等)。在一些实施例中,意图标签确定子单元222可以分别确定每个消息内容对应的意图标签,并将每个历史会话消息中出现次数最多的意图标签确定为该历史会话消息对应的意图标签。例如,若某个历史会话消息102包括5个消息内容且其中有4个消息内容对应与意图标签A,1个消息内容对应于意图标签B,则意图标签确定子单元222可以确定该历史会话消息102对应的意图标签为意图标签A。

[0051] 标签确定单元220确定一组用户101各自对应的用户偏好标签225后,可以确定一组用户101对应的一组用户偏好标签225。可以理解,存在不同用户101对应于相同的用户偏好标签225的情况,因此,这一组用户偏好标签225中存在重复的用户偏好标签225。用户选择单元250可以基于这一组用户偏好标签225,从一组用户101中选择与目标用户偏好标签对应的目标用户子组112。这里的目标用户偏好标签为一组用户偏好标签225中的任意用户偏好标签225,其可以是用户自行确定好并提供给用户选择单元250的。

[0052] 在一些实施例中,用户选择单元250可以基于目标用户偏好标签对这一组用户101进行查询,并将查找到的至少一个用户确定为目标用户子组。在一些实施例中,用户选择单元250还可以获取到基于不同用户偏好标签225进行划分所得到的多个用户子组,用户选择单元250可以直接基于目标用户偏好标签,从多个用户子组中确定与目标用户偏好标签相匹配的用户子组,该用户子组即为目标用户子组112。在这种情况下,标签确定单元220所确定的一组用户101对应的一组用户偏好标签225被提供给划分单元230。划分单元230可以基于一组用户101各自对应的用户偏好标签225,,将一组用户划分为多个用户子组235(包括用户子组235-1、235-2、……、235-N,其中N为正小于等于M的正整数),每个用户子组中的用户对应相同的用户偏好标签225。

[0053] 在一些实施例中,架构200还可以包括文本生成单元240。文本生成单元240可以基于多个用户子组235各自对应的历史会话消息102,确定用于描述每个用户子组235对应的用户偏好标签225的描述文本245。也即文本生成单元240可以确定与多个用户子组235分别对应的多个描述文本245(包括描述文本245-1、245-2、……、245-N)。

[0054] 关于确定描述文本的具体方式,针对每个用户子组235对应的历史会话消息102

(若该用户子组235包括多个用户,则该用户子组235对应的历史会话消息102包括多个历史会话消息102),在一些实施例中,文本生成单元240可以先生成对应的历史对话消息102的最长公共文本序列。示例性地,若历史对话消息102包括的消息内容对应于多个文本序列,这多个文本序列包括本身即为文本类型的消息内容(也即本身即为文本序列)和其他类型(例如音频)的消息内容转化后所得到的文本序列。若这多个文本序列中包括文本序列"我喜欢听音乐,这款设备支持吗"和文本序列"你也喜欢听音乐吗",则这两个文本序列的最长公共文本序列为"喜欢听音乐"。文本生成单元240进而可以将最长公共文本序列从文本序列中删除以更新文本序列,并对更新后的文本序列多次重复寻找最长公共文本序列和删除最长公共文本序列这样的过程直到没有最长公共文本序列。

[0055] 示例性地,对于文本序列"我喜欢听音乐,这款设备支持吗"而言,最长公共文本序列为"喜欢听音乐",删除最长公共文本序列的剩余部分为"我,这款手机支持吗"。对于文本序列"你也喜欢听音乐吗",最长公共文本序列为"喜欢听音乐",删除最长公共文本序列的剩余部分为"你也吗"。文本生成单元240可以针对这两个文本序列的剩余部分"我,这款手机支持吗"和"你也吗"再次寻找最长公共文本序列"吗",并重复这样的步骤直到没有最长公共文本序列。

[0056] 文本生成单元240进而可以基于如下公式计算不同文本序列的相似度分数:

[0057] $$Sim_{score} = \frac{2 * L_{match}}{L_{total}} \quad (5)$$

[0058] 其中,L_{match}为最长公共文本序列的总长度,如上述示例所示,最长公共文本序列为"喜欢听音乐"、"吗",则$L_{match}=6$。L_{total}为两个文本序列的总文本长度,其例如可以为14+8=22,则相似度分数$Sim_{score}=2*6/22\approx0.55$。

[0059] 文本生成单元240可以基于预先确定的相似度分数阈值,将对应相似度分数高于相似度分数阈值的多个文本序列进行过滤以仅保留1个可以代表这些相似的文本序列的文本序列。也即仅保留一个具有代表性的文本序列,该文本序列即为对应于历史会话消息102的用户偏好标签225的描述文本。

[0060] 在一些实施例中,对于每个用户子组235,文本生成单元240还可以采用类似的方式,基于该用户子组235对应的历史会话消息102中每个意图标签对应的历史会话消息,确定用于描述该用户子组235对应的用户偏好标签225所对应的多个意图标签中的每个意图标签的描述文本。

[0061] 以此方式,可以通过提供描述文本来使得用户知晓不同意图标签和/或不同用户偏好标签的含义,有助于帮助用户确定其想要选择的目标用户偏好标签,可以提升用户的用户体验。

[0062] 如图2所示,架构200还包括用户确定单元260。用户选择单元250基于目标用户偏好标签确定目标用户子组112后,用户确定单元260可以基于目标用户子组112中各个用户101的目标特征信息,确定目标用户群114。这里的目标特征信息例如可以包括用户画像特征、用户行为特征等多个适当的维度的特征。在一些实施例中,用户确定单元260可以利用相似人群扩展算法(Looklike算法),基于目标用户子组112中各个用户101的目标特征信息来确定目标用户群114。目标用户群114的特征信息与目标特征信息相匹配,并且目标用户

群114具有目标用户偏好标签所指示的偏好。

[0063] 在一些实施例中,用户确定单元260可以直接将利用Looklike算法所得到的包括目标用户子组112的用户群确定为目标用户群114。在一些实施例中,用户确定单元260还可以利用Looklike算法确定与目标特征信息相匹配第一用户群,第一用户群包括目标用户子组112和另外的多个用户261(多个用户261例如可以包括用户203-1、203-2、……、203-K,K为与M和N均无关的正整数,用户203例如可以对应于图1中的用户103)。由于电子设备110是基于历史会话信息来实现的目标用户群的确定,而具有历史会话信息的一组用户101往往不再需要被进行对象推广(例如具有历史会话信息的用户往往已经购买商品,不需要再对其进行商品推广)。为了保证目标用户群114中所包括的多个用户为与一组用户101不同的用户,以提升后续对目标用户群114提供对象推荐信息的推荐效果,在一些实施例中,用户确定单元260还可以从第一用户群中删除目标用户子组112以确定目标用户群114。可以理解,目标用户群114即为多个用户261。

[0064] 在一些实施例中,电子设备110获取到目标用户群114后,可以基于目标用户偏好标签所指示的偏好,向目标用户群114提供对象推荐信息(例如广告、短消息、视频等),以实现对对象的推广。示例性地,目标用户偏好标签可以指示具有该目标用户偏好标签的用户(例如目标用户子组112和/或目标用户群114)偏好与某一类型的对象。例如,目标用户偏好标签可以指示具有该目标用户偏好标签的用户偏好服装类的商品。电子设备110可以将确定的目标用户群114提供给与目标用户偏好标签所指示的偏好相匹配的对象推广方。例如,可以将偏好服装类商品的用户提供给服装类商品的推广方。推广方可以提供自身的对象的对象推荐信息。例如,推广方可以提供针对自身的服装类商品的广告。电子设备110进而可以将该对象推荐信息提供给目标用户群114。

[0065] 在一些实施例中,对于目标用户偏好标签中对应不同意图标签的不同用户,推广方还可以提供不同的对象推荐信息。例如,推广方可以针对目标用户群114中对对应于意图标签A的至少一个用户提供对象推荐信息A,对对应于意图标签B的至少一个用户提供对象推荐信息B。在一些实施例中,目标用户偏好标签也可以包括多个用户偏好标签,电子设备110可以分别确定这多个用户偏好标签对应的多个目标用户群。推广方可以对这多个目标用户群分别提供不同的对象推荐信息。

[0066] 以此方式,推广方可以针对不同的用户(例如不同意图标签对应的不同用户以及不同用户偏好标签对应的不同用户)制定并提供不同的对象推荐信息,可以实现对不同用户的差异化推广,有助于提升对象的推广效果。

[0067] 综上所述,本公开的实施例可以基于用户在目标平台上的历史会话消息确定用户对应的用户偏好标签,可以保证确定的标签的准确性。可以基于与目标用户偏好标签相对应的用户组,确定特征信息与这个用户组所包含的用户的特征信息相匹配的目标用户群,有助于提升确定的目标用户群的准确性。

[0068] 上文已经描述了信息处理的各个步骤的具体细节,提供了一种用于信息处理的方法。图3示出了根据本公开的一些实施例的用于信息处理的过程300的流程图。过程300可以被实现在电子设备110处。下面参考图1描述过程300。

[0069] 如图3所示,在框310,电子设备110获取一组用户在目标平台上的一组历史会话消息,每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的。

[0070] 在框320,电子设备110基于一组历史会话消息确定一组用户各自对应的用户偏好标签,每个用户偏好标签指示相应用户对对象的偏好。

[0071] 在框330,电子设备110基于一组用户各自对应的用户偏好标签,从一组用户中选择与目标用户偏好标签对应的目标用户子组。

[0072] 在框340,电子设备110基于目标用户子组中各个用户的目标特征信息,确定目标用户群,目标用户群的特征信息与目标特征信息相匹配,并且目标用户群具有目标用户偏好标签所指示的偏好。

[0073] 在一些实施例中,基于一组历史会话消息确定一组用户各自对应的用户偏好标签包括:基于一组历史会话消息各自所包括的消息内容,确定一组历史会话消息各自对应的意图标签,每个意图标签指示在历史会话消息中相应用户的意图;以及基于意图标签与用户偏好标签之间的映射关系以及一组历史会话消息各自对应的意图标签,确定一组用户各自对应的用户偏好标签。

[0074] 在一些实施例中,确定一组历史会话消息各自对应的意图标签包括:对于一组历史会话消息中的每个历史会话消息,确定该历史会话消息所包括的文本是否包括意图词;以及如果该历史会话消息所包括的文本包括意图词,基于意图词与意图标签之间的映射关系,确定该历史会话消息对应的意图标签。

[0075] 在一些实施例中,确定一组历史会话消息各自对应的意图标签包括:利用机器学习模型处理一组历史会话消息各自所包括的消息内容,以确定一组历史会话消息各自对应的意图标签。

[0076] 在一些实施例中,过程300还包括:基于一组用户各自对应的用户偏好标签,将一组用户划分为多个用户子组,每个用户子组中的用户对应相同的用户偏好标签;以及基于多个用户子组各自对应的历史会话消息,确定用于描述每个用户子组对应的用户偏好标签的描述文本。

[0077] 在一些实施例中,确定目标用户群包括:利用相似人群扩展算法,基于目标用户子组中各个用户的目标特征信息来确定目标用户群。

[0078] 在一些实施例中,利用相似人群扩展算法确定目标用户群包括:利用相似人群扩展算法,确定与目标特征信息相匹配第一用户群,第一用户群包括目标用户子组和另外的多个用户;以及从第一用户群中删除目标用户子组以确定目标用户群。

[0079] 在一些实施例中,过程300还包括:基于目标用户偏好标签所指示的偏好,向目标用户群提供对象推荐信息。

[0080] 本公开的实施例还提供了用于实现上述方法或过程的相应装置。图4示出了根据本公开的某些实施例的用于信息处理的装置400的示意性结构框图。装置400可以被实现为或者被包括在电子设备110中。装置400中的各个模块/组件可以由硬件、软件、固件或者它们的任意组合来实现。

[0081] 如图4所示,装置400包括消息获取模块410,被配置为获取一组用户在目标平台上的一组历史会话消息,每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的。装置400还包括标签确定模块420,被配置为基于一组历史会话消息确定一组用户各自对应的用户偏好标签,每个用户偏好标签指示相应用户对对象的偏好。装置400还包括用户选择模块430,被配置为基于一组用户各自对应的用户偏好标签,从一组用户中选择

与目标用户偏好标签对应的目标用户子组。装置400还包括用户确定模块440,被配置为基于目标用户子组中各个用户的目标特征信息,确定目标用户群,目标用户群的特征信息与目标特征信息相匹配,并且目标用户群具有目标用户偏好标签所指示的偏好。

[0082] 在一些实施例中,标签确定模块420包括:意图标签确定模块,被配置为基于一组历史会话消息各自所包括的消息内容,确定一组历史会话消息各自对应的意图标签,每个意图标签指示在历史会话消息中相应用户的意图;以及偏好标签确定模块,被配置为基于意图标签与用户偏好标签之间的映射关系以及一组历史会话消息各自对应的意图标签,确定一组用户各自对应的用户偏好标签。

[0083] 在一些实施例中,意图标签确定模块被进一步配置为:对于一组历史会话消息中的每个历史会话消息,确定该历史会话消息所包括的文本是否包括意图词;以及如果该历史会话消息所包括的文本包括意图词,基于意图词与意图标签之间的映射关系,确定该历史会话消息对应的意图标签。

[0084] 在一些实施例中,意图标签确定模块被进一步配置为:利用机器学习模型处理一组历史会话消息各自所包括的消息内容,以确定一组历史会话消息各自对应的意图标签。

[0085] 在一些实施例中,装置400还包括:划分模块,被配置为基于一组用户各自对应的用户偏好标签,将一组用户划分为多个用户子组,每个用户子组中的用户对应相同的用户偏好标签;以及描述文本确定模块,被配置为基于多个用户子组各自对应的历史会话消息,确定用于描述每个用户子组对应的用户偏好标签的描述文本。

[0086] 在一些实施例中,用户确定模块440还被配置为:利用相似人群扩展算法,基于目标用户子组中各个用户的目标特征信息来确定目标用户群。

[0087] 在一些实施例中,用户确定模块440包括:第一用户群确定模块,被配置为利用相似人群扩展算法,确定与目标特征信息相匹配第一用户群,第一用户群包括目标用户子组和另外的多个用户;以及删除模块,被配置为从第一用户群中删除目标用户子组以确定目标用户群。

[0088] 在一些实施例中,装置400还包括:信息提供模块,被配置为基于目标用户偏好标签所指示的偏好,向目标用户群提供对象推荐信息。

[0089] 装置400中所包括的单元和/或模块可以利用各种方式来实现,包括软件、硬件、固件或其任意组合。在一些实施例中,一个或多个单元和/或模块可以使用软件和/或固件来实现,例如存储在存储介质上的机器可执行指令。除了机器可执行指令之外或者作为替代,装置400中的部分或者全部单元和/或模块可以至少部分地由一个或多个硬件逻辑组件来实现。作为示例而非限制,可以使用的示范类型的硬件逻辑组件包括现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准品(ASSP)、片上系统(SOC)、复杂可编程逻辑器件(CPLD),等等。

[0090] 应当理解,以上方法中的一个或多个步骤可以由适当的电子设备或电子设备的组合来执行。这样的电子设备或电子设备的组合例如可以包括图1中的电子设备110。

[0091] 图5示出了其中可以实施本公开的一个或多个实施例的电子设备500的框图。应当理解,图5所示出的电子设备500仅仅是示例性的,而不应当构成对本文所描述的实施例的功能和范围的任何限制。图5所示出的电子设备500可以用于实现图1的电子设备110。

[0092] 如图5所示,电子设备500是通用电子设备的形式。电子设备500的组件可以包括但

不限于一个或多个处理器或处理单元510、存储器520、存储设备530、一个或多个通信单元540、一个或多个输入设备550以及一个或多个输出设备560。处理单元510可以是实际或虚拟处理器并且能够根据存储器520中存储的程序来执行各种处理。在多处理器系统中,多个处理单元并行执行计算机可执行指令,以提高电子设备500的并行处理能力。

[0093] 电子设备500通常包括多个计算机存储介质。这样的介质可以是电子设备500可访问的任何可以获得的介质,包括但不限于易失性和非易失性介质、可拆卸和不可拆卸介质。存储器520可以是易失性存储器(例如寄存器、高速缓存、随机访问存储器(RAM))、非易失性存储器(例如,只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、闪存)或它们的某种组合。存储设备530可以是可拆卸或不可拆卸的介质,并且可以包括机器可读介质,诸如闪存驱动、磁盘或者任何其他介质,其可以能够用于存储信息和/或数据并且可以在电子设备500内被访问。

[0094] 电子设备500可以进一步包括另外的可拆卸/不可拆卸、易失性/非易失性存储介质。尽管未在图5中示出,可以提供用于从可拆卸、非易失性磁盘(例如“软盘”)进行读取或写入的磁盘驱动和用于从可拆卸、非易失性光盘进行读取或写入的光盘驱动。在这些情况中,每个驱动可以由一个或多个数据介质接口被连接至总线(未示出)。存储器520可以包括计算机程序产品525,其具有一个或多个程序模块,这些程序模块被配置为执行本公开的各种实施例的各种方法或动作。

[0095] 通信单元540实现通过通信介质与其他电子设备进行通信。附加地,电子设备500的组件的功能可以以单个计算集群或多个计算机器来实现,这些计算机器能够通过通信连接进行通信。因此,电子设备500可以使用与一个或多个其他服务器、网络个人计算机(PC)或者另一个网络节点的逻辑连接来在联网环境中进行操作。

[0096] 输入设备550可以是一个或多个输入设备,例如鼠标、键盘、追踪球等。输出设备560可以是一个或多个输出设备,例如显示器、扬声器、打印机等。电子设备500还可以根据需要通过通信单元540与一个或多个外部设备(未示出)进行通信,外部设备诸如存储设备、显示设备等,与一个或多个使得用户与电子设备500交互的设备进行通信,或者与使得电子设备500与一个或多个其他电子设备通信的任何设备(例如,网卡、调制解调器等)进行通信。这样的通信可以经由输入/输出(I/O)接口(未示出)来执行。

[0097] 根据本公开的示例性实现方式,提供了一种计算机可读存储介质,其上存储有计算机可执行指令,其中计算机可执行指令被处理器执行以实现上文描述的方法。根据本公开的示例性实现方式,还提供了一种计算机程序产品,计算机程序产品被有形地存储在非瞬态计算机可读介质上并且包括计算机可执行指令,而计算机可执行指令被处理器执行以实现上文描述的方法。

[0098] 这里参照根据本公开实现的方法、装置、设备和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0099] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理单元,从而生产出一种机器,使得这些指令在通过计算机或其他可编程数据处理装置的处理单元执行时,产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中,这

些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作,从而,存储有指令的计算机可读介质则包括一个制造品,其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0100] 可以把计算机可读程序指令加载到计算机、其他可编程数据处理装置、或其他设备上,使得在计算机、其他可编程数据处理装置或其他设备上执行一系列操作步骤,以产生计算机实现的过程,从而使得在计算机、其他可编程数据处理装置、或其他设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0101] 附图中的流程图和框图显示了根据本公开的多个实现的系统、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、程序段或指令的一部分,模块、程序段或指令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。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,两个连续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。也要注意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动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。

[0102] 以上已经描述了本公开的各实现,上述说明是示例性的,并非穷尽性的,并且也不限于所公开的各实现。在不偏离所说明的各实现的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实现的原理、实际应用或对市场中的技术的改进,或者使本技术领域的其他普通技术人员能理解本文公开的各个实现方式。

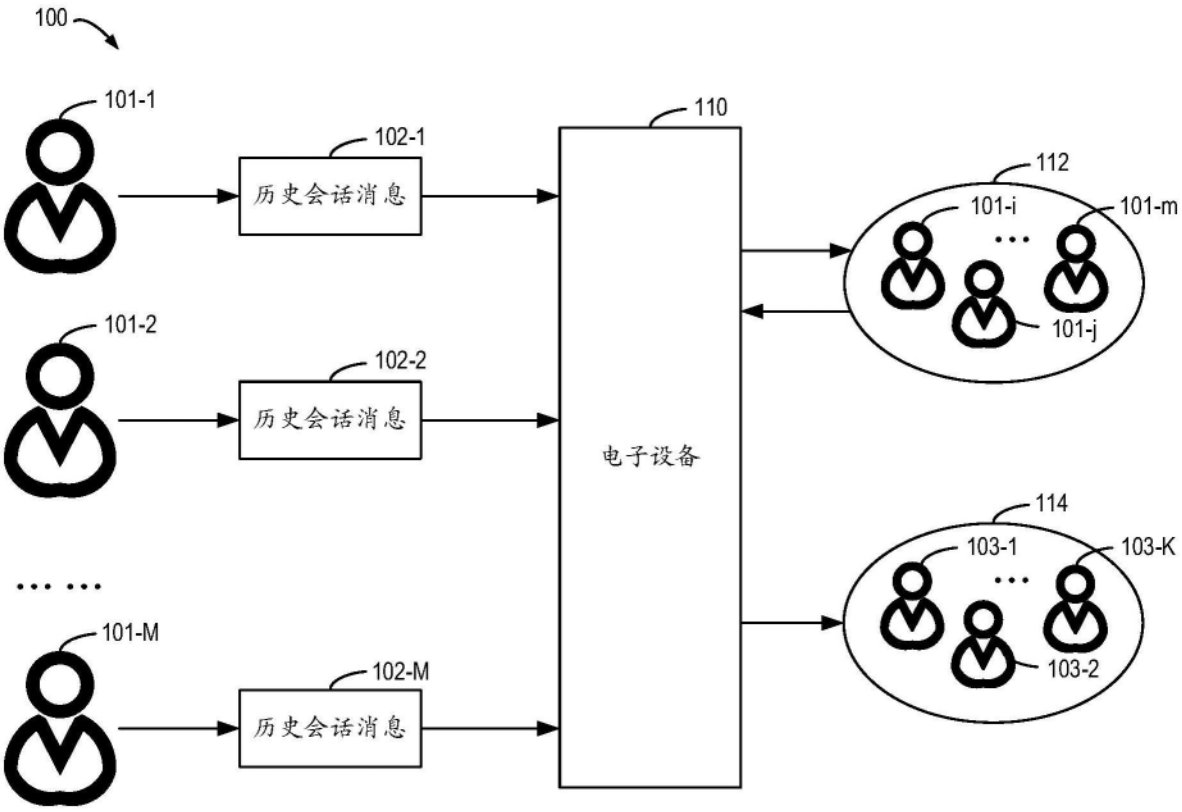

100
101-1
102-1
历史会话消息
110
112
101-i
101-m
101-j
101-2
102-2
历史会话消息
电子设备
114
103-1
103-K
103-2
101-M
102-M
历史会话消息

图1

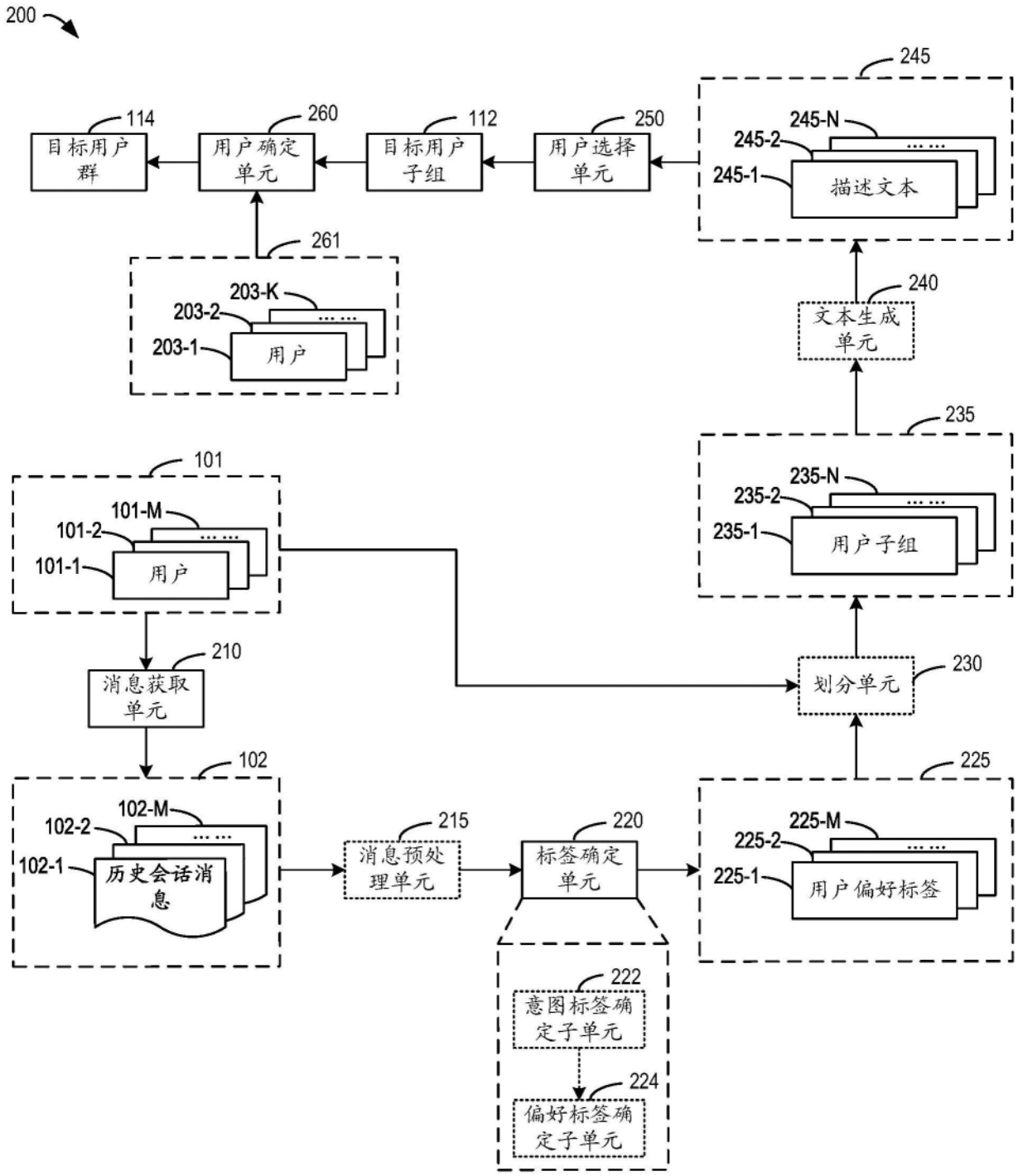

200
245
245-N
245-2
245-1
描述文本
250
用户选择单元
112
目标用户子组
260
用户确定单元
114
目标用户群
261
203-K
203-2
203-1
用户
240
文本生成单元
235
235-N
235-2
235-1
用户子组
101
101-M
101-2
101-1
用户
210
消息获取单元
230
划分单元
102
102-M
102-2
102-1
历史会话消息
215
消息预处理单元
220
标签确定单元
225
225-M
225-2
225-1
用户偏好标签
222
意图标签确定子单元
224
偏好标签确定子单元

图2

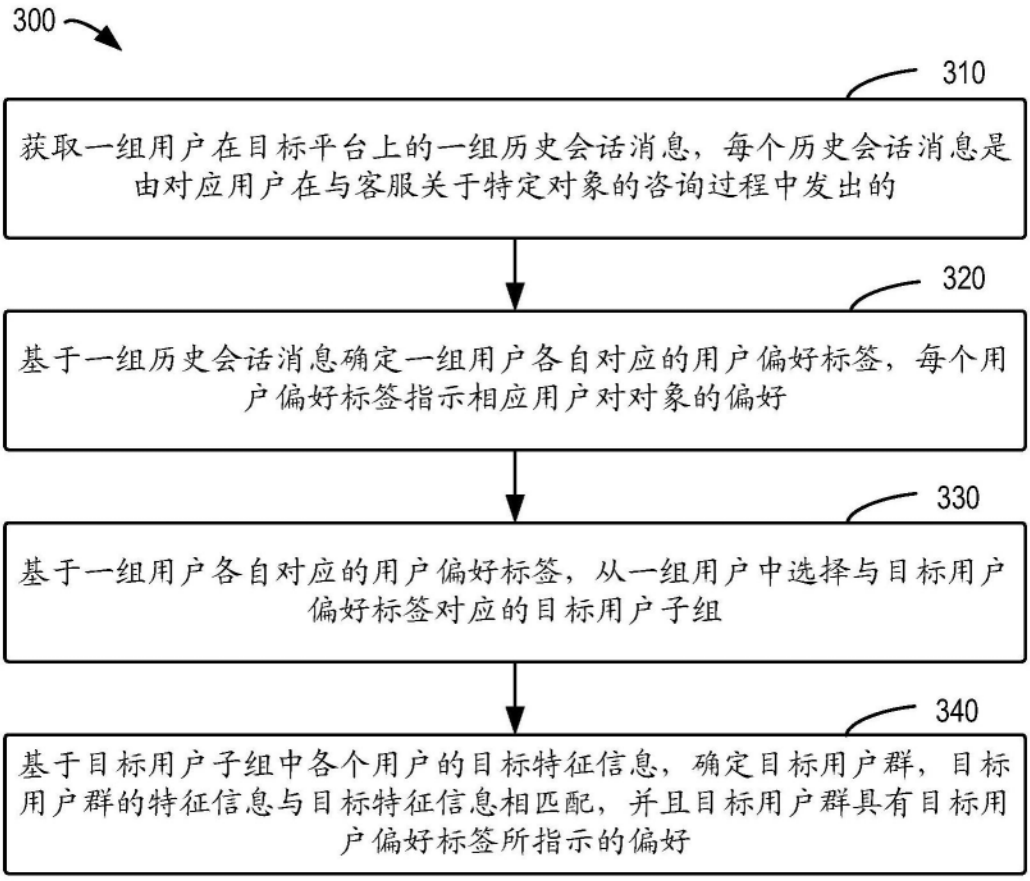

300
310
获取一组用户在目标平台上的一组历史会话消息，每个历史会话消息是由对应用户在与客服关于特定对象的咨询过程中发出的
320
基于一组历史会话消息确定一组用户各自对应的用户偏好标签，每个用户偏好标签指示相应用户对对象的偏好
330
基于一组用户各自对应的用户偏好标签，从一组用户中选择与目标用户偏好标签对应的目标用户子组
340
基于目标用户子组中各个用户的目标特征信息，确定目标用户群，目标用户群的特征信息与目标特征信息相匹配，并且目标用户群具有目标用户偏好标签所指示的偏好

图3

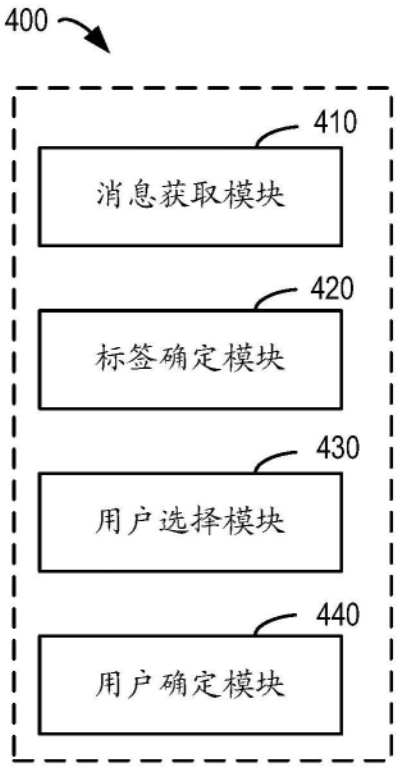

400
410
消息获取模块
420
标签确定模块
430
用户选择模块
440
用户确定模块

图4

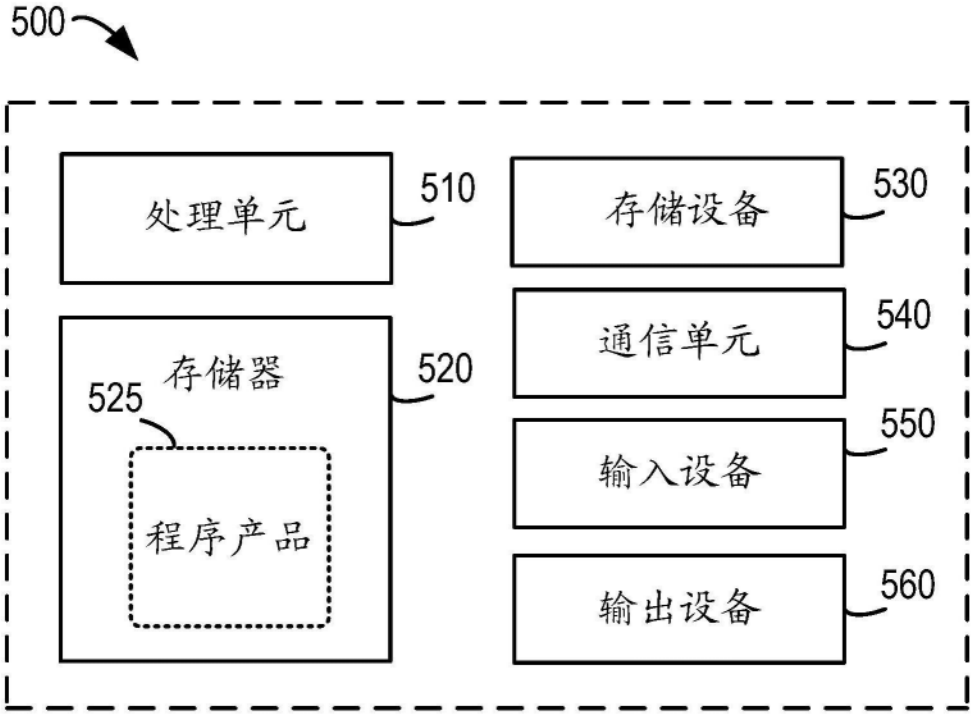

500
处理单元
510
存储设备
530
存储器
525
520
通信单元
540
程序产品
输入设备
550
输出设备
560

图5