



(12) 发明专利申请

(10) 申请公布号 CN 114490947 A

(43) 申请公布日 2022. 05. 13

(21) 申请号 202210142940.9

G06V 40/16 (2022.01)

(22) 申请日 2022.02.16

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

(71) 申请人 平安国际智慧城市科技股份有限公司

G10L 15/22 (2006.01)

G10L 15/26 (2006.01)

G10L 25/63 (2013.01)

地址 518000 广东省深圳市前海深港合作区妈湾兴海大道3048号前海自贸大厦1-34层

(72) 发明人 刘荣荣

(74) 专利代理机构 深圳中一联合知识产权代理有限公司 44414

专利代理师 姚泽鑫

(51) Int. Cl.

G06F 16/33 (2019.01)

G06F 16/332 (2019.01)

G06F 40/30 (2020.01)

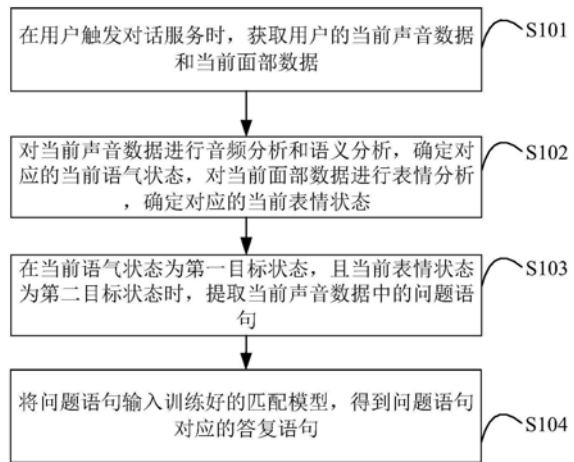
权利要求书2页 说明书11页 附图2页

(54) 发明名称

基于人工智能的对话服务方法、装置、服务器及介质

(57) 摘要

本申请适用于人工智能领域,尤其涉及一种基于人工智能的对话服务方法、装置、服务器及介质。该方法在触发对话服务时,获取用户的当前声音数据和当前面部数据,对当前声音数据进行音频分析和语义分析,确定当前语气状态,对当前面部数据进行表情分析,确定当前表情状态,在当前语气状态为第一目标状态,且当前表情状态为第二目标状态时,提取当前声音数据中的问题语句,将问题语句输入训练好的匹配模型,得到问题语句对应的答复语句,将语气与表情相结合来识别用户意图,并在语气和表情对应的状态均满足条件时生成相应的答复语句,避免在用户出现不满等意图的情况下仍继续进行对话,从而可以准确把握用户意图,提升对话服务的服务质量。



1. 一种基于人工智能的对话服务方法,其特征在于,所述对话服务方法包括:
 - 在用户触发对话服务时,获取所述用户的当前声音数据和当前面部数据;
 - 对所述当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对所述当前面部数据进行表情分析,确定对应的当前表情状态;
 - 在所述当前语气状态为第一目标状态,且所述当前表情状态为第二目标状态时,提取所述当前声音数据中的问题语句,所述第一目标状态表征所述用户未出现消极语气,所述第二目标状态表征所述用户未出现消极表情;
 - 将所述问题语句输入训练好的匹配模型,得到所述问题语句对应的答复语句。
2. 根据权利要求1所述的对话服务方法,其特征在于,在所述提取所述当前声音数据中的问题语句之前,还包括:
 - 若所述当前声音数据存在前一声音数据,则对所述前一声音数据进行音频分析和语义分析,确定对应的前一语气状态;
 - 获取训练好的提取模型;
 - 根据所述前一语气状态,调整所述训练好的提取模型,得到调整后的提取模型;
 - 相应地,所述提取所述当前声音数据中的问题语句包括:
 - 使用所述调整后的提取模型,提取所述当前声音数据中的问题语句。
3. 根据权利要求1所述的对话服务方法,其特征在于,在所述对所述当前声音数据进行音频分析和语义分析,确定对应的当前语气状态之后,还包括:
 - 在所述当前语气状态为第三目标状态和/或所述当前表情状态为第四目标状态时,获取人工服务的人员状态信息,所述第三目标状态表征所述用户出现消极语气,所述第四目标状态表征所述用户出现消极表情;
 - 根据所述人员状态信息,确定目标人员,并将所述当前声音数据发送给所述目标人员。
4. 根据权利要求3所述的对话服务方法,其特征在于,在所述将所述当前声音数据发送给所述目标人员之后,还包括:
 - 基于所述用户触发的对话服务,将所述目标人员与所述用户对接;
 - 在对接成功的情况下,获取所述目标人员发送的回复信息,将所述回复信息发送给所述用户。
5. 根据权利要求1所述的对话服务方法,其特征在于,在所述将所述问题语句输入训练好的匹配模型,得到所述问题语句对应的答复语句之后,还包括:
 - 从答复话术数据库中获取所述问题语句对应的答复话术;
 - 将所述答复语句与所述答复话术结合,得到所述问题语句对应的答复。
6. 根据权利要求1至5任一项所述的对话服务方法,其特征在于,所述对所述当前声音数据进行音频分析和语义分析,确定对应的当前语气状态包括:
 - 检测所述当前声音数据中声音的振幅和频率;
 - 识别所述当前声音数据的语义,确定所述当前声音数据对应的态度;
 - 在所述振幅低于第一阈值或者所述频率低于第二阈值,且所述当前声音数据对应的态度为积极时,确定所述当前语气状态为第一目标状态;
 - 在所述当前声音数据对应的态度为消极和/或所述振幅不低于所述第一阈值且所述频率不低于所述第二阈值时,确定所述当前语气状态不为第一目标状态。

7. 根据权利要求6所述的对话服务方法,其特征在于,所述识别所述当前声音数据的语义,确定所述当前声音数据对应的态度包括:

对所述当前声音数据进行降噪处理,得到降噪后的声音数据;

从降噪后的声音数据中提取人声数据;

将所述人声数据转换为文字数据,使用自然语言处理识别所述文字数据,确定所述文字数据对应的态度为所述当前声音数据对应的态度。

8. 一种基于人工智能的对话服务装置,其特征在于,所述对话服务装置包括:

数据获取模块,用于在用户触发对话服务时,获取所述用户的当前声音数据和当前面部数据;

数据分析模块,用于对所述当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对所述当前面部数据进行表情分析,确定对应的当前表情状态,所述第一目标状态表征所述用户未出现消极语气,所述第二目标状态表征所述用户未出现消极表情;

问题提取模块,用于在所述当前语气状态为第一目标状态,且所述当前表情状态为第二目标状态时,提取所述当前声音数据中的问题语句;

问题答复模块,用于将所述问题语句输入训练好的匹配模型,得到所述问题语句对应的答复语句。

9. 一种服务器,其特征在于,所述服务器包括处理器、存储器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至7任一项所述的对话服务方法。

10. 一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至7任一项所述的对话服务方法。

基于人工智能的对话服务方法、装置、服务器及介质

技术领域

[0001] 本申请属于人工智能领域,尤其涉及一种基于人工智能的对话服务方法、装置、服务器及介质。

背景技术

[0002] 目前,通过具有交互功能的线上服务平台,可以让用户获得相关资讯、信息或服务。用户在使用线上服务平台上进行咨询等服务时,该线上服务平台通常采用人工智能的方式对用户提出的问题进行解析,并自动且智能生成回答语句,从而可以提高了线上服务平台的回复效率,但是人工智能并不能完全代替人工处理所有问题,并且仅通过对语句进行语义分析并不能准确的表征用户的意图,容易出现用户对线上服务平台的回复不满意的情况,影响用户体验,导致用户投诉,因此,如何准确识别用户意图以提高服务质量成为亟待解决的问题。

发明内容

[0003] 有鉴于此,本申请实施例提供了一种基于人工智能的对话服务方法、装置、服务器及介质,以解决现有技术对用户意图识别不准,影响服务质量的问题。

[0004] 第一方面,本申请实施例提供一种基于人工智能的对话服务方法,所述对话服务方法包括:

[0005] 在用户触发对话服务时,获取所述用户的当前声音数据和当前面部数据;

[0006] 对所述当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对所述当前面部数据进行表情分析,确定对应的当前表情状态;

[0007] 在所述当前语气状态为第一目标状态,且所述当前表情状态为第二目标状态时,提取所述当前声音数据中的问题语句,所述第一目标状态表征所述用户未出现消极语气,所述第二目标状态表征所述用户未出现消极表情;

[0008] 将所述问题语句输入训练好的匹配模型,得到所述问题语句对应的答复语句。

[0009] 第二方面,本申请实施例提供一种基于人工智能的对话服务装置,所述对话服务装置包括:

[0010] 数据获取模块,用于在用户触发对话服务时,获取所述用户的当前声音数据和当前面部数据;

[0011] 数据分析模块,用于对所述当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对所述当前面部数据进行表情分析,确定对应的当前表情状态;

[0012] 问题提取模块,用于在所述当前语气状态为第一目标状态,且所述当前表情状态为第二目标状态时,提取所述当前声音数据中的问题语句,所述第一目标状态表征所述用户未出现消极语气,所述第二目标状态表征所述用户未出现消极表情;

[0013] 问题答复模块,用于将所述问题语句输入训练好的匹配模型,得到所述问题语句对应的答复语句。

[0014] 第三方面,本申请实施例提供一种服务器,所述服务器包括处理器、存储器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如第一方面所述的对话服务方法。

[0015] 第四方面,本申请实施例提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现如第一方面所述的对话服务方法。

[0016] 第五方面,本申请实施例提供一种计算机程序产品,当计算机程序产品在服务器上运行时,使得服务器执行上述第一方面所述的对话服务方法。

[0017] 本申请实施例与现有技术相比存在的有益效果是:本申请在用户触发对话服务时,获取用户的当前声音数据和当前面部数据,对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对当前面部数据进行表情分析,确定对应的当前表情状态,在当前语气状态为第一目标状态,且当前表情状态为第二目标状态时,提取当前声音数据中的问题语句,将问题语句输入训练好的匹配模型,得到问题语句对应的答复语句,将语气与表情相结合来识别用户意图,并在语气和表情对应的状态均满足条件时生成相应的答复语句,避免在用户出现不满等意图的情况下仍继续进行对话,从而可以准确把握用户意图,提升对话服务的服务质量。

附图说明

[0018] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0019] 图1是本申请实施例一提供的一种基于人工智能的对话服务方法的流程示意图;

[0020] 图2是本申请实施例二提供的一种基于人工智能的对话服务方法的流程示意图;

[0021] 图3是本申请实施例三提供的一种基于人工智能的对话服务装置的结构示意图;

[0022] 图4是本申请实施例四提供的一种服务器的结构示意图。

具体实施方式

[0023] 以下描述中,为了说明而不是为了限定,提出了诸如特定系统结构、技术之类的具体细节,以便透彻理解本申请实施例。然而,本领域的技术人员应当清楚,在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本申请的描述。

[0024] 应当理解,当在本申请说明书和所附权利要求书中使用时,术语“包括”指示所描述特征、整体、步骤、操作、元素和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0025] 还应当理解,在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合,并且包括这些组合。

[0026] 如在本申请说明书和所附权利要求书中所使用的那样,术语“如果”可以依据上下文被解释为“当...时”或“一旦”或“响应于确定”或“响应于检测到”。类似地,短语“如果确

定”或“如果检测到[所描述条件或事件]”可以依据上下文被解释为意指“一旦确定”或“响应于确定”或“一旦检测到[所描述条件或事件]”或“响应于检测到[所描述条件或事件]”。

[0027] 另外,在本申请说明书和所附权利要求书的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0028] 在本申请说明书中描述的参考“一个实施例”或“一些实施例”等意味着在本申请的一个或多个实施例中包括结合该实施例描述的特定特征、结构或特点。由此,在本说明书中的不同之处出现的语句“在一个实施例中”、“在一些实施例中”、“在其他一些实施例中”、“在另外一些实施例中”等不是必然都参考相同的实施例,而是意味着“一个或多个但不是所有的实施例”,除非是以其他方式另外特别强调。术语“包括”、“包含”、“具有”及它们的变形都意味着“包括但不限于”,除非是以其他方式另外特别强调。

[0029] 本申请实施例中的服务器可以是掌上电脑、桌上型计算机、笔记本电脑、超级移动个人计算机(ultra-mobile personal computer,UMPC)、上网本、云端终端设备、个人数字助理(personal digital assistant,PDA)等,本申请实施例对服务器的具体类型不作任何限制。

[0030] 本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中,人工智能(Artificial Intelligence,AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0031] 人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、机器人技术、生物识别技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0032] 应理解,以下实施例中各步骤的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不对本申请实施例的实施过程构成任何限定。

[0033] 为了说明本申请的技术方案,下面通过具体实施例来进行说明。

[0034] 参见图1,是本申请实施例一提供的一种基于人工智能的对话服务方法的流程示意图,上述对话服务方法应用于服务器,服务器连接相应的用户端的终端设备,为用户端的终端设备配置相应的人机交互服务界面,以提供对话服务。该服务器连接相应的数据库,以获取相应的数据。如图1所示,该对话服务方法可以包括以下步骤:

[0035] 步骤S101,在用户触发对话服务时,获取用户的当前声音数据和当前面部数据。

[0036] 其中,用户在其终端设备上通过在服务器配置的人机交互服务界面中点击对话服务的按钮以触发对话服务,同时,用户端的终端设备上还配置有声音采集器和图像采集器,即麦克风和摄像头,使用声音采集器和图像采集器可以获取该触发对话服务的用户的语音数据和面部数据。上述终端设备将获取的声音数据和面部数据发给上述服务器,即认为上述服务器获取到了当前声音数据和当前面部数据。

[0037] 上述终端设备在向服务器发送声音数据和面部数据之前,可先询问触发对话服务的用户是否允许该终端设备向服务器发送该用户的语音数据和面部数据,经该用户允许方可向服务器提供用户的语音数据和面部数据,否则结束当次对话服务。

[0038] 若该用户不允许终端设备向服务器发送用户的语音数据和面部数据,则还可以输出放权提醒,以提醒用户需允许向终端设备服务器发送用户的语音数据和面部数据方可进行对话服务,或者若用户不允许终端设备向服务器发送用户的语音数据和面部数据,则还可以输出限制服务提醒,以提醒用户当前服务可能存在不准确的情况。例如,在上述终端设备上弹出“放权提醒”的对话框,用户可在对话框中进行选择操作,以实现放权或拒绝放权,或者在上述终端设备上弹出“限制提醒”的对话框,显示“当前对话服务,可能不准确”。

[0039] 在一种实施方式中,上述服务器连接相应的显示设备,并能够直接输出人机交互界面,因此,可供用户直接使用,其中,该服务器连接显示设备和输入设备,显示设备用于显示相应的人机交互服务界面,用户通过输入设备触发对话服务,上述服务器上还配置有声音采集器和图像采集器,即麦克风和摄像头,可直接获取触发对话服务的用户的语音数据和面部数据。

[0040] 步骤S102,对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对当前面部数据进行表情分析,确定对应的当前表情状态。

[0041] 本申请中,对声音数据的处理可以分为音频分析和语义分析,其中,音频分析可以是分析声音的音轨、音色、音量等,进而判断声音是否为尖锐、急促等情况,而语义分析可以是采用自然语言处理的方式对声音对应的文字进行分析,进而判断该文字中是否存在责怪、责问等语义,将音频分析的结果与语义分析的结果相结合,可以确定语气状态。语气状态可以是指用户出现负面情绪对应的状态或者用户未出现负面情绪对应的状态。例如,当声音数据中的声音比较急促和尖锐,同时声音数据对应的文字出现责怪和责问等,通过两者结合,可以确定该声音数据对应的语气状态为负面情绪对应的语气。

[0042] 基于声音的情绪识别可以分为两个大类,依据的是情绪的不同表示方式。第一种表示方式是情感的种类,最常用的六种基本情感包括开心(happiness)、难过(sadness)、生气(anger)、恶心(disgust)、害怕(fear)、惊讶(surprise);第二种表示方式是基于几个维度向量,最常用的是arousal和valence,arousal代表唤起程度的高低,valence代表积极情绪的高低,这两个维度都是通过数值来代表其自身的高低程度,例如,维度对应的数值区间为 $[-1,1]$,其中,-1可以代表非常低迷/消极,1可以代表非常激动/积极。通过对声音的情绪识别,可以将声音数据最终定性为不同情感种类,规定上述六种类型的情感中的一部分为不满的语气状态,或者通过对声音的情绪识别,可以将声音数据最终定性为得分,规定得分在一定阈值下对应为不满的语气状态。

[0043] 针对文字的情绪识别即是对文字的语义进行分析,本申请中可以采用自然语言处理(Natural Language Processing,NLP)分析文字数据进行语义分析,也即意见挖掘(Opinion Mining),以提取一种观点,该观点用于表征该文字的是否积极或者是否消极的表述。

[0044] 基于面部的情绪识别可以是指对表情进行分析,是基于人脸识别的图像分析,可以通过鉴别表情、微表情等并进行相应的表情相似度匹配,进而确定表情对应的状态,其中,表情对应状态可以是指用户出现负面情绪对应的表情或者用户未出现负面情绪对应的状态。

[0045] 表情分析可以是分为人脸识别和表情分类两个部分,其中,可以对获取的面部数据进行Opencv人脸识别,然后,再使用Keras进行表情分类管理,以及情绪识别,最终得到表

情状态。

[0046] 可选的是,对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态包括:

[0047] 检测当前声音数据中声音的振幅和频率;

[0048] 识别当前声音数据的语义,确定当前声音数据对应的态度;

[0049] 在振幅低于第一阈值或者频率低于第二阈值,且当前声音数据对应的态度为积极时,确定当前语气状态为第一目标状态;

[0050] 在所述当前声音数据对应的态度为消极和/或所述振幅不低于所述第一阈值且所述频率不低于所述第二阈值时,确定所述当前语气状态不为第一目标状态。

[0051] 其中,通过对声音数据的声波进行处理,来检测声音数据中声音的振幅和频率,振幅用于表征声音的响度大小,频率用于表征声音的音调高低。基于对响度和音调的分析,可以给出振幅对应的第一阈值和频率对应的第二阈值,第一阈值以上表征响度较大,第二阈值以上表征音调较高。

[0052] 识别声音数据的语义,确定该声音数据对应的态度,定义第一目标状态表征为用户未出现消极语气,因此,第一目标状态要求是振幅低于第一阈值或者频率低于第二阈值,且当前声音数据对应的态度为积极,其他情况均不是第一目标状态。

[0053] 可选的是,识别当前声音数据的语义,确定当前声音数据对应的态度包括:

[0054] 对当前声音数据进行降噪处理,得到降噪后的声音数据;

[0055] 从降噪后的声音数据中提取人声数据;

[0056] 将人声数据转换为文字数据,使用自然语言处理识别文字数据,确定文字数据对应的态度为当前声音数据对应的态度。

[0057] 其中,针对语义的识别,具体可以通过对声音数据对应的文字进行识别,采用上述NPL技术,但语义识别的准确度收到声音转文字的影响,因此,在对文字数据进行转换之前,需要对声音数据进行降噪处理,提取用户的声音进行转换。

[0058] 其中,语音文本转换技术可以用语音识别(Automatic Speech Recognition, ASR)技术对人声数据进行转换,该ASR技术是以语音为研究对象,通过语音信号处理和模式识别转换为相应文本。

[0059] 步骤S103,在当前语气状态为第一目标状态,且当前表情状态为第二目标状态时,提取当前声音数据中的问题语句。

[0060] 其中,第一目标状态表征所述用户未出现消极语气,即语气表达的是中性或者满意的,第二目标状态表征用户未出现消极表情,即表情表现的是中性或者满意的。

[0061] 当前语气状态为第一目标状态,且当前表情状态为第二目标状态时,表明用户当前情绪并未出现不满等情况,可以继续使用自动回复的方式应对用户提出的问题。

[0062] 当前语气状态不为第一目标状态或者当前表情状态不为第二目标状态时,表明用户当前情绪出现不满等情况,可能是由于自动回答无法准确解答用户问题,因此,不宜再使用自动回复的方式应对用户提出的问题,避免激化用户情绪。

[0063] 本申请中,提取当前声音数据中的问题语句,即是对声音数据进行分析,确定声音数据中的目标字段或关键字段,将该目标字段或者关键字段作为该用户在当前声音数据中所想要咨询的问题。在提取声音数据中的问题之前,还将上述声音数据转换为文字数,可采

用上述ASR技术。

[0064] 本申请中,基于提取模型对声音数据中问题语句进行提取,该提取模型重点关注疑问语气和疑问词,并基于此来对其中的问题语句进行提取。

[0065] 可选的是,在提取当前声音数据中的问题语句之前,还包括:

[0066] 若当前声音数据存在前一声音数据,则对前一声音数据进行音频分析和语义分析,确定对应的前一语气状态;

[0067] 获取训练好的提取模型;

[0068] 根据前一语气状态,调整训练好的提取模型,得到调整后的提取模型;

[0069] 相应地,提取当前声音数据中的问题语句包括:

[0070] 使用调整后的提取模型,提取当前声音数据中的问题语句。

[0071] 其中,在当前声音数据不是在一个对话服务中用户发出的首个声音数据时,该当前声音数据均会存在前一声音数据,而前一声音数据中的语气状态会与当前声音数据有一定程度的联系,因此,在对当前语音数据进行问题语句的提取之前,还可以将前一语气状态作为提取模型的输入,进而对提取模型中语气状态的参数进行调整,从而调整提取模型能够更加准确地提取当前声音数据中的问题语句。例如,前一声音数据的前一语气状态为不满时,当前声音数据中问题语句伴随的语气可能会加重,因此,将提取模型中语气的权重增加后,再执行问题语句的提取。

[0072] 步骤S104,将问题语句输入训练好的匹配模型,得到问题语句对应的答复语句。

[0073] 匹配模型可以采用机器学习、神经网络等,本申请中,匹配模型可以是基于双向长短期记忆网络(Bi-directional Long Short-Term Memory,BLSTM)算法的神经网络模型。

[0074] 其中,使用BLSTM算法构建的网络模型,首先,需要对用户的问题语句及其对应的候选答案句进行分词、去停顿词和词性标注等处理,然后,通过word2vec将所有的分词转换为词向量,进而得到问题语句对应的文本的特征表示,对于训练集中问题语句用 $Q = \{q_1, q_2, \dots, q_n\}$ 表示,答案定义为 $A = \{a_1, a_2, \dots, a_n\}$,具体步骤包括构建引入注意力机制的BLSTM模型,将问题语句Q与候选答案句A分别转化为特征向量 Q' 和 A' , Q' 和 A' 作为BLSTM模型的输入信息,然后在神经网络的隐层计算得到输出向量,再将BLSTM模型的输出信息注Attention模型对其加权计算,最后得到对应答案语句的向量表示。

[0075] 其中,使用word2vec训练词向量,将词向量的维度设置为150维,数据集来自网络搜索的WebQA数据集,分为训练集和测试集,采用两种评价标准对提出的问答进行验证:平均精度均值(Mean Average Precasion,MAP)和平均倒数排序(Mean Reciprocal Rank,MRR),其中,

$$[0076] \quad MAP = \frac{1}{n} \sum_{q=1}^n \text{avg}(P(q))$$

$$[0077] \quad MRR = \frac{1}{n} \sum_{q=1}^n \frac{1}{\text{rank}(q)}$$

[0078] 式中, $\text{avg}(P(q))$ 表示平均精度值, $\text{rank}(q)$ 表示候选答案中第一个正确答案的排名位置,MRR代表第一个正确答案的排名,而MAP仅测试所有正确答案的排名。

[0079] 在引入注意力机制的BLSTM模型中,对于每一个词向量都是同等对待的,注意力机制的思想是模仿人脑对不同事物的注意力不同的思维,通过自动加权的方式对整体信息进

行局部重点关注,注意力机制的概念是允许模型关注过去的输出。

[0080] 基于注意力机制的BLSTM模型将双向LSTM的每个输出结果送入注意力机制层,自动计算每个输出的结果,从而得到加权后的句子的向量表示;

[0081] $M = \tanh(H+B)$

[0082] $\alpha = \text{softmax}(W^T M)$

[0083] $\gamma = H\alpha^T$

[0084] 式中,H表示输出向量 $[h_1, h_2, \dots, h_T]$ 组成的矩阵,T代表输入语句的长度,B表示相对于H偏置向量,首先,利用 \tanh 函数将H的隐层表示为M,通过已知的上下文向量W和M,利用 softmax 函数可以计算出句子H的权重表征 α ,然后,再用输出向量矩阵H与权重矩阵 α 的转置的相乘,计算出该句子的向量表示 γ ,这里的向量W是对语句信息里比较重要部分的一种表示,它的值会在模型训练过程中随情况不同进行调整。

[0085] 上述提出的引入注意力机制的BLSTM模型相比LSTM的准确率取得了明显提升,该模型结合了注意力机制,忽略无意义的信息,获取了更多的上下文语义信息,因此上述自然语言处理算法中提出的模型的能显著提高答案的有效性。

[0086] 举例说明,针对环境服务平台,用户启动环境服务平台进行服务咨询,该环境服务平台通过相应的声音采集器和图像采集器采集用户的声音和面部图像,当声音对应的当前语气状态为正常状态(声音平缓、音调平缓等),且面部图像对应的当前表情状态为正常状态(微笑、未皱眉等)时,表明用户没有出现不满,因此,可以继续对话服务,对声音进行提取,确定问题为“A场景下的环保材料”,将问题输入训练好的BLSTM模型,可以得到对应的答复语句为:“为您找到了以下结果:Z工程下使用Y材料”。

[0087] 可选的是,在将问题语句输入训练好的匹配模型,得到问题语句对应的答复语句之后,还包括:

[0088] 从答复话术数据库中获取问题语句对应的答复话术;

[0089] 将答复语句与答复话术结合,得到问题语句对应的答复。

[0090] 其中,答复话术是在答复话术数据库中预先根据对话场景设置的话术,在该话术可以结合相应的答复语句,最终形成要答复的答复内容。例如,问题语句为:A场景下的环保材料都有哪些,据此调用对应的答复话术为:“为您找到以下信息:‘答复语句’”,答复语句为:Y材料、Z材料,因此,答复的内容为:为您找到以下信息:Y材料、Z材料。

[0091] 在一种实施方式中,如果需要采用语音答复,则将答复的内容按照预设的声音转化为语音,并输出。

[0092] 本申请实施例在用户触发对话服务时,获取用户的当前声音数据和当前面部数据,对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对当前面部数据进行表情分析,确定对应的当前表情状态,在当前语气状态为第一目标状态,且当前表情状态为第二目标状态时,提取当前声音数据中的问题语句,将问题语句输入训练好的匹配模型,得到问题语句对应的答复语句,将语气与表情相结合来识别用户意图,并在语气和表情对应的状态均满足条件时生成相应的答复语句,避免在用户出现不满等意图的情况下仍继续进行对话,从而可以准确把握用户意图,提升对话服务的服务质量。

[0093] 参见图2,是本申请实施例二提供的一种基于人工智能的对话服务方法的流程示意图,如图2所示,该对话服务方法可以包括以下步骤:

[0094] 步骤S201,在用户触发对话服务时,获取用户的当前声音数据和当前面部数据。

[0095] 步骤S202,对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对当前面部数据进行表情分析,确定对应的当前表情状态。

[0096] 其中,步骤S201至步骤S204与上述步骤S101至步骤S104的内容型相同,可参考步骤S101至步骤S104的描述,在此不再赘述。

[0097] 步骤S203,在当前语气状态为第三目标状态和/或当前表情状态为第四目标状态时,获取人工服务的人员状态信息。

[0098] 其中,第三目标状态表征用户出现消极语气,即语气表达的是不满意的,第四目标状态表征用户出现消极表情,即表情表现的是不满意的。

[0099] 第三目标状态是与第一目标状态相对立的,第四目标状态是与第二目标状态相对立的,当前语气状态为第三目标状态或者当前表情状态为第四目标状态时,均表明用户当前情绪出现了不满等情况,可能是由于自动回答无法准确解答用户问题,因此,不宜再使用自动回复的方式应对用户提出的问题,避免激化用户情绪。

[0100] 本申请中,在无法自动回答用户的问题时,还可以采用人工介入的方式,由人工回答相应的问题,从而提升用户的体验。基于此,上述服务器还可以与人工服务平台连接,从而获取人工服务平台中每个人工服务的人员的人员状态信息,包括是否在线、是否忙碌、是否能够解答相关场景等。

[0101] 在一种实施方式中,在获取人工服务的人员状态信息之后,如果当前暂无人员能够提供人工对话服务,则还向用户输出人工转接提醒,以提醒用户正在帮其转接人工。

[0102] 步骤S204,根据人员状态信息,确定目标人员,并将当前声音数据发送给目标人员。

[0103] 其中,根据人员状态信息,确定在线、且不忙碌、且能够解答对应场景问题的人员为目标人员,将当前声音数据经过人工服务平台发送至目标人员处,使得该目标人员能够浏览该当前声音数据,实现人工服务的调用。例如,在确定目标人员之后,向其发送接入请求,并在该目标人员确认后将数据发送给该目标人员。

[0104] 在将当前声音数据发送给目标人员之后还可以获取目标人员针对该当前声音数据做出的答复,并将答复发送给用户,即为用户提供人工对话服务。在一种实施方式中,将该目标人员与用户对接,使得目标人员与用户可以直接进行对话。例如,使用该对话服务界面将目标人员与用户对接,并可以使用语音和/或文字进行对话。

[0105] 可选的是,在将当前声音数据发送给目标人员之后,还包括:

[0106] 基于用户触发的对话服务,将目标人员与用户对接;

[0107] 在对接成功的情况下,获取目标人员发送的回复信息,将回复信息发送给用户。

[0108] 其中,服务器提供的对话服务能够与目标人员进行对接,即将对话服务的一端由机器人修改为对应目标人员,对话服务的另一端为用户,从而实现目标人员与用户的对接。

[0109] 本申请实施例在用户触发对话服务时,获取用户的当前声音数据和当前面部数据,对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对当前面部数据进行表情分析,确定对应的当前表情状态,在当前语气状态为第三目标状态或者当前表情状态为第四目标状态时,调用人工服务进行对话,避免由于自动回答无法准确解答用户问题引起用户不满,停止使用自动对话,避免激化用户情绪。

[0110] 对应于上文实施例的对话服务方法,图3示出了本申请实施例三提供的基于人工智能的对话服务装置的结构框图,上述对话服务装置应用于服务器,服务器连接相应的用户端的终端设备,为用户端的终端设备配置相应的人机交互服务界面,以提供对话服务。该服务器连接相应的数据库,以获取相应的数据。为了便于说明,仅示出了与本申请实施例相关的部分。

[0111] 参见图3,该对话服务装置包括:

[0112] 数据获取模块31,用于在用户触发对话服务时,获取用户的当前声音数据和当前面部数据;

[0113] 数据分析模块32,用于对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态,对当前面部数据进行表情分析,确定对应的当前表情状态,第一目标状态表征用户未出现消极语气,第二目标状态表征用户未出现消极表情;

[0114] 问题提取模块33,用于在当前语气状态为第一目标状态,且当前表情状态为第二目标状态时,提取当前声音数据中的问题语句;

[0115] 问题答复模块34,用于将问题语句输入训练好的匹配模型,得到问题语句对应的答复语句。

[0116] 可选的是,上述对话服务装置还包括:

[0117] 语气状态获取模块,用于在提取当前声音数据中的问题语句之前,若当前声音数据存在前一声音数据,则对前一声音数据进行音频分析和语义分析,确定对应的前一语气状态;

[0118] 提取模型获取模块,用于获取训练好的提取模型;

[0119] 提取模型调整模块,用于根据前一语气状态,调整训练好的提取模型,得到调整后的提取模型;

[0120] 相应地,上述问题提取模块33包括:

[0121] 问题提取单元,用于使用调整后的提取模型,提取当前声音数据中的问题语句。

[0122] 可选的是,上述对话服务装置还包括:

[0123] 状态信息获取模块,用于在对当前声音数据进行音频分析和语义分析,确定对应的当前语气状态之后,在当前语气状态为第三目标状态和/或当前表情状态为第四目标状态时,获取人工服务的人员状态信息,第三目标状态表征用户出现消极语气,第四目标状态表征用户出现消极表情;

[0124] 第一数据发送模块,用于根据人员状态信息,确定目标人员,并将当前声音数据发送给目标人员。

[0125] 可选的是,上述对话服务装置还包括:

[0126] 对接模块,用于在将当前声音数据发送给目标人员之后,基于用户触发的对话服务,将目标人员与用户对接;

[0127] 第二数据发送模块,用于在对接成功的情况下,获取目标人员发送的回复信息,将回复信息发送给用户。

[0128] 可选的是,上述对话服务装置还包括:

[0129] 话术获取模块,用于在将问题语句输入训练好的匹配模型,得到问题语句对应的答复语句之后,从答复话术数据库中获取问题语句对应的答复话术;

- [0130] 答复确定模块,用于将答复语句与答复话术结合,得到问题语句对应的答复。
- [0131] 可选的是,上述数据分析模块32包括:
- [0132] 检测单元,用于检测当前声音数据中声音的振幅和频率;
- [0133] 态度确定单元,用于识别当前声音数据的语义,确定当前声音数据对应的态度;
- [0134] 第一状态确定单元,用于在振幅低于第一阈值或者频率低于第二阈值,且当前声音数据对应的态度为积极时,确定当前语气状态为第一目标状态;
- [0135] 第二状态确定单元,用于在所述当前声音数据对应的态度为消极和/或所述振幅不低于所述第一阈值且所述频率不低于所述第二阈值时,确定所述当前语气状态不为第一目标状态。
- [0136] 可选的是,上述态度确定单元包括:
- [0137] 降噪处理单元,用于对当前声音数据进行降噪处理,得到降噪后的声音数据;
- [0138] 人声提取子单元,用于从降噪后的声音数据中提取人声数据;
- [0139] 态度确定子单元,用于将人声数据转换为文字数据,使用自然语言处理识别文字数据,确定文字数据对应的态度为当前声音数据对应的态度。
- [0140] 需要说明的是,上述模块之间的信息交互、执行过程等内容,由于与本申请方法实施例基于同一构思,其具体功能及带来的技术效果,具体可参见方法实施例部分,此处不再赘述。
- [0141] 图4为本申请实施例四提供的一种服务器的结构示意图。如图4所示,该实施例的服务器4包括:至少一个处理器40(图4中仅示出一个)、存储器41以及存储在存储器41中并可在至少一个处理器40上运行的计算机程序42,处理器40执行计算机程序42时实现上述任意各个对话服务方法实施例中的步骤。
- [0142] 该服务器4可包括,但不仅限于,处理器40、存储器41。本领域技术人员可以理解,图4仅仅是服务器4的举例,并不构成对服务器4的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如还可以包括输入输出设备、网络接入设备等。
- [0143] 所称处理器40可以是CPU,该处理器40还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。
- [0144] 存储器41在一些实施例中可以是服务器4的内部存储单元,例如服务器4的硬盘或内存。存储器41在另一些实施例中也可以是服务器4的外部存储设备,例如服务器4上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。进一步地,存储器41还可以既包括服务器4的内部存储单元也包括外部存储设备。存储器41用于存储操作系统、应用程序、引导装载程序(BootLoader)、数据以及其他程序等,例如计算机程序的程序代码等。存储器41还可以用于暂时地存储已经输出或者将要输出的数据。
- [0145] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将装置的内部结构划分成不同的功能单元或模块,以完成以上描述

的全部或者部分功能。实施例中的各功能单元、模块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,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述装置中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实现上述实施例方法中的全部或部分流程,可以通过计算机程序来指令相关的硬件来完成,计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述方法实施例的步骤。其中,计算机程序包括计算机程序代码,计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。计算机可读介质至少可以包括:能够携带计算机程序代码的任何实体或装置、记录介质、计算机存储器、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、电载波信号、电信信号以及软件分发介质。例如U盘、移动硬盘、磁碟或者光盘等。在某些司法管辖区,根据立法和专利实践,计算机可读介质不可以是电载波信号和电信信号。

[0146] 本申请实现上述实施例方法中的全部或部分流程,也可以通过一种计算机程序产品来完成,当计算机程序产品在服务器上运行时,使得服务器执行时实现可实现上述方法实施例中的步骤。

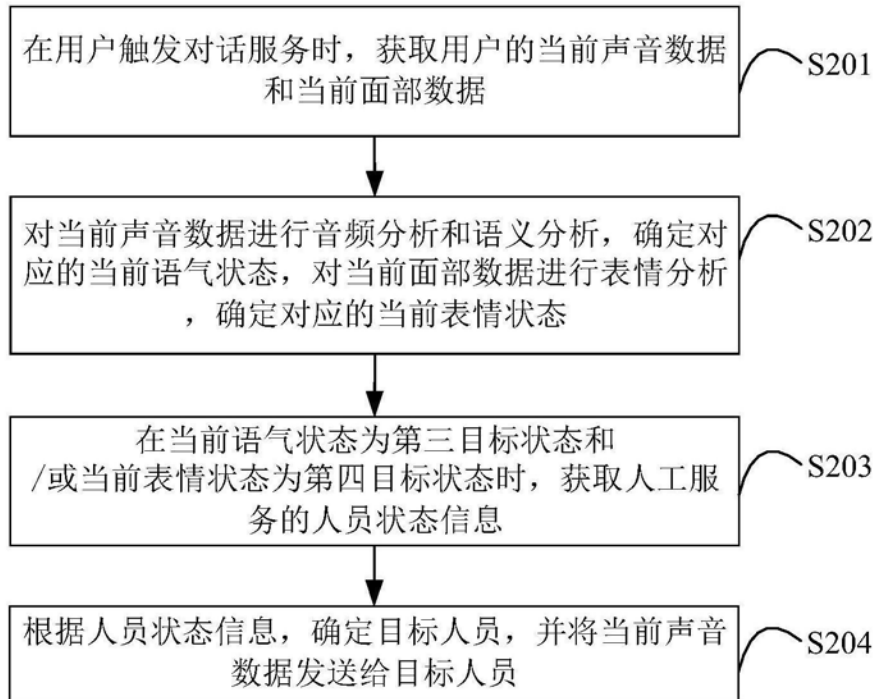
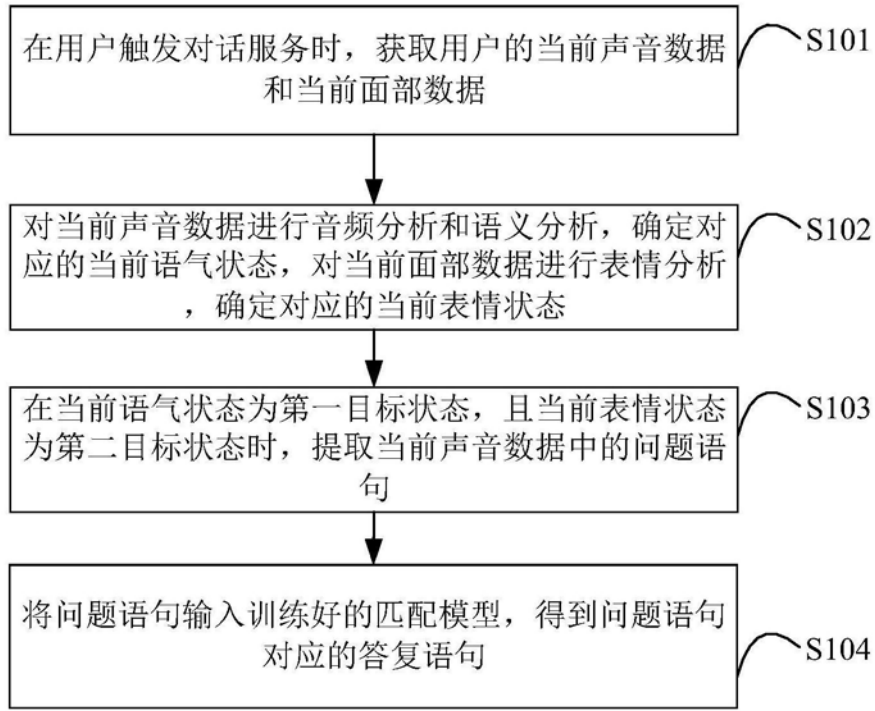
[0147] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述或记载的部分,可以参见其它实施例的相关描述。

[0148] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0149] 在本申请所提供的实施例中,应该理解到,所揭露的装置/服务器和方法,可以通过其它的方式实现。例如,以上所描述的装置/服务器实施例仅仅是示意性的,例如,模块或单元的划分,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,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以通过一些接口,装置或单元的间接耦合或通讯连接,可以是电性,机械或其它的形式。

[0150] 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。

[0151] 以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围,均应包含在本申请的保护范围之内。



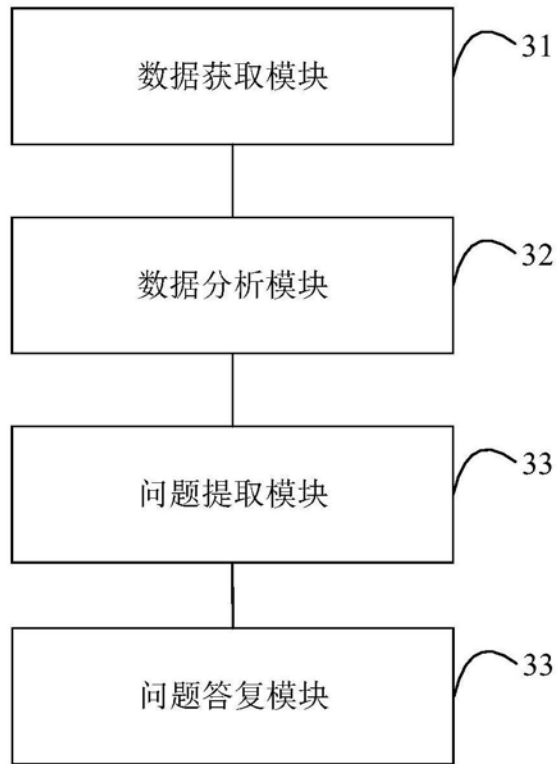


图3

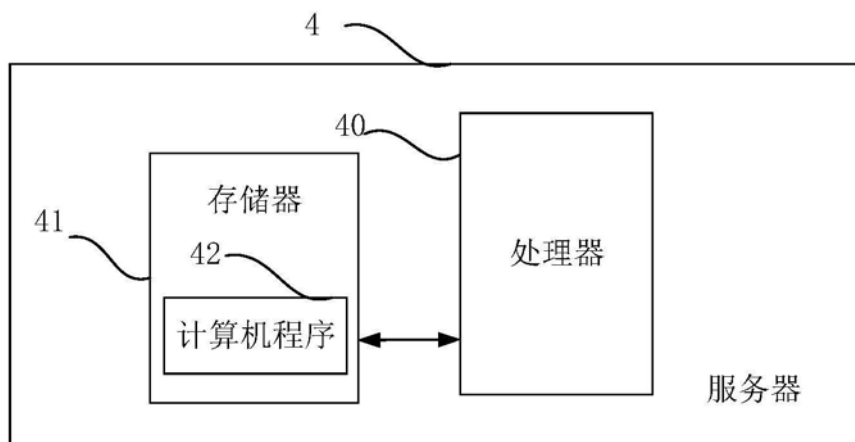


图4