



(12) 发明专利

(10) 授权公告号 CN 108363706 B

(45) 授权公告日 2023.07.18

(21) 申请号 201710056801.3

G10L 13/033 (2013.01)

(22) 申请日 2017.01.25

G10L 15/06 (2013.01)

(65) 同一申请的已公布的文献号

G10L 15/07 (2013.01)

申请公布号 CN 108363706 A

G10L 25/63 (2013.01)

(43) 申请公布日 2018.08.03

(56) 对比文件

CN 105280183 A, 2016.01.27

(73) 专利权人 北京搜狗科技发展有限公司

审查员 陈海霞

地址 100084 北京市海淀区中关村东路1号

院9号楼搜狐网络大厦9层01房间

(72) 发明人 赵海舟 许静芳

(74) 专利代理机构 深圳市深佳知识产权代理事

务所(普通合伙) 44285

专利代理师 王仲凯

(51) Int. Cl.

G06F 16/332 (2019.01)

G10L 15/22 (2006.01)

权利要求书4页 说明书12页 附图4页

(54) 发明名称

人机对话交互的方法和装置、用于人机对话交互的装置

(57) 摘要

本发明实施例提供了一种人机对话交互的方法和装置,其中所述方法包括:获取交互方的语音数据、图像数据、以及场景数据;依据所述场景数据获取对应的场景特征模型;将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;采用所述目标人物特征属性和场景数据确定目标对话策略;基于所述目标对话策略控制机器人的表情、语音和/或动作输出。本发明实施例使得在人机交互的过程中,机器可以根据目标对话策略配合交互方当前对话的特征,与交互方进行拟人化的对话,从而提高交互方交互体验。



1. 一种人机对话交互的方法,其特征在于,包括:
 - 获取交互方的语音数据、图像数据、以及场景数据;
 - 基于不同场景对应的人物特征属性设置对应的场景特征模型,依据所述场景数据获取对应的场景特征模型;
 - 将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;
 - 采用所述目标人物特征属性和场景数据确定目标对话策略,其中,基于不同的人物特征属性和场景特征设置不同的对话策略,所述目标对话策略为指导机器人在当前场景下采取与对话内容对应的动作、表情和语调的策略;
 - 基于所述目标对话策略控制机器人的表情、语音和/或动作输出;
 - 其中,所述场景特征模型通过如下方式训练:
 - 获取各个场景特征模型下的训练样本以及每个训练样本对应的人物特征属性;所述训练样本包括训练语音数据和训练图像数据,所述训练数据的构成方式包括:对各个场景下的不同人物进行语音采集和图像采集,并依据采集得到的人物的语音数据和图像数据建立与该人物的人物特征属性之间的对应关系;
 - 从所述训练语音数据中提取出训练语调特征数据和训练纹路特征数据;
 - 从所述训练图像数据中提取出训练表情特征数据和训练动作特征数据;
 - 采用每个场景特征下的训练样本包括的所述语调特征数据、训练纹路特征数据、训练表情特征数据和/或训练动作特征数据,以及对应的人物特征属性,训练得到各场景特征模型,所述人物特征属性包括年龄属性、性格属性以及情绪属性。
2. 根据权利要求1所述的方法,其特征在于,所述获取交互方的语音数据、图像数据、以及场景数据的步骤包括:
 - 基于麦克风采集所述交互方的语音数据;
 - 基于摄像头采集所述交互方的图像数据;
 - 以及,基于传感器采集所述场景数据。
3. 根据权利要求2所述的方法,其特征在于,所述获取交互方的语音数据、图像数据、以及场景数据的步骤包括:
 - 展示交互界面;
 - 基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。
4. 根据权利要求1所述的方法,其特征在于,所述依据所述场景数据获取对应的场景特征模型的步骤包括:
 - 从所述场景数据中提取出场景特征属性;
 - 获取所述场景特征属性对应的场景特征模型。
5. 根据权利要求1所述的方法,其特征在于,所述将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性,包括:
 - 从所述语音数据中提取出语调特征数据和纹路特征数据;
 - 从所述图像数据中提取出表情特征数据和动作特征数据;
 - 基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据,结合所述场景特征模型,得到所述目标人物特征属性。
6. 根据权利要求1-5任一所述的方法,其特征在于,所述基于所述目标对话策略控制机

机器人的表情、语音和/或动作输出,包括:

获取与所述目标对话策略相应的文字信息、表情指令、语音指令、和/或动作指令;
基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。

7. 一种人机对话交互的装置,其特征在于,包括:

交互方数据获取模块,用于获取交互方的语音数据、图像数据、以及场景数据;

场景特征模型获取模块,用于基于不同场景对应的人物特征属性设置对应的场景特征模型,依据所述场景数据获取对应的场景特征模型;

人物特征属性获得模块,用于将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

目标对话策略确定模块,用于采用所述目标人物特征属性和场景数据确定目标对话策略,其中,基于不同的人物特征属性和场景特征设置不同的对话策略,所述目标对话策略为指导机器人在当前场景下采取与对话内容对应的动作、表情和语调的策略;

人机交互对话模块,用于基于所述目标对话策略控制机器人的表情、语音和/或动作输出;

训练样本获取模块,用于获取各个场景特征模型下的训练样本以及每个训练样本对应的人物特征属性;所述训练样本包括训练语音数据和训练图像数据,所述训练数据的构成方式包括:对各个场景下的不同人物进行语音采集和图像采集,并依据采集得到的人物的语音数据和图像数据建立与该人物的人物特征属性之间的对应关系;

第一训练特征数据提取模块,用于从所述训练语音数据中提取出训练语调特征数据和训练纹路特征数据;

第二训练特征数据提取模块,用于从所述训练图像数据中提取出训练表情特征数据和训练动作特征数据;

场景特征模型训练模块,用于采用每个场景特征下的训练样本包括的所述语调特征数据、训练纹路特征数据、训练表情特征数据和/或训练动作特征数据,以及对应的人物特征属性,训练得到各场景特征模型,所述人物特征属性包括年龄属性、性格属性以及情绪属性。

8. 根据权利要求7所述的装置,其特征在于,所述交互方数据获取模块包括:

第一交互方数据采集子模块,用于基于麦克风采集所述交互方的语音数据;基于摄像头采集所述交互方的图像数据;以及,基于传感器采集所述场景数据。

9. 根据权利要求8所述的装置,其特征在于,所述交互方数据获取模块包括:

交互界面展示子模块,用于展示交互界面;

第二交互方数据采集子模块,用于基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。

10. 根据权利要求7所述的装置,其特征在于,所述场景特征模型获取模块包括:

场景特征属性提取子模块,用于从所述场景数据中提取出场景特征属性;

场景特征模型确定子模块,用于获取所述场景特征属性对应的场景特征模型。

11. 根据权利要求7所述的装置,其特征在于,所述人物特征属性获得模块包括:

第一特征数据提取子模块,用于从所述语音数据中提取出语调特征数据和纹路特征数据;

第二特征数据提取子模块,用于从所述图像数据中提取出表情特征数据和动作特征数据;

目标人物特征属性得到子模块,用于基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据,结合所述场景特征模型,得到所述目标人物特征属性。

12. 根据权利要求7-11任一所述的装置,其特征在于,所述目标对话策略设置有对应的文字、表情、动作,所述人机交互对话模块包括:

指令获取子模块,用于获取与所述目标对话策略相应的文字信息、表情指令、语音指令、和/或动作指令;

指令执行子模块,用于基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。

13. 一种用于人机对话交互的装置,其特征在于,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令:

获取交互方的语音数据、图像数据、以及场景数据;

基于不同场景对应的人物特征属性设置对应的场景特征模型,依据所述场景数据获取对应的场景特征模型;

将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

采用所述目标人物特征属性和场景数据确定目标对话策略,其中,基于不同的人物特征属性和场景特征设置不同的对话策略,所述目标对话策略为指导机器人在当前场景下采取与对话内容对应的动作、表情和语调的策略;

基于所述目标对话策略控制机器人的表情、语音和/或动作输出;

其中,所述场景特征模型通过如下方式训练:

获取各个场景特征模型下的训练样本以及每个训练样本对应的人物特征属性;所述训练样本包括训练语音数据和训练图像数据,所述训练数据的构成方式包括:对各个场景下的不同人物进行语音采集和图像采集,并依据采集得到的人物的语音数据和图像数据建立与该人物的人物特征属性之间的对应关系;

从所述训练语音数据中提取出训练语调特征数据和训练纹路特征数据;

从所述训练图像数据中提取出训练表情特征数据和训练动作特征数据;

采用每个场景特征下的训练样本包括的所述语调特征数据、训练纹路特征数据、训练表情特征数据和/或训练动作特征数据,以及对应的人物特征属性,训练得到各场景特征模型,所述训练数据的构成方式包括:对各个场景下的不同人物进行语音采集和图像采集,并依据采集得到的人物的语音数据和图像数据建立与该人物的人物特征属性之间的对应关系。

14. 根据权利要求13所述的用于人机对话交互的装置,其特征在于,所述获取交互方的语音数据、图像数据、以及场景数据的步骤包括:

基于麦克风采集所述交互方的语音数据;

基于摄像头采集所述交互方的图像数据;

以及,基于传感器采集所述场景数据。

15. 根据权利要求14所述的用于人机对话交互的装置,其特征在于,所述获取交互方的

语音数据、图像数据、以及场景数据的步骤包括：

展示交互界面；

基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。

16. 根据权利要求13所述的用于人机对话交互的装置，其特征在于，所述依据所述场景数据获取对应的场景特征模型的步骤包括：

从所述场景数据中提取出场景特征属性；

获取所述场景特征属性对应的场景特征模型。

17. 根据权利要求13所述的用于人机对话交互的装置，其特征在于，所述将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性，包括：

从所述语音数据中提取出语调特征数据和纹路特征数据；

从所述图像数据中提取出表情特征数据和动作特征数据；

基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据，结合所述场景特征模型，得到所述目标人物特征属性。

18. 根据权利要求13-17任一所述的用于人机对话交互的装置，其特征在于，所述基于所述目标对话策略控制机器人的表情、语音和/或动作输出，包括：

获取与所述目标对话策略相应的文字信息、表情指令、语音指令、和/或动作指令；

基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。

19. 一种可读存储介质，其特征在于，当所述存储介质中的指令由用于人机对话交互的装置的处理器的处理器执行时，使得用于人机对话交互的装置能够执行如方法权利要求1-6任一所述的人机对话交互的方法。

人机对话交互的方法和装置、用于人机对话交互的装置

技术领域

[0001] 本发明涉及数据处理技术领域,特别是涉及一种人机对话交互的方法和一种人机对话交互的装置、以及用户人机对话交互的装置。

背景技术

[0002] 人与机器的交互,就是人与机器进行信息交换的过程。信息交换效率是衡量人机对话交互最重要的指标。人与机器的交互,将追随人与人交互的演化路径,对话交互是最高效的人人交互方式,也将成为最高效的人机对话交互方式。

[0003] 现有的人机对话技术,只能将交互方的语音信息转成文字信息,不能识别更多的信息,这样造成机器在回复交互方时只能根据单一的参数、模型去产生回复。另外,现有的人机对话系统回复时通常是合成语音信息为主,对话形态单调,交互方体验效果不佳。

发明内容

[0004] 鉴于上述问题,提出了本发明实施例以便提供一种克服上述问题或者至少部分地解决上述问题的一种人机对话交互的方法和相应的一种人机对话交互的装置。

[0005] 为了解决上述问题,本发明实施例公开了一种人机对话交互的方法,包括:

[0006] 获取交互方的语音数据、图像数据、以及场景数据;

[0007] 依据所述场景数据获取对应的场景特征模型;

[0008] 将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

[0009] 采用所述目标人物特征属性和场景数据确定目标对话策略;

[0010] 基于所述目标对话策略控制机器人的表情、语音和/或动作输出。

[0011] 可选地,所述获取交互方的语音数据、图像数据、以及场景数据的步骤包括:

[0012] 基于麦克风采集所述交互方的语音数据;

[0013] 基于摄像头采集所述交互方的图像数据;

[0014] 以及,基于传感器采集所述场景数据。

[0015] 可选地,所述获取交互方的语音数据、图像数据、以及场景数据的步骤包括:

[0016] 展示交互界面;

[0017] 基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。

[0018] 可选地,所述依据所述场景数据获取对应的场景特征模型的步骤包括:

[0019] 从所述场景数据中提取出场景特征属性;

[0020] 获取所述场景特征属性对应的场景特征模型。

[0021] 可选地,所述场景特征模型通过如下方式训练:

[0022] 获取各个场景特征模型下的训练样本以及每个训练样本对应的人物特征属性;所述训练样本包括训练语音数据和训练图像数据;

[0023] 从所述训练语音数据中提取出训练语调特征数据和训练纹路特征数据;

[0024] 从所述训练图像数据中提取出训练表情特征数据和训练动作特征数据;

[0025] 采用每个场景特征下的训练样本包括的所述语调特征数据、训练纹路特征数据、训练表情特征数据和/或训练动作特征数据,以及对应的人物特征属性,训练得到各场景特征模型。

[0026] 可选地,所述将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性,包括:

[0027] 从所述语音数据中提取出语调特征数据和纹路特征数据;

[0028] 从所述图像数据中提取出表情特征数据和动作特征数据;

[0029] 基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据,结合所述场景特征模型,得到所述目标人物特征属性。

[0030] 可选地,所述基于所述目标对话策略控制机器人的表情、语音和/或动作输出,包括:

[0031] 获取与所述目标对话策略相应的文字信息、表情指令、语音指令、和/或动作指令;

[0032] 基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。

[0033] 本发明实施例还公开了一种人机对话交互的装置,包括:

[0034] 交互方数据获取模块,用于获取交互方的语音数据、图像数据、以及场景数据;

[0035] 场景特征模型获取模块,用于依据所述场景数据获取对应的场景特征模型;

[0036] 人物特征属性获得模块,用于将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

[0037] 目标对话策略确定模块,用于采用所述目标人物特征属性和场景数据确定目标对话策略;

[0038] 人机交互对话模块,用于基于所述目标对话策略控制机器人的表情、语音和/或动作输出。

[0039] 可选地,所述交互方数据获取模块包括:

[0040] 第一交互方数据采集子模块,用于基于麦克风采集所述交互方的语音数据;基于摄像头采集所述交互方的图像数据;以及,基于传感器采集所述场景数据。

[0041] 可选地,所述交互方数据获取模块包括:

[0042] 交互界面展示子模块,用于展示交互界面;

[0043] 第二交互方数据采集子模块,用于基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。

[0044] 可选地,所述场景特征模型获取模块包括:

[0045] 场景特征属性提取子模块,用于从所述场景数据中提取出场景特征属性;

[0046] 场景特征模型确定子模块,用于获取所述场景特征属性对应的场景特征模型。

[0047] 可选地,所述装置还包括:

[0048] 训练样本获取模块,用于获取各个场景特征模型下的训练样本以及每个训练样本对应的人物特征属性;所述训练样本包括训练语音数据和训练图像数据;

[0049] 第一训练特征数据提取模块,用于从所述训练语音数据中提取出训练语调特征数据和训练纹路特征数据;

[0050] 第二训练特征数据提取模块,用于从所述训练图像数据中提取出训练表情特征数据和训练动作特征数据;

[0051] 场景特征模型训练模块,用于采用每个场景特征下的训练样本包括的所述语调特征数据、训练纹路特征数据、训练表情特征数据和/或训练动作特征数据,以及对应的人物特征属性,训练得到各场景特征模型。

[0052] 可选地,所述人物特征属性获得模块包括:

[0053] 第一特征数据提取子模块,用于从所述语音数据中提取出语调特征数据和纹路特征数据;

[0054] 第二特征数据提取子模块,用于从所述图像数据中提取出表情特征数据和动作特征数据;

[0055] 目标人物特征属性得到子模块,用于基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据,结合所述场景特征模型,得到所述目标人物特征属性。

[0056] 可选地,所述目标对话策略设置有对应的文字、表情、动作,所述人机交互对话模块包括:

[0057] 指令获取子模块,用于获取与所述目标对话策略相应的文字信息、表情指令、语音指令、和/或动作指令;

[0058] 指令执行子模块,用于基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。

[0059] 本发明实施例还提供一种用于人机对话交互的装置,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令:

[0060] 获取交互方的语音数据、图像数据、以及场景数据;

[0061] 依据所述场景数据获取对应的场景特征模型;

[0062] 将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

[0063] 采用所述目标人物特征属性和场景数据确定目标对话策略;

[0064] 基于所述目标对话策略控制机器人的表情、语音和/或动作输出。

[0065] 本发明实施例包括以下优点:

[0066] 本发明实施例在人机交互的过程中,获取交互方的语音数据、图像数据和场景数据,并基于场景数据获取符合当下对话场景的场景特征模型,然后将交互方的语音数据和/或图像数据输入到场景特征模型得到人物特征属性,并基于该人物特征属性制定相应的目标对话策略,使得在人机交互的过程中,机器可以根据目标对话策略配合交互方当前对话的特征,与交互方进行拟人化的对话。

[0067] 本发明实施例可以依据采集得到的场景数据选择符合相应场景下的场景特征模型,依据该场景特征模型确定与当前交互方的语音特征和/图像特征相匹配的人物特征属性,其中,人物特征属性可以反映说话方的意图、情绪,在不同的场景下不用人物特征属性的说话方的语音特征和图像特征会略有不同。因此,选择与当前场景数据对应的场景特征模型来确定人物特征属性,可以使得说话方的意图和情绪表达更加准确。本发明实施例进一步可以根据人物特征属性来制定与交互方交互的目标对话策略,可以使得机器与交互方的交互更加个性化,能够更好地服务交互方。

附图说明

- [0068] 图1是本发明的一种人机对话交互的方法实施例一的步骤流程图；
- [0069] 图2是本发明的一种人机对话交互的方法实施例二的步骤流程图；
- [0070] 图3是本发明的一种人机对话交互的装置实施例的结构框图；
- [0071] 图4是根据一示例性实施例示出的一种人机交互的装置的框图；
- [0072] 图5是根据一示例性实施例示出的一种用于人机对话交互的装置作为服务器时的框图。

具体实施方式

[0073] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0074] 参照图1，示出了本发明的一种人机对话交互的方法实施例一的步骤流程图，具体可以包括如下步骤：

[0075] 步骤101，获取交互方的语音数据和图像数据、以及场景数据；

[0076] 在交互方与机器人进行对话交互的过程中，实时获取交互方的关联数据。其中，该关联数据可以包括交互方的语音数据、图像数据，以及，对话所在交互场景的场景数据等等。其中，该关联数据可以通过数据采集设备获取，也可以由人工输入。

[0077] 在本发明实施例中，上述关联数据可以是主动采集的。即为，一旦确定当前机器人处于人机交互状态，则自动通过机器人内置或外部连接的数据采集设备对当前交互的对象（即为交互方）进行语音数据和/图像数据、以及场景数据的采集。

[0078] 具体地，所述步骤101可以包括如下子步骤：

[0079] 子步骤S11，基于麦克风采集交互方的语音数据；基于摄像头采集交互方的图像数据；以及基于传感器采集当前对话的场景数据。

[0080] 与交互方进行交互的机器人上可以安装有多种数据采集设备，包括机器人内部安装以及外部连接。基于不同的数据采集设备可以采集到不同的交互方关联数据，例如交互方说话时的语音数据，交互方的面部表情数据，交互方的手势动作数据，交互方当前所处的场景（环境）数据等等。

[0081] 在本发明的一种优选实施例中，所述步骤101可以包括如下子步骤：

[0082] 子步骤S12，展示交互界面；

[0083] 子步骤S13，基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。可选地，本发明实施例还可以通过询问交互方或通过交互界面引导交互方输入该关联数据，具体地，可以为交互方展示交互界面，并在该交互界面上逐项提示交互方输入相应的数据，例如提示对准麦克风输入语音数据，对准相机输入图像数据，根据交互方当前所在场景确定场景数据。进一步地，还可以提示交互方输入诸如包括年龄、性别信息等数据。

[0084] 当然，在实际中，通过交互方输入的方式相对机器人自行采集的方式较为不及时，并且所能获取的数据较少，无法应对交互方的变化，因此在人机交互过程中，应当以机器自动采集的数据为主，交互方输入的数据为辅。

[0085] 需要说明的是，本发明实施例还可以通过可穿戴式设备来采集交互方的人体生理数据，例如心跳、呼吸、消化、体温等等，基于这些数据，机器人能够分析和识别处理，从而判

断出交互方的情绪特征。

[0086] 步骤102,依据所述场景数据获取对应的场景特征模型;

[0087] 在本发明实施例中,考虑到在不同的场景下,交互方的人物特征属性会有所不同,使得采集得到的交互方的关联数据也略有差异。例如在室内人物表现可能会较为拘谨,在室外则可能会较为放开,在行车时则可能会较为沉闷,因此,本发明实施例根据不同的场景对应的人物特征属性,设置对应的场景特征模型。例如,在室内可以设置对应的与室内环境较为匹配的人物特征属性对应的室内特征模型,在室外可以设置对应的室外人物特征属性对应的室外特征模型,在行车可以设置与行车时环境较为匹配的人物特征属性对应的行车特征模型。

[0088] 在本发明的一种优选实施例中,所述202可以包括如下子步骤:

[0089] 子步骤S21,从所述场景数据中提取出场景特征属性;

[0090] 子步骤S22,获取所述场景特征属性对应的场景特征模型。

[0091] 本发明实施例可以通过传感器采集人机对话发生场景的具体环境信息,即采集到场景数据。具体来说,可以通过温湿度传感器识别温度、湿度,通过速度传感器识别是在移动还是静止,通过光传感器识别是白天还是晚上,通过环境探测器识别是室内还是室外等。对于传感器识别到的数据,将从中提取出相应的场景特征属性,基于这些场景特征属性,可以简单分析出是否在室内、是否在行车等等。例如,速度传感器识别出的速度数据,可以从中提取到交互方当前的速度信息,如果速度达到预设的车辆行驶速度,就可以认为交互方处于行车场景下。

[0092] 当然,在具体实施本发明实施例时,还可以利用其他传感器或者其他方式来识别出交互方当前所处的场景,本发明实施例对此不加以限制。

[0093] 在本发明的一种优选实施例中,所述场景特征模型可以通过如下方式进行训练:

[0094] 获取各个场景特征下的训练样本以及每个训练样本对应的人物特征属性;所述训练样本包括训练语音数据和训练图像数据;

[0095] 从所述训练语音数据中提取出训练语调特征数据和训练纹路特征数据;

[0096] 从所述训练图像数据中提取出训练表情特征数据和训练动作特征数据;

[0097] 采用每个场景特征下的训练样本包括的所述语调特征数据、训练纹路特征数据、训练表情特征数据和/或训练动作特征数据,以及对应的人物特征属性,训练得到各场景特征模型。

[0098] 本发明实施例可以利用预先收集的各个场景特征下的海量训练样本以及训练样本对应的人物特征属性作为训练数据。训练样本可以包括训练语音数据和训练图像数据。例如,对某个场景下的不同人物进行语音采集和图像采集,并依据采集得到的某一人物的语音数据和图像数据建立与该人物的人物特征属性之间的对应关系,构成一训练数据。例如,对某个场景下,获取一小孩的语音数据和图像数据,并从该小孩的语音数据中提取出小孩的语调特征数据和纹路特征数据,并从小孩的图像数据中提取出小孩的面部表情特征数据和手势特征数据,将该部分数据作为训练样本,建立与小孩的人物特征属性、以及当前场景特征的关联,作为一训练样本,保存在训练样本库中。基于此,使得训练样本库中保存有海量的针对不同场景的训练样本。

[0099] 通过深度学习,依据各场景特征下海量的训练样本,可以训练得到针对各场景下

的场景特征模型。其中,训练样本对应的人物特征属性可以包括年龄属性(小孩/青年/老人)、性格属性(开朗/内敛/腼腆)以及情绪属性(悲伤/平静/兴奋)等。训练完成后可以对语音数据、图像数据做出相应的分类。

[0100] 场景特征模型可以基于输入的语音数据、图像数据确定对应的人物特征属性。

[0101] 在本发明的具体应用中,可以通过训练语调特征数据来训练交互方的情绪的场景特征模型,可以通过训练特征数据来训练年龄的场景特征模型,进一步地,还可以添加其他数据来训练,使得模型精度更加高。

[0102] 步骤103,将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性。

[0103] 具体的,本发明实施例步骤103可以包括:

[0104] 从所述语音数据中提取出语调特征数据和纹路特征数据;

[0105] 从所述图像数据中提取出表情特征数据和动作特征数据;

[0106] 基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据,结合所述场景特征模型,得到所述目标人物特征属性。

[0107] 对于采集到的交互方关联数据,例如语音数据和/或图像数据将会输入到对应的场景特征模型,基于该场景特征模型得到与当前场景和对应交互方特征比较匹配的人物特征属性。具体来说,人物特征属性可以包括基本特征、情绪特征、性格特征等等,其中,情绪特征可以包括激动、平静、悲伤,基本特征可以包括老人、儿童、男性、女性等,性格特征可以包括开朗、活泼、内敛、腼腆等。

[0108] 需要说明的是,对于人物特征属性的划分和判断,可以根据实际需求设置上述一种或多种,或者是添加其他的人物特征属性,本发明实施例对此无需加以限制。

[0109] 步骤104,采用所述目标人物特征属性和场景数据确定目标对话策略;

[0110] 本发明实施例基于不同的人物特征属性和场景特征设置不同的对话策略,例如,如果识别出当前场景为比较压抑的氛围,交互方的情绪是悲伤的,可以在对话过程中采用抚慰的对话策略。

[0111] 步骤105,基于所述目标对话策略控制机器人的表情、语音和/或动作输出。

[0112] 在人机对话过程中,可以基于目标对话策略的控制机器人的表情、语音、和/或动作输出,与交互方进行对话。其中,表情可以为面部表情,例如眼睛的表现特征、嘴巴的表现特征等;语音可以为机器人语音输出的语调、声音的高低等;动作可以为机器人的手势、头部动作、以及其他肢体的动作等。

[0113] 本发明实施例中可以通过声纹识别采集语音数据中的纹路特征数据。声纹识别,也称为说话人识别,有两类,即说话人辨认和说话人确认。前者用以判断某段语音是若干人中的哪一个所说的,是“多选一”问题,而后者用以确认某段语音是否是指定的某个人所说的。

[0114] 本发明实施例可以利用纹路特征数据作为交互方的身份标识,当识别到有两个或者两个以上的交互方进行人机对话时,可以基于纹路特征数据判断是哪个交互方产生的语音数据,并基于该交互方产生的语音数据和/或图像数来确定的针对该交互方的目标对话策略,与该交互方进行对话,而不是对于两个交互方多个交互方使用相同的对话策略,由此可以满足交互方的个性化需求,使得对话更加有趣。

[0115] 本发明实施例在人机交互的过程中,获取交互方的语音数据和图像数据、以及场景数据,并基于场景数据获取符合当下场景的场景特征模型,然后将交互方的语音数据和图像数据输入到场景特征模型得到与当前场景和交互方特征相匹配的人物特征属性,并基于该人物特征属性制定相应的目标对话策略,控制机器人的表情和动作输出,使得交互方在人机交互的过程中,机器人可以根据目标对话策略配合与交互方进行对话。

[0116] 本发明实施例可以选择符合相应场景下的场景特征模型,来确定当前场景下机器人需要表现的人物特征属性,其中,人物特征属性能够反映人类的性格、意图、情绪等,在不同的场景下不同人物特征属性的人类产生的语音数据和图像数据会略有不同,因此,通过相应场景下的场景特征模型来确定当前场景下机器人需要表现出的人物特征属性,可以使得机器人的拟人化更加准确。本发明实施例进一步根据获取的目标人物特征属性来控制机器人的表情和动作输出,可以使得机器与人类的交互更加个性化,能够更好地服务用户。

[0117] 参照图2,示出了本发明的一种人机对话交互的方法实施例二的步骤流程图,具体可以包括如下步骤:

[0118] 步骤201,获取交互方的语音数据、图像数据、以及场景数据;

[0119] 步骤202,依据所述场景数据获取对应的场景特征模型;

[0120] 步骤203,将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

[0121] 步骤204,采用所述目标人物特征属性和场景数据确定目标对话策略;

[0122] 步骤205,获取与所述目标对话策略相应的文字信息、表情指令、语音指令和/或动作指令;

[0123] 由于方法实施例二中的步骤201-步骤205的具体实施方式基本相应于前述的方法实施例一的具体实施方式,故本实施例对于步骤201-步骤205的描述中未详尽之处,可以参见前述实施例一中的相关说明,在此就不赘述了。

[0124] 步骤206,基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。

[0125] 在本发明实施例中,基于人物特征属性确定目标对话策略,目标对话策略中设置有相应的文字信息、表情指令、语音指令以及动作指令,用以指导机器人在当前场景下如何通过一定的对话和表情、动作来表现该人物特征属性。

[0126] 例如,如果识别出交互方的情绪是悲伤的,则可能依据模型得到的人物特征属性对应的目标对话策略为安抚型,以及该目标对话策略获取相应的文字信息、表情指令、语音指令和/或动作指令,例如,文字信息为安慰性话语,表情指令为悲伤、面部柔和等,语音指令为声音较低、语调轻柔等,动作指令为抚摸交互方头部等,并基于该表情指令、语音指令和/或动作指令控制机器人输出所述文字信息。

[0127] 由此可以使得人机交互更加贴近交互方当前的实际情况,提升用户体验效果。

[0128] 现有的技术只能将交互方的语音转成文字,不能识别更多的信息,这样造成机器人在对话时只能根据单一的参数、模型去产生回复。

[0129] 另外现有的机器人对话时通常是合成语音为主,而且缺乏针对不同输入的语调变化,本发明实施例结合语音、语调、表情、动作,提升了交互体验。也即是说,本发明实施例实现了多模输入和多模输出的人机交互方法,使得机器的对话不再单一。

[0130] 综上可知,本发明实施例将当前所处场景和面对的人物对象的状态(语音数据、图像数据、以及场景数据)等作为参考,获取机器人在当前场景下需要表现的表情和动作等多模态输出,实现机器的多模态人机交互。即,本发明实施例提出一种综合利用语音识别、情感识别、人脸识别、场景识别等技术构成多模输入;使得机器人语音配合表情、动作构成多模输出,来提升对话系统体验的。其中,本发明实施例的实现过程可以分为两种:

[0131] 1、离线过程

[0132] 离线过程即之前提到的收集数据、训练的过程,本发明实施例依据各个场景中人物特征属性以及对话内容等信息,针对不同类型的人物,在不同场景下,发生的表情、动作等进行统计分析,建立场景特征模型。并且,设置在各个场景和人物特征属性对应的对话策略。例如,在某种场景下,与当前聊天内容对应的动作、表情、语调等;或者是针对老人、小孩,在某种场景下,与当前聊天内容对应的动作、表情、语调等。

[0133] 2、在线过程

[0134] 依据当前场景数据,以及交互方的聊天内容和图像数据等等,对当前情景进行判断,确定场景特征模型,基于场景特征模型确定该场景下的交互方特征数据并确定目标对话策略,然后基于目标对话策略,获取与当前场景以及聊天内容相匹配的表情、动作等,实现机器人的多模态输出。

[0135] 需要说明的是,对于方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明实施例并不受所描述的动作顺序的限制,因为依据本发明实施例,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作并不一定是本发明实施例所必须的。

[0136] 参照图3,示出了本发明的一种人机对话交互的装置实施例的结构框图,具体可以包括如下模块:

[0137] 交互方数据获取模块301,用于获取交互方的语音数据、图像数据、以及场景数据;

[0138] 场景特征模型获取模块302,用于依据所述场景数据获取对应的场景特征模型;

[0139] 人物特征属性获得模块303,用于将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

[0140] 目标对话策略确定模块304,用于采用所述目标人物特征属性和场景数据确定目标对话策略;

[0141] 人机交互对话模块305,用于基于所述目标对话策略控制机器人的表情、语音和/或动作输出。

[0142] 在本发明的一种优选实施例中,所述交互方数据获取模块301可以包括:

[0143] 第一交互方数据采集子模块,用于基于麦克风采集所述交互方的语音数据;基于摄像头采集所述交互方的图像数据;以及,基于传感器采集所述场景数据。

[0144] 在本发明的一种优选实施例中,所述交互方数据获取模块301可以包括:

[0145] 交互界面展示子模块,用于展示交互界面;

[0146] 第二交互方数据采集子模块,用于基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。

[0147] 在本发明的一种优选实施例中,所述场景特征模型获取模块302可以包括:

- [0148] 场景特征属性提取子模块,用于从所述场景数据中提取出场景特征属性;
- [0149] 场景特征模型确定子模块,用于获取所述场景特征属性对应的场景特征模型。
- [0150] 在本发明的一种优选实施例中,所述装置还可以包括:
- [0151] 场景特征属性提取子模块,用于从所述场景数据中提取出场景特征属性;
- [0152] 场景特征模型确定子模块,用于获取所述场景特征属性对应的场景特征模型。
- [0153] 在本发明的一种优选实施例中,所述人物特征属性获得模块303可以包括:
- [0154] 第一特征数据提取子模块,用于从所述语音数据中提取出语调特征数据和纹路特征数据;
- [0155] 第二特征数据提取子模块,用于从所述图像数据中提取出表情特征数据和动作特征数据;
- [0156] 目标人物特征属性得到子模块,用于基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据,结合所述场景特征模型,得到所述目标人物特征属性。
- [0157] 在本发明的一种优选实施例中,所述目标对话策略设置有对应的文字、表情、动作,所述人机交互对话模块305可以包括:
- [0158] 指令获取子模块,用于获取与所述目标对话策略相应的文字信息、表情指令、语音指令、和/或动作指令;
- [0159] 指令执行子模块,用于基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。
- [0160] 对于装置实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。
- [0161] 图4是根据一示例性实施例示出的一种人机交互的装置500的框图。例如,装置500可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理等。
- [0162] 参照图4,装置500可以包括以下一个或多个组件:处理组件502,存储器504,电源组件506,多媒体组件508,音频组件510,输入/输出(I/O)的接口512,传感器组件514,以及通信组件516。
- [0163] 处理组件502通常控制装置500的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理元件502可以包括一个或多个处理器520来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件502可以包括一个或多个模块,便于处理组件502和其他组件之间的交互。例如,处理部件502可以包括多媒体模块,以方便多媒体组件508和处理组件502之间的交互。
- [0164] 存储器504被配置为存储各种类型的数据以支持在装置500的操作。这些数据的示例包括用于在装置500上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器504可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。
- [0165] 电源组件506为装置500的各种组件提供电力。电源组件506可以包括电源管理系统,一个或多个电源,及其他与为装置500生成、管理和分配电力相关联的组件。

[0166] 多媒体组件508包括在所述装置500和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件508包括一个前置摄像头和/或后置摄像头。当设备500处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0167] 音频组件510被配置为输出和/或输入音频信号。例如,音频组件510包括一个麦克风(MIC),当装置500处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器504或经由通信组件516发送。在一些实施例中,音频组件510还包括一个扬声器,用于输出音频信号。

[0168] I/O接口512为处理组件502和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0169] 传感器组件514包括一个或多个传感器,用于为装置500提供各个方面的状态评估。例如,传感器组件514可以检测到设备500的打开/关闭状态,组件的相对定位,例如所述组件为装置500的显示器和小键盘,传感器组件514还可以检测装置500或装置500一个组件的位置改变,用户与装置500接触的存在或不存在,装置500方位或加速/减速和装置500的温度变化。传感器组件514可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件514还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件514还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0170] 通信组件516被配置为便于装置500和其他设备之间有线或无线方式的通信。装置500可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个示例性实施例中,通信部件514经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信部件514还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RFID)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0171] 在示例性实施例中,装置500可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述方法。

[0172] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器504,上述指令可由装置500的处理器520执行以完成上述方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0173] 图5是根据一示例性实施例示出的一种用于人机对话交互的装置作为服务器时的框图。该服务器1900可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上中央处理器(central processing units,CPU)1922(例如,一个或一个以上处理器)和存储

器1932,一个或一个以上存储应用程序1942或数据1944的存储介质1930(例如一个或一个以上海量存储设备)。其中,存储器1932和存储介质1930可以是短暂存储或持久存储。存储在存储介质1930的程序可以包括一个或一个以上模块(图示没标出),每个模块可以包括对服务器中的一系列指令操作。更进一步地,中央处理器1922可以设置为与存储介质1930通信,在服务器1900上执行存储介质1930中的一系列指令操作。

[0174] 服务器1900还可以包括一个或一个以上电源1926,一个或一个以上有线或无线网络接口1950,一个或一个以上输入输出接口1958,一个或一个以上键盘1956,和/或,一个或一个以上操作系统1941,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™等等。

[0175] 一种非临时性计算机可读存储介质,当所述存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行一种人机对话交互的方法,所述方法包括:

[0176] 获取交互方的语音数据、图像数据、以及场景数据;

[0177] 依据所述场景数据获取对应的场景特征模型;

[0178] 将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性;

[0179] 采用所述目标人物特征属性和场景数据确定目标对话策略;

[0180] 基于所述目标对话策略控制机器人的表情、语音和/或动作输出。

[0181] 可选地,所述获取交互方的语音数据、图像数据、以及场景数据的步骤包括:

[0182] 基于麦克风采集所述交互方的语音数据;

[0183] 基于摄像头采集所述交互方的图像数据;

[0184] 以及,基于传感器采集所述场景数据。

[0185] 可选地,所述获取交互方的语音数据、图像数据、以及场景数据的步骤包括:

[0186] 展示交互界面;

[0187] 基于所述交互界面提示交互方输入语音数据、图像数据、以及场景数据。

[0188] 可选地,所述依据所述场景数据获取对应的场景特征模型的步骤包括:

[0189] 从所述场景数据中提取出场景特征属性;

[0190] 获取所述场景特征属性对应的场景特征模型。

[0191] 可选地,所述场景特征模型通过如下方式训练:

[0192] 获取各个场景特征模型下的训练样本以及每个训练样本对应的人物特征属性;所述训练样本包括训练语音数据和训练图像数据;

[0193] 从所述训练语音数据中提取出训练语调特征数据和训练纹路特征数据;

[0194] 从所述训练图像数据中提取出训练表情特征数据和训练动作特征数据;

[0195] 采用每个场景特征下的训练样本包括的所述语调特征数据、训练纹路特征数据、训练表情特征数据和/或训练动作特征数据,以及对应的人物特征属性,训练得到各场景特征模型。

[0196] 可选地,所述将所述语音数据和图像数据输入至所述场景特征模型得到目标人物特征属性,包括:

[0197] 从所述语音数据中提取出语调特征数据和纹路特征数据;

[0198] 从所述图像数据中提取出表情特征数据和动作特征数据;

[0199] 基于所述语调特征数据、纹路特征数据、表情特征数据和/或动作特征数据,结合

所述场景特征模型,得到所述目标人物特征属性。

[0200] 可选地,所述基于所述目标对话策略控制机器人的表情、语音和/或动作输出,包括:

[0201] 获取与所述目标对话策略相应的文字信息、表情指令、语音指令、和/或动作指令;

[0202] 基于所述表情指令、语音指令和/或动作指令控制所述机器人输出所述文字信息。

[0203] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发明的其它实施方案。本发明旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本发明的真正范围和精神由下面的权利要求指出。

[0204] 应当理解的是,本发明并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

[0205] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

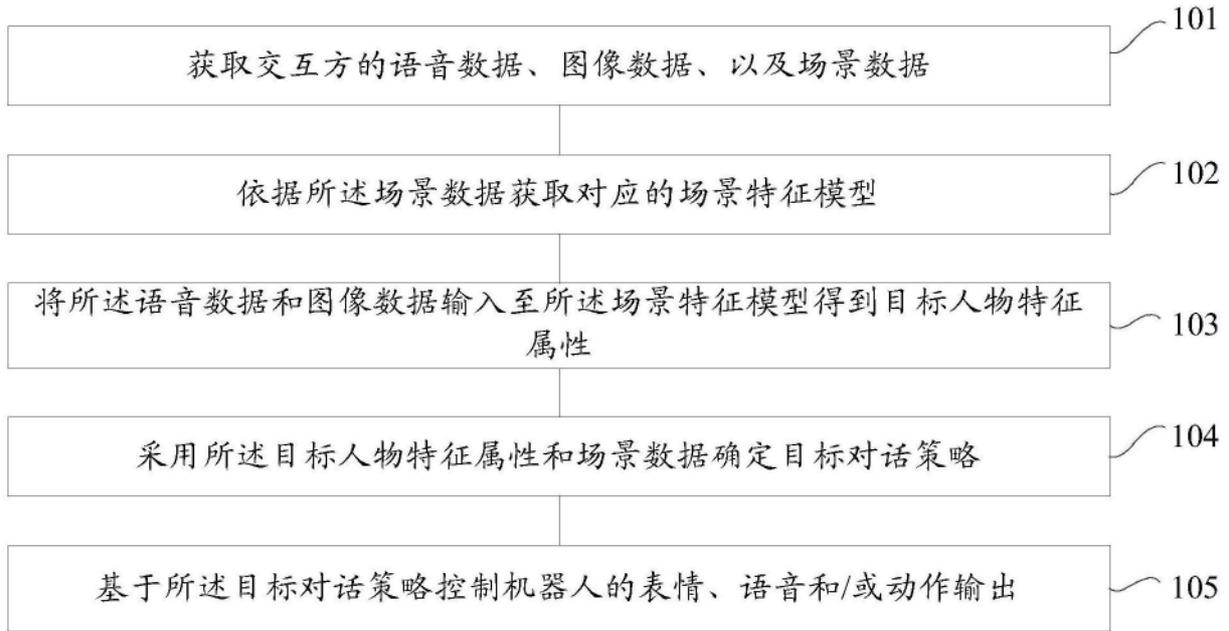


图1

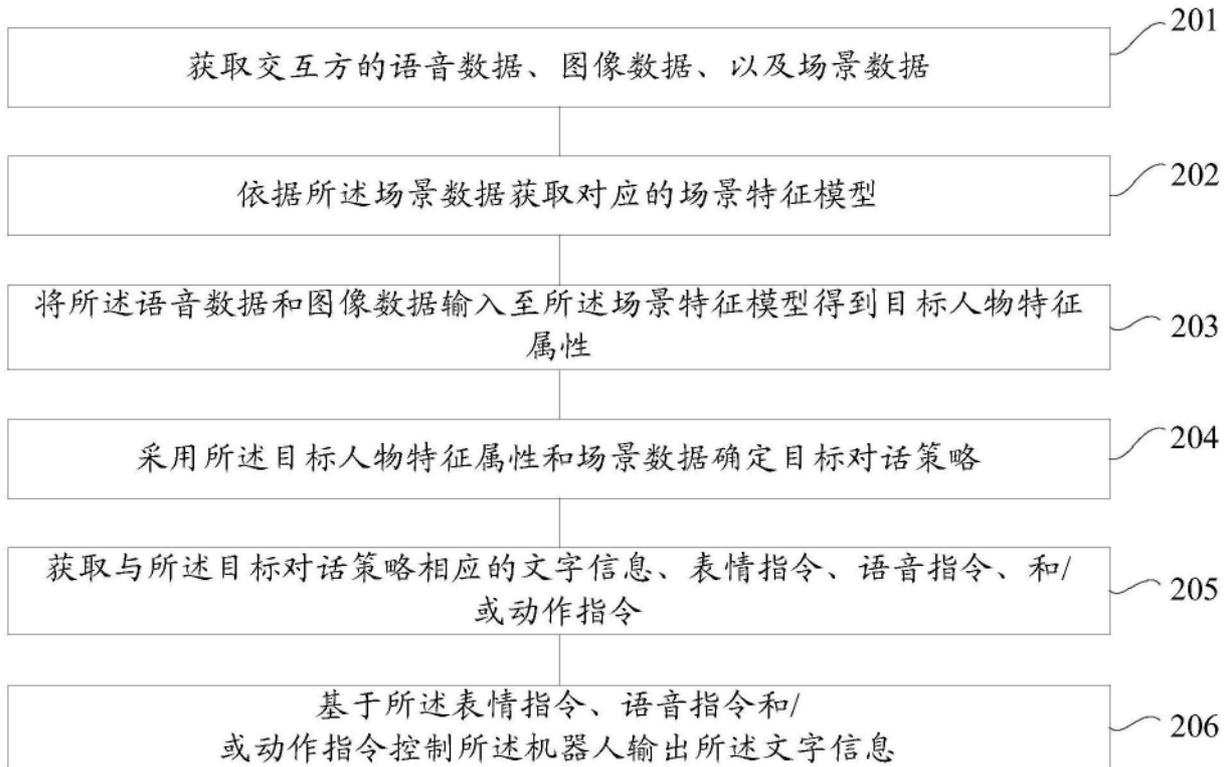


图2

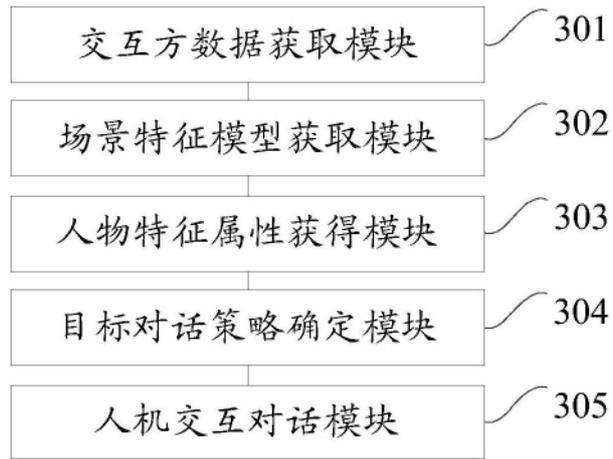


图3

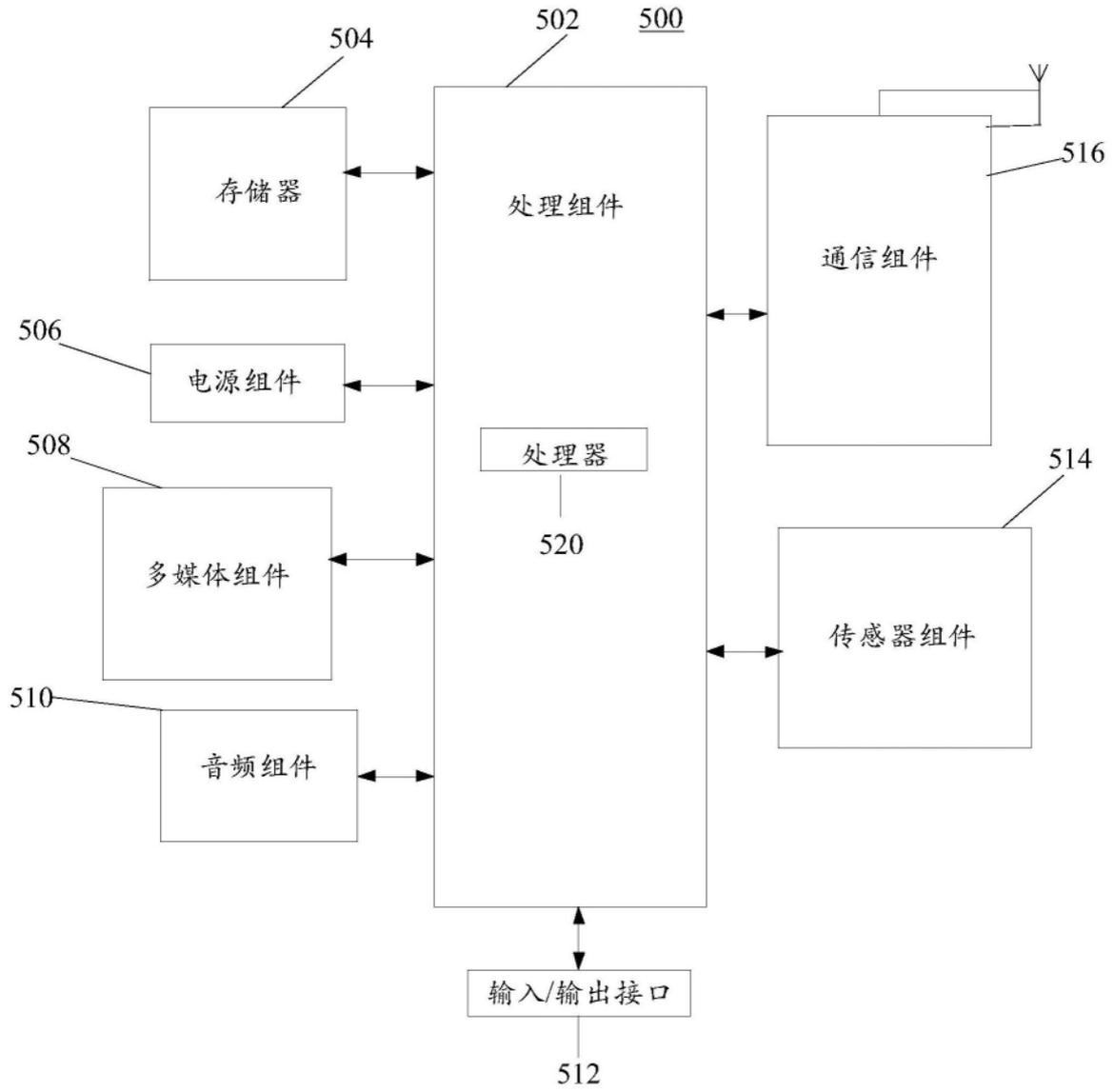


图4

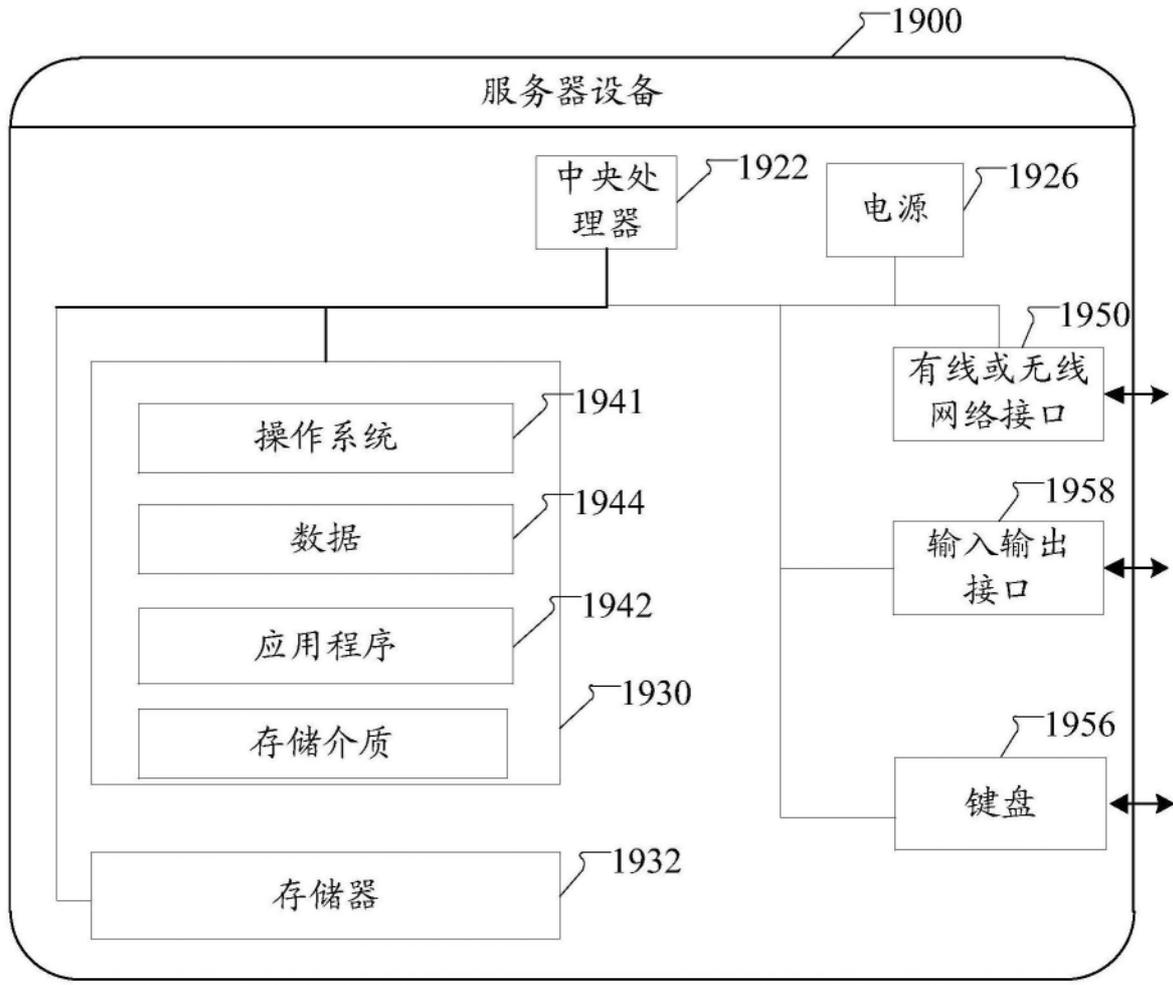


图5