



(12) 发明专利

(10) 授权公告号 CN 108538294 B

(45) 授权公告日 2020.11.17

(21) 申请号 201810392723.9

G10L 15/26 (2006.01)

(22) 申请日 2018.04.27

G10L 15/18 (2013.01)

G10L 15/197 (2013.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 108538294 A

(56) 对比文件

(43) 申请公布日 2018.09.14

CN 107665706 A, 2018.02.06

CN 103020047 A, 2013.04.03

(73) 专利权人 科大讯飞股份有限公司

CN 107665708 A, 2018.02.06

地址 230031 安徽省合肥市高新区望江西路666号

CN 107240398 A, 2017.10.10

CN 105575386 A, 2016.05.11

(72) 发明人 杨强 夏涛 黄鑫 陈志刚 章继东

CN 107423363 A, 2017.12.01

US 2017213546 A1, 2017.07.27

审查员 薛彦杰

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 张小娜 王宝筠

(51) Int. Cl.

G10L 15/22 (2006.01)

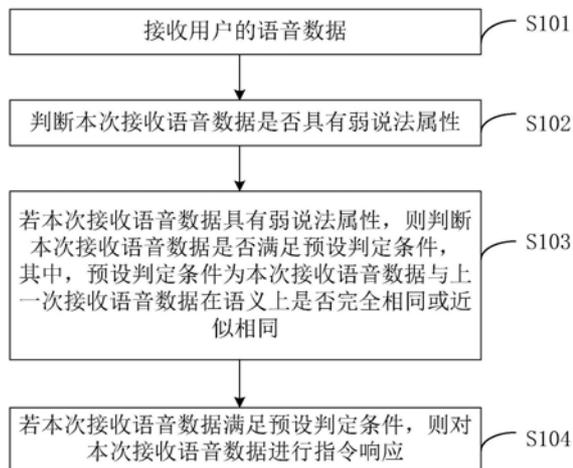
权利要求书3页 说明书19页 附图3页

(54) 发明名称

一种语音交互方法及装置

(57) 摘要

本申请公开了一种语音交互方法,先判断本次接收语音数据是否具有弱说法属性,若本次接收语音数据具有弱说法属性,则进一步判断本次接收语音数据是指令语音还是噪声语音。由于连续两次识别到同一个噪声语音的可能性极低,以及在上一次语音指令不被响应的情况下,用户很可能再次给出相同或相似的语音指令,基于此,如果本次接收语音数据与上一次接收语音数据在语义上完全相同或近似相同,则可以认为本次接收语音数据是指令语音并对其进行指令响应。可见,对于具有弱说法属性的用户语音,能够判断其是否为指令语音,从而对这类指令语音进行响应,进而提升了用户体验。



1. 一种语音交互方法,其特征在于,包括:
接收用户的语音数据;
判断本次接收语音数据是否具有弱说法属性,其中,所述弱说法属性指的是,说法随意、句式结构随意、不易判断出是否有交互需求;
若所述本次接收语音数据具有弱说法属性,则判断所述本次接收语音数据是否满足预设判定条件,其中,所述预设判定条件包括所述本次接收语音数据与上一次接收语音数据在语义上是否完全相同或近似相同;
若所述本次接收语音数据满足所述预设判定条件,则对所述本次接收语音数据进行指令响应。
2. 根据权利要求1所述的方法,其特征在于,所述预设判定条件还包括:
所述本次接收语音数据与所述上一次接收语音数据之间的时间间隔是否小于预设时长。
3. 根据权利要求1所述的方法,其特征在于,所述判断本次接收语音数据是否具有弱说法属性,包括:
获取所述本次接收语音数据的识别文本,作为本次识别文本;
生成所述本次识别文本对应的说法判定特征;
根据所述本次识别文本对应的说法判定特征,判断所述本次接收语音数据是否具有弱说法属性。
4. 根据权利要求3所述的方法,其特征在于,所述根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性,包括:
利用预先构建的强弱说法判定模型,来根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性。
5. 根据权利要求4所述的方法,其特征在于,所述强弱说法判定模型按照以下方式构建:
将当前获取的一条样本文本作为训练文本,并生成所述训练文本对应的说法判定特征;
将所述训练文本对应的说法判定特征输入至当前强弱说法判定模型中,以便当前强弱说法判定模型预测所述训练文本的说法类型,所述说法类型为强说法或弱说法;
获取预先为所述训练文本标注的说法类型;
若所述训练文本的说法类型预测结果和说法类型标注结果不同,则更新当前强弱说法判定模型的模型参数,完成本轮模型训练;
判断是否满足训练结束条件;若是,则结束训练;若否,则获取下一条样本文本并将所述下一条样本文本作为所述训练文本,继续下一轮模型训练。
6. 根据权利要求5所述的方法,其特征在于,所述利用预先构建的强弱说法判定模型,来根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性,包括:
将所述本次识别文本对应的说法判定特征输入至预先构建的强弱说法判定模型,以获取所述强弱说法判定模型的判定结果,其中,所述判定结果包括第一概率和第二概率,所述第一概率为所述本次接收语音数据具有弱说法属性的概率,所述第二概率为所述本次接收

语音数据具有强说法属性的概率；

若所述第一概率大于或等于所述第二概率，则认为所述本次接收语音数据具有弱说法属性；

若所述第一概率小于所述第二概率，则认为所述本次接收语音数据具有强说法属性。

7. 根据权利要求3至6任一项所述的方法，其特征在于，所述说法判定特征的对应文本作为待判定文本时，所述说法判定特征包括以下至少一种特征值：

根据所述待判定文本的文本长度生成的第一特征值；

根据所述待判定文本的分词个数生成的第二特征值；

根据所述待判定文本的文本成分之间的句法关系生成的第三特征值；

根据所述待判定文本中的业务关键词生成的第四特征值；

根据所述待判定文本与上一待判定文本的相关度生成的第五特征值；

根据所述待判定文本与人机交互设备所提问题的相关度生成的第六特征值，所述待判定文本是对所述人机交互设备所提问题的回答文本。

8. 根据权利要求7所述的方法，其特征在于，所述第五特征值按照下述方式生成：

确定所述待判定文本所属的业务说法模板，作为本次说法模板；

确定所述上一待判定文本所属的业务说法模板，作为上一次说法模板；

确定第一条件概率并将所述第一条件概率作为第五特征值，所述第一条件概率是出现所述上一次说法模板的情况下出现所述本次说法模板的概率。

9. 根据权利要求7所述的方法，其特征在于，所述第六特征值按照下述方式生成：

确定所述待判定文本所属的业务说法模板，作为本次说法模板；

确定所述人机交互设备对所述待判定文本所提问题所属的业务说法模板，作为提问说法模板；

确定第二条件概率并将所述第二条件概率作为第六特征值，所述第二条件概率是出现所述提问说法模板的情况下出现所述本次说法模板的概率。

10. 一种语音交互装置，其特征在于，包括：

语音接收单元，用于接收用户的语音数据；

说法判定单元，用于判断本次接收语音数据是否具有弱说法属性，其中，所述弱说法属性指的是，说法随意、句式结构随意、不易判断出是否有交互需求；

指令判定单元，用于若所述本次接收语音数据具有弱说法属性，则判断所述本次接收语音数据是否满足预设判定条件，其中，所述预设判定条件包括所述本次接收语音数据与上一次接收语音数据在语义上是否完全相同或近似相同；

指令响应单元，用于若所述本次接收语音数据满足所述预设判定条件，则对所述本次接收语音数据进行指令响应。

11. 根据权利要求10所述的装置，其特征在于，所述预设判定条件还包括：

所述本次接收语音数据与所述上一次接收语音数据之间的时间间隔是否小于预设时长。

12. 根据权利要求10所述的装置，其特征在于，所述说法判定单元包括：

文本获取子单元，用于获取所述本次接收语音数据的识别文本，作为本次识别文本；

特征生成子单元，用于生成所述本次识别文本对应的说法判定特征；

说法判定子单元,用于根据所述本次识别文本对应的说法判定特征,判断所述本次接收语音数据是否具有弱说法属性。

13. 根据权利要求10至12任一项所述的装置,其特征在于,所述说法判定特征的对应文本作为待判定文本时,所述说法判定特征包括以下至少一种特征值:

根据所述待判定文本的文本长度生成的第一特征值;

根据所述待判定文本的分词个数生成的第二特征值;

根据所述待判定文本的文本成分之间的句法关系生成的第三特征值;

根据所述待判定文本中的业务关键词生成的第四特征值;

根据所述待判定文本与上一待判定文本的相关度生成的第五特征值;

根据所述待判定文本与人机交互设备所提问题的相关度生成的第六特征值,所述待判定文本是对人机交互设备所述所提问题的回答文本。

14. 一种语音交互装置,其特征在于,包括:处理器、存储器、系统总线;

所述处理器以及所述存储器通过所述系统总线相连;

所述存储器用于存储一个或多个程序,所述一个或多个程序包括指令,所述指令当被所述处理器执行时使所述处理器执行如权利要求1-9任一项所述的方法。

15. 一种计算机可读存储介质,包括指令,当其在计算机上运行时,使得计算机执行如权利要求1-9任意一项所述的方法。

一种语音交互方法及装置

技术领域

[0001] 本申请涉及人工智能技术领域,尤其涉及一种语音交互方法及装置。

背景技术

[0002] 随着语音技术的发展,智能语音技术已经被广泛的应用于生产生活的各个领域,当前的人机语音交互,已经由原先类似指令的语音交互,逐步的发展为自然语言的交互。

[0003] 用户语音可以分为两类,一类是具有强说法属性的语音,另一类是具有弱说法属性的语音。但是,对于用户的某些指令语音,现有的人机交互设备可能会因为该用户语音是具有弱说法属性的语音,而选择拒识该用户语音,也即不对该用户语音进行指令响应,从而降低了用户体验。

发明内容

[0004] 有鉴于此,本申请实施例的主要目的在于提供一种语音交互方法及装置,对于具有弱说法属性的用户语音,能够判断其是否为指令语音,从而对这类指令语音进行响应,进而提升了用户体验。

[0005] 本申请实施例提供了一种语音交互方法,包括:

[0006] 接收用户的语音数据;

[0007] 判断本次接收语音数据是否具有弱说法属性;

[0008] 若所述本次接收语音数据具有弱说法属性,则判断所述本次接收语音数据是否满足预设判定条件,其中,所述预设判定条件包括所述本次接收语音数据与上一次接收语音数据在语义上是否完全相同或近似相同;

[0009] 若所述本次接收语音数据满足所述预设判定条件,则对所述本次接收语音数据进行指令响应。

[0010] 可选的,所述预设判定条件还包括:

[0011] 所述本次接收语音数据与所述上一次接收语音数据之间的时间间隔是否小于预设时长。

[0012] 可选的,所述判断本次接收语音数据是否具有弱说法属性,包括:

[0013] 获取所述本次接收语音数据的识别文本,作为本次识别文本;

[0014] 生成所述本次识别文本对应的说法判定特征;

[0015] 根据所述本次识别文本对应的说法判定特征,判断所述本次接收语音数据是否具有弱说法属性。

[0016] 可选的,所述根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性,包括:

[0017] 利用预先构建的强弱说法判定模型,来根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性。

[0018] 可选的,所述强弱说法判定模型按照以下方式构建:

- [0019] 将当前获取的一条样本文本作为训练文本,并生成所述训练文本对应的说法判定特征;
- [0020] 将所述训练文本对应的说法判定特征输入至当前强弱说法判定模型中,以便当前强弱说法判定模型预测所述训练文本的说法类型,所述说法类型为强说法或弱说法;
- [0021] 获取预先为所述训练文本标注的说法类型;
- [0022] 若所述训练文本的说法类型预测结果和说法类型标注结果不同,则更新当前强弱说法判定模型的模型参数,完成本轮模型训练;
- [0023] 判断是否满足训练结束条件;若是,则结束训练;若否,则获取下一条样本文本并将所述下一条样本文本作为所述训练文本,继续下一轮模型训练。
- [0024] 可选的,所述利用预先构建的强弱说法判定模型,来根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性,包括:
- [0025] 将所述本次识别文本对应的说法判定特征输入至预先构建的强弱说法判定模型,以获取所述强弱说法判定模型的判定结果,其中,所述判定结果包括第一概率和第二概率,所述第一概率为所述本次接收语音数据具有弱说法属性的概率,所述第二概率为所述本次接收语音数据具有强说法属性的概率;
- [0026] 若所述第一概率大于或等于所述第二概率,则认为所述本次接收语音数据具有弱说法属性;
- [0027] 若所述第一概率小于所述第二概率,则认为所述本次接收语音数据具有强说法属性。
- [0028] 可选的,所述说法判定特征的对应文本作为待判定文本时,所述说法判定特征包括以下至少一种特征值:
- [0029] 根据所述待判定文本的文本长度生成的第一特征值;
- [0030] 根据所述待判定文本的分词个数生成的第二特征值;
- [0031] 根据所述待判定文本的文本成分之间的句法关系生成的第三特征值;
- [0032] 根据所述待判定文本中的业务关键词生成的第四特征值;
- [0033] 根据所述待判定文本与上一待判定文本的相关度生成的第五特征值;
- [0034] 根据所述待判定文本与人机交互设备所提问题的相关度生成的第六特征值,所述待判定文本是对所述人机交互设备所提问题的回答文本。
- [0035] 可选的,所述第五特征值按照下述方式生成:
- [0036] 确定所述待判定文本所属的业务说法模板,作为本次说法模板;
- [0037] 确定所述上一待判定文本所属的业务说法模板,作为上一次说法模板;
- [0038] 确定第一条件概率并将所述第一条件概率作为第五特征值,所述第一条件概率是出现所述上一次说法模板的情况下出现所述本次说法模板的概率。
- [0039] 可选的,所述第六特征值按照下述方式生成:
- [0040] 确定所述待判定文本所属的业务说法模板,作为本次说法模板;
- [0041] 确定所述人机交互设备对所述待判定文本所提问题所属的业务说法模板,作为提问说法模板;
- [0042] 确定第二条件概率并将所述第二条件概率作为第六特征值,所述第二条件概率是出现所述提问说法模板的情况下出现所述本次说法模板的概率。

- [0043] 本申请实施例还提供了一种语音交互装置,包括:
- [0044] 语音接收单元,用于接收用户的语音数据;
- [0045] 说法判定单元,用于判断本次接收语音数据是否具有弱说法属性;
- [0046] 指令判定单元,用于若所述本次接收语音数据具有弱说法属性,则判断所述本次接收语音数据是否满足预设判定条件,其中,所述预设判定条件包括所述本次接收语音数据与上一次接收语音数据在语义上是否完全相同或近似相同;
- [0047] 指令响应单元,用于若所述本次接收语音数据满足所述预设判定条件,则对所述本次接收语音数据进行指令响应。
- [0048] 可选的,所述预设判定条件还包括:
- [0049] 所述本次接收语音数据与所述上一次接收语音数据之间的时间间隔是否小于预设时长。
- [0050] 可选的,所述说法判定单元包括:
- [0051] 文本获取子单元,用于获取所述本次接收语音数据的识别文本,作为本次识别文本;
- [0052] 特征生成子单元,用于生成所述本次识别文本对应的说法判定特征;
- [0053] 说法判定子单元,用于根据所述本次识别文本对应的说法判定特征,判断所述本次接收语音数据是否具有弱说法属性。
- [0054] 可选的,所述说法判定特征的对应文本作为待判定文本时,所述说法判定特征包括以下至少一种特征值:
- [0055] 根据所述待判定文本的文本长度生成的第一特征值;
- [0056] 根据所述待判定文本的分词个数生成的第二特征值;
- [0057] 根据所述待判定文本的文本成分之间的句法关系生成的第三特征值;
- [0058] 根据所述待判定文本中的业务关键词生成的第四特征值;
- [0059] 根据所述待判定文本与上一待判定文本的相关度生成的第五特征值;
- [0060] 根据所述待判定文本与人机交互设备所提问题的相关度生成的第六特征值,所述待判定文本是对人机交互设备所述所提问题的回答文本。
- [0061] 本申请实施例还提供了一种语音交互装置,包括:处理器、存储器、系统总线;
- [0062] 所述处理器以及所述存储器通过所述系统总线相连;
- [0063] 所述存储器用于存储一个或多个程序,所述一个或多个程序包括指令,所述指令当被所述处理器执行时使所述处理器执行上述语音交互方法中的任意一种实现方式。。
- [0064] 本申请实施例还提供了一种计算机可读存储介质,包括指令,当其在计算机上运行时,使得计算机执行上述语音交互方法中的任意一种实现方式。
- [0065] 本实施例提供的一种语音交互方法及装置,先判断本次接收语音数据是否具有弱说法属性,若本次接收语音数据具有弱说法属性,则进一步判断本次接收语音数据是指令语音还是噪声语音。由于连续两次识别到同一个噪声语音的可能性极低,以及在上一次语音指令不被响应的情况下,用户很可能再次给出相同或相似的语音指令,基于此,如果本次接收语音数据与上一次接收语音数据在语义上完全相同或近似相同,则可以认为本次接收语音数据是指令语音并对其进行指令响应。可见,对于具有弱说法属性的用户语音,能够判断其是否为指令语音,从而对这类指令语音进行响应,进而提升了用户体验。

附图说明

[0066] 为了更清楚地说明本申请实施例或现有技术中的技术方案,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的前提下,还可以根据这些附图获得其他的附图。

[0067] 图1为本申请实施例提供的一种语音交互方法的流程示意图;

[0068] 图2a为本申请实施例提供的说法判定方法的流程示意图之一;

[0069] 图2b为本申请实施例提供的说法判定方法的流程示意图之二;

[0070] 图3为本申请实施例提供的模型训练方法的流程示意图;

[0071] 图4为本申请实施例提供的一种语音交互装置的组成示意图

[0072] 图5为本申请实施例提供的一种语音交互装置的硬件结构示意图。

具体实施方式

[0073] 用户语音可以分为两类,一类是具有强说法属性的语音,另一类是具有弱说法属性的语音。通常情况下,将用户表达中带有交互需求、句式结构比较严格的说法,称之为强说法,比如“我想看中央一套的电视节目”;反之,那些说法比较随意,句式结构要求较低,很难判断出是否有交互需求的说法,称之为弱说法,比如“太好了”。

[0074] 为了给用户提供更人性化的交互体验,人机交互设备需要支持弱说法,但是,具有弱说法属性的用户语音可能是噪音语音也可能是指令语音,因此,在人机交互设备支持弱说法的情况下,具有弱说法属性的噪音语音很可能会触发交互业务。例如,对于支持歌曲单说的音乐业务,由于歌曲资源很多而且很杂,一旦人机交互设备接收到能匹配歌曲名的噪音语音,比如在交谈中比较容易出现的称谓语“父亲”,由于“父亲”是一首歌曲名,所以将会触发音乐业务,但这可能不是用户的真实意图。

[0075] 为此,人机交互设备可以不支持或者限制支持弱说法,即尽量减少支持弱说法,以减少噪音语音误触发交互业务的可能性。但是,如果具有弱说法属性的用户语音是指令语音而非噪音语音,这种不支持或者限制支持弱说法的方式,可能不能根据该用户语音触发交互业务,即不能对该用户语音进行及时响应,从而降低了用户体验。

[0076] 为了解决上述缺陷,本申请实施例提供了一种语音交互方法及装置,对于具有弱说法属性的用户语音,能够判断其是否为指令语音,从而对这类指令语音进行响应,进而提升了用户体验。

[0077] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0078] 第一实施例

[0079] 参见图1,为本实施例提供的一种语音交互方法的流程示意图,该方法可以应用于任意一种具有语音交互功能的智能设备上,比如手机、平板等,后文将该智能设备称为人机交互设备。该语音交互方法可以包括以下步骤:

[0080] S101:接收用户的语音数据。

[0081] 对于接收到的语音数据,可能是用户输入的用于交互的指令语音,也可能是非指令语音也即噪声语音,比如,噪声语音可以是用户与其他人交谈时的人人交谈语音。具有弱说法属性的语音通常情况下为噪声语音。

[0082] S102:判断本次接收语音数据是否具有弱说法属性。

[0083] 针对支持弱说法的人机交互设备,需要有一种措施来判断本次接收语音数据是否具有弱说法属性,具体判断方式请参见第二实施例。

[0084] 若判断结果为本次接收语音数据具有弱说法属性,则执行后续步骤S103。反之,若判断结果为本次接收语音数据不具有弱说法属性,即具有强说法属性,此时,直接对本次接收语音数据进行指令响应,例如,假设本次接收语音数据为“我想听中央一套的电视节目”,人机交互设备可以对其进行指令响应,并将响应结果反馈给用户,该响应结果可以是播放中央一套的当前电视节目或是回放中央一套的某个电视节目。

[0085] S103:若本次接收语音数据具有弱说法属性,则判断本次接收语音数据是否满足预设判定条件。

[0086] 在本实施例中,由于具有弱说法属性的用户语音,其可能是指令语音也可能是噪声语音,因此,当判断本次接收语音数据具有弱说法属性后,还需要进一步判断本次接收语音数据是指令语音还是噪声语音。

[0087] 通常情况下,噪声语音是随机出现的,因此,人机交互设备连续两次识别到同一个噪声语音的可能性极低。此外,如果用户的一条语音是指令语音而非噪声语音,则当该条语音被人机交互设备拒识而得不到正常响应之后,用户在不知道人机交互设备的工作原理的情况下,很自然的认为是人机交互设备没有接收到该条语音,从而会很自然的重复该条语音、或以近似相同的语言表达该条语音,以期望被人机交互设备接收到。

[0088] 基于上述理由,本实施例可以判断本次接收语音数据与上一次接收语音数据在语义上是否完全相同或近似相同,这个判断条件即为S103中的预设判断条件。如果判断结果为“是”,则认为本次接收语音数据是指令语音,因此,可以通过执行后续步骤S104对其进行指令响应,否则,认为本次接收语音数据是噪声语音,可以不对本次接收语音数据进行指令响应。

[0089] S104:若本次接收语音数据满足预设判定条件,则对本次接收语音数据进行指令响应。

[0090] 现举例说明S103-S104,当人机交互设备接收到用户输入语音“父亲”后,在判断其具有弱说法属性的情况下,若在该条语音之前,人机交互设备同样接收到用户输入语音“父亲”或接收到与之语义近似的语音“歌曲父亲”,人机交互设备可以对其进行指令响应,并将响应结果反馈给用户,该响应结果可以是播放歌曲“父亲”。

[0091] 进一步地,通过对人们的行为揣测可知,实际场景中,用户前后给出上述两条相同语音或类似语音的时间间隔通常会比较短。因此,在本实施例的一种实现方式中,S103中的预设判断条件还可以包括“判断本次接收语音数据与上一次接收语音数据之间的时间间隔是否小于预设时长”,使这两个判断条件组成S103中的预设判断条件。现举例说明,当人机交互设备接收到用户输入语音“父亲”后,在判断其具有弱说法属性的情况下,若在该条语音之前的预设时长(比如5秒)内,人机交互设备同样接收到用户输入语音“父亲”或接收到与之语义近似的语音“歌曲父亲”,人机交互设备可以对其进行指令响应,并将响应结果反

馈给用户,该响应结果可以是播放歌曲“父亲”。

[0092] 在本实施例的一种实现方式中,为了实现步骤S104,即实现“对本次接收语音数据进行指令响应”,需要对本次接收语音数据的识别文本进行语义理解,得到语义理解结果,从而根据语义理解结果对本次接收语音数据进行指令响应,并将响应结果反馈给用户。

[0093] 具体实现时,先对本次接收语音数据进行语音识别,得到本次识别文本,本实施例不对语音识别方法进行限制,可以采用现有的或未来出现的任一语音识别方法进行语音识别;然后,对本次识别文本进行语义理解,根据语义理解结果确定本次接收语音数据所指的交互请求,以便针对该交互请求进行指令响应。需要说明的是,本实施例不限制语音识别与语义理解步骤的执行顺序,只要在根据语义理解结果进行指令响应之前执行即可,比如在步骤S101之后进行语音识别和语义理解。

[0094] 在进行语义理解时,可以基于句法的语义分析,比如基于语法规则网络的语义分析方法得到本次识别文本的语义理解结果;也可以通过预先训练的神经网络模型得到本次识别文本的语义理解结果,其中,该神经网络模型可以是卷积神经网络(Convolutional Neural Network,简称CNN)、或循环神经网络(Recurrent neural Network,简称RNN)等。

[0095] 其中,语义理解结果可以包含如下重要信息:

[0096] 业务及业务取值,用于表示当前语义理解结果所属具体业务;

[0097] 操作及操作取值,用于表示当前语义理解结果触发所属具体业务的具体操作;

[0098] 语义槽及语义槽取值,用于表示当前语义理解结果所属具体业务中的语义槽以及对应语义槽的取值。

[0099] 举例说明,假设本次接收语音数据为“我想听刘德华的歌曲”,得到的语义理解结果为“{service:music,operation:play,artist:刘德华}”,其中,service为业务,music为业务取值,operation为操作,play为操作取值,artist为语义槽,刘德华为语义槽取值。

[0100] 综上,本实施例提供的一种语音交互方法,先判断本次接收语音数据是否具有弱说法属性,若本次接收语音数据具有弱说法属性,则进一步判断本次接收语音数据是指令语音还是噪声语音。由于连续两次识别到同一个噪声语音的可能性极低,以及在上一次语音指令不被响应的情况下,用户很可能再次给出相同或相似的语音指令,基于此,如果本次接收语音数据与上一次接收语音数据在语义上完全相同或近似相同,则可以认为本次接收语音数据是指令语音并对其进行指令响应。可见,对于具有弱说法属性的用户语音,能够判断其是否为指令语音,从而对这类指令语音进行响应,进而提升了用户体验。

[0101] 第二实施例

[0102] 本实施例将对第一实施例中S102的具体实施方式进行介绍。

[0103] 参见图2a所示的说法判定方法的流程示意图,上述步骤S102即“判断本次接收语音数据是否具有弱说法属性”,可以包括以下步骤:

[0104] S201:获取本次接收语音数据的识别文本,作为本次识别文本。

[0105] 如第一实施例所述,本实施例不对语音识别方法进行限制,可以采用现有的或未来出现的任一语音识别方法,对本次接收语音数据进行语音识别,得到本次识别文本。

[0106] S202:生成本次识别文本对应的说法判定特征。

[0107] 在本实施例中,该说法判定特征可以理解为基于本次识别文本生成的、与强弱说

法特性相关的文本特征。在生成说法判定特征时,可以按照预设生成规则,来根据本次识别文本生成一条或多条不同类型的特征值。

[0108] 需要说明的是,当需要生成本次识别文本对应的说法判定特征时,可以将本次识别文本作为待判定文本,按照第三实施例中的具体实施方式,生成本次识别文本对应的说法判定特征。

[0109] S203:根据本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性。

[0110] 当根据本次识别文本生成相应的说法判定特征后,便可以基于该说法判定特征来判断本次接收语音数据具有弱说法属性还是具有强说法属性。

[0111] 参见图2b所示,在本实施例的一种实现方式中,本步骤S203具体可以包括:利用预先构建的强弱说法判定模型,来根据本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性。在本实现方式中,可以预先构建一种强弱说法判定模型,该强弱说法判定模型可以是模式识别领域中的常用分类模型,比如支持向量机、深度神经网络、决策树等,然后,利用该强弱说法判定模型来判断本次接收语音数据是具有弱说法属性还是具有强说法属性。

[0112] 为了构建得到强弱说法判定模型,可以建立一初始的强弱说法判定模型,并为模型参数赋予初值。此外,还需要收集大量的样本文本形成样本集合,该样本集合中的这些样本文本,可以直接是文本形式的样本文本、也可以是通过语音识别得到的样本文本,并且,为每条样本文本标注说法类型,说法类型包括两种,一种是“强说法”,另一种是“弱说法”,以便基于这些样本文本以及对这些样本文本的标注结果,对初始建立的强弱说法判定模型进行训练。

[0113] 参见图3所示的模型训练方法的流程示意图,可以按照以下方式训练强弱说法判定模型:

[0114] S301:将当前获取的一条样本文本作为训练文本,并生成该训练文本对应的说法判定特征。

[0115] 从上述样本集合中提取一条样本文本,将该样本文本作为训练样本,可以按照预设生成规则,来根据训练样本生成一条或多条不同类型的特征值。

[0116] 需要说明的是,训练文本对应的说法判定特征的生成方式,应与步骤S202中本次识别文本对应的说法判定特征的生成方式相同。此外,当需要生成训练文本对应的说法判定特征时,可以将该训练文本作为待判定文本,按照第三实施例中的具体实施方式,生成该训练文本对应的说法判定特征。

[0117] S302:将训练文本对应的说法判定特征输入至当前强弱说法判定模型中,以便当前强弱说法判定模型预测训练文本的说法类型,所述说法类型为强说法或弱说法。

[0118] S303:获取预先为训练文本标注的说法类型。

[0119] S304:若训练文本的说法类型预测结果和说法类型标注结果不同,则更新当前强弱说法判定模型的模型参数,完成本轮模型训练。

[0120] 现举例说明步骤S302-S304,假设训练文本为“播放歌曲”,将“播放歌曲”对应的说法判定特征输入至当前强弱说法判定模型中后,强弱说法判定模型便会基于输入特征预测“播放歌曲”的说法类型,比如,预测其为“强说法”。假设预先为“播放歌曲”标注的说法类型

为“弱说法”，那么，预测结果与标注结果不同，则可以更新当前强弱说法判定模型的模型参数，比如，可以基于预测其为“强说法”的概率值进行参数更新。

[0121] S305:判断是否满足训练结束条件；若是，则结束训练；若否，则获取下一条样本文本并将下一条样本文本作为训练文本，继续下一轮模型训练。

[0122] 在本实施例中，所述训练结束条件，可以是达到预设训练次数、也可以是模型参数的变化幅度小于预设阈值，等等。如果没有满足训练结束条件，则从上述样本集合中提取下一条样本文本，并将其作为训练文本以开始下一轮模型训练，即，继续执行步骤S301中的“生成该训练文本对应的说法判定特征”及其后续步骤。

[0123] 在一种实现方式中，当训练得到强弱说法判定模型后，可以利用该强弱说法判定模型按照以下方式实现步骤S203，具体可以包括以下步骤：

[0124] S2031:将本次识别文本对应的说法判定特征输入至预先构建的强弱说法判定模型中，以获取该强弱说法判定模型的判定结果。

[0125] 在本实施例中，直接将本次识别文本对应的说法判定特征作为强弱说法判定模型的输入特征，强弱说法判定模型的输出可以是本次接收语音数据分别为弱说法和强说法的概率，例如，弱说法的概率为0.3，强说法的概率为0.7。

[0126] 为便于区分，本实施例定义了第一概率和第二概率，第一概率为本次接收语音数据具有弱说法属性的概率，第二概率为本次接收语音数据具有强说法属性的概率，第一概率和第二概率共同作为强弱说法判定模型的判定结果。

[0127] S2032:若第一概率大于或等于第二概率，则认为本次接收语音数据具有弱说法属性。

[0128] 例如，假设第一概率为0.7、第二概率为0.3，则可以认为本次接收语音数据具有弱说法属性。

[0129] S2033:若第一概率小于第二概率，则认为本次接收语音数据具有强说法属性。

[0130] 例如，假设第一概率为0.3、第二概率为0.7，则可以认为本次接收语音数据具有强说法属性。

[0131] 综上，本实施例提供的一种语音交互方法，可以预先构建一种强弱说法判定模型，在实际交互过程中，先利用强弱说法判定模型判断本次接收语音数据是否具有弱说法属性，若本次接收语音数据具有弱说法属性，则进一步判断本次接收语音数据是指令语音还是噪声语音。可见，在对本次接收语音数据的说法类型进行判断时，本实施例可以通过强弱说法判定模型实现判断。

[0132] 第三实施例

[0133] 第二实施例中提到，当需要生成所述本次识别文本对应的说法判定特征时，可以将本次识别文本作为待判定文本；当需要生成所述训练文本对应的说法判定特征时，可以将训练文本作为待判定文本。具体生成时，可以使所述说法判定特征包括以下6种特征值中的至少一种特征值：

[0134] 1、根据待判定文本的文本长度生成的第一特征值。

[0135] 可以先确定待判定文本的文本长度L，本实施例不限制文本长度的度量方式，例如，若待判定文本为中文文本，其文本长度L可以是字的个数，若待判定文本为英文文本，其文本长度L可以是字母的个数。然后，根据文本长度L生成一特征值，本实施例将该特征值称

为第一特征值。

[0136] 在一种实现方式中,可以按照下述方式生成第一特征值:

[0137] 若待判定文本的文本长度小于或等于预设的最大文本长度,则利用待判定文本的文本长度与最大文本长度的比值确定第一特征值,具体可以将该比值与预设系数的乘积作为第一特征值;若待判定文本的文本长度大于最大文本长度,则将预设长度特征值作为第一特征值。

[0138] 在本实现方式中,可以定义第一特征值的取值范围 $[0, a]$, a 为大于0的数,比如 $a=1$,并可以将该取值范围的上限值 a 作为所述预设长度特征值。假设待判定文本的文本长度为 L ,最大文本长度为 L_{\max} ;当 $L \leq L_{\max}$ 时,将 $L/L_{\max} * a$ 作为第一特征值,当 $L > L_{\max}$ 时,将 a 作为第一特征值。例如,假设为中文文本设置的最大文本长度 $L_{\max}=20$ 以及 $a=1$,若待判定文本为“请播放张学友的歌”,则其文本长度 $L=8$,因此, $L/L_{\max} * a=8/20$,即 $8/20$ 为第一特征值。

[0139] 2、根据所述待判定文本的分词个数生成的第二特征值。

[0140] 可以先确定待判定文本的分词个数 N ,本实施例不限制分词个数的度量方式,例如,若待判定文本为中文文本,其分词个数 N 可以是采用现有的或未来出现的分词方法对待判定文本进行分词得到的分词个数,若待判定文本为英文文本,其分词个数可以是单词的个数。然后,根据分词个数 N 生成一特征值,本实施例将该特征值称为第二特征值。

[0141] 在一种实现方式中,可以按照下述方式生成第二特征值:

[0142] 若待判定文本的分词个数小于或等于预设的最大分词个数,则利用待判定文本的分词个数与最大分词个数的比值确定第二特征值,具体可以将该比值与预设系数的乘积作为第二特征值;若所述待判定文本的分词个数大于所述最大分词个数,则将预设个数特征值作为第二特征值。

[0143] 在本实现方式中,可以定义第二特征值的取值范围 $[0, b]$, b 为大于0的数,比如 $b=1$,并可以将该取值范围的上限值 b 作为所述预设个数特征值。假设待判定文本的分词个数为 N ,最大分词个数为 N_{\max} ;当 $N \leq N_{\max}$ 时,将 $N/N_{\max} * b$ 作为第二特征值,当 $N > N_{\max}$ 时,将 b 作为第二特征值。例如,假设为中文文本设置的最大分词个数 $N_{\max}=10$ 以及 $b=1$,若待判定文本为“播放歌曲”,则其分词个数 $N=2$,因此, $N/N_{\max} * a=2/10$,即 $2/10$ 为第二特征值。

[0144] 3、根据待判定文本的文本成分之间的句法关系生成的第三特征值。

[0145] 当待判定文本是一个句子或几个句子的组合时,待判定文本的不同文本成分之间可能存在一种或多种句法关系,比如,表1所示的“主谓关系”、“动宾关系”、“间宾关系”、“前置宾语”、“兼语”、“定中关系”、“状中结构”、“动补结构”、“并列关系”、“介宾关系”、“左附加关系”、“右附加关系”、“独立结构”、“核心关系”,等等。可以基于待判定文本中存在的这些句法关系生成一特征值,本实施例将该特征值称为第三特征值。

[0146] 在一种实现方式中,可以按照下述方式生成第三特征值:

[0147] 步骤a1:确定待判定文本的文本成分之间具有的各种句法关系;

[0148] 可以预先收集每一语种的句法关系,形成句法关系集合。

[0149] 在本步骤中,可以先确定待判定文本所属的语种,从而确定该语种对应的句法关系集合;然后,基于该句法关系集合中的句法关系类型,对待判定文本做依存句法分析,得到待判定文本的句法结构,从而确定待判定文本的各个文本成分之间的句法关系。

[0150] 步骤a2:为各种句法关系分别进行赋值,并利用各种句法关系的赋值结果生成第三特征值。

[0151] 在本步骤中,需要为待判定文本的各种句法关系分别赋值,例如表1所示,假设待判定文本存在“主谓关系”,可以为其赋值1,对于待判定文本不存在的句法关系,可以为其赋值0。

[0152] 表1

句法关系编号	句法关系	句法关系赋值结果
1	主谓关系	1
2	动宾关系	0
3	间宾关系	0
4	前置宾语	0
5	兼语	0
6	定中关系	0
7	状中结构	0
8	动补结构	0
9	并列关系	0
10	介宾关系	0
11	左附加关系	0
12	右附加关系	0
13	独立结构	0
14	核心关系	0

[0154] 在一种实现方式中,步骤a2具体可以包括:利用各种句法关系的赋值结果,更新初始M维向量中与各种句法关系对应的初始向量值,其中,初始M维向量是由M种句法关系分别对应的初始向量值构成的;将更新后的向量值作为第三特征值。

[0155] 举例说明,假设上述句法关系集合中共存在在表1所示的14种句法关系,则可以使 $M=14$,预先构建初始14维向量 $[0,0,0,0,0,0,0,0,0,0,0,0,0,0]$,该初始14维向量可以是行向量也可以是列向量,在该初始14维向量中,从左到右的初始向量值0依次对应于表1中编号1-14的句法关系。基于此,假设待判定文本只存在“主谓关系”,可以为“主谓关系”赋值1并更新上述初始14维向量,得到更新后的向量值为 $[1,0,0,0,0,0,0,0,0,0,0,0,0,0]$,其中,1代表待判定文本存在表1中编号为1的“主谓关系”,0代表待判定文本不存在表1中编号2-14的句法关系。

[0156] 在另一种实现方式中,步骤a2可以直接生成一个M维向量作为第三特征值,M表示上述句法关系集合中的句法关系的总个数,使该M维向量表示待判定文本中是否出现相应句法关系。其中,该M维向量中每个元素表示对应句法关系是否出现在待判定文本中,即,当待判定文本中出现相应句法关系时,相应向量元素的取值为1,当待判定文本中未出现相应句法关系时,相应向量元素的取值为0。

[0157] 需要说明的是,本实施例不限定使用0代表不存在相应句法关系、也不限定使用1代表存在相应句法关系,还可以使用其它的两个不同数值来代替。

[0158] 4、根据所述待判定文本的业务关键词生成的第四特征值。

[0159] 可以预先为人机交互业务划分交互业务类型,比如音乐业务、天气业务等。并且,为每一种交互业务类型配置业务关键词,比如音乐业务的关键词可以包括“歌手”、“歌曲”、“专辑”、“想听”、“播放”、“演唱”等,比如天气业务的关键词可以包括“查询”、“天气”、“温度”、“风向”,“PM25”,“雨”、“雪”、“阴”、“晴”等。基于此,可以从待判定文本中提取其对应交互业务类型下的业务关键词,基于这些业务关键词生成一特征值,本实施例将该特征值称为第四特征值。

[0160] 在一种实现方式中,可以按照下述方式生成第四特征值:

[0161] 步骤b1:确定待判定文本所属的交互业务类型,所述交互业务类型下包括至少一个关键词。

[0162] 可以采用第一实施例S104介绍的语义理解方法,对待判定文本进行语义理解,从而确定待判定文本所属的交互业务类型,比如音乐业务、天气业务等。例如,假设待判定文本为“我想听张学友的吻别”,其交互业务类型为音乐业务。

[0163] 步骤b2:确定待判定文本中属于所述交互业务类型的各个业务关键词。

[0164] 可以预先收集每一交互业务类型下的业务关键词,形成关键词集合。

[0165] 可以依次检测待判定文本中的词,确定该词是否出现对应关键词集合(即待判定文本所属的交互业务类型下的关键词集合)中。例如,假设待判定文本为“我想听张学友的吻别”,其涉及的业务关键词包括“歌手”和“歌曲”。

[0166] 步骤b3:为各个业务关键词分别进行赋值,并利用各个业务关键词的赋值结果生成第四特征值。

[0167] 在本步骤中,需要为待判定文本的各个业务关键词分别赋值,例如表2所示,假设待判定文本涉及的业务关键词为“歌手”和“歌曲”,可以为其赋值1,对于待判定文本不涉及的业务关键词,可以为其赋值0。

[0168] 表2

[0169]

业务关键词编号	业务关键词	业务关键词赋值结果
1	歌手	1
2	歌曲	1
3	专辑	0
4	想听	0
5	播放	0
6	演唱	0

[0170] 在一种实现方式中,步骤b3具体可以包括:利用各个业务关键词的赋值结果,更新初始N维向量中与各个业务关键词对应的初始向量值,其中,初始N维向量是所述交互业务类型下的N个业务关键词分别对应的初始向量值构成的;将更新后的向量值作为第四特征值。

[0171] 举例说明,假设待判定文本所属业务交互类型下的关键词集合中,共存在表2所示的6个业务关键词,则可以使 $N=6$,预先构建初始6维向量 $[0,0,0,0,0,0]$,该初始6维向量可以是行向量也可以是列向量,在该初始6维向量中,从左到右的初始向量值0依次对应于表2中编号1-6的业务关键词。基于此,假设待判定文本涉及的业务关键词包括“歌手”和“歌曲”,可以为“歌手”和“歌曲”分别赋值1并更新初始6维向量,得到更新后的向量值为 $[1,1,$

0,0,0,0],其中,1代表待判定文本涉及表2中编号为1-2的业务关键词,0代表待判定文本不涉及表2中编号为3-6的业务关键词。

[0172] 在另一种实现方式中,步骤b2可以直接生成N维向量作为第四征值,N表示上述对应关键词集中的业务关键词的总个数,使该N维向量表示待判定文本中是否涉及相应的业务关键词。其中,该N维向量中每个元素表示对应业务关键词是否出现在待判定文本中,即,当待判定文本中涉及对应业务关键词时,相应向量元素的取值为1,当待判定文本中未涉及对应业务关键词时,相应向量元素的取值为0。

[0173] 需要说明的是,本实施例不限定使用0代表不涉及相应业务关键词、也不限定使用1代表涉及相应业务关键词,还可以使用其它的两个不同数值来代替。

[0174] 5、根据所述待判定文本与上一待判定文本的相关度生成的第五特征值。

[0175] 当所述待判定文本为本次识别文本(即本次接收语音数据的识别文本)时,则上一待判定文本为上一次识别文本(即上一次接收语音数据的识别文本);当所述待判定文本为本轮模型训练使用的训练文本时,则上一待判定文本为上一轮模型训练使用的训练文本。

[0176] 需要说明的是,如果不存在上一待判定文本,可以使第五特征值为默认值,比如0。

[0177] 在一种实现方式中,可以按照下述方式生成第五特征值:

[0178] 步骤c1:确定待判定文本所属的业务说法模板,作为本次说法模板。

[0179] 可以采用第一实施例S104介绍的语义理解方法,对待判定文本进行语义理解,从而确定待判定文本所属的交互业务类型,比如音乐业务、天气业务等。然后,从该交互业务类型下找到待判定文本所属的业务说法模板。

[0180] 例如,假设待判定文本为“我想听刘德华的歌曲”,对其进行语义理解后的语义理解结果为“{service:music,operation:play,artist:刘德华}”,其中“music”为业务取值,表示音乐业务,因此,找到“music”下的说法模板“我想听{\$artist}的歌曲”作为待判定文本所属的业务说法模板,其中,{\$artist}表示业务说法模板中的语义槽,语义槽取值为“刘德华”。

[0181] 步骤c2:确定上一待判定文本所属的业务说法模板,作为上一次说法模板。

[0182] 同样地,可以采用第一实施例S104介绍的语义理解方法,对上一待判定文本进行语义理解,从而确定上一待判定文本所属的交互业务类型,然后,从该交互业务类型下找到上一待判定文本所属的业务说法模板。

[0183] 步骤c3:确定第一条件概率并将第一条件概率作为第五特征值,所述第一条件概率是出现上一次说法模板的情况下出现本次说法模板的概率。

[0184] 在本实施例中,可以使用条件概率 $P(B|A)$ 作为第一条件概率。其中,B表示当前待判定文本所属的业务说法模板,即本次说法模板;A表示上一待判定文本所属的业务说法模板,即上一次说法模板; $P(B|A)$ 表示在所述上一次说法模板出现的情况下,出现所述本次说法模板的概率。

[0185] 在一种实现方式中,所述确定第一条件概率,具体可以包括:在第一样本语料中,确定所述上一次说法模板的出现次数,并确定所述上一次说法模板出现的情况下所述本次说法模板的出现次数;根据确定的出现次数计算第一条件概率。

[0186] 在本实现方式中,可以预先收集大量的用户语音数据和/或人机交互设备给出的反馈结果,形成第一样本语料,并从第一样本语料中抽象出不同交互业务类型下的业务说

法模板。在本实施例中，第一样本语料可以是用于训练强弱说法判定模型的训练数据。

[0187] 基于第一样本语料以及从中抽象出的业务说法模板，可以按照公式(1)计算第一条件概率：

$$[0188] \quad P(B|A) = \frac{M_{AB}}{M_A} \quad (1)$$

[0189] 其中， M_{AB} 表示在所述上一次说法模板出现在第一样本语料的情况下，第一样本语料中出现所述本次说法模板的次数； M_A 表示第一样本语料中出现所述上一次说法模板的次数。

[0190] 例如，假设上一待判定文本为“我想查询明天合肥的天气情况”，对其进行语义理解后的语义理解结果为“{service:weather,operation:query,time:明天,city:合肥}”，其中“weather”为业务取值，表示天气业务，因此，找到“weather”下的业务说法模板“我想查询{date}{city}的天气情况”作为当前待判定文本所属的业务说法模板，其中，{date}和{city}表示业务说法模板中的语义槽，语义槽取值为“明天”和“合肥”。假设当前待判定文本为“明天的呢”，其所属的业务说法模板为“{date}的呢”。

[0191] 可见，“我想查询{date}{city}的天气情况”即为所述上一次说法模板，简称说法模板A，“{date}的呢”即为所述本次说法模板，简称说法模板B。在第一样本语料中，假设说法模板A出现的总数为8000次，在说法模板A出现的情况下，说法模板B出现的总数为6000次，则根据公式(1)进行计算，即 $P(B|A) = 6000/8000 = 0.75$ ，该值即为第一条件概率。

[0192] 6、根据待判定文本与人机交互设备所提问题的相关度生成的第六特征值，其中，待判定文本是对所述人机交互设备所提问题的回答文本。

[0193] 例如，当人机交互设备向用户提出问题时，比如“你要去哪”，如果用户对该问题进行了语音回答，比如答复语音为“我要去科大”，当答复语音作为所述待判定文本时，可以根据“我要去科大”与“你要去哪”的相关度生成第六特征值。

[0194] 需要说明的是，如果所述待判定文本不是对所述人机交互设备所提问题的回答文本，可以使第六特征值为默认值，比如0。

[0195] 在本实施例中，需要判断所述待判定文本是否是对所述人机交互设备所提问题的回答文本。具体地，可以先判断所述待判定文本之前是否存在一条人机交互设备所提的问题，即该问题与所述待判定文本是否相继出现；如果是，则对该问题以及所述待判定文本进行语义分析，根据语义分析结果进行判断。

[0196] 例如，可以确定该问题中与询问内容相关的语义槽，比如该问题是“你要去哪”，其相关语义槽则为“地理位置”，此时，再分析所述待判定文本是否对相关语义槽“地理位置”进行了填充，比如所述待判定文本为“我要去科大”，而“科大”是对“地理位置”的填充结果，因此，可以判定“我要去科大”是对“你要去哪”的回答文本。

[0197] 在一种实现方式中，可以按照下述方式生成第六特征值：

[0198] 步骤d1：确定待判定文本所属的业务说法模板，作为本次说法模板。

[0199] 需要说明的是，本步骤与上述步骤c1相同，具体实现请参见上述步骤c1的介绍，在此不再赘述。

[0200] 步骤d2：确定人机交互设备对所述待判定文本所提问题所属的业务说法模板，作为提问说法模板。

[0201] 需要说明的是,本步骤与上述步骤c1类似,只要将待判定文本替换为人机交互设备所提问题(即问题文本)即可,具体实现请参见上述步骤c1的介绍,在此不再赘述。

[0202] 步骤d3:确定第二条件概率并将所述第二条件概率作为第六特征值,所述第二条件概率是出现所述提问说法模板的情况下出现所述本次说法模板的概率。

[0203] 在本实施例中,可以使用条件概率 $P(B|C)$ 作为第一条件概率。其中,B表示当前待判定文本所属的业务说法模板,即本次说法模板;C表示人机交互设备所提问题所属的业务说法模板,即提问说法模板; $P(B|C)$ 表示在所述提问说法模板出现的情况下,出现所述本次说法模板的概率。

[0204] 在一种实现方式中,所述确定第二条件概率,具体可以包括:在第二样本语料中,确定所述提问说法模板的出现次数,并确定所述提问说法模板出现的情况下所述本次说法模板的出现次数;根据确定的出现次数计算第二条件概率。

[0205] 在本实现方式中,可以预先收集大量的问答组合,该问答组合包括人机交互设备所提的问题以及用户针对该问题的回答,形成第二样本语料,进一步地,该第二样本语料还可以包括与该问答组合相关的其它用户语音数据和/或人机交互设备给出的反馈结果。并且,还需要从该第二样本语料中抽象出不同交互业务类型下的业务说法模板。

[0206] 基于第二样本语料以及从中抽象出的业务说法模板,可以按照公式(2)计算第二条件概率:

$$[0207] \quad P(B|C) = \frac{M_{CB}}{M_C} \quad (2)$$

[0208] 其中, M_{CB} 表示在所述提问说法模板出现在第二样本语料的情况下,第一样本语料中出现所述本次说法模板的次数; M_C 表示第二样本语料中出现所述提问说法模板的次数。

[0209] 例如,假设人机交互设备所提问题为“你要去哪”,其所属的业务说法模板为“你要去\$where”,假设当前待判定文本为“我要去科大”,其所属的业务说法模板为“去\$where”。可见,“你要去哪”即为所述提问说法模板,简称说法模板C,“去\$where”即为所述本次说法模板,简称说法模板B,在第二样本语料中,假设说法模板C出现的总数为8000次,在说法模板C出现的情况下,说法模板B出现的总数为5000次,则根据公式(2)进行计算,即 $P(B|C) = 5000/8000 = 0.625$,该值即为第二条件概率。

[0210] 综上,本实施例提供的一种语音交互方法,先生成本次接收语音数据的识别文本对应的说法判定特征,然后将这些特征作为强弱说法判定模型的输入特征,以利用强弱说法判定模型判断本次接收语音数据是否具有弱说法属性,若本次接收语音数据具有弱说法属性,则进一步判断本次接收语音数据是指令语音还是噪声语音。可见,在对本次接收语音数据的说法类型进行判断时,本实施例可以生成上述6种特征值中的至少一个特征值,使强弱说法判定模型利用这些特征值进行判断。

[0211] 第四实施例

[0212] 本实施例将对一种语音交互装置进行介绍,相关内容请参见上述方法实施例。其中,该语音交互装置可以是上述方法实施例中提及的人机交互设备、也可以是上述人机交互设备中的组成部分。

[0213] 参见图4,为本实施例提供的一种语音交互装置的组成示意图,该语音交互设备400包括:

- [0214] 语音接收单元401,用于接收用户的语音数据;
- [0215] 说法判定单元402,用于判断本次接收语音数据是否具有弱说法属性;
- [0216] 指令判定单元403,用于若所述本次接收语音数据具有弱说法属性,则判断所述本次接收语音数据是否满足预设判定条件,其中,所述预设判定条件包括所述本次接收语音数据与上一次接收语音数据在语义上是否完全相同或近似相同;
- [0217] 指令响应单元404,用于若所述本次接收语音数据满足所述预设判定条件,则对所述本次接收语音数据进行指令响应。
- [0218] 在本实施例的一种实现方式中,所述预设判定条件还包括:
- [0219] 所述本次接收语音数据与所述上一次接收语音数据之间的时间间隔是否小于预设时长。
- [0220] 在本实施例的一种实现方式中,所述说法判定单元包括:
- [0221] 文本获取子单元,用于获取所述本次接收语音数据的识别文本,作为本次识别文本;
- [0222] 特征生成子单元,用于生成所述本次识别文本对应的说法判定特征;
- [0223] 说法判定子单元,用于根据所述本次识别文本对应的说法判定特征,判断所述本次接收语音数据是否具有弱说法属性。
- [0224] 在本实施例的一种实现方式中,所述说法判定子单元,具体用于利用预先构建的强弱说法判定模型,来根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性。
- [0225] 在本实施例的一种实现方式中,所述强弱说法判定模型通过模型构建单元进行构建;所述模型构建单元包括:
- [0226] 说法特征生成子单元,用于将当前获取的一条样本文本作为训练文本,并生成所述训练文本对应的说法判定特征;
- [0227] 说法类型预测子单元,用于将所述训练文本对应的说法判定特征输入至当前强弱说法判定模型中,以便当前强弱说法判定模型预测所述训练文本的说法类型,所述说法类型为强说法或弱说法;
- [0228] 标注类型获取子单元,用于获取预先为所述训练文本标注的说法类型;
- [0229] 模型参数更新子单元,用于若所述训练文本的说法类型预测结果和说法类型标注结果不同,则更新当前强弱说法判定模型的模型参数,完成本轮模型训练;
- [0230] 结束条件判断子单元,用于判断是否满足训练结束条件;若是,则结束训练;若否,则获取下一条样本文本并将所述下一条样本文本作为所述训练文本,继续下一轮模型训练。
- [0231] 在本实施例的一种实现方式中,所述说法判定子单元包括:
- [0232] 判定结果获取子单元,用于将所述本次识别文本对应的说法判定特征输入至预先构建的强弱说法判定模型,以获取所述强弱说法判定模型的判定结果,其中,所述判定结果包括第一概率和第二概率,所述第一概率为所述本次接收语音数据具有弱说法属性的概率,所述第二概率为所述本次接收语音数据具有强说法属性的概率;
- [0233] 弱说法确定子单元,用于若所述第一概率大于或等于所述第二概率,则认为所述本次接收语音数据具有弱说法属性;

[0234] 强说法确定子单元,用于若所述第一概率小于所述第二概率,则认为所述本次接收语音数据具有强说法属性。

[0235] 在本实施例的一种实现方式中,所述说法判定特征的对应文本作为待判定文本时,所述说法判定特征包括以下至少一种特征值:

[0236] 根据所述待判定文本的文本长度生成的第一特征值;

[0237] 根据所述待判定文本的分词个数生成的第二特征值;

[0238] 根据所述待判定文本的文本成分之间的句法关系生成的第三特征值;

[0239] 根据所述待判定文本中的业务关键词生成的第四特征值;

[0240] 根据所述待判定文本与上一待判定文本的相关度生成的第五特征值;

[0241] 根据所述待判定文本与人机交互设备所提问题的相关度生成的第六特征值,所述待判定文本是对人机交互设备所述所提问题的回答文本。

[0242] 在本实施例的一种实现方式中,所述第五特征值按照下述方式生成:

[0243] 确定所述待判定文本所属的业务说法模板,作为本次说法模板;

[0244] 确定所述上一待判定文本所属的业务说法模板,作为上一次说法模板;

[0245] 确定第一条件概率并将所述第一条件概率作为第五特征值,所述第一条件概率是出现所述上一次说法模板的情况下出现所述本次说法模板的概率。

[0246] 在本实施例的一种实现方式中,所述第六特征值按照下述方式生成:

[0247] 确定所述待判定文本所属的业务说法模板,作为本次说法模板;

[0248] 确定所述人机交互设备对所述待判定文本所提问题所属的业务说法模板,作为提问说法模板;

[0249] 确定第二条件概率并将所述第二条件概率作为第六特征值,所述第二条件概率是出现所述提问说法模板的情况下出现所述本次说法模板的概率。

[0250] 第五实施例

[0251] 本实施例将对另一种语音交互装置进行介绍,相关内容请参见上述方法实施例。其中,该语音交互装置可以是上述方法实施例中提及的人机交互设备、也可以是上述人机交互设备中的组成部分。

[0252] 参见图5,为本实施例提供的一种语音交互装置的硬件结构示意图,所述语音交互装置500包括存储器501和接收器502,以及分别与所述存储器501和所述接收器502连接的处理器503,所述存储器501用于存储一组程序指令,所述处理器503用于调用所述存储器501存储的程序指令执行如下操作:

[0253] 接收用户的语音数据;

[0254] 判断本次接收语音数据是否具有弱说法属性;

[0255] 若所述本次接收语音数据具有弱说法属性,则判断所述本次接收语音数据是否满足预设判定条件,其中,所述预设判定条件包括所述本次接收语音数据与上一次接收语音数据在语义上是否完全相同或近似相同;

[0256] 若所述本次接收语音数据满足所述预设判定条件,则对所述本次接收语音数据进行指令响应。

[0257] 在本实施例的一种实现方式中,所述预设判定条件还包括:

[0258] 所述本次接收语音数据与所述上一次接收语音数据之间的时间间隔是否小于预

设时长。

[0259] 在本实施例的一种实现方式中,所述处理器503还用于调用所述存储器501存储的程序指令执行如下操作:

[0260] 获取所述本次接收语音数据的识别文本,作为本次识别文本;

[0261] 生成所述本次识别文本对应的说法判定特征;

[0262] 根据所述本次识别文本对应的说法判定特征,判断所述本次接收语音数据是否具有弱说法属性。

[0263] 在本实施例的一种实现方式中,所述处理器503还用于调用所述存储器501存储的程序指令执行如下操作:

[0264] 利用预先构建的强弱说法判定模型,来根据所述本次识别文本对应的说法判定特征,判断本次接收语音数据是否具有弱说法属性。

[0265] 在本实施例的一种实现方式中,所述处理器503还用于调用所述存储器501存储的程序指令执行如下操作:

[0266] 将当前获取的一条样本文本作为训练文本,并生成所述训练文本对应的说法判定特征;

[0267] 将所述训练文本对应的说法判定特征输入至当前强弱说法判定模型中,以便当前强弱说法判定模型预测所述训练文本的说法类型,所述说法类型为强说法或弱说法;

[0268] 获取预先为所述训练文本标注的说法类型;

[0269] 若所述训练文本的说法类型预测结果和说法类型标注结果不同,则更新当前强弱说法判定模型的模型参数,完成本轮模型训练;

[0270] 判断是否满足训练结束条件;若是,则结束训练;若否,则获取下一条样本文本并将所述下一条样本文本作为所述训练文本,继续下一轮模型训练。

[0271] 在本实施例的一种实现方式中,所述处理器503还用于调用所述存储器501存储的程序指令执行如下操作:

[0272] 将所述本次识别文本对应的说法判定特征输入至预先构建的强弱说法判定模型,以获取所述强弱说法判定模型的判定结果,其中,所述判定结果包括第一概率和第二概率,所述第一概率为所述本次接收语音数据具有弱说法属性的概率,所述第二概率为所述本次接收语音数据具有强说法属性的概率;

[0273] 若所述第一概率大于或等于所述第二概率,则认为所述本次接收语音数据具有弱说法属性;

[0274] 若所述第一概率小于所述第二概率,则认为所述本次接收语音数据具有强说法属性。

[0275] 在本实施例的一种实现方式中,所述说法判定特征的对应文本作为待判定文本时,所述说法判定特征包括以下至少一种特征值:

[0276] 根据所述待判定文本的文本长度生成的第一特征值;

[0277] 根据所述待判定文本的分词个数生成的第二特征值;

[0278] 根据所述待判定文本的文本成分之间的句法关系生成的第三特征值;

[0279] 根据所述待判定文本中的业务关键词生成的第四特征值;

[0280] 根据所述待判定文本与上一待判定文本的相关度生成的第五特征值;

[0281] 根据所述待判定文本与人机交互设备所提问题的相关度生成的第六特征值,所述待判定文本是对所述人机交互设备所提问题的回答文本。

[0282] 在本实施例的一种实现方式中,所述处理器503还用于调用所述存储器501存储的程序指令执行如下操作:

[0283] 确定所述待判定文本所属的业务说法模板,作为本次说法模板;

[0284] 确定所述上一待判定文本所属的业务说法模板,作为上一次说法模板;

[0285] 确定第一条件概率并将所述第一条件概率作为第五特征值,所述第一条件概率是出现所述上一次说法模板的情况下出现所述本次说法模板的概率。

[0286] 在本实施例的一种实现方式中,所述处理器503还用于调用所述存储器501存储的程序指令执行如下操作:

[0287] 确定所述待判定文本所属的业务说法模板,作为本次说法模板;

[0288] 确定所述人机交互设备对所述待判定文本所提问题所属的业务说法模板,作为提问说法模板;

[0289] 确定第二条件概率并将所述第二条件概率作为第六特征值,所述第二条件概率是出现所述提问说法模板的情况下出现所述本次说法模板的概率。

[0290] 在一些实施方式中,所述处理器503可以为中央处理器(Central Processing Unit,CPU),所述存储器501可以为随机存取存储器(Random Access Memory,RAM)类型的内部存储器,所述接收器502可以包含普通物理接口,所述物理接口可以为以太(Ethernet)接口或异步传输模式(Asynchronous Transfer Mode,ATM)接口。所述处理器503、接收器502和存储器501可以集成为一个或多个独立的电路或硬件,如:专用集成电路(Application Specific Integrated Circuit,ASIC)。

[0291] 进一步地,本实施例还提供了一种计算机可读存储介质,包括指令,当其在计算机上运行时,使得计算机执行上述语音交互方法中的任意一种实现方式。

[0292] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到上述实施例方法中的全部或部分步骤可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者诸如媒体网关等网络通信设备,等等)执行本申请各个实施例或者实施例的某些部分所述的方法。

[0293] 需要说明的是,本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0294] 还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在

包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0295] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下,在其它实施例中实现。因此,本申请将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

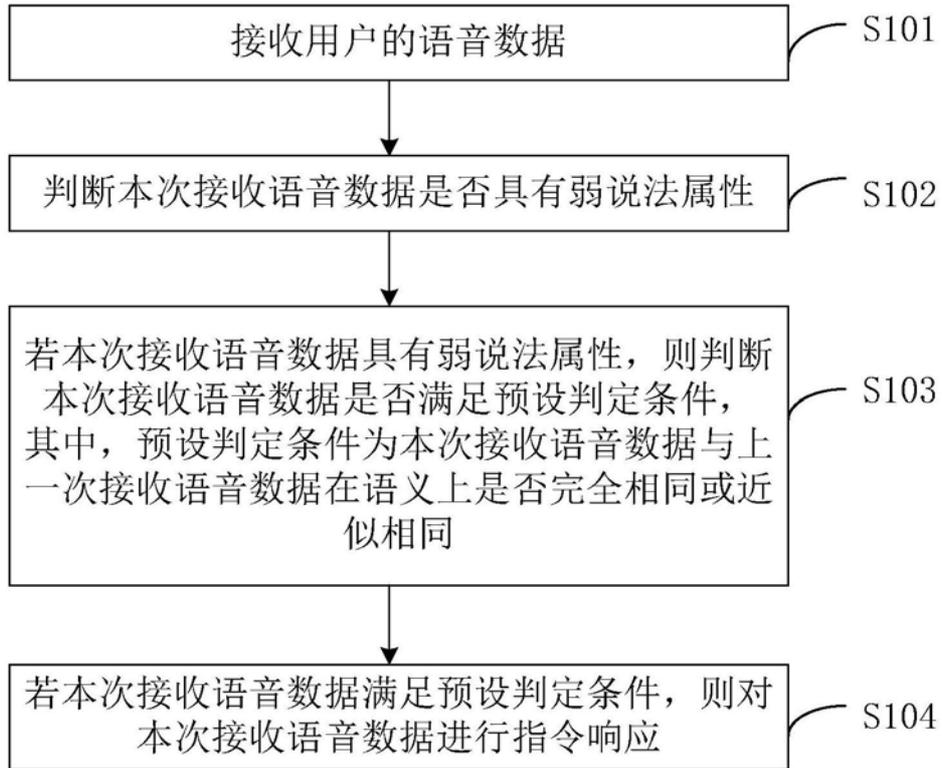


图1

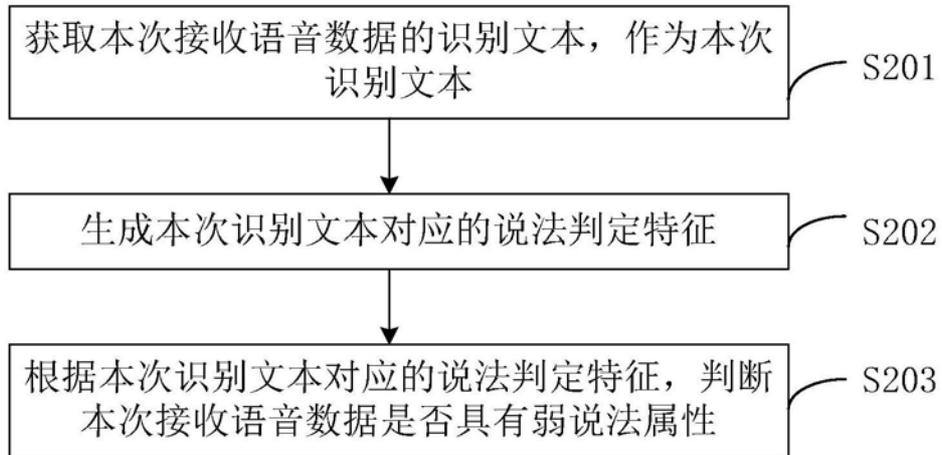


图2a

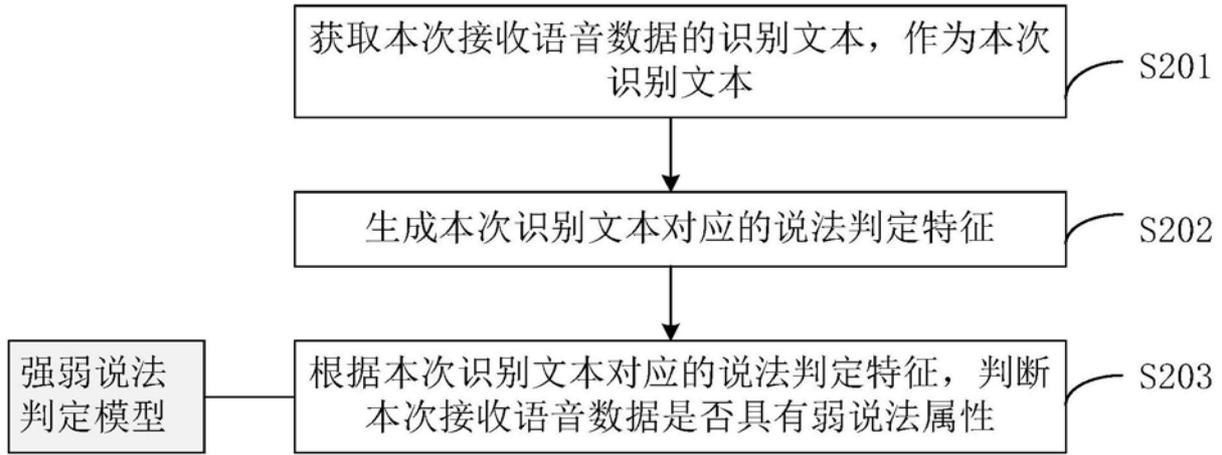


图2b

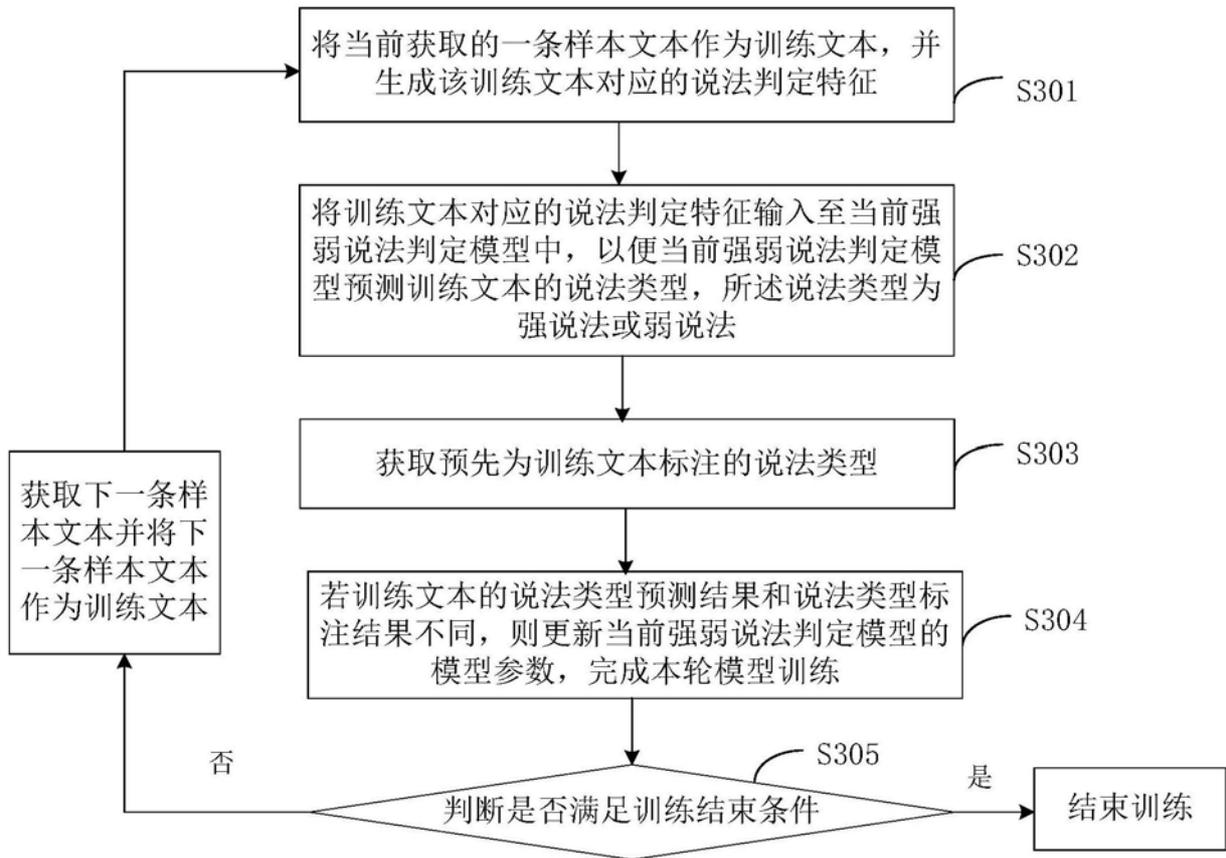


图3

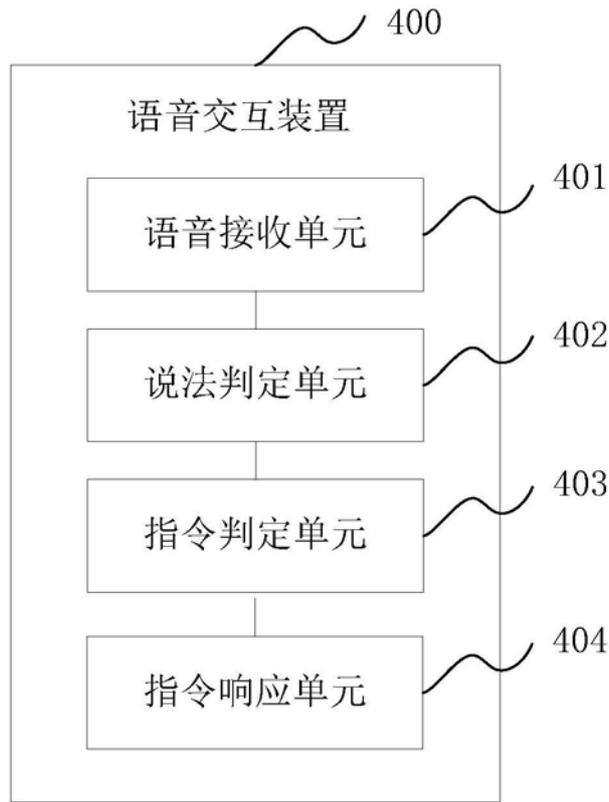


图4

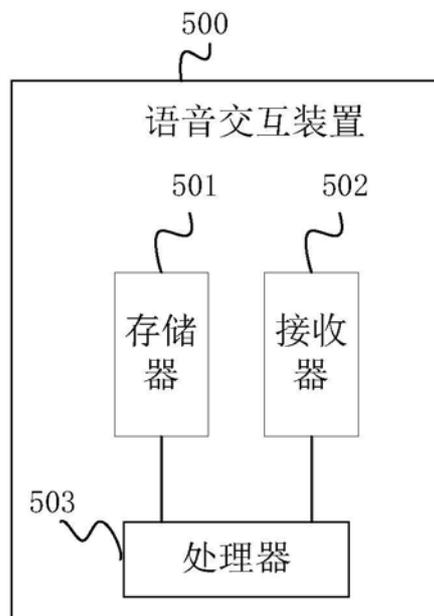


图5