



(12) 发明专利申请

(10) 申请公布号 CN 116469389 A

(43) 申请公布日 2023. 07. 21

(21) 申请号 202310318553.0

(22) 申请日 2023.03.28

(71) 申请人 京东科技信息技术有限公司

地址 100176 北京市大兴区北京经济技术
开发区科创十一街18号院2号楼6层
601(72) 发明人 冯明超 黎陈 王佳 王佳
张政臣 陈蒙 何晓冬(74) 专利代理机构 中原信达知识产权代理有限
责任公司 11219
专利代理师 韩黎捷 张效荣

(51) Int. Cl.

G10L 15/26 (2006.01)

G10L 15/06 (2013.01)

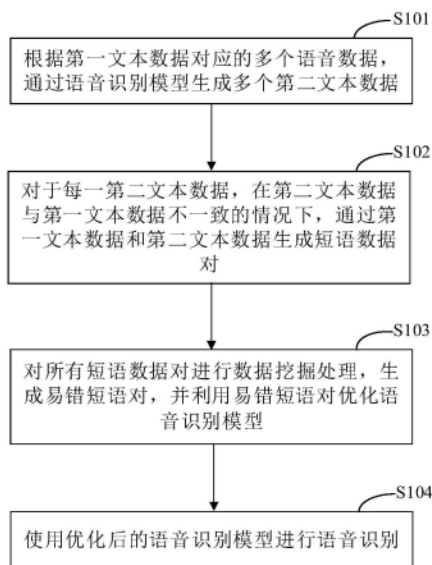
权利要求书2页 说明书10页 附图4页

(54) 发明名称

一种语音识别的方法和装置

(57) 摘要

本发明公开了一种语音识别的方法和装置，涉及人工智能技术领域。该方法的一具体实施方式包括：根据第一文本数据对应的多个语音数据，通过语音识别模型生成多个第二文本数据；对于每一第二文本数据，在第二文本数据与第一文本数据不一致的情况下，通过第一文本数据和第二文本数据生成短语数据对；对所有短语数据对进行数据挖掘处理，生成易错短语对，并利用易错短语对优化语音识别模型；使用优化后的语音识别模型进行语音识别。该实施方式能够通过挖掘异常数据得到高质量的训练数据，降低优化模型所耗费的成本，提高优化模型的效率，并且在使用时可以提高语音识别模型的准确率和效率，从而优化人机语言交互系统的效果，提高用户的使用体验。



1. 一种语音识别的方法,其特征在于,包括:

根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;

对于每一所述第二文本数据,在所述第二文本数据与所述第一文本数据不一致的情况下,通过所述第一文本数据和所述第二文本数据生成短语数据对;

对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用所述易错短语对优化所述语音识别模型;

使用优化后的语音识别模型进行语音识别。

2. 根据权利要求1所述的方法,其特征在于,所述通过所述第一文本数据和所述第二文本数据生成短语数据对,包括:

对所述第一文本数据和所述第二文本数据进行差异片段对齐处理,生成短语数据对,所述短语数据对包括:与所述第一文本数据对应的第一短语数据,和与所述第二文本数据对应的第二短语数据。

3. 根据权利要求2所述的方法,其特征在于,所述对所有短语数据对进行数据挖掘处理,生成易错短语对,包括:

对于每一所述短语数据对,生成所述短语数据对的两个差异片段,并将所述两个差异片段作为一个差异单元;

统计所有短语数据对中每个差异片段出现的频次,以及每个差异单元出现的频次;

按照预设的数据挖掘算法,根据所述每个差异片段出现的频次和所述每个差异单元出现的频次,得到高频短语数据对,并将所述高频短语数据对作为所述易错短语对。

4. 根据权利要求1或3所述的方法,其特征在于,所述利用所述易错短语对优化所述语音识别模型,包括:

对于每一所述易错短语对,根据所述易错短语对对应的第二文本数据对应的语音数据,生成与所述第二短语数据对应的短语语音数据,并将所述短语语音数据作为输入,所述第一短语数据作为输出,通过模型训练来优化所述语音识别模型。

5. 根据权利要求1所述的方法,其特征在于,所述利用所述易错短语对优化所述语音识别模型,包括:

根据所述易错短语对确定待识别的热度词,并通过提高所述热度词的识别率以优化所述语音识别模型。

6. 根据权利要求1所述的方法,其特征在于,所述利用所述易错短语对优化所述语音识别模型,包括:

根据所述易错短语对训练文本纠错模型,并结合所述文本纠错模型来优化所述语音识别模型。

7. 根据权利要求1所述的方法,其特征在于,所述使用优化后的语音识别模型进行语音识别之后,还包括:

响应于用户意图识别请求,获取通过所述优化后的语音识别模型生成的用户文本数据;

将所述用户文本数据输入至语言模型中,生成用户文本向量;

根据所述用户文本向量,通过意图识别模型得到用户意图识别结果。

8. 根据权利要求7所述的方法,其特征在于,所述语言模型是通过如下方式训练的:

对于每一所述易错短语对,分别将所述易错短语对对应的第一短语数据和第二短语数据输入至所述语言模型,生成所述第一短语数据的文本向量和所述第二短语数据的文本向量;

根据所述第一短语数据的文本向量和所述第二短语数据的文本向量,得到文本向量相似度的损失函数;

以所述文本向量相似度的损失函数趋近于0为训练目标,训练所述语言模型。

9.根据权利要求7所述的方法,其特征在于,所述意图识别模型是通过如下方式训练的:

对于每一所述高频短语数据对,将所述第二文本数据的文本向量作为输入,所述第一文本数据的意图识别结果作为输出,训练所述意图识别模型。

10.一种语音识别的装置,其特征在于,包括:

文本数据生成模块,用于根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;

短语数据对生成模块,用于对于每一所述第二文本数据,在所述第二文本数据与所述第一文本数据不一致的情况下,通过所述第一文本数据和所述第二文本数据生成短语数据对;

语音识别模型优化模块,用于对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用所述易错短语对优化所述语音识别模型;

语音识别模块,用于使用优化后的语音识别模型进行语音识别。

11.一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-9中任一所述的方法。

12.一种计算机可读介质,其上存储有计算机程序,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1-9中任一所述的方法。

一种语音识别的方法和装置

技术领域

[0001] 本发明涉及人工智能技术领域,尤其涉及一种语音识别的方法和装置。

背景技术

[0002] 自动语音识别(ASR)是人工智能领域中的一项基础性技术,其目标是将语音转换为对应的文字信息,自动语音识别作为人机语音交互系统的前置基础模块,也影响着人机语言交互系统的整体效果。目前优化自动语音识别的方案为基于深度学习设计复杂度高的网络模型,使用大规模的训练数据集来对网络模型进行训练。

[0003] 在实现本发明过程中,发明人发现现有技术中至少存在如下问题:

[0004] 优化模型所耗费的成本较高,优化模型的效率低,并且在使用时语音识别模型的准确率和效率低,从而影响了人机语言交互系统的效果,降低了用户的使用体验。

发明内容

[0005] 有鉴于此,本发明实施例提供一种语音识别的方法和装置,能够通过挖掘异常数据得到高质量的训练数据,降低优化模型所耗费的成本,提高优化模型的效率,并且在使用时可以提高语音识别模型的准确率和效率,从而优化人机语言交互系统的效果,提高用户的使用体验。

[0006] 为实现上述目的,根据本发明实施例的一个方面,提供了一种语音识别的方法。

[0007] 一种语音识别的方法,包括:根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;对于每一所述第二文本数据,在所述第二文本数据与所述第一文本数据不一致的情况下,通过所述第一文本数据和所述第二文本数据生成短语数据对;对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用所述易错短语对优化所述语音识别模型;使用优化后的语音识别模型进行语音识别。

[0008] 可选地,所述通过所述第一文本数据和所述第二文本数据生成短语数据对,包括:对所述第一文本数据和所述第二文本数据进行差异片段对齐处理,生成短语数据对,所述短语数据对包括:与所述第一文本数据对应的第一短语数据,和与所述第二文本数据对应的第二短语数据。

[0009] 可选地,所述对所有短语数据对进行数据挖掘处理,生成易错短语对,包括:对于每一所述短语数据对,生成所述短语数据对的两个差异片段,并将所述两个差异片段作为一个差异单元;统计所有短语数据对中每个差异片段出现的频次,以及每个差异单元出现的频次;按照预设的数据挖掘算法,根据所述每个差异片段出现的频次和所述每个差异单元出现的频次,得到高频短语数据对,并将所述高频短语数据对作为所述易错短语对。

[0010] 可选地,所述利用所述易错短语对优化所述语音识别模型,包括:对于每一所述易错短语对,根据所述易错短语对对应的第二文本数据对应的语音数据,生成与所述第二短语数据对应的短语语音数据,并将所述短语语音数据作为输入,所述第一短语数据作为输出,通过模型训练来优化所述语音识别模型。

[0011] 可选地,所述利用所述易错短语对优化所述语音识别模型,包括:根据所述易错短语对确定待识别的热度词,并通过提高所述热度词的识别率以优化所述语音识别模型。

[0012] 可选地,所述利用所述易错短语对优化所述语音识别模型,包括:根据所述易错短语对训练文本纠错模型,并结合所述文本纠错模型来优化所述语音识别模型。

[0013] 可选地,所述使用优化后的语音识别模型进行语音识别之后,还包括:响应于用户意图识别请求,获取通过所述优化后的语音识别模型生成的用户文本数据;将所述用户文本数据输入至语言模型中,生成用户文本向量;根据所述用户文本向量,通过意图识别模型得到用户意图识别结果。

[0014] 可选地,所述语言模型是通过如下方式训练的:对于每一所述易错短语对,分别将所述易错短语对对应的第一短语数据和第二短语数据输入至所述语言模型,生成所述第一短语数据的文本向量和所述第二短语数据的文本向量;根据所述第一短语数据的文本向量和所述第二短语数据的文本向量,得到文本向量相似度的损失函数;以所述文本向量相似度的损失函数趋近于0为训练目标,训练所述语言模型。

[0015] 可选地,所述意图识别模型是通过如下方式训练的:对于每一所述高频短语数据对,将所述第二文本数据的文本向量作为输入,所述第一文本数据的意图识别结果作为输出,训练所述意图识别模型。

[0016] 根据本发明实施例的另一方面,提供了一种语音识别的装置。

[0017] 一种语音识别的装置,包括:文本数据生成模块,用于根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;短语数据对生成模块,用于对于每一所述第二文本数据,在所述第二文本数据与所述第一文本数据不一致的情况下,通过所述第一文本数据和所述第二文本数据生成短语数据对;语音识别模型优化模块,用于对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用所述易错短语对优化所述语音识别模型;语音识别模块,用于使用优化后的语音识别模型进行语音识别。

[0018] 可选地,所述短语数据对生成模块还用于:对所述第一文本数据和所述第二文本数据进行差异片段对齐处理,生成短语数据对,所述短语数据对包括:与所述第一文本数据对应的第一短语数据,和与所述第二文本数据对应的第二短语数据。

[0019] 可选地,所述语音识别模型优化模块还用于:对于每一所述短语数据对,生成所述短语数据对的两个差异片段,并将所述两个差异片段作为一个差异单元;统计所有短语数据对中每个差异片段出现的频次,以及每个差异单元出现的频次;按照预设的数据挖掘算法,根据所述每个差异片段出现的频次和所述每个差异单元出现的频次,得到高频短语数据对,并将所述高频短语数据对作为所述易错短语对。

[0020] 可选地,所述语音识别模型优化模块还用于:对于每一所述易错短语对,根据所述易错短语对对应的第二文本数据对应的语音数据,生成与所述第二短语数据对应的短语语音数据,并将所述短语语音数据作为输入,所述第一短语数据作为输出,通过模型训练来优化所述语音识别模型。

[0021] 可选地,所述语音识别模型优化模块还用于:根据所述易错短语对确定待识别的热度词,并通过提高所述热度词的识别率以优化所述语音识别模型。

[0022] 可选地,所述语音识别模型优化模块还用于:根据所述易错短语对训练文本纠错模型,并结合所述文本纠错模型来优化所述语音识别模型。

[0023] 可选地,还包括用户意图识别模块,用于:响应于用户意图识别请求,获取通过所述优化后的语音识别模型生成的用户文本数据;将所述用户文本数据输入至语言模型中,生成用户文本向量;根据所述用户文本向量,通过意图识别模型得到用户意图识别结果。

[0024] 可选地,还包括语言模型训练模块,用于:对于每一所述易错短语对,分别将所述易错短语对对应的第一短语数据和第二短语数据输入至所述语言模型,生成所述第一短语数据的文本向量和所述第二短语数据的文本向量;根据所述第一短语数据的文本向量和所述第二短语数据的文本向量,得到文本向量相似度的损失函数;以所述文本向量相似度的损失函数趋近于0为训练目标,训练所述语言模型。

[0025] 可选地,还包括意图识别模型训练模块,用于:对于每一所述高频短语数据对,将所述第二文本数据的文本向量作为输入,所述第一文本数据的意图识别结果作为输出,训练所述意图识别模型。

[0026] 根据本发明实施例的又一方面,提供了一种电子设备。

[0027] 一种电子设备,包括:一个或多个处理器;存储器,用于存储一个或多个程序,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本发明实施例所提供的语音识别的方法。

[0028] 根据本发明实施例的又一方面,提供了一种计算机可读介质。

[0029] 一种计算机可读介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现本发明实施例所提供的语音识别的方法。

[0030] 上述发明中的一个实施例具有如下优点或有益效果:通过根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;对于每一第二文本数据,在第二文本数据与第一文本数据不一致的情况下,通过第一文本数据和第二文本数据生成短语数据对;对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用易错短语对优化语音识别模型;使用优化后的语音识别模型进行语音识别的技术方案,能够通过挖掘异常数据得到高质量的训练数据,降低优化模型所耗费的成本,提高优化模型的效率,并且在使用时可以提高语音识别模型的准确率和效率,从而优化人机语言交互系统的效果,提高用户的使用体验。

[0031] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

附图说明

[0032] 附图用于更好地理解本发明,不构成对本发明的不当限定。其中:

[0033] 图1是根据本发明一个实施例的语音识别的方法的主要步骤示意图;

[0034] 图2是根据本发明一个实施例的语音识别的方法的流程示意图;

[0035] 图3是根据本发明一个实施例的人机语言交互系统的流程示意图;

[0036] 图4是根据本发明一个实施例的语音识别的装置的主要模块示意图;

[0037] 图5是本发明实施例可以应用于其中的示例性系统架构图;

[0038] 图6是适于用来实现本发明实施例的终端设备或服务器的计算机系统的结构示意图。

具体实施方式

[0039] 以下结合附图对本发明的示范性实施例做出说明,其中包括本发明实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本发明的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0040] 需要说明的是,本发明的技术方案中,所涉及的用户个人信息的采集、收集、更新、分析、处理、使用、传输、存储等方面,均符合相关法律法规的规定,被用于合法的用途,且不违背公序良俗。对用户个人信息采取必要措施,防止对用户个人信息数据的非法访问,维护用户个人信息安全、网络安全和国家安全。

[0041] 图1是根据本发明一个实施例的语音识别的方法的主要步骤示意图。

[0042] 如图1所示,本发明一个实施例的语音识别的方法主要包括如下的步骤S101至步骤S104。

[0043] 步骤S101:根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据。

[0044] 图2是根据本发明一个实施例的语音识别的方法的流程示意图。

[0045] 如图2所示,本发明实施例首先将第一文本数据 text_{in} (例如 text_{in} 可以为“领什么东西”)输入到TTS算法模型(一种将文本转化为语音的算法模型)中,选取不同的音色和不同的噪声大小来得到第一文本数据对应的多个语音数据,然后将每个语音数据输入到ASR模型(即语音识别模型)中来得到各自对应的文本句子作为第二文本数据 $\text{text}_{\text{out}}^i$,其中 i 用于区分不同的第二文本数据。通过第一文本数据和第二文本数据生成短语数据对,并对短语数据对进行数据挖掘得到易错短语对,从而优化语言识别模型并进行意图识别。

[0046] 步骤S102:对于每一第二文本数据,在第二文本数据与第一文本数据不一致的情况下,通过第一文本数据和第二文本数据生成短语数据对。

[0047] 在一个实施例中,通过第一文本数据和第二文本数据生成短语数据对,可以包括:对第一文本数据和第二文本数据进行差异片段对齐处理,生成短语数据对,短语数据对包括:与第一文本数据对应的第一短语数据,和与第二文本数据对应的第二短语数据。

[0048] 具体地,对于每一第二文本数据 $\text{text}_{\text{out}}^i$,对比该第二文本数据 $\text{text}_{\text{out}}^i$ 和对应的第一文本数据 text_{in} ,若该 $\text{text}_{\text{out}}^i$ 与 text_{in} 一致,例如 $\text{text}_{\text{out}}^i$ 为“领什么东西”,则说明ASR模型识别正确,若该 $\text{text}_{\text{out}}^i$ 与 text_{in} 不一致,例如 $\text{text}_{\text{out}}^i$ 可以为“宁什么东西”、“您什么东西”等,则说明ASR模型识别错误,通过该 $\text{text}_{\text{out}}^i$ 和 text_{in} 生成短语数据对 $\langle \text{text}_{\text{in}}, \text{text}_{\text{out}}^i \rangle$ 。由此可见,对于任一第一文本数据,可以得到多个对应的短语数据对,通过基于多个第一文本数据得到的短语数据对,生成短语数据对库。

[0049] 步骤S103:对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用易错短语对优化语音识别模型。

[0050] 在一个实施例中,对所有短语数据对进行数据挖掘处理,生成易错短语对,可以包括:对于每一短语数据对,生成短语数据对的两个差异片段,并将两个差异片段作为一个差异单元;统计所有短语数据对中每个差异片段出现的频次,以及每个差异单元出现的频次;按照预设的数据挖掘算法,根据每个差异片段出现的频次和每个差异单元出现的频次,得到高频短语数据对,并将高频短语数据对作为易错短语对。

[0051] 具体地,遍历短语数据对库中所有的短语数据对,对于每一短语数据对 $\langle \text{text}_{\text{in}}, \text{text}_{\text{out}}^i \rangle$,确定 text_{in} 和 $\text{text}_{\text{out}}^i$ 的差异文本,基于相邻两侧的共同片段或音节间相似性对差异文本进行对齐,得到差异片段,并将两个对齐后的差异片段作为一个差异单元,例如 $\langle \text{text}_{\text{in}}, \text{text}_{\text{out}}^i \rangle$ 为 $\langle$ 领什么东西,宁什么东西 $\rangle$,则 text_{in} 和 $\text{text}_{\text{out}}^i$ 的差异文本可以为“领”和“宁”,在对齐处理之后 text_{in} 和 $\text{text}_{\text{out}}^i$ 的差异片段可以为“领什么”和“宁什么”,差异单元为 $\langle$ 领什么,宁什么 $\rangle$ 。

[0052] 统计所有短语数据对中每个差异片段出现的频次,以及每个差异单元出现的频次,例如所有的差异单元包括 $\langle$ 领什么,宁什么 $\rangle$ 和 $\langle$ 领什么,您什么 $\rangle$,那么“领”、“宁”、“您”的频次分别为2、1、1,差异单元 $\langle$ 领什么,宁什么 $\rangle$ 和 $\langle$ 领什么,您什么 $\rangle$ 的频次分别为1、1。按照预设的数据挖掘算法,根据每个差异片段出现的频次和每个差异单元出现的频次,得到高频短语数据对,并将高频短语数据对作为易错短语对,其中,数据挖掘算法可以为FP Growth算法,FP Growth算法是一种以自底向上方式探索树,由FP树产生频繁项集的计算。

[0053] 在一个实施例中,利用易错短语对优化语音识别模型,可以包括:对于每一易错短语对,根据易错短语对对应的第二文本数据对应的语音数据,生成与第二短语数据对应的短语语音数据,并将短语语音数据作为输入,第一短语数据作为输出,通过模型训练来优化语音识别模型。

[0054] 具体地,对于每一易错短语对,根据易错短语对对应的第二文本数据对应的语音数据,例如易错短语对为 $\langle$ 领什么,宁什么 $\rangle$,则易错短语对对应的第二文本数据即为“宁什么东西”,可以根据“宁什么东西”的语音数据得到“宁什么”的语音数据,也可以直接生成“宁什么”的语音数据。通过将“宁什么”的语音数据作为输入,将“领什么”作为输出,构建训练数据并补充输入到语音识别模型中进行监督训练,从而让语音识别模型能够自动学习这些高频的、容易识别错误的异常案例的特征,以此提高语音识别模型的准确率。

[0055] 在另一个实施例中,可以直接将易错短语对对应的第二文本数据对应的语音数据作为输入,将第一文本数据作为输出,通过模型训练来优化语音识别模型。例如对于易错短语对 $\langle$ 领什么,宁什么 $\rangle$,可以将“宁什么东西”的语音数据作为语音模型的输入,将“领什么东西”作为语音模型的输出,通过模型训练来优化语音识别模型。

[0056] 在一个实施例中,利用易错短语对优化语音识别模型,可以包括:根据易错短语对确定待识别的热度词,并通过提高热度词的识别率以优化语音识别模型。

[0057] 具体地,在挖掘得到易错短语对之后,可以通过配置异常案例相关的热词功能来提高特定内容的识别率,并最终在一定程度上提高语音识别模型的准确率。例如对于易错短语对 $\langle$ 领什么,宁什么 $\rangle$,表示实际用户在询问“领什么”的时候,语音识别模型容易错误的识别为“宁什么”,此时可以将“领什么”作为待识别的热度词加入到语音识别模型的训练数据中,进而让语音识别模型提高对于“领什么”的识别概率,从而防止误识别情况。

[0058] 在一个实施例中,利用易错短语对优化语音识别模型,可以包括:根据易错短语对训练文本纠错模型,并结合文本纠错模型来优化语音识别模型。

[0059] 具体地,根据易错短语对训练文本纠错模型,通过开发相应的文本纠错功能来将语音识别模型的识别结果进行高频错误的文本纠正,例如将“宁”纠正为“领”,从而提高语音识别模型的准确率。

[0060] 步骤S104:使用优化后的语音识别模型进行语音识别。

[0061] 在一个实施例中,使用优化后的语音识别模型进行语音识别之后,还可以包括:响应于用户意图识别请求,获取通过优化后的语音识别模型生成的用户文本数据;将用户文本数据输入至语言模型中,生成用户文本向量;根据用户文本向量,通过意图识别模型得到用户意图识别结果。

[0062] 图3是根据本发明一个实施例的人机语言交互系统的流程示意图。

[0063] 如图3所示,首先由机器人发送开启外呼服务指令到调度中心执行TTS任务来将预先准备好的文本话术转换成语音,或者直接使用预先人工录好的音频数据,然后通过调度中心的电话系统开始给用户拨打电话并进行语音消息播报,调度中心的电话系统会记录用户的语音内容并将用户的语音内容输入到语音识别模块中将其转换成文本内容,并将文本内容发送至SLU(口语语言理解)模块中进行意图识别。将意图识别的结果发送至NLG(自然语言生成)模块中进行文本答案的生成,再由机器人将此文本答案依照前面的流程发送给用户,从而实现人机的多轮语音交互。其中,SLU模块中包括语言模型和意图识别模型,语言模型用于文本内容转换为文本向量,意图识别模型用于识别文本向量生成意图识别结果。本发明实施例中的语言识别模型可以预先通过易错短语进行离线优化,语言模型和意图识别模型也可以预先通过易错短语进行离线优化,从而提高语言识别模型的准确率和SLU模块的准确率,提升人机交互系统的智能化水平。

[0064] 在一个实施例中,语言模型可以通过如下方式训练的:对于每一易错短语对,分别将易错短语对对应的第一短语数据和第二短语数据输入至语言模型,生成第一短语数据的文本向量和第二短语数据的文本向量;根据第一短语数据的文本向量和第二短语数据的文本向量,得到文本向量相似度的损失函数;以文本向量相似度的损失函数趋近于0为训练目标,训练语言模型。

[0065] 具体地,将易错短语对中的两个短语数据输入到语言模型中,从而分别得到对应的文本向量,并计算两个文本向量的余弦相似度,然后通过深度学习中的梯度的反向传播算法来将相似度的损失函数趋近于0,进而优化语言模型,使得易错短语对中的两个短语数据对于语言模型能够具有相同的语义信息,从而在下游的意图识别任务中能够具备相似或者相同的意图类别。

[0066] 在一个实施例中,意图识别模型可以通过如下方式训练的:对于每一高频短语数据对,将第二文本数据的文本向量作为输入,第一文本数据的意图识别结果作为输出,训练意图识别模型。

[0067] 图4是根据本发明一个实施例的语音识别的装置的主要模块示意图。

[0068] 如图4所示,本发明一个实施例的语音识别的装置400主要包括:文本数据生成模块401、短语数据对生成模块402、语音识别模型优化模块403、语音识别模块404。

[0069] 文本数据生成模块401,用于根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据。

[0070] 短语数据对生成模块402,用于对于每一第二文本数据,在第二文本数据与第一文本数据不一致的情况下,通过第一文本数据和第二文本数据生成短语数据对。

[0071] 语音识别模型优化模块403,用于对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用易错短语对优化语音识别模型。

[0072] 语音识别模块404,用于使用优化后的语音识别模型进行语音识别。

[0073] 在一个实施例中,短语数据对生成模块401具体用于:对第一文本数据和第二文本数据进行差异片段对齐处理,生成短语数据对,短语数据对包括:与第一文本数据对应的第一短语数据,和与第二文本数据对应的第二短语数据。

[0074] 在一个实施例中,语音识别模型优化模块403具体用于:对于每一短语数据对,生成短语数据对的两个差异片段,并将两个差异片段作为一个差异单元;统计所有短语数据对中每个差异片段出现的频次,以及每个差异单元出现的频次;按照预设的数据挖掘算法,根据每个差异片段出现的频次和每个差异单元出现的频次,得到高频短语数据对,并将高频短语数据对作为易错短语对。

[0075] 在一个实施例中,语音识别模型优化模块403具体用于:对于每一易错短语对,根据易错短语对对应的第二文本数据对应的语音数据,生成与第二短语数据对应的短语语音数据,并将短语语音数据作为输入,第一短语数据作为输出,通过模型训练来优化语音识别模型。

[0076] 在一个实施例中,语音识别模型优化模块403具体用于:根据易错短语对确定待识别的热度词,并通过提高热度词的识别率以优化语音识别模型。

[0077] 在一个实施例中,语音识别模型优化模块403具体用于:根据易错短语对训练文本纠错模型,并结合文本纠错模型来优化语音识别模型。

[0078] 在一个实施例中,还可以包括用户意图识别模块(图中未示出),用于:响应于用户意图识别请求,获取通过优化后的语音识别模型生成的用户文本数据;将用户文本数据输入至语言模型中,生成用户文本向量;根据用户文本向量,通过意图识别模型得到用户意图识别结果。

[0079] 在一个实施例中,还包括语言模型训练模块(图中未示出),用于:对于每一易错短语对,分别将易错短语对对应的第一短语数据和第二短语数据输入至语言模型,生成第一短语数据的文本向量和第二短语数据的文本向量;根据第一短语数据的文本向量和第二短语数据的文本向量,得到文本向量相似度的损失函数;以文本向量相似度的损失函数趋近于0为训练目标,训练语言模型。

[0080] 在一个实施例中,还包括意图识别模型训练模块(图中未示出),用于:对于每一高频短语数据对,将第二文本数据的文本向量作为输入,第一文本数据的意图识别结果作为输出,训练意图识别模型。

[0081] 另外,在本发明实施例中语音识别的装置的具体实施内容,在上面语音识别的方法中已经详细说明了,故在此重复内容不再说明。

[0082] 图5示出了可以应用本发明实施例的语音识别的方法或语音识别的装置的示例性系统架构500。

[0083] 如图5所示,系统架构500可以包括终端设备501、502、503,网络504和服务器505。网络504用以在终端设备501、502、503和服务器505之间提供通信链路的介质。网络504可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0084] 用户可以使用终端设备501、502、503通过网络504与服务器505交互,以接收或发送消息等。终端设备501、502、503上可以安装有各种通讯客户端应用,例如语音识别类应用、人机交互应用、意图识别类应用、即时通信工具、邮箱客户端、社交平台软件等(仅为示例)。

[0085] 终端设备501、502、503可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0086] 服务器505可以是提供各种服务的服务器,例如对用户利用终端设备501、502、503所浏览的语音识别类网站提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的语音识别请求等数据进行根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;对于每一第二文本数据,在第二文本数据与第一文本数据不一致的情况下,通过第一文本数据和第二文本数据生成短语数据对;对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用易错短语对优化语音识别模型;使用优化后的语音识别模型进行语音识别等处理,并将处理结果(例如语音识别结果--仅为示例)反馈给终端设备。

[0087] 需要说明的是,本发明实施例所提供的语音识别的方法一般由服务器505执行,相应地,语音识别的装置一般设置于服务器505中。

[0088] 应该理解,图5中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0089] 下面参考图6,其示出了适于用来实现本发明实施例的终端设备或服务器的计算机系统600的结构示意图。图6示出的终端设备或服务器仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0090] 如图6所示,计算机系统600包括中央处理单元(CPU)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储部分608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中,还存储有系统600操作所需的各种程序和数据。CPU 601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0091] 以下部件连接至I/O接口605:包括键盘、鼠标等的输入部分606;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607;包括硬盘等的存储部分608;以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分609。通信部分609经由诸如因特网的网络执行通信处理。驱动器610也根据需要连接至I/O接口605。可拆卸介质611,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610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608。

[0092] 特别地,根据本发明公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本发明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609从网络上被下载和安装,和/或从可拆卸介质611被安装。在该计算机程序被中央处理单元(CPU)601执行时,执行本发明的系统中限定的上述功能。

[0093] 需要说明的是,本发明所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、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(EPR0M或闪存)、光纤、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。在本发明中,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。而在本发明中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0094] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、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。

[0095] 描述于本发明实施例中所涉及到的模块可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块也可以设置在处理器中,例如,可以描述为:一种处理器包括文本数据生成模块401、短语数据对生成模块402、语音识别模型优化模块403、语音识别模块。其中,这些模块的名称在某种情况下并不构成对该模块本身的限定,例如,文本数据生成模块还可以被描述为“用于根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据的模块”。

[0096] 作为另一方面,本发明还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读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一个该设备执行时,使得该设备包括:根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;对于每一第二文本数据,在第二文本数据与第一文本数据不一致的情况下,通过第一文本数据和第二文本数据生成短语数据对;对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用易错短语对优化语音识别模型;使用优化后的语音识别模型进行语音识别。

[0097] 根据本发明实施例的技术方案,根据第一文本数据对应的多个语音数据,通过语音识别模型生成多个第二文本数据;对于每一第二文本数据,在第二文本数据与第一文本数据不一致的情况下,通过第一文本数据和第二文本数据生成短语数据对;对所有短语数据对进行数据挖掘处理,生成易错短语对,并利用易错短语对优化语音识别模型;使用优化后的语音识别模型进行语音识别。能够通过挖掘异常数据得到高质量的训练数据,降低优化模型所耗费的成本,提高优化模型的效率,并且在使用时可以提高语音识别模型的准确

率和效率,从而优化人机语言交互系统的效果,提高用户的使用体验。

[0098] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,取决于设计要求和因素,可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

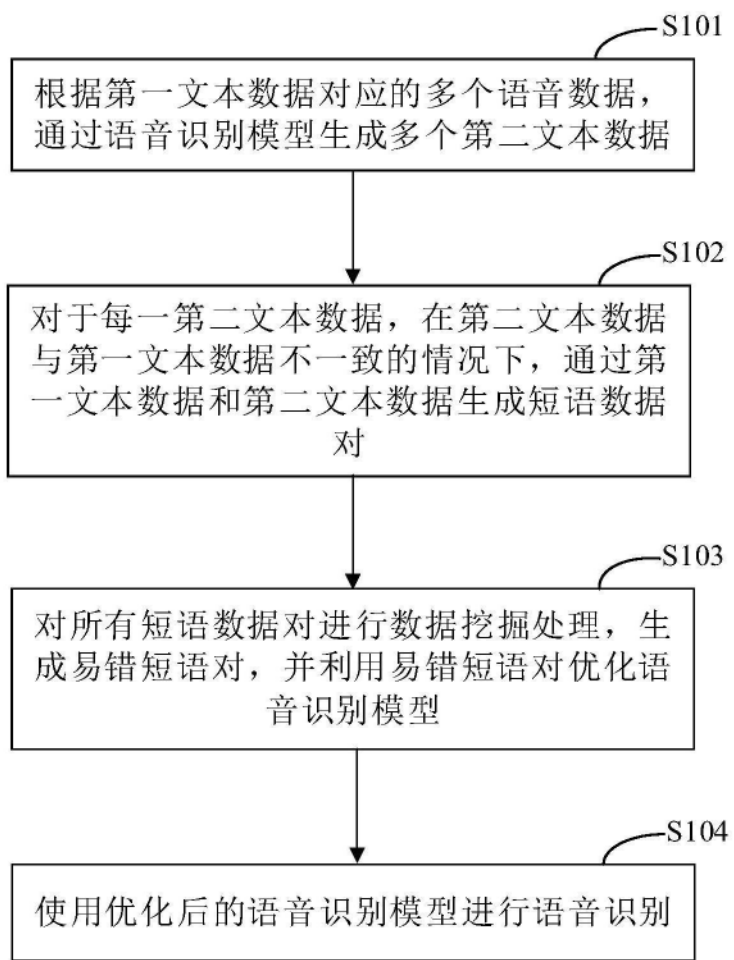


图1

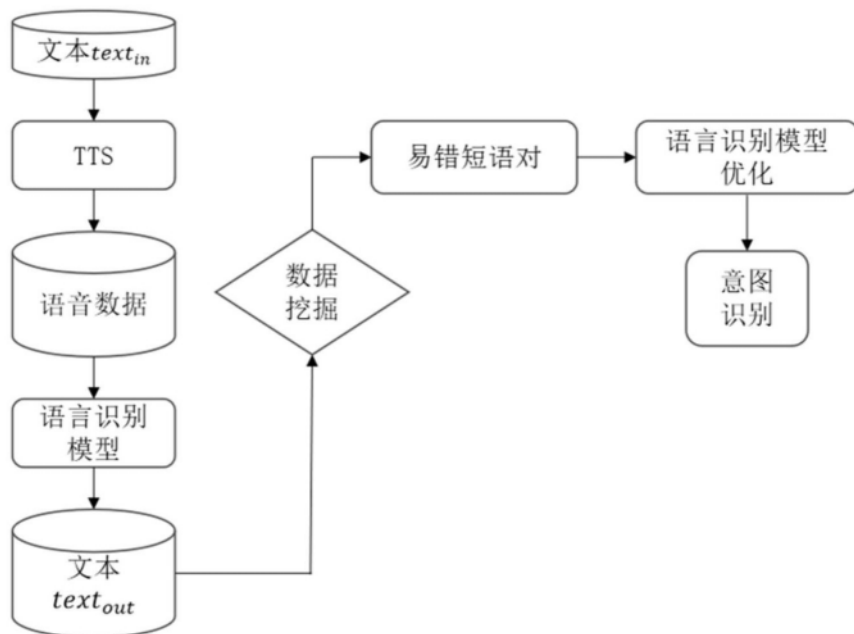


图2

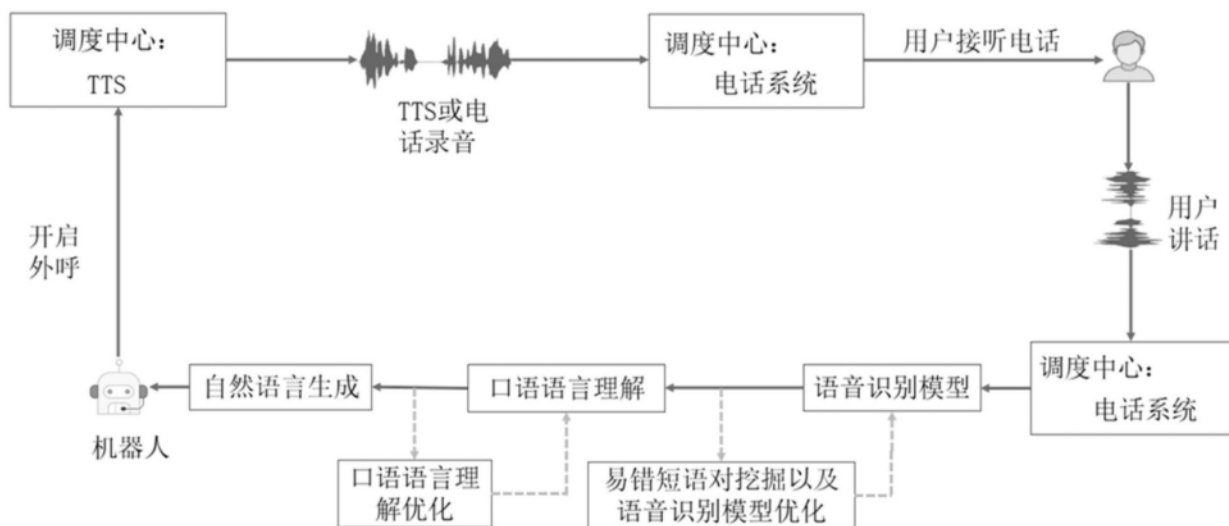


图3

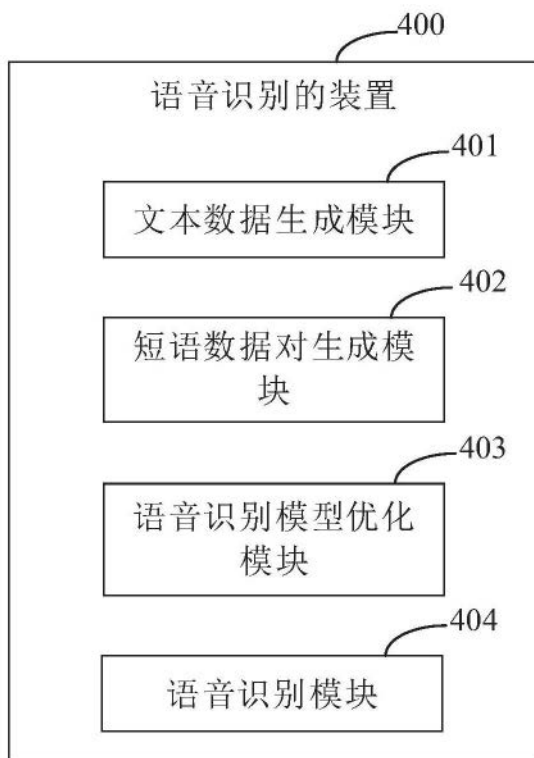


图4

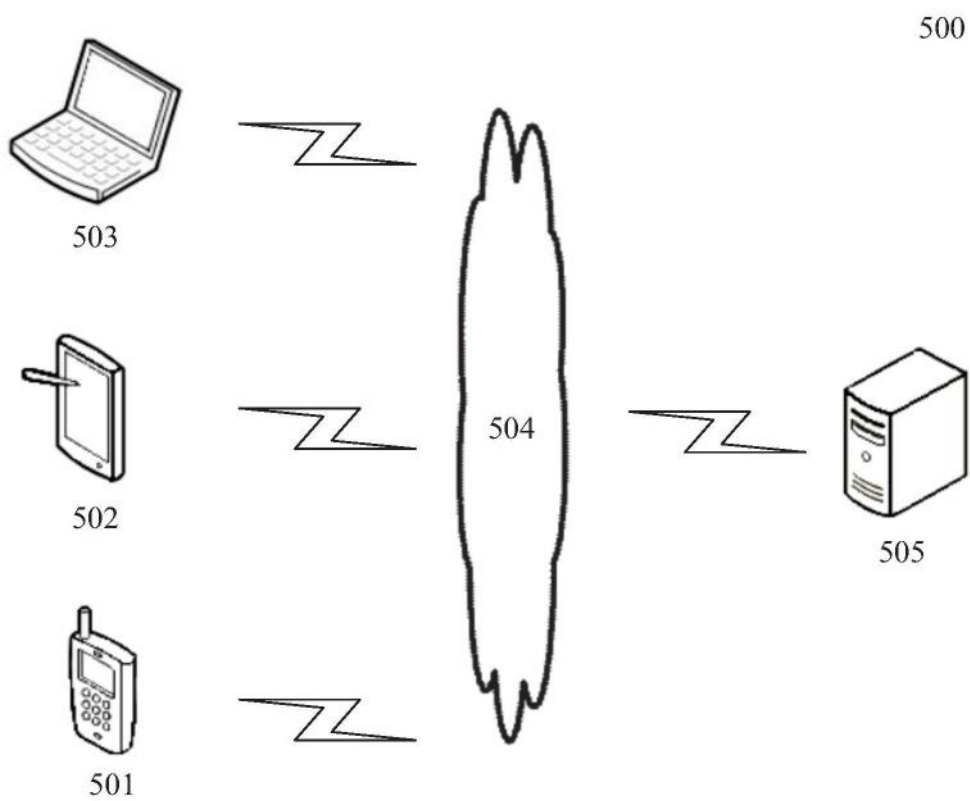


图5

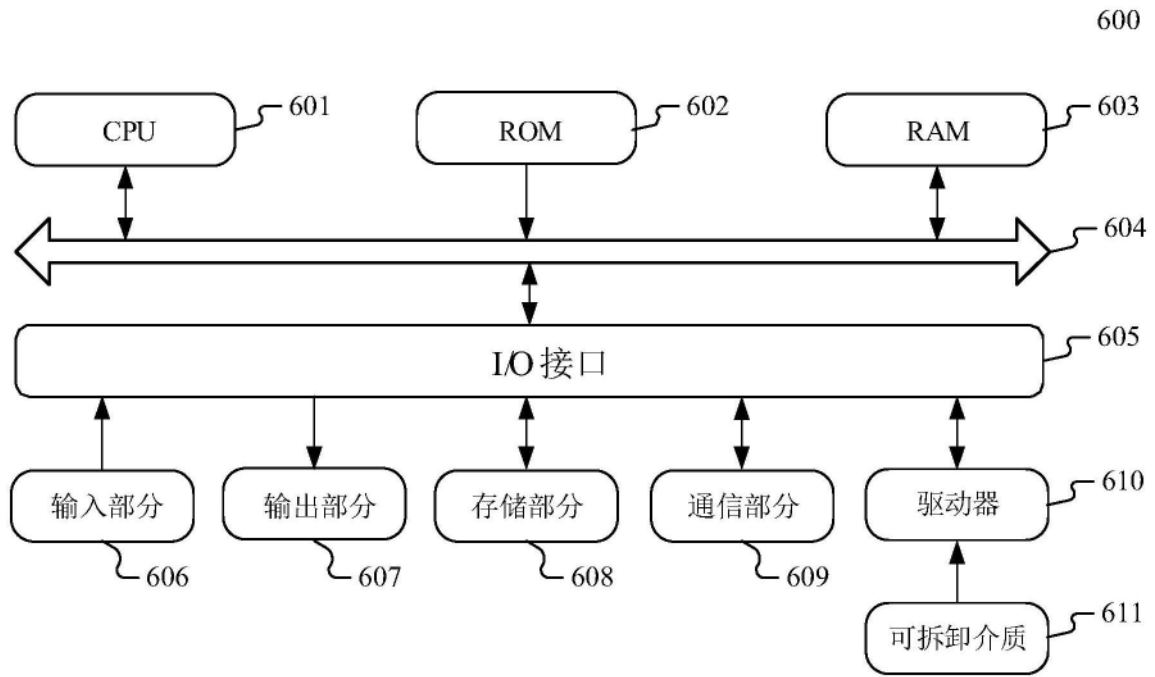


图6