



(12) 发明专利

(10) 授权公告号 CN 109754809 B

(45) 授权公告日 2021.02.09

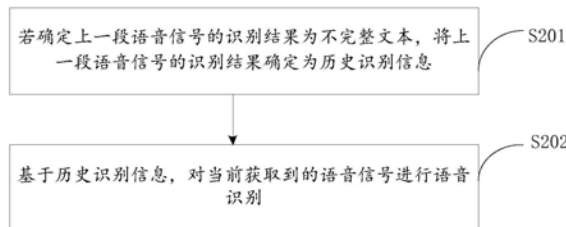
(21) 申请号 201910085677.2	G10L 15/183 (2013.01)
(22) 申请日 2019.01.29	G10L 15/02 (2006.01)
(65) 同一申请的已公布的文献号	G10L 15/06 (2013.01)
申请公布号 CN 109754809 A	G10L 15/08 (2006.01)
(43) 申请公布日 2019.05.14	G10L 15/14 (2006.01)
(73) 专利权人 北京猎户星空科技有限公司	G10L 15/30 (2013.01)
地址 100025 北京市朝阳区姚家园南路一	(56) 对比文件
号惠通时代广场8号	CN 107146618 A, 2017.09.08
(72) 发明人 李宝祥 钟贵平 李家魁	CN 107146618 A, 2017.09.08
(74) 专利代理机构 北京同达信恒知识产权代理	CN 102486801 A, 2012.06.06
有限公司 11291	CN 1226327 A, 1999.08.18
代理人 黄志华	CN 103035243 A, 2013.04.10
(51) Int. Cl.	CN 107146618 A, 2017.09.08
G10L 15/26 (2006.01)	CN 105244022 A, 2016.01.13
G10L 15/18 (2013.01)	WO 2007067878 A3, 2008.05.15
	审查员 薛彦杰
	权利要求书1页 说明书9页 附图3页

(54) 发明名称

语音识别方法、装置、电子设备及存储介质

(57) 摘要

本发明公开了一种语音识别方法、装置、电子设备及存储介质,所述方法包括:若确定上一段语音信号的识别结果为不完整文本,将上一段语音信号的识别结果确定为历史识别信息;基于历史识别信息,对当前获取到的语音信号进行语音识别。本发明实施例提供的技术方案,在确定上一段语音信号的识别结果不是完整文本后,将上一段语音信号的识别结果作为识别当前获取到的语音信号时的历史识别信息,在对当前获取到的语音信号计算语言模型得分时,增加了历史识别信息带来的影响,从而提升语音识别准确率。



1. 一种语音识别方法,其特征在于,包括:

若确定上一段语音信号的识别结果为不完整文本,将所述上一段语音信号的识别结果确定为历史识别信息;

计算当前获取到的语音信号对应的各条假设词序路径的概率得分,所述假设词序路径是基于所述历史识别信息对应的历史词序路径得到的;

根据概率得分最高的假设词序路径,确定所述当前获取到的语音信号的识别结果。

2. 根据权利要求1所述的方法,其特征在于,所述确定上一段语音信号的识别结果为不完整文本,包括:

对所述上一段语音信号的识别结果,进行断句处理;

若断句处理后的识别结果中包含的标点符号为预设标点符号,确定所述上一段语音信号的识别结果为不完整文本。

3. 根据权利要求1所述的方法,其特征在于,所述确定上一段语音信号的识别结果为不完整文本,包括:

对所述上一段语音信号的识别结果,进行语义解析;

根据语义解析结果,确定所述上一段语音信号的识别结果为不完整文本。

4. 根据权利要求1所述的方法,其特征在于,所述确定上一段语音信号的识别结果为不完整文本,包括:

对所述上一段语音信号的识别结果,进行句法分析;

若句法分析结果不符合预设句法模板,确定所述上一段语音信号的识别结果为不完整文本。

5. 根据权利要求1-4任一项所述的方法,其特征在于,还包括:

从所述历史识别信息对应的各条假设词序路径中,根据所述各条假设词序路径的概率得分,选择预设数量的假设词序路径,确定为所述历史识别信息对应的历史词序路径。

6. 根据权利要求5所述的方法,其特征在于,所述方法还包括:

根据所述概率得分最高的假设词序路径对应的历史词序路径,更新所述历史识别信息。

7. 一种语音识别装置,其特征在于,包括:

确定模块,用于若确定上一段语音信号的识别结果为不完整文本,将所述上一段语音信号的识别结果确定为历史识别信息;

识别模块,用于计算当前获取到的语音信号对应的各条假设词序路径的概率得分,所述假设词序路径是基于所述历史识别信息对应的历史词序路径得到的;

根据概率得分最高的假设词序路径,确定所述当前获取到的语音信号的识别结果。

8. 一种电子设备,包括收发机、存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,其特征在于,所述收发机用于在所述处理器的控制下接收和发送数据,所述处理器执行所述程序时实现权利要求1至6任一项所述方法的步骤。

9. 一种计算机可读存储介质,其上存储有计算机程序指令,其特征在于,该程序指令被处理器执行时实现权利要求1至6任一项所述方法的步骤。

## 语音识别方法、装置、电子设备及存储介质

### 技术领域

[0001] 本发明涉及语音识别技术领域,尤其涉及一种语音识别方法、装置、电子设备及存储介质。

### 背景技术

[0002] 语音识别是指通过机器学习等方法让机器能够自动的将语音转换成对应的文字,语音识别过程即基于训练好的声学模型,并结合词典、语言模型,对输入的语音帧序列识别的过程。语音识别结果的准确率影响语音交互方式的普及,如果语音识别结果的准确率过低,语音交互的方式就不可用。

[0003] 语言模型用于估计假设词序列的可能性。利用语言模型,可以确定哪个词序列的可能性更大,或者给定若干个词,可以预测下一个最可能出现的词语。例如,输入拼音串为 nixianzaiganshenme,对应的输出可以有多种形式,如“你现在干什么”、“你西安再赶什么”等,利用语言模型,就可知道前者的概率大于后者。因此,在对一段完整的语音进行识别时,语言模型能够基于上下文关系,从多种词序列中选出一个可能性最大的一个词序列。

[0004] 但是,当用户说话习惯性的停顿时,会将同一段语言拆分为两段语音进行识别,例如,用户发出的语音为“我来昊天、、、星空面试”,由于“昊天”和“星空”之间存在足够长度的静音帧,此时会将“我来昊天”和“星空面试”分成两段语音分别进行识别,因此,会先对第一段语音进行识别,得到识别结果“我来昊天”,在识别第二段语音时,会得到多个序列,如“清空面试”、“星空面试”,语言模型会输出概率较高的“清空面试”,导致语音识别结果的准确率过低。

### 发明内容

[0005] 本发明实施例提供一种语音识别方法、装置、电子设备及存储介质,以解决现有技术中语音识别准确率较低的问题。

[0006] 第一方面,本发明一实施例提供了一种语音识别方法,包括:

[0007] 若确定上一段语音信号的识别结果为不完整文本,将上一段语音信号的识别结果确定为历史识别信息;

[0008] 基于历史识别信息,对当前获取到的语音信号进行语音识别。

[0009] 第二方面,本发明一实施例提供了一种语音识别装置,包括:

[0010] 确定模块,用于若确定上一段语音信号的识别结果为不完整文本,将上一段语音信号的识别结果确定为历史识别信息;

[0011] 识别模块,用于基于历史识别信息,对当前获取到的语音信号进行语音识别。

[0012] 第三方面,本发明一实施例提供了一种电子设备,包括收发机、存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,其中,收发机用于在控制器的控制下接收和发送数据,处理器执行程序时实现上述任一种方法的步骤。

[0013] 第四方面,本发明一实施例提供了一种计算机可读存储介质,其上存储有计算机

程序指令,该程序指令被处理器执行时实现上述任一种方法的步骤。

[0014] 本发明实施例提供的技术方案,在识别当前获取到的语音信号前,先判断上一段语音信号的识别结果是否为完整文本,在确定上一段语音信号的识别结果不是完整文本后,将上一段语音信号的识别结果作为识别当前获取到的语音信号时的历史识别信息,在对当前获取到的语音信号计算语言模型得分时,增加了历史识别信息带来的影响,使得与历史识别信息关联度更高的假设词序路径的概率得分高于其它关联度较低的假设词序路径的概率得分,进而从当前获取到的语音信号对应的多个假设词序路径中找出与历史识别信息匹配度最高的假设词序路径,作为当前获取到的语音信号的识别结果,提高语音识别的准确率。

## 附图说明

[0015] 为了更清楚地说明本发明实施例的技术方案,下面将对本发明实施例中所需要使用的附图作简单地介绍,显而易见地,下面所介绍的附图仅仅是本发明的一些实施例,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。

[0016] 图1为本发明实施例提供的语音识别方法的应用场景示意图;

[0017] 图2为本发明一实施例提供的语音识别方法的流程示意图;

[0018] 图3为本发明一实施例提供的语音识别方法的流程又一示意图;

[0019] 图4为本发明一实施例提供的语音识别装置的结构示意图;

[0020] 图5为本发明一实施例提供的电子设备的结构示意图。

## 具体实施方式

[0021] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述。

[0022] 为了方便理解,下面对本发明实施例中涉及的名词进行解释:

[0023] 语言模型(Language Model,LM)的目的是建立一个能够描述给定词序列在语言中的出现的概率的分布。也就是说,语言模型是描述词汇概率分布的模型,一个能可靠反应语言识别时用词的概率分布的模型。语言模型在自然语言处理中占有重要的地位,在语音识别、机器翻译等领域得到了广泛应用。例如,利用语言模型能够得到语音信号对应的多种假设词序列中可能性最大的一个词序列,或者给定若干词,预测下一个最可能出现的词语等。常用的语言模型包括N-Gram LM(N元语言模型)、Big-Gram LM(二元语言模型)、Tri-Gram LM(三元语言模型)。

[0024] 音素(phone),是语音中的最小的单位,依据音节里的发音动作来分析,一个动作构成一个音素。汉语中的音素分为声母、韵母两大类,例如,声母包括:b、p、m、f、d、t、等,韵母包括:a、o、e、i、u、ü、ai、ei、ao、an、ian、ong、iong等。英语中的音素分为元音、辅音两大类,例如,元音有a、e、ai等,辅音有p、t、h等。

[0025] 声学模型(AM,Acoustic model)是语音识别系统中最为重要的部分之一,是把语音的声学特征分类对应到音素的模型。

[0026] 词典是字词对应的音素集合,描述了字词和音素之间的映射关系。

[0027] 附图中的任何元素数量均用于示例而非限制,以及任何命名都仅用于区分,而不具有任何限制含义。

[0028] 在具体实践过程中,现有的语音识别方法的准确率较低,尤其是当用户说话习惯性的停顿时,会将同一段语言拆分为两段语音进行识别,例如,用户发出的语音为“我来昊天、、、、星空面试”,由于“昊天”和“星空”之间存在足够长度的静音帧,此时会将“我来昊天”和“星空面试”分成两段语音分别进行识别,因此,会先对第一段语音得到识别结果“我来昊天”,识别第二段语音时会得到多个序列,如“清空面试”、“星空面试”,语言模型会输出概率较高的“清空面试”,导致语音识别结果的准确率过低。

[0029] 为此,本发明的发明人考虑到,在识别当前获取到的语音信号前,先判断上一段语音信号的识别结果是否为完整文本,在确定上一段语音信号的识别结果不是完整文本后,将上一段语音信号的识别结果作为识别当前获取到的语音信号时的历史识别信息,在对当前获取到的语音信号计算语言模型得分时,增加了历史识别信息带来的影响,使得与历史识别信息关联度更高的假设词序路径的概率得分高于其它关联度较低的假设词序路径的概率得分,进而从当前获取到的语音信号对应的多个假设词序路径中找出与历史识别信息匹配度最高的假设词序路径,作为当前获取到的语音信号的识别结果,提高语音识别的准确率。

[0030] 在介绍了本发明的基本原理之后,下面具体介绍本发明的各种非限制性实施方式。

[0031] 首先参考图1,其为本发明实施例提供的语音识别方法的应用场景示意图。用户10与智能设备11交互过程中,智能设备11将用户10输入的语音信号发送给服务器12,服务器12通过语音识别方法进行语音信号识别,将语音信号的识别结果反馈给智能设备11。

[0032] 这种应用场景下,智能设备11和服务器12之间通过网络进行通信连接,该网络可以为局域网、广域网等。智能设备11可以为智能音箱、机器人等,也可以为便携设备(例如:手机、平板、笔记本电脑等),还可以为个人电脑(PC,Personal Computer),服务器12可以为任何能够提供语音识别服务的服务器设备。

[0033] 下面结合图1所示的应用场景,对本发明实施例提供的技术方案进行说明。

[0034] 参考图2,本发明实施例提供了一种语音识别方法,包括以下步骤:

[0035] S201、若确定上一段语音信号的识别结果为不完整文本,将上一段语音信号的识别结果确定为历史识别信息。

[0036] 具体实施时,可通过多种方式确定上一段语音信号的识别结果是否为不完整文本,下面介绍本发明实施例采用的三种实施方式:

[0037] 第一种方式、预测识别结果对应的标点符号,确定识别结果是否为不完整文本。

[0038] 具体地,通过如下步骤确定上一段语音信号的识别结果是否为不完整文本:对上一段语音信号的识别结果,进行断句处理;若断句处理后的识别结果中包含的标点符号为预设标点符号,确定上一段语音信号的识别结果为不完整文本,否则,确定上一段语音信号的识别结果为完整文本。

[0039] 具体实施时,预设标点符号可以包括句号、分号、感叹号、问号等表示一句话已经结束的标点符号。如果通过断句处理得到多个标点符号,选取识别结果结尾处的标点符号与预设标点符号进行比对,若识别结果结尾处的标点符号为预设标点符号,则确定该识别

结果为不完整文本,否则,确定该识别结果为完整文本。

[0040] 具体实施时,可通过标点预测模型对识别结果进行断句处理,得到识别结果对应的标点符号。标点预测模型是一种能够自动为文本标注标点符号的模型。例如,现有的标点预测模型可通过条件随机场(CRF,conditional random field algorithm)算法实现,其主要通过建立概率模型来进行标点预测,标点预测模型为现有技术,不再赘述。

[0041] 第二种方式、通过语义分析确定识别结果是否为不完整文本。

[0042] 具体地,通过如下步骤确定上一段语音信号的识别结果是否为不完整文本:对上一段语音信号的识别结果,进行语义解析;根据语义解析结果,确定上一段语音信号的识别结果是否为不完整文本。

[0043] 具体实施时,可通过NLP(Natural Language Processing,自然语言处理)方法对识别结果进行语义解析,若语义解析结果中不包括识别结果对应的意图(intent),则确定上一段语音信号的识别结果为不完整文本,若语义解析结果中包括意图,则根据语义解析结果中的其它信息进一步判断上一段语音信号的识别结果是否为完整文本。以语义解析结果中的槽位(slot)信息为例,若语义解析结果中包含识别出的意图对应的所有槽位信息,则确定上一段语音信号的识别结果为完整文本,否则确定上一段语音信号的识别结果为不完整文本。其中,意图是用户通过交互输入所要表达的目的,槽位信息是将用户意图转化为明确用户指令所需要补充的信息,每个意图对应的槽位信息可根据实际应用场景进行配置,只有获取到意图对应的所有槽位信息后,才能将根据槽位信息将用户意图转化为明确用户指令。

[0044] 举例说明,上一段语音信号的识别结果为“我来”,显然用户还没有表述清楚自己的意图,此时无法识别到“我来”对应的意图,表明上一段语音信号的识别结果为不完整文本。上一段语音信号的识别结果为“我想听刘德华的”,通过语义解析可得到用户的意图为听音乐,得到的槽位信息包括“刘德华”,根据解析出的槽位信息判断还缺少必要的槽位信息,如歌名,确定上一段语音信号的识别结果为不完整文本。

[0045] 第三种方式、通过句法分析确定识别结果是否为不完整文本。

[0046] 具体地,通过如下步骤确定上一段语音信号的识别结果是否为不完整文本:对上一段语音信号的识别结果,进行句法分析;若句法分析结果不符合预设句法模板,确定上一段语音信号的识别结果为不完整文本,否则,确定上一段语音信号的识别结果为完整文本。

[0047] 具体实施时,识别上一段语音信号的识别结果中各个词的词性,根据识别出的各个词的词性对上一段语音信号的识别结果进行句法分析,确定上一段语音信号的识别结果对应的句式结构;若上一段语音信号的识别结果对应的句式结构符合预设句法模板,则确定上一段语音信号的识别结果为完整文本,否则,确定上一段语音信号的识别结果为不完整文本。

[0048] 汉语中的词可以分为两类14种词性。一类是实词,包括:名词、动词、形容词、区别词、代词、数词、量词;一类是虚词,包括:副词、介词、连词、助词、语气词、拟声词、叹词。本实施例中,可仅标注常用的名词、动词、形容词、形容词、副词等。

[0049] 具体实施时,可先对上一段语音信号的识别结果进行分词处理,可利用分词算法(如jieba分词算法等)实现分词处理。然后,基于字符串匹配的字典查找算法或基于统计的算法,标注识别结果中各个词的词性。其中,基于字符串匹配的字典查找算法即从字典中查

找每个词的词性,对每个词进行标注,基于统计的算法即通过HMM隐马尔科夫模型来进行词性标注。接着,通过对标注好词性的识别结果进行句法分析,确定识别结果对应的句式结构,最后,将识别结果的句式结构与预设句法模板进行比对,若识别结果对应的句式结构符合预设句法模板,则确定识别结果为完整文本,否则,确定识别结果为不完整文本。句法分析为现有技术,例如,可采用哈工大LTP或斯坦福句法分析工具Stanford Parser,不再赘述。

[0050] 具体实施时,预设句法模板包括但不限于以下类型:主语+谓语+宾语、谓语+宾语等。预设句法模板可根据实际应用场景进行配置。假设语音信号的识别结果为“播放音乐”,则分词结果为“播放”、“音乐”,词性标注结果为“播放(动词)”、“音乐(名词)”,句式分析结果为谓语+宾语(“播放”为谓语,“音乐”为宾语),在预设句法模板内,因此,识别结果“播放音乐”为完整文本。例如,语音信号的识别结果为“我要听”,则分词结果为“我”、“要”、“听”,词性标注结果为“我(名词)”、“要(助动词)”、“听(动词)”,句式分析结果为主语+谓语,该句式结构不在预设句法模板内,因此,识别结果“我要听”是不完整文本。

[0051] 如果上一段语音信号的识别结果为完整文本,表示上一段语音信号和当前获取到的语音信号分别属于两句话,则直接对当前获取到的语音信号进行识别,无需基于上一段语音信号的识别结果进行识别。

[0052] S202、基于历史识别信息,对当前获取到的语音信号进行语音识别。

[0053] 具体实施时,步骤S202具体包括以下步骤:计算当前获取到的语音信号对应的各条假设词序路径的概率得分,假设词序路径是基于历史识别信息对应的历史词序路径得到的;根据概率得分最高的假设词序路径,确定当前获取到的语音信号的识别结果。

[0054] 本实施例中,假设词序列是指语音信号对应的音素序列可能对应的词序列。语音识别过程大致为:对语音信号进行预处理,提取语音信号的声学特征向量,然后,将声学特征向量输入声学模型,得到音素序列,例如,“nixianzaiganshenme”,接着,基于语言模型和词典,得到音素序列对应的多个假设词序列中可能性最大的一个词序列,例如,音素序列“nixianzaiganshenme”可能对应多个假设词序列,如{你-现在-干-什么}、{你-现在-赶-什么}、{你-西安-在-干-什么}、{你-先-在-干-神-么}等。具体地,语音信号对应的一个假设词序列对应解码网络中的一条假设词序路径,在基于语言模型和词典构建的解码网络中,搜索与音素序列最匹配的一条假设词序路径,该假设词序路径对应的假设词序列即为语音信号对应的识别结果。假设词序路径的概率得分表征其对应的该假设词序列出现的概率,具体地,可通过以下公式计算假设词序路径的概率得分: $Score = \sum_{j \in L} \log SL_j$ ,其中,L为词序列在解码网络中对应的路径, $SL_j$ 为路径L上的第j个词的概率得分, $SL_j = P(W_j | W_{j-1})$ ,即根据语言模型得到的第j-1个词之后出现第j个词的概率,当j=1时, $SL_1 = P(W_1)$ 表示路径L上的第1个词作为词序列中的第一个词出现的概率。以Big-Gram语言模型为例,词序列{你-现在-干-什么}对应的概率得分为 $(\log P(\text{你}) + \log P(\text{现在} | \text{你}) + \log P(\text{干} | \text{现在}) + \log P(\text{什么} | \text{干}))$ 。

[0055] 例如,历史识别信息对应的历史词序路径为 $\{W_1-W_2-W_3\}$ ,其概率得分为 $A_1$ 。基于历史词序路径 $\{W_1-W_2-W_3\}$ ,得到当前获取到的语音信号对应的假设词序路径包括 $\{W_4-W_5\}$ 、 $\{W_6-W_7-W_8\}$ 。以Big-Gram语言模型为例,基于历史词序路径 $\{W_1-W_2-W_3\}$ , $\{W_4-W_5\}$ 的概率得分为 $A'_1 = P(W_4 | W_3) + P(W_5 | W_4)$ , $\{W_6-W_7-W_8\}$ 的概率得分为 $A'_2 = P(W_6 | W_3) + P(W_7 | W_6) + P(W_8 | W_7)$ 。在没有历

史识别信息的情况下,  $\{W_4-W_5\}$  的概率得分为  $A_1=P(W_4)+P(W_5|W_4)$ ,  $\{W_6-W_7-W_8\}$  的概率得分为  $A_2=P(W_6)+P(W_7|W_6)+P(W_8|W_7)$ 。假设  $\{W_1-W_2-W_3\}$  和  $W_4$  的关联度远大于  $\{W_1-W_2-W_3\}$  和  $W_6$  的关联度, 则  $P(W_4|W_3)$  要远高于  $P(W_6|W_3)$ , 因此, 即使  $A_1$  小于  $A_2$ , 由于增加了历史词序路径带来的影响,  $A'_1$  会大于  $A'_2$ , 从而针对当前获取到的语音信号得到更加准确的识别结果  $\{W_4-W_5\}$ , 将  $\{W_4-W_5\}$  作为当前获取到的语音信号的识别结果。

[0056] 例如, 用户想表达“我想听刘德华的忘情水”, 当说到“刘德华的”时候犹豫了一下, 因此, 语音识别时, 将“我想听刘德华的忘情水”截取为两段语音信号, 分别是: “我想听刘德华的”和“忘情水”。在语音识别时, 先识别出上一段语音信号“我想听刘德华的”, 在识别“忘情水”时, 识别出文本“我想听刘德华的”为不完整文本, 因此, 将“我想听刘德华的”作为历史识别信息, 由于在语言模型中, “刘德华”和“忘情水”这两个词的关联度较高, 因此, 在识别“忘情水”这段语音信号时, “我想听刘德华的忘情水”这一词序列的概率得分要高于“我想听刘德华的”与其他词组成的词序列的概率得分。而如果没有“我想听刘德华的”作为历史识别信息, 则“忘情水”的概率得分可能会低于其他词。

[0057] 又比如, 当用户说话习惯性的停顿时, 用户发出的语音为“我来昊天、、、、星空面试”, 由于“昊天”和“星空”之间存在足够长度的静音帧, 此时会将“我来昊天”和“星空面试”分成两段语音信号分别进行识别, 因此, 先识别第一段语音信号, 得到的识别结果为“我来昊天”, 识别第二段语音信号时会得到多个假设词序路径, 如“清空面试”、“星空面试”, 假设“清空面试”的概率得分更高, 则会将“清空面试”作为第二段语音的识别结果, 导致最终得到的识别结果错误。采用本发明实施例的方法后, 在识别出第一段语音信号为“我来昊天”后, 判断“我来昊天”为不完整文本, 此时将“我来昊天”作为历史识别信息, 在识别第二段语音信号时, 由于语言模型学习过“昊天星空”这个实体词, 因此, 在基于历史识别信息“我来昊天”搜索路径时“星空面试”的概率得分会高于“清空面试”, 因此, 将“星空面试”作为第二段语音信号的识别结果。

[0058] 本实施例的语音识别方法, 在识别当前获取到的语音信号前, 先判断上一段语音信号的识别结果是否为完整文本, 在确定上一段语音信号的识别结果不是完整文本后, 将上一段语音信号的识别结果作为识别当前获取到的语音信号时的历史识别信息, 在对当前获取到的语音信号计算语言模型得分时, 增加了历史识别信息带来的影响, 使得与历史识别信息关联度更高的假设词序路径的概率得分高于其它关联度较低的假设词序路径的概率得分, 进而从当前获取到的语音信号对应的多个假设词序路径中找出与历史识别信息匹配度最高的假设词序路径, 作为当前获取到的语音信号的识别结果, 提高语音识别的准确率。

[0059] 实际应用中, 假设用户输入的语音为“我来昊天、、、、星空面试、、、、我是张三”, 语音识别时, 分成三段语音“我来昊天”“星空面试”“我是张三”。在识别“星空面试”时, 由于上一句“我来昊天”是不完整文本, 因此, 将“我来昊天”作为识别语音信号“星空面试”时的历史识别信息, 得到正确识别结果“星空面试”。当识别“我是张三”时, 上一句“星空面试”是不完整文本, 但是实际上, “我来昊天星空面试”是完整文本, 而“我是张三”与“我来昊天星空面试”分属两个句子, 若将继续将“星空面试”作为“我是张三”的历史识别信息, 有可能会导导致识别结果发生错误。

[0060] 为此, 具体实施时, 在确定上一段语音信号的识别结果是否为不完整文本时, 可基



于历史识别信息和上一段语音信号的识别结果,来确定上一段语音信号的识别结果是否为不完整文本,即合并历史识别信息和上一段语音信号的识别结果,确定合并后的文本是否为不完整文本。具体实施时,可通过上述实施例中的三种实施方式来确定合并后的文本是否为不完整文本,若确定合并后的文本为不完整文本,将上一段语音信号的识别结果确定为历史识别信息,基于历史识别信息,对当前获取到的语音信号进行语音识别;若确定合并后的文本为完整文本,则直接对当前获取到的语音信号进行识别,同时,可清空历史识别信息。

[0061] 例如,在识别语音信号“星空面试”时,由于上一段语音信号的识别结果“我来昊天”为不完整文本,因此,将其作为历史识别信息,基于历史识别信息对语音信号“星空面试”进行识别。然后,在识别下一段语音信号“我是张三”时,历史识别信息“我来昊天”和上一段语音信号的识别结果“星空面试”合并成一个文本“我来昊天星空面试”,判断“我来昊天星空面试”为完整文本,因此,无需使用历史识别信息,直接对语音信号“我是张三”进行识别,同时,清空历史识别信息“我来昊天”,防止其干扰后续的语音识别。

[0062] 实际应用中,通过语言模型可得到语音信号对应的多个假设词序路径的概率得分,然后选取概率得分最高的假设词序路径,作为该语音信号的识别结果。由于,一个完整的句子可能会因为用户说话过程中的停顿,被分成两段语音,这会导致前后两段语音信息的识别结果都产生误差。为此,在图2所示的语音识别方法的基础上,本发明实施例还提供了另一种语音识别方法,如图3所示,包括以下步骤:

[0063] S301、若确定上一段语音信号的识别结果为不完整文本,将上一段语音信号的识别结果确定为历史识别信息。

[0064] 步骤S301的具体实施方式可参考步骤S201,不再赘述。

[0065] S302、从历史识别信息对应的各条假设词序路径中,根据各条假设词序路径的概率得分,选择预设数量的假设词序路径,确定为历史识别信息对应的历史词序路径。

[0066] 具体实施时,预设数量可根据实际需求确定,此处不做限定。

[0067] 具体实施时,按照路径的概率得分从大到小,将历史识别信息对应的各条假设词序路径进行排序,选择前预设数量的假设词序路径,确定为历史识别信息对应的历史词序路径。

[0068] S303、计算当前获取到的语音信号对应的各条假设词序路径的概率得分,假设词序路径是基于历史识别信息对应的历史词序路径得到的。

[0069] 具体的,基于S302中每条历史词序路径,计算当前获取到的语音信号对应的各条假设词序路径的概率得分。

[0070] S304、根据概率得分最高的假设词序路径,确定当前获取到的语音信号的识别结果。

[0071] 具体的,根据S303中计算得到的各条假设词序路径的概率得分,选择概率得分最高的假设词序路径,确定当前获取到的语音信号的识别结果。

[0072] 进一步的,该方法还包括如下步骤:

[0073] S305、根据概率得分最高的假设词序路径对应的历史词序路径,更新历史识别信息。

[0074] 举例说明,假设确定出的历史识别信息对应的历史词序路径为 $\{W_1-W_2-W_3\}$ 和 $\{W_4-$

$W_5$ }, 基于历史词序路径得到当前获取到的语音信号对应的假设词路径包括  $\{W_6-W_7-W_8\}$  和  $\{W_9-W_{10}\}$ ,  $\{W_6-W_7-W_8\}$  的概率得分为  $A_3$ ,  $\{W_9-W_{10}\}$  的概率得分为  $A_4$ 。以 Big-Gram 语言模型为例, 基于历史词序路径  $\{W_1-W_2-W_3\}$ ,  $\{W_6-W_7-W_8\}$  的概率得分为  $A'_1 = P(W_6|W_3) + P(W_7|W_6) + P(W_8|W_7)$ ; 基于历史词序路径  $\{W_1-W_2-W_3\}$ ,  $\{W_9-W_{10}\}$  的概率得分为  $A'_2 = P(W_9|W_3) + P(W_{10}|W_9)$ ; 基于历史词序路径  $\{W_4-W_5\}$ ,  $\{W_6-W_7-W_8\}$  的概率得分为  $A''_1 = P(W_6|W_5) + P(W_7|W_6) + P(W_8|W_7)$ ; 基于历史词序路径  $\{W_4-W_5\}$ ,  $\{W_9-W_{10}\}$  的概率得分为  $A''_2 = P(W_9|W_5) + P(W_{10}|W_9)$ 。在没有历史识别信息的情况下,  $\{W_6-W_7-W_8\}$  的概率得分为  $A_1 = P(W_6) + P(W_7|W_6) + P(W_8|W_7)$ ,  $\{W_9-W_{10}\}$  的概率得分为  $A_2 = P(W_9) + P(W_{10}|W_9)$ 。假设  $\{W_1-W_2-W_3\}$  和  $W_6$  的关联度远大于其他组合的关联度, 则  $P(W_6|W_3)$  要远高于  $P(W_9|W_3)$ 、 $P(W_6|W_5)$ 、 $P(W_9|W_5)$ , 因此, 即使  $A_1$  小于  $A_2$ , 由于增加了历史词序路径带来的影响,  $A'_1$  会大于  $A'_2$ 、 $A''_1$  和  $A''_2$ , 则概率得分最高的假设词序路径为  $\{W_6-W_7-W_8\}$ , 将  $\{W_6-W_7-W_8\}$  确定为当前获取到的语音信号的识别结果。进一步地, 假设在识别上一段语音信号时,  $\{W_4-W_5\}$  的概率得分最高, 则在识别当前获取到的语音信号前, 上一段语音信号的识别结果为  $\{W_4-W_5\}$ ; 在识别当前获取到的语音信号过程中, 最高概率得分  $A'_1$  对应的历史词序路径为  $\{W_1-W_2-W_3\}$ , 将上一段语音信号的识别结果更新为  $\{W_1-W_2-W_3\}$ , 实现基于当前获取到的语音信号对上一段语音信号的识别结果进行更新。

[0075] 例如, 第一段语音信号“我来昊天”对应的假设词序路径包括“我来昊天”、“我来航天”等, 将“我来昊天”、“我来航天”作为历史识别信息, 在识别“星空面试”时, 会基于历史识别信息计算概率得分, 此时, 由于语言模型学习过“昊天星空”这个词, 所以, 即便第一段语音信号的识别结果为“我来航天”, 在识别第二段语音信号“星空面试”时, “我来昊天星空面试”的概率得分会高于“我来航天星空面试”的概率得分, 因此, 能够将第一段语音信号的识别结果更新为“我来昊天”。

[0076] 本发明实施例的语音识别方法, 保留上一段语音信号的识别结果中概率得分较高的预设数量个假设词序路径作为历史识别信息, 在识别当前获取到的语音信号时, 结合多个历史识别信息, 可以基于上一段语音信号对应的多个历史词序路径和当前获取到的语音信号对应的假设词序路径得到各种可能的词序路径, 在上一段语音信号和当前获取到的语音信号的相互影响下, 从各种可能的词序路径中选取概率得分最高的词序路径作为最终的识别结果, 不仅提高了识别当前语音的准确率, 还能够对上一段语音信号的识别结果进行更新。

[0077] 本发明实施例的语音识别方法, 可以由智能设备内的控制器执行, 也可以由服务器执行, 本实施例不作限定。

[0078] 本发明实施例的语音识别方法, 可用于识别任意一门语言, 例如汉语、英语、日语、德语等。本发明实施例中主要是以对汉语的语音识别为例进行说明的, 对其他语言的语音识别方法与此类似, 本发明实施例中不再一一举例说明。

[0079] 如图4所示, 基于与上述语音识别方法相同的发明构思, 本发明实施例还提供了一种语音识别装置40, 包括: 确定模块401和识别模块402。

[0080] 确定模块401, 用于若确定上一段语音信号的识别结果为不完整文本, 将上一段语音信号的识别结果确定为历史识别信息。

[0081] 识别模块402, 用于基于历史识别信息, 对当前获取到的语音信号进行语音识别。

[0082] 进一步地, 确定模块401具体用于: 对上一段语音信号的识别结果, 进行断句处理;

若断句处理后的识别结果中包含的标点符号为预设标点符号,确定上一段语音信号的识别结果为不完整文本。

[0083] 进一步地,确定模块401具体用于:对上一段语音信号的识别结果,进行语义解析;根据语义解析结果,确定上一段语音信号的识别结果为不完整文本。

[0084] 进一步地,确定模块401具体用于:对上一段语音信号的识别结果,进行句法分析;若句法分析结果不符合预设句法模板,确定上一段语音信号的识别结果为不完整文本。

[0085] 基于上述任一实施例,识别模块402具体用于:计算当前获取到的语音信号对应的各条假设词序路径的概率得分,假设词序路径是基于历史识别信息对应的历史词序路径得到的;根据概率得分最高的假设词序路径,确定当前获取到的语音信号的识别结果。

[0086] 基于上述任一实施例,识别模块402还用于:从历史识别信息对应的各条假设词序路径中,根据各条假设词序路径的概率得分,选择预设数量的假设词序路径,确定为历史识别信息对应的历史词序路径。

[0087] 进一步地,识别模块402还用于:根据概率得分最高的假设词序路径对应的历史词序路径,更新历史识别信息。

[0088] 本发明实施例提的语音识别装置与上述语音识别方法采用了相同的发明构思,能够取得相同的有益效果,在此不再赘述。

[0089] 基于与上述语音识别方法相同的发明构思,本发明实施例还提供了一种电子设备,该电子设备具体可以为智能音箱、机器人等智能设备内的控制器,也可以为桌面计算机、便携式计算机、智能手机、平板电脑、个人数字助理(Personal Digital Assistant, PDA)、服务器等。如图5所示,该电子设备50可以包括处理器501、存储器502和收发机503。收发机503用于在处理器501的控制下接收和发送数据。

[0090] 存储器502可以包括只读存储器(ROM)和随机存取存储器(RAM),并向处理器提供存储器中存储的程序指令和数据。在本发明实施例中,存储器可以用于存储语音识别方法的程序。

[0091] 处理器501可以是CPU(中央处理器)、ASIC(Application Specific Integrated Circuit,专用集成电路)、FPGA(Field-Programmable Gate Array,现场可编程门阵列)或CPLD(Complex Programmable Logic Device,复杂可编程逻辑器件)处理器通过调用存储器存储的程序指令,按照获得的程序指令实现上述任一实施例中的语音识别方法。

[0092] 本发明实施例提供了一种计算机可读存储介质,用于储存为上述电子设备所用的计算机程序指令,其包含用于执行上述语音识别方法的程序。

[0093] 上述计算机存储介质可以是计算机能够存取的任何可用介质或数据存储设备,包括但不限于磁性存储器(例如软盘、硬盘、磁带、磁光盘(MO)等)、光学存储器(例如CD、DVD、BD、HVD等)、以及半导体存储器(例如ROM、EPROM、EEPROM、非易失性存储器(NAND FLASH)、固态硬盘(SSD))等。

[0094] 以上所述,以上实施例仅用以对本申请的技术方案进行了详细介绍,但以上实施例的说明只是用于帮助理解本发明实施例的方法,不应理解为对本发明实施例的限制。本技术领域的技术人员可轻易想到的变化或替换,都应涵盖在本发明实施例的保护范围之内。

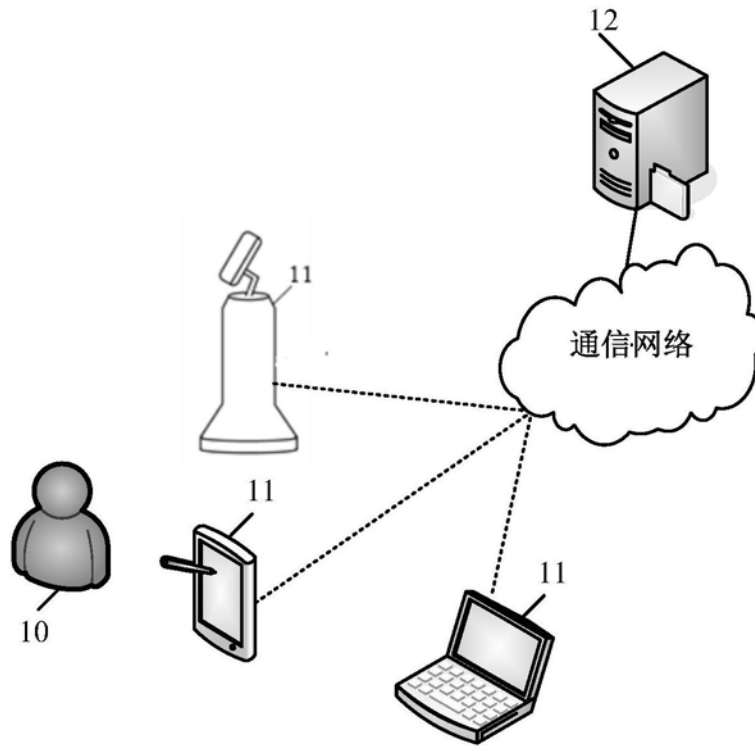


图1

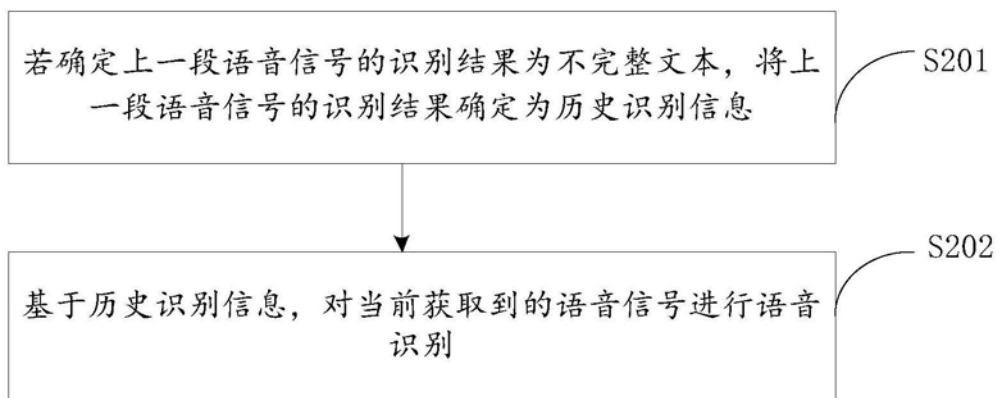


图2

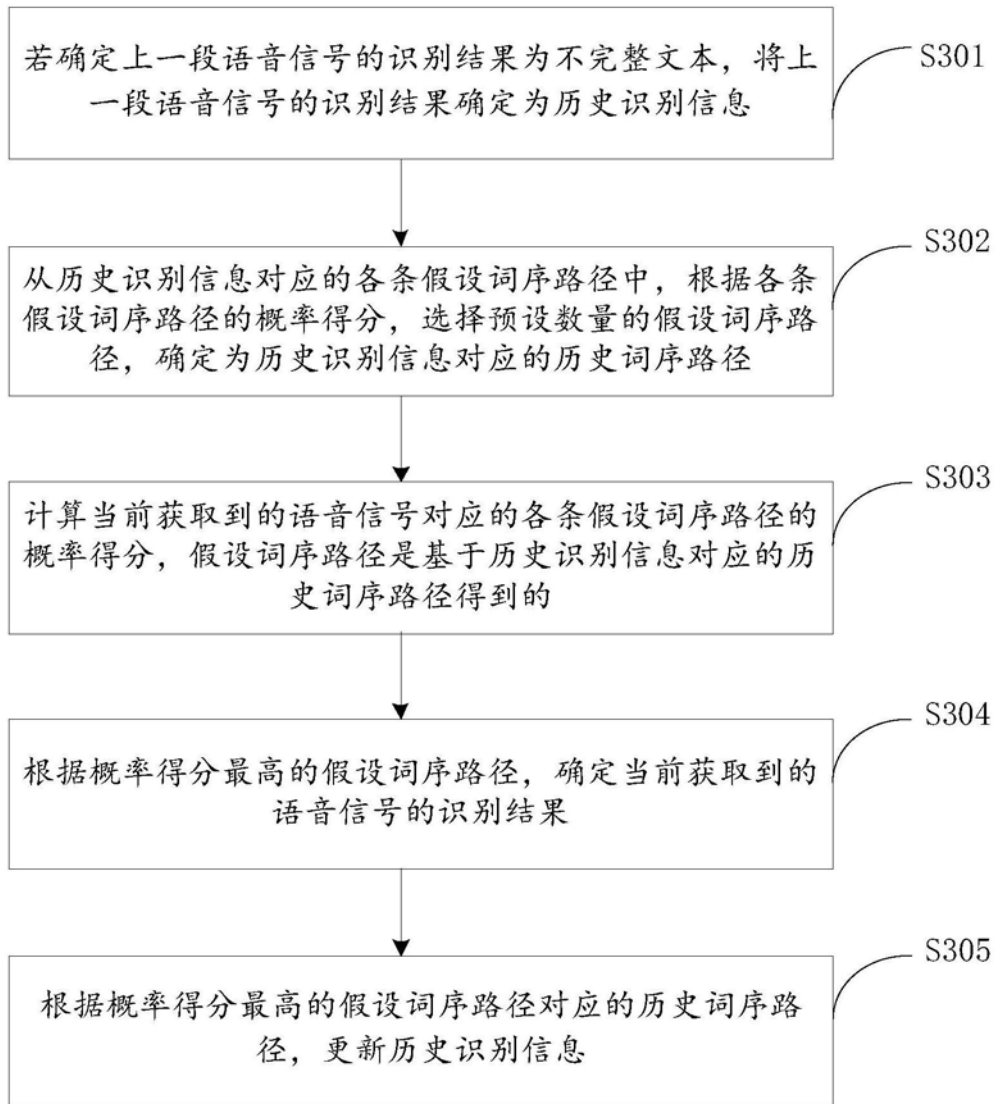


图3

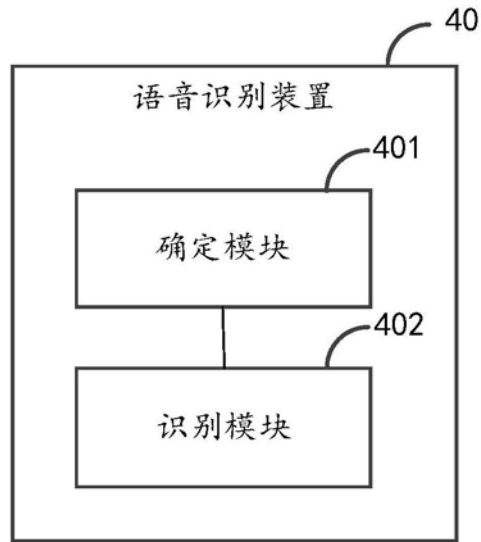


图4

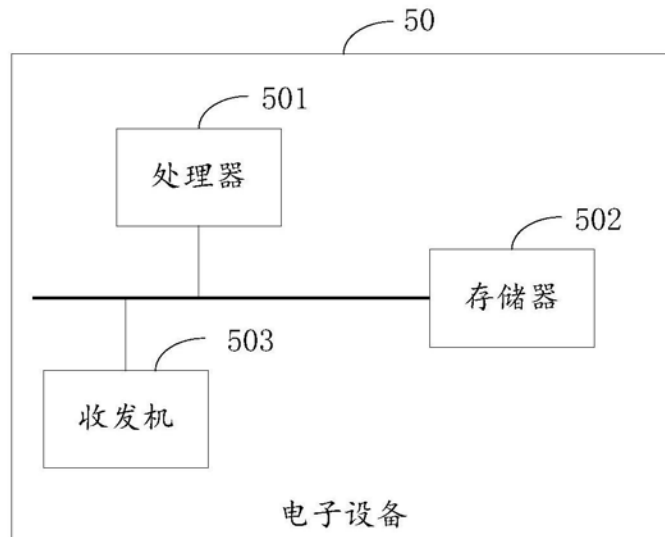


图5