



(12)发明专利

(10)授权公告号 CN 108288468 B

(45)授权公告日 2019.07.19

(21)申请号 201710517737.4

G10L 15/08(2006.01)

(22)申请日 2017.06.29

G10L 15/28(2013.01)

G10L 15/30(2013.01)

(65)同一申请的已公布的文献号

申请公布号 CN 108288468 A

(43)申请公布日 2018.07.17

(73)专利权人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区  
科技中一路腾讯大厦35层

(72)发明人 郑平 饶丰 卢鲤 李涛

(74)专利代理机构 北京三高永信知识产权代理  
有限责任公司 11138

代理人 朱雅男

(56)对比文件

CN 105654946 A,2016.06.08,

CN 106126714 A,2016.11.16,

CN 102395013 A,2012.03.28,

CN 103500579 A,2014.01.08,

US 2016078866 A1,2016.03.17,

CN 106649514 A,2017.05.10,

CN 101345051 A,2009.01.14,

CN 106531160 A,2017.03.22,

审查员 喻新

(51)Int.Cl.

G10L 15/22(2006.01)

G10L 15/10(2006.01)

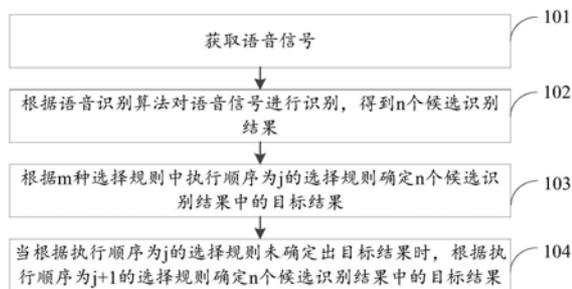
权利要求书3页 说明书14页 附图4页

(54)发明名称

语音识别方法及装置

(57)摘要

本发明公开了一种语音识别方法及装置,属于计算机领域。该方法包括:获取语音信号;根据语音识别算法对所述语音信号进行识别,得到n个候选识别结果;根据m种选择规则中执行顺序为j的选择规则确定所述n个候选识别结果中的目标结果;当根据所述执行顺序为j的选择规则未确定出所述目标结果时,根据执行顺序为j+1的选择规则确定所述n个候选识别结果中的所述目标结果;解决了根据RNN语言模型计算困惑度消耗的时间较长,导致从多个候选识别结果中选择目标结果的实时性较差的问题;提高了从n个候选识别结果中选择目标结果的实时性。



1. 一种语音识别方法,其特征在于,所述方法包括:

获取语音信号;

根据语音识别算法对所述语音信号进行识别,得到n个候选识别结果,所述候选识别结果是指所述语音信号对应的文本信息,所述n为大于1的整数;

根据m种选择规则中执行顺序为j的选择规则确定所述n个候选识别结果中的目标结果,所述目标结果是指所述n个候选识别结果中与所述语音信号匹配度最高的候选识别结果,所述m为大于1的整数,所述j的初始值为1;

当根据所述执行顺序为j的选择规则未确定出所述目标结果时,根据执行顺序为j+1的选择规则确定所述n个候选识别结果中的所述目标结果。

2. 根据权利要求1所述的方法,其特征在于,所述m种选择规则的执行顺序根据各自的算法复杂程度确定,所述执行顺序与所述算法复杂程度呈正相关关系。

3. 根据权利要求1所述的方法,其特征在于,所述m种选择规则包括命令选择规则、功能选择规则和对话选择规则中的至少两种,所述命令选择规则的算法复杂程度低于所述功能选择规则的算法复杂程度,所述功能选择规则的算法复杂程度低于所述对话选择规则的算法复杂程度,

所述命令选择规则用于指示语音识别设备根据命令词库中是否包括与第i个候选识别结果相匹配的命令关键词来检测所述第i个候选识别结果是否为所述目标结果, $1 \leq i \leq n$ ;

所述功能选择规则用于指示所述语音识别设备根据语音词库中是否包括与语音关键词相匹配的词库关键词来检测所述第i个候选识别结果是否为所述目标结果,所述语音关键词是所述第i个候选识别结果中的至少一个关键词;

所述对话选择规则用于指示所述语音识别设备根据训练出的语言模型确定每个候选识别结果与所述语音信号的相似程度来选择所述目标结果。

4. 根据权利要求3所述的方法,其特征在于,根据所述命令选择规则,确定所述n个候选识别结果中的目标结果,包括:

检测所述命令词库的第一对应关系是否包括所述与第i个候选识别结果相匹配的命令关键词, $1 \leq i \leq n$ ;

在所述第一对应关系包括所述与第i个候选识别结果相匹配的命令关键词时,确定所述第i个候选识别结果为所述目标结果;

其中,所述第一对应关系至少包括所述命令关键词。

5. 根据权利要求4所述的方法,其特征在于,所述检测所述命令词库的第一对应关系是否包括所述与第i个候选识别结果相匹配的命令关键词之后,还包括:

在所述第一对应关系不包括与所述n个候选识别结果中的任意一个候选识别结果相匹配的命令关键词时,检测所述命令词库中的第二对应关系是否包括与所述第i个候选识别结果中的任意一个字相匹配的關鍵字;

在所述第二对应关系包括与所述第i个候选识别结果中的字相匹配的關鍵字时,根据所述第二对应关系中所述关键字对应的索引值,在所述第一对应关系中查找所述索引值对应的命令关键词;

确定所述第i个候选识别结果与所述命令关键词之间的编辑距离,所述编辑距离用于指示所述第i个候选识别结果转换为所述命令关键词所需执行的操作次数;

在所述编辑距离小于预设数值时,确定所述第*i*个候选识别结果为所述目标结果;

其中,所述第一对应关系包括所述索引值与所述命令关键词之间的对应关系,所述第二对应关系包括所述索引值与所述关键字之间的对应关系。

6. 根据权利要求3所述的方法,其特征在于,根据所述功能选择规则,选择所述*n*个候选识别结果中的目标结果,包括:

分析第*i*个候选识别结果的功能模板, $1 \leq i \leq n$ ;

检测所述语音词库是否包括与所述第*i*个候选识别结果中的所述语音关键词相匹配的所述词库关键词;

在所述语音词库包括与所述第*i*个候选识别结果中的语音关键词相匹配的所述词库关键词时,将所述第*i*个候选识别结果确定为所述目标结果,所述语音关键词是所述第*i*个候选识别结果中的至少一个关键词;

其中,所述第*i*个候选识别结果包括所述功能模板和所述语音关键词。

7. 根据权利要求3所述的方法,其特征在于,根据所述对话选择规则,选择所述*n*个候选识别结果中的目标结果,包括:

根据所述语言模型计算每个所述候选识别结果的困惑度;

确定所述*n*个候选识别结果中所述困惑度的最小值,将所述最小值对应的所述第*i*个候选识别结果确定为所述目标结果;

其中,所述困惑度用于指示所述候选识别结果与所述语音信号的所述相似程度,且所述困惑度与所述相似程度呈负相关关系;所述语言模型是根据至少一个领域对应的专用语料库生成的*N*-gram语言模型,所述*N*-gram语言模型用于根据当前词的前*N*-1个词的出现概率确定所述当前词的出现概率,所述*N*为正整数。

8. 一种语音识别装置,其特征在于,所述装置包括:

信号获取模块,用于获取语音信号;

语音识别模块,用于根据语音识别算法对所述信号获取模块获取到的所述语音信号进行识别,得到*n*个候选识别结果,所述候选识别结果是指所述语音信号对应的文本信息,所述*n*为大于1的整数;

第一确定模块,用于根据*m*种选择规则中执行顺序为*j*的选择规则确定所述语音识别模块识别出的所述*n*个候选识别结果中的目标结果,所述目标结果是指所述*n*个候选识别结果中与所述语音信号匹配度最高的候选识别结果,所述*m*为大于1的整数,所述*j*的初始值为1;

第二确定模块,用于当所述第一确定模块根据所述执行顺序为*j*的选择规则未确定出所述目标结果时,根据执行顺序为*j*+1的选择规则确定所述*n*个候选识别结果中的所述目标结果。

9. 根据权利要求8所述的装置,其特征在于,所述*m*种选择规则的执行顺序根据各自的算法复杂程度确定,所述执行顺序与所述算法复杂程度呈正相关关系。

10. 根据权利要求8所述的装置,其特征在于,所述*m*种选择规则包括命令选择规则、功能选择规则和对话选择规则中的至少两种,所述命令选择规则的算法复杂程度低于所述功能选择规则的算法复杂程度,所述功能选择规则的算法复杂程度低于所述对话选择规则的算法复杂程度,

所述命令选择规则用于指示语音识别设备根据命令词库中是否包括与第*i*个候选识别

结果相匹配的命令关键词来检测所述第 $i$ 个候选识别结果是否为所述目标结果,  $1 \leq i \leq n$ ;

所述功能选择规则用于指示所述语音识别设备根据语音词库中是否包括与语音关键词相匹配的词库关键词来检测所述第 $i$ 个候选识别结果是否为所述目标结果, 所述语音关键词是所述第 $i$ 个候选识别结果中的至少一个关键词;

所述对话选择规则用于指示所述语音识别设备根据训练出的语言模型确定每个候选识别结果与所述语音信号的相似程度来选择所述目标结果。

11. 一种语音识别设备, 其特征在于, 所述语音识别设备包括处理器和存储器, 所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集, 所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如权利要求1至7任一所述的语音识别方法。

12. 一种计算机可读存储介质, 其特征在于, 所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集, 所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现如权利要求1至7任一所述的语音识别方法。

## 语音识别方法及装置

### 技术领域

[0001] 本发明实施例涉及计算机领域,特别涉及一种语音识别方法及装置。

### 背景技术

[0002] 语音识别技术是指通过语音识别设备将语音信息识别为文本信息的技术,语音识别技术广泛应用于语音拨号、语音导航、智能家居控制、语音搜索、听写数据录入等场景。由于同一发音的语音信号可能对应多组不同字的组合,比如:nihao对应“你好”、“拟好”、“倪浩”这三种组合,因此,语音识别设备根据语音信号可能识别出多个候选识别结果。当语音识别设备识别出多个候选识别结果时,如何选择出与语音信号匹配度最高的候选识别结果显得尤为重要。

[0003] 在一种典型的语音识别方法中,语音识别设备获取到n个候选识别结果后,根据循环神经网络(Recurrent Neural Network,RNN)语言模型,计算每个候选识别结果的困惑度,将困惑度的最小值对应的候选识别结果确定为目标结果。其中,RNN语言模型是根据一个通用的语料库训练得到的,困惑度用于指示候选识别结果与语音信号的相似程度,且困惑度与该相似程度呈负相关关系;目标结果是指n个候选识别结果中与实际接收到的语音信号匹配度最大的候选识别结果,n为大于1的整数。

[0004] 由于根据RNN语言模型计算困惑度消耗的时间较长,因此,从n个候选识别结果中选择目标结果的实时性较差。

### 发明内容

[0005] 为了解决由于语音识别设备根据RNN语言模型计算困惑度消耗的时间较长,导致的从多个候选识别结果中选择目标结果的实时性较差的问题,本发明实施例提供了一种语音识别方法及装置。所述技术方案如下:

[0006] 第一方面,提供了一种语音识别方法,所述方法包括:

[0007] 获取语音信号;

[0008] 根据语音识别算法对所述语音信号进行识别,得到n个候选识别结果,所述候选识别结果是指所述语音信号对应的文本信息,所述n为大于1的整数;

[0009] 根据m种选择规则中执行顺序为j的选择规则确定所述n个候选识别结果中的目标结果,所述目标结果是指所述n个候选识别结果中与所述语音信号匹配度最高的候选识别结果,所述m为大于1的整数,所述j的初始值为1;

[0010] 当根据所述执行顺序为j的选择规则未确定出所述目标结果时,根据执行顺序为j+1的选择规则确定所述n个候选识别结果中的所述目标结果。

[0011] 第二方面,提供了一种候选识别结果选择装置,所述装置包括:

[0012] 信号获取模块,用于获取语音信号;

[0013] 语音识别模块,用于根据语音识别算法对所述信号获取模块获取到的所述语音信号进行识别,得到n个候选识别结果,所述候选识别结果是指所述语音信号对应的文本信

息,所述n为大于1的整数;

[0014] 第一确定模块,用于根据m种选择规则中执行顺序为j的选择规则确定所述语音识别模块识别出的所述n个候选识别结果中的目标结果,所述目标结果是指所述n个候选识别结果中与所述语音信号匹配度最高的候选识别结果,所述m为大于1的整数,所述j的初始值为1;

[0015] 第二确定模块,用于当所述第一确定模块根据所述执行顺序为j的选择规则未确定出所述目标结果时,根据执行顺序为j+1的选择规则确定所述n个候选识别结果中的所述目标结果。

[0016] 第三方面,提供了语音识别设备,其特征在于,所述语音识别设备包括处理器和存储器,所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现第一方面提供的语音识别方法。

[0017] 第四方面,提供了一种计算机可读存储介质,所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现第一方面提供的语音识别方法。

[0018] 本发明实施例提供的技术方案带来的有益效果包括:通过依次执行m选择规则中的至少一种来从语音识别出的n个候选识别结果中选择目标结果,其中,每种选择规则的算法复杂程度均低于根据RNN语言模型计算困惑度的算法复杂程度;解决了根据RNN语言模型计算困惑度消耗的时间较长,导致从多个候选识别结果中选择目标结果的实时性较差的问题;当仅执行了一种选择规则就能够确定出目标结果时,由于该选择规则的算法复杂程度低于根据RNN语言模型计算困惑度的算法复杂程度,因此,提高了从n个候选识别结果中选择目标结果的实时性。

## 附图说明

[0019] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例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。

[0020] 图1A是本发明一个实施例提供的语音识别系统的结构示意图;

[0021] 图1B是本发明一个实施例提供的语音识别方法的流程图;

[0022] 图2是本发明另一个实施例提供的语音识别的流程图;

[0023] 图3是本发明一个实施例提供的第一对应关系示意图和第二对应关系的示意图;

[0024] 图4是本发明另一个实施例提供的语音识别方法的流程图;

[0025] 图5是本发明另一个实施例提供的语音识别的流程图;

[0026] 图6是本发明一个实施例提供的语音识别装置的框图;

[0027] 图7是本发明一个实施例提供的语音识别设备的结构示意图。

## 具体实施方式

[0028] 为使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明实施方

式作进一步地详细描述。

[0029] 首先,对本发明实施例涉及的若干个名词进行介绍。

[0030] 语音识别设备:具有将语音信号识别为文本信息的功能的电子设备。

[0031] 可选地,语音识别设备可以是安装有语音识别引擎的服务器,语音识别设备通过该语音识别引擎将语音信号识别为文本信息。

[0032] 其中,语音识别设备接收到的语音信号可以是该语音识别设备通过音频采集组件采集到的;或者,也可以是语音接收设备通过音频采集组件采集后发送给语音识别设备的。其中,语音接收设备是与语音识别设备相独立的电子设备,比如:语音接收设备可以为手机、平板电脑、智能音箱、智能电视、智能空气净化器、智能空调、电子书阅读器、MP3 (Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)播放器、MP4 (Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机、台式计算机等等。

[0033] 可选地,语音识别设备也可以是手机、平板电脑、智能音箱、智能电视、智能空气净化器、智能空调等,本实施例对此不作限定。

[0034] 可选地,下文中以语音识别设备接收语音接收设备发送的语音信号为例进行说明。

[0035] 候选识别结果:对于某一条语音信号来说,语音识别设备识别出的至少一条文本信息。

[0036] 可选地,当语音识别设备得到的候选识别结果为至少两条时,需要从该至少两条候选识别结果中选择出目标结果。其中,目标结果是指与语音信号的匹配度最高的候选识别结果。

[0037] 请参考图1A,其示出了本发明一个实施例提供的语音识别系统的结构示意图。该系统包括至少一个语音接收设备110和语音识别设备120。

[0038] 语音接收设备110可以是手机、平板电脑、智能音箱、智能电视、智能空气净化器、智能空调、电子书阅读器、MP3播放器、MP4播放器、膝上型便携计算机、台式计算机,本实施例对此不作限定。

[0039] 语音接收设备110中安装有音频采集组件111。音频采集组件111用于采集语音信号。

[0040] 语音接收设备110与语音识别设备120之间通过无线网络或有线网络建立连接。语音接收设备110通过音频采集组件111采集到语音信号后,通过该连接将语音信号发送至语音识别设备120。

[0041] 语音识别设备120用于将语音信号识别为文本信息(候选识别结果)。

[0042] 可选地,语音识别设备120在识别出多个候选识别结果时,还可以从该多个候选识别结果中选择出目标结果。

[0043] 可选地,语音识别设备120选择出目标结果后,将该目标结果反馈给语音接收设备110。

[0044] 可选地,语音识别设备120可以实现为服务器或者服务器集群,本实施例对此不作限定。

[0045] 可选地,在手机、平板电脑、智能音箱、智能电视、智能空气净化器、智能空调、电子

书阅读器、MP3播放器、MP4播放器、膝上型便携计算机等移动终端的物理硬件支持运行复杂算法时,语音识别设备120可以实现为上述移动终端的至少一种,本实施例对此不作限定。

[0046] 可选地,上述的无线网络或有线网络使用标准通信技术和/或协议。网络通常为因特网、但也可以是任何网络,包括但不限于局域网(Local Area Network,LAN)、城域网(Metropolitan Area Network,MAN)、广域网(Wide Area Network,WAN)、移动、有线或者无线网络、专用网络或者虚拟专用网络的任何组合)。在一些实施例中,使用包括超文本标记语言(HyperText Mark-up Language,HTML)、可扩展标记语言(Extensible Markup Language,XML)等的技术和/或格式来代表通过网络交换的数据。此外还可以使用诸如安全套接字层(Secure Socket Layer,SSL)、传输层安全(Trassport Layer Security,TLS)、虚拟专用网络(Virtual Private Network,VPN)、网际协议安全(Internet Protocol Security,IPsec)等常规加密技术来加密所有或者一些链路。在另一些实施例中,还可以使用定制和/或专用数据通信技术取代或者补充上述数据通信技术。

[0047] 可选地,本发明实施例以各个实施例的执行主体为语音识别设备为例进行说明。

[0048] 请参考图1B,其示出了本发明一个实施例提供的语音识别方法的流程图。该方法可以包括以下几个步骤:

[0049] 步骤101,获取语音信号。

[0050] 可选地,语音信号是语音接收设备发送至语音识别设备的;或者,是语音识别设备采集到的;或者,是通过移动存储装置输入到语音识别设备的。

[0051] 步骤102,根据语音识别算法对语音信号进行识别,得到n个候选识别结果。

[0052] 其中,候选识别结果是指语音信号对应的文本信息,n为大于1的整数。

[0053] 语音识别算法用于将语音信号识别为至少一条文本信息。语音识别算法可以是基于对维特比(Viterbi)算法进行改进得到的并行算法;或者,也可以是基于对维特比算法进行改进得到的串行算法;或者,还可以是树-格(Tree-Trellis)算法,本实施例对此不作限定。

[0054] 可选地,语音识别算法具有对n个候选识别结果进行初步排序的功能,此时,语音识别设备获取到的n个候选识别结果具有顺序标识,这样,语音识别设备选择目标结果时,按照顺序标识指示顺序依次检测是否为目标结果。

[0055] 需要补充说明的是,语音识别设备可以仅识别出一个候选识别结果,本实施例对此不作限定。

[0056] 步骤103,根据m种选择规则中执行顺序为j的选择规则确定n个候选识别结果中的目标结果。

[0057] 目标结果是指n个候选识别结果中与语音信号匹配度最高的候选识别结果,m为大于1的整数,j的初始值为1; $1 \leq j \leq m-1$ 。

[0058] 可选地,m种选择规则的执行顺序是根据每种选择规则的算法复杂程度确定的,且算法复杂程度与执行顺序呈正相关关系。即,算法复杂程度越低,执行顺序的序号越小,执行顺序越靠前;算法复杂程度越高,执行顺序的序号越大,执行顺序越靠后。

[0059] 其中,选择规则的算法复杂程度与选择目标结果的速度呈负相关关系。即,算法复杂程度越高,选择目标结果的速度越慢;算法复杂程度越低,选择目标结果的速度越快。

[0060] 可选地,每种选择规则的算法复杂程度通过复杂程度标识表示。示意性地,算法复

杂程度标识为1、2、3,其中,数值越小算法复杂程度越低。

[0061] 可选地,m种选择规则的执行顺序是由开发人员指定的,由于m种选择规则的算法复杂程度均低于根据RNN语言模型计算困惑度的算法复杂程度,因此,无论优先执行哪一种选择规则,语音识别设备选择目标结果的速度均比通过根据RNN语言模型计算困惑度来选择目标结果的速度快。

[0062] 在这种情况下,执行顺序可以通过执行顺序标识表示。示意性地,执行顺序标识为#1、#2、#3,其中,#1表示执行顺序为1,#2表示执行顺序为2,#3表示执行顺序为3。

[0063] 可选地,m种选择规则的执行顺序是随机选取的。

[0064] 步骤104,当根据执行顺序为j的选择规则未确定出目标结果时,根据执行顺序为j+1的选择规则确定n个候选识别结果中的目标结果。

[0065] 语音识别设备根据执行顺序为j的选择规则可能未确定出目标结果,此时,语音识别设备根据执行顺序为j+1的选择规则继续确定目标结果,直至确定出n个候选识别中的目标结果时,流程结束。

[0066] 可选地,语音识别设备对n个候选识别结果重新排序,其中,n个候选识别结果中的目标结果的排列顺序为第一位;除第一位结果之外的剩余n-1个候选识别结果中的目标结果的排列顺序为第二位;除第一位和第二位结果之外的剩余n-2个候选识别结果中的目标结果的排列顺序为第三位,如此循环。

[0067] 综上所述,本发明提供的语音识别方法,通过依次执行m选择规则中的至少一种来从语音识别出的n个候选识别结果中选择目标结果,其中,每种选择规则的算法复杂程度均低于根据RNN语言模型计算困惑度的算法复杂程度;解决了根据RNN语言模型计算困惑度消耗的时间较长,导致从多个候选识别结果中选择目标结果的实时性较差的问题;当仅执行了一种选择规则就能够确定出目标结果时,由于该选择规则的算法复杂程度低于根据RNN语言模型计算困惑度的算法复杂程度,因此,提高了从n个候选识别结果中选择目标结果的实时性。

[0068] 可选地,本实施例中的m种选择规则是根据不同使用场景确定的选择规则。m种选择规则包括命令选择规则、功能选择规则和对话选择规则中的至少两种。在命令场景下(即,语音信号为命令形式的消息),通过m种选择规则中的命令选择规则能够识别出目标结果;在功能场景下(即,语音信号为功能性消息),通过m种选择规则中的功能选择规则能够识别出目标结果;在对话场景下(即,语音信号为对话形式的消息),通过m种选择规则中的对话选择规则能够识别出目标结果。

[0069] 其中,命令形式的消息用于指示语音接收设备执行某个命令,比如:当语音接收设备为智能音箱时,命令形式的消息可以为上一首、下一首、暂停、播放等消息。

[0070] 通常命令形式的消息无规律、且数量有限,比如:对于上一首这个命令形式的消息来说,可以变化为上首、请播放上一首、请播放上首、请切换上首、请切换上一首等,上述各种变化无规律,且变化的种类有限。

[0071] 由于命令形式的消息无规律、且数量有限,因此,本实施例中在语音识别设备中预设有命令词库,该命令词库包括多个命令关键词,命令选择规则用于指示语音识别设备根据命令词库中是否包括与第i个候选识别结果相匹配的命令关键词来检测该第i个候选识别结果是否为目标结果, $1 \leq i \leq n$ 。

[0072] 功能性消息用于指示的语音接收设备根据至少一个语音关键词执行某个命令,比如:功能性消息为“播放周杰伦的歌”。

[0073] 通常功能性消息具有固定形式的功能模板和可变化的语音关键词。比如:在“播放周杰伦的歌”中,功能模板为“播放()的歌”,语音关键词为“周杰伦”。

[0074] 由于通常功能性消息具有固定形式的功能模板和可变化的语音关键词,因此,本实施例在语音识别设备中预设功能模板库和语音词库,功能选择规则用于指示语音识别设备根据语音词库中是否包括与语音关键词相匹配的词库关键词来检测第*i*个候选识别结果是否为目标结果,该语音关键词是第*i*个候选识别结果中的至少一个关键词。

[0075] 对话形式的消息是指无规律、且变化数量未知的消息。比如:对话消息为“你在做什么”、“今天有空吗”、“电影真好看”等。

[0076] 由于对话形式的消息无规律、且变化数量未知,因此,本实施例中,在语音识别中设置有预先训练出语言模型,对话选择规则用于指示语音识别设备根据训练出的语言模型确定每个候选识别结果与语音信号的相似程度来选择目标结果。

[0077] 可选地,在本实施例中,命令选择规则的算法复杂程度低于功能选择规则的算法复杂程度,功能选择规则的算法复杂程度低于对话选择规则的算法复杂程度。相应地,语音识别设备优先执行命令选择规则来选择目标结果;在根据命令选择规则未选择出目标结果时,再执行功能选择规则来选择目标结果;在根据功能选择规则未选择出目标结果时,再执行对话选择规则来选择目标结果。

[0078] 可选地,本实施例中,命令选择规则的算法复杂程度、功能选择规则的算法复杂程度和对话选择规则的算法复杂程度均远远小于根据RNN语言模型选择目标结果的算法复杂程度,因此,若语音识别设备依次执行命令选择规则、功能选择规则和对话选择规则才确定出目标结果,语音识别设备消耗的总时长也小于根据RNN语言模型选择目标结果消耗的总时长。

[0079] 下面对根据命令选择规则来选择目标结果(参见图2所示的实施例)、根据功能选择规则选择目标结果(参见图4所示的实施例)、根据对话选择规则选择目标结果(参见图5所示的实施例)分别作介绍。

[0080] 请参考图2,其示出了本发明另一个实施例提供的语音识别方法的流程图。该方法可以包括以下几个步骤:

[0081] 步骤201,检测命令词库的第一对应关系是否包括与第*i*个候选识别结果相匹配的命令关键词。

[0082] 第一对应关系包括索引值与命令关键词之间的对应关系。

[0083] 可选地,第一对应关系通过正排表实现,该正排表包括至少一个键值对,每个键值对中的键为哈希值(索引值),每个键值对中的值为一个命令关键词。

[0084] 本实施例不对第一对应关系中的键值对的数量作限定,示意性地,第一对应关系中的键值对的数量为1000。

[0085] 语音识别设备检测命令词库的第一对应关系是否包括与第*i*个候选识别结果相匹配的命令关键词,包括:计算第*i*个候选识别结果的哈希值,在第一对应关系中检测是否存在与该哈希值相等的键;如果存在,确定第一对应关系包括与第*i*个候选识别结果相匹配的命令关键词,执行步骤202;如果不包括,则令*i* = *i* + 1,继续执行本步骤。

[0086] 可选地,第一对应关系可以指包括至少一个命令关键词,语音识别设备将第i个候选识别结果与每个命令关键词进行匹配;如果第一对应关系中存在与第i个候选识别结果完全匹配的命令关键词,则执行步骤202,如果不包括,则令 $i=i+1$ ,继续执行本步骤。

[0087] 步骤202,确定第i个候选识别结果为目标结果,流程结束。

[0088] 可选地,当第一对应关系包括至少两个候选识别结果对应的命令关键词时,语音识别设备可以将第一个候选识别结果作为目标结果,或者,语音识别设备执行步骤203,从该至少两个候选识别结果中再次选择目标结果。

[0089] 步骤203,在第一对应关系不包括与n个候选识别结果中的任意一个候选识别结果相匹配的命令关键词时,检测命令词库中的第二对应关系是否包括与第i个候选识别结果中的任意一个字相匹配的关键字。

[0090] 第二对应关系包括索引值与关键字之间的对应关系,命令关键词包括关键字。

[0091] 可选地,第二对应关系通过倒排表实现,该倒排表包括至少一个键值对,每个键值对中的键为关键字的哈希值,每个键值对中的值为该关键字对应在第一对应关系中的至少一个索引值。

[0092] 语音识别设备检测命令词库中的第二对应关系是否包括与第i个候选识别结果中的任意一个字相匹配的关键字,包括:计算第i个候选识别结果中每个字的哈希值;检测第二对应关系中是否包括与任意一个字的哈希值相等的键;如果包括,确定第二对应关系包括与第i个候选识别结果中的字相匹配的关键字,执行步骤204;如果不包括,则令 $i=i+1$ ,继续执行本步骤。

[0093] 可选地,第二对应关系中每个键值对的键也可以为关键字。

[0094] 步骤204,根据第二对应关系中关键字对应的索引值,在第一对应关系中查找索引值对应的命令关键词。

[0095] 由于命令关键词是由关键字组成的,而不同的命令关键词可能包含同一关键字,因此,语音识别设备根据关键字对应的索引值,即,第二对应关系中关键字对应键值对中的值,查找到的命令关键词的数量为至少一个。

[0096] 本实施例中,通过第一对应关系和第二对应关系结合来检测与第i个候选识别结果相匹配的命令关键词,使得语音识别设备无需存储所有的命令关键词的变化形式,只需要存储所有变化形式均包括的关键字,就能够确定出对应的命令关键词,节省了语音识别设备的存储空间。

[0097] 步骤205,确定第i个候选识别结果与命令关键词之间的编辑距离。

[0098] 编辑距离(或称莱文斯坦(Levenshtein)距离)用于指示第i个候选识别结果转换为命令关键词所需执行的操作次数。其中,转换的操作包括但不限于:替换、插入和删除。

[0099] 语音识别设备可能确定出多个命令关键词,此时,确定第i个候选识别结果与每个命令关键词之间的编辑距离。

[0100] 比如:第i个候选识别结果为“在停”,语音识别设备确定出的命令关键词为“暂停”,语音识别设备只需要将“在”替换为“暂”就可以将“在停”转换为“暂停”,则第i个候选识别结果与命令关键词之间的编辑距离为1。

[0101] 步骤206,在编辑距离小于预设数值时,确定第i个候选识别结果为目标结果。

[0102] 在编辑距离小于预设数值时,说明第i个候选识别结果与命令关键词相似程度较

高,此时,将该第*i*个候选识别结果确定为目标结果。

[0103] 预设数值的取值通常较小,且本实施例不对预设数值的取值作限定,示意性地,预设数值为2。

[0104] 参考图3所示的第一对应关系和第二对应关系的示意图,其中,第一对应关系包括3个键值对,每个的键值对由索引值和命令关键词组成,第二对应关系包括3个键值对,每个键值对由哈希值和索引值组成。

[0105] 若语音识别设备识别出4个候选识别结果,该4个候选识别结果分别为:再填、在田、在填和暂停。语音识别设备分别计算这4个候选识别结果的哈希值,其中,再填的哈希值为1、在田的哈希值为2、在填的哈希值为3、暂停的哈希值为4,而第一对应关系中的键包括4,则将暂停确定为目标结果。

[0106] 若语音识别设备识别出4个候选识别结果,该4个候选识别结果分别为:再填、在田、在填和在停。语音识别设备分别计算这4个候选识别结果的哈希值,其中,再填的哈希值为1、在田的哈希值为2、在填的哈希值为3、在停的哈希值为5,此时,第一对应关系中的键不包括1、2、3、5,则语音识别设备计算每个候选识别结果中每个字的哈希值。对于候选识别结果“在停”来说,“在”的哈希值为11,“停”的哈希值为12,第二对应关系中的键包括12,语音识别设备根据第二对应关系中12对应的索引值4,在第一对应关系中查找索引值4对应的命令关键词“暂停”,“在停”与“暂停”之间的编辑距离为1,小于预设数值2,则将“在停”确定为目标结果。

[0107] 可选地,当所有的候选识别结果与命令关键词之间的编辑距离均大于或等于预设数值时,则未根据命令选择规则选择出目标结果,此时,语音识别设备根据其它选择规则继续选择目标结果;或者,将第一条候选识别结果确定为目标结果;或者,不选择目标结果,流程结束。其中,其它选择规则为功能选择规则或对话选择规则。

[0108] 可选地,语音识别设备也可以将编辑距离最小的候选识别结果确定为目标结果。

[0109] 综上所述,本发明提供的语音识别方法,通过命令选择规则来选择*n*个候选识别结果中的目标结果;当仅执行了命令选择规则就能够确定出目标结果时,由于命令选择规则的算法复杂程度低于根据RNN语言模型计算困惑度的算法复杂程度,因此,提高了从*n*个候选识别结果中选择目标结果的实时性。

[0110] 另外,通过第一对应关系和第二对应关系结合来检测与第*i*个候选识别结果相匹配的命令关键词,使得语音识别设备无需存储所有的命令关键词的变化形式,只需要存储所有变化形式均包括的关键字,就能够确定出对应的命令关键词,节省了语音识别设备的存储空间。

[0111] 可选地,语音识别设备将目标结果发送至语音接收设备,该语音接收设备根据目标结果对应的命令执行相应的操作。比如:语音接收设备为智能音箱,且目标结果为暂停,则智能音箱接收到该目标结果后暂停播放当前播放的音频信息。

[0112] 请参考图4,其示出了本发明另一个实施例提供的语音识别方法的流程图。该方法可以包括以下几个步骤:

[0113] 步骤401,分析第*i*个候选识别结果的功能模板, $1 \leq i \leq n$ 。

[0114] 可选地,语音识别设备中预设功能模板库,该功能模板库包括至少一个功能模板。

[0115] 可选地,功能模板通过正则表达式(或称规则表达式)来表示。比如:功能模板为“一首(.+)的歌”。本实施例不对功能模板库中的功能模板的数量作限定,示意性地,功能模板库中的功能模板的数量为540个。

[0116] 其中,正则表达式用于检索和/或替换符合某个功能模板的文本信息。

[0117] 语音识别设备通过将第*i*个候选识别结果与功能模板库中的每个功能模板进行匹配来分析第*i*个候选识别结果中的功能模板。

[0118] 步骤402,检测语音词库是否包括与第*i*个候选识别结果中的语音关键词相匹配的词库关键词。

[0119] 第*i*个候选识别结果包括功能模板和至少一个语音关键词,语音识别设备分析出第*i*个候选识别结果的功能模板后,将第*i*个候选识别结果中剩余的关键词作为语音关键词。

[0120] 语音识别设备中预设有语音词库,语音词库包括至少一个词库关键词。本实施例不对语音词库中词库关键词的数量作限定,示意性地,语音词库中词库关键词的数量为100万。

[0121] 语音识别设备将第*i*个候选识别结果中的语音关键词与语音词库中的至少一个词库关键词逐个进行匹配,在语音词库包括与第*i*个候选识别结果中的语音关键词相匹配的词库关键词时,执行步骤403;在语音词库不包括与第*i*个候选识别结果中的语音关键词相匹配的词库关键词时,令*i* = *i* + 1,继续执行本步骤。

[0122] 步骤403,将第*i*个候选识别结果确定为目标结果,流程结束。

[0123] 可选地,当根据功能选择规则未选择出目标结果时,语音识别设备根据其它选择规则继续选择目标结果;或者,将第一条候选识别结果确定为目标结果;或者,不选择目标结果,流程结束。其中,其它选择规则为命令选择规则或对话选择规则。

[0124] 其中,根据功能选择规则未选择出目标结果包括但不限于如下几种情况:语音识别设备未分析出各个候选识别结果的功能模板;或者,语音识别设备未在语音词库中查找到与各个候选识别结果中的语音关键词相匹配的词库关键词。

[0125] 假设语音识别设备得到3个候选识别结果,分别为:1、我想听图案歌的歌;2、我想听童安格的咯;3、我想听童安格的歌。语音识别设备将上述3个候选识别结果分别与功能模板库中的功能模板相匹配,得到第1个候选识别结果的功能模板为“我想听(.+)的歌”、第2个候选识别结果的功能模板为“我想听(.+)的(.+)”、第3个候选识别结果的功能模板为“我想听(.+)的歌”。

[0126] 对于第1个候选识别结果来说,语音关键词为图案歌;对于第2个候选识别结果来说,语音识别设备将第一个关键词作为语音关键词,即,语音关键词为童安格;对于第3个候选识别结果来说,语音关键词为童安格。

[0127] 语音识别设备依次将候选识别结果中的语音关键词与语音词库中的词库关键词进行匹配,在将第2个候选识别结果中的语音关键词与词库关键词进行匹配时,语音识别设备能够确定出与语音关键词相匹配的词库关键词,则将第2个候选识别结果确定为目标结果。

[0128] 可选地,对于第2个候选识别结果来说,语音识别设备也可以将所有关键词都作为语音关键词,即,语音关键词为童安格和咯,此时,语音词库虽然包括与“童安格”相匹配的

词库关键词,但是不包括与“咯”相匹配的词库关键词,此时,语音识别设备依次将候选识别结果中的语音关键词与语音词库中的词库关键词进行匹配,在将第3个候选识别结果中的语音关键词与词库关键词进行匹配时,语音识别设备能够确定出与语音关键词相匹配的词库关键词,则将第3个候选识别结果确定为目标结果。

[0129] 综上所述,本发明提供的语音识别方法,通过功能选择规则来选择n个候选识别结果中的目标结果;当仅执行了功能选择规则就能够确定出目标结果时,由于功能选择规则的算法复杂程度低于根据RNN语言模型计算困惑度的算法复杂程度,因此,提高了从n个候选识别结果中选择目标结果的实时性。

[0130] 可选地,语音识别设备将目标结果发送至语音接收设备,该语音接收设备根据目标结果中的语音关键词执行相应的操作。比如:语音接收设备为智能音箱,且目标结果为播放周杰伦的歌,则智能音箱接收到该目标结果后,搜索周杰伦的歌,并播放搜索结果对应的音频信息。

[0131] 可选地,语音识别设备根据目标结果中的语音关键词进行搜索,并将搜索结果发送至语音接收设备,该语音接收设备播放该搜索结果对应的音频信息。比如:语音接收设备为智能音箱,且目标结果为播放周杰伦的歌,则语音识别设备根据目标结果中的语音关键字周杰伦,搜索周杰伦的歌,并将搜索结果发送至智能音箱,智能音箱播放搜索结果对应的音频信息。

[0132] 请参考图5,其示出了本发明另一个实施例提供的语音识别方法的流程图。该方法可以包括以下几个步骤:

[0133] 步骤501,根据语言模型计算每个候选识别结果的困惑度。

[0134] 困惑度(preplexity)用于指示候选识别结果与语音信号的相似程度。困惑度与相似程度呈负相关关系。

[0135] 语言模型是用来描述自然语言内在规律的数学模型。

[0136] 可选地,本实施例中,语言模型是根据至少一个领域对应的专用语料库生成的N-gram语言模型,N-gram语言模型用于根据当前词的前N-1个词的出现概率确定当前词的出现概率,N为正整数。本实施例不对N的取值作限定,示意性地,N为3,3-gram语言模型也称为Tri-gram语言模型。示意性地,N为2,2-gram语言模型也称为Bi-gram语言模型。

[0137] N-gram语言模型通过概率和分布函数来描述词、词组及句子等自然语言基本单位的性质和关系,体现了自然语言中存在的基于统计原理的生成和处理规则。

[0138] 本实施例中,以语音识别设备根据3-gram语言模型或2-gram语言模型计算每个候选识别结果的困惑度为例进行说明。

[0139] 可选地,3-gram语言模型通过下述公式表示:

$$[0140] \quad p(S) = p(w_1) p(w_2|w_1) p(w_3|w_1, w_2) \dots p(w_n|w_1, w_2, \dots, w_{n-1})$$

$$[0141] \quad = p(w_1) p(w_2|w_1) p(w_3|w_1, w_2) \dots p(w_n|w_{n-1}, w_{n-2})$$

[0142] 其中, $p(S)$ 表示候选识别结果出现的概率, $p(w_1)$ 表示候选识别结果中第1个词出现的概率、 $p(w_2|w_1)$ 表示候选识别结果中第2个词依赖第1个词的出现而出现的概率、 $p(w_3|w_1, w_2)$ 表示候选识别结果中第3个词依赖第1个词和第2个词的出现而出现的概率、 $p(w_n|w_{n-1}, w_{n-2})$ 表示候选识别结果中第n个词依赖前1个词(第n-1个词)和前2个词(第n-2个词)的出现而出现的概率。

[0143] 可选地,2-gram语言模型通过下述公式表示:

$$[0144] \quad p(S) = p(w_1) p(w_2 | w_1) p(w_3 | w_1, w_2) \dots p(w_n | w_1, w_2, \dots, w_{n-1})$$

$$[0145] \quad = p(w_1) p(w_2 | w_1) p(w_3 | w_2) \dots p(w_n | w_{n-1})$$

[0146] 其中, $p(S)$ 表示候选识别结果出现的概率, $p(w_1)$ 表示候选识别结果中第1个词出现的概率、 $p(w_2 | w_1)$ 表示候选识别结果中第2个词依赖第1个词的出现而出现的概率、 $p(w_3 | w_2)$ 表示候选识别结果中第3个词依赖第2个词的出现而出现的概率、 $p(w_n | w_{n-1})$ 表示候选识别结果中第n个词依赖前1个词(第n-1个词)的出现而出现的概率。

[0147] 其中,至少一个领域包括但不限于以下几种:天气领域、音乐领域、数学领域、体育领域、计算机领域、家居领域、地理领域、自然领域。

[0148] 当然,至少一个领域也可以包括其它领域,本实施例对此不作限定。

[0149] 语音识别设备通过预设公式根据语言模型计算每个候选识别结果的困惑度。

[0150] 困惑度可以视为语言模型预测出的每个词后的候选词的出现概率的几何平均数。通常候选识别结果出现的概率与困惑度呈负相关关系,即,候选识别结果出现的概率越大,则困惑度越低;候选识别结果出现的概率越小,则困惑度越高。

[0151] 可选地,语音识别设备通过预设公式根据语言模型计算每个候选识别结果的困惑度时,先计算每个候选识别结果的交叉熵,根据该交叉熵和预设公式确定该语言识别结果的困惑度。

[0152] 其中,交叉熵用于表示语音模型确定出的模型语言与候选识别结果的差异情况。交叉熵越小,模型语言与候选识别结果的差异越小,候选识别结果与语音信号的匹配度越高;交叉熵越大,模型语言与候选识别结果的差异越大,候选识别结果与语音信号的匹配度越低。

[0153] 可选地,语言模型也可以为其它类型,比如神经网络语言模型,本实施例对此不作限定。

[0154] 步骤502,确定n个候选识别结果中困惑度的最小值,将最小值对应的第i个候选识别结果确定为目标结果。

[0155] 由于困惑度越小,说明候选识别结果与语音信号的相似程度越高,因此,将困惑度的最小值对应的第i个候选识别结果确定为目标结果。

[0156] 综上所述,本发明提供的语音识别方法,通过对话选择规则来选择n个候选识别结果中的目标结果;当仅执行了对话选择规则就能够确定出目标结果时,由于对话选择规则的算法复杂程度低于根据RNN语言模型计算困惑度的算法复杂程度,因此,提高了从n个候选识别结果中选择目标结果的实时性。

[0157] 可选地,语音识别设备将目标结果发送至语音接收设备,该语音接收设备根据目标结果获取对话信息。比如:语音接收设备为智能音箱,且目标结果为你在做什么,则智能音箱接收到该目标结果后,根据对话模型生成对话信息。

[0158] 可选地,语音识别设备根据目标结果生成对话信息,并将对话信息发送至语音接收设备,该语音接收设备播放该对话信息对应的音频信息。比如:语音接收设备为智能音箱,且目标结果为你在做什么,则语音识别设备根据目标结果生成对话信息,并将对话信息发送至智能音箱,智能音箱播放对话信息对应的音频信息。

[0159] 下述为本发明装置实施例,可以用于执行本发明方法实施例。对于本发明装置实

施例中未披露的细节,请参照本发明方法实施例。

[0160] 请参考图6,其示出了本发明一个实施例提供的语音识别装置的框图。该装置具有执行上述方法示例的功能,功能可以由硬件实现,也可以由硬件执行相应的软件实现。该装置可以包括:信号获取模块610、语音识别模块620、第一确定模块630和第二确定模块640。

[0161] 信号获取模块610,用于获取语音信号;

[0162] 语音识别模块620,用于根据语音识别算法对所述信号获取模块610获取到的所述语音信号进行识别,得到n个候选识别结果,所述候选识别结果是指所述语音信号对应的文本信息,所述n为大于1的整数;

[0163] 第一确定模块630,用于根据m种选择规则中执行顺序为j的选择规则确定所述语音识别模块620识别出的所述n个候选识别结果中的目标结果,所述目标结果是指所述n个候选识别结果中与所述语音信号匹配度最高的候选识别结果,所述m为大于1的整数,所述j的初始值为1;

[0164] 第二确定模块640,用于当所述第一确定模块630根据所述执行顺序为j的选择规则未确定出所述目标结果时,根据执行顺序为j+1的选择规则确定所述n个候选识别结果中的所述目标结果。

[0165] 可选地,m种选择规则的执行顺序根据各自的算法复杂程度确定,执行顺序与算法复杂程度呈正相关关系。

[0166] 可选地,m种选择规则包括命令选择规则、功能选择规则和对话选择规则中的至少两种,命令选择规则的算法复杂程度低于功能选择规则的算法复杂程度,功能选择规则的算法复杂程度低于对话选择规则的算法复杂程度,

[0167] 命令选择规则用于指示语音识别设备根据命令词库中是否包括与第i个候选识别结果相匹配的命令关键词来检测第i个候选识别结果是否为目标结果, $1 \leq i \leq n$ ;

[0168] 功能选择规则用于指示语音识别设备根据语音词库中是否包括与语音关键词相匹配的词库关键词来检测第i个候选识别结果是否为目标结果,语音关键词是第i个候选识别结果中的至少一个关键词;

[0169] 对话选择规则用于指示语音识别设备根据训练出的语言模型确定每个候选识别结果与语音信号的相似程度来选择目标结果。

[0170] 可选地,第一确定模块630,包括:第一检测单元和第一确定单元。

[0171] 第一检测单元,用于检测命令词库的第一对应关系是否包括与第i个候选识别结果相匹配的命令关键词, $1 \leq i \leq n$ ;

[0172] 第一确定单元,用于在第一对应关系包括与第i个候选识别结果相匹配的命令关键词时,确定第i个候选识别结果为目标结果;

[0173] 其中,第一对应关系至少包括命令关键词。

[0174] 可选地,该装置还包括:第二检测模块、关键字查找模块、第二确定模块和第三确定模块。

[0175] 第二检测模块,用于在第一对应关系不包括与n个候选识别结果中的任意一个候选识别结果相匹配的命令关键词时,检测命令词库中的第二对应关系是否包括与第i个候选识别结果中的任意一个字相匹配的关键字;

[0176] 关键字查找模块,用于在第二对应关系包括与第i个候选识别结果中的字相匹配

的关键字时,根据第二对应关系中关键字对应的索引值,在第一对应关系中查找索引值对应的命令关键词;

[0177] 第二确定模块,用于确定第*i*个候选识别结果与命令关键词之间的编辑距离,编辑距离用于指示第*i*个候选识别结果转换为命令关键词所需执行的操作次数;

[0178] 第三确定模块,用于在编辑距离小于预设数值时,确定第*i*个候选识别结果为目标结果;

[0179] 其中,第一对应关系包括索引值与命令关键词之间的对应关系,第二对应关系包括索引值与关键字之间的对应关系。

[0180] 可选地,第一确定模块630,包括:模板分析单元、第三检测单元和第四确定单元。

[0181] 模板分析单元,用于分析第*i*个候选识别结果的功能模板, $1 \leq i \leq n$ ;

[0182] 第三检测单元,用于检测语音词库是否包括与第*i*个候选识别结果中的语音关键词相匹配的词库关键词;

[0183] 第四确定单元,用于在语音词库包括与第*i*个候选识别结果中的语音关键词相匹配的词库关键词时,将第*i*个候选识别结果确定为目标结果,语音关键词是第*i*个候选识别结果中的至少一个关键词;

[0184] 其中,第*i*个候选识别结果包括功能模板和语音关键词。

[0185] 可选地,第一确定模块630,包括:困惑度计算单元和第五确定单元。

[0186] 困惑度计算单元,用于根据语言模型计算每个候选识别结果的困惑度;

[0187] 第五确定单元,用于确定*n*个候选识别结果中困惑度的最小值,将最小值对应的第*i*个候选识别结果确定为目标结果;

[0188] 其中,困惑度用于指示候选识别结果与语音信号的相似程度,且困惑度与相似程度呈负相关关系;语言模型是根据至少一个领域对应的专用语料库生成的*N*-gram语言模型,*N*-gram语言模型用于根据当前词的前*N*-1个词的出现概率确定当前词的出现概率,*N*为正整数。

[0189] 本发明实施例还提供了一种计算机可读存储介质,该计算机可读存储介质可以是存储器中所包含的计算机可读存储介质;也可以是单独存在,未装配入语音识别设备中的计算机可读存储介质。计算机可读存储介质存储有至少一条指令、至少一段程序、代码集或指令集,该至少一条指令、至少一段程序、代码集或指令集由处理器加载并执行以实现上述各个方法实施例提供的语音识别方法。

[0190] 图7是本发明一个实施例提供的语音识别设备的结构示意图。语音识别设备700包括中央处理单元(英文:Central Processing Unit,简称:CPU)701、包括随机存取存储器(英文:random access memory,简称:RAM)702和只读存储器(英文:read-only memory,简称:ROM)703的系统存储器704,以及连接系统存储器704和中央处理单元701的系统总线705。所述语音识别设备700还包括帮助计算机内的各个器件之间传输信息的基本输入/输出系统(I/O系统)706,和用于存储操作系统713、应用程序714和其他程序模块715的大容量存储设备707。

[0191] 所述基本输入/输出系统706包括有用于显示信息的显示器708和用于用户输入信息的诸如鼠标、键盘之类的输入设备709。其中所述显示器708和输入设备709都通过连接到系统总线705的输入/输出控制器710连接到中央处理单元701。所述基本输入/输出系统706

还可以包括输入输出控制器710以用于接收和处理来自键盘、鼠标、或电子触控笔等多个其他设备的输入。类似地，输入/输出控制器710还提供输出到显示屏、打印机或其他类型的输出设备。

[0192] 所述大容量存储设备707通过连接到系统总线705的大容量存储控制器(未示出)连接到中央处理单元701。所述大容量存储设备707及其相关联的计算机可读介质为语音识别设备700提供非易失性存储。也就是说，所述大容量存储设备707可以包括诸如硬盘或者只读光盘(英文:Compact Disc Read-Only Memory,简称:CD-ROM)驱动器之类的计算机可读介质(未示出)。

[0193] 不失一般性，所述计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其他数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括RAM、ROM、可擦除可编程只读存储器(英文:erasable programmable read-only memory,简称:EPR0M)、电可擦除可编程只读存储器(英文:electrically erasable programmable read-only memory,简称:EEPR0M)、闪存或其他固态存储其技术,CD-ROM、数字通用光盘(英文:Digital Versatile Disc,简称:DVD)或其他光学存储、磁带盒、磁带、磁盘存储或其他磁性存储设备。当然，本领域技术人员可知所述计算机存储介质不局限于上述几种。上述的系统存储器704和大容量存储设备707可以统称为存储器。

[0194] 根据本发明的各种实施例，所述语音识别设备700还可以通过诸如因特网等网络连接网络上的远程计算机运行。也即语音识别设备700可以通过连接在所述系统总线705上的网络接口单元711连接到网络712，或者说，也可以使用网络接口单元711来连接到其他类型的网络或远程计算机系统(未示出)。

[0195] 具体在本发明实施例中，语音识别设备700还包括有存储器，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。上述一个或者一个以上程序包含用于执行上述语音识别方法的指令。

[0196] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0197] 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。

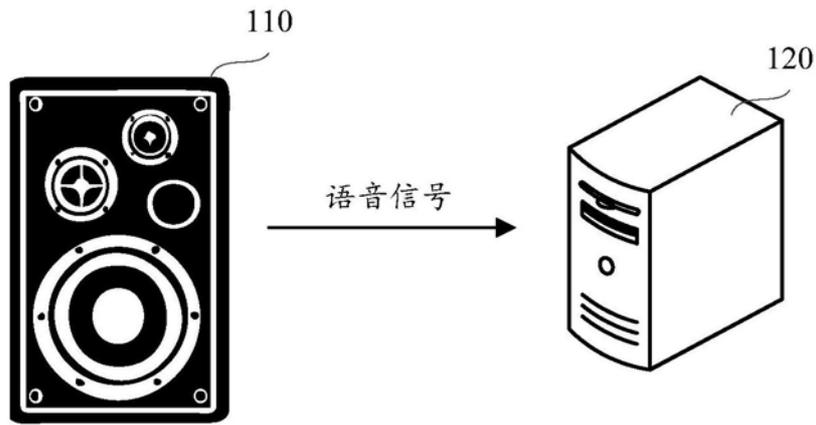


图1A

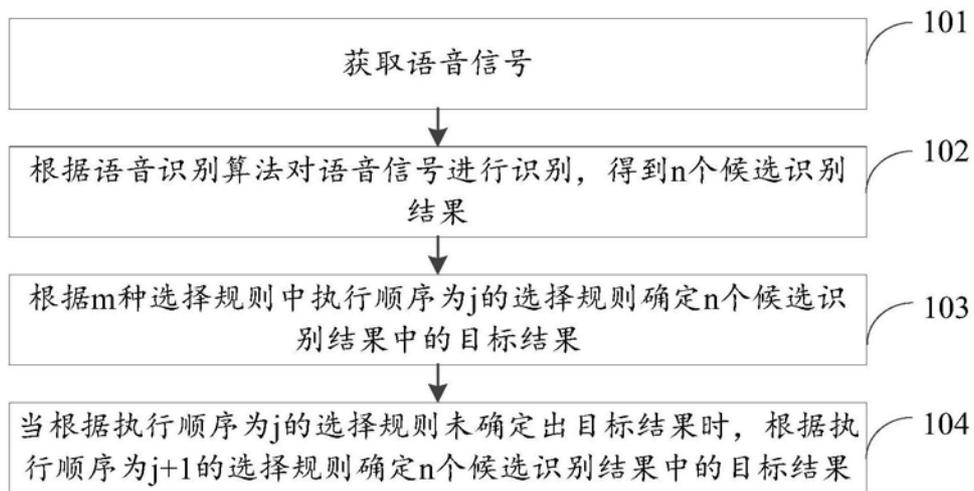


图1B

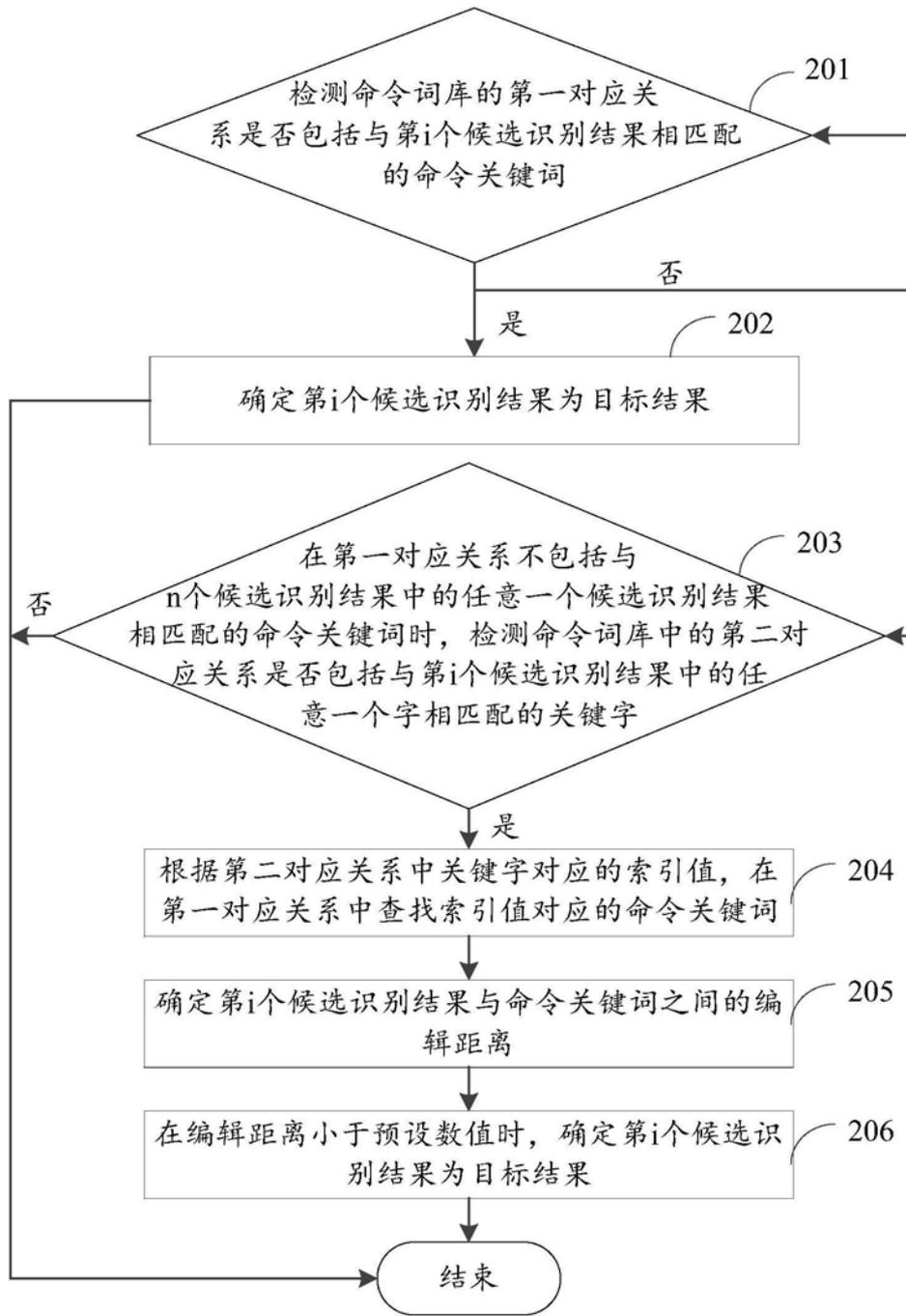


图2

第一对应关系		第二对应关系	
键	值	键	值
4	暂停	12	4
6	上首	13	6 7
7	上一首	14	6 7

图3

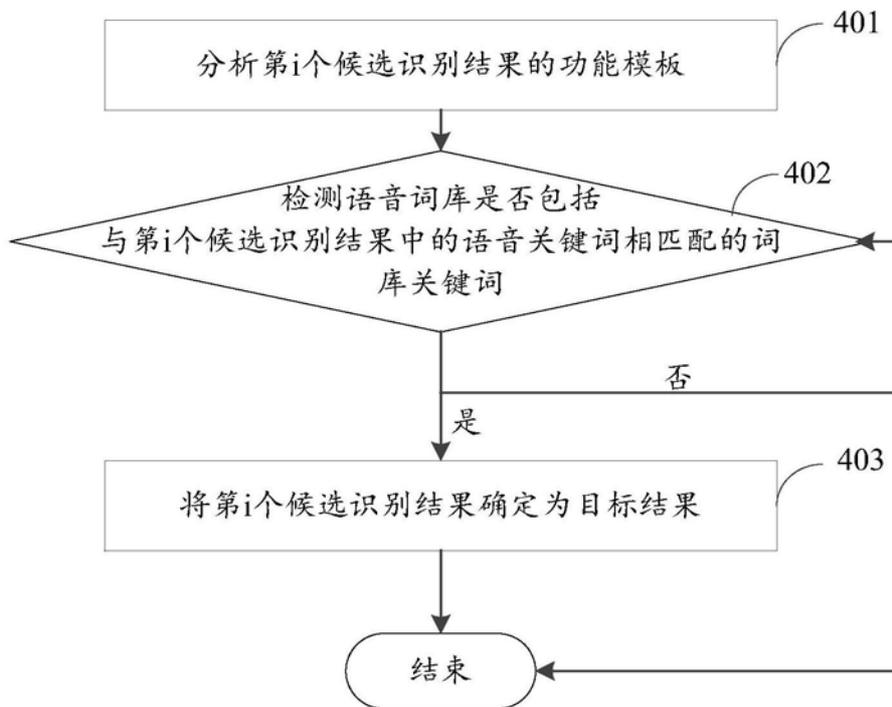


图4

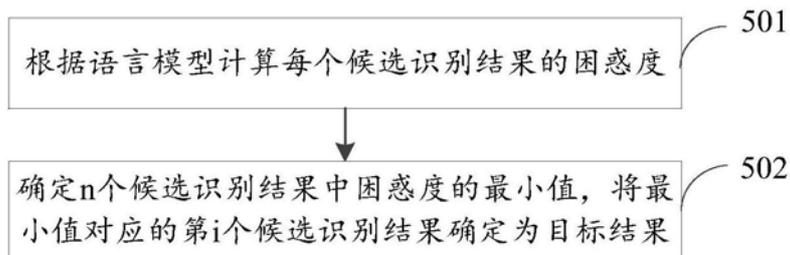


图5

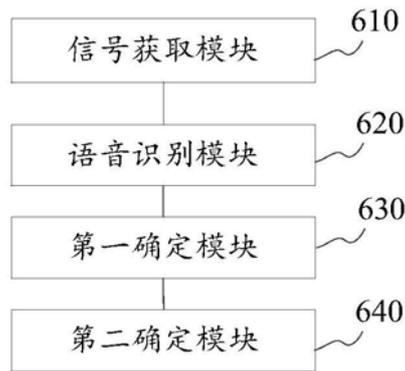


图6

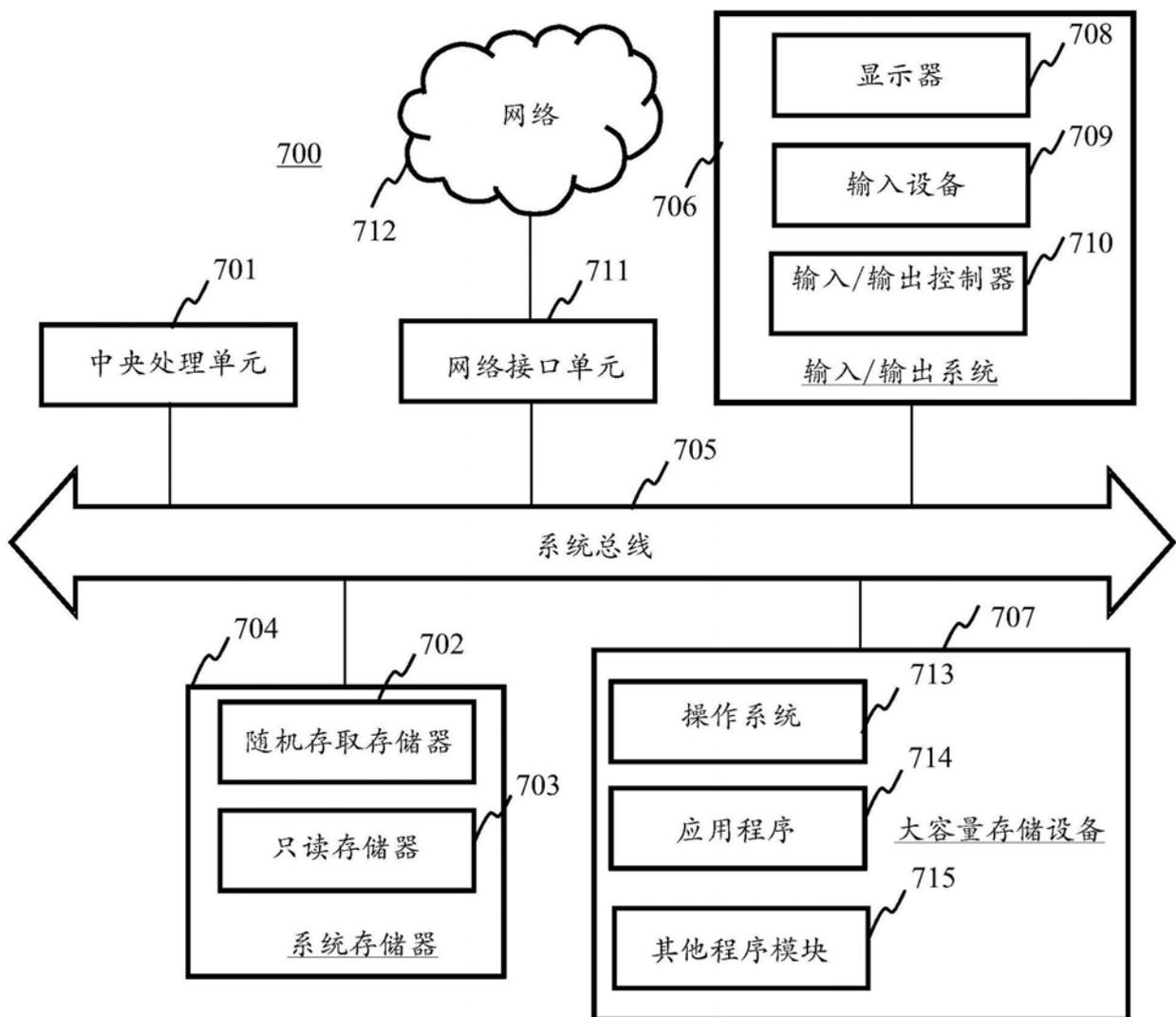


图7