(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112802476 A
(43)申请公布日 2021.05.14

(21)申请号 202011607654.2

(22)申请日 2020.12.30

(71)申请人 深圳追一科技有限公司
地址 518051 广东省深圳市南山区粤海街道科技园社区科苑路8号讯美科技广场3号楼23A、23B

(72)发明人 周维聪 袁丁 赵金昊 吴悦

(74)专利代理机构 广州华进联合专利商标代理有限公司 44224
代理人 方高明

(51)Int.Cl.
G10L 15/26(2006.01)
G10L 15/02(2006.01)
G10L 25/24(2013.01)
G10L 15/14(2006.01)
G10L 15/16(2006.01)

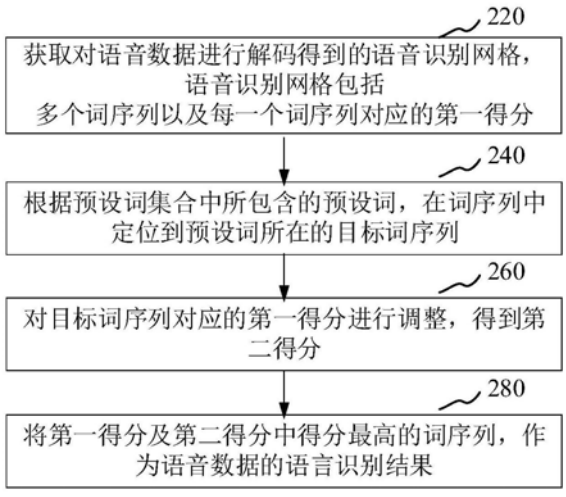
权利要求书2页 说明书12页 附图6页

(54)发明名称

语音识别方法和装置、服务器、计算机可读存储介质

(57)摘要

本申请涉及一种语音识别方法和装置、服务器、计算机可读存储介质，包括：获取对语音数据进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分。根据预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列。对目标词序列对应的第一得分进行调整得到第二得分，将第一得分及第二得分中得分最高的词序列，作为语音数据的语言识别结果。可以基于预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列，并采用对目标词序列的得分进行调整的方式，实现了对解码得到语音识别结果的过程的干预，进而提高所得到的语音识别结果的准确性。



CN 112802476 A

1.一种语音识别方法,其特征在于,所述方法包括:

获取对语音数据进行解码得到语音识别网格lattice,所述语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分;

根据预设词集合中所包含的预设词,在所述词序列中定位到所述预设词所在的目标词序列;

对所述目标词序列对应的第一得分进行调整,得到第二得分;

将所述第一得分及所述第二得分中得分最高的词序列,作为所述语音数据的语言识别结果。

2.根据权利要求1所述的方法,其特征在于,所述根据预设词集合中所包含的预设词,在所述词序列中定位到所述预设词所在的目标词序列,包括:

获取所述预设词集合中所包含的预设词的音素信息;

将所述预设词的音素信息与所述词序列中的音素信息进行匹配;

若匹配成功,则在所述词序列中定位到匹配成功的音素信息所在的目标跳转边。

3.根据权利要求2所述的方法,其特征在于,所述对所述词序列对应的第一得分进行调整,得到第二得分,包括:

对所述目标跳转边上的模型得分进行调整得到新的模型得分;

若所述目标跳转边上的词信息与所述预设词相同,则将所述目标跳转边上的模型得分更新为所述新的模型得分;

基于所述目标跳转边上所述新的模型得分,计算得到所述词序列的第二得分。

4.根据权利要求3所述的方法,其特征在于,在所述对所述目标跳转边上的模型得分进行调整得到新的模型得分之后,包括:

若所述目标跳转边上的词信息与所述预设词不同,则在所述目标跳转边的起始节点与终止节点之间增加新的跳转边;

配置所述新的跳转边上的词信息为所述预设词,并配置所述新的跳转边上的模型得分为所述新的模型得分;

基于所述新的跳转边上所述新的模型得分,计算得到包含所述新的跳转边的词序列的第二得分。

5.根据权利要求3所述的方法,其特征在于,所述对所述目标跳转边上的模型得分进行调整得到新的模型得分,包括:

对所述目标跳转边上的模型得分进行增大预设比例得到新的模型得分。

6.根据权利要求1所述的方法,其特征在于,所述方法还包括:

从预设训练语料中获取识别错误率高于预设错误率阈值的预设词;

基于所述预设词得到所述预设词集合。

7.根据权利要求6所述的方法,其特征在于,所述预设词包括基础词及所述基础词的相似词,所述基础词的相似词为与所述基础词的音素的相似度高于预设相似度阈值的词。

8.根据权利要求1所述的方法,其特征在于,所述方法还包括:

对待处理的语音数据进行声学特征提取;

将所提取的声学特征输入声学模型,计算所述声学特征的声学模型得分;

采用解码算法调用主解码网络及子解码网络,对所述声学特征及所述声学特征的声学

模型得分进行解码得到语音识别网格lattice,所述语音识别网格lattice中包括多个词序列以及每一个所述词序列对应的第一得分;所述主解码网络为对原始文本训练语料进行训练所得的解码图,所述子解码图为对待识别场景中的命名实体进行训练所得的解码图。

9.一种语音识别装置,其特征在于,所述装置包括:

语音识别网格获取模块,用于获取对语音数据进行解码得到语音识别网格lattice,所述语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分;

目标词序列定位模块,用于根据预设词集合中所包含的预设词,在所述词序列中定位到所述预设词所在的目标词序列;

得分调整模块,用于对所述目标词序列对应的第一得分进行调整,得到第二得分;

语言识别结果生成模块,用于将所述第一得分及所述第二得分中得分最高的词序列,作为所述语音数据的语言识别结果。

10.根据权利要求9所述的装置,其特征在于,所述装置还包括:

声学特征提取模块,用于对待处理的语音数据进行声学特征提取;

声学模型得分计算模块,用于将所提取的声学特征输入声学模型,计算所述声学特征的声学模型得分;

解码模块,用于采用解码算法调用主解码网络及子解码网络,对所述声学特征及所述声学特征的声学模型得分进行解码得到语音识别网格lattice,所述语音识别网格lattice中包括多个词序列以及每一个所述词序列对应的第一得分;所述主解码网络为对原始文本训练语料进行训练所得的解码图,所述子解码图为对待识别场景中的命名实体进行训练所得的解码图。

11.一种服务器,包括存储器及处理器,所述存储器中储存有计算机程序,其特征在于,所述计算机程序被所述处理器执行时,使得所述处理器执行如权利要求1至8中任一项所述的语音识别方法的步骤。

12.一种计算机可读存储介质,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至8中任一项所述的语音识别方法的步骤。

语音识别方法和装置、服务器、计算机可读存储介质

技术领域

[0001] 本申请涉及自然语言处理技术领域，特别是涉及一种语音识别方法和装置、服务器、计算机可读存储介质。

背景技术

[0002] 随着人工智能和自然语言处理技术的不断发展，语音识别技术也得到了快速地发展。语音识别技术能够将语音转换成为对应的字符或编码，广泛应用于智能家居、实时语音转写、机器同传等领域。

[0003] 但是，传统的语音识别技术在语音识别过程中也会或多或少地出现一些错误，而这些错误往往大大降低了语音识别的准确率。因此，在人们对语音识别效果的要求不断提高的情况下，就亟需提高语音识别的准确率。

发明内容

[0004] 本申请实施例提供一种语音识别方法、装置、服务器、计算机可读存储介质，可以提高所得到的语音识别结果的准确性。

[0005] 一种语音识别方法，所述方法包括：

[0006] 获取对语音数据进行解码得到语音识别网格lattice，所述语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分；

[0007] 根据预设词集合中所包含的预设词，在所述词序列中定位到所述预设词所在的目标词序列；

[0008] 对所述目标词序列对应的第一得分进行调整，得到第二得分；

[0009] 将所述第一得分及所述第二得分中得分最高的词序列，作为所述语音数据的语言识别结果。

[0010] 一种语音识别装置，所述装置包括：

[0011] 语音识别网格获取模块，用于获取对语音数据进行解码得到语音识别网格lattice，所述语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分；

[0012] 目标词序列定位模块，用于根据预设词集合中所包含的预设词，在所述词序列中定位到所述预设词所在的目标词序列；

[0013] 得分调整模块，用于对所述目标词序列对应的第一得分进行调整，得到第二得分；

[0014] 语言识别结果生成模块，用于将所述第一得分及所述第二得分中得分最高的词序列，作为所述语音数据的语言识别结果。

[0015] 一种服务器，包括存储器及处理器，所述存储器中储存有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如上方法的步骤。

[0016] 一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现如上方法的步骤。

[0017] 上述语音识别方法、装置、服务器、计算机可读存储介质，获取对语音数据进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分。根据预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列。对目标词序列对应的第一得分进行调整，得到第二得分，将第一得分及第二得分中得分最高的词序列，作为语音数据的语言识别结果。可以基于预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列，并对目标词序列对应的第一得分进行调整，进而，从得分调整之后的词序列中筛选出得分最高的词序列作为语音识别结果。最终，采用对目标词序列的得分进行调整的方式，实现了对解码得到语音识别结果的过程的干预，进而提高所得到的语音识别结果的准确性。

附图说明

[0018] 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，下面描述中的附图仅仅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。

[0019] 图1为一个实施例中语音识别方法的应用场景图；

[0020] 图2为一个实施例中语音识别方法的流程图；

[0021] 图3为图2中根据预设词集合中所包含的预设词，在所述词序列中定位到所述预设词所在的目标词序列方法的流程图；

[0022] 图4为一个实施例中语音识别网格lattice的结构示意图；

[0023] 图5为图2中对词序列对应的第一得分进行调整得到第二得分方法的流程图；

[0024] 图6为另一个实施例中语音识别网格lattice的结构示意图；

[0025] 图7为图2中获取对语音数据进行解码得到语音识别网格lattice方法的流程图；

[0026] 图8为一个实施例中语音识别装置的结构框图；

[0027] 图9为图8中得分调整模块的结构框图；

[0028] 图10为另一个实施例中语音识别装置的结构框图；

[0029] 图11为一个实施例中服务器的内部结构示意图。

具体实施方式

[0030] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0031] 可以理解，本申请所使用的术语“第一”、“第二”等可在本文中用于描述各种元件，但这些元件不受这些术语限制。这些术语仅用于将第一个元件与另一个元件区分。

[0032] 图1为一个实施例中语音识别方法的应用场景图。如图1所示，该应用环境包括终端120和服务器140，该终端120与服务器140之间通过网络连接。服务器140通过本申请中的语音识别方法，获取对语音数据进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分；根据预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列；对目标词序列对应的第一得分

进行调整,得到第二得分;将第一得分及第二得分中得分最高的词序列,作为语音数据的语言识别结果。这里,终端120可以是手机、平板电脑、PDA(Personal Digital Assistant,个人数字助理)、车载电脑、穿戴式设备、智能家居等任意终端设备。

[0033] 图2为一个实施例中语音识别方法的流程图,如图2所示,提供了一种语音识别方法,应用于服务器,包括步骤220至步骤280。

[0034] 步骤220,获取对语音数据进行解码得到语音识别网格lattice,语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分。

[0035] 其中,语音数据可以是指所获取到的音频信号。具体可以是在语音输入场景、智能聊天场景、语音翻译场景中所获取到的音频信号。对待处理的语音数据进行声学特征提取。声学特征提取的具体过程可以是:将获取到的一维音频信号,通过特征提取算法转化为一组高维向量。所得到的高维向量即为声学特征,常见的声学特征有MFCC,Fbank,ivector等,本申请对此不做限定。Fbank(FilterBank)就是一种前端处理算法,以类似于人耳的方式对音频进行处理,可以提高语音识别的性能。其中,获得语音信号的Fbank特征的一般步骤是:预加重、分帧、加窗、短时傅里叶变换(STFT)、mel滤波、去均值等。且通过对Fbank做离散余弦变换(DCT)即可获得MFCC特征。

[0036] 其中,MFCC(梅尔频率倒谱系数,Mel-frequency cepstral coefficients),梅尔频率是基于人耳听觉特性提出来的,因此,梅尔频率与Hz频率成非线性对应关系。梅尔频率倒谱系数(MFCC)则是利用梅尔频率与Hz频率之间的这种非线性对应关系,计算得到的Hz频谱特征。MFCC主要用于语音数据特征提取和降低运算维度。例如:对于一帧有512维(采样点)数据,经过MFCC后可以提取出最重要的40维数据,从而达到了降维的目的。其中,ivector为描述每个说话人的特征向量。

[0037] 将所提取的声学特征输入声学模型,计算声学特征的声学模型得分。其中,声学模型可以包括神经网络模型和隐马尔可夫模型。采用解码网络对声学特征及声学特征的声学模型得分进行解码得到语音识别网格lattice,语音识别网格lattice中包括多个词序列以及每一个词序列对应的第一得分。这里的第一得分包括声学模型得分及语言模型得分。

[0038] 其中,语音识别网格lattice包括多条备选词序列。其中,每条备选词序列包括多个词语以及多个路径,lattice本质上是一个有向无环(directed acyclic graph)图,图上的每个节点代表一个词的结束时间点,每条跳转边代表一个可能的词,以及该词发生的声学模型得分和语言模型得分。对语音识别的结果进行表示时,每个节点存放当前位置上语音识别的结果,包括声学概率、语言概率等信息。

[0039] 步骤240,根据预设词集合中所包含的预设词,在词序列中定位到预设词所在的目标词序列。

[0040] 具体的,在一种情况下,预设词可以是人工在待识别场景下所指定的一些语音识别错误率超过预设错误率阈值的命名实体,本申请对此不做限定。例如,对于“北京”这个词语,被识别错误的概率超过了预设错误阈值(例如预设错误阈值为80%),则将“北京”作为第一预设词。由这些第一预设词就得到了第一预设词集合,其中,可以针对各个不同待识别场景,可以基于各待识别场景中的第一预设词分别形成对应的第一预设词集合。也可以不区分待识别场景,基于很多个待识别场景中的第一预设词形成一个统一的第一预设词集合。若在某个特定识别场景下进行语音识别,则根据该特定识别场景对应的第一预设词集

合中的第一预设词，在语音数据进行解码得到的多个词序列中定位到第一预设词所在的目标词序列。也可以根据该统一的第一预设词集合中的第一预设词，在语音数据进行解码得到的多个词序列中定位到第一预设词所在的目标词序列。

[0041] 在另一种情况下，预设词可以是被错误识别出的词语，且被错误识别出的概率超过了预设错误识别阈值(例如预设错误是被阈值为50％)的词语。例如，对于“北京”这个词语，被错误识别为音素不相同的“北极”的概率为60％，超过了预设错误识别阈值。则将“北极”作为第二预设词。可以理解为，第二预设词为第一预设词被错误识别到的词语，且第二预设词与第一预设词的音素信息不同或音素信息的相似度低于预设相似度阈值。由这些第二预设词就得到了第二预设词集合，其中，可以针对各个不同待识别场景，可以基于各待识别场景中的第二预设词分别形成对应的第二预设词集合。也可以不区分待识别场景，基于很多个待识别场景中的第二预设词形成一个统一的第二预设词集合。

[0042] 若在某个特定识别场景下进行语音识别，则根据该特定识别场景对应的预设词集合中的第二预设词，在语音数据进行解码得到的多个词序列中定位到第二预设词所在的目标词序列。也可以根据该统一的第二预设词集合中的第二预设词，在语音数据进行解码得到的多个词序列中定位到第二预设词所在的目标词序列。

[0043] 步骤260，对目标词序列对应的第一得分进行调整，得到第二得分。

[0044] 因为，这里的第一得分包括声学模型得分及语言模型得分，所以在对目标词序列对应的第一得分进行调整时，可以对目标词序列对应的声学模型得分和/或语言模型得分进行调整，本申请对此不做限定。且因为目标词序列的第一得分表示了该目标词序列出现的概率，且目标词序列中包含了预设词，所以为了提高预设词的语音识别准确率，一般可以对目标词序列对应的第一得分进行调整，以提高目标词序列中预设词被成功识别的概率。

[0045] 在这里，若是基于第一预设词集合中的第一预设词，在语音数据进行解码得到的多个词序列中定位到第一预设词所在的目标词序列。因为此时目标词序列中的第一预设词是识别错误率较高的词语，所以为了提高该第一预设词的识别准确率，就需要对目标词序列对应的第一得分进行调大或升高处理，并在处理之后得到第二得分。显然，所得到的第二得分是大于第一得分的。

[0046] 若是基于第二预设词集合中的第二预设词，在语音数据进行解码得到的多个词序列中定位到第二预设词所在的目标词序列。因为此时目标词序列中的第二预设词是被错误识别出的词语，所以为了降低该第二预设词被识别出的概率，就需要对目标词序列对应的第一得分进行调小或降低处理之后，得到第二得分。显然，所得到的第二得分是小于第一得分的。因为第二预设词为第一预设词被错误识别为的词语，所以采用该方法可以从另一个角度提高第一预设词的识别准确率。

[0047] 步骤280，将第一得分及第二得分中得分最高的词序列，作为语音数据的语言识别结果。

[0048] 最终，未进行得分调整的词序列的得分依然为第一得分，进行了得分调整的目标词序列的得分为第二得分。因此，将所有的词序列按照得分高低进行排序，进而获取第一得分及第二得分中得分最高的词序列，将该得分最高的词序列，作为语音数据的语言识别结果即输出的词序列。

[0049] 本申请实施例中，获取对语音数据进行解码得到语音识别网格lattice，语音识别

网格lattice中包括多个词序列以及每个所述词序列对应的第一得分。根据预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列。对目标词序列对应的第一得分进行调整，得到第二得分，将第一得分及第二得分中得分最高的词序列，作为语音数据的语言识别结果。可以基于预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列，并对目标词序列对应的第一得分进行调整，进而，从得分调整之后的词序列中筛选出得分最高的词序列作为语音识别结果。最终，采用对目标词序列的得分进行调整的方式，实现了对解码得到语音识别结果的过程的干预，进而提高所得到的语音识别结果的准确性。

[0050] 在一个实施例中，如图3所示，步骤240，根据预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列，包括：

[0051] 步骤242，获取预设词集合中所包含的预设词的音素信息。

[0052] 预设词集合中包含人工指定的多个预设词，例如，可以是在语音识别过程中筛选出的100个词语作为预设词，本申请对预设词集合的数量不做限定。其中，音素是一种基本声学单元，是根据语音的自然属性所划分出来的最小语音单位。例如，预设词集合中的预设词“北京”的音素信息为“b”“ei”“j”“ing”。

[0053] 步骤244，将预设词的音素信息与词序列中的音素信息进行匹配。

[0054] 在获取预设词集合中所包含的预设词的音素信息之后，可以将预设词的音素信息与词序列中的音素信息进行匹配。例如，将预设词“北京”的音素信息均为“b”“ei”“j”“ing”，与词序列中的音素信息进行匹配。

[0055] 步骤246，若匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边。

[0056] 因为词序列包括节点与跳转边，且跳转边携带了声学特征的词信息。若预设词的音素信息与词序列中的音素信息进行匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边。

[0057] 如图4所示，为一个实施例中语音识别网格lattice的结构示意图。对待处理的语音数据进行声学特征提取。将所提取的声学特征输入声学模型，计算声学特征的声学模型得分。这里的声学特征包括音素，当然本申请对此不做限定。例如，接收到一段音频信号，并从中依次提取到的声学特征为n，i，h，ao，b，ei，j，ing这八个音素，并从主解码网络中依次获取这八个音素对应的词序列。

[0058] 从主解码网络中依次获取这八个音素对应的词序列的过程中得到过个词序列(图中所示出了3个词序列)，其中一个词序列中包括节点1为起始节点，节点2、3、4、5、6、7、8为中间节点，节点9为终止节点。起始节点与终止节点之间连接了跳转边，跳转边上携带了词信息与音素信息。其中，节点1与节点2之间的跳转边携带了词信息：你好；携带了音素信息为：n。节点2与节点3之间的跳转边携带了词信息：blank；携带了音素信息为：i。节点3与节点4之间的跳转边携带了词信息：blank；携带了音素信息为：h。节点4与节点5之间的跳转边携带了词信息：blank；携带了音素信息为：ao。节点5与节点6之间的跳转边携带了词信息：北京；携带了音素信息为：b。节点6与节点7之间的跳转边携带了词信息：blank；携带了音素信息为：ei。节点7与节点8之间的跳转边携带了词信息：blank；携带了音素信息为：j。节点8与节点9之间的跳转边携带了词信息：blank；携带了音素信息为：ing。

[0059] 将预设词“北京”的音素信息均为“b”“ei”“j”“ing”，与词序列中的音素信息进行匹配。此时，预设词的音素信息与词序列中的音素信息进行匹配成功，则在该词序列中定位到匹配成功的音素信息所在的目标跳转边。即定位至节点5与节点6之间的跳转边、节点6与节点7之间的跳转边、节点7与节点8之间的跳转边及节点8与节点9之间的跳转边，这些跳转边为目标跳转边。

[0060] 本申请实施例中，获取预设词集合中所包含的预设词的音素信息，将预设词的音素信息与词序列中的音素信息进行匹配。若匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边。从而，就可以针对性地对目标跳转边上所携带的模型得分进行调整。以最终达到对目标词序列的得分进行调整的目的，进而采用对目标词序列的得分进行调整的方式，实现了对解码得到语音识别结果的过程的干预，提高所得到的语音识别结果的准确性。

[0061] 在一个实施例中，如图5所示，步骤260，对词序列对应的第一得分进行调整，得到第二得分，包括：

[0062] 步骤402，对目标跳转边上的模型得分进行调整得到新的模型得分。

[0063] 获取预设词集合中所包含的预设词的音素信息，将预设词的音素信息与词序列中的音素信息进行匹配。若匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边。然后，就可以对目标跳转边上的模型得分进行调整得到新的模型得分。具体的，可以对目标跳转边上的模型得分进行增大得到新的模型得分。

[0064] 步骤404，判断目标跳转边上的词信息与预设词是否相同；

[0065] 步骤406，若目标跳转边上的词信息与预设词相同，则将目标跳转边上的模型得分更新为新的模型得分。

[0066] 然后，判断目标跳转边上的词信息与预设词是否相同，若目标跳转边上的词信息与预设词相同，则将目标跳转边上的模型得分更新为新的模型得分。

[0067] 结合图4所示，即为判断节点5与节点6之间的跳转边、节点6与节点7之间的跳转边、节点7与节点8之间的跳转边及节点8与节点9之间的跳转边上的词信息，与预设词是否相同。因为节点5与节点6之间的跳转边上携带的词信息为“北京”，所以得出目标跳转边上的词信息与预设词相同。进而，将目标跳转边上的模型得分进行增大得到新的模型得分。

[0068] 例如，假设跳转边的模型得分均在[0,1]之间，则假设节点5与节点6之间的跳转边的模型得分为0.5，则将节点5与节点6之间的跳转边的模型得分增大至0.55；假设节点6与节点7之间的跳转边的模型得分为0.6，则将节点6与节点7之间的跳转边的模型得分增大至0.7。以此类推，对目标跳转边上的模型得分进行调整，得到新的模型得分。

[0069] 步骤408，基于目标跳转边上新的模型得分，计算得到词序列的第二得分。

[0070] 将各词序列中每个跳转边的模型得分进行求和，得到该词序列的总分。因此，基于各词序列中未调整的跳转边上的模型得分及目标跳转边上调整后新的模型得分进行求和，计算得到词序列的第二得分。

[0071] 本申请实施例中，获取预设词集合中所包含的预设词的音素信息，将预设词的音素信息与词序列中的音素信息进行匹配。若匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边。针对目标跳转边上的词信息与预设词相同的情况，将目标跳转边上的模型得分更新为新的模型得分。最后，基于目标跳转边上新的模型得分，计算得到词

序列的第二得分，以最终达到对目标词序列的得分进行调整的目的。

[0072] 在一个实施例中，在对目标跳转边上的模型得分进行调整得到新的模型得分之后，包括：

[0073] 步骤410，若目标跳转边上的词信息与预设词不同，则在目标跳转边的起始节点与终止节点之间增加新的跳转边。

[0074] 步骤412，配置新的跳转边上的词信息为预设词，并配置新的跳转边上的模型得分为新的模型得分；

[0075] 获取预设词集合中所包含的预设词的音素信息，将预设词的音素信息与词序列中的音素信息进行匹配。若匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边。然后，判断目标跳转边上的词信息与预设词是否相同，若目标跳转边上的词信息与预设词不同，则说明此时目标跳转边上所识别出的词信息为识别错误的词语。因此，在目标跳转边的起始节点与终止节点之间增加新的跳转边，并配置新的跳转边上的词信息为预设词，并配置新的跳转边上的模型得分为新的模型得分。

[0076] 当然，在目标跳转边上的词信息与预设词相同的情况下，也可以在目标跳转边的起始节点与终止节点之间增加新的跳转边，配置新的跳转边上的词信息为预设词，其实也就是目标跳转边上原来的词信息，并配置新的跳转边上的模型得分为新的模型得分。

[0077] 其中，此时新的模型得分可以是根据各个目标跳转边上原来携带的模型得分进行调整所得，具体调整的方法可以与上一个实施例中的方法相同，在此不再赘述。

[0078] 步骤414，基于新的跳转边上新的模型得分，计算得到包含新的跳转边的词序列的第二得分。

[0079] 将各词序列中每个跳转边的模型得分进行求和，得到该词序列的总分。因此，包含新的跳转边的词序列的第二得分为基于该词序列中未调整的跳转边上的模型得分及新的跳转边上调整后新的模型得分进行求和，计算得到包含新的跳转边的词序列的第二得分。

[0080] 如图6所示，为一个实施例中语音识别网格lattice的结构示意图。对待处理的语音数据进行声学特征提取。将所提取的声学特征输入声学模型，计算声学特征的声学模型得分。这里的声学特征包括音素，当然本申请对此不做限定。例如，接收到一段音频信号，并从中依次提取到的声学特征为n,i,h,ao,b,ei,j,ing这八个音素，并从主解码网络中依次获取这八个音素对应的词序列。

[0081] 从主解码网络中依次获取这八个音素对应的词序列的过程中得到，其中一个词序列中包括节点1为起始节点，节点2、3、4、5’、6’、7’、8’为中间节点，节点9为终止节点。起始节点与终止节点之间连接了跳转边，跳转边上携带了词信息与音素信息。其中，节点1与节点2之间的跳转边携带了词信息：你好；携带了音素信息为：n。节点2与节点3之间的跳转边携带了词信息：blank；携带了音素信息为：i。节点3与节点4之间的跳转边携带了词信息：blank；携带了音素信息为：h。节点4与节点5’之间的跳转边携带了词信息：blank；携带了音素信息为：ao。节点5’与节点6’之间的跳转边携带了词信息：北极；携带了音素信息为：b。节点6’与节点7’之间的跳转边携带了词信息：blank；携带了音素信息为：ei。节点7’与节点8’之间的跳转边携带了词信息：blank；携带了音素信息为：j。节点8’与节点9之间的跳转边携带了词信息：blank；携带了音素信息为：ing。

[0082] 然后，判断出目标跳转边上的词信息与预设词不同，则说明此时目标跳转边上所

识别出的词信息“北极”为识别错误的词语。因此，在目标跳转边的起始节点与终止节点之间增加新的跳转边，并配置新的跳转边上的词信息为预设词“北京”，并配置新的跳转边上的模型得分为新的模型得分。该新的跳转边包括节点5”与节点6”之间的跳转边携带了词信息：北京；携带了音素信息为：b。节点6”与节点7”之间的跳转边携带了词信息：blank；携带了音素信息为：ei。节点7”与节点8”之间的跳转边携带了词信息：blank；携带了音素信息为：j。节点8”与节点9之间的跳转边携带了词信息：blank；携带了音素信息为：ing。

[0083] 本申请实施例中，获取预设词集合中所包含的预设词的音素信息，将预设词的音素信息与词序列中的音素信息进行匹配。若匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边。针对目标跳转边上的词信息与预设词不相同的情况，在目标跳转边的起始节点与终止节点之间增加新的跳转边，并配置新的跳转边上的词信息为预设词，并配置新的跳转边上的模型得分为新的模型得分。最后，基于新的跳转边上新的模型得分，计算得到包含新的跳转边的词序列(新的词序列)的第二得分，以最终达到生成了准确率更高的词序列，并对新的词序列的得分进行调整的目的，以提高该新的词序列能够最终被筛选为语音识别结果的概率。

[0084] 在一个实施例中，对目标跳转边上的模型得分进行调整得到新的模型得分，包括：

[0085] 对目标跳转边上的模型得分进行增大预设比例得到新的模型得分。

[0086] 本申请实施例中，目标跳转边的数目可能为多个，因此，在对述目标跳转边上的模型得分进行增大时，可以对多个目标跳转边上的模型得分均进行增大预设比例得到新的模型得分。当然，在其他实施例中，也可以对多个目标跳转边上的模型得分分别进行增大不同预设比例得到新的模型得分。本申请对此不做限定。对目标跳转边上的模型得分进行增大，可以实现提高包含目标跳转边的词序列的模型得分，以最终提高包含目标跳转边的词序列作为语音识别结果的概率。

[0087] 在一个实施例中，提供了一种语音识别方法，还包括：

[0088] 从预设训练语料中获取识别错误率高于预设错误率阈值的预设词；

[0089] 基于预设词得到预设词集合。

[0090] 具体的，在一种情况下，预设词可以是人工在待识别场景下从预设训练语料中所指定的一些语音识别错误率超过预设错误率阈值的命名实体，本申请对此不做限定。例如，对于“北京”这个词语，被识别错误的概率超过了预设错误阈值(例如预设错误阈值为80％)，则将“北京”作为第一预设词。由这些第一预设词就得到了第一预设词集合，其中，可以针对各个不同待识别场景，可以基于各待识别场景中的第一预设词分别形成对应的第一预设词集合。也可以不区分待识别场景，基于很多个待识别场景中的第一预设词形成一个统一的第一预设词集合。

[0091] 本申请实施例中，从预设训练语料中获取识别错误率高于预设错误率阈值的预设词，基于预设词得到预设词集合。然后，可以基于预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列，并对目标词序列对应的第一得分进行调整，进而，从得分调整之后的词序列中筛选出得分最高的词序列作为语音识别结果。最终，采用对目标词序列的得分进行调整的方式，实现了对解码得到语音识别结果的过程的干预，进而提高所得到的语音识别结果的准确性。

[0092] 在一个实施例中，预设词包括基础词及基础词的相似词，基础词的相似词为与基

础词的音素的相似度高于预设相似度阈值的词。

[0093] 其中，基础词为人工从预设训练语料中获取识别错误率高于预设错误率阈值的词语，基础词的相似词为与基础词的音素的相似度高于预设相似度阈值的词。然后，定义预设词集合中不仅包括基础词还包括基础词的相似词。

[0094] 例如，与预设词“北京”的音素相似度高于预设相似度阈值的词为“背景”、“北境”、“倍镜”等，本申请对此不做限定。

[0095] 本申请实施例中，预设词集合中不仅包括基础词还包括基础词的相似词，如此，便对预设词集合进行了扩充，使得预设词集合可以覆盖到更多音素相似度高于预设相似度阈值的词语。从而，提高了音素相似的基础词及基础词的相似词所在的目标词序列，最终被筛选作为语音识别结果的概率。相对而言，则降低了与基础词及基础词的相似词音素不相同的词所在的词序列被筛选作为语音识别结果的概率，进而提高所得到的语音识别结果的准确性。

[0096] 在一个实施例中，如图7所示，步骤220，获取对语音数据进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分，包括：

[0097] 步骤222，对待处理的语音数据进行声学特征提取。

[0098] 其中，语音数据可以是指所获取到的音频信号。具体可以是在语音输入场景、智能聊天场景、语音翻译场景中所获取到的音频信号。对待处理的语音数据进行声学特征提取。声学特征提取的具体过程可以是：将获取到的一维音频信号，通过特征提取算法转化为一组高维向量。所得到的高维向量即为声学特征，常见的声学特征有MFCC，Fbank，ivector等，本申请对此不做限定。Fbank(FilterBank)就是一种前端处理算法，以类似于人耳的方式对音频进行处理，可以提高语音识别的性能。其中，获得语音信号的Fbank特征的一般步骤是：预加重、分帧、加窗、短时傅里叶变换(STFT)、mel滤波、去均值等。且通过对Fbank做离散余弦变换(DCT)即可获得MFCC特征。

[0099] 其中，MFCC(梅尔频率倒谱系数，Mel-frequency cepstral coefficients)，梅尔频率是基于人耳听觉特性提出来的，因此，梅尔频率与Hz频率成非线性对应关系。梅尔频率倒谱系数(MFCC)则是利用梅尔频率与Hz频率之间的这种非线性对应关系，计算得到的Hz频谱特征。MFCC主要用于语音数据特征提取和降低运算维度。例如：对于一帧有512维(采样点)数据，经过MFCC后可以提取出最重要的40维数据，从而达到了降维的目的。其中，ivector为描述每个说话人的特征向量。

[0100] 步骤224，将所提取的声学特征输入声学模型，计算声学特征的声学模型得分。

[0101] 具体的，声学模型可以包括神经网络模型和隐马尔可夫模型，其中，神经网络模型可以向隐马尔可夫模型提供声学建模单元，声学建模单元的粒度可以包括：字、音节、音素、或者状态等。而隐马尔可夫模型可以依据神经网络模型提供的声学建模单元，确定音素序列。一个状态在数学上表征一个马尔科夫过程的状态。且声学模型是预先根据音频训练语料进行训练，所得到的模型。

[0102] 将所提取的声学特征输入声学模型，可以计算得到声学特征的声学模型得分。其中，声学模型得分可以看作是根据每个声学特征下每个音素出现的概率计算得到的分数。

[0103] 步骤226，采用解码算法调用主解码网络及子解码网络，对声学特征及声学特征的

声学模型得分进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每一个词序列对应的第一得分；主解码网络为对原始文本训练语料进行训练所得的解码图，子解码图为对待识别场景中的命名实体进行训练所得的解码图。

[0104] 其中，解码网络用于在给定音素序列的情况下，找到最佳的解码路径，进而可以得到多个词序列以及每一个词序列对应的第一得分。在本申请实施例中，所采用的解码网络包括主解码网络及子解码网络，主解码网络为对原始文本训练语料进行训练所得的解码图，子解码图为对待识别场景中的目标命名实体进行训练所得的解码图。如此，采用主解码网络可以对除去命名实体的音素序列进行解码，采用子解码网络可以对目标命名实体的音素序列进行解码。从而，采用主解码网络及子解码网络，对声学特征及声学特征的声学模型得分进行解码，就得到了多个词序列以及每一个词序列对应的第一得分。

[0105] 其中，待识别场景中的目标命名实体包括待识别场景中的专业词汇，本申请对此不做限定。

[0106] 本申请实施例中，在获取对语音数据进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分时，通过对待处理的语音数据进行声学特征提取，将所提取的声学特征输入声学模型，计算声学特征的声学模型得分。再采用主解码网络及子解码网络，对声学特征及声学特征的声学模型得分进行解码得到语音识别网格lattice。语音识别网格lattice中包括多个词序列以及每一个词序列对应的第一得分。在解码的过程中，并未对待识别场景重新训练解码网络，而是对待识别场景中的目标命名实体进行训练得到子解码图，再采用主解码网络及子解码网络，对声学特征及声学特征的声学模型得分进行解码得到多个词序列以及每一个词序列对应的第一得分。所以，针对待识别场景中的目标命名实体，基于子解码网络就可以对目标命名实体进行准确地解码。且因为未对待识别场景重新训练解码网络，所以大大缩短训练时间长，提高语音识别效率。

[0107] 在一个实施例中，如图8所示，提供了一种语音识别装置800，包括：

[0108] 语音识别网格lattice获取模块820，用于获取对语音数据进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每个所述词序列对应的第一得分；

[0109] 目标词序列定位模块840，用于根据预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列；

[0110] 得分调整模块860，用于对目标词序列对应的第一得分进行调整，得到第二得分；

[0111] 语言识别结果生成模块880，用于将第一得分及第二得分中得分最高的词序列，作为语音数据的语言识别结果。

[0112] 在一个实施例中，目标词序列定位模块840，还用于获取所述预设词集合中所包含的预设词的音素信息；将所述预设词的音素信息与所述词序列中的音素信息进行匹配；若匹配成功，则在所述词序列中定位到匹配成功的音素信息所在的目标跳转边。

[0113] 在一个实施例中，如图9所示，得分调整模块860，包括：

[0114] 模型得分计算单元862，用于对所述目标跳转边上的模型得分进行调整得到新的模型得分；

[0115] 模型得分更新单元864，用于若所述目标跳转边上的词信息与所述预设词相同，则

将所述目标跳转边上的模型得分更新为所述新的模型得分；

[0116] 第二得分计算单元866，用于基于所述目标跳转边上所述新的模型得分，计算得到所述词序列的第二得分。

[0117] 在一个实施例中，得分调整模块860，还包括：

[0118] 跳转边增加单元，用于若所述目标跳转边上的词信息与所述预设词不同，则在所述目标跳转边的起始节点与终止节点之间增加新的跳转边；

[0119] 模型得分配置单元，用于配置所述新的跳转边上的词信息为所述预设词，并配置所述新的跳转边上的模型得分为所述新的模型得分；

[0120] 第二得分计算单元，还用于基于所述新的跳转边上所述新的模型得分，计算得到包含所述新的跳转边的词序列的第二得分。

[0121] 在一个实施例中，模型得分计算单元862，还用于对所述目标跳转边上的模型得分进行增大预设比例得到新的模型得分。

[0122] 在一个实施例中，如图10所示，提供了一种语音识别装置800，还包括：预设词集合生成模块890，用于从预设训练语料中获取识别错误率高于预设错误率阈值的预设词；基于所述预设词得到所述预设词集合。

[0123] 在一个实施例中，所述预设词包括基础词及所述基础词的相似词，所述基础词的相似词为与所述基础词的音素的相似度高于预设相似度阈值的词。

[0124] 在一个实施例中，语音识别网格获取模块220，包括：

[0125] 声学特征提取单元，用于对待处理的语音数据进行声学特征提取；

[0126] 声学模型得分计算单元，用于将所提取的声学特征输入声学模型，计算声学特征的声学模型得分；

[0127] 解码单元，用于采用解码算法调用主解码网络及子解码网络，对声学特征及声学特征的声学模型得分进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每一个词序列对应的第一得分；主解码网络为对原始文本训练语料进行训练所得的解码图，子解码图为对待识别场景中的命名实体进行训练所得的解码图。

[0128] 上述语音识别装置中各个模块的划分仅用于举例说明，在其他实施例中，可将语音识别装置按照需要划分为不同的模块，以完成上述语音识别装置的全部或部分功能。

[0129] 图11为一个实施例中服务器的内部结构示意图。如图11所示，该服务器包括通过系统总线连接的处理器和存储器。其中，该处理器用于提供计算和控制能力，支撑整个服务器的运行。存储器可包括非易失性存储介质及内存储器。非易失性存储介质存储有操作系统和计算机程序。该计算机程序可被处理器所执行，以用于实现以下各个实施例所提供的一种语音识别方法。内存储器为非易失性存储介质中的操作系统计算机程序提供高速缓存的运行环境。该服务器可以是手机、平板电脑或者个人数字助理或穿戴式设备等。

[0130] 本申请实施例中提供的语音识别装置中的各个模块的实现可为计算机程序的形式。该计算机程序可在终端或服务器上运行。该计算机程序构成的程序模块可存储在终端或服务器的存储器上。该计算机程序被处理器执行时，实现本申请实施例中所描述方法的步骤。

[0131] 本申请实施例还提供了一种计算机可读存储介质。一个或多个包含计算机可执行指令的非易失性计算机可读存储介质，当计算机可执行指令被一个或多个处理器执行时，

使得处理器执行语音识别方法的步骤。

[0132] 一种包含指令的计算机程序产品，当其在计算机上运行时，使得计算机执行语音识别方法。

[0133] 本申请实施例所使用的对存储器、存储、数据库或其它介质的任何引用可包括非易失性和/或易失性存储器。合适的非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)，它用作外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDR SDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)。

[0134] 以上实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

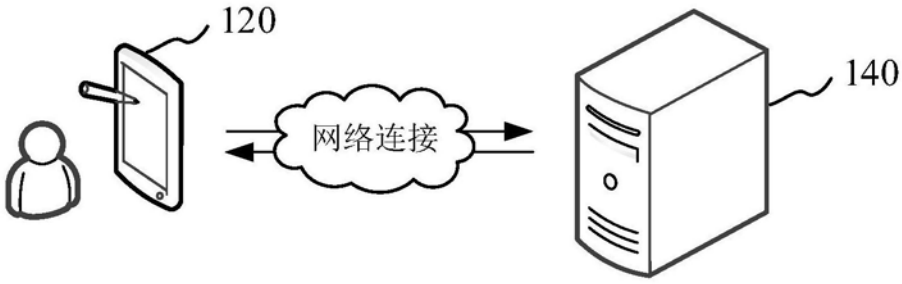

120
网络连接
140

图1

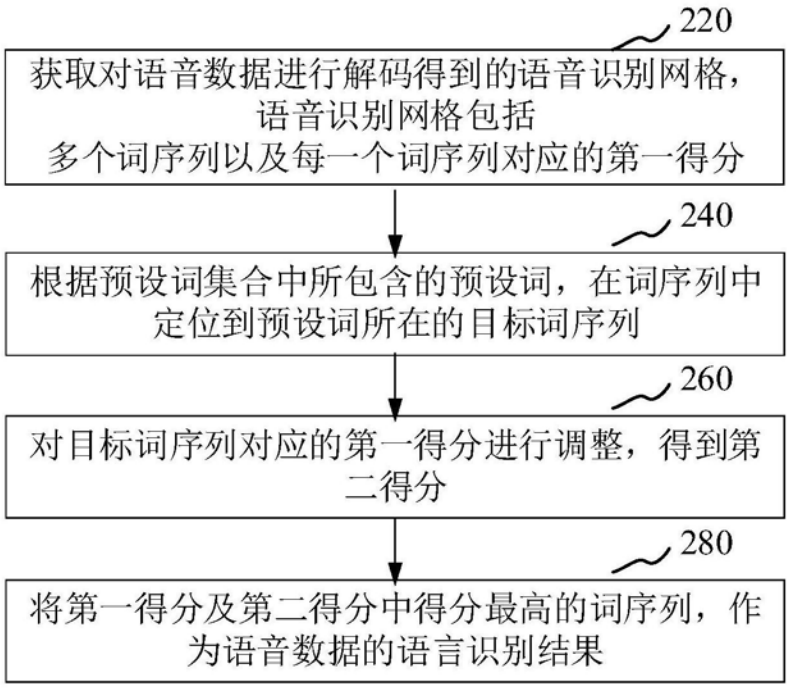

220
获取对语音数据进行解码得到的语音识别网格，语音识别网格包括多个词序列以及每一个词序列对应的第一得分
240
根据预设词集合中所包含的预设词，在词序列中定位到预设词所在的目标词序列
260
对目标词序列对应的第一得分进行调整，得到第二得分
280
将第一得分及第二得分中得分最高的词序列，作为语音数据的语言识别结果

图2

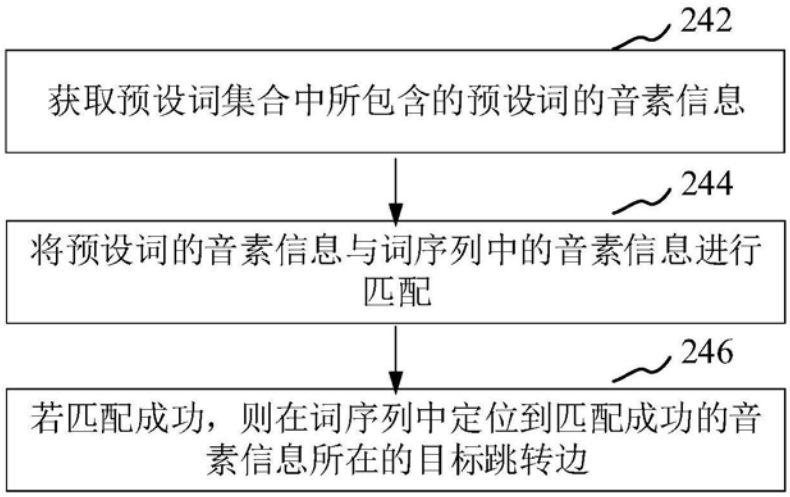

242
获取预设词集合中所包含的预设词的音素信息
244
将预设词的音素信息与词序列中的音素信息进行匹配
246
若匹配成功，则在词序列中定位到匹配成功的音素信息所在的目标跳转边

图3

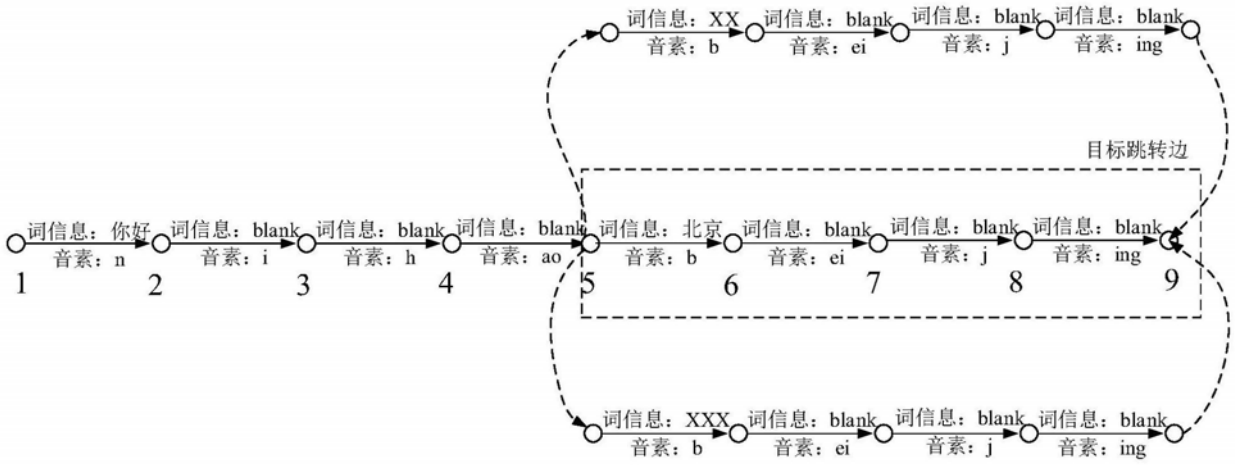

词信息：你好
音素：n
词信息：blank
音素：i
词信息：blank
音素：h
词信息：blank
音素：ao
词信息：北京
音素：b
词信息：blank
音素：ei
词信息：blank
音素：j
词信息：blank
音素：ing
词信息：XX
音素：b
词信息：blank
音素：ei
词信息：blank
音素：j
词信息：blank
音素：ing
词信息：XXX
音素：b
词信息：blank
音素：ei
词信息：blank
音素：j
词信息：blank
音素：ing
目标跳转边
1
2
3
4
5
6
7
8
9

图4

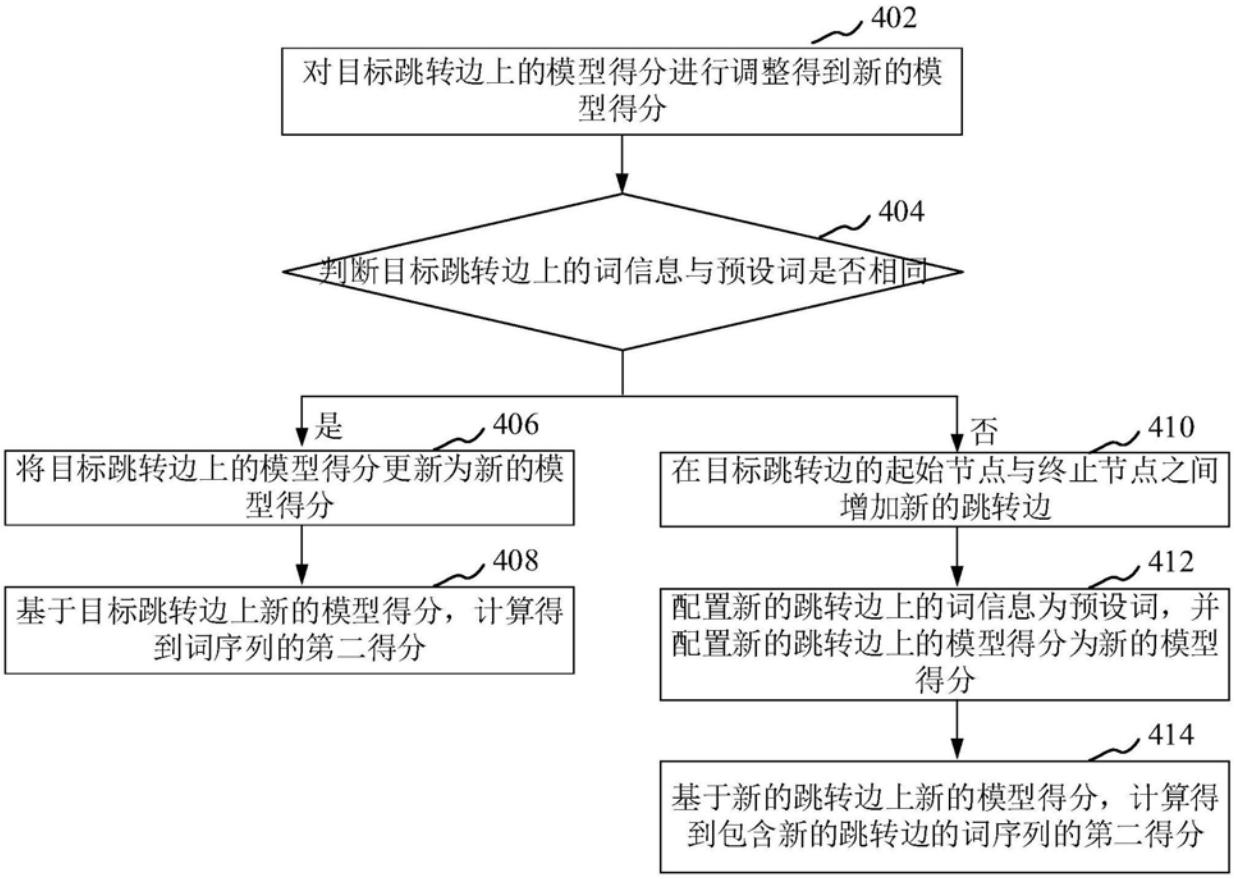

402
对目标跳转边上的模型得分进行调整得到新的模型得分
404
判断目标跳转边上的词信息与预设词是否相同
是
406
将目标跳转边上的模型得分更新为新的模型得分
408
基于目标跳转边上新的模型得分，计算得到词序列的第二得分
否
410
在目标跳转边的起始节点与终止节点之间增加新的跳转边
412
配置新的跳转边上的词信息为预设词，并配置新的跳转边上的模型得分为新的模型得分
414
基于新的跳转边上新的模型得分，计算得到包含新的跳转边的词序列的第二得分

图5

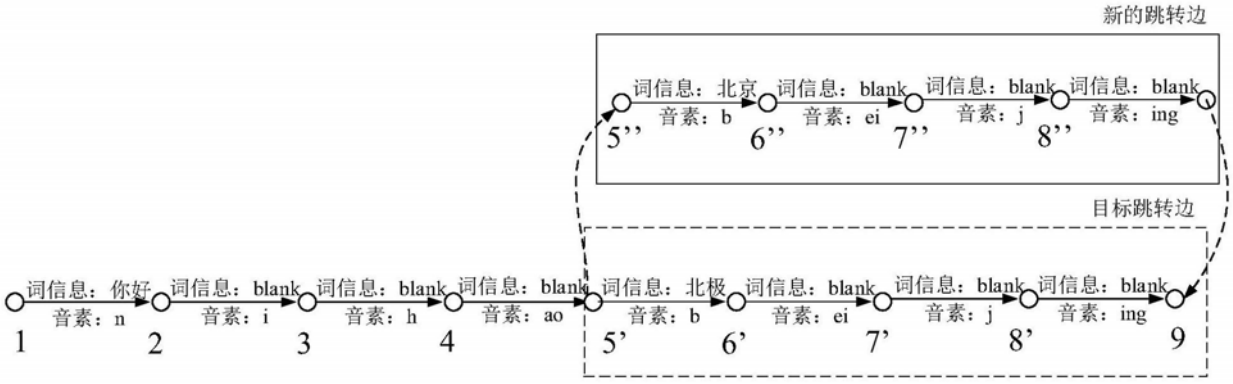

新的跳转边
词信息：北京
音素：b
词信息：blank
音素：ei
词信息：blank
音素：j
词信息：blank
音素：ing
5''
6''
7''
8''
目标跳转边
词信息：你好
音素：n
词信息：blank
音素：i
词信息：blank
音素：h
词信息：blank
音素：ao
词信息：北极
音素：b
词信息：blank
音素：ei
词信息：blank
音素：j
词信息：blank
音素：ing
1
2
3
4
5'
6'
7'
8'
9

图6

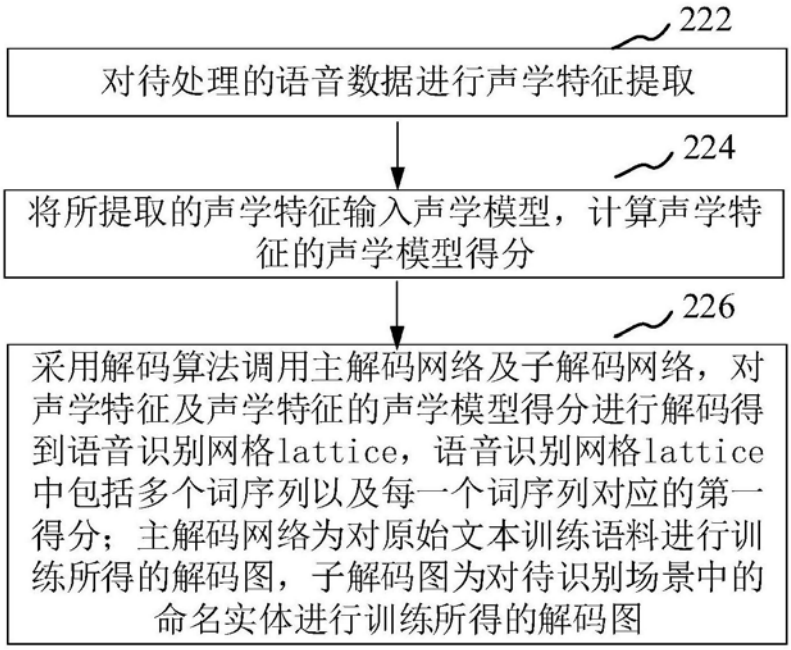

222
对待处理的语音数据进行声学特征提取
224
将所提取的声学特征输入声学模型，计算声学特征的声学模型得分
226
采用解码算法调用主解码网络及子解码网络，对声学特征及声学特征的声学模型得分进行解码得到语音识别网格lattice，语音识别网格lattice中包括多个词序列以及每一个词序列对应的第一得分；主解码网络为对原始文本训练语料进行训练所得的解码图，子解码图为对待识别场景中的命名实体进行训练所得的解码图

图7

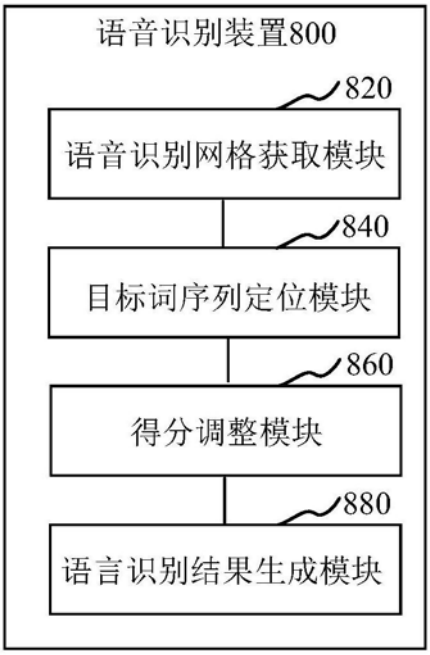

语音识别装置800
820
语音识别网格获取模块
840
目标词序列定位模块
860
得分调整模块
880
语言识别结果生成模块

图8

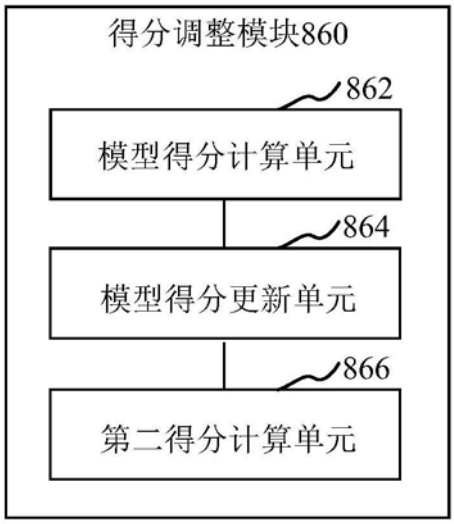

得分调整模块860
862
模型得分计算单元
864
模型得分更新单元
866
第二得分计算单元

图9

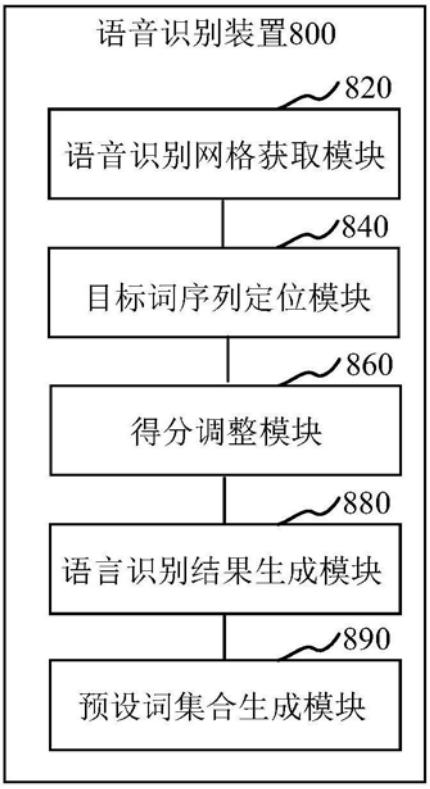

语音识别装置800
820
语音识别网格获取模块
840
目标词序列定位模块
860
得分调整模块
880
语言识别结果生成模块
890
预设词集合生成模块

图10

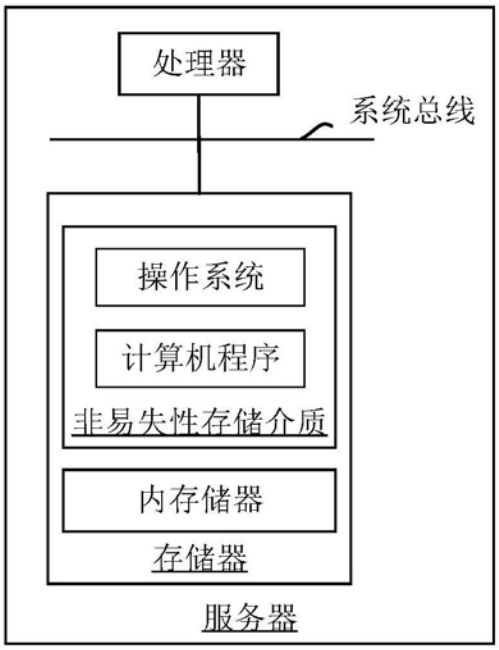


图11