



(12)发明专利

(10)授权公告号 CN 104064183 B

(45)授权公告日 2017.12.08

(21)申请号 201410279788.4

(22)申请日 2014.06.20

(65)同一申请的已公布的文献号

申请公布号 CN 104064183 A

(43)申请公布日 2014.09.24

(73)专利权人 哈尔滨工业大学深圳研究生院

地址 518000 广东省深圳市南山区西丽镇

深圳大学城哈工大校区

(72)发明人 刘明 王明江

(74)专利代理机构 深圳市科吉华烽知识产权事

务所(普通合伙) 44248

代理人 陈本发 刘显扬

(51)Int.Cl.

G10L 15/14(2006.01)

G10L 15/06(2013.01)

(56)对比文件

CN 101030369 A, 2007.09.05,

US 5920839 A, 1999.07.06,

CN 102254087 A, 2011.11.23,

刘德.孤立词语音识别算法优化的研究和实现.《中国优秀硕士学位论文全文数据库 信息科技辑》.2012,(第03期),

胡磊等.一种改进的隐马尔可夫模型在语音识别中的应用.《信息与控制》.2007,第36卷(第6期),

审查员 陈红红

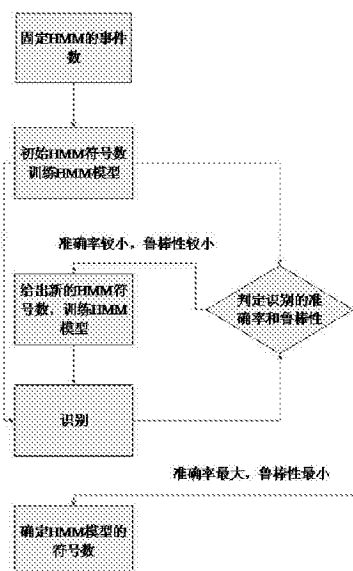
权利要求书1页 说明书3页 附图1页

(54)发明名称

一种基于动态HMM观察符号数的提高语音识别准确率的方法

(57)摘要

本发明为大规模孤立词语音识别提供了一种提高识别准确率的方法,本发明的方法针对不同的孤立词动态的建立了隐马尔科夫模型观察符号数,解决了不同的孤立词因观察符号数相同而识别准确率低的问题。实验结果表明,本发明的方法在稍许增加识别计算量的前提下,有效地提高了大规模孤立词语音识别的准确率。本发明的方法可以动态调整识别模型的参数,相比于传统的基于统计概率的静态模型的语音识别,采用本发明方法的优势在于针对不同用户自适应的调整识别模型的参数,从而提高识别的准确率。待识别孤立词为10240词,实验结果表明,本发明的方法将总的识别率的平均值由96.3%提高到了99.2%。



1. 一种基于动态HMM观察符号数的提高语音识别准确率的方法,其特征在于,所述方法包括以下步骤:

A. 给出初始的HMM模型的参数,所述参数包括事件数和观察符号数,并对模型进行训练,经语音识别过程后,得到一个初始的HMM模型;其中,初始事件数为40,观察符号数目为32,观察序列个数为20,HMM事件跳转概率矩阵为 40×20 ,由观察序列的个数和观察符号数目可以得到一个 20×32 的观察序列概率矩阵;初始事件概率矢量是一个 1×20 的行矩阵;

B. 初始的事件数N取40,动态地改变N值,步长为2,训练过程中改变HMM事件数来观察每个词汇的识别准确率并进行记录,每个词汇识别准确率百分比最高的HMM事件数N作为该词汇的最优HMM事件数;

C. 对于训练所用词库中的每一个词汇,固定步骤B中的HMM的事件数N;动态的改变HMM观察符号数M,步长为2,继续训练得到新的HMM模型,并用训练所用词库中的语音进行孤立词语音识别,待所有词都识别完成后,统计每次改变HMM观察符号数所得到的识别准确率;重复该步骤,找到准确率最大时所对应的HMM观察符号数M;

D. 用户的语音录入后经过特征参数提取,结合步骤C得到的HMM模型参数,经过前向概率计算并给出识别结果后;然后,自动将该用户录入的非标准语音词汇的语音结合词库中对应词汇的语音进行训练,重新改变HMM的观察符号数M,并计算得到针对特定人的最佳HMM观察符号数。

2. 根据权利要求1所述的方法,其特征在于:用待识别孤立词中识别正确的词的数目除以所有的待识别孤立词,得到的结果以百分比来表示所述准确率。

3. 根据权利要求1所述的方法,其特征在于:所述步骤C中,找到准确率最大时所对应的HMM观察符号数M,具体为:待识别孤立词为10240词时,多次识别准确率的最大值为99.2%。

一种基于动态HMM观察符号数的提高语音识别准确率的方法

技术领域

[0001] 本发明涉及孤立词语音识别领域,具体涉及一种提高大规模孤立词语音识别的准确率的方法。

背景技术

[0002] 隐马尔科夫(HMM)模型是一种反映事件跳转概率、观察样本出现概率的非常好的数学模型,因此将语音特征参数按照一定的算法进行处理,得到HMM概率模型。HMM模型由马尔可夫链演变而来,在语音处理的各个领域获得广泛应用。语音的HMM概率模板的建立,需要语音的特征参数向量的聚类编码,语音向量编码,概率模板训练过程进行前向、后向概率计算,直到得到一个收敛的概率模型。

[0003] 声学模型通常是将获取的语音特征参数使用特定的概率算法进行训练后产生。在基于HMM的语音识别中,一个声学模型就是一个HMM模型,通常是将获取的语音特征参数使用HMM概率跳转算法进行训练后产生HMM模型集合。待识别语音经过提取和HMM模型一致的特征参数,采用后向贝叶斯概率算法,计算出后验概率,产生最大的后验概率的HMM概率模板所代表的语音样本即为待识别语音。

[0004] 对语音数据而言,主要是时频采样和频谱变换,有略微时频特性差异的语音都可以建立相应的HMM模型。其次,模型训练就是利用现有的样本对HMM的参数进行调整,使之能够准确描述不同语音对应的语音概率特征。建立模型的过程实际上是对语音做数学建模,并且假定相应的语音特识别概率由这些数学模型计算得到,并且有一个极值。对HMM来说,主要是确定模型的基本拓扑结构,包括事件数目、事件的跳转模式和跳转概率、观察序列概率等。

[0005] 每个汉语孤立词的HMM观察符号数目是不相同的。孤立词的数量越大,数学模型中概率统计的状态越多,对应到HMM参数中,即不同的词汇其对应的HMM符号数目越多。随着孤立词个数的增加,若使用同一个HMM符号数来代表语音帧向量的内在联系,显然会使识别的准确率下降。

发明内容

[0006] 为解决现有技术中存在的问题,本发明提出了一种通过动态改变HMM模型的观察符号数目来提高大规模孤立词语音识别准确率的方法,解决了随着识别孤立词数量的增加而识别准确率下降的问题。

[0007] 本发明通过以下技术方案实现:

[0008] 一种基于动态HMM观察符号数的提高大规模孤立词语音识别准确率的方法,包括以下步骤:

[0009] A. 给出初始的HMM模型的事件数和观察符号数,并对模型进行训练,经语音识别过程后,得到一个初始的HMM模型;其中,初始事件数为40,观察符号数目为32,观察序列个数为20,HMM事件跳转概率矩阵为 40×20 ,由观察序列的个数和观察符号数目可以得到一个20

$\times 32$ 的观察序列概率矩阵;初始事件概率矢量是一个 1×20 的行矩阵;

[0010] B. 初始的事件数 N 取40,动态地改变 N 值,步长为2,训练过程中改变HMM事件数来观察每个词汇的识别准确率并进行记录,每个词汇识别准确率百分比最高的HMM事件数 N 作为该词汇的最优HMM事件数;

[0011] C. 对于训练所用词库中的每一个词汇,固定步骤B中的HMM的事件数 N ;动态的改变HMM观察符号数 M ,步长为2,继续训练得到新的HMM模型,并用训练所用词库中的语音进行孤立词语音识别,待所有词都识别完成后,统计每次改变HMM观察符号数所得到的识别准确率;重复该步骤,找到准确率最大时所对应的HMM观察符号数 M ;

[0012] D. 用户的语音录入后经过特征参数提取,结合步骤C得到的HMM模型参数,经过前向概率计算并给出识别结果;然后,自动将该用户录入的词汇的语音结合词库中对应词汇的语音进行训练,重新改变HMM的观察符号数 M ,并计算得到针对特定人的最佳HMM观察符号数。

[0013] 本发明的有益效果是:根据初始HMM事件数和符号数训练得到的HMM模型进行孤立词语音识别,观察识别的准确率;在固定HMM事件数的情况下,改变HMM观察符号数,继续训练得到新的HMM模型并进行孤立词语音识别,重复上述过程,对比每次不同观察符号数时的识别率,最大值所对应的观察符号数既是最优的;用户通过输入非标准语音词汇,算法通过学习,自适应改变HMM的观察符号数这一参数来使识别的准确率最大。相比于传统的基于统计概率的静态模型的语音识别,采用本发明方法的优势在于针对不同用户自适应的调整识别模型的参数,从而提高识别的准确率。待识别孤立词为10240词,实验结果表明,本发明的方法将总的识别率的平均值由96.3%提高到了99.2%。

附图说明

[0014] 图1是本发明的基于动态HMM观察符号数的提高语音识别准确率的方法的流程图。

具体实施方式

[0015] 下面结合附图说明及具体实施方式对本发明进一步说明。

[0016] 本发明使用的隐马尔可夫HMM模型概率参数如下:

[0017] (1) N , HMM模型中的事件数。HMM模型中事件数是隐含的,在以后的表述中,标记模型中的各个事件为 $\{S_1, S_2, \dots, S_N\}$,在 t 时刻所处的事件为 q_t 。

[0018] (2) M , HMM模型中每个事件下可以观察到序列中的元素的数目,即观察符号数。标记各个观察符号为 $V = \{v_1, v_2, \dots, v_M\}$,观察序列为 $O = \{o_1, o_2, \dots, o_T\}$,其中 o_t 为集合 V 中的一种观察符号, T 为观察序列长度。

[0019] (3) 事件转移概率分布 $A = [a_{ij}]$,其中

[0020] $a_{ij} = p[q_{t+1} = S_j | q_t = S_i] \quad 1 \leq i \leq N, 1 \leq j \leq N$ 。

[0021] (4) 观察序列概率分布 $B = [b_j(k)]$,其中

[0022] $b_j(k) = p[o_t = v_k | q_t = S_j] \quad 1 \leq k \leq M, 1 \leq j \leq N$ 。

[0023] (5) 初始事件概率分布 $\pi = [\pi_i]$,其中

[0024] $\pi_i = P[q_1 = S_i] \quad 1 \leq i \leq N$ 。

[0025] 待识别孤立词中识别正确的词的数目除以所有的待识别孤立词,得到的百分比结

果用于表示识别的准确率。

[0026] 根据Baum-Welch算法由语音特征参数经过聚类计算后得到的编码来计算HMM模型参数时,初始事件概率分布不重要,只要满足概率的和为1即可,只会对计算过程中的迭代次数有轻微影响。因此本发明的初始事件概率分布 $\pi_i = 1/N$ 。

[0027] 在具体的计算实现过程中,本发明采用的算法为贝叶斯前向、后向概率计算和Baum-welch算法,附图1是本发明的方法的实现的流程图,详述如下:

[0028] 1.首先给出初始的HMM模型的事件数和观察符号数;然后,开始对模型进行训练。经语音识别过程后,得到一个初始的HMM模型,此时,该概率模型对不同的词汇不是最优的。模型的参数包括事件数N,每个事件对应的观察符号数M。对离散HMM,观察符号数原则上由样本空间决定,但受计算量的限制,一般可取16~64,经实验,除了某些词,大部分词的M在24~50之间识别的准确率不会有太大波动。

[0029] 本发明采用HMM的初始事件数为40,观察符号数目为32,观察序列个数为20,HMM事件跳转概率矩阵为 40×20 维,由观察序列的个数和观察符号数目可以得到一个 20×32 的观察序列概率矩阵;初始事件概率矢量是一个 1×20 的行矩阵。

[0030] 2.获得语音识别结果的准确率,用待识别孤立词中识别正确的词的数目除以所有的待识别孤立词,得到的结果以百分比来标识所述准确率。初始的事件数N取40,动态改变中步长为2。训练过程中改变HMM事件数来观察每个词汇的识别准确率并进行记录。每个词汇识别准确率百分比最高的HMM事件数作为该词汇的最优HMM事件数。

[0031] 3.对于每个词汇,固定步骤2中的HMM模型的事件数,动态的改变HMM观察符号数M,M的值的改变步长为2。继续训练得到新的HMM模型,并用训练所用词库中的语音进行孤立词语音识别。待所有词都识别完成后,统计每次改变HMM观察符号数所得到的识别准确率。重复该步骤,找到准确率最大时所对应的HMM观察符号数。

[0032] 实验表明,在固定了每个词汇最优的HMM事件数目后,通过改变HMM观察符号数,待识别孤立词为10240词时,多次识别准确率的最大值为99.2%;此时,每个词所对应的HMM的观察符号数就是最优的,若继续变化HMM的观察符号数,不论增大还是减小,准确率都会低于99.2%。

[0033] 4.用户输入语音词汇实现自适应学习,用户的语音录入后,经过参数提取后,结合步骤3得到的词库中的每个词汇的HMM模型参数,经前向概率计算后,得到所有的概率,经过排序,找出最大的概率值,那么此时用户录入的词汇就是最大概率值所对应的那个词库中的词汇。之后,系统使用该用户录入的词汇的语音结合词库中的这个词汇的语音,重新训练得到该词汇的新的HMM模型,即新的观察符号数。

[0034] 至此,本发明的方法实现了语音识别算法自适应地学习调整模型参数。

[0035] 以上内容是结合具体的优选实施方式对本发明所作的进一步详细说明,不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干简单推演或替换,都应当视为属于本发明的保护范围。

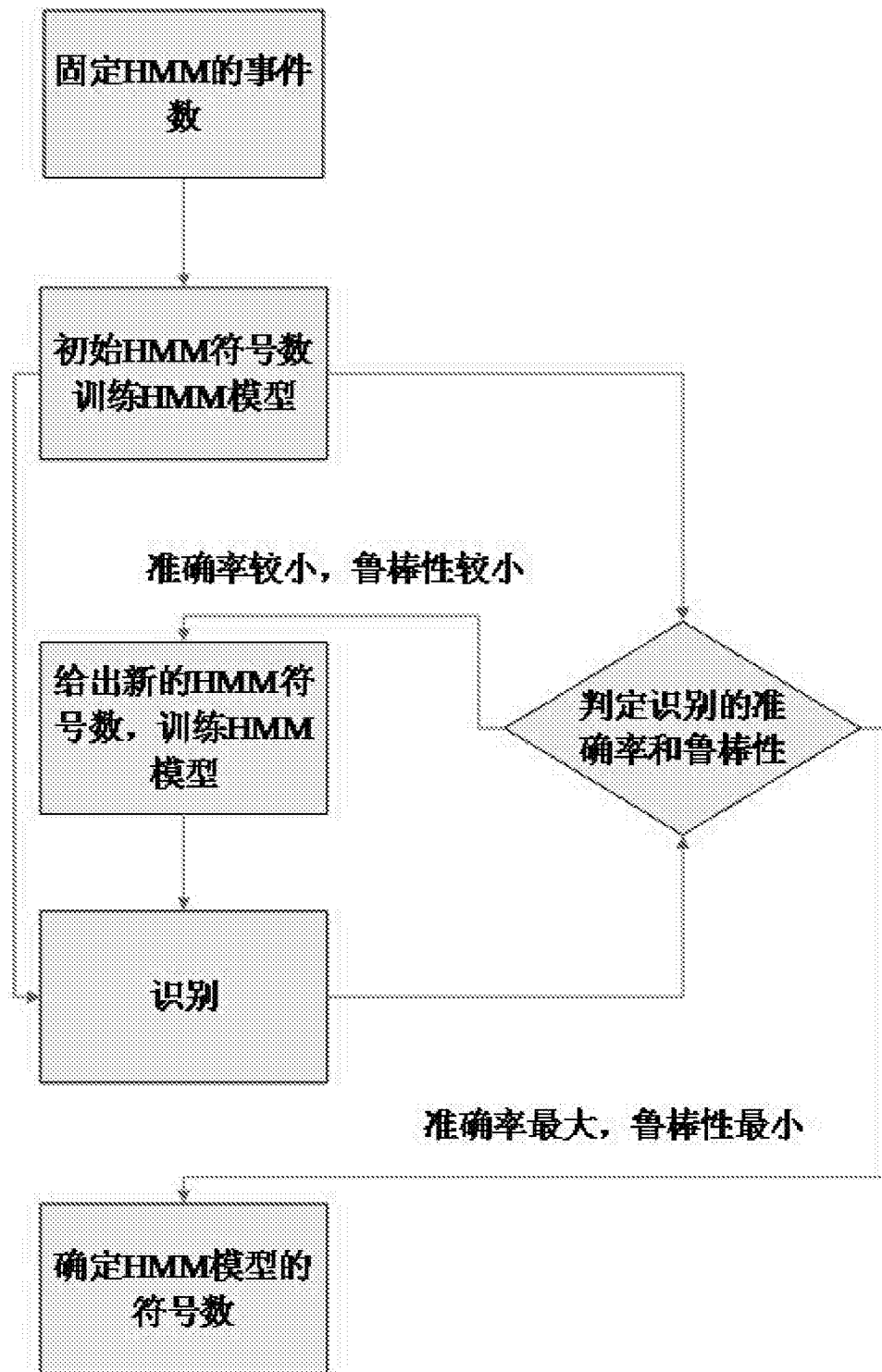


图1