基于HTK的语音识别系统设计

石现峰,张学智,张 峰
(西安工业大学 陕西 西安 710032)

摘 要:HTK是英国剑桥大学开发的一套基于C语言的语音处理工具箱,广泛应用于语音识别、语音合成、字符识别和DNA排序等领域。文中主要介绍了HTK的基本原理和软件结构,并且针对HTK工具箱进行了二次开发,设计开发了一套完整的语音识别输入系统及其相应的测试平台,并验证了该语音识别系统的识别率,实验表明,该系统取得了较好的语音输入效果。

关键词:HTK;语音识别;HMM
中图分类号:TP18 **文献标识码**:A **文章编号**:1673-629X(2006)10-0037-02

Design of Speech Recognition System Based on HTK

SHI Xian-feng, ZHANG Xue-zhi, ZHANG Feng
(Xi'an Technological University, Xi'an 710032, China)

Abstract: HTK is a C language-based toolkit developed by CUED meanly used for speech signal reorganization, speech synthesis, character reorganization, DNA compositor and so on. HTK's general principles and software architecture is discussed in this paper and a suit of speech recognition system is designed based on HTK using further development technology. A test platform is also designed to test this system and gives the correct rate. Experimental results are satisfied.

Key words: HTK; speech recognition; HMM

0 引 言

语音识别是指机器通过学习实现从语音信号到文字符号的理解过程,是一种十分重要的人机交互方式。信息产业的迅速发展促使许多研究机构投入了大量的人力、物力和财力来研究语音识别,这一领域的突破也具有重大的现实意义,让机器能够听懂人类的自然语音可以解决诸如智能机器人、语音输入、低码率语音编码等问题,突破信息处理的一个瓶颈。

HTK(HMM Tools Kit)是一个剑桥大学开发的专门用于建立和处理HMM的实验工具包[1],主要应用于语音识别领域,也可以应用于语音合成、字符识别和DNA排序等领域。HTK经过剑桥大学、Entropic公司及Microsoft公司的不断增强和改进,使其在语音识别领域处于世界领先水平,另外,HTK还是一套源代码开放的工具箱,其基于ANSI C的模块化设计方式可以方便地嵌入到用户系统中。文中介绍了HTK的原理、特点及使用,并在VC环境下设计了一套完整的语音识别及测试系统。

1 HTK原理

HTK工具箱是使用HMM模型作为语音识别的核心的。当HMM应用于孤立词语音识别时,它用不同的隐含状态转移来描述不同的语音发音。对于连续语音识别系统,多个孤立词HMM子模型按一定的语言模型组成的复合HMM模型序列来刻画连续的语音信号,在序列中的每个模型直接对应于相关的发音,并且,每一个模型都有进入和退出状态,这两个状态没有对应的观察矢量,只用于不同模型的连接。

在孤立词语音识别中,对于训练数据,需要为每一个发音单元提供边界信息,常使用手工标注的方法实现。这种方法对于少量的训练数据还可以,对于大量训练数据是不可行的,需要的工作量太大,而且手工标注有时并不是很准确,这会直接影响系统的识别率。但是,对于大量词汇、连续语音识别系统来讲,大量的训练数据是必需的,所以,一般情况下,在连续语音识别的模型训练中,发音单元的边界信息是不需要的,只需要包含相应的发音序列的描述文件。在HTK中使用MLF格式的文件来描述发音序列。训练方法也必需使用嵌入式训练算法,这种算法把样本中前一个模型的退出状态和后一个模型的进入状态按照某种方式连接起来,这样,每一个训练样本就成为了一个组合的HMM模型,在训练时,同时对样本中所有模型的参数进行调整。

HTK的许多功能被编译为一序列的函数库模块,这些模块可以使用相同的接口方式和外界进行交互。HTK的主要函数模块的功能如下:用户的输入输出和与操作系

收稿日期:2006-03-18
作者简介:石现峰(1978-),男,河南人,助教,博士研究生,研究方向为信号与信息处理、嵌入式系统。

统的接口由函数模块 HShell 控制,图形用户界面由 HGraf 提供;内存的管理由 HMem 负责;语音分析所需的处理操作由 HSigP 提供;数学函数的支持由 HMath 提供;HLabel 提供标签文件的接口;HLM 负责语言模型的建立;HNet 负责语法网络的建立;HDict 负责建立词汇的发音词典;HVQ 负责矢量量化 VQ 码本的建立;HModel 负责 HMM 模型的定义和建立。针对所有语音数据的输入输出,HAudio 模块提供了直接音频输入(从声卡)的支持;HWave 模块提供了各种波形数据格式的支持;HParm 提供了几种特征参数文件的格式支持。HUtil 提供了许多有关 HMM 模型的应用例程;HTrain 和 HFB 提供了对多种 HTK 训练工具的支持;HAdapt 为多种 HTK 的自适应工具提供了支持;HRec 包含了 HTK 中主要的识别处理函数[2]。

2 基于 HTK 的语音识别系统设计

2.1 中文语音数据库的建立

语音识别是建立在一定的语料库的基础上的,所以必需构建用于建立、测试语音识别系统的汉语语音数据库。建立语音数据库的工作量巨大,必需使用一定的辅助工具,以提高工作效率,文中使用自行设计的建库软件 WaveManager 构建语音库。语音数据库的目录树结构设计如图 1 所示。

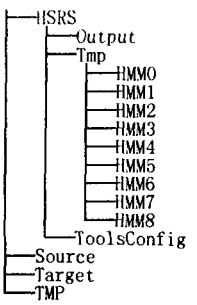


图 1 语音数据库的目录结构

2.2 训练步骤及训练工具的设计

根据 HTK 原理,笔者设计了本语音识别系统的训练步骤:根据语法规则建立发音网络;将波形文件编码;创建语素的标音(Transcription)文件;初始化 HMM 原型;建立 MMF 文件 hmmdefs;使用 HERest 训练 HMM 模型;建立 sp 模型;将 sil 模型和 sp 模型的中间状态捆绑并训练;识别;结果分析;结果输出[3,4]。

语音识别系统的训练若使用手工方式进行,工作量很大。为了提高效率,可以使用脚本使其自动执行,但由于 HTK 的源代码开放,笔者使用 VC 对其进行二次开发,设计了一个功能更为强大的语音识别系统自动建立工具,它能够根据一定的语音数据库及相关的系统配置,按照设计好的训练步骤自动建立一个语音识别系统,并将系统文件输出[5]。其运行主界面如图 2 所示。

2.3 语音识别系统的结构设计及优化

语音识别系统的结构对系统的识别率有很大的影响,为了分析方便,这里只使用 34 个中国省区的语音数据库内容进行分析,其结果具有代表性。

语音识别系统的结构主要表现在以下几个方面:识别单元,可以选取音节、声韵母作为识别单元;HMM 模型,主要是指模型的状态数的不同;语法规则,分为以词为基础的语法规则和以字为单位的语法规则;特征参数。

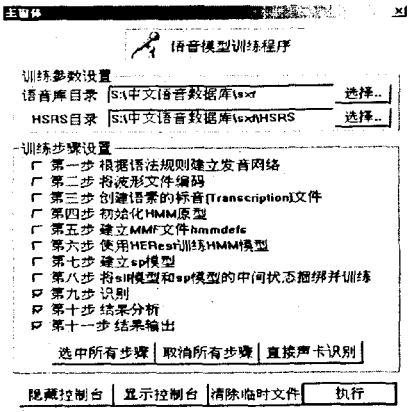


图 2 训练程序运行主界面

通过实验可以得出系统最佳的配置:识别单元使用声韵母,语法规则使用字,HMM 模型状态数为 7,特征参数使用 MFCC_D。在此配置下,对于 34 个省区数据库的识别率可以达到 100%(训练集内)。

2.4 系统测试

使用前面的语音识别系统训练出来的 HMM 模型库及相关配置文件可以直接用于语音输入,为了进一步测试语音识别系统的运行效果,这里使用 VC 设计了语音输入测试平台 SpeechCenter,该平台能够实时检测声卡输入的语音信号,并将其检测到的语音信号识别为汉语拼音,输入到相应的测试区域内。

测试平台的运行主界面如图 3 所示。

经过实际的语音输入测试,该语音识别系统在安静的环境下具有较高的识别率,若配合相应的中文输入法,能够很方便地输入汉字,汉字的语音输入效果良好。

图 3 SpeechCenter 的运行主界面

3 结 语

由于汉语的发音特点,使得汉语的语音输入具有更高的可行性和实用性,汉语的语音输入和处理也是中国进入信息化社会的一个必然要求。由于 HTK 在语音识别方面处于国际领先的地位,并且由于其开放性也更适合于进行二次开发,所以是进行汉语的语音识别研究的一个理想平台,本识别系统的建立和测试也说明了使用 HTK 进行汉语的语音识别研究的优势。

参考文献:

[1] Young S. HTKHistory[EB/OL]. 2005. http://htk.eng.cam.

(下转第 41 页)

$B(4,2)$ $B(1,1)$ 的绝对值都小于等于 5 时,$C(u,v)$ 的定义如下:

$$\begin{bmatrix} * & * & * & * & * & * & 0 & 0 \\ * & * & * & * & * & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & B(1,2) & B(1,3) & B(1,4) & B(1,5) & A(1,1) & 0 & 0 \\ B(2,1) & B(2,2) & B(2,3) & B(2,4) & A(1,2) & 0 & 0 & 0 \\ B(3,1) & B(3,2) & B(3,3) & A(2,1) & 0 & 0 & 0 & 0 \\ B(4,1) & B(4,2) & A(1,3) & 0 & 0 & 0 & 0 & 0 \\ B(1,1) & A(2,2) & 0 & 0 & 0 & 0 & 0 & 0 \\ A(3,1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (6)$$

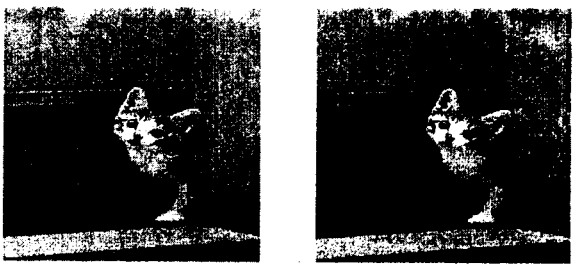
(2) 若当 $A(1,1)$ $A(1,2)$ $A(2,1)$ $A(1,3)$ $A(2,2)$ $A(3,1)$和$B(1,2)$ $B(1,3)$ $B(1,4)$ $B(1,5)$ $B(2,1)$ $B(2,2)$ $B(2,3)$ $B(2,4)$ $B(3,1)$ $B(3,2)$ $B(3,3)$ $B(4,1)$ $B(4,2)$ $B(1,1)$ 的绝对值有任一个是大于 5 的,若为正数则对应 $C(u,v)$ 的位置就用 $10-B(u,v)$ 或 $10-A(u,v)$ 代替,若为负数则对应 $C(u,v)$ 的位置就用 $-(10+B(u,v))$ 或 $-(10+A(u,v))$ 代替。$C(u,v)$ 除了式(6)中的打 * 的部分其他都为 0,由此就得出 $C(u,v)$。由于对大于 5 的都用 10 减了,为了知道哪些数字是减过的,需要附加一个矩阵,这个矩阵和要隐藏的图像的 DCT 系数大小一样。这个矩阵除在减过的地方值为 1 其他地方都为 0,也就是附加的矩阵就只有 0 和 1。这个附加的矩阵可以先保存起来,等要恢复隐藏的图像时再用。这样可以保证存储的值都在 5 以下,对原始图像几乎没有什么影响。以上的式(5.2)就是式(5.1)的附加的矩阵。

3　实验和结论

将以上得到的处理过的 DCT 系数加到要隐藏这幅图像的量化过的 DCT 系数上,就可以达到隐藏的目的了。

下面的几幅图像就是进行实验后的结果,其中图 1 中的图像是 256 * 256 大小的,图 2 中是 512 * 512 大小的。将图 1 隐藏在图 2 中,图 2 是隐藏了图 1 后的图像,可看出它与原图几乎一样,这也与预期的效果一样。图 1 是恢复后的,由此可看出此方法的隐藏效果很好。

因为 JPEG 压缩在量化时失真很大,该隐藏方法在量化后隐藏信息,且将信息隐藏在低、中频段,隐藏信息的数据值也很小,隐蔽性很高。也是这个原因,待隐藏的图像的质量因子可以很高。此方法适用于对灰度图像的隐藏。



(a)　　(b)

图 1　待隐藏的图像(a)和恢复后的图像(b)

(a)　　(b)

图 2　未隐藏信息的图像(a)和隐藏了信息的图像(b)

4　结束语

此算法所隐藏的图像的大小可以与原始图像的一样,且对原始图像的视觉影响很小。既可以隐藏信息,也可以保护信息不被发现。

它还可以有很多扩展,上面只介绍了一种隐藏图像的方法,读者也可以尝试着隐藏文本内容。以上只针对了灰度图像,也可尝试使用彩色图像。

参考文献:

[1] 李　霞.一种基于 JPEG 压缩的信息隐藏方法[J].计算机工程与应用,2003,39(3):164-166.

[2] Cachin C. An information Theoretic Model for Steganography [A]. In: Second International Workshop on Information Hiding, IH'98[C]. Portland, Oregon, USA: [s. n.], 1998. 307-308.

[3] Wang Z. blind measurement of blocking artifacts in images [D]. Austin: the university of texas, 2003.

[4] 林福宗.多媒体技术基础[M].北京:清华大学出版社,2003.

[5] 吴乐南.数据压缩原理与应用[M].北京:电子工业出版社,2003.

(上接第 38 页)

ac.uk/docs/history.html.

[2] Young S, Evermann G, Kershaw D. The HTK Book [EB/OL]. 2005. http://htk.eng.cam.ac.uk/.

[3] 傅国康.语音识别的马尔可夫理论研究[D].西安:西北工业大学,1999.

[4] 尉　洪.基于说话人自适应的连接数字串语音识别[D].昆明:云南大学,2002.

[5] 贲　俊,余小清,万旺根.基于音素的非特定人英语命令词识别算法研究[J].信号处理,2002(6):535-538.