

孤立词语音识别系统的 MATLAB 实现

The MATLAB Realization of Isolated-word Speech Recognition

(湖南大学)杨 熙 苏 娟 彭勇群

YANG XI SU JUAN PENG YONGQUN

摘要:本文建立了一个孤立词语音识别系统,并利用 MATLAB 的语音工具箱 voice box 对系统进行了仿真和分析。实验结果表明,对 20 个孤立词的非特定人识别,准确率在 95%左右。

关键词:孤立词;MATLAB;语音识别

中图分类号:TP391.42 文献标识码:A

Abstract: This article has established an isolated-word speech recognition system, and has carried on the simulation and analysis using MATLAB pronunciation toolbox voice box to the system. Experiments show that the rate of accuracy is about 95% with 20 isolated words of speaker-independent pronunciation.

Key words: isolated-word; MATLAB; speech recognition

1 引言

MATLAB以强大的计算和绘图功能著称,除此之外,还具有易扩性,如语音处理工具箱 voice box 就是由第三方软件开发者免费提供的。

2 系统设计

语音识别的基本流程如图 1 所示:



图 1 语音识别系统框图

从图 1 可知,整个语音识别系统包括四部分:预处理、特征参数提取、训练和识别。其中预处理包括加重和端点检测。预加重一般通过一个数字滤波器 $1-\mu z^{-1}$ 来实现, μ 值接近于 1。

2.1 端点检测

端点检测就是从一段信号中检测出语音信号的起始点和结束点。这里采用基于短时能量和过零率的端点检测算法。这种算法不是实时的,是先读取录制好的 wav 文件,获取信号后再将其分帧并计算短时能量和过零率参数。

$$e(i) = \sum_{n=1}^N |x_i(n)| \quad \text{语音信号第 } i \text{ 帧的短时能量}$$

$$ZCR(i) = \sum_{n=1}^N |x_i(n) - x_i(n+1)| \quad \text{语音信号第 } i \text{ 帧的过零率}$$

为了处理方便,可先将语音信号进行幅度归一化。用命令 $x=x/\max(\text{abs}(x))$ 即可实现。图 2 为孤立词“关机”的端点检测结果。从图 2 能看出,端点检测比较准确,这有利于后面的参数提取等进一步处理。

2.2 特征参数提取

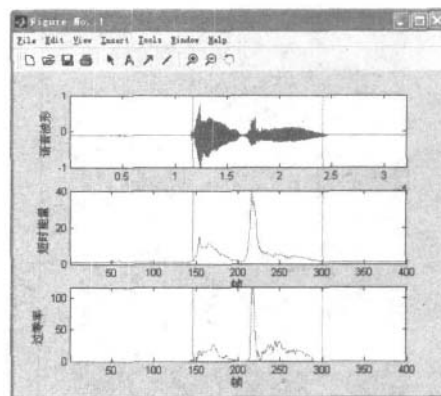


图 2 语音端点检测

常用的参数有线性预测的倒谱系数(LPCC)和 Mel 频率的倒谱系数(MFCC)[1]。LPCC 参数是一种基于合成的参数,而 MFCC 参数考虑了人耳的听觉特性,且没有任何前提假设。大量实验表明,MFCC 参数的性能优于 LPCC 参数。图 3 给出了女声“关机”孤立词的 LPCC 参数和 MFCC 参数的仿真结果。

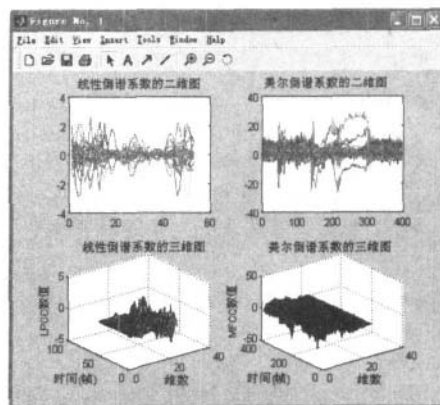


图 3 “关机”孤立词的特征提取

图 3 中仿真用的参数都是 24 维(加上了一阶差分)的语音

杨熙:硕士

特征参数向量。本系统采用 MFCC 系数, 24 个滤波器。

2.3 训练和识别——离散马尔科夫模型(DHMM)

HMM 模型是一种统计模型, 一般分为连续 HMM(CHMM) 和离散 HMM(DHMM)。前者的识别率比较高, DHMM 略低些。但前者的高识别率是以运算量的增加为代价的, 而 DHMM 的模型参数少, 对训练数据量要求不高, 且计算量少, 易于实现, 所以这里采用 DHMM。一个 DHMM 系统可用 A, B 三项参数来描述。

2.3.1 VQ- 码本的设定

在 DHMM 语音识别系统的训练、识别和 CHMM 语音识别系统的训练中, 都要用到矢量量化(VQ)。VQ 的关键技术包括码本设计和搜索策略两方面。

生成码本最基本的算法是 LBG 算法。在利用 LBG 算法之前, 首先需要得到一个初始码本, 采取的方法是分裂法。码本大小为 128, $2^7=128$, 所以经过 7 次分裂就可以了。

对于码本搜索, 简单的方法是比较所有码字和输入矢量的距离, 距离最小的便是目标码字, 这种搜索方法称为“全搜索”。由于本文是进行孤立词识别, 因此选择全搜索算法。

这里说一下的是, 在实践中, 经常会出现空腔, 这对后面的计算是有害的。避免产生空腔的一种方法是在分裂生成码本的算法中, 每次挑最大的腔, 将其一分为二。也有直接舍去空腔的做法。

2.3.2 DHMM 模型初始参数的设定

考虑语音信号的时序特点和系统的实时性, 选用自左向右的无跨越 HMM。对于模型状态数的选择, 状态数太多, 则计算量太大; 太少, 又得不到较好的识别效果。在很多汉语识别实验中, 状态数为 5 时, 识别的效果最好, 本文选取状态数为 5。

初始概率矩阵 π 的初值较易确定, 通常设置为均匀分布之值或非零的随机数, 对识别率的影响不是太大。

$$\pi = \{1, 0, 0, 0, 0\}$$

无跨越自左向右的模型, 状态转移矩阵 A 比较稀疏。当 a_{ij+1} 为状态数的倒数时, 可得到最好的结果。

$$A = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

对 DHMM 等简单的情况, B 的设置较容易, 可采取均匀或随机的设置每一字符的概率初值。码书大小为 128。

$$B = \begin{bmatrix} 1/128 & 1/128 & \dots & 1/128 \\ 1/128 & 1/128 & \dots & 1/128 \\ 1/128 & 1/128 & \dots & 1/128 \\ 1/128 & 1/128 & \dots & 1/128 \\ 1/128 & 1/128 & \dots & 1/128 \end{bmatrix}$$

2.3.3 模板的训练

设定 DHMM 模型的初始参数后, 就可以进行语音模板的训练。由于有大量观察序列参与训练, 在此, 采用修正后的 Baum-Welch 算法进行参数的优化重估。

收敛条件控制着何时结束训练过程。有两种判定方法可以决定模板的 HMM 参数模型是否达到局部最优。一种是前后两

次的输出概率的差值小于一定阈值或模型参数几乎不变为止; 另一种是采用固定训练次数的办法。

2.3.3 语音的识别

识别过程其实就是模板的匹配过程, 而匹配的过程就是利用对数形式的 Viterbi 算法, 计算测试语音序列在每个 HMM 下的最大输出概率。取输出概率最大的 HMM 对应的模型, 即为最后的输出结果。

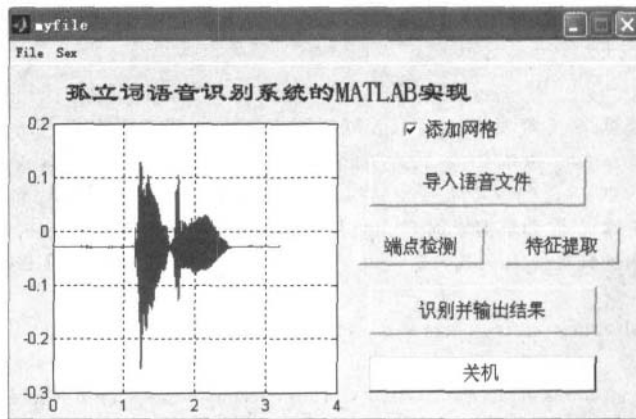


图4 “关机”孤立词的识别

图4的界面是在 MATLAB 里面做的。坐标轴里的波形是“关机”的语音波形。

所有的语音数据都是在普通环境下录制, 以 .wav 文件的格式存储。采样频率 8kHz, 16 位采样, 32ms 一帧, 帧移为 16ms。语音库分为训练集和测试集两部分。训练集包括 32 人语音数据 (16 男 16 女), 20 个普通命令词 (2-3 个词) 组成, 每个人发 6 遍音, 用于训练码本和 DHMM 模型; 测试集包括 16 人 (男 8 人 (4 人参加过训练), 女 8 人 (4 人参加过训练)) 的语音数据。男生和女生的平均正确识别率都在 95% 左右。

3 结束语

对于一个适应性强的语音系统, 一般会加语音增强等环节, 还会有自适应等功能, 也可以采用更好的端点检测算法[6]。本文中的系统比较简单, 主要注重用 MATLAB 来分析和仿真。

本文作者创新点: 尝试设计 MATLAB 环境下的一个完整的孤立词语音识别系统, 并用 MATLAB 进行了仿真和分析。展示了 MATLAB 软件应用于语音处理中的优越性, 利用此软件, 能方便验证系统算法的正确性及整体性能, 有利于成功移植整个系统到 DSP 等嵌入式硬件上, 实现系统的实际应用。

参考文献:

- [1] 何强, 何英. MATLAB 扩展编程[M]. 北京: 清华大学出版社, 2002.
- [2] 杨行峻, 迟惠生. 语音信号数字处理[M]. 北京: 电子工业出版社, 1995.
- [3] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2003.
- [4] 王志飞. 小词汇量非特定人孤立词语音识别的 FPGA 实现[D]. 大连: 大连理工大学, 2005.
- [5] R. L. Joshi et al. A new MMSE encoding algorithm for Vector Quantization. ICASSP, 1991.1: 645-648.
- [6] 江官星, 王建英. 一种改进的检测语音端点的方法[J]. 微计算机信息, 2005, 5(1): 138-139 (下转第 310 页)

5 结论

很多研究只从同一个家族来比较常温蛋白质和高温蛋白质的差别,他们认为不同的家族蛋白质有不同的特性,而作者认为不管是何种蛋白质只要是耐热的,它们在某一方面就有相同的特性,共性的特性使它们都具有耐高温的特点。从预测率可以看出蛋白质一级结构对蛋白质耐热性的影响最大。

尽管耐热微生物是获得热稳定蛋白质的主要来源,但对于那些不能在嗜热微生物中发现的酶类,单点突变和基因敲除可以将常温酶改造为耐热酶,在实验之前可以利用贝叶斯方法进行预测,既可以缩短试验成本又可以节省时间。

本文作者创新点:(1)研究对象是具有全基因组生物的蛋白质,所得到的结论更具有说服力;(2)利用贝叶斯这种方法进行蛋白质耐热性的研究;(3)以氨基酸含量作为输入的特征向量,这样预测率不仅反映贝叶斯方法是否适合蛋白质耐热性预测,而且可以验证氨基酸含量是蛋白质耐热的主要影响因素的结论。

参考文献:

- [1]Karshikoff A, Ladenstein R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule' for hot roads. *Trends Biochem Sci*, 2001, 26(9): 550-556.
- [2]Kumar S, Tsai C J, Nussinov R. Factors enhancing protein thermostability[J]. *Protein Eng*, 2000, 3(3): 179-191.
- [3]Vielle C, Zeikus G J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability[J]. *Microbiol Mol Biol Rev*, 2001, 65(1): 1-43.
- [4]郑启鹏 李秀 刘文煌 李兵. 支持向量机在银行贷款客户分类中的应用研究[J]. *微计算机信息*, 2005, 33: 68-70.
- [5]刘安斐 李弼程. 一种新的 ART 网络遥感影像分类方法[J]. *微计算机信息*, 2005, 31:96-97,123.
- [6]Gu S D, Nie S D, Chen Y, et al. Application study of segmentation algorithm of head MRI based on K-nearest Neighbor rule[J]. *J Shanghai Med Univ*, 2000, 27(2): 108-112.
- [7]Tatusov R L, Koonin E V, Lipman D J. A genomic perspective on protein families[J]. *Science*, 1997, 24; 278(5338): 631-637.
- [8]Tatusov R L, Galperin M Y, Natale D A, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution[J]. *Nucleic Acids Res*, 2000, 1, 28(1): 33-36.
- [9]Tatusov R L, Natale D A, Garkavtsev I V, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes[J]. *Nucleic Acids Res*, 2001, 1, 29(1): 22-28.
- [10]Berman H M, Westbrook J, Feng Z, et al. The Protein Data Bank[J]. *Nucleic Acids Res*, 2000, 1, 28(1): 235-242.
- [11]Li P, Li X M, Tan J L, et al. Non-destructive classification to microstructure in 20 steel high temperature aged based on Bayesian law[J]. *Petro-chemical equipment*, 2004, 33(4): 8-10.
- [12]黄捷 彭宏 林嘉宜. 一种新的正态分布实例的贝叶斯分类算法[J]. *华南理工大学学报(自然科学版)*, 2001, 29(12): 47-49.
- [13]Park K J, Kanehisa M. Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs[J]. *Bioinformatics*, 2003, 1, 19(13):

1656-1663.

[14]Yano J K, Poulos T L. New understandings of thermostable and peizostable enzymes[J]. *Curr Opin Biotechnol*, 2003, 14(4): 360-365.

作者简介:丁彦蕊(1976-),女,山东青岛人,讲师,博士,主要从事生物信息学,人工智能方面的研究;蔡宇杰(1973-),男,江苏无锡人,副教授,博士,主要从事色谱分离,发酵工程方面的研究;孙俊(1971-),男,江苏无锡人,讲师,副教授,主要从事人工智能优化;须文波(1946-),男,江苏无锡人,教授,博士生导师,主要从事生物信息学及系统控制方面的研究。

Biography:Ding Yanrui (1976-), female, Qingdao Shandong province, lecturer, the doctor, is engaged in bioinformatics and artificial intelligence; Cai Yujie (1973-), male, Wuxi Jiangsu province, associate professor, the doctor, is engaged in separating in chromatogram and in research of fermentation; Sun Jun (1971-), male, Wuxi Jiangsu province, lecturer, is engaged in artificial intelligence; Xu Wenbo(1946-), male, Wuxi Jiangsu province, professor, doctoral supervisor, is engaged in the research of bioinformatics and systematic control mainly.

(214122 江苏无锡 江南大学 信息工程学院)丁彦蕊 孙俊 须文波

(214036 江苏无锡 江南大学 工业生物技术教育部重点实验室)丁彦蕊 蔡宇杰

(214036 江南 江南大学 生物工程学院)蔡宇杰
(School of information technology, Southern Yangtze University, Jiangsu Wuxi 214122, China)Ding Yan-rui Sun Jun Xu Wen-Bo

(Key Laboratory of Industrial Biotechnology, Jiangsu Wuxi 214036, China)Ding Yan-rui Cai Yu-jie

(School of biotechnology, Jiangsu Wuxi 214036, China)Cai Yu-jie

通讯地址:(214122 江苏 江苏省无锡市蠡湖大道 1800 号)丁彦蕊
(收稿日期:2007.1.12)(修稿日期:2007.2.15)

(上接第 289 页)

作者简介:杨熙(1982-),女,湖南长沙人,硕士,主要研究方向:语音增强、语音识别;苏娟,湖南大学电子与信息工程学院,副教授,硕士生导师,主要研究方向:单片机,嵌入式系统,语音识别等;彭勇群,硕士,主要研究方向:DSP 应用,语音识别。

Biography:YANG Xi(1982-), master, major in speech recognition and speech enhancement;SU Juan, associate professor, master tutor, major in microprocessor, embedded system and speech recognition;PENG Yongqun, master, major in the application of DSP and speech recognition.

(410082 湖南长沙 湖南大学电气与信息工程学院)杨熙 苏娟 彭勇群

(Institute of Electronics & Information Engineering, Hunan University, Changsha 410082,China) YANG Xi, SU Juan, PENG Yongqun

通讯地址:(410082 湖南 湖南长沙湖南大学 18 舍 406-5)杨熙
(收稿日期:2007.1.12)(修稿日期:2007.2.15)