

文章编号:1001-9081(2006)04-0883-03

一种基于 MFCC 和 LPCC 的文本相关说话人识别方法

于明,袁玉倩,董浩,王哲

(河北工业大学信息工程学院,天津 300130)

(yuming@hebut.edu.cn)

摘要:在说话人识别的建模过程中,为传统矢量量化模型的码字增加了方差分量,形成了一种新的连续码字分布的矢量量化模型。同时采用美尔倒谱系数及其差分和线性预测倒谱系数及其差分相结合作为识别的特征参数,来进行与文本有关的说话人识别。通过与动态时间规整算法和传统的矢量量化方法进行比较表明,在系统响应时间并未明显增加的基础上,该模型识别率有一定提高。

关键词:说话人识别;线性预测倒谱系数;美尔倒谱系数;矢量量化;动态时间规整

中图分类号: TP18; TP391.42 **文献标识码:** A

Text-dependent speaker recognition method using MFCC and LPCC features

YU Ming, YUAN Yu-qian, DONG Hao, WANG Zhe

(School of Information Engineering, Hebei University of Technology, Tianjin 300130, China)

Abstract: In the process of feature extraction of a text-dependent speaker recognition system, the difference of Mel Frequency Cepstrum Coefficient (MFCC) and Linear Prediction Cepstrum Coefficient (LPCC) was chosen to be the speech characteristic parameters, and in the process of speech modeling, a variance was added to the code word of Vector Quantization (VQ) and got continuous vector quantization, then compared it with Dynamic Time Warping (DTW) method and VQ method in text-dependent speaker recognition experiment. The results of identification show that the recognition efficiency is proved without any obvious increasing of responds time.

Key words: speaker recognition; Linear Prediction Cepstrum Coefficient (LPCC); Mel Frequency Cepstrum Coefficient (MFCC); Vector Quantization (VQ); Dynamic Time Warping (DTW)

0 引言

说话人识别是语音识别的一个分支,它和语音识别一样,都是通过对所收到的语音信号进行处理,提取相应的特征或建立相应的模型,然后据此做出判断。说话人识别可以分为两个范畴,即说话人辨认和说话人确认。前者是把要检测的语句判为 N 个训练说话人之一所说,是一个多择一的问题;后者则是把带检测人的语句与其参考说话人的相比较,相符的即得到肯定(确认),不相符的则得到否定(拒绝承认),是二择一的问题。

说话人识别随着统计学的发展和计算机速度的提高,矢量量化(Vector Quantization, VQ),矩阵和分段量化(Matrix and Segment Quantization),马尔可夫模型(HMM)等系统模型得到了极大的重视和发展。

在采用传统的矢量量化方法进行识别系统建模的过程中,码本的生成是特征矢量经过矢量量化聚类而成的,即在训练过程中使用 LBG 算法产生一个离散码字码本,每一个码字为一个均值矢量,即为所有选择这个码字的输入矢量集合的中心。然而在 LBG 算法当中依然存在很多不完善的地方,如初始码本的设置,采用何种畸变度量准则等。这些缺陷对说话人识别系统的识别率势必会造成一定影响,导致识别率的降低。本文从另一个角度出发,为矢量量化模型的码字增加了方差分量,形成连续码字分布的矢量量化(Continuous-code

Vector Quantization, CVQ) 模型,从而,码本的码字由一对矢量来表达,能够更好的反映特征分布的离散程度。经过实验证实,在系统响应时间并未明显增加的基础上,这种连续码字分布的矢量量化模型较之离散码字分布的矢量量化(Distributed Vector Quantization, DVQ)模型要好,使与文本有关的说话人识别系统的识别率有明显提高。

1 特征参量

本文中,与文本相关的说话人识别方法采用了线性预测倒谱系数(Linear Prediction Cepstrum Coefficient, LPCC)及其差分和美尔频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)及其差分相组合作为说话人识别的特征参数。LPCC 体现了每个人特定的声道特性, MFCC 则利用了人耳听觉频率非线性特性,在噪声环境中更能体现其优势^[1]。

1.1 美尔倒谱系数 MFCC 及其差分倒谱系数

MFCC 参数的计算通常采用如下的流程:

(1) 确定每一帧语音采样序列的点数,系统中取 $N = 256$ 点。对每帧序列 $s(n)$ 进行预加重处理后再经过离散 FFT (Fast Fourier Transform) 变换,取模的平方得到离散功率谱 $S(n)$ 。

(2) 计算 $S(n)$ 通过 M 个滤波器 $H_m(n)$ 后得到的功率值,即计算 $S(n)$ 和 $H_m(n)$ 在各离散频率点上乘积之和,得到 M 个参数 $p_m, m = 0, 1, \dots, M-1$ 。

(3) 计算 p_m 的自然对数,得到 $L_m, m = 0, 1, \dots, M-1$ 。

收稿日期:2005-10-19;修订日期:2005-12-23 基金项目:河北省教育厅博士基金资助项目(B2003202)

作者简介:于明(1964-),男,河北秦皇岛人,教授,博士,主要研究方向:数字语音与图像处理、生物特征识别;袁玉倩(1981-),女,天津人,硕士研究生,主要研究方向:数字语音与模式识别;董浩(1977-),男,河北任丘人,硕士研究生,主要研究方向:通信与电子测控技术;王哲(1981-),男,河北石家庄人,硕士研究生,主要研究方向:图像处理与融合。

(4) 对 L_0, L_1, \dots, L_{m-1} 计算其离散余弦变换, 得到 D_m , $m = 0, 1, \dots, M-1$ 。

舍去代表直流成分的 D_0 , 取 D_1, D_2, \dots, D_K 作为 MFCC 参数。最后, 对 MFCC 进行一阶差分, 得到一组新的 MFCC 差分系数, 作为特征矢量的一组分量。

差分参数的计算采用下面的公式^[2,3]:

$$d(n) = \frac{1}{\sqrt{\sum_{i=-k}^k i^2}} \sum_{i=-k}^k i \cdot c(n+i) \quad (1)$$

这里的 c 和 d 都表示一帧语音参数, k 为常数, 通常取 2, 这时差分参数就称为当前帧的前两帧和后两帧的线性组合。

1.2 线性预测倒谱系数 LPCC 及其差分倒谱系数

本文采用的 LPCC 的计算方法是依据全极点模型对 LPC 参数进行递推, 形成 LPC 倒谱, 它的递推式如下:

$$\begin{cases} c_1 = a_1 \\ c_n = a_n + \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k} & 1 < n \leq p \\ c_n = \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k} & n > p \end{cases} \quad (2)$$

式中 a_1, \dots, a_p 为 p 阶 LPC 特征向量, $c_n, n = 1, \dots, p$ 为倒谱的前 p 个值, 当 LPCC 的阶数不超过 LPC 阶数 p 的时候, 用第二式进行计算; 如果 LPCC 阶数大于 p , 则用第三式进行计算, 此时实际上是一种外推。本系统的线性预测模型选择的阶数为 10, LPCC 的阶数选择为 16 阶。则前 10 阶 LPCC 系数通过 10 阶 LPC 迭代计算, 后 6 阶 LPCC 则是通过外推得到的。

线性预测差分倒谱的定义^[2,4]为:

$$c_m(t) = \sum_{i=-k}^k i \cdot c_m(t+i) / \sum_{i=-k}^k i^2 \quad (3)$$

这里的 $c_m(t)$ 和 $c_m(t+i)$ 都表示一帧语音参数, k 为常数, 通常取 2, 这时差分参数就称为当前帧的前两帧和后两帧参数的线性组合。

2 说话人识别的建模

目前使用的说话人识别的方法通常有: 模板匹配法, 概率模型法, 矢量量化方法和神经网络法。

2.1 矢量量化进行说话人识别

在基于矢量量化的说话人识别方法中, 每个说话人的特征用相应的码书来表征, 而码书由从说话人的训练语音序列中提取的特征矢量聚类而成。只要训练序列足够长, 就可以认为该码书包含了这个说话人的个人特征。

采用矢量量化用于说话人识别的过程是: 在训练阶段, 假设系统中第 i 个说话人的训练语音特征矢量集 $Y(i) = \{y_1^{(i)}, y_2^{(i)}, \dots, y_T^{(i)}\}$, 则可用 LBG 算法设计该说话人的码书 $C(i) = \{C_1^{(i)}, C_2^{(i)}, \dots, C_M^{(i)}\}$ 。如果系统中共有 N 个人, 则在训练阶段需要生成 N 个码书 $C(i) (i = 1, 2, \dots, N)$ 。在识别阶段, 首先要从测试语音中提取出特征矢量集 $X = \{x_1, x_2, \dots, x_L\}$, 其中 L 为用于测试的特征矢量的个数。然后需要用系统中 N 个人各自对应的码书对特征矢量集 X 进行矢量量化, 则第 i 个码书对应的总的量化失真定义为^[5]:

$$d^{(i)} = \sum_{l=1}^L \min_{1 \leq m \leq M} (x_l, C_m^{(i)}) \quad (4)$$

根据最小量化失真准则, 总的量化失真最小的码书对应的说话人即为判决结果。本试验中的失真畸变采用的是欧式距离。基于矢量量化的说话人识别系统在识别阶段的流程如

图 1 所示。

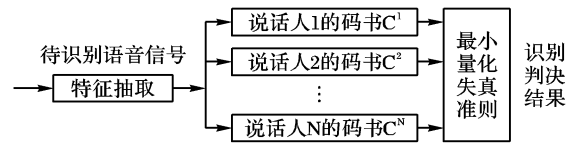


图1 识别阶段流程

2.2 连续分布的矢量量化码本

为了能够提高识别率, 系统为传统的矢量量化模型的码字增加了方差分量, 形成了连续码字分布的矢量量化模型, 经实验证实比原有的离散码字矢量量化模型识别率有所提高。

在系统中首先采用离散码本码字。训练时使用 LBG 算法产生一个离散码字码本, 则每一个码字是一个均值矢量, 代表一个“质心”^[6]。在此基础上, 再计算一下主对角线方差, 就能更好地反映特征分布的离散程度。因此, 码本的第 i 个码字由一对矢量来表达 (Y_i, \sum_i) , 主对角线方差的计算公式为:

$$\sum_i = \sum_{x \in S^{(m)}} (x - y_i^{(m)})^2 \quad 1 \leq i \leq M \quad (5)$$

式中: M 为码字容量。

识别时失真测度用马式距离, 对某一矢量序列 $\{x_1, x_2, \dots, x_T\}$, 对第 k 个码本的总失真如下计算:

$$D_k = \sum_{1 \leq i \leq M} \min \{ (X_i - Y_i^{(k)}) \sum_i^{(k)-1} (X_i - Y_i^{(k)})^T \} \quad (6)$$

式中: $Y_i^{(k)}$ 分别是第 k 个码本的第 i 个码字的均值和方差, M 是码字容量。最后, 选择具有最小 D_k 的码本所对应的人作为识别结果。实验中码本长度选择为 64, 经证实增加了方差分量的码本, 在提高系统识别率方面有一定效果。

2.3 动态时间规整 (Dynamic Time Warping, DTW) 算法

在实验中将语音处理提取的特征参数 LPCC 和 MFCC 按帧存储, 参考模板和测试模板采用同样的方法获取后, 在模板匹配阶段用 DTW 算法计算测试模板和参考模板之间的距离, 其中距离值最小所对应的模板为所求最佳匹配。采用 DTW 算法的依据是倒谱失真测度。

3 试验流程及结果分析

语音数据库: 共 45 人, 30 男、15 女, 均为普通话发音, 说话内容包括 20 个常用的汉语词句, 选用的词语考虑到了汉语中各个元音, 辅音, 摩擦音, 爆破音和鼻音等各个不同的汉语因素, 进行与文本有关的说话人识别。语音信号主要集中在 300Hz ~ 3400Hz, 采用 44100Hz 的采样频率, 采样位数 16 位, 采样通道选用立体声。

实验过程包括语音输入、小波去噪、端点检测和语音特征参数提取, 利用模板进行匹配计算相似度, 最后得出判决结果。其中预处理包括对语音信号进行预加重, 目的在于提升语音高频部分, 使语音信号的频谱区域平坦; 分帧, 取 256 点为一帧, 帧移 128 点; 加汉明窗。至于语音特征参数, 本文选取的是 16 阶 LPCC 及其差分作为一种特征参量, 16 阶 MFCC 及其差分作为另外一种特征参量, 两者结合使用, 有助于提高系统的识别率。

首先选取 DTW 算法进行识别, 并采用了两种工作方式:

(1) 采用 16 阶 MFCC 及其差分, 训练模板, 测试系统形成完整的说话人识别系统。

实验一:录制同一个人的两组发音。

实验二:录制 45 个人的两组发音,每人两组内容相同,不同人有的说相同的内容。

实验三:录制 45 个人的两组发音,每人每组发音内容不同,但是用于识别的发音样本中有内容与用于训练的发音样本内容相同,只是二者不属于同一个发音人。

(2) 采用 16 阶 MFCC 和 16 阶 LPCC 及其差分相结合,训练模板,测试系统形成完整的说话人识别系统。再次进行第一种工作方式中的三个实验步骤。

表 1 采用 DTW 算法建模且不同实验过程时识别率对比

特征参数	实验一	实验二	实验三
MFCC + Δ MFCC	91.1%	88.9%	22.2%
MFCC + Δ MFCC + LPCC + Δ LPCC	91.1%	88.9%	88.9%

通过以上的几次实验,发现当仅采用 MFCC 及其差分作为特征参数进行识别时,如果待测试语音所有者不属于模板数据库中的人员,而该测试者又发出了同模板库中某一合法使用者相同的语音,则系统将其识别为库内的合法者。

基于这个原因,本文引入 LPCC 及其差分,将其作为系统的另一个特征参数。因为 LPCC 体现了每个人特定的声道特性,而不同说话人的声道特性是不同的,所以提取出的 LPCC 也是有差异的。如果有人非法盗用了合法者的口令,则用 MFCC 来识别说话人语音听觉频率特性的时候,将有可能被误识为合法者,但是 LPCC 则可利用说话人的声道特性来识别出其非法者的身份。当有人模仿说话人的声道特性讲话的时候,MFCC 也将判断该说话人的语音听觉频率非线性特性,以加强系统的安全。因此两者相结合使用,达到了比较好的识别效果。

选取 LPCC 和 MFCC 及其差分作为特征参量,并分别采用 DTW, DVQ, CVQ 这三种方法建模来进行上述实验三中的识别:

表 2 采用同种特征参量不同模型时识别率对比

特征参数	DTW	DVQ	CVQ
MFCC + Δ MFCC + LPCC + Δ LPCC	88.9%	91.1%	93.3%

从实验结果中可以看出,DTW 算法较之 VQ 方法的识别率略低,主要原因在于其识别性能过于依赖端点检测,而端点检测的精度随不同的音素而有不同,有些音素的端点检测精度较低,由此影响了识别率。且 DTW 算法由于计算量较大,

在识别过程中需要花费更长的时间。

在码字中增加了方差分量的 CVQ 识别方法使识别率较之传统 VQ 方法有所提高,因此这种证明改进后的方法是比较有效的。

选取 LPCC 和 MFCC 及其差分作为特征参量,观察不同长度的测试语音对 DVQ 和 CVQ 识别率的影响:

表 3 短语长度对系统识别率的影响

文本长度	DVQ	CVQ
1s	88.9%	91.1%
2s	90%	92.5%
4s	94.3%	94.3%
4s 以上	94.3%	94.3%

通过实验,可以看出在语音长度为 4s 左右的时候识别效果好于语音长度为 1s 或 2s 的情况。语音长度在 4s 以上时识别率基本没有明显提高,因此在矢量量化方法当中,选取较长的 4s 左右的语音进行识别,识别效果较好。

4 结语

文中通过提取输入语音的美尔倒谱系数及其差分,线性预测倒谱系数及其差分的双重方法,在说话人辨认的建模过程中,在原有矢量量化模型的码本码字中加入一方差分量,形成一种新的连续码字分布的矢量量化模型,并与传统的动态时间规整算法和矢量量化方法做出比较,进行与文本有关的说话人识别实验,实验结果证明此种方法在识别率和识别所用时间方面都比传统方法有所改善。

参考文献:

- [1] 林宝成,陈永彬. 基于 ARMA 模型的汉语讲话者识别[J]. 声学学报, 1998, 23(3): 229 - 234.
- [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] 何英,何强. Matlab 扩展编程[M]. 北京: 清华大学出版社, 2002.
- [4] FAKOTAKIS N, SIRIGOS J. A high performance text - independent speaker identification system based on vowel spotting and neural nets [A]. Proceedings of IEEE Int Conf on Acoustics, Speech and Signal Processing[C]. Atlanta, GA, USA, 1996.
- [5] 朱民雄,闻新,黄健群,等. 计算机语音技术[M]. 北京: 航空航天大学出版社, 2002.
- [6] 易克初,田斌,付强. 语音信号处理[M]. 北京: 国防工业出版社, 2000.

《计算机应用》重点征稿内容

- 1、数据库技术:XML 数据管理、数据流管理、Web、数据集成、数据仓库、商务智能、数据挖掘等;
- 2、新一代计算:移动计算、网格计算、高性能计算、普适计算、社交计算(Social computing)、仿生计算等;
- 3、信息安全:网络系统防御技术、基于内容的过滤、高可信软件、新一代密码技术、无线安全技术、多方安全计算技术、病毒与垃圾邮件防范技术、网络可生成技术等;
- 4、软件技术与开发:软件重用、软件更新、软件平台、敏捷式软件开发、软件质量保证、软件环境、进化管理、软件测试等;
- 5、系统集成技术:集成系统体系结构、基于 Internet 的应用模式、领域应用模式、中间件、Web 服务技术、应用系统架构、系统协同设计与验证、系统评价与优化、信息系统集成技术等;
- 6、典型应用:移动应用、WLAN、IPv4/IPv6 过渡技术、分布式系统、图形图像处理、多媒体技术与应用、自动识别技术、Web Service 应用、虚拟现实技术、嵌入式技术、计算机在教育领域的应用、数字娱乐技术、先进制造技术、智能交通、地理信息系统、生物医学计算等;
- 7、其他相关新技术与应用。