

一种新的抗噪语音特征的提取

黄亚娟¹, 郑 林²

(1 江苏省产品质量监督检验研究院, 江苏 南京 210007; 2 河海大学 计算机与信息学院, 江苏 南京 210098)

摘 要: 针对 LPCC 易受噪声干扰和不能反映人耳听觉特性的缺点, 提出了新的抗噪声特征, 实验表明, 在各种信噪比(SNR)环境中, 该方法的精度高于线性预测倒谱系数(LPCC)和梅尔倒谱系数(MFCC)。

关键词: 抗噪; 语音特征; 信噪比; 线性预测倒谱系数; 梅尔倒谱系数

中图分类号: TP3

文献标识码: A

文章编号: 1000-7180(2011)10-0215-02

A New Antinoise Feature Extraction of Speech

HUANG Ya-juan¹, ZHENG Lin²

(1 Jiangsu Provincial Supervising & Testing Research Institute for Product Quality, Nanjing 210007, China;

2 College of Computer and Information Engineering, Hohai University, Nanjing 210098, China)

Abstract: As LPCC is easily disturbed by noise and it doesn't contain human auditory characteristics, Proposing a new antinoise feature. The experiment showed that in all kinds of SNR environment, the method has more accuracy than LPCC and MFCC.

Key words: anti-noise; voice feature; SNR; LPCC; MFCC

1 引言

特征参数的选取是语音识别系统中的关键步骤, 人们研究了很多表示语音的特征参数, 其中最常用的特征参数是梅尔倒谱系数和 LPCC 倒谱系数, LPCC 能较好地反映声道的共振峰特性, 但没有充分考虑人耳听觉的非线性, 而且容易受到噪声干扰。根据人类听觉系统对频率及幅度的感知实验结果, 梅尔频率尺度能将声音的频率非线性映射(或称弯折)到一个新的频率尺度, 在此尺度下提取更符合人耳的听觉特性的语音特征; 而语音信号的自相关函数能较好的抗噪能力, 因此本文提出了一个新的抗噪特征, 该特征结合信号的自相关函数, 梅尔频率与 LPCC 形成的, 用 GMM 模型在不同的 SNR 下对说话人进行识别, 获得了较好的识别效果。

2 新的抗噪特征的提取

2.1 自相关函数

语音信号 $s(m)$ 的自相关函数的定义^[1] 为

$$R_n(k) = \frac{1}{N} \sum_{m=0}^{N-1} s_n(m) s_n(m)' \quad (1)$$

其中

$$s_n(m) = w(m)x(n+m) \quad (0 \leq m \leq N-1+k) \quad (2)$$

$$w(m) = \begin{cases} 1, & 0 \leq m \leq N-1 \\ 0, & \text{else} \end{cases} \quad (3)$$

$$s_n(m)' = w(m)'x(n+m) \quad 0 \leq m \leq N-1+k \quad (4)$$

$$w(m)' = \begin{cases} 1, & 0 \leq m \leq N-1+k \\ 0, & \text{else} \end{cases} \quad (5)$$

N 一般取基音周期的 2 ~ 2.5 倍。

2.2 对自相关函数进行线性预测

线性预测的原理是 $R_n(k)$ 的数值, 是前 $m-1$ 个语音点数值的线性组合得到的, 又因为 $R_n(k)$ 的数值是由气流 $X(n)$ 冲激声门发出的语音信号 $S(n)$ 计算得到的, 所以 $R_n(k)$ 的数值也与气流 $X(n)$ 相关, 于是得到以下公式:

$$R_n(k) = G^* X(k) + \sum_{l=1}^m a_l R_n(k-l) \quad (6)$$

两边同时进行 Z 变换, 得到

$$R_n(Z) = G^* X(Z) + \sum_{l=1}^k a_l R_n(Z) Z^{-l} \quad (7)$$

$$H(Z) = \frac{R_n(Z)}{X(Z)} = \frac{G}{1 - \sum_{l=1}^k a_l Z^{-l}} \quad (8)$$

根据 z 变换的公式, 有

$$H(Z) = \sum_{n=-\infty}^{+\infty} h(n) z^{-n} = \sum_{n=0}^{+\infty} h(n) z^{-n} \quad (9)$$

以上两式联立得到

$$h(n) = \begin{cases} a_n + \sum_{l=1}^{n-1} (1-l/n) a_l h(n-l) & (1 \leq n \leq k) \\ \sum_{l=1}^{n-1} (1-l/n) a_l h(n-l) & (n > k) \end{cases} \quad (10)$$

上述公式中 a_l 通过最小二乘法得到, 所计算得到的 $h(n)$ 即为线性预测系数。

2.3 线性美尔尺度变换

$$MC_K(n) =$$

$$\begin{cases} h(n) + \sum_{p=1}^m \alpha_p MC_0(n+p) & k=0 \\ \sum_{p=1}^m \alpha_p MC_{k-1}(n+p) + \sum_{q=1}^m \beta_q MC_k(n+p) & k>0 \end{cases} \quad (11)$$

其中

$$\sum_{q=1}^m (\alpha_p + \beta_p) = 1 \quad (12)$$

$$\alpha_p = \frac{\alpha'_p}{\sum_p \alpha'_p} \quad (13)$$

$$\beta_p = \frac{\beta'_p}{\sum_p \beta'_p} \quad (14)$$

$$\alpha'_p = \frac{1}{1 + \exp(p)} \quad (15)$$

$$\beta'_p = 1 - \alpha_p \quad (16)$$

MC_k 为 Mel 系数, n 为迭代次数, K 为 Mel 系数的阶数, p, q 为尺度变换的阶数, 一般而言, 阶数越大, 越符合人耳的听觉特性, 但并且 p 增加到一定程度之后, 效果就不明显了, 所以一般 K 值为 3~7。取迭代中 n 从大到小取值。上最后求得 ALMCC 的新的语音特征即为 $(MC_0(0), MC_1(0), \dots, MC_{\text{阶数}}(0))$ 。

3 GMM 模型

高斯混合模型是由多个高维高斯密度函数线性

组合得到的^[2], 即:

$$p(x | \lambda) = \sum_{i=1}^M \omega_i b_i(x) \quad i = 1, 2, \dots, M \quad (17)$$

其中

$$b_i(x) =$$

$$\frac{1}{(2\pi)^{D/2} |\sum_i|^{1/2}} \exp\left\{-\frac{1}{2}(X - \mu_i)^t \sum_i^{-1} (X - \mu_i)\right\} \quad (18)$$

X 是一个 L 维的随机变量, ω_i 是 L 维高斯函数 $b_i(x)$ 的权重, 并且满足:

$$\sum_{i=1}^M \omega_i = 1 \quad (19)$$

(ω, μ, Σ) 由 EM^[3] 算法估计出来。

在说话人识别中, λ 表示说话人, X 表示说话人 λ 的语音特征, 则 $p(x | \lambda)$ 表示说话人为 λ 的情况下, 语音特征为 X 概率。

4 识别实验

用 TIMIT 的数据库中的数据, 对说话人进行识别。参与识别的共有 34 个人, 其中 24 男人, 10 女人, 记录 10 s 的语音信息, 分别用 MFCC, LPCC, 新的抗噪特征在不同的 SNR 下识别说话人, 比较识别率的高低。见表 1。

表 1 不同信噪比下, MFCC, LPCC, ALMCC 识别精度

	SMN		准确率
	ALMCC	LPCC	MPCC
1	0.558 82	0.529 41	0.529 41
2	0.647 06	0.470 59	0.558 82
3	0.617 65	0.529 41	0.529 41
4	0.647 06	0.558 82	0.5
5	0.676 47	0.558 82	0.617 65
6	0.705 88	0.529 41	0.558 82
7	0.735 29	0.617 65	0.647 06
8	0.764 71	0.676 47	0.676 47
9	0.794 12	0.676 47	0.735 29
10	0.764 71	0.647 06	0.735 29
11	0.735 29	0.705 88	0.735 29
12	0.794 12	0.705 88	0.705 58
13	0.823 53	0.647 06	0.794 12
14	0.794 12	0.794 12	0.794 12
15	0.882 53	0.823 53	0.802 3
16	0.911 76	0.823 53	0.823 53
17	0.852 94	0.852 94	0.882 35
18	0.932 07	0.852 94	0.823 53
19	0.911 76	0.852 94	0.882 35

(下转第 220 页)

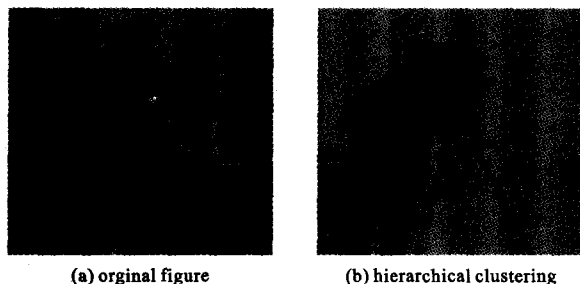


图2 原始图像与传统层次聚类分割效果图

5 结束语

文中对层次聚类算法中心点存在偏移这一不足之处进行了分析研究,并在矫正类中心过程中优化了距离计算方法,最后通过实验证明了传统的层次聚类算法与改进后的算法在标准数据集和图像分割上的有效性.比较发现,改进算法的聚类效果更加准确,符合客观实际.但是,两种算法计算时间与内存开销相当,对于较大数据集存在局限性,因而如何在有限内存开销范围内处理更大规模数据集有待进一步研究.

参考文献:

- [1] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报, 2008,19(1):48-51.

- [2] 史变霞,张明新. 一种改进的层次聚类算法[J]. 微电子学与计算机, 2010,27(12):55-56.
 [3] 胡晓庆,马儒宁,钟宝江. 层次聚类算法的有效性研究[J]. 山东大学学报:工学版, 2010,40(5):146-147.
 [4] Kantardzic M. Data mining concepts, models, methods and algorithms[M]. 北京:清华大学出版社, 2003.
 [5] 段明秀,杨路明. 对层次聚类算法的改进[J]. 湖南理工学院学报:自然科学版, 2008,21(2):28-30.
 [6] 郭晓娟,刘晓霞,李晓玲. 层次聚类算法的改进与分析[J]. 计算机应用与软件, 2008,25(6):243-244.
 [7] Beyer K S, Goldstein J, Ramakrishnan R, et al. When is 'nearest neighbor' meaningful? [C] // Proceedings of the 7th International Conference On Data Theory. London, UK: Springer-Verlag, 1999:217-235.
 [8] Flsher A. Iris plants database[EB/OL]. [2011-04-16]. <http://archive.ics.uci.edu/ml/datasets/Iris>. Html. Authorized license.

作者简介:

王 娟 女, (1988-), 硕士研究生. 研究方向为模式识别、图像处理.

杨绪兵 男, (1973-), 博士, 副教授. 研究方向为模式识别、神经计算和图像处理.

周 宇 女, (1972-), 硕士, 副教授. 研究方向为图像处理、信息融合、模式识别.

周溜溜 男, (1987-), 硕士研究生. 研究方向为数据挖掘.

(上接第 216 页)

5 结束语

提出了一个新的抗噪特征,该特征的提取过程为:先计算语音信号的自相关函数,然后对自相关函数进行线性预测,最后对线性预测后的自相关函数用线性美尔尺度进行变换.用该特征在 GMM 模型中进行说话人识别实验.结果表明,在各种信噪比(SNR)环境中,该方法的精度高于线性预测倒谱系数(LPCC)和美尔倒谱系数(MFCC).

参考文献:

- [1] 赵力. 语音信号处理[M]. 北京:机械工业出版社, 2003.
 [2] 李建文,张晋平. 基于改进语音特征提取方法的语音识别[J]. 微电子学与计算机, 2009,26(007):230-233.
 [3] 朱周华. 期望最大(EM)算法及其在混合高斯模型中的应用[J]. 现代电子技术, 2003(24):16-18

作者简介:

黄亚娟 女, 高级工程师. 研究方向为电子设备检测, 模式识别;

郑 林 男, (1984-), 硕士研究生. 研究方向为语音识别.