

基于 MFCC 特征组合参数的说话人识别研究

曾霞霞^{1,2}, 徐 戈^{1,2}, 吴征远^{1,2}

1. 闽江学院计算机科学系, 福建 福州 350121; 2. 福建省信息处理与智能控制重点实验室, 福建 福州 350121

[摘要] 为提高说话人识别系统的识别率, 提出了一种提取 Mel 频率倒谱系数 (MFCC) 与差分特征组合参数的方法: 先对传统的 MFCC 参数进行特征分量归一化处理, 提升 MFCC 系数的噪声鲁棒性; 再用高斯混合模型 (GMM) 构建了说话人识别系统。使用 TIMIT 语音库进行实验测试, 并比较了不同高斯混合数的 MFCC 特征参数组合对识别率的影响。结果表明: 使用改进的 MFCC 混合参数明显地提高了说话人的识别率。

[关键词] 说话人识别; Mel 频率倒谱系数; 高斯混合模型; 特征参数; 特征向量

[中图分类号] TN 912. 34

Speaker Recognition Research Based on the Combination of MFCC Feature Parameters

ZENG Xia-xia^{1,2}, XU Ge^{1,2}, WU Zheng-yuan^{1,2}

(1. Department of Computer Science, Minjiang University, Fuzhou 350121, China;

2. Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Fuzhou 350121, China)

Abstract: To improve the recognition rate of speaker recognition system, this paper presents a method to extract the combination parameter of Mel frequency cepstral coefficients (MFCC) and accelerated coefficient. The normalized eigenvector of the traditional MFCC feature parameters, improved noise robustness of MFCC coefficient. The paper constructs a speaker recognition system using Gaussian mixture model (GMM), and conducts on TIMIT corpus, and analyzs recognition ratios under different Gaussian mixed numbers (GMN) and the combination of MFCC feature parameters. Experiments show that feature parameters combined improved MFCC with accelerated coefficient are beneficial to improve the speaker recognition rate.

Keywords: speaker recognition; Mel frequency cepstral coefficients (MFCC); Gaussian mixture model (GMM); feature parameter; eigenvector

0 引言

说话人识别是指利用计算机对说话人的声音数据进行分析, 提取声音信号中能表征说话人的特征参数来自动识别说话人身份的技术^[1]。与其他身份认证技术相比, 说话人识别技术利用语音特征的个体独特性的优势, 在考勤系统、安全控制、司法系统等诸多领域得到广泛应用。说话人识别方法中, 常见的流程是先将语音信号经过特征提取, 再对提取的特征进行模式识别。所提取的特征对于后

面的识别结果非常关键,目前常用的特征参数有基音频率 (pitch)^[2]、线性预测系数^[3] (linear prediction coefficients, LPC)、线性预测倒谱系数 (linear prediction cepstral coefficients, LPCC)^[2-3] 以及 Mel 频率倒谱系数 (Mel frequency cepstral coefficients, MFCC)^[3-4] 等。Mel 频率是一种能够反映人耳对不同频率语音的感知能力的一种频率表示方式^[5]。MFCC 就是利用了人耳的听觉原理和倒谱的解的相关特性^[6],因此广泛应用于语音信号的表示中。MFCC 在纯净语音的说话人识别中获得了很好的识别率,但对于带噪语音的识别率还不很理想^[7]。近年来很多学者对传统的 MFCC 提出改进:文献[5]通过重建浊音帧频谱来提取 MFCC;文献[6]用 MFCC 与感知加权线性预测参数的组合特征来提取;文献[7]利用 MFCC 相似度改进了端点检测的准确率;文献[8]使用动态 MFCC 参数来提高说话人识别的效率。由于语音倒谱特征分布的形状在噪声环境下比干净环境下变化大,本文探讨对 MFCC 系数特征分量进行规整来增加 MFCC 对识别率的贡献问题。

1 特征参数的提取

1.1 传统的 MFCC 特征参数提取

提取 MFCC 倒谱参数的常规流程是先对经端点检测后的语音信号进行预加重、分帧、加窗处理,然后通过快速傅里叶变化 (FFT)^[8] 得到信号在频谱上的能量分布,再通过三角 Mel 滤波器组^[9] 得到对应的对数能量,最后经离散傅里叶变换 (DCT) 得出 MFCC。MFCC 是一个含有多阶的矢量集 $C_i, 1 \leq i \leq Z$, 其中 Z 为语音信号的帧数,将所有矢量进行组合 $[C_1, C_2, \dots, C_Z]$ 便得到 MFCC 倒谱系数特征向量组。

1.2 改进 MFCC 特征的组合参数的提取

经传统 MFCC 特征提取后,利用 MFCC 多维特征概率分布的不均衡性和矢量组合 $[C_1, C_2, \dots, C_Z]$ 的关联性,使用倒谱分量均值对倒谱矢量进行方差归一化 (MVN)^[7] 处理来提升 MFCC 系数的噪声鲁棒性,减少信道卷积噪声的影响。经归一化改进特征分量后的 MFCC (MFCC_G) 参数主要反映音频数据的静态特征,而人耳对语音的动态特征更为敏感^[9],因而要提取一阶 MFCC (Δ MFCC) 和二阶 MFCC ($\Delta\Delta$ MFCC),来获取音频信号动态变化的差分特征参数。具体提取方法如下:

1) 某帧语音信号经常规流程提取出的 MFCC 系数,其 T 阶的特征分量表示为 $[X_1, X_2, \dots, X_T]$ 。

首先计算分量各阶的均值 \bar{X} ,再计算方差 $\partial_X = \sqrt{\sum_{i=1}^T (X_i - \bar{X})^2 / T}$,然后使用均值与方差进行归一化处理: $\hat{X} = (X_i - \bar{X}) / \partial_X$ 。

2) 对归一化得到的 MFCC_G 参数,再利用方差计算公式提取其 Δ MFCC 和 $\Delta\Delta$ MFCC。

$$d_i = \begin{cases} C_{i+1} - C_i, & i < k, \\ C_i - C_{i+1}, & i \geq I - K, \\ \sum_{k=1}^K k(C_{i+k} - C_{i-k}) / 2 \sum_{k=1}^K k^2, & \text{其他}, \end{cases} \quad (1)$$

其中: d_i 表示第 i 个一阶差分倒谱系数, I 为维数, C_i 表示第 i 个 MFCC, K 表示一阶倒数的时间差,取值 1 或 2, $1 \leq k \leq K$ 。

3) 将改进后的 MFCC_G 参数与 Δ MFCC 合并为一帧音频信号的参数,即 MFCC_G + Δ MFCC; 以及将 MFCC_G + Δ MFCC + $\Delta\Delta$ MFCC 合为一个矢量作为特征参数。至此组合参数提取结束。

2 基于 GMM 模型构建说话人识别系统

2.1 说话人识别系统的功能

本说话人识别系统分别基于传统的 MFCC 和改进的 MFCC_G 提取特征参数,并引入差分特征参

数组组合重建特征参数矢量，构建了6套特征参数不同的说话人识别方法，每套流程都实现了声纹特征提取和声纹识别两块功能。声音信号特征提取后的识别过程分训练和识别两个阶段进行。在训练阶段，系统读取已知说话人语料库的声音数据，提取声音中能够反映该说话人特征的信息，根据所提取的特征参数为该说话人建立一个模型，将其保存在模式库中。在识别阶段，系统读取待识别的说话人的声音数据，提取特征参数，然后与模式库中的模型进行比较，根据一定的相似性准则求得模式库中与其最匹配的模型，并将对应的说话人名字作为识别结果输出。这里需要强调的是用于训练的声音数据和识别的声音数据使用的是同一说话人的不同内容的声音素材。

2.2 基于 GMM 模型的说话人识别方法

采用高斯混合模型（Gaussian mixture model, GMM）的声纹识别方法，先通过特征参数训练建立 GMM 模型，然后根据提取的说话人特征序列从模式库中进行匹配。GMM 参数模型的训练过程也就是描述说话人的声音特征矢量在特征空间分布的过程。假设某个说话人用于训练的特征参数矢量序列 $\mathbf{X} = \{\mathbf{x}_t, t = 1, 2, \dots, T\}$ 中各矢量 \mathbf{x}_t 是独立不相关的，寻找一个模型参数使得相应的似然函数值最大^[10]。基于 GMM 的说话人识别阶段，首先需要提取待识别说话人声音信号中的特征参数矢量序列，然后从模式库中找出与之最匹配的模型。假设共有 S 个说话人，在模式库中对应的 GMM 模型分别为 $\lambda_1, \lambda_2, \dots, \lambda_S$ 。识别的具体做法就是对所提取得到的特征参数 \mathbf{X} 从下述模型中找到具有最大后验概率的模型所对应的说话人。说话人识别的目标转化为计算
$$\hat{S} = \underset{1 \leq k \leq S}{\operatorname{argmax}} \sum_{t=1}^T \log P(\mathbf{x}_t | \lambda_k) = \underset{1 \leq k \leq S}{\operatorname{argmax}} \sum_{t=1}^T \left[\log \sum_{i=1}^M \omega_i p_i(\mathbf{x}_t; \mu_i, \Sigma_i) \right]$$
，其中： ω_i 为混合权重，满足 $\sum_{i=1}^M \omega_i = 1$ ； $P_i(x, \mu_i, \Sigma_i)$ 为高斯函数，表示 GMM 的第 i 个高斯分量； μ_i 为第 i 个高斯分量的均值矢量； Σ_i 为第 i 个高斯分量的协方差矩阵；GMM 模型 $\lambda_k = \{\omega_{ik}, \mu_{ik}, \Sigma_{ik}\}, i = 1, 2, \dots, M$ ， M 为高斯模型混合数。

3 实验分析

3.1 实验设置

基于 MFCC 和 GMM 构建说话人识别系统，提出6种条件的测试方案，使用 TIMIT 语音库中部分说话人的语音文件进行仿真实验。从 TIMIT 不同方言库中抽取100个说话人（50 男性 50 女性）每人10条语音素材，采样率为8 kHz，采样精度为16 bit，其中7条语句用于训练和3条语句用于识别测试。特征参数提取了12 维的 MFCC，并进一步提取其差分一阶与二阶 MFCC。GMM 模型构建中通过期望最大化算法（EM 算法）^[11]进行最大似然估计，数据训练中使用的高斯混合数（Guassian mixture number, GMN）分别取值8、16、32 进行不同的实验测试。

3.2 实验结果分析

3.2.1 基于传统 MFCC 参数的识别率

对语音库中100人的语音数据进行常规流程处理，分别提取 MFCC 倒谱系数，并计算 Δ MFCC 和 $\Delta\Delta$ MFCC，使用差分参数组合（如 MFCC + Δ MFCC）进行识别测试，结果见表1。

表 1 基于传统与改进 MFCC 的组合参数的识别率

Tab. 1 Recognition rates based on combination parameters of traditional and improved MFCC

组合参数 Combination parameters	识别率 Rate/%		
	GMN = 8	GMN = 16	GMN = 32
MFCC	86	91	92
MFCC + Δ MFCC	89	92	94
MFCC + Δ MFCC + $\Delta\Delta$ MFCC	87	92	93
MFCC_G	90	92	94
MFCC_G + Δ MFCC	92	93	95
MFCC_G + Δ MFCC + $\Delta\Delta$ MFCC	93	95	96

从表 1 中可以看出, 基于传统 MFCC 和 GMM 的语音识别系统可以取得较好的说话人识别效果, 使用 MFCC + Δ MFCC 组合参数比 MFCC 系统的识别率明显提高, 但 MFCC + Δ MFCC + $\Delta\Delta$ MFCC 组合参数并没有给系统带来优越性, 反而增加了计算量并降低了识别的性能。从另外一个角度, 高斯混合数的取值也会影响系统的识别率, 其中取值 32 效果最好。

3.2.2 改进 MFCC 特征分量的识别结果

对语音库中 100 人的语音数据分别提取改进 MFCC_G 倒谱系数, 同样使用其一阶差分参数组合 (如 MFCC_G + Δ MFCC) 和二阶差分参数组合 (MFCC_G + Δ MFCC + $\Delta\Delta$ MFCC) 分别进行识别测试。

对表 1 数据分析, 认为在提升系统识别率上 MFCC_G 优越于 MFCC。使用 MFCC_G + Δ MFCC 与 MFCC_G + Δ MFCC + $\Delta\Delta$ MFCC 参数组合比单独使用 MFCC_G 参数能获得更高的识别率。实验结果验证了, 在说话人识别系统中采用改进的 MFCC_G 组合参数有益于提高系统的识别率。

4 结束语

介绍了一种 MFCC 与差分特征参数组合的混合参数的提取方法, 以及基于高斯混合模型构建了说话人识别系统。实验结果验证了改进分量的 MFCC_G 混合参数用于说话人识别中的有效性。MFCC 特征参数在语音识别领域应用非常广泛, 引入差分特征参数的组合可以在一定程度上提高识别率, 但特征维数的叠加必然增加了计算复杂度, 如何提升识别的时效性还有待于进一步的研究。

[参考文献]

- [1] BERITELLI F, CASALE S, RUGGERI G, et al. Performances evaluation and comparison of G.729/AMR/fuzzy voice activity detectors. IEEE Signal Processing Letters, 2002, 9(3): 85-88. DOI:10.1109/97.995824.
- [2] HASAN M R, JAMIL M. Speaker identification using Mel Frequency Cepstral Coefficients//ICECE: 3rd International Conference on Electrical & Computer Engineering. Dhaka, Bangladesh: IEEE Press, 2004: 28-30.
- [3] BURGET L, MATEJKA P, SCHWARZ P. Analysis of feature extraction and channel compensation in a GMM speaker recognition system. IEEE Transactions on Audio, Speech and Language Processing, 2007, 15(7): 1979-1986. DOI:10.1109/TASL.2007.902499.
- [4] ITO A, AIBA A, ITO M, et al. Detection of abnormal sound using multi-stage GMM for surveillance microphone//Proceedings of the 2009 Fifth International Conference on Information Assurance and Security. Xian, China: IEEE Computer Society, 2009(1): 733-736. DOI:10.1109/IAS.2009.160.
- [5] 胡政权, 曾毓敏, 宗原, 等. 说话人识别中的 MFCC 参数提取的改进. 计算机工程与应用, 2014, 50(7): 217-220. DOI:10.3778/j.issn.1002-8331.1206-0216.
- [6] 周国鑫, 高勇. 基于 GMM-UBM 模型的说话人辨识研究. 无线电工程, 2014, 44(12): 14-17. DOI:10.3969/j.issn.1003-3106.2014.12.05.
- [7] 邓瑞, 肖纯智, 高勇. 基于 MFCC 相似度和谱熵的端点检测算法. 现代电子技术, 2013, 36(21): 67-69. DOI:10.3969/j.issn.1004-373X.2013.21.018.
- [8] 王华朋, 杨军, 吴鸣, 等. 基于自适应同源方差控制的法庭自动说话人识别. 应用科学学报, 2014(6): 582-586. DOI:10.3969/j.issn.0255-8297.2014.06.006.
- [9] ZHANG J, WEI G. Robust multi-stream speech recognition based on weighting the output probabilities of feature components. 声学学报 (英文版), 2009, 33(3): 269-279.
- [10] PELECANOS J, POVEY D, RAMASWAMY G. Secondary classification for GMM based speaker recognition//2006 IEEE International Conference on Acoustics, Speech, and Signal Processings. Toulouse, France: IEEE Press, 2006: 109-112.
- [11] 蒋晔. 基于短语音和信道变化的说话人识别研究. 南京: 南京理工大学, 2013.

(责任编辑 朱雪莲 英文审校 黄振坤)