短语音说话人辨认的研究

蒋 晔,唐振民
(南京理工大学计算机科学与技术学院,江苏南京 210094)

摘 要: 针对短语音说话人辨认训练语料不充分的特点,对特征参数和GMM模型进行优化和改进,提出一种基于局部模糊PCA的GMM说话人辨认方法.该方法采用特征组合代替单一特征,以提高有效特征维数来弥补特征样本的不足,并用局部模糊PCA对组合特征进行有效降维,在对识别率影响很小的前提下,降低了系统的时空复杂度.本文还对GMM参数初始化方法进行改进,采用分裂法与模糊k均值聚类相结合方法.实验表明,与传统初始化方法相比该方法能有效提高短语音说话人辨认性能.

关键词: 说话人辨认; 短语音; 局部模糊主成分分析; 分裂法与模糊k均值聚类相结合
中图分类号: TN913.2 **文献标识码:** A **文章编号:** 0372-2112 (2011) 04-0953-05

Research on the Speaker Identification Based on Short Utterance

JIANG Ye, TANG Zhen-min
(*School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu* 210094, *China*)

Abstract: For the inadequate training speech data of speaker identification based on short utterance, feature vectors and GMM models are optimized and improved, an efficient GMM based on local PCA with fuzzy clustering is presented. To compensate for the limited feature samples, the effective feature dimensions are increased with feature combinations instead of single feature. Furthermore, the time and space complexity of the system can be compressed by reducing dimensions of feature combinations with local fuzzy PCA in the premise of little effect on recognition rate. Finally, a new approach which combines division and fuzzy k-means clustering is used, in order to optimize GMM initialization parameters. The experiments show that the improved method is more effective in improving performance of the system than traditional initialization methods.

Key words: speaker identification; short utterance; local fuzzy principal component analysis; combined division and fuzzy k-means clustering

1 引言

说话人识别是利用人类的声音来确定或鉴别说话人身份的技术.按其识别任务可分为:说话人辨认和说话人确认[1,2].近年来,说话人识别技术逐步向实际应用发展,但由于无法找到仅反映说话人个性的特征,需要提高训练语料时间来弥补这一缺陷.长时的语音文本、用户不愿长时训练、有时缺乏长时语料等,已成为说话人识别技术商业化的一大阻碍.因此利用尽可能少的训练数据建立有效的说话人模型,实现高性能的说话人识别,更具现实意义[3].

短语音说话人辨认,因其训练语料时长较短,导致特征样本不足,识别性能下降.本文采用特征组合形成高维特征来弥补这一不足.实验表明,增加特征有效维数能提高识别性能.然而,特征维数的增加,也意味着需要更多的模型参数来描述说话人的特征分布.从而加大了时空复杂度.

为了减少特征维数和特征各维之间的相关性,Jolloffe等人提出了主成分分析(Principal Component Analysis,PCA)这一理论[4].PCA是一种特征提取方法,通过变换把原始空间投影到更小的子空间,从而降低特征维数[5].Kambhatla和Leen首先提出VQPCA模型,用VQ把数据分割成不相交的几个类,然后对每个聚类中心进行局部PCA处理[6].说话人辨认方面,Seo等人提出基于局部PCA的高斯混合模型(Gaussian Mixture Model,GMM)[7].本文在语料短缺情况下,引进分类隶属度因子,提出基于局部模糊PCA的GMM说话人辨认方法.实验表明,该方法在3s训练的条件下识别率达80%左右.

2 特征提取

Mel频率倒谱系数(Mel Frequency Cepstrum Coefficient,MFCC)和线性预测倒谱系数(Linear Prediction Cepstrum Coefficient,LPCC)是说话人辨认中最常用的两种特征参数[8].但这两种特征都只考虑到语音帧内的信息,而没有考虑到语音帧之间的信息.获取语音帧之间的时变信息,能够提高说话人辨认的性能[9].

收稿日期:2010-04-01;修回日期:2010-06-11

Delta 特征是一种能够反映语音帧之间时变信息的动态特征,其计算如下:

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta(c_{t+\theta} - c_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{1}$$

其中,d_t 表示第 t 帧特征的 Delta 特征,Θ 表示第 t 帧时序变化的语音帧的数量(本文取 2).

针对短语音说话人辨认训练样本不足的特点,我们先计算 LPCC 和 MFCC 的最佳识别维数,然后基于一个自然的考虑,把这两组有效特征及 Delta 特征组合在一起,从而得到在一定特征范围内的最优特征组合.

3 局部模糊 PCA 降维

3.1 模糊 k 均值聚类算法及其改进

假设有一 P 维空间特征矢量集 $\boldsymbol{X} = \{x_1, \cdots, x_T\}$,模糊 k 均值聚类指定了每一特征矢量在不同类中的隶属程度,可用 $K \times T$ 的矩阵 $\boldsymbol{U} = [u_{jt}]$ 来表示,其中 u_{jt} 表示 x_t 在第 j 类 R^j 的隶属度函数.其函数有如下性质[10]:

$$0 \leqslant u_{jt} \leqslant 1, \quad j = 1,2,\cdots,K, \quad t = 1,2,\cdots,T \tag{2}$$

$$\sum_{j=1}^{K} u_{jt} = 1, \forall t, 0 < \sum_{t=1}^{T} u_{jt} < T, \forall j \tag{3}$$

式(2)表明每一个特征样本 x_t 在 K 个聚类中都存在一个隶属度函数,式(3)要求每个 x_t 对于各个聚类的隶属度之和为 1, u_{jt} 可大致理解为 x_t 属于 j 类中的概率[11,12].

模糊 k 均值聚类算法是基于聚类损失函数的最小化,其公式如下:

$$J_m = \sum_{t=1}^{T}\sum_{j=1}^{K} (u_{jt})^m d^2(x_t, c_j), K \leqslant T \tag{4}$$

其中,$m > 1$ 是一个可以控制聚类结果的模糊程度的常数;c_j 是第 j 个聚类的中心;$d^2(x_t, c_j)$ 代表 x_t 与 c_j 之间的距离,定义如下:

$$d^2(x_t, c_j) = \| x_t - c_j \|_F^2 = (x_t - c_j)^T F_j^{-1} (x_t - c_j) \tag{5}$$

其中,$\boldsymbol{F}_j$ 是第 j 个聚类的模糊协方差矩阵,定义如下:

$$\boldsymbol{F}_j = \frac{\sum_{t=1}^{T} u_{jt}(x_t - c_j)(x_t - c_j)^T}{\sum_{t=1}^{T} u_{jt}} \tag{6}$$

为了得到最后的模糊集可在条件式(3)下求式(4)的极小值,令 J_m 对 c_j 和 u_{jt} 的偏导数为 0,可得必要条件:

$$u_{jt} = \frac{\left[\frac{1}{d^2(x_t, c_j)}\right]^{\frac{1}{(m-1)}}}{\sum_{i=1}^{K}\left[\frac{1}{d^2(x_t, c_j)}\right]^{\frac{1}{(m-1)}}} \tag{7}$$

$$c_j = \frac{\sum_{t=1}^{T} (u_{jt})^m x_t}{\sum_{t=1}^{T} (u_{jt})^m} \tag{8}$$

用迭代法求解式(7)和式(8),就是模糊 k 均值算法.算法步骤如下:

Step1 设定聚类数目 K 和参数 m.

Step2 初始化各个聚类中心 c_j.

Step3 重复下面的计算,直到各个样本的隶属度值稳定.

用当前的聚类中心按式(7)计算隶属度函数.用当前的隶属度函数按式(8)更新计算各类聚类中心.

当算法收敛时,就得到了各类聚类中心和各个样本对于各类的隶属度值,从而完成模糊聚类划分.

3.2 初始化聚类中心及其改进

传统的聚类中心初始化方法有随机法和重心法,都需要任意选择聚类中心,没有用到特征矢量序列分布的先验信息,导致 GMM 模型精度欠佳.本文用分裂法和模糊 k 均值聚类相结合的方法初始化聚类中心.该方法契合了特征矢量的分布函数由多个高斯分布函数线性组合的原理,对样本聚类后得到的初始参数能通过 EM 算法较快收敛,并使样本分布能较好地拟合高斯分布.初始化算法如下:

Step1 把提取的每个说话人特征参数集作为训练样本集.形成一个 $T \times P$ 的矩阵(T 为帧数,P 为特征维数).

Step2 由 $\mu[j] = \frac{\sum_{i=1}^{T} X_{[i][j]}}{T}$ $(j = 1,2,\cdots,P)$ 得到一个 P 维的均值矢量.根据 $\mu[j]$ 计算方差,形成 P 维的方差矢量 $\boldsymbol{\eta}[k]$ $(k = 1,2,\cdots p)$,然后根据 $\mu[j] \pm 0.01\boldsymbol{\eta}[k]$ 分裂成 2 个聚类中心.

Step3 按最小距离准则计算每一帧(训练样本)与聚类中心的距离,把样本集分为 i 类(i 为当前聚类中心个数).

Step4 更新聚类中心,对属同一类的样本集进行均值矢量计算,把不同类的均值矢量作为新的聚类中心.

Step5 若 $\frac{|lastdist - mindist|}{lastdist} > 0.0001$($mindist$ 为更新聚类中心后样本与离它最近的聚类中心的距离和,$lastdist$ 为上一次聚类后样本与离它最近的聚类中心的距离和),则转到 Step3.

Step6 根据 Step1 和 Step2,用更新好的 2 个聚类中心分成 4 个聚类中心,然后按 3、4、5 的步骤把训练矢量聚成 4 类.依次类推,可分成 8 类、16 类等.

Step7 假设 GMM 的阶数为 M,则最后把训练样本

集分为 M 类,由每一类的均值矢量作为模糊 K 均值聚类的聚类中心.

3.3 PCA 降维

模糊 PCA 转换矩阵由计算模糊协方差矩阵(式(6))的特征值和特征向量获得.特征值从大到小排列,计算其对应的特征向量,即主成分.用前 k 个主成分的方差在全部方差中所占比重来描述累积贡献率.当累积贡献率大于 80% 时,确定主成分的个数(L).形成一个最优化特征矢量维数的 $L\times P$ 转化矩阵.

在训练和测试时,每一帧特征矢量转化为:

$$y_{t_j}=\boldsymbol{\Phi}_j x_t^T,\text{if}\quad x_t\in R^j \tag{9}$$

其中 $\boldsymbol{\Phi}_j=(\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_L)_j$ 是 $L\times P$ 加权矩阵,其行代表是第 j 聚类中的 L 个主特征向量,$\boldsymbol{\varphi}_i$ 是 $\boldsymbol{F}_j$ 中第 i 个最大特征值所对应的特征向量.取式(9)的协方差矩阵的对角阵形式作为 GMM 的初始化参数.

4 基于模糊 PCA 的 GMM

假设 $\boldsymbol{Y}=\{Y_1,\cdots,Y_K\}=\{y_1,y_2,\cdots y_T\}$($k$ 为聚类数,T 为语音总帧数)是所有原特征参数经模糊 PCA 处理后的特征矢量集,其中 $Y_j=\{y_{t_{j=1}},\cdots,y_{t_{j=T_j}}\}$ 表示属于第 j 聚类(R^j)的特征矢量集.为每个说话人建立一个 M 阶 GMM(一般使 $K=M$),其实质是通过训练,估计 GMM 的参数集 λ.它由各均值矢量、协方差矩阵及混合分量的权值组成,表示成如下三元组的形式:

$$\lambda=\{c_j,u_j,\Sigma_j\},j=1,2,\cdots,M \tag{10}$$

利用 $\boldsymbol{Y}_j(1\leqslant j\leqslant K(M))$计算初始参数集 λ.

这样,GMM 的似然函数可表示为:

$$\begin{aligned}p(Y|\lambda)&=\prod_{t=1}^{T}p(y_t|\lambda)\\&=\prod_{t_1=1}^{T_1}p(y_{t_1}|\lambda)\cdots\prod_{t_K=1}^{T_K}p(y_{t_K}|\lambda)\end{aligned} \tag{11}$$

其中,$p(y_t|\lambda)$是第 t 帧特征参数在模型 λ 下的概率密度,它由 M 个单高斯分布的线性组合来描述.形式如下:

$$\begin{aligned}p(y_t|\lambda)&=\sum_{j=1}^{M}p(y_t,j|\lambda)=\sum_{j=1}^{M}c_j p(y_t|j,\lambda)\\p(y_t|j,\lambda)&=\frac{1}{(2\pi)^{\frac{1}{2}}|\Sigma_j|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y_t-\mu_j)^T\Sigma_j^{-1}(y_t-\mu_j)\right\}\end{aligned} \tag{12}$$

式中,j 为隐状态号,也就是高斯分量的序号,M 阶 GMM 就有 M 个隐状态,c_j 为第 j 个分量的混合权值,其值对应为隐状态 j 的先验概率.满足 $\sum_{j=1}^{M}c_j=1$.$p(y_t|j,\lambda)$为高斯混合分量,当 Σ_j 取对角阵,即 $\Sigma_j=\text{diag}\{\sigma_{j0}^2,\sigma_{j1}^2,\cdots,\sigma_{jL-1}^2\}$($L$ 为特征维数).代入式(12),可得:

$$p(y_t|j,\lambda)=\prod_{k=0}^{L-1}\frac{1}{\sqrt{2\pi}\sigma_{jk}}\exp\left[-\frac{(y_t-\mu_{jk})^2}{2\sigma_{jk}^2}\right] \tag{13}$$

参数集 λ 可通过对式(11)的最大似然估计获得,但无法直接计算,我们采用 EM 算法来估计 GMM 的参数 λ[13].它是一种迭代算法,每次迭代由求期望(E-step)和求最大值(M-step)组成,估计出一个新的模型参数 $\bar{\lambda}$,使 $P(Y|\lambda)\leqslant P(Y|\bar{\lambda})$,然后再以 $\bar{\lambda}$ 作为模型的参数开始下一次的迭代,反复迭代,直到满足收敛条件.定义 Q 函数为:

$$\begin{aligned}Q(\lambda,\bar{\lambda})&=\sum_{t=1}^{T}Q_t(\lambda,\bar{\lambda})\\&=\sum_{t=1}^{T}\sum_{j=1}^{M}\frac{p(y_t,j|\lambda)}{p(y_t|\lambda)}\log p(y_t,j|\bar{\lambda})\\&=\sum_{j=1}^{M}\sum_{t=1}^{T}\frac{c_j p(y_t|j,\lambda)}{p(y_t|\lambda)}(\log\bar{c}_j+\log p(y_t|j,\bar{\lambda})\end{aligned} \tag{14}$$

欲估计 $\bar{c}_j$,令 $\frac{\partial Q(\lambda,\bar{\lambda})}{\partial\bar{c}_j}=0$,可求得:

$$\bar{c}_j=\frac{1}{T}\sum_{t=1}^{T}\frac{c_j p(y_t|j,\lambda)}{p(y_t|\lambda)} \tag{15}$$

E-step:求训练数据落在假定的隐状态 j 的概率 $p(q_t=j|y_t,\lambda)$表示为:

$$p(q_t=j|y_t,\lambda)=\frac{c_j p(y_t|j,\lambda)}{p(y_t|\lambda)} \tag{16}$$

M-step:找出式(14)中估计参数的最大值,即 $\lambda=\arg\max_{\lambda}Q(\bar{\lambda};\lambda)$.可分别求式(14)对于三个参数$\{c_j,\mu_j,\Sigma_j\}$偏导为 0 时的参数值:

混合权值:
$$\bar{c}_j=\frac{1}{T}\sum_{j=1}^{T}p(q_t=j|y_t,\lambda) \tag{17}$$

均值矢量:
$$\bar{\boldsymbol{\mu}}_j=\frac{\sum_{t=1}^{T}p(q_t=j|y_t,\lambda)y_t}{\sum_{t=1}^{T}p(q_t=j|y_t,\lambda)} \tag{18}$$

协方差矩阵:

$$\overline{\boldsymbol{\sigma}_{jk}^2}=\frac{\sum_{t=1}^{T}p(q_t=j|y_t,\lambda)(y_{tk}-\mu_{tk})^2}{\sum_{t=1}^{T}p(q_t=j|y_t,\lambda)},k=0,1,\cdots,f-1 \tag{19}$$

与传统 GMM 的说话人辨认相比,本文作了如下改进:(1)对原始特征矢量集进行模糊 PCA 处理,使特征维数远小于原始特征维数,不仅减小时空复杂度,还去除了各维特征之间的相关性;(2)用分裂法和模糊 k 均值聚类相结合的方法对特征矢量集进行聚类.分裂法的引入,使聚类中心的选择加入了特征先验信息,而无需随机选择聚类中心,提高了 GMM 的精度.

5 说话人辨认

对于有 N 个人的说话人辨认系统,其中每个说话

人可分别用 GMM($\lambda_1,\lambda_2,\cdots,\lambda_N$)来表示.识别时,找出使测试语音的特征矢量 $\boldsymbol{O}=\{o_1,o_2,\cdots,o_T\}$ 产生最大后验概率的 GMM.即:

$$n^* = \arg\max_{1\leqslant n\leqslant N} p(\lambda_n \mid \boldsymbol{O}) = \arg\max_{1\leqslant n\leqslant N} \frac{p(\boldsymbol{O}\mid\lambda_n)p(\lambda_n)}{p(\boldsymbol{O})} \quad (20)$$

其中,$p(\lambda_n)$为第 n 个人说话的先验概率,$p(\boldsymbol{O})$为所有说话人条件下特征矢量集 $\boldsymbol{O}$ 的概率,对于每个说话人,$p(\lambda_n)$和 $p(\boldsymbol{O})$都可视为相等.故式(20)可简化为:

$$n^* = \arg\max_{1\leqslant n\leqslant N} p(O \mid \lambda_n) = \arg\max_{1\leqslant n\leqslant N} \sum_{t=1}^{T} Inp(o_t \mid \lambda_n) \quad (21)$$

根据式(21)求判决结果,即计算测试语音的每一帧特征矢量$\{o_1,o_2,\cdots,o_T\}$在模型 λ_n 下的对数得分和,得分最高者所对应的 λ_n 作为最后识别结果.

6 实验结果与分析

6.1 实验语音库

实验语音数据取自 TIMIT 语音库.采样率为 16KHz,单声道录音,采用 16Bit 量化,共包括 630 个说话人(438 位男性,192 位女性).每个说话人分别朗读 10 个语句,为实验方便,本文标记为 sa1-sa10,每句长度大约 3 秒.

6.2 语音预处理

在特征提取之前,对训练和测试语音做预处理工作.预加重:预加重系数为 0.9375;分帧:帧长取 512 个采样点(32ms),帧移取 128 个采样点(8ms);加窗:加 hamming 窗;端点检测:用过零率来检测静音,用短时能量来检测浊音,两者配合采用双门限的方法实现可靠的端点检测.

6.3 短语音说话人辨认中特征参数的研究

6.3.1 单一特征参数的研究

实验条件:实验样本取自 TIMIT 库中 440 个说话人(316 位男性、124 位女性);训练语料 sa1,时长约为 3 秒;测试语料 sa10,时长约为 2 秒;特征参数有 LPCC、MFCC;模型采用 4 阶、8 阶、16 阶、32 阶 GMM.实验结果见表 1、表 2.

表 1 LPCC 维数和 GMM 阶数对识别率的影响

特征维数 \ 模型阶数	GMM(4 阶)	GMM(8 阶)	GMM(16 阶)	GMM(32 阶)
LPCC(5 阶)	35.682%	42.273%	33.227%	29.773%
LPCC(7 阶)	46.818%	52.955%	46.045%	36.272%
LPCC(9 阶)	57.727%	57.045%	52.955%	37.227%
LPCC(12 阶)	65%	62.955%	55.682%	40.682%
LPCC(14 阶)	66.818%	66.364%	57.045%	39.318%
LPCC(16 阶)	67.272%	70.909%	58.182%	42.818%
LPCC(18 阶)	70.227%	69.545%	59.773%	40.227%
LPCC(19 阶)	69.773%	67.727%	56.545%	39.318%
LPCC(25 阶)	66.818%	68.045%	53.864%	34.091%

由表 1、表 2 可以看出,当 GMM 取 8 阶时,识别率相对较高,这主要是由特征样本的数量决定的.短语音说话人辨认样本较少,采用过高的模型阶数来描述特征空间,会造成过拟合使识别率下降;采用过低的模型阶数,会使 GMM 不能充分表达特征空间.当特征参数维数增加,普遍会使识别率提高,不论 LPCC 还是 MFCC 当维数达到 16 附近以后,识别率的变化就不明显了,甚至下降,这是符合倒谱特征的性质的.

表 2 MFCC 维数和 GMM 阶数对识别率的影响

特征维数 \ 模型阶数	GMM(4 阶)	GMM(8 阶)	GMM(16 阶)	GMM(32 阶)
MFCC(5 阶)	38.181%	42.272%	37.045%	30.909%
MFCC(7 阶)	46.363%	58.864%	47.272%	38.863%
MFCC(9 阶)	58.227%	60.909%	53.864%	40%
MFCC(12 阶)	62.955%	65.455%	55%	43.863%
MFCC(14 阶)	65%	69.091%	58.181%	43.409%
MFCC(16 阶)	71.363%	72.955%	57.045%	42.272%
MFCC(18 阶)	70.227%	70.454%	56.136%	41.363%
MFCC(19 阶)	70%	70.227%	55.909%	40.682%
MFCC(25 阶)	67.955%	67.955%	52.954%	39.318%

6.3.2 特征组合的研究

一个自然的考虑是,如果把单一有效特征组合起来,提高特征的有效维数,应该能取得更好的识别性能.实验条件如 6.3.1,MFCC 和 LPCC 都取最佳维数 16,GMM 取最佳阶数 8,测试人数分别取 140、240、340、440.实验结果如表 3.

表 3 不同特征组合对识别率的影响

特征维数 \ 测试人数	140	240	340	440
MFCC + LPCC	90.227%	83.181%	77.955%	74.090%
LPCC + MFCC	89.090%	80.454%	77.045%	73.863%
MFCC_D_LPCC	92.954%	83.863%	79.318%	75.909%
MFCC_LPCC_D	91.428%	82.045%	78.636%	70%
MFCC_D_LPCC_D	92.5%	83.863%	78.181%	71.136%

表 3 数据表明:有效特征的组合,提高了有效特征维数,其中 MFCC_D_LPCC 组合最优,比单一特征 LPCC 和 MFCC 在同等条件下提高了 5% 和 2.954%.

6.4 模糊 PCA 降维实验

针对表 3 得出的最佳特征组合进行局部模糊 PCA 降维,我们采用累积贡献率达 80% 时所对应的主成分个数,通过计算得主成分个数 $L=16$.表 4 为最佳特征组合降维后与降维前的对比结果.

表 4 模糊 PCA 降维对识别率的影响

特征维数 \ 测试人数	140	240	340	440
MFCC_D_LPCC(48 维)	92.954%	83.863%	79.318%	75.909%
MFCC_D_LPCC(16 维)	92.272%	83.409%	78.636%	75.227%

从表 4 可以看出,模糊 PCA 降维确实发挥了作用.当由 48 维特征降低到 16 维,混合阶数为 8 时,识别率下降很小,以 440 个说话人样本为例,仅下降 0.682%.

却比单一特征 LPCC(16 维)和 MFCC(16 维)提高了 4.318% 和 2.272%.

6.5 不同 GMM 参数初始化方法对识别率的影响

实验条件:实验样本分别取 140、240、340、440 个说话人;训练语料 sa1;测试语料 sa10;特征为 16 维 MFCC;GMM 阶数为 8.表 5 为几种参数初始化方法的对比结果.

表 5　不同 GMM 参数初始化方法对识别率的影响

测试人数 / 参数初始化	140	240	340	440
随机法	82.045%	78.181%	70%	61.818%
K 均值聚类	89.090%	85%	78.181%	72.954%
模糊 k 均值聚类	90%	87.955%	80.454%	75.909%
分裂法与模糊 k 均值聚类相结合	93.181%	91.591%	83.863%	79.545%

由表 5 可以看出,分裂法与模糊 k 均值聚类相结合方法明显优于其它三种方法.前三种方法都是随机选择聚类中心,若选择不当,会直接影响最后 GMM 参数的生成.本文采用的方法,用特征样本的均值和方差作为先验信息来对聚类中心进行精确选择,从而精确聚类,使得最后训练得到的 GMM 能更好地刻画说话人特征空间.

7 结论

本文针对短语音说话人辨认因训练语料不充分而导致识别率下降的问题,采用了三大补偿技术:用特征组合代替单一有效特征,增加有效特征维数来提高识别性能;用局部模糊 PCA 的 GMM 方法来优化特征维数,达到去相关性、降维的目的;对传统初始化聚类中心算法进行改进,采用分裂法与模糊 k 均值聚类相结合算法.实验表明,采用本文的方法在 TIMIT 库上用 440 个人进行 3 秒左右的训练,最后识别率达到 80% 左右.与传统基于 GMM 的方法相比其识别率有 8% 左右的提高.

参考文献

[1] P Joseph, JR Campbell. Speaker recognition: A tutorial[J]. Proceedings of the IEEE, 1997, 85(9): 1437 – 1462.

[2] Tomi Kinnunen, Li Haizhou. An overview of text-independent speaker recognition: From feature to super vectors[J]. Speech Communication, 2009, 52(2): 12 – 40.

[3] 林琳,王树勋,陈建.基于可区分性加权的模糊核说话人识别[J].电子学报,2008,36(7):1446 – 1450.
LIN Lin, WANG Shu-xun, CHEN Jian. A fuzzy kernel with discriminative weighted method for speaker recognition[J]. Acta Electronica Sinica, 2008, 36(7): 1446 – 1450. (in Chinese)

[4] I T Jolliffe. Principal Component Analysis [M]. Springer: Berlin, 1986.

[5] 安冬,王守觉.基于仿生模式识别和 PCA/ICA 的 DOA 估计方法[J].电子学报,2004,32(9):1448 – 1451.
AN Dong, WANG shou-jue. A DOA estimation method based on Biomimetic pattern recognition and PCA/ICA [J]. Acta Electronica Sinica, 2004, 32(9): 1448 – 1451. (in Chinese)

[6] N Kambhatla. Dimension reduction by local PCA[J]. Neural Computing, 1997, 9(7): 1493 – 1516.

[7] C W Seo, K Y Lee. GMM based on local PCA for speaker identification[J]. Electronics Letters, 2001, 37(24): 1486 – 1488.

[8] S Molau, M Pitz, R Schluter. Computing Mel-frequency cepstral coefficients on the power spectrum [A]. Proceedings of the 2001 IEEE International Conference on Acoustics, Speech and Signal Processing[C]. USA: IEEE Press, 2001. 73 – 76.

[9] S Furui. Cepstral analysis technique for automatic speaker verification[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1981, 29(2): 254 – 271.

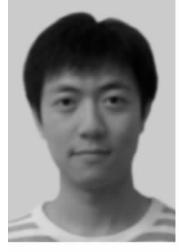
[10] J C Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithm[M]. New York: Plenum Press, 1981.

[11] E E Gustafson, W C Kessel. Fuzzy clustering with a fuzzy covariance matrix[A]. Proceedings of 18th IEEE Conference on Decision and Control[C]. USA: IEEE Press, 1979. 761 – 766.

[12] 武小红,周建江.可能性模糊 c-均值聚类新算法[J].电子学报,2008,36(10):1996 – 2000.
WU Xiao-hong, ZHOU Jian-jiang. A novel possibilistic fuzzy C-means clustering [J]. Acta Electronica Sinica, 2008, 36(10): 1996 – 2000. (in Chiese)

[13] D A Douglas. Robust text-independent speaker identification using gaussian mixture speaker models[J]. IEEE Transactions on Speech and Audio Processing, 1995, 3(1): 72 – 83.

作者简介



蒋　晔　男,1982 年 8 月出生于江苏江阴.2008 年毕业于南京理工大学计算机系,获工学硕士学位.现为南京理工大学博士研究生,从事模式识别、语音识别和说话人识别方面的有关研究.　E-mail: guyujiangrihua@126.com

唐振民　男,1961 年 4 月出生于陕西咸阳.教授、博士生导师.1982 年、1988 年和 2002 年分别在哈尔滨船舶工程学院、华东工学院和南京理大学获得工学学士、工学硕士和工学博士学位.现为南京理工大学计算机学院院长,主要从事语音识别、图像处理和智能机器人等方面的研究工作.