

文章编号: 1005—0523(2007)05—0077—04

一种改进的基于正交 GMM 的说话人辨识方法

李 蓉

(华东交通大学 信息工程学院,江西 南昌 330013)

摘要:在分析期望最大化(EM)算法的基础上,针对 EM 算法搜索的局限性把进化算法引入到正交高斯混合模型(OGMM)的参数训练中.实验结果表明以进化算法进行参数训练的正交 GMM 说话人辨认系统与用传统 EM 算法作为参数训练的说话人辨认系统相比有更好的识别效果,并具有良好的应用前景.

关 键 词:说话人辨认;正交 GMM;进化算法;EM 算法

中图分类号:TP391.42 **文献标识码:**A

1 引言

说话人辨认是利用说话人所说的话提取出来的语音信号特征,进行身份辨认的一种模式识别方法.它在金融、证券、公安、军队等需要身份安全认证的领域均具有较高的应用价值.说话人识别系统由特征提取和识别模型两部分组成.

基于高斯混合模型(Gaussian Mixture Models, GMM)的说话人识别方法是现代说话人识别技术的重要方法之一.这种方法的基本原理是对说话人集合中的每一个体建立一个概率模型,将说话人的个性特征在特征空间的分布抽象为该概率模型随机产生的结果.由于该概率模型采用的是多个单高斯分布的线性组合,因此被称为高斯混合模型^[1].而正交高斯混合模型^[3]针对提取出来的说话人语音特征矢量元素间存在相关性这一特点,先把特征矢量进行线性变换,得到元素间不相关的新特征,再进行 GMM 的训练与识别.相比传统的 GMM,具有较高的识别率和较少的计算机量和存储量,但仍存在对模型初值十分敏感、在实际训练中极易得到局部最优模型参数等问题.进化策略作为一种全局搜索算法,本文在分析期望最大化(EM)算法的基础上,针对 EM 算法搜索的局限性即搜索过程中会落入局部最优解的问题,把进化算法应用到正交高斯混合模型(GMM)的参数训练中,组成一种新的混合算法,对模型参数进行全局优化,对提高参数估计的准确性具有很好的作用.最后,进行了与文本无关的说话人辨认实验,详细比较了利用新的混合算法训练正交 GMM 与用 EM 算法训练的 GMM 得到的识别结果.

2 识别过程

2.1 GMM 的结构与概述^[1]

高斯混合模型(GMM)是用 M 个单高斯分布的加权和组合来描述语音特征在特征空间中的概率分布.每个说话人对应一个 GMM,模型如下:

$$p(x|\lambda)=\sum_{i=1}^M W_i g_i \tag{1}$$

其中, M 是 GMM 的混合元素数目,又叫阶数; x 是一个 D 维特征矢量; $W_i (i=1, 2, \dots, M)$ 为混合权值,即是高斯函数的个数,符合 $\sum_{i=1}^M W_i=1$; $g_i(x)$ 为 D 维高斯函数,即:

$$g_i(x)=\frac{1}{(2\pi)^{D/2}|\sum_i|^{1/2}}\exp\{- (x-\mu_i)^T \sum_i^{-1} (x-\mu_i)\} \tag{2}$$

其中, μ_i 为均值矢量也叫此密度函数的中心点; $\sum_i$ 为协方差矩阵.至此每个人可用各混合权值、均值矢量、协方差三类模型参数来描述,用 λ 表示这样一组参数集,则, $\lambda=\{W_i, \mu_i, \sum_i\}, (i=1, 2, \dots, M)$.

2.2 正交 GMM 的结构与概述^{[2][3]}

由矩阵理论可知,用原始协方差矩阵的本征向量组成的变换矩阵可将其变换成对角矩阵,语音特征矢量进行相同的变换可减少语音特征矢量之间的相关性,且可以减少 GMM 的计算量和存储量.正交高斯混合模型即是由一个线性变换矩阵和一个一般的对角线 GMM 模型构成,该线性变换矩阵是与说话人相关的,即每个模型都有各自的线性变换矩阵.

在求出线性变换矩阵后,下一步是将所有的该模型的全部训练矢量与该线性变换矩阵相乘作线性变换。

假设当前的说话人的特征所在的空间设为 x 空间,说话人的协方差矩阵记为 Σ_x , Ω 是由 Σ_x 的本征向量组成的变换矩阵,则经过线性变换 $y = \Omega^T x$ 后,将公式(2)中的 x 用 y 来代替,得:

$$g_i(y) = \frac{1}{(2\pi)^{D/2} |\Sigma_{y_i}|^{1/2}} \exp \left\{ - (y - \mu_{y_i})^T \Sigma_{y_i}^{-1} (y - \mu_{y_i}) \right\} \quad (3)$$

其中, Σ_{y_i} 为 y 空间的 GMM 的第 i 个高斯函数的协方差矩阵,且 $\Sigma_{y_i} = \Omega^T \Sigma_{x_i} \Omega$; μ_{y_i} 为 y 空间的 GMM 的第 i 个高斯函数的均值矢量,且 $\mu_{y_i} = \Omega^T \mu_{x_i}$ 。

2.3 EM 算法:

定义辅助函数 Q : $Q(\lambda, \lambda) = \sum_{i=1}^M p(X, i | \lambda) \log(p(X, i | \lambda))$ (4)

其中: i 为第 i 个高斯分布分量序号, X 为训练语音特征列向量序列。对上式求偏导,可得到迭代公式,由两步组成:

第一步,期望过程:

$$p(i | x_j, \lambda) = \frac{W_i g_i(x_j)}{\sum_{k=1}^M W_k g_k(x_j)} \quad i=1, 2, \dots, M; j=1, 2, \dots, T \quad (5)$$

其中: $p(i | x_j, \lambda)$ 为某一说话人的模型下特征 x_j 落在第 i 个高斯分布的概率。

第二步:最大化过程

求混合权值:

$$W_i = \frac{1}{T} \sum_{j=1}^T p(i | x_j, \lambda) \quad i=1, 2, \dots, M \quad (6)$$

求均值矢量:

$$\bar{\mu}_i = \frac{\sum_{j=1}^T p(i | x_j, \lambda) x_j}{\sum_{j=1}^T p(i | x_j, \lambda)} \quad i=1, 2, \dots, M \quad (7)$$

求方差矩阵:

$$\Sigma_i = \frac{\sum_{j=1}^T p(i | x_j, \lambda) (x_j - \bar{\mu}_i)(x_j - \bar{\mu}_i)^T}{\sum_{j=1}^T p(i | x_j, \lambda)} \quad i=1, 2, \dots, M \quad (8)$$

先通过设定 $\lambda = \{W_i, \mu_i, \Sigma_i\}$ 的初始值,并使 W_i 满足 $\sum_{i=1}^M W_i = 1$,通过期望过程和最大化过程反复迭代,算法的终止可以设定迭代的步数或计算两次迭代的参数集合 λ 和 λ' 的距离是否小于给定的一个值来判断。

2.4 识别过程

设有 S 个说话人,对应每一个人的 GMM 分别为 $\lambda_i (i=1, 2, \dots, S)$,输入某一个人的一句话时,提取出特征列向量序列 $X = \{x_j, j=1, 2, \dots, T\}$,其中的 x_j 相互独立,通过线性变换 $Y = \Omega^T X$ 得到 $Y = \{y_j, j=1, 2, \dots, T\}$,说话人识别的目的是要找到具有最大后验概率的模型参数 λ_m , m 即为对应的说话人。

$$m = \arg \max_{1 \leq i \leq S} \sum_{j=1}^T \log p(y_j | \lambda_i) \quad (9)$$

3 基于进化计算的 EM 算法

GMM 的传统训练方法 EM 算法是基于梯度上升的算法,它可以保证每一次迭代后模型对训练数据的似然概率是递增的,但作为一个基于梯度的爬山算法容易陷入一个局部最优解,因此对于高维的语音特征,很难保证在高维的特征空间里搜索到参数的最优位置。本文引入进化策略,再结合 EM 算法,组成混合算法来训练 GMM 的参数,步骤如下:

(1) 建立种群

进化算法中每个个体对应解空间的一个解。每个人对应 GMM 的一组参数集的解为一个个体,用实数编码。由于变异用的是正态性变异,每个个体都需要一组同结构的摄动因子来控制其变异,同时,这组摄动因子也在变异。一个个体用 $\lambda = \{W_i, \mu_i, \Sigma_i | i=1, 2, \dots, M\}$ 表示,对于 GMM, Σ_i 用满矩阵,正交 GMM 的 Σ_i 取对角阵。

(2) 适应函数的选取

我们要解决的问题是寻找 GMM 的最优参数,使得训练数据对模型的对数似然概率最大。所以个体的目标函数值用训练数据对个体模型的对数似然概率表示。

(3) 选择策略

本文采用了非线性排名的选择策略^[4]来缓和算法的过早收敛。

(4) 变异

采用正态变异,即高斯变异。每次变异的计算如下:

$$\sigma'_i = \sigma_i \cdot \exp(N_i(0, \Delta\sigma)) \quad (10)$$

$$\nu'_i = \nu_i \cdot N_i(0, \Delta\sigma'_i), i=1, 2, \dots, C \quad (11)$$

其中, σ'_i 为 σ_i 的变异值; ν'_i 为 ν_i 的变异值。

(5) 取代

对于进化策略,则分 $(\mu + \lambda)$ 策略和 (μ, λ) 策略。这里用 $(\mu + \lambda)$ 策略。在 μ 父体和 λ 个体组成的集合中选择 μ 个最后的个体,组成新一代种群。

(6) 加入传统的 EM 算法

在进入下一次进化前,在新一代种群中随机选一个个体进行传统的 EM 算法操作,使得算法的局部搜索能力得到加强。

(7) 算法终止

采用预置最大进化代数,在进化代数达到这一给定最大值时,终止迭代,保留最优个体和函数值。

4 仿真实验

整个实验在 Windows XP 环境下用 Matlab6.0 仿真完成,语音库是在实验室环境下录制的语音数据,人数为 30 个人,每人 15 句话。语音特征用 12 阶 MFCC。为了减少计算量和增加识别难度,我们先用 GMM 进行 30 个人的识别实验,再选

取识别率最低的 12 个说话人的语音数据做以下三个实验.

斯混合模型(GMM)的阶数为 2—8;每个人的训练语音为 2 句话,测试语音为 10 句;每一阶数的识别做 7 次. 外部识别率如表 1 所示.

4.1 实验一

用进化的 EM 混合算法代替 GMM 中的传统 EM 算法. 高

表 1 实验一的外部识别率

阶 数		2	3	4	5	6	7	8
次 数	No.1	69.17	70	74.17	74.17	70.83	68.33	67.50
	No.2	74.17	65.83	68.33	67.50	70.83	80.00	73.33
	No.3	71.67	70.83	80.00	78.33	73.33	70.83	75.00
	No.4	72.50	68.33	72.50	74.17	74.17	72.50	75.00
	No.5	67.50	74.17	70.83	75.00	72.50	74.17	73.33
	No.6	65.83	67.50	78.33	72.50	62.50	70.83	68.33
	No.7	67.5	71.67	75.83	76.67	71.67	75.83	75.83
	average	69.7629	69.7614	74.2843	74.0486	70.8329	73.2129	72.6171
EM 算法		47.50	55.83	66.67	73.33	73.33	77.50	70.83

从进化 EM 混合算法的 GMM 的外部识别率在训练数据不足的情况下,同阶数的每次实验的外部识别率都有较大波动,平均的波动达到 9.64%,但除了 6 阶和 7 阶外,它的外部平均识别率都大于传统的 EM 算法. 说明较少的数据限制了进化算法的搜索能力.

4.2 实验二

增加训练语音数据,每个人的训练语音增加到 4 句话,测试语音为 10 句;高斯混合模型(GMM)的阶数为 2—10;其它条件同实验 1,结果如表 2 所示. 同时,在 GMM 阶数为 6 时,我们将识别率为 92.5%对应的 12 个人的对数似然概率(即进化 EM 混合算法中的目标函数值)和 EM 算法识别率为 91.67%的对数似然概率进行比较,如表 3 所示.

表 2 实验二的外部识别率

阶 数		2	3	4	5	6	7	8	9	10
次 数	No.1	74.17	84.17	87.5	92.5	94.17	92.5	94.17	94.17	91.67
	No.2	75.00	80.00	87.5	90.83	92.5	93.33	94.17	95.83	93.33
	No.3	73.33	85.00	86.67	91.67	92.5	93.33	94.17	93.33	91.67
	No.4	74.17	82.50	87.5	90.00	94.17	92.5	94.17	93.33	93.33
	No.5	76.67	82.50	88.33	90.00	92.5	95.00	94.17	95.00	95.00
	No.6	73.33	80.00	88.33	89.17	92.5	92.50	94.17	95.00	94.17
	No.7	74.17	86.67	90.00	93.33	92.5	93.33	92.5	92.5	92.5
	average	74.406	82.977	87.976	91.071	92.977	93.213	93.93	94.166	93.096
EM 算法		58.33	67.5	81.67	88.33	91.67	91.67	90.00	90.83	89.17

表 3 进化 EM 混合算法与 EM 算法的目标函数值比较

说话人	GMM	混合算法	说话人	GMM	混合算法
No.1	—13342	—12925	No.7	—11984	—11060
No.2	—14537	—13994	No.8	—11788	—11101
No.3	—11627	—11192	No.9	—15723	—15002
No.4	—13187	—12595	No.10	—13917	—13111
No.5	—15115	—14657	No.11	—15929	—15015
No.6	—11406	—10748	No.12	—11466	—10635

进化 EM 混合算法的 GMM 的外部识别率在足够的训练

中国知网 <https://www.cnki.net>

数据的情况下,同阶数的每次实验的外部识别率的波动(见表 2、表 3)已减少到 3.52%,稳定性得到很大的提高. 而且它的每一阶的外部识别率的平均值和最小值都大于传统的 EM 算法. 进化算法的搜索能力得到充分发挥. 混合算法的搜索精度超过了传统的 EM 算法.

4.3 实验三

训练语音数据同实验二,用正交 GMM,阶数为 2—10,用进化的 EM 混合算法代替正交 GMM 中传统的 EM 算法,结果见表 4. 同时,在 GMM 阶数为 9 时,我们将识别率为 95.83%对应的 12 个人的对数似然概率(即进化 EM 混合算法中的目标函数值)与 EM 算法识别率为 90.0%的对数似然概率进行比较,如表 5 所示.

表 4 进化 EM 混合算法的 GMM 的外部识别率

阶数 \ 次数	2	3	4	5	6	7	8	9	10
No.1	90.00	91.67	92.50	94.17	95.00	95.00	95.83	95.00	94.17
No.2	88.33	90.83	93.33	94.17	95.00	95.83	95.83	95.00	93.33
No.3	90.00	91.67	93.33	92.50	94.17	93.33	95.00	94.17	93.33
No.4	90.00	91.67	93.33	94.17	95.00	94.17	95.83	94.17	93.33
No.5	88.33	91.67	92.50	92.5	94.17	95.00	95.83	95.00	92.5
No.6	90.00	90.00	93.33	93.33	94.17	94.17	95.83	94.17	93.33
No.7	90.00	90.83	93.33	94.17	94.17	95.83	95.00	95.83	94.17
average	89.529	91.191	93.093	93.573	94.526	94.761	95.593	94.763	93.451
EM 算法	83.33	87.5	91.67	92.50	94.17	93.33	91.67	90.00	90.00

表 5 进化 EM 混合算法与 EM 算法的目标函数值比较

说话人	正交 GMM 混合算法		说话人	正交 GMM 混合算法	
No.1	-13342	-12759	No.7	-11580	-10719
No.2	-14537	-13707	No.8	-11498	-10761
No.3	-11627	-10945	No.9	-15507	-14789
No.4	-13187	-12165	No.10	-13766	-12644
No.5	-15115	-14382	No.11	-15711	-14804
No.6	-11406	-10622	No.12	-11146	-10282

基于进化混合算法的正交 GMM 同样在每一阶数的识别率都要高于基于 EM 算法的正交 GMM(表 4,5). 而且在 7 阶和 8 阶时达到了本文所有实验的最高外部识别率 95.83%; 在 8 阶时达到了实验的最高平均识别率 95.593%. 同阶数的每次实验的外部识别率的波动为 1.482%, 也是最小的.

5 结论

本文在正交 GMM 的训练中引入进化算法, 对用于 GMM 训练中的 EM 算法进行了改进. 改进后的混合算法在语音训练数据充足的情况下, 平均识别率达到了 95.593%, 比起用 EM 算法训练的正交 GMM 有所提高. 但是本文用的语音数据

都是在实验室环境下完成的, 在实际使用中算法的鲁棒性有待进一步研究. 本文只是对两种方法的正确性和可行性进行了仿真实验, 从实用性考虑, 可用 C++ 对本算法进行重写, 进一步提高系统实时性和实用性, 此外也可以采用新的信号处理方法来提取更能表征说话人的新特征.

参考文献:

[1] Reynolds D, Rose R.C. Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Trans on speech and Audio Processing[J]. 1995

[2] Kuo-Hwei You, Hsiao-Chuan Wang. Joint estimation of feature transformation parameters and Gaussian mixture model for speaker identification [J]. Speech Communication, 1999, (28), 227-241.

[3] 候风雷, 等. 基于正交高斯混合模型的说话人识别研究[J]. 信息工程大学学报, 2002, 13, (2): 45-47;

[4] Andrew Chipperfield. Genetic Algorithm Toolbox User's Guide [N]. 1994.

[5] 岳喜才. 说话人识别使用遗传 RBF 网络[J]. 应用声学学报, 1999, 35-38.

[6] 潘正君, 康立山, 陈毓屏. 演化计算[M]. 北京: 清华大学出版社, 1998.

An Improved Speaker Recognition Method Based on Orthogonal GMM

LI Rong

(School of Information Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: This paper applies the evolutionary algorithm in the parameters training of orthogonal GMM based on analyzing the searching limitation of EM algorithm. The experiment result shows that the recognition rate based on the evolutionary algorithm used in orthogonal GMM is higher than EM algorithm.

Key words: speaker recognition; orthogonal GMM; evolutionary algorithm; EM algorithm