

混合粒子对优化算法在说话人识别中的应用

薛丽萍^{①②} 尹俊勋^① 周家锐^② 纪震^②

^①(华南理工大学电子与信息学院 广州 510640)

^②(深圳大学信息工程学院 深圳 518060)

摘要: 在粒子群优化(Particle Swarm Optimization, PSO)和混合蛙跳算法(Shuffled Frog-Leaping Algorithm, SFLA)的基础上, 该文提出了一种新的混合粒子对优化(Shuffled Particle-Pair Optimizer, SPPO)算法, 应用于矢量量化的说话人识别。该算法将全局信息交换和局部深度搜索相结合寻求最佳的说话人码本。群体按适应值分为 3 个粒子对, 每个粒子对由两个粒子构成, 按先后顺序执行 PSO 算法中的速度位置更新和 LBG 算法以实现局部细致搜索, 间隔一定的迭代次数通过 SFLA 混合策略实现粒子对间的信息交换, 从而使群体向全局最优解靠近。实验结果表明, 本算法始终稳定地取得显著优于 LBG, FCM, FRLVQ-FVQ 和 PSO 算法的说话人识别性能, 较好地解决了初始码本影响的识别性能的问题, 且在计算时间和收敛速度方面有相当的优势。

关键词: 说话人识别; 粒子群优化; 混合蛙跳算法; 矢量量化; 与文本无关

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2009)06-1359-04

A Novel Shuffled Particle-pair Optimizer for Speaker Recognition

Xue Li-ping^{①②} Yin Jun-xun^① Zhou Jia-rui^② Ji Zhen^②

^①(School of Electronics and Information Engineering, South China University of Technology, Guangzhou 510640, China)

^②(Faculty of Information Engineering, Shenzhen University, Shenzhen 518060, China)

Abstract: A novel Shuffled Particle-Pair Optimizer (SPPO) is proposed for speaker recognition based on vector quantization, which combines the advantage both in Particle Swarm Optimization (PSO) and Shuffled Frog-Leaping Algorithm (SFLA). The SPPO contains elements of local exploration and global information exchange to get global optimized speaker codebook. In this algorithm, the population is partitioned into 3 particle-pairs according to the performance, and each particle-pair consists of two particles. The particle-pairs perform simultaneously local exploration using basic operations of PSO (velocity updating and position updating) and LBG algorithm in sequence. A shuffling strategy, in which the particles are periodically shuffled and reorganized into new particle-pairs, allows for the exchange of information between particle-pairs to move toward the global optimum. Experimental results demonstrate that the performance of this new method is much better than that of LBG, FCM, FRLVQ-FVQ, and PSO consistently with lower speaker recognition error rates, shorter computational time and higher convergence rate. The dependence of the final codebook on the selection of the initial codebook is also reduced effectively.

Key words: Speaker recognition; Particle Swarm Optimization(PSO); Shuffled Frog-Leaping Algorithm(SFLA); Vector quantization; Text-independent

1 引言

短语音(训练语音时间为 10s 左右)的说话人识别由于其良好的应用前景受到广泛关注, 矢量量化的说话人识别是其较好的选择。矢量量化的关键是码本的优化设计, 文献[1]提出了 LBG 算法, 其算法理论严密, 易于实现, 但它对初始码本的选择非常敏感。文献[2]采用模糊 C 均值(Fuzzy C-Means, FCM), 识别性能有所提高, 但计算模糊隶属度

十分耗时。文献[3]提出了利用加强学习作为预处理的模糊矢量量化(FRLVQ-FVQ), 有效地改善了 FCM 的性能。智能优化算法推动了全局最优码本的研究, 如文献[4]利用蚁群算法改善识别性能, 但大部分优化算法对参数的设置敏感, 计算量大, 求优过程耗时, 收敛速度慢。

粒子群优化(Particle Swarm Optimization, PSO)算法^[5]在搜索复杂变量、高维空间的最优解方面具有很好的适应性和敏捷性。收敛速度快、设置参数少。文献[6-8]成功地将 PSO 应用于语音和图像处理。不过 PSO 也面临着与其他进化算法类似的问题: 早熟收敛和算法后期搜索效率降低。混

2008-06-02 收到, 2008-10-27 改回

国家自然科学基金(60572100)和深圳大学科研启动基金(200637)资助课题

合蛙跳算法(Shuffled Frog-Leaping Algorithm, SFLA)是一种基于群体的协同搜索算法^[9],其局部搜索和混合策略为与其他优化算法融合提供了思路,文献[10]将 SFLA 与细菌优化结合解决装配线调度问题。本文针对 PSO 算法的问题,结合 SFLA 算法的操作策略,提出了一种新的混合粒子对优化(Shuffled Particle-Pair Optimizer, SPPO)算法。此算法利用 3 个粒子对,按 SFLA 的局部深度搜索和全局信息交换的混合策略,粒子在每次迭代中执行 PSO 速度位置更新和标准 LBG 算法的混合更新操作,在训练矢量空间搜索最优码本。本算法参数设置简单,既能提高群体的多样性,又不增加算法的复杂性,寻优能力较强。与 LBG, FCM, FRLVQ-FVQ, PSO 相比, SPPO 算法提高了码本设计质量,在与文本无关的说话人识别实验中表现出了优异的性能。

2 矢量量化的说话人识别

令说话人训练语音的 L 维特征矢量集为 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$, $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{iL}\}$, \mathbf{Y} 是由 M 个 L 维的码字组成的说话人码本模型,即 $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}$ 。矢量量化就是把 T 个训练矢量分配到 M 簇中,每一簇由一个码字代表。码本的设计质量通常是由训练矢量与对应的最近码字之间的均方误差(MSE)来表示,简称为 \tilde{D} 。

$$\tilde{D} = \frac{1}{T} \sum_{i=1}^T [d_{\min}(\mathbf{x}_i)]^2 \quad (1)$$

其中 $d_{\min}(\mathbf{x}_i) = \min_{\mathbf{y}_j \in \mathbf{Y}} d(\mathbf{x}_i, \mathbf{y}_j)$, $d(\mathbf{x}_i, \mathbf{y}_j) = \|\mathbf{x}_i - \mathbf{y}_j\|$ 为

Euclidean 距离。

在识别阶段,用系统中建立的 N 个说话人码本依次对测试语音特征矢量序列进行矢量量化,计算各自的量化误差,选择平均量化误差最小的码本所对应的说话人作为识别的结果。

3 粒子群优化算法

在一个 D 维的目标搜索空间中,随机生成 P 个粒子,第 i 个粒子的位置可表示为 $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iD})$,速度为 $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{iD})$,根据适应度函数计算 \mathbf{z}_i 当前的适应值,来衡量粒子位置的优劣。粒子 i 迄今为止搜索到的最优位置为 $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{iD})$,整个粒子群迄今为止搜索到的最优位置为 $\mathbf{p}_g = (p_{g1}, p_{g2}, \dots, p_{gD})$ 。每次迭代中粒子 i 的粒子根据以下公式更新速度和位置:

$$v_{id}^{k+1} = wv_{id}^k + c_1r_1(p_{id} - z_{id}^k) + c_2r_2(p_{gd} - z_{id}^k) \quad (2)$$

$$z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1} \quad (3)$$

其中 $d = 1, 2, \dots, D$; k 是迭代次数; r_1 和 r_2 为均匀分布在 $[0, 1]$ 之间的随机数; w 为惯性权重; c_1 , c_2 为学习因子。粒子在目标搜索空间中不断跟踪 \mathbf{p}_i 和 \mathbf{p}_g 进行搜索,直到达到预定

的迭代次数为止。

4 混合蛙跳算法

混合蛙跳算法(SFLA)模拟青蛙群体的觅食特性。在一个 D 维的目标搜索空间中,随机生成 P 只青蛙(问题的解)组成初始群体,第 i 只青蛙表示问题的解为 $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ 。青蛙个体按适应度值从优到劣排列,将整个群体分成 m 个子群体。其中,排名第 1 的青蛙分入第 1 子群,第 2 分入第 2 子群,第 m 的青蛙分入第 m 子群,第 $m+1$ 又分入第 1 子群,依次类推,直到全部青蛙划分完毕。

每个子群体进行局部深度搜索,即每次迭代中,该子群中的当前最差个体 X_w 追随该子群的当前最好个体 X_b 或整个群体中的最好个体 X_g 进行更新操作,若 X_w 的适应度值无改进,则随机产生一个新的 X_w 。

当所有子群都完成一定迭代次数局部搜索后,混合策略将所有个体混合并排序和重新划分子群体,然后再进行局部深度搜索,如此反复直到满足终止条件。

5 混合粒子对优化算法

为提高 PSO 的优化性能,本文引入 SFLA 算法的混合和局部搜索策略,提出一种新的混合粒子对优化(SPPO)算法。它将群体按 SFLA 算法划分规则分为 3 个粒子对,每个粒子对由两个粒子构成,3 个粒子对分别进行局部搜索,在每次迭代中执行 PSO 的速度更新和位置更新以及迭代次数为 3 的 LBG 算法的混合更新操作。当 3 个粒子对每完成一定迭代次数的局部搜索后,混合所有粒子并重组粒子对,完成远距离的信息快速传递,实现全局信息交换。如此反复,在训练矢量空间中搜索最优码字。

SPPO 利用 3 个粒子对增加搜索过程的多样性,克服早熟,每个粒子对又能集中搜索空间中有希望的区域,提高求解精度。无需局部版 PSO 复杂的领域更新操作,计算代价低。SPPO 采用 PSO 和 LBG 混合更新策略,使粒子能够尽快移动到最优点上,加快算法的收敛速度,取得计算代价与求解质量之间的较好平衡。

5.1 粒子结构和适应度函数的选取

在 SPPO 算法中,粒子结构的设计是基于码字的。每个粒子代表着一个码本。把说话人训练特征矢量集聚类成 M 簇,每个码字 \mathbf{y}_j 代表一簇,粒子结构如表 1 所示。

粒子的维数为 D 维,其中 $D = M \times L$ 。粒子位置的取值范围一般分别取语音特征矢量集中每一维的最小值和最大值。

SPPO 选取均方误差(MSE)作为适应度函数,利用式(1)计算。

表 1 粒子结构

$y_{11}, y_{12}, \dots, y_{1L}$	$y_{21}, y_{22}, \dots, y_{2L}$...	$y_{j1}, y_{j2}, \dots, y_{jL}$...	$y_{M1}, y_{M2}, \dots, y_{ML}$
---------------------------------	---------------------------------	-----	---------------------------------	-----	---------------------------------

5.2 SPPO 的码本优化设计

一般 PSO 和 SFLA 的群体规模较大, 群体规模大会提高求解精度, 但影响算法的运算速度和收敛性。基于码本的优化方案, 利用粒子对形成协作工作关系, 在优化效率上有优势^[8], 因此 SPPO 群体规模仅为 6 个粒子, 每个子群体只使用 1 个粒子对。

SPPO 算法码本优化设计的主要步骤如下:

(1) 随机初始化群体, 从训练特征矢量集随机地选取 M 个矢量作为每个粒子的初始码字。设置最大迭代次数 K_{max} , 粒子对更新次数 G_{max} , 迭代次数计数器 $k=1$ 。

(2) 将群体中所有粒子按适应值从优到劣排列, 按 SFLA 划分方法分为 3 个粒子对。

(3) 对每个粒子对:

(a) 按适应值更新确定 p_i 和粒子对的 p_j ;

(b) 粒子按式(2), 式(3)更新速度和位置;

(c) 粒子执行迭代次数为 3 的 LBG 操作, 并处理处理空腔, 用具有较大误差的训练矢量来代替具有越界问题的码字;

粒子对按(a), (b), (c)步执行 G_{max} 次。

(4) 对更新后的所有粒子对进行混合, 取代原群体;

(5) $k = k + 1$, 当 $k > K_{max}$ 时, 转向步骤(6), 否则转向步骤(2)。

(6) 结束迭代, 找出群体中的最优粒子, 作为所训练的说话人的码本模型。

6 实验结果

实验仿真平台为 Windows XP, Matlab 7.0, CPU 为 P4 3.0GHz, 内存为 1GB。实验对比测试 LBG, FCM, FRLVQ-FVQ, PSO 和 SPPO 的说话人训练和识别效果。

实验语音数据取自 TIMIT 语音数据库^[11], 随机抽取 100 个说话人作为辨认的集合, 30 个女性, 70 个男性。语音信号经过预加重系数为 0.95 的滤波, 采用汉明窗进行分帧和加窗, 提取 15 维 MFCC 作为语音特征矢量。每个人的训练和识别语句分别为 5 句, 每句的有效语音时间平均约 2.5s。每次识别实验共进行 500 次与文本无关的说话人识别测试。

PSO 算法的群体规模为 10 个粒子, 采用与 SPPO 同样的混合更新操作。PSO 和 SPPO 的中止条件设为最大迭代次数。FRLVQ-FVQ 算法的参数设置见文献[3]。LBG, FCM 和 FRLVQ-FVQ 算法的中止条件为 $(D^{(k-1)} - D^{(k)}) / D^{(k)} \leq \epsilon$ 。参数设置如表 2, 其中 m 为模糊度。

码本尺寸为 $M=8$, 分别应用各算法进行 10 次训练和识别实验并比较误识率。各算法的初始码本均采用随机生成。表 3 给出了 10 次实验所得的平均误识率。

各算法 10 次训练和识别实验的误识率的变化范围如图 1 所示。LBG, FCM, FRLVQ-FVQ, PSO 和 SPPO 算法的误识率的变化范围分别为 1.4%, 2.2%, 0.4%, 1.8% 和 0.4%。

表 2 LBG, FCM, PSO 和 SPPO 算法的参数设置

算法	ϵ	m	K_{max}	G_{max}	w	c_1	c_2
LBG	0.001	—	—	—	—	—	—
FCM	0.001	1.1	—	—	—	—	—
PSO	—	—	35	—	0.1	0.3	0.5
SPPO	—	—	7	5	0.1	0.3	0.5

表 3 说话人误识率比较

误识率	算法				
	LBG	FCM	FRLVQ-FVQ	PSO	SPPO
ERR(%)	7.30	6.24	7.10	5.34	4.28

FRLVQ-FVQ 和 SPPO 受初始码本的影响最小。PSO 最好的结果与 SPPO 一样, 但误识率的变化范围大, 存在早熟收敛。SPPO 表现稳定, 其识别性能受初始码本的影响小。

在图 2 中 SPPO 的 MSE 值取自于每次迭代中 3 个粒子对的最优值。从图 2 整个收敛过程看, LBG, FCM, FRLVQ-FVQ 和 PSO 算法均在不同程度上收敛到了局部最优值。SPPO 的搜索策略保证了其收敛速度快, 且能跳出局部最优。

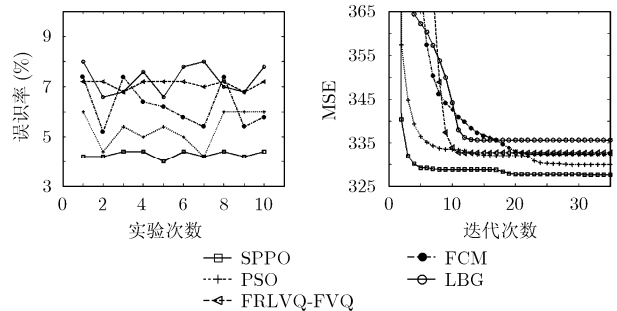


图 1 各算法的 10 次实验结果比较

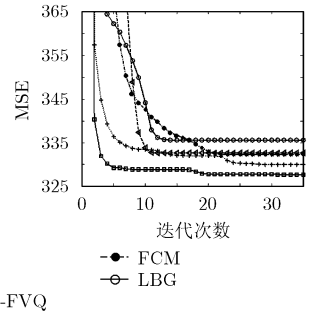


图 2 各算法的收敛特性比较

在码本尺寸分别为 $M=8, 16, 32, 64$ 的条件下, 表 4 给出了 100 个说话人训练的平均失真测度和每个人的平均计算时间 $T(s)$ 。相对于 LBG、FCM、FRLVQ-FVQ 和 PSO 算法, SPPO 在不同码本条件下均取得了最小的均方误差, 有效地改善了码本设计质量。在计算时间方面, SPPO 比 FRLVQ-FVQ 和 PSO 快, 随着码本尺寸的增加计算时间优势明显。

图 3 给出了不同码本各算法平均误识率。SPPO 均能得到最好的识别性能, 误识率最低。

7 结束语

本文提出了一种新的混合粒子对优化算法(SPPO), 3 个粒子对分别通过在迭代过程中交替执行 PSO 和 LBG 算法实现局部深度探索, 通过粒子对混合和重组实现粒子间信息快速交换, 从而较好地平衡全局优化和局部探索, 使得算法能

表4 平均失真测度和计算时间比较

算法	M=8		M=16		M=32		M=64	
	\bar{D}	T (s)	\bar{D}	T (s)	\bar{D}	T (s)	\bar{D}	T (s)
LBG	247.6	0.6	192.3	1	149.9	2.3	114.7	1.5
FCM	246.3	7	190.7	15.2	147.2	51.9	111.2	162.9
FRLVQ-FVQ	247.0	23.8	191.7	29.3	148.5	95.4	112.7	215.4
PSO	243.2	33.6	188.4	49.2	145.9	79.8	109.8	149.8
SPPO	242.0	23.4	186.9	28.8	144.6	63.6	109.1	124.2

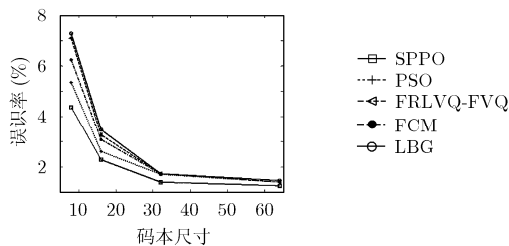


图3 码本尺寸与平均误识率曲线图

够跳出局部最优。实验结果表明新算法在与文本无关的说话人识别性能方面显著优于 LBG、FCM、FRLVQ-FVQ 和 PSO 算法,在码本大小分别为 8, 16, 32, 64 时,能够稳定地取得最好的码本设计质量和识别效果。新算法寻优能力较强,有效地解决了说话人初始码本的选取影响识别性能的问题,且收敛速度快,在计算时间方面优于 FRLVQ-FVQ 和 PSO 算法,参数设置简单,便于实际应用。

参考文献

- [1] Soong F K, Rosenberg A E, and Rabiner L R, *et al.* A vector quantization approach to speaker recognition[C]. International Conference on Acoustics, Speech, and Signal Processing, Tampa, 1985: 387-390.
 - [2] Tran D, Wagner M, and Van Le T. A proposed decision rule for speaker recognition based on fuzzy C-Means clustering [C]. 5th International Conference on Spoken Language Processing, Sydney Australia, 1998: 755-758.
 - [3] Xu Wen-huan, Nandi A K, and Zhang Ji-hong. Novel vector quantizer design using reinforced learning as a pre-process [J]. *Signal Processing*, 2005, 85(7): 1315-1333.
 - [4] 胡恒滔, 龙建忠. 基于蚁群算法的模糊C-均值聚类算法在声纹识别中的应用[J]. *四川大学学报(自然科学版)*, 2007, 44(3): 543-547.
Hu Heng-tao and Long Jian-zhong. Speaker identification using fuzzy C-means clustering algorithm based on improved ant colony algorithm. *Journal of Sichuan University (Natural Science Edition)*, 2007, 44(3): 543-547.
 - [5] Kennedy J and Eberhart R. Particle swarm optimization [C]. Proceedings of IEEE International Conference on Neural Networks, Piscataway, 1995: 1942-1948.
 - [6] Xue Li-ping, Yin Jun-xun, and Ji Zhen, *et al.* A particle swarm optimization for hidden Markov model training [C]. 8th International Conference on Signal Processing, Guilin, 2006, Vol(1-4): 791-794.
 - [7] 薛丽萍, 尹俊勋, 纪震. 基于粒子群优化-模糊聚类的说话人识别[J]. *深圳大学学报(理工版)*, 2008, 25(2): 178-183.
Xue Li-ping, Yin Jun-xun, and Ji Zhen. Speaker recognition based on particle swarm optimization and fuzzy clustering analysis [J]. *Journal of Shenzhen University (Science and Engineering)*, 2008, 25(2): 178-183.
 - [8] 纪震, 廖惠连, 许文焕, 等. 粒子对算法在图像矢量量化中的应用[J]. *电子学报*, 2007, 38(10): 1916-1920.
Ji Zhen, Liao Hui-lian, and Xu Wen-huan, *et al.* A strategy of particle-pair for vector quantization in image coding [J]. *Acta Electronics Sinica*, 2007, 38(10): 1916-1920.
 - [9] Eusuffin M and Lansey K E. Optimization of water distribution network design using shuffled frog leaping algorithm[J]. *Journal of Water Resources Planning and Management*, 2003, 129(3): 210-225.
 - [10] Alireza R V and Ali H M. A hybrid multi-objective shuffled frog-leaping algorithm for a mixed-model assembly line sequencing problem [J]. *Computers & Industrial Engineering*, 2007, 53(1): 642-666.
 - [11] Garofolo J S and Lamel L F. TIMIT acoustic-phonetic continuous speech corpus [DB/CD]. <http://www.ldc.upenn.edu/Catalog/>, 2007-12.
- 薛丽萍: 女, 1962年生, 博士生, 副教授, 研究方向为语音信号处理、智能计算、嵌入式系统。
尹俊勋: 男, 1942年生, 教授, 博士生导师, 主要研究方向为音视频信号处理、通信、嵌入式系统。
周家锐: 男, 1973年生, 研究生, 研究方向为智能计算。
纪震: 男, 1973年生, 教授, 主要研究方向为智能计算、图像处理、数字水印以及数字信号处理硬件系统。