

优化的 HMM 算法在文本相关的说话人识别中的研究*

徐惠红

(辽东学院 信息技术学院, 辽宁 丹东 118000)

摘要:对 HMM 算法进行了优化,采用遗传算法与 LBG 算法相结合的方法生成码本,通过实验验证了优化后算法在文本相关说话人身份认证方面效率有所提高。

关键词: LBG 算法;隐马尔可夫;遗传算法;说话人识别

中图分类号: TP391

文献标识码: A

Recognition study of text dependent speakers based on optimized HMM algorithm

XU Hui Hong

(Faculty of Information and Technology, Liaodong University, Dandong 118000, China)

Abstract:The article has carried on the optimization to the HMM algorithm,the codebook which generated from genetic algorithm and LBG algorithm .It can be proved that the optimized algorithm improved the text dependent recognition efficiency through experiment.

Key words: LBG algorithm ;HMM;genetic algorithm;speaker recognition

1 HMM 算法的优化

Viterbi 算法^[1]是一种帧同步算法,对于每一帧计算所得的路径都保留在相应的寄存器(栈结构)中,在下一帧时对每一条路径都要继续扩展,尚未到模型出口状态的路径,会扩展成两条,而处于出口状态的路径扩展将更复杂,具体要视上层提供的候选转移路径数目而定。所以各帧对应的路径数是以指数级上涨,这是造成计算负担过大的重要原因。而实际上,大多数的路径得分都极低,在后续计算中都是没有竞争力的,因此完全可以将其排除在外。如果处理每一帧时都将其中大多数较低得分的路径删除掉,就可以减少很大一部分计算量。

优化的 Viterbi 算法是设一个常量范围,找到其中累计得分最高的路径,其得分为:

$$\Phi_i(s_i^*) = \max_{1 \leq j \leq N} [\Phi_i(j)] \quad (1)$$

其他的路径若累计得分远小于 $\Phi_i(j) \geq v_b$ (2) 则它在下一帧中要成为最佳候选路径的可能性很小。所以可以认定凡是累计得分低于 $b\Phi_i(s_i^*)$ 的路径都可以事先被删除。这里的 b 是个预定义好的常量。判断是否保留的门限值为:

$$v_b = b\Phi_i(s_i^*) \quad (3)$$

实际上,只有满足:

$$\Phi_i(j) \geq v_b \quad (4)$$

的路径才需要保留到下一帧的处理中去。实验中,为了防止下溢,所有的概率值都取了对数,因此乘法都用取对数后的加法代替。从实验的中间数据可以发现,这个常数没有普遍适应性,在某几帧数据比较集中的情况下,用这种剪枝方法会保留下来很多路径,这是比较合理的;但当出现某一路径特别突出的情况时,就可能只保留这一路径,而把全局真正最优路径排除在外,从而严重影响识别的精度。解决的办法是采用动态生成调整门限值,在数学上通常用黄金分割点的方法来取这个门限,具体做法是,比较得到当前帧所有路径中得分最高和最低的路径,记为:

$$\Phi_{\max}(s_i^*) = \max_{1 \leq j \leq N} [\Phi_i(j)] \quad (5)$$

$$\Phi_{\min}(s_i^*) = \min_{1 \leq j \leq N} [\Phi_i(j)] \quad (6)$$

则 $b = 0.618[\Phi_{\max}(s_i^*) \times \Phi_{\min}(s_i^*)]$ (7)

为解决实验中出现的保留路径值过度集中的情况,即会有很多在门限范围内的路径都被保留,定义一个辅助变量 k 和路径深度检测变量 y , k 为一个根据实验精度和计算量定义的一个值,当出现保留路径个数大于 k ,且这些路径没有在当前情况下有明显的将来会较优的表现时,将只保留 k 条这些中最优的路径;当出现路径深度值超过 y ,就要将动态门限值 b 重新计算减少 Δb

* 基金项目:辽宁省自然科学基金资助(20072009)

技术与方法

Technique and Method

值,这样就可将未来非优的路径早一点淘汰,以减少计算量。如果出现路径深度小于 y ,保留最优路径数小于 k 时,也可将动态门限值 b 增加 Δb ,增加保留的路径数、减少误淘汰率。这里 k 、 y 和 Δb 的取值不是固定的,要在实验中反复试验才能得到较满意的实验效果。

2 用遗传算法改进码本生成算法

算法将给定种群中的每一个个体(即染色体)认定为一个码书。设给定码书为 $C=\{y_1, y_2, \dots, y_n\}$,大小为 N ,训练矢量集为 $X=\{x_1, x_2, \dots, x_n\}$,每次迭代中都采用了LBG算法,实现遗传算法与LBG算法的结合^[2]。

(1) 编码方案

算法采用了二进制码书作为染色体的浮点数编码方法,它更适用于大样本空间的搜索,且局部搜索能力强,又不易陷入局部极值,收敛速度快。每个染色体对应一个码书,染色体中的一个基因对应一个码字的特征,所以一个种群由若干个码书组成。

(2) 适应值函数

对于每个染色体,都可以得到一个相应的码书 $C=\{y_1, y_2, \dots, y_n\}$,通过 $p_i = \operatorname{argmin}_{1 \leq j \leq M} \|x_i - y_j\|^2$ 算式即可得到训练矢量集的一个划分,然后由上式可计算出此码书的失真了 $f(p, c)$,因此在本算法中染色体的适应值函数取为:

$$\text{fitness} = \frac{1}{1+f(p, c)}$$

(3) 初始群体的形成

初始群体可以随机生成,为了缩短优化时间,算法随机从训练序列中挑选 M 个矢量构成一个含有 M 个码字的码书,一共产生 n 个初始个体组成一个初始群体。

群体应有一定的规模,以保持群体内个体的多样性,但为了减少运算时间,初始群体也不易选得太大。

(4) 选择算子

算法采取了最优保存策略和比例选择法相结合的思路。

首先在每一代开始的时候,把当前种群中适应度值最大,即最优的个体记录下来,然后对种群中的个体进行比例选择操作及以后的各种操作来产生新的种群,并从其中选出一个适应度值最小,即当前最差个体,用前面所记录的最优个体来替换它,这样就产生了进入下一代的种群。这样既能不断提高群体的平均适应度值,又能保证最优个体的适应度值不减小。

(5) 用遗传算法实现码本步骤

用一个变量 savebest 来保留最好的结果, $f(t-1)$ 为上代种群中最大适应度, $f(t)$ 为当前代的最大适应度。根据总失真长期在某个值上下震荡来作为终止条件,即 $(|f(t-1)-f(t)|/f(t)) < 0.0001$ 。也可以用进化代数达到多少代作为进化终止条件。算法描述为:

①生成初始种群,种群规模为 n ,放入集合 Set ,子代种群集合 Set1 赋值为 \emptyset ;

② Set 中随机抽取两个染色体 $X1$ 和 $X2$,按一定概率进行杂交运算,得到 X' ;

③对杂交后的个体 X' ,按一定概率分别对各单元进行变异得到 X'' ,把变异后的新染色体放入种群①中;

④如果 Set1 中的数目少于种群大小 n ,转入②;

⑤在形成新种群①后,采用了父子竞争的择优选择的方式,从父代群体 Set 集和子代集①中进行竞争,得到父子两代中较优的种群集②,种群规模为 n ;

⑥结合LBG算法,根据种群集中个体为含有 N 个单元的矢量池,可以得到有 N 个码字的码书,各个染色体中的 N 个码字根据最小平方失真对训练序列重新分类,这样就可以得到了改进后的染色体编码。把所有 Set2 中的染色体进行这样的处理,并替换掉该染色体;

⑦计算种群集 Set2 中各个体的适应值,选择其中一个适应度最好的染色体保存在 savebest 中,即 $\text{Set}=\text{Set2}$;

⑧判断终止条件,不符合终止条件则执行②至⑦;否则,结束。

3 实验结果及分析

实验采用在实验室无噪音环境下,10个志愿者:其中5名男性,5名女性分别录制40份同样内容wav格式的声音文件,20份用来做测试生成每个人的特征状态序列,另20份用来测试。采样频率为11025 Hz,量化值为16 bit,采用单声道语音进行识别测试,录音数据按帧长512点(采样点个数),帧移为256点,预加重为 $1-0.95Z^{-1}$,加哈明窗逐帧提取语音特征,模型的状态数 $(N=6)$ ^[8]。

3.1 声音的训练

实验选择了连续的左-右型HMM对声音进行建模。输入每个人的声音文件分别用改进的LBG算法和没改进算法生成特征值序列,用优化的和没优化的Viterbi算法确定状态序列,以Baum-Welch算法训练,则每个声音得到一个HMM $(\lambda=(\pi, A, B))$,为每个用户建模。

3.2 声音的认证

通过训练,输入待测声音特征向量 Q ,将 Q 和对应声音模型HMM $(\lambda=(\pi, A, B))$ 作为前向-后向算法(Forward-Backward算法)的入口参数,通过前向-后向算法计算待认证的声音特征向量 Q 及对应声音模型HMM $(\lambda=(\pi, A, B))$ 的匹配概率 $P(O|\lambda)$ 。当其值大于给定的阈值 ε_0 (经验值)^[8]时,则为真匹配,否则为伪匹配。

3.3 实验结果及分析

为了使得到的HMM能更好地描述特征值序列,分别测试了3转移、4转移的连续的左-右型HMM结构^[9],在不同参数提取、不同码本生成算法组合的实验结果如下表1。

实验结论:

(1)用遗传算法改进码本生成算法的方法明显提高

(下转第74页)

表 1 不同 HMM 参数、不同算法的识别率

选择的算法	MFCC 维数	3 转移1%		4 转移1%	
		识别率	误纳率	识别率	误纳率
LBG(优化)+Viterbi 优化	12	91.2	7.8	92.0	5.3
LBG(优化)+Viterbi 优化	24	92.5	6.0	93.6	4.6
LBG+Viterbi 优化	12	90.6	10.0	91.0	8.5
LBG+Viterbi 优化	24	91.5	9.8	92.3	7.3

了识别率, 最大量消除空包腔的方法很有效。

(2) 对于不同的左-右型 HMM 结构, 转移的状态数对实验结果有很大的影响, 通过实验数据能够发现 4 转移的左-右型明显优于 3 转移的左-右型。

(3) 在状态数、HMM 模型一样的情况下, MFCC 维数不同对实验结果影响很大, 当维数取 24 时识别率高于维数为 12 的情况, 即当码本的维数增加时对识别率的提高有一定的帮助, 但不同参数的情况下维数取多少, 还需要更多的实验来验证。在所有试验中, 4 转移的左-右型、LBG 码本优化、MFCC 参数为 24 时的实验结果最优。

由于实验是在无噪音的实验室环境下进行的, 从实验数据看, 采取码本优化和 HMM 技术优化相结合的方法

式应用于词汇量小、特定内容的语音识别方面取得较满意的实验结果。但是对于 HMM 模型参数和 MFCC 维数最佳组合在实际应用中及在词汇量大的语音识别方面还需进一步研究。

参考文献

- [1] 袁俊. HMM 连续语音识别中 Viterbi 算法的优化及应用[J]. 电子技术, 2001(2): 48-49.
- [2] 李廷, 王东进, 刘发林. 基于混和遗传算法的码书设计方法[J]. 电讯技术, 2007, 2(1): 151-153.
- [3] 陈良柱. 采用遗传算法的码本设计及说话人识别 [D]. 大连理工大学, 2008, 43-45.
- [4] 曹凯. 使用进化算法的矢量量化 [D]. 陕西师范大学, 2007: 19-21.
- [5] 程开东, 栾方军, 马驯良. 一种基于隐马尔科夫模型的在线手写签名认证算法[J]. 吉林大学学报(理学版), 2008, 46(5): 940-943.

(收稿日期: 2009-09-17)

作者简介:

徐惠红, 女, 1974 年生, 硕士, 主要研究方向: 模式识别。

《微型机与应用》2010 年第 02 期