

# 一种改进型 HMM 说话人识别算法

陶洁, 张会林

(上海理工大学 光电信息与计算机工程学院, 上海 200090)

**摘要:** 针对 Baum-Welch 算法依赖于初始值的选取而容易陷入局部最优解的问题, 基于全局优化的思想, 提出了一种改进的 HMM 语音识别算法。该算法将遗传算法应用到 HMM 模型训练中, 得到了全局最优解。实验结果表明, 所提出的算法使用有效, 识别率显著提高。

**关键词:** 遗传算法; HMM 模型; 语音识别

中图分类号: TP391.4

文献标识码: A

文章编号: 1674-7720(2012)21-0031-02

## Speech recognition based on GA and HMM

Tao Jie, Zhang Huilin

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200090, China)

**Abstract:** In order to avoid Baum-Welch algorithm falling into local optimum, an improved algorithm based on global optimization about training HMM-genetic algorithms is proposed. Experimental result shows that the improved algorithm is practical and effective, the recognition rate is improved significantly.

**Key words:** genetic algorithms; HMM model; speech recognition

近年来,随着语音识别技术的不断发展,语音识别系统的性能不断提高,隐马尔可夫模型方法是其中的一个主要识别方法。语音识别系统的识别率十分依赖于模型的训练,而经典的训练算法(Baum-Welch 算法)有一个致命的弱点,即最终所得的解十分依赖于初始值的选取,所以总是局部最优解,影响了整个系统的识别率。本文将遗传算法植入 HMM 模型参数的优化当中,使得整个语音识别系统的识别效果大大改善。

### 1 遗传算法

遗传算法将自然界的生物进化原理引入待优化参数形成的编码串联群体中,按所选择的适应度函数通过遗传中的复制、交叉及变异对个体进行筛选,使得适应度高的个体被保留下来,组成新的群体。通过不断迭代,保留下的个体适应度不断提高,直到满足一定的条件。从数学的角度解释,可以简单地认为,基因重组使子代基因趋向于局部最优解,而基因变异能使子代基因突破局部范畴,经过多代的交叉和变异,达到全局最优解。

根据待优化问题的数学模型,定义适应度函数  $F(a_i)$ , 其中  $a_i$  为其中的一条染色体,则  $F(a_i)$  就是判断该染色体优劣的依据。对于每一代基因,计算所有染色体的适

应度函数,进行排序,选择一定数目的优秀染色体,用于产生子代的父代样本。

复制操作只能从旧种群中选择出较优秀的染色体,但是不能创造出新的染色体。交叉操作模拟了生物进化过程中的交配过程,通过两条染色体的分离重组,产生新的优良物种。图 1 为多点交叉重组的示意图。交叉过程为:在匹配池中任选两条染色体,随机选择一点或多点交换点,然后交换双亲染色体交换点右边的部分,即可得到两条新的子代染色体。



图 1 多点交叉重组

最佳基因是在一代一代的基因重组和基因变异中产生的。基因突变用来模拟生物在自然的遗传环境中由于各种偶然因素引起的变异突变,它以很小的概率随机地改变染色体编码串中的某一位。变异算法有利于局部最优处跳出,防止算法过早地收敛。

遗传算法的具体实现步骤如下:

(1) 随机产生最初的染色体群体  $p=\{a_1, a_2, \dots, a_L\}$ ,  $a_i$  为其中的一条染色体。

(2) 计算各个染色体的适应度函数  $F(a_i)$ , 并对适应度函数  $F(a_i)$  进行排序, 根据设定的门限选取一定数目的优秀染色体作为产生子代染色体的父代样本。

(3) 以一点或多点交叉产生新的子代染色体, 交叉点随机生成。

(4) 设定变异概率门限, 根据生成的随机数决定染色体中的某个变量是否有突变发生。

## 2 HMM 说话人识别系统

识别主要有两个任务: 一个是对每个 HMM 模型的训练, 也就是计算 HMM 参数; 另一个是识别任务, 也就是已知了参考模型的 HMM 参数, 决定未知模式与哪个参考模式是最佳匹配。训练采用 Baum-Welch 算法, 识别采用 Viterbi 算法。

### 2.1 Baum-Welch 估计

给定一个观察值序列  $O=o_1, o_2, \dots, o_N$ , 以及一个需要通过训练进行重估参数的 HMM 模型  $M=\{A, B, \pi\}$ 。重估公式如下:

$$\hat{a}_{ij} = \sum_{k=1}^{N-1} \xi_k(i, j) / \sum_{k=1}^{N-1} \gamma_k(i) \quad (1)$$

$$\hat{b}_i(r) = \sum_{(k=1 \& o_k=r)}^N \gamma_k(i) / \sum_{k=1}^N \gamma_k(i) \quad (2)$$

其中,  $\sum_{k=1}^{N-1} \gamma_k(i)$  是由状态  $i$  期望转移的次数,  $\sum_{k=1}^{N-1} \xi_k(i, j)$  是由状态  $i$  转移到状态  $j$  的期望个数,  $\gamma_k(i)$  表示  $k$  阶段在状态  $i$  的概率,  $\xi_k(i, j)$  表示  $k$  阶段在状态  $i, k+1$  阶段在状态  $j$  的概率。

利用式(1)、(2), 训练步骤如下:

(1) 给未知模型假设初始条件, 计算  $P(X|S)$ 。

(2) 使用式(1)、(2), 由当前的模型参数的估计值, 重新估计新的模型  $\bar{S}$  的参数。

(3) 计算  $P(X|\bar{S})$ 。如果  $P(X|\bar{S}) - P(X|S) > \epsilon$ ,  $S = \bar{S}$  并且回到步骤(1), 否则停止。

### 2.2 Viterbi 识别

给定一个观察值序列  $O=o_1, o_2, \dots, o_N$ , 以及一个 HMM 模型  $M=\{A, B, \pi\}$ , 在最佳的意义上确定一个状态序列  $S=s_1, s_2, \dots, s_N$ 。这里的“最佳”是指使  $P(O, SIM)$  最大。 $P(O, SIM)$  计算式为:

$$P(O, SIM) = P(s_1)P(o_1|s_1) \prod_{k=2}^N P(s_k|s_{k-1})P(o_k|s_k) \quad (3)$$

观察式(3), 定义每步转移的代价为

$$d(s_k|s_{k-1}) = \ln P(s_k|s_{k-1}) + \ln P(o_k|s_k) \quad (4)$$

$$d(s_1|s_0) = \ln P(s_1) + \ln P(o_1|s_1) \quad (5)$$

定义经过  $k$  阶段的代价为  $D(K) = \sum_{r=1}^k d(s_r, s_{r-1})$ , 则

在每一步转移中使  $D(k)$  最大的状态  $i$  构成的序列即为所求的最佳状态序列。

## 3 用遗传算法改进 HMM

基于遗传算法改进的 HMM 系统的设计模型如图 2 所示。

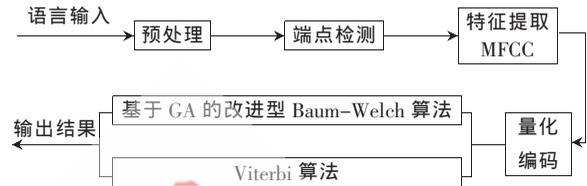


图 2 改进型 HMM 系统

在将遗传算法引入 HMM 训练的过程中, 首先要解决的是染色体的构造问题。将 HMM 模型的所有参数排列成一串, 构成染色体。对于语音识别, 采用自左向右的 HMM 模型, 本文中为 5 状态自左向右只含一阶跳转的 HMM 模型。HMM 模型  $\lambda=\{A, B, \pi\}$ , 参数由初始状态向量  $\pi$ 、状态转移矩阵  $A$  及每个状态的输出概率密度函数组成。向量  $\pi$  含有 5 个元素, 转移矩阵  $A$  中共含有元素 25 个, 其中不为 0 的参数为 9 个。各状态的输出概率密

度函数  $b_j(o) = \sum c_{jk} N(o, \mu_{jk}, \Sigma_{jk})$ 。其中  $j$  代表状态,  $c_{jk}$

为混合系数,  $N(\cdot)$  为高斯分布,  $\mu_{jk}$  为平均矢量,  $\Sigma_{jk}$  为协方差矩阵。语音特征参数采用 24 阶 Mel 倒谱系数, 选取为 5 个高斯概率密度函数的混合。将初始状态向量  $\pi$ 、转移矩阵  $A$  和混合系数矩阵  $c$  的参数共  $5+9+25=39$  个按行组成一串, 形成染色体的前一部分, 将平均矢量  $\mu_{jk}$  和协方差矩阵  $\Sigma_{jk}$  共  $5 \times 5 \times (10+24 \times 24) = 14\ 650$  个参数按行组成一串, 形成染色体的后一部分。在染色体前一部分的行向量之和均为 1。因此, 需要在生成每一代染色体前, 对这一部分行向量所对应的每一段染色体进行归一化, 以满足 HMM 的约束条件。

在遗传算法中, 适应度函数作为区分个体优劣的标准, 需保证优秀个体的适应度比差的个体适应度要高。这里个体的适应度用各个训练样本的对数似然概率表

示, 即  $f(\lambda) = \sum_{i=1}^T \ln(P(o_i|\lambda))$ 。

实验中, 染色体的前一部分依概率进行二点或多点交叉, 而后一部分染色体只进行多点交叉, 多点交叉概率为  $\rho_c=0.8$ 。染色体前一部分的变异概率为  $\rho_m=0.1$ , 后一部分变异概率为  $\rho_m=0.04$ 。此外, 设置系统最大进化代数为 100。

## 4 实验结果分析

实验基于 HMM 进行语音识别, 训练数据取自 10 人, 在不同 SNR 下获取语音数据, 时间长度为 5 s, 采样

率为 8 kHz, A/D 转换精度为 16 bit, 单声道。遗传算法的优化过程中适应度函数  $F$  的变化过程如图 3 所示。语音识别系统识别结果如表 1 所示。

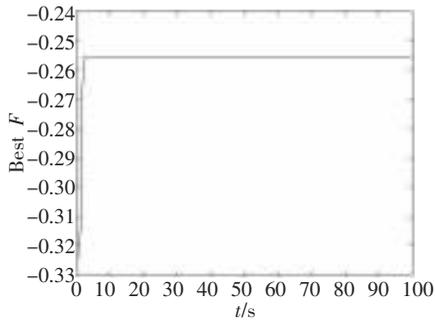


图 3 适应度函数  $F$  变化过程

表 1 两种系统不同 SNR 下识别率比较

SNR/dB	10	15	20	40	60
系统 I	47.8	83.4	85.0	86.7	87.6
系统 II	53.0	85.5	86.9	87.2	87.7

表 1 中, 系统 I 为基于 HMM 的语音识别系统, 系统 II 为基于遗传算法改进的 HMM 语音识别系统。从实验结果可以看出, 基于遗传算法改进的 HMM 模型对语音识别率有较大改善, 平均提高了 4~5 个百分点。

本应用基于 HMM 语音识别算法, 结合遗传算法进

行 HMM 模型训练并将其应用于语音识别中。经过实际验证, 计算简单, 效果好, 基本达到了设计目的和现实要求。

#### 参考文献

- [1] 张思才, 张方晓. 一种遗传算法适应度函数的改进方法[J]. 计算机应用与软件, 2011, 23(2): 108-110.
- [2] MANIEZZO V. Genetic evolution of the topology and weight distribution of neural networks [J]. IEEE Transactions on Neural Networks, 1994, 5(6): 900-909.
- [3] TERASHIMA R, YOSHIMURA T, WAKITA T. Prediction method of speech recognition performance based on HMM-based speech synthesis technique [J]. IEEE Transactions on Electronics, Information and Systems, 2010, 130: 557-564.
- [4] 胡广书. 现代信号处理教程[M]. 北京: 清华大学出版社, 2004.

(收稿日期: 2012-08-03)

#### 作者简介:

陶洁, 女, 1988 年生, 硕士研究生, 主要研究方向: 嵌入式语音信号处理。

张会林, 男, 1970 年生, 博士, 副教授, 主要研究方向: 信息采集与处理、图像处理。