

# 一种 DHMM 的混合训练方法

茅晓泉, 胡光锐, 唐 斌

(上海交通大学电子工程系, 上海 200030)

**摘 要:** 隐马尔柯夫模型(HMM)作为描述语音信号的一个工具,按输出概率分布的不同,可分为连续 HMM(CHMM)和离散 HMM(DHMM)。经典的训练方法 Baumr Welch 算法虽然收敛迅速,但是这类基于爬山的算法只能取得局部最优解,从而影响了系统的识别率。对于 CHMM,借助于分类 K 平均方法可以取得可靠的初始点以保证迅速准确的收敛。而对于 DHMM,该方法收益不大,最终所得的仍是局部最优解。由于进化计算一个最重要的特点便是全局搜索,这样可得全局最优解或次优解。本文将进化计算应用到 DHMM 的训练中,提出了一个把传统算法和进化计算相结合的混合算法。实验结果表明该方法既保证了全局搜索又实现了快速收敛,最终所得的模型优于传统方法和简单进化计算方法。

**关键词:** 隐马尔柯夫模型; 进化计算; 语音识别

**中图分类号:** TN912.34

**文献标识码:** A

**文章编号:** 0372-2112 (2002) 01-0148-03

## A Hybrid Training Method for DHMMs

MAO Xiao-quan, HU Guang-rui, TANG Bin

(Dept. of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract:** Hidden Markov Models are very successful in modeling the acoustic behavior of speech. They may be classified into two groups, continuous HMMs (CHMMs) and discrete HMMs (DHMMs), according to the output probability distribution. Traditional training methods such as Baumr Welch algorithm are noted for the rapid convergence. However, these methods are hill-climbing based algorithms and they just lead to locally optimal solutions, which might deteriorate the recognition rate. For CHMMs, a segmental k-means method has been developed to get reliable initial estimate and thus guaranteed the rapid and proper convergence. For DHMMs, this offers little help and the final solution is a locally optimal solution. While one outstanding character of evolutionary computation is global search, it can converge to a globally optimal solution or at least a sub-optimal solution. In this paper evolutionary computation is applied to training DHMMs. A hybrid training method that combines the traditional method and evolutionary computation is proposed. Experimental results show that the proposed method has both qualities of global search and rapid convergence and the resulting models are superior to those obtained with traditional methods or simple evolutionary computation and eventually contribute to the increase of recognition rate.

**Key words:** HMMs; evolutionary computation; speech recognition

## 1 引言

隐马尔柯夫模型是语音识别领域的一个很有效的数学工具,它采用统计建模方法来描述语音的时域特性。其输出概率采用离散概率分布(DHMM),或连续概率密度函数(CHMM)。具体的实现包括两部分:训练和识别。经典的训练算法便是 Baumr Welch 算法<sup>[1]</sup>,它作为基于爬山的算法,有两个优点,一是收敛迅速,二是每一步迭代后似然概率都增大。但是它有一个致命的缺陷,最终所得解依赖于初始值的选取,从而往往只是局部最优,影响了最终系统的识别率。

基于此,出现了分类 k 平均以获取一个好的初始估计<sup>[1]</sup>。它对于 CHMM 很有效,但是对于 DHMM 帮助不大。由于进化计算中使用的是基于种群的全局搜索,获取全局最优解的机

会要大许多<sup>[2]</sup>。然而进化计算需要大量代的进化方能接近全局最优<sup>[3]</sup>,这将意味着很大的计算量。

本文提出了一个混合训练算法,把进化计算和 Baumr Welch 算法结合起来。实验结果表明了该方法既能全局搜索且收敛迅速,产生的模型最优。

## 2 经典的 Baumr Welch 算法

HMM 是基于这样的假设:语音信号可以用一个参数随机过程来描述。它用一个双随机过程来描述语音的时域特性。DHMM 和 CHMM 的不同之处在于输出的分布。在本文的实验中,采用的是一个 5 状态从左到右无跳跃的 DHMM<sup>[1]</sup>。码本的大小设为 128。模型中的观察值和状态的转移相关。这里用

$= (A, B)$  来表示模型的所有参数.  $A = [a_{ij}]$  是状态转移概率矩阵,  $B = [b_{ijk}]$  是输出分布矩阵.  $O = O_1 O_2 \dots O_T$  表示一个观察序列.

实现中主要有训练和识别两个模块. 首先是从训练数据中取得模型的参数, 然后便由训练好的模型来进行识别未知的数据. 因此, 训练过程至关重要, 不准确的训练生成不准确的模型, 从而直接影响系统的性能. 在传统的 Baum-Welch 算法, 采用了迭代算法, 每一次迭代均使得似然概率  $P(O|\lambda)$  增大, 直至  $P(O|\lambda)$  的增大程度低于预设的门限值. 在该方法中, 收敛速度快, 每次迭代后的参数无需做归一化处理. 但是它有一个致命缺陷, 最终解依赖于初始参数的设置, 只是初始点附近的一个局部最优解. 这也是所有基于爬山的算法的共有的缺陷. 用这样得到的局部最优的模型去做识别, 系统的性能自然受到了影响. 进化计算由于它的很多优点, 如全局搜索、操作简单及自学习<sup>[2]</sup>, 在各个领域得到了广泛的应用. 基于此, 本文尝试了把它引入到 DHMM 的训练中.

### 3 进化计算的设计

进化计算的灵感来自于自然选择的物竞天择的机制. 它用一组参数来表示个体, 这些参数也就是个体的基因. 一群个体组成一个种群, 对种群的简单遗传操作生成后代, 再由适应度来引导搜索. 这样, 由目标函数引出的适应函数不再必须是可微或连续的. 具体的算法设计, 包括编码表示、适应函数、选择策略、控制参数的选取以及遗传算子和终止准则的设计.

#### 3.1 编码表示

编码表示的设计是整体设计的一个重要部分. 简明准确的编码可以提高算子的效率, 而冗余的表示会增加运算的负担. 这里每一个个体对应于一个 HMM, 染色体对应于模型的参数. 参数包括两部分: A 和 B. 对于 5 状态从左到右无跳跃的 DHMM, A 中有 9 个非零值. 因为  $\sum_{j=1}^5 a_{ji} = 1, j = 1 \dots 5$ , 所以实验中只取了四个独立的参数来组成染色体的前一部分, 便是  $a_{i, i+1}$ . 对于 B, 由  $b_{i, i, k}, i = 1 \dots 5, k = 1 \dots 128$  和  $b_{i, i+1, k}, i = 1 \dots 4, k = 1 \dots 128$  组成了染色体的后一部分. 因为  $\sum_{k=1}^{128} b_{i, i, k} = 1, i = 1 \dots 5$  和  $\sum_{k=1}^{128} b_{i, i+1, k} = 1, i = 1 \dots 4$ , 为保证每次产生的新一代种群的个体参数仍满足此二约束, 需要做归一化的工作.

#### 3.2 适应函数

在自然选择中, 适者生存, 不适者淘汰. 进化计算中, 个体的适应度由适应函数来度量, 适应函数和目标函数相关. 必须保证越符合目标的个体其适应度越高. DHMM 的训练中希望训练数据对模型的似然概率越高越好, 因此这里个体的适应度用各个训练样本的对数似然概率的和来表示,  $f(\lambda) = \sum_{k=1}^K \ln(P(O^{(k)}|\lambda))$ , 这里  $O^{(k)}$  表示用于训练模型的第  $k$  个观察序列.  $K$  表示训练样本的数目. 对于 DHMM, 由于似然概率均是在 0 和 1 之间, 显然这一适应度是一个负值. 因此下面选择策略的设计, 便不能直接采用基于适应值比例的选择策略.

### 3.3 选择策略

选择策略的选取直接影响着算法的收敛. 最常用的是基于适应值比例的选择和基于排名的选择. 前者还要求适应值均为正值. 由于使用基于适应值比例的选择常常会出现过早收敛和停滞现象<sup>[4]</sup>, 实验中采用了非线性排名选择. 对于每一代种群, 个体按其适应度排序, 再按排名分配选择概率, 适应度高的选择概率也高. 这样也避免了基于适应值比例选择的对适应值调整的工作. 但是也多出了一个排序工作, 实验中采用了快速排序算法.

### 3.4 遗传算子和控制参数

遗传算子包含杂交算子和变异算子. 在一定意义上, 杂交算子相当于一个局部搜索操作, 它产生父代附近的两个子代. 而变异算子则使得个体能够跳出当前的局部搜索区域. 二者的结合正是进化计算的进化所在. 针对 DHMM 的参数结构, 分别对 A 和 B 中的两部分进行单点杂交. 每一个点对应于一个状态. 每一个点随机抽取, 将父代对应的  $a_{i, i+1}, b_{i, i, k}, k = 1 \dots 128, b_{i, i+1, k}, k = 1 \dots 128$ , 进行互换. 变异算子则采用了均匀性变异, 对 A 和 B 中随机抽出的点进行变异. 至于控制参数, 种群的大小设为 30, 进化的最大代数设为 10000, 杂交的概率取 0.9.

### 3.5 终止准则

常用的终止准则是预先设置最大进化的代数或预先设置一个适应度改善的门限值. 对于前一种准则, 在进化代数到达预置值时进化终止. 在后种情况下, 在适应度的改善低于该门限值时进化停止. 实验中采用了第一种终止准则, 最大的进化代数取 10000.

### 3.6 混合算法

进化计算的智能性在于杂交算子使得子代趋近局部最优解而变异算子使得子代能够跳出当前的局部搜索区域, 从而增大逼近全局最优的可能性<sup>[4]</sup>. 换句话说, 杂交属局部搜索, 而变异和基于种群的操作带来全局搜索.

但是在实验中, 发现杂交算子产生的子代经常并不理想<sup>[5]</sup>. 也就是说, 局部搜索的能力并不够. 而 Baum-Welch 算法是 HMM 训练的局部搜索的经典算法, 由此我们考虑把它嵌入到遗传操作中. 同时还要避免出现过早收敛和保持种群的多样性. 由此, 在每一代子代产生以后, 随机抽取一个个体, 把它的参数用 Baum-Welch 算法进行一次迭代更新, 这样新的个体取代原个体. 为区别于不加入 Baum-Welch 算法的进化计算方法, 该方法称为混合训练方法, 不加入 Baum-Welch 算法的称为简单进化计算方法. 下面的实验结果表明使用混合训练方法获得的模型要优于 Baum-Welch 方法和简单进化计算方法.

## 4 实验结果

分别用上述算法对汉语孤立数字识别进行了实验. 语音训练数据取自 2 人. 语音信号经过 8kHz 采样器采样, A/D 变换精度为 16 比特. 采样信号首先经由一个一阶数字预加重滤波器处理 (预加重系数为 0.95). 再通过 Hamming 窗, 窗长为 45ms, 帧移为 15ms. 然后逐帧求取特征矢量, 包括 11 阶 LPC 倒

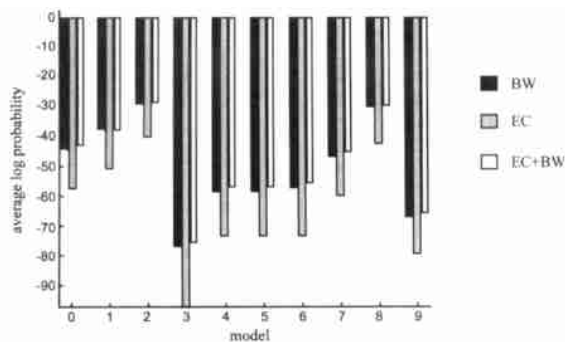


图1 三种方法所得模型平均对数概率的比较

谱系数和对数能量。

分别用上述三种方法来训练 DHMM。在 Baumr Welch 方法中,迭代至对数似然概率的增加小于 0.00001。由于进化计算的随机性,简单进化计算方法和混合训练方法均采用运行 5 次求平均值的方法。表 1 中列出了各个方法求得各个模型的平均对数似然概率。

表1 三种方法的训练结果

	BW	EC	BW + EC
# 0	- 43.9210	- 56.7932	- 42.9159
# 1	- 37.4857	- 51.1048	- 38.3626
# 2	- 28.8802	- 39.5739	- 28.2368
# 3	- 76.7652	- 97.3913	- 75.9631
# 4	- 57.6254	- 73.3609	- 57.0752
# 5	- 58.4486	- 73.9349	- 57.4176
# 6	- 57.5764	- 73.7499	- 55.8659
# 7	- 47.1669	- 59.6641	- 45.2452
# 8	- 30.1656	- 42.6226	- 30.1625
# 9	- 67.0622	- 80.0566	- 66.3491

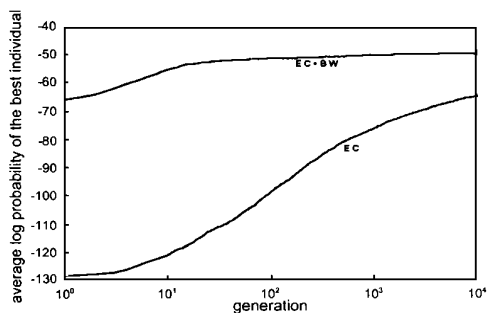


图2 简单进化计算方法和混合训练方法的收敛性能比较

显然混合方法生成的模型是最优的。由于最大进化代数设为 10000,简单进化计算生成的模型尚未收敛至全局最优解。

图 2 显示了简单进化算法和混合算法训练中的收敛过程,纵轴代表每代中最优个体的平均对数似然概率,横轴代表迭代的代数。由于混合算法改进了局部搜索的能力,由图 2 中可以看出,混合训练方法的收敛速度要比简单进化算法快得多,同时生成的模型的似然概率也要高许多。另外,混合训练在进化至第 2000 代左右已大致收敛至最终解。这样,在训练中可以预设训练的最大进化代数为 2000,从而大大节省了训练时间。

对于简单进化计算训练方法,由图 2 可以看出在进化至第 10000 代时,训练尚未收敛,模型还有很大的改善趋势。如果增大最大进化代数,模型方能收敛至全局最优解,但是这样造成计算量进一步增大,训练时间延长。

## 5 结论

本文把进化计算引入到 DHMM 的训练中,已获得更优的模型。为改善算法的收敛速度,提出了一个把 Baumr Welch 算法和进化计算相结合的混合算法。实验结果表明,该方法生成的模型对训练数据的描述更准确,同时又提高了收敛速度。它兼具了 Baumr Welch 算法和进化计算的优点。

## 参考文献:

- [1] L Rabiner, B H Juang. Fundamentals of Speech Recognition [M]. Englewood: Prentice Hall, 1993: 342 - 344.
- [2] K F Man, K S Tang, S Kwong. Genetic Algorithms Concepts and Designs [M]. Berlin: Springer, 1999: 219 - 225.
- [3] C W Chau, S Kwong, C K Diu, et al. Optimization of HMM by a genetic algorithm [C]. Proc. ICASSP 97: 1727 - 1730.
- [4] 潘正君, 康立山, 陈毓屏. 进化计算 [M]. 北京: 清华大学出版社, 1998: 16 - 29.
- [5] K Rasheed. Guided Crossover: A new operator for genetic algorithm based optimization [C]. Proc. ICEC '99: 1535 - 1541.

## 作者简介:



茅晓泉 男, 1974 年生, 1996 年获上海交通大学学士学位, 现在上海交通大学电子工程系攻读博士学位。主要从事进化计算与语音识别研究。



胡光锐 男, 上海交通大学电子工程系教授, 博士生导师。主要的研究兴趣包括: 语音识别, 神经网络, 通信系统的抗干扰研究, 协同学和混沌神经网络在语音识别中的应用等。E-mail: grhu@mail.sjtu.edu.cn



唐斌 男, 1968 年生, 1989 年、1997 年分获西安电子科技大学学士、硕士学位, 1998 年入上海交通大学电子工程系攻读博士学位。主要从事系统仿真及电子战信号处理研究。