汉语连续语音识别中多项式拟合语音轨迹模型的研究

欧智坚,王作英
(清华大学电子工程系,北京 100084)

摘 要: 尽管作为当前最为流行的语音识别模型, HMM 由于采用状态输出独立同分布假设,忽略了对语音轨迹动态特性的描述. 本文基于一个更为灵活的语音描述统计框架——广义 DDBHMM,提出了一个具体的多项式拟合语音轨迹模型,以及新的训练和识别算法,更好地刻划了真实的语音特性. 本文还给出了一种有效的剪枝算法,得到一个实用化模型. 汉语大词汇量非特定人连续语音识别的实验表明,这种剪枝的多项式拟合语音轨迹模型以较少的计算量明显改善了识别系统的性能.

关键词: 连续语音识别; 隐马尔可夫模型; 基于段长分布的隐马尔可夫模型
中图分类号: TN912.34 **文献标识码**: A **文章编号**: 0372-2112 (2003) 04-0608-04

Research on Polynomial-Fitting Speech-Trajectory Model in Chinese Continuous Speech Recognition

OU Zhi-jian, WANG Zuo-ying
(*Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*)

Abstract: Although as the most popular model for speech recognition, HMM takes no account of the dynamics of the speech trajectory, since it assumes the outputs of a state to be independent and identically distributed. In this paper, based on a more flexible statistical framework for speech description-the generalized DDBHMM, a particular polynomial-fitting speech-trajectory model is proposed with new algorithms for training and recognition. It describes the real characteristics of speech more reasonably. With the effective path-pruning algorithm additionally proposed, it becomes a practicable model. Experiments on Chinese large-vocabulary speaker-independent continuous speech recognition showed that with this path-pruned polynomial-fitting speech-trajectory model, the recognition performance is improved distinctively at relatively low computational cost.

Key words: continuous speech recognition; hidden markov model (HMM); duration distribution based HMM (DDBHMM)

1 引言

现今成功的连续语音识别系统几乎都基于隐马尔可夫模型(HMM)构建. 但是 HMM 引入的状态输出独立同分布假设忽略了对语音轨迹动态特性的描述,发声器官作为一动力学系统所固有的时变特性未予考虑,从而阻碍了识别系统性能的进一步提高.

一个最早的做法是,在特征中加入一、二阶倒谱与能量的差分,使当前帧包含其前后若干帧的动态信息,从而提高识别性能. 但这还不够,目前的研究方法主要有: (1) 把对状态输出变长帧序列 $o_1 \cdots o_T$ 的统计描述通过某种轨迹采样化归为对定长帧序列 $\tilde{o}_1 \cdots \tilde{o}_R$ 的联合高斯描述[1],但估计这么大的协方差阵又成了一个问题. (2) 通过假设状态输出帧序列 $o_1 \cdots o_T$ 独立不同分布,建立起参数化的语音轨迹模型去显式描述语音轨迹的均值缓变[2,3],但计算量急剧增加,需要合理的剪枝算法. (3) 在状态输出随机过程中引入 Markov 性,表示为自回归过程[4~6],但实用中并没达到意想中的好效果. 以上这些做法大多是对较小实验集(非特定人英语 E 集字母[4,5]或特定人小词表的英语孤立字[1~3,6])作出的.

总的看来,还没能有一个能应用于汉语大词汇量非特定人连续语音识别中的有效的实用化模型. 本文基于一个更为灵活的语音描述统计框架——广义 DDBHMM,提出了一个具体的多项式拟合语音轨迹模型,给出了新的训练和识别算法,研究了实用化的剪枝算法. 我们进行的汉语大词汇量非特定人连续语音识别的实验表明,这种剪枝的多项式拟合语音轨迹模型以较少的计算量明显改善了识别系统的性能.

2 广义 DDBHMM

一般地,对于语音模型的一个合理提法是:一个输出特征为 $O = o_1 o_2 \cdots o_T$,具有状态段长分布的 N 状态 $(s_1 \cdots s_N)$ 语音过程,其统计特性(概率分布)为

收稿日期:2001-08-20;修回日期:2002-08-12
基金项目:863 基金(No. 2001AA114071)

$$P(\tau_1 \cdots \tau_N; o_1 \cdots o_T | s_1 \cdots s_N) = P(\tau_1 \cdots \tau_N | s_1 \cdots s_N) \cdot P(o_1 \cdots o_T | s_1 \cdots s_N, \tau_1 \cdots \tau_N) \quad (1)$$

其中 τ_n 是第 n 个状态(段) s_n ($n = 1, \cdots, N$) 对应的段长,$1 = P_1 \le P_2 \le \cdots \le P_N \le P_{N+1} = T + 1$ 是状态分割点序列,$\tau_n = P_{n+1} - P_n$, $n = 1, \cdots, N$, $\sum_{n=1}^{N} \tau_n = T$;状态 s_n 对应的输出特征为 $o_{P_n} o_{P_n+1} \cdots o_{P_{n+1}-1}$. 记时刻 t 系统驻留在状态 $q_t \in \{s_1, \cdots, s_N\}$,则状态段长序列 $\tau_1 \cdots \tau_N$,状态分割点序列 $1 = P_1 \le P_2 \le \cdots \le P_N \le P_{N+1} = T + 1$,以及状态序列 $Q = q_1 \cdots q_T$,三者是等价的. 于是

$$P(\tau_1 \cdots \tau_N; o_1 \cdots o_T | s_1 \cdots s_N) = P(\tau_1 \cdots \tau_N | s_1 \cdots s_N) \cdot P(o_1 \cdots o_T | q_1 \cdots q_T) \quad (2)$$

上式是对 N 段语音过程的完备描述,称为广义 DDBHMM[7]. 实用时要作一些简化. 根据不同的建模思想,引入一些假设,做不同的展开,从而建立起不同形式的模型.

假设各状态段之间独立,状态内帧间特征相互独立,并且状态段长为均匀分布,则有

$$P(\tau_1 \cdots \tau_N; o_1 \cdots o_T | s_1 \cdots s_N) = \text{常数} \cdot \prod_{n=1}^{N} \prod_{t=P_n}^{P_{n+1}-1} P(o_t | s_n) \quad (3)$$

这是目前我们所用的"基线模型".

式(3)中引入的独立性假设忽略了语音特征沿时间轴的相互联系,因此为了描述语音轨迹的动态特性,需要回到式(2). 下面假设状态段长是均匀分布,而着力研究状态输出模型

$$P(\tau_1 \cdots \tau_N; o_1 \cdots o_T | s_1 \cdots s_N) = \text{常数} \cdot P(o_1 \cdots o_T | q_1 \cdots q_T) \quad (4)$$

即归结为概率分布 $P(o_1 \cdots o_T | q_1 \cdots q_T)$,即 $P(O | Q)$.

3 多项式拟合语音轨迹模型

一个状态(段) s 代表一定颗粒度层次上的语音单元,状态输出帧序列随时间在特征空间划出一条轨迹,称为"语音轨迹". 每段语音轨迹都是状态 s 的输出随机过程的一个实现. 现有 HMM 假设状态输出独立同分布,这样直观地看,是用方波(常数矢量 ——均值)去拟合属于 s 的语音轨迹,描述的也只有这些语音轨迹的静态特性;从而一句话(很多状态的级联)就是用一分段方波 ——阶梯波去拟合. 但这并不够. 我们可以用时间的多项式函数去描述语音轨迹缓变的动态特性,而这正是现有 HMM 所忽略的一种重要的识别信息.

3.1 模型描述

下面的多项式拟合语音轨迹模型基于式(4). 进一步假设各状态段之间统计独立,于是 $P(o_1 \cdots o_T | q_1 \cdots q_T) = \prod_{n=1}^{N} P(y_n | s_n)$,其中 $y_n = o_{P_n} o_{P_n+1} \cdots o_{P_{n+1}-1}$ 是第 n 个状态 s_n 对应的输出特征. 这样就可以对每个状态段分开进行描述.

假设状态 s 对应的 D 维语音轨迹 $o_0 o_1 \cdots o_{\tau-1}$ 可用时间 t 的一个 m 阶多项式拟合如下

$$o_t = \sum_{i=0}^{m} k_i^s t^i + v_t, \quad t = 0, \cdots, \tau - 1 \quad (5)$$

拟合误差 v_t 假设为状态内帧间不相关零均值高斯白噪声,$v_t \sim N(0, \sum_s)$; $k_i^s \in R^D$ ($i = 0, \cdots, m$), $\sum_s$ 是与状态 s 有关的模型参数. 语音轨迹在不同时刻有不一样的均值 $\bar{o}_t = \sum_{i=0}^{m} k_i^s t^i$,描述了一种缓变的动态特性,这正是我们所希望的. 于是,

$$P(o_0 o_1 \cdots o_{\tau-1} | s) = \prod_{t=0}^{\tau-1} \frac{1}{(2\pi)^{D/2} |\sum_s|^{1/2}} \exp\{ -\frac{1}{2} (o_t - \sum_{i=0}^{m} k_i^s t^i)^{Tr} \cdot \sum_s^{-1} (o_t - \sum_{i=0}^{m} k_i^s t^i) \} \quad (6)$$

可以看出,现有 HMM 是上述多项式拟合语音轨迹模型在 $m = 0$ 的一个特例.

3.2 训练和识别算法

记系统的参数为 $\lambda = \{ k_i^s (i = 0, \cdots, m), \sum_s | s \in S \}$, S 为系统的所有状态的集合.

训练即最大似然参数估计,是一个迭代过程. 每一步迭代包括一个根据已有状态标注对训练语音的分割过程,和一个利用分割得到的统计量进行逐状态模型参数重估过程. 识别即最大似然译码,是一个寻找对输入特征进行最优状态分割过程.

训练中的分割过程与识别的分割过程很相似,可以使用相同的算法(只是训练时的状态序列被限制是已知的标注,更简单一些). 在废除状态输出独立同分布假设后,标准的 Viterbi 分割算法不再适用. 依据动态规划(DP)原理,我们提出一个更灵活的分割算法.

输入特征序列 $O = o_1 o_2 \cdots o_T$,分割过程即 $\max_{Q = q_1 \cdots q_T} P(o_1 \cdots o_T | q_1 \cdots q_T)$.

由时刻 t 和状态 s 组成了一个大的网格点空间 (t, s),$1 \le t \le T$, $s \in S$. 一个状态序列(一个分割)就对应于网格点按照时间顺序连接成的一条路径. 网格点之间的连接关系由字表、发音规则决定(训练时,还受标注的约束). 记集合 $\Omega(t, s) = \{ q_{t-1} | q_t = s \}$,即当时刻 t 驻留在状态 s 时,时刻 $t - 1$ 的所有可能驻留状态集合;则网格点 (t, s) 前面可能接的网格点集合为 $\{ (t - 1, s') | s' \in \Omega(t, s) \}$. $\Omega(t, s)$ 表示了网格点的连接关系. 如果 $q_t = s$,记状态序列 Q 开始进入状态 s 的时刻为 ρ_t.

定义 $D(t, s, \rho) = \max_{q_1 \cdots q_{t-1}} P(o_1 \cdots o_t | q_1 \cdots q_{t-1}, q_t = s, \rho_t = \rho)$, $0 \le \rho \le t$ (7)

即在时刻 ρ 进入状态 s 且一直驻留至时刻 t 的一条最优路径的似然值(匹配得分).

用 $Q(t, s, \rho) = \arg\max_{q_1 \cdots q_{t-1}} P(o_1 \cdots o_t | q_1 \cdots q_{t-1}, q_t = s, \rho_t = \rho)$, $0 \le \rho \le t$ (8)

记录这条最优路径;则状态序列 $Q(t, s, \rho)$ 的第 ρ 个状态到第 t 个状态都是 s.

定义 $\delta(t, s) = \max_{\rho} D(t, s, \rho)$ (9)

$$\Phi(t, s) = Q(t, s, \arg\max_{\rho} D(t, s, \rho)) \quad (10)$$

即达到 $\delta(t, s)$ 时的最优状态序列.

认为在条件 $q_1, \cdots, q_{t-2}, q_{t-1} = s', q_t = s, \rho_t = \rho, o_1 \cdots o_{t-1}$ 下 o_t 的分布完全取决于 $q_t = s, \rho_t = \rho, o_1 \cdots o_{t-1}$,其余条件 $q_1 \cdots q_{t-2}, q_{t-1} = s$ 已被隐含或不起作用. 假设

$$P(o_t | q_1 \cdots q_{t-2}, q_{t-1} = s', q_t = s, \rho_t = \rho, o_1 \cdots o_{t-1}) = P(o_t | q_t = s, \rho_t = \rho, o_1 \cdots, o_{t-1}), \quad s' \in \Omega(t, s) \quad (11)$$

可以得到如下递归公式:

$$D(t,s,\rho)=\begin{cases}\left[\max\limits_{s'\in\Omega(t,s)\atop s'\neq s}\delta(t-1,s')\right]\\ \qquad\cdot P(o_t|\ q_t=s,\rho_t=\rho,o_1\cdots o_{t-1}),\rho=t,\\ D(t-1,s,\rho)\cdot P(o_t|\ q_t=s,\rho_t=\rho,o_1\cdots o_{t-1}),\\ \qquad 0\le\rho<t\end{cases}\tag{12}$$

$$Q(t,s,\rho)=\begin{cases}\Phi\left(t-1,\arg\max\limits_{s'\in\Omega(t,s)\atop s'\neq s}\delta(t-1,s')\right)\cup\{s\},\rho=t,\\ Q(t-1,s,\rho)\cup\{s\},\qquad 0\le\rho<t\end{cases}\tag{13}$$

其中 $\cup$ 表示状态级联运算.

最后得到,最优状态序列(也即最优状态分割点)为

$$Q^*=\Phi(T,\arg\max_s\delta(T,s))\tag{14}$$

3.3 剪枝算法

上述分割过程在每一个网格点 (t,s) 需保留多条路径 $Q(t,s,\rho)$,$0\le\rho<t$,从而更新路径所需概率计算项 $P(o_t|\ q_t=s,\rho_t=\rho,o_1\cdots o_{t-1})$ 的数目也成倍增加. 由于分割过程占据了训练和识别的大部分时间,分割算法的复杂度就约为训练和识别的复杂度. 因此从概率计算项的数目估计,上述分割算法的计算复杂度约为 $O(T^2)$,而基线模型的复杂度约为 $O(T)$. 其实在每一个网格点,我们可以只保留部分的 $\delta(t,s,\rho)$ ——路径剪枝,记这样的 ρ 集合为 $A_{t,s}$.

(Ⅰ)基于最大段长约束,定义 $A_{t,s}=\{\rho|\ t-\tau^s_{\max}<\rho\le t\}$,其中 $\tau^s_{\max}$ 表示状态 s 的最大可能段长值. 这相当于全路径搜索,称为 AllPath 方法. 计算复杂度约为 $O(\overline{\tau^s_{\max}}\cdot T)$,其中 $\overline{\tau^s_{\max}}$ 是 $\tau^s_{\max}$ 的平均值.

(Ⅱ)类似文[3]中方法,根据基线模型式(3)应用标准 Viterbi 分割,设进行到时刻 t 时,状态 s 的进入点为 $\eta_{t,s}\le t$,定义 $A_{t,s}=\{\rho|\ \eta_{t,s}-r\le\rho\le\min(\eta_{t,s}+r,t)\}$,即只保留分割点在与基线模型的松弛几帧的范围内的那些路径,称为 Around 方法. 计算复杂度约为 $O((2r+1)\cdot T)$.

(Ⅲ)定义 $A_{t,s}=\{\rho|\ \delta(t,s,\rho)$ 是前 M 得分,$\rho=t$ 或 $\rho\in A_{t-1,s}\}$,这里 $|A_{t,s}|=M$,称为 MBest 方法. 计算复杂度约为 $O((M+1)\cdot T)$.

AllPaht 方法的计算量很大,但差不多可以给出多项式拟合语音轨迹模型的一个最优性能. MBest 方法要好于 Around 方法,后面有实验和分析.

3.4 参数估值公式

下面给出训练中具体的参数重估公式. 已知一段训练语音 $O=o_1o_2\cdots o_T$ 和状态标注 $s_1\cdots s_N$. 经过分割过程得到当前参数 λ 下的最优状态分割 Q_λ:$1=P_1\le\cdots\le P_N\le P_{N+1}=T+1$. 定义 $\Gamma_s=\{n|\ s_n,n=1,\cdots,N\}$.

对数似然函数

$$\begin{aligned}L(\hat\lambda)&=\log P(O|\ Q_\lambda,\hat\lambda)=\sum_{n=1}^{N}\log P(o_{P_n}\cdots o_{P_{n+1}-1}|\ s_n,\hat\lambda)\\&=\sum_{s\in S}\sum_{n\in\Gamma_s}\log P(o_{P_n}\cdots o_{P_{n+1}-1}|\ s_n,\hat\lambda)\\&=\sum_s\sum_{n\in\Gamma_s}\sum_{t=P_n}^{P_{n+1}-1}\Big\{-\frac{D}{2}\log(2\pi)-\frac{1}{2}\log|\overline{\Sigma}_s|\\&\quad-\frac{1}{2}(o_t-\sum_{i=0}^{m}\overline{k}_i^s(t-P_n)^i)^{Tr}\overline{\Sigma}_s^{-1}(o_t\\&\quad-\sum_{i=0}^{m}\overline{k}_i^s(t-P_n)^i)\Big\}\end{aligned}$$

问题成为最优化求解 $\lambda=\arg\max_\lambda P(O|\ Q_\lambda,\hat\lambda)$,其中 $\hat\lambda=\{\overline{k}_i^s(i=0,\cdots,m),\overline{\Sigma}_s|\ s\in S\}$.

参数的新估值 λ 应使目标函数 $L(\hat\lambda)$ 对 $\hat\lambda$ 的导数为 0,从而 k_i^s,$i=0,\cdots,m$ 满足

$$\sum_{j=0}^{m}\Big(\sum_{n\in\Gamma_s}\sum_{t=P_n}^{P_{n+1}-1}(t-P_n)^{j+i}\Big)\cdot k_j^s=\sum_{n\in\Gamma_s}\sum_{t=P_n}^{P_{n+1}-1}(t-P_n)^i o_t\tag{15}$$

而 $\hat\Sigma_s=$

$$\frac{\sum_{n\in\Gamma_s}\sum_{t=P_n}^{P_{n+1}-1}(o_t-\sum_{i=0}^{m}k_i^s(t-P_n)^i)(o_t-\sum_{i=0}^{m}k_i^s(t-P_n)^i)^{Tr}}{\sum_{n\in\Gamma_s}\sum_{t=P_n}^{P_{n+1}-1}(1)}\tag{16}$$

更清楚一点,$m+1$ 个向量 k_i^s,$i=0,\cdots,m$ 的第 d 维 $k_i^{s(d)}$ 的求解归结为以下联立方程组:

$$\begin{pmatrix}ct_0&ct_1&\cdots&ct_m\\ct_1&ct_2&\cdots&ct_{m+1}\\\cdots&\cdots&\cdots&\cdots\\ct_m&ct_{m+1}&\cdots&ct_{2m}\end{pmatrix}\begin{pmatrix}k_0^{s(d)}\\k_1^{s(d)}\\\cdots\\k_m^{s(d)}\end{pmatrix}=\begin{pmatrix}\Omega_{t_0}^{(d)}\\\Omega_{t_1}^{(d)}\\\cdots\\\Omega_{t_m}^{(d)}\end{pmatrix}$$

,其中定义

$$ct_l=\sum_{n\in\Gamma_s}\sum_{t=P_n}^{P_{n+1}-1}(t-P_n)^l,\ l=0,\cdots,2m$$

$$\Omega_{t_i}=\sum_{n\in\Gamma_s}\sum_{t=P_n}^{P_{n+1}-1}(t-P_n)^i o_t,\ i=0,\cdots,m$$

它们就是最优状态分割下得到的统计量. 依上用这些统计量去更新模型参数.

4 实验结果及分析

实验所用数据来源于 863 提供的连续语音数据库. 使用了其中的男声部分,共 83 个文件,每个文件是一男声录音,每人约 520 句话. 以下各项实验均以其中 76 个文件作训练数据,另外 7 个文件(与训练集的说话人不同)作集外识别数据,基于此完成大词汇量非特定人连续语音识别的实验. 语音以帧长 20ms,帧叠 10ms 分帧处理求取 14 维 MFCC、1 维归一化能量以及它们的一阶和二阶差分,共 45 维特征. 基于汉字的 CV 结构,一个字表示为 2 个状态(辅音部分)和 4 个状态(元音部分)的级联,由此从句子标注得到状态标注.

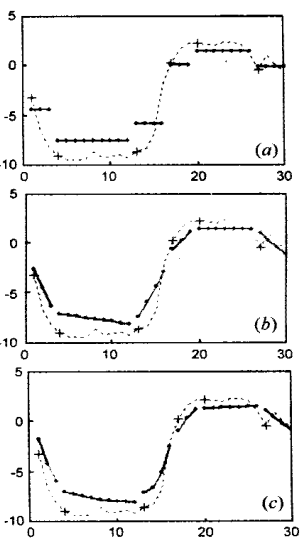

图 1 音节“shi4”的一段语音轨迹在基线模型.

我们先直观看一下语音

轨迹在基线模型式(3)和在多项式拟合模型式(5)中被描述的情况. 图 1 给出了从训练集中抽取出来的对应于音节"shi4"的一段语音轨迹,图示的是 14 维 MFCC 的第一维分量(MFCC1),横轴是时间,纵轴是 MFCC1 的值. 虚线是实际观察到的语音轨迹,根据训练基线模型时得到的分割点将这段轨迹分成 6 段("+"表示了每个状态段的开始点);实线给出了各模型在不同时刻的均值——时间的 0 阶(图 1(a))、1 阶(图 1(b))、2 阶(图 1(c))多项式函数. 可以对照看出,较图 1(a)而言,图 1(b)、(c)中语音轨迹得到更好的描述.

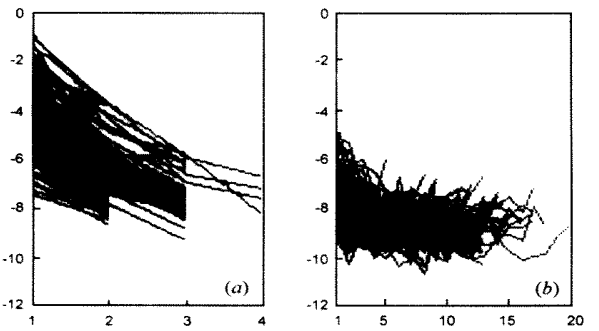

图 2　(a)"shi4"的辅音部分的 1th状态,2th状态;(b)对应的语音轨迹

为了更清楚地看出语音轨迹呈现的动态特性,根据训练基线模型时得到的分割点,对训练集中的每个人抽取出三段对应于"shi4"的辅音部分(2 个状态)的语音轨迹(MFCC1),这样得到非特定人条件下的多条(76 ×3 条)语音轨迹. 图 2(a)、(b)示出了对应状态的语音轨迹. 可以看出,时间的多项式函数(至少是斜线,而不是简单的平直线)更好地描述了这些轨迹的缓变趋势(动态特性),从而也能更好地识别语音.

表 1 给出了不同模型下声学层音节的识别结果. 可以看出:

(1)采用多项式拟合模型,在不同的多项式阶数(m = 1, 2)、剪枝方法下,都使识别性能较基线模型有一致提高.

(2)综合考虑计算量和识别性能,三种剪枝算法中,MBest 方法最好. 我们统计得到 $\overline{\tau_{\max}} = 15$,所以 AllPath 方法基本上不实用. Around 方法由于其分割点的取法限制(与基线模型相差无几),所以并没得到很好的识别效果. Around 方法在 $r = 1$、计算量约 $O(4T)$时的误识率,与同等计算量的 MBest 方法($M = 3$)比要差一些.

(3)如图 3 所示,采用 MBest 方法,随着 M 增大,误识率一致降低;当 $M \geq 3$ 后,下降趋势很缓. 在 $M = 4$ 时已经很接近甚少剪枝的 AllPath 方法所提供的性能.

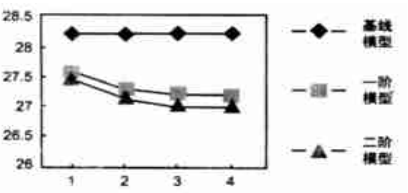

图 3　不同 M(横轴)下 MBest 方法的误识率(纵轴)

(4)二阶多项式拟合模型的性能,在相同剪枝条件下一致好于一阶多项式拟合模型.

总的来说,由于更好地描述了语音轨迹,多项式拟合语音轨迹模型明显改善了识别系统的性能. 二阶多项式模型配以实用化的 MBest 剪枝方法(例如 M = 3),使误识率较基线模型降低了 4.43 %.

5　总结

本文基于一个更为灵活的语音描述统计框架——广义 DDBHMM,提出了一个具体的多项式拟合语音轨迹模型,以及新的训练和识别算法,更好地刻划了真实的语音特性,明显改善了识别系统的性能. 同时提出的 MBest 剪枝方法以较少的计算量使上述模型得以实用.

表 1　不同模型下声学层的误识率(%)

	AllPath	MBest				Around
		M = 1	M = 2	M = 3	M = 4	r = 1
基线模型	28.24					
一阶模型	27.15	27.59	27.27	27.18	27.16	27.52
二阶模型	26.92	27.45	27.13	26.99	26.96	27.21

参考文献:

[1] M Ostendorf, S Roukos. A stochastic segment model for phoneme-based continuous speech recognition [J]. IEEE Trans, 1989, ASSP-37(12): 1857 - 1869.

[2] H Gish, K Ng. A segmental speech model with applications to word spotting [A]. Proceedings of ICASSP [C]. ICASSP, 1993. 447 - 450.

[3] L Deng, et al. Speech recognition using hidden Markov models with polynomial regression functions as nonstationary states [J]. IEEE Trans, 1994, SAP-2(4): 507 - 520.

[4] K K Paliwal. Use of temporal correlation between successive frames in a hidden Markov model based speech recognizer [A]. Proceedings of ICASSP [C]. ICASSP, 1993. 215 - 218.

[5] P C Woodland. Hidden markov models using vector linear prediction and discriminative distributions [A]. Proceedings of ICASSP [C]. 1992. 509 - 512.

[6] P Kenny, et al. A linear predictive HMM for vector-valued observation with application to speech recognition [J]. IEEE Trans, 1990, ASSP-38(2): 220 - 225.

[7] 王作英,曹洪. 语音识别的改进隐含马尔可夫模型 [A]. 863 智能计算机系统主题学术会议 [C]. 1988.

作者简介:

欧智坚　男,1975 年 10 月出生于湖南省耒阳市,1998 年毕业于上海交通大学电子工程系,获学士学位,同年免试进入清华大学电子工程系硕博连读,研究方向为语音信号处理.

王作英　男,1935 年出生于江西省赣县,1959 年毕业于清华大学无线电电子学系,1963 年毕业于苏联莫斯科鲍曼高等工业学校制造系,获博士学位,自 1963 年至今在清华大学电子工程系任教,现为该系教授,博士生导师,中国通信学会通信理论委员会副主任,获国务院特殊津贴专家,研究领域为信号和信息处理,近年来主要从事语音信号处理研究,主持和参加国家 863 高科技项目语音识别的研究.