从线性预测 HMM 到一种新的语音识别的混合模型

欧智坚,王作英
(清华大学电子工程系,北京 100084)

摘 要: 线性预测 HMM(Linear Prediction HMM,LPHMM)并没有象传统 HMM 那样引入状态输出独立同分布假设,但实用中识别性能并不佳.通过分析两种 HMM 的各自优劣,本文提出了一种新的语音识别的混合模型,将语音静态特性(基于传统 HMM)和动态特性(基于 LPHMM)分别描述又有机结合在一起,更为精确地刻划了真实的语音现象,同时又继承使系统的实现改动很小和较小的计算量.汉语大词汇量非特定人连续语音识别的实验表明,混合模型的识别性能显著好于 LPHMM 和传统 HMM.理论上,本文还给出了 LPHMM 的一组闭式参数重估公式.

关键词: 连续语音识别;隐马尔可夫模型;线性预测隐马尔可夫模型
中图分类号: TN912 **文献标识码:** A **文章编号:** 0372-2112 (2002) 09-1313-04

From Linear Prediction HMM to a New Combined Model for Speech Recognition

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

Abstract: *Linear prediction* (LP) HMM does not make the independent and identical distribution (IID) assumption as traditional HMM; however it often produces unsatisfactory results in practice. In this paper, a new combined model for speech recognition is proposed, based on a new analysis of both HMMs' modeling strengths and weaknesses. The new model works with LPHMM as the dynamic part and traditional IID-based HMM as the static part; in addition, easy implementation and low cost are preserved. Experiments on speaker-independent continuous speech recognition demonstrated that the combined model performed much better than both LPHMM and traditional HMM. Theoretically, a new closed-form parameter re-estimation formula is suggested for training LPHMM.

Key words: continuous speech recognition; hidden markov model(HMM); linear prediction HMM

1 引言

尽管作为当前最为流行的语音识别模型,HMM 仍引入了某些与语音实际特性不够协调的假设,阻碍了识别系统性能的进一步提高.传统 HMM 假设状态输出独立同分布(Independent and Identical Distribution,IID),忽略了对语音固有动态特性的描述.在特征中加入一、二阶倒谱与能量的差分已有效地得到广泛应用,正因为它使当前语音帧包含了前后若干帧的动态信息.进一步,各种改进模型[1]被提出以在 HMM 中引入对语音动态特性的描述.但是当应用于大词汇量连续语音识别中,它们受困于急剧增加的运算量,实用中往往采取次优的多遍搜索策略,识别性能并不佳.

值得注意的是,如果将当前帧的概率分布依赖于周围帧的变化情况以线性预测的方式表示,这些方法[2-6]较其它的模型假设更有吸引力,基本上不增加运算量也更直观些.它们都能归纳到我们后面介绍的更一般的线性预测 HMM(Linear Prediction HMM, LPHMM)中.在 LPHMM 中,识别过程中的 Viterbi 搜索与训练过程中的 Viterbi 对齐,甚至是前向-后向算法仍然适用,只要将现有的状态输出分布 $p(o_t)$ 替换以一个帧间相关的输出分布 $p(o_t|o_{t-1},\cdots)$.应该说,利用 LPHMM 是一种直观的考虑帧间相关性的想法.早期的研究出现在文献[2](其中没有实验结果的报道)和[3](给出了糟糕的实验结果).文献[4]中"惊奇地"(surprisingly)发现,LPHMM 只在使用简单倒谱特征时有效,如果加上差分,性能会变差;并且"是非而是地"(paradoxically),LPHMM 识别性能变坏,却得到比传统 HMM 高得多的似然值.我们先前应用 LPHMM 于大词汇量汉语连续语音识别中也有类似的结果,单就 LPHMM 描述语音特征并不充分,效果不佳.当结合使用"鉴别输出分布",LPHMM 在英语 E 集字母识别中有效[5].文献[6]中报道了从 11.8% 到 11.4% 的微弱英语字误识率的下降.

使用 LPHMM,在理论推导上不会遇到太大的困难,但在实用中都没达到意想中的好效果,对其不一致性能缺乏理论上的分析和理解.本文从分析 LPHMM 入手,提出了一种新的语音识别的混合模型——一个能应用于大词汇量连续语音识别中的有效的实用化模型.LPHMM 中的相关输出分布 $p(o_t|o_{t-1},\cdots)$ 作为一条件概率,描述的是当前帧围绕着前后帧给出的预测值变化的情况,仅反映了语音的动态特性;另一方面,传统 HMM 基于状态输出独立同分布(IID),用状态均值矢量表示状态输出矢量在特征空间中的聚集中心位置,仅描

收稿日期:2002-03-20;修回日期:2002-06-20
基金项目:国家 863 计划基金(No.2001AA114071)

© 1995-2004 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

述了语音的静态特性.因此,如果将这两种模型描述的互补信息融合在一起,得到一个混合模型,则能更为精确地刻划真实的语音现象.文献[7]中,已经得到了初步验证.

本文将更全面地阐述上述混合模型以及实验结果,同时给出有关 *LPHMM* 的一个新的结果——参数训练中的一组闭式重估公式.

2 线性预测 HMM(LPHMM)

一般地,假设状态 S 对应的 D 维观测矢量可如下描述:

$$\boldsymbol{o}_t = \sum_{i=1}^{m} \beta_i^s \boldsymbol{O}_{t+l_i} + \mu_s + v_t, \tag{1}$$

其中 l_i 是与第 i 个预测子(predictor)相关的帧偏离,$\beta_i^s \in R^{D\times D}$ 是第 i 个预测子的预测矩阵,$\mu_s \in R^D$ 考虑了观测矢量的非零均值,$v_t \sim N(0, \sum_s)$ 为帧间不相关零均值高斯白噪声.上述表示方式使得预测子可以非常灵活地选择对当前帧进行预测的若干帧的偏离(前或后).选择不同的预测结构 $\{l_1, \cdots, l_m\}$ 就形成不同的 LPHMM 模型.

于是状态 s 对观测 $\boldsymbol{o}_t$ 的概率密度函数(Probability Density Function,PDF)为

$$\tilde{b}_s(\boldsymbol{o}_t) \triangleq p(\boldsymbol{o}_t \mid \boldsymbol{o}_{t+l_i}, i=1,\cdots,m, s) = \frac{1}{(2\pi)^{D/2}\left|\sum_s\right|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(w_t^s - \mu_s)^T \sum_s^{-1}(w_t^s - \mu_s)\right\}, \tag{2}$$

其中 $w_t^s = \boldsymbol{o}_t - \sum_{i=1}^{m} \beta_i^s \boldsymbol{o}_{t+l_i}$.LPHMM 中对一句话的概率计算与传统 HMM 中一样,只要将上述"帧间相关的输出 PDF" $\tilde{b}_s(o_t)$ 代替传统的"IID 的输出 PDF" $b_s(o_t) \sim N(m_s, \Lambda_s)$.使用 Viterbi 训练时,与 LPHMM 有关的模型参数的重估公式如下*.

根据标注 $s_1 \cdots s_N$ 对训练数据 $\boldsymbol{O} = o_1 o_2 \cdots o_T$ 进行 Viterbi 对齐,我们得到当前模型参数 $\lambda = \{\mu_s, \sum_s, \beta_i^s, i=1,\cdots,m\}$ 下的最大似然状态序列 Q_λ,以及对每个状态 s 而言的一个帧集合 Γ_s.Γ_s 包含了分割后对齐到状态 s 的那些帧.于是问题成为最大化如下对数似然值:

$$L(\hat{\lambda}) = \log P(O \mid Q_\lambda, \hat{\lambda}) = \sum_s \sum_{t\in\Gamma_s} \log \tilde{b}_s(o_t \mid \hat{\lambda}). \tag{3}$$

将目标函数 $L(\hat{\lambda})$ 对模型参数 $\hat{\lambda}$ 求 Gateaux 导数,得到如下估值公式:

$$\hat{\mu}_s = \frac{\Phi_0^s - \sum_{i=1}^{m} \hat{\beta}_i^s \Phi_i^s}{|\Gamma_s|} \tag{4}$$

$$\hat{\sum}_s = \frac{R_{00}^s - \sum_{i=1}^{m}(R_{0i}^s \hat{\beta}_i^{sT} + \hat{\beta}_i^s R_{0i}^{sT}) + \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\beta}_i^s R_{ij}^s \hat{\beta}_j^{sT}}{|\Gamma_s|} - \hat{\mu}_s \hat{\mu}_s^T \tag{5}$$

$$(\hat{\beta}_1^s, \cdots, \hat{\beta}_m^s) = (\hat{B}_1^s, \cdots, \hat{B}_m^s) \begin{pmatrix} R_{11}^s & \cdots & R_{1m}^s \\ \vdots & \ddots & \vdots \\ R_{m1}^s & \cdots & R_{mm}^s \end{pmatrix} \tag{6}$$

其中 $\Phi_i^s = \sum_{t\in\Gamma_s} O_{t+l_i}$, $R_{ij}^s = \sum_{t\in\Gamma_s} O_{t+l_i} O_{t+l_j}^T$, $\hat{B}_i^s = \sum_{t\in\Gamma_s}(O_t - \hat{\mu}_s) O_{t+l_i}^0$, $0 \leqslant i, j \leqslant m$,它们是最优状态分割下得到的有关状态的统计量.依上用这些统计量去更新模型参数.

3 混合模型

3.1 分析 LPHMM 和基于 IID 的传统 HMM

如果用另一种方法求解最大化式(3),我们能进一步理解 LPHMM.方便起见,设 $x_t = (o_t^T, o_{t+l_1}^T, \cdots, o_{t+l_m}^T)^T$——即时刻 t 处的扩展帧,$\theta_s = (I, -\beta_1^s, \cdots, -\beta_m^s)$,于是 $w_t^s = \theta_s x_t$,$L(\hat{\lambda})$ 可重写成

$$L(\hat{\lambda}) = -\sum_s \frac{|\Gamma_s|}{2}\{\log|\hat{\sum}_s| + \mathrm{trace}[\hat{\sum}_s^{-1}(\hat{\theta}_s C_x^s \hat{\theta}_s^T)] + (\hat{\theta}_s \eta_x^s - \hat{\mu}_s)\hat{\sum}_s^{-1}(\hat{\theta}_s \eta_x^s - \hat{\mu}_s)^T + D\log(2\pi)\} \tag{7}$$

其中 η_x^s, C_x^s 分别是在数据集 $\{x_t \mid t \in \Gamma_s\}$(即对齐到 s 的扩展帧)上计算的样本均值和样本协方差阵.

首先,固定 $\hat{\theta}_s$ 得到最大化 $L(\hat{\lambda})$ 时的 $\hat{\mu}_s = \hat{\theta}_s \eta_x^s$ 和 $\hat{\sum}_s = \hat{\theta}_s C_x^s \hat{\theta}_s^T$.然后,将上述 $\hat{\mu}_s$ 和 $\hat{\sum}_s$ 代入(7),得到只含 $\hat{\theta}_s$ 变量的对数似然值:

$$L(\{\hat{\theta}_s\}) = -\sum_s \frac{|\Gamma_s|}{2}\{\log|\hat{\theta}_s C_x^s \hat{\theta}_s^T| + D\log(2\pi) + D\} \tag{8}$$

为估计预测矩阵,需最大化似然函数(8),等价地,最小化 $|\hat{\theta}_s C_x^s \hat{\theta}_s^T|$.注意到 $|\hat{\theta}_s C_x^s \hat{\theta}_s^T|$ 是数据 $o_t - \sum_{i=1}^{m} \beta_i^s o_{t+l_i}$ 的样本协方差阵的行列式值.而一个随机变量的样本协方差阵的行列式值是对其紧凑分布程度的一个很好的度量,于是最小化 $|\hat{\theta}_s C_x^s \hat{\theta}_s^T|$,就是去找这样的 β_i^s,使得 o_t 最紧凑地条件分布于它周围帧给出的范围内,即围绕着 $\sum_{i=1}^{m} \beta_i^s o_{t+l_i}$ 分布.通过这种方式,状态 s 输出帧的动态特性在 LPHMM 中通过"帧间相关的输出 PDF"得到了很好的刻划.另一方面,基于 IID 的传统 HMM 仍在实用中表现出有效的识别性能,这归于它在一定程度上很好地描述了语音的静态特性.状态输出特征矢量很好地静态(无条件)散布在以输出分布 PDF $b_s(o_t)$ 的均值矢量为中心位置周围.

问题在于,在将特征 o_t 分类到状态时,单依据 $\tilde{b}_s(o_t)$ 计算的似然得分进行判决是不够的,如果不考虑由 $b_s(o_t)$ 计算的似然得分(见图 1);反之也对(见图 2).图示中给出了视 o_t 为一维时两个状态的情形.每个椭圆是 $p(o_t, o_{t-1} \mid s)$ 的等概率线,分别描述了状态 $s=1,2$ 的输出.沿 o_t 轴和斜线 l_1(垂直于 $o_t = \beta^1 o_{t-1}$)的高斯 PDF 曲线分别代表了 $\{b_1(o_t), b_2(o_t)\}$ 和 $\{\tilde{b}_1(o_t), \tilde{b}_2(o_t)\}$(为清楚起见,后者沿 l_1 线放在了一起).两条高斯 PDF 曲线的重叠区域就给出了分类错误率.单独依赖 $b_s(o_t)$ 或 $\tilde{b}_s(o_t)$ 进行统计判决分别有错误率 Err_s 和 Err_d.当 $Err_s < Err_d$,由 $b_s(o_t)$ 代表的语音静态特性的分布在分类语音帧时就比由 $\tilde{b}_s(o_t)$ 代表的语音动态特性的分布更有鉴别力;反之有类似的结论.

---

* 注意到,上述定义的 LPHMM 其实只对状态输出分布作了假设,其中状态转移分布既可以是齐次的(如传统 HMM 那样使用转移矩阵)又可以是非齐次的(如 DDBHMM[1]那样使用段长分布);有关这部分参数的估值沿用原有的做法.下面仅描述与状态输出分布有关的参数估计.

© 1995-2004 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

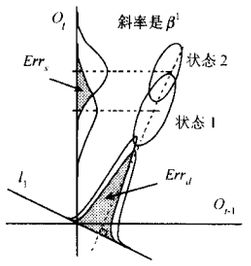


图 1　在 LPHMM 中两个状态得不到很好鉴别的情形，$Err_s < Err_d$

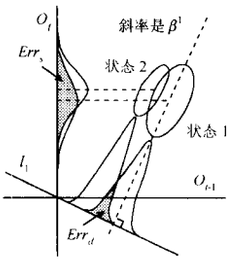


图 2　在基于 IID 的传统 HMM 中两个状态得不到很好鉴别的情形，$Err_d < Err_s$

3.2　结合 LPHMM 和基于 IID 的传统 HMM

由此，利用两种 HMM 对语音静态特性和动态特性的互补描述能力建立一个混合模型是更合理，也更有效的. 事实上，准平稳的静态特性和变化的动态特性在任何一段语音中都是混在一起的. 新的“混合输出 PDF”定义为[7]

$$\tilde{b}_s(o_t) = b_s(o_t)^{1-\alpha} \cdot \tilde{b}_s(o_t)^{\alpha}, \tag{9}$$

其中 α 是混合权重. 当 $=0,1$ 时，传统 HMM 和 LPHMM 就分别成为上述混合模型的一个特例.

上述混合模型从 LPHMM 直接继承了既容纳语音帧间相关性，又不带来过大计算量的一个非常好的性质：只要将 $\tilde{b}_s(o_t)$ 代替 $b_s(o_t)$，我们同样可以在上述混合模型中使用已有的训练和识别算法. 事实上，一旦得到每个状态的有关统计量，模型参数 $\{m_s, \Lambda_s\}$ 和 $\{\mu_s, \sum_s, \beta_i^s, i=1,\cdots,m\}$ 则分别进行估值. 因此，为实现上述混合模型，训练器和识别器只要作出很小的改动.

3.3　讨论

上述这种对数线性组合几种信息的方式在语音识别中不乏应用. 如声学模型与语言模型的组合；建立在不同特征参数集上的多码本混合[8]. 文献[9]中的一种 Bigram-Constrained HMM 其实是上述混合模型在 $\alpha=0.5, m=1, l_1=-1$ 并且所有状态的预测矩阵共享一个的情况下的特例. 与上述混合模型类似地，文献[10]中使用一种概率分布(Probability Distribution，PD)先验－后验结合的方式去改进传统 HMM；后验 PD 类似于我们前面的 $\tilde{b}_s(o_t)$ 但基于 ELP(Extended Logarithmic Pool)计算，先验 PD 即是 IID 的 $b_s(o_t)$. 这种结合也是因为发现单用后验 PD 来作识别并不够；如果是这样的话，则基于 ELP 计算后验 PD 的复杂步骤就显得多余了. 文献[10]中的实验仅基于离散 HMM 作出. 本文提出的混合模型就显得更简单灵活和有效些.

3.4　LPHMM 中的一组闭式参数重估公式

注意到参数重估式(6)的右边出现有待估的 $\hat{\mu}_s$，式(4,5,6)其实是一联立方程组. 实际上，通常将出现在右边的待估值(如 $\hat{\mu}_s$)替换以现有值(或者说上一遍迭代的值，如 μ_s)，然后应用式(6)先得出 $\hat{\beta}_i^s$，然后应用式(4,5). 这其实将 EM 算法中的 M 步(最大化)变成了分步的极大化. 下面我们给出一组新的闭式公式，它避免了上面这种轮换作法.

继续 3.1 中的分析，有

$$C_x^s = \frac{\sum_{t\in\Gamma_s} x_t x_t^T}{|\Gamma_s|} - \eta_x^s \eta_x^{sT} = \begin{pmatrix} V_{00}^s & V_{01}^s & \cdots & V_{0m}^s \\ V_{10}^s & V_{11}^s & \cdots & V_{1m}^s \\ \vdots & \vdots & \vdots & \vdots \\ V_{m0}^s & V_{m1}^s & \cdots & V_{mm}^s \end{pmatrix},$$

其中 $V_{ij}^s = \frac{R_{ij}^s}{|\Gamma_s|} - \frac{\Phi_i^s}{|\Gamma_s|}\frac{\Phi_j^{sT}}{|\Gamma_s|}$.

且有

$$\frac{\partial L(\{\hat{\theta}_s\})}{\partial(\hat{\theta}_1^s,\cdots,\hat{\beta}_m^s)} = 2(\hat{\theta}_s C_x^s \hat{\theta}_s^T)^{-1} \cdot \left(V_{01}^s - \sum_{i=1}^m \hat{\theta}_i^s V_{i1}^s, \cdots, V_{0m}^s - \sum_{i=1}^m \hat{\beta} V_{im}^s\right).$$

使上述导数为零，从而 $\hat{\beta}_i^s$ 满足

$$(\hat{\beta}_1^s, \hat{\beta}_2^s, \cdots, \hat{\beta}_m^s) = (V_{01}^s, V_{02}^s, \cdots, V_{0m}^s) \begin{pmatrix} V_{11}^s & V_{12}^s & \cdots & V_{1m}^s \\ V_{21}^s & V_{22}^s & \cdots & V_{2m}^s \\ \vdots & \vdots & \vdots & \vdots \\ V_{m1}^s & V_{m2}^s & \cdots & V_{mm}^s \end{pmatrix}^{-1}. \tag{10}$$

式(10)右边不再出现有待估值，所以先应用(10)然后(4,5)就得到 LPHMM 模型参数重估公式的闭式解.

表 1　特征矢量分别是 15,30 和 45 维时基于 IID 的传统 HMM，LPHMM 以及混合模型的误识率(%)

维数/模型		{−5}	{−4}	{−3}	{−2}	{−1}	{+1}	{+2}	{+3}	{+4}	{+5}	{−2,+2}	{−3,+3}	{−4,+4}	{−5,+5}
15 维	IID	59.20													
	LP	46.97	45.50	45.28	49.10	63.44	65.15	53.92	45.87	43.51	43.67	86.87	61.43	48.08	44.45
	混合	49.17	47.24	45.01	43.48	43.75	45.13	43.37	44.06	45.04	46.86	41.87	39.67	39.92	41.40
30 维	IID	29.59													
	LP	28.21	28.21	29.85	33.90	44.11	43.43	34.11	29.70	28.20	28.11	58.88	35.42	29.81	28.87
	混合	27.54	27.21	27.49	27.86	28.59	28.61	27.74	27.27	27.19	27.50	28.51	26.83	25.86	26.16
45	IID	26.30													

4　实验结果及分析

实验所用语音数据来源于 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，每个文件是一个男声录音，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. 采用半音节模型(CV 结构)，基本语音单元为 100 个由 2 个状态组成的辅音单元，164 个由 4 个状态组成的元音单元，以及一个由一个状态组成的静音单元. 以下识别实验给出声学层[11]音节的识别结果. 考虑到训练时充分估计全预测矩阵对数据量的要求过高，同时也注意到相邻帧主要在特征矢量的相同维间表现出相关性，我们下面假定诸 β_i^s 是对角阵.

语音以 20ms 帧长，10ms 帧叠分帧处理求取 14 维 MFCC、1

© 1995-2004 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

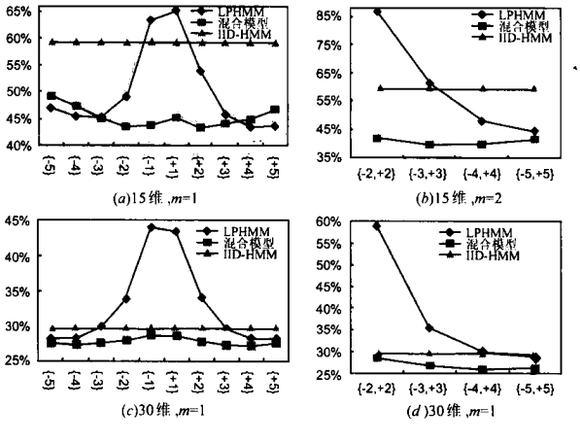


图3 特征矢量分别是15维(3.a)(3.b)和30维(3.c)(3.d)时各种模型随不同预测结构$\{l_1,\cdots,l_m\}$(横轴)的误识率(纵轴)

维归一化能量,计15维特征;加上一阶差分形成30维特征,再加上二阶差分形成45维特征.我们在特征矢量的不同维数下进行了多组实验.表1给出了特征矢量分别是15,30和45维时各种模型的误识率.由于预测结构$\{l_1,\cdots,l_m\}$的选择不同,我们又区分得到多种不同的LPHMM和混合模型.比如说,$\{-4,+4\}$就表示$m=2,l_1=-4,l_2=+4$的LPHMM.图3示出了不同预测结构$\{l_1,\cdots,l_m\}$对误识率的影响.在15维时有(3.a)(3.b),在30维时有(3.c)(3.d).横轴表示各种模型的不同预测结构$\{l_1,\cdots,l_m\}$.由于IID-HMM(即传统HMM)与$\{l_1,\cdots,l_m\}$无关,故其误识率曲线是一平直线.我们取$\alpha=0.5$.总得说来,我们有:

(1)LPHMM的性能与预测结构$\{l_1,\cdots,l_m\}$的选择有很大关系.在15维时使用预测结构$\{-1\}$或$\{+1\}$,30维时使用$\{-2\}$,$\{-1\}$,$\{+1\}$或$\{+2\}$,LPHMM的性能很差.这是由于帧叠处理的影响,使得相邻帧有人为引入的强相关性,而这并不提供鉴别力.

(2)混合模型的性能比LPHMM和传统HMM都要好,说明这种互补描述的确很大地改进了系统性能.一个情况是,15维时使用$\{-5\}$,$\{-4\}$,$\{+4\}$或$\{+5\}$的LPHMM表现出很好甚至超过相应混合模型的性能.这是由于我们在实验中固定使用$\alpha=0.5$,此时我们应调节以增大LPHMM部分的作用.

(3)使用15维特征时,混合模型能得到的最好性能是39.67%的误识率($\{-3,+3\}$时),较传统HMM下降了$\frac{59.20-39.67}{59.20}\approx33\%$.使用30维特征时,混合模型能得到的最好性能是25.86%的误识率($\{-4,+4\}$时),较传统HMM下降了$\frac{29.59-25.86}{29.59}\approx13\%$;并且已经好于45维时传统HMM的性能(26.30%的误识率),而存储量和运算量都较45维时下降了约$1/9\approx11\%$.

5 总结

通过分析LPHMM和基于IID的传统HMM的各自优劣,本文提出了一种新的语音识别的混合模型,同时克服了传统HMM不合理的IID假设以及LPHMM缺乏对语音静态特性描述从而在实用中性能不佳的缺点,又基本上使新系统的实现改动很小和较小计算量.汉语大词汇量非特定人连续语音识别的实验表明,混合模型的性能显著好于LPHMM和传统HMM.理论上我们还给出了LPHMM参数训练中的一组闭式重估公式.进一步,预测矩阵在不同状态间共享既能减少运算量又能部分克服最大似然训练的鉴别力不够;混合权重α与状态相关以及优化得到,都是有益的值得研究的方向.

参考文献:

[1] M Ostendorf, et al. From HMM's to segment models: a unified view of stachastic modeling for speech recognition [J]. IEEE Trans, 1996, SAP-4(5):360-378.

[2] C J Wellenkens. Explicit correlation in hidden Markov model for speech recognition [A]. Proceedings of ICASSP [C]. USA:1987.384-386.

[3] P F Brown. The acoustic modeling problem in automatic speech recognition [R]. USA: IBM Tech Report, No. RC 12750, 1987.

[4] 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.

[5] P C Woodland. Hidden Markov models using vector linear prediction and discriminative distributions [A]. Proceedings of ICASSP [C]. USA:1992.509-512.

[6] Y Jia, J Li. Relax frame independence assumption for standard HMMs by state dependent auto-regressive feature models [A]. Proceedings of ICASSP [C]. USA:2001.485-488.

[7] Zhijian Ou, Zuoying Wang. A new combined model of statics-dynamics of speech [A]. Proceedings of ICASSP [C]. USA:2002.965-968.

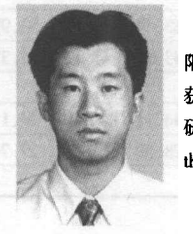
[8] Y Normandin, et al. High-performance connected digit recognition using maximum mutual information estimation [J]. IEEE Trans, 1994, SAP-2(2):299-311.

[9] S Takahashi, et al. Phoneme HMM's constrained by frame correlations [A]. Proceedings of ICASSP [C]. USA:1993.219-222.

[10] N S Kim, et al. Frame-correlated hidden Markov model based on extended logarithmic pool [J]. IEEE Trans, 1997, SAP-5(2):149-160.

[11] 王作英.基于段长分布的HMM语音识别模型 [A].第二届全国汉字、汉语识别会议论文集 [C].中国:1989.

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.现为清华大学电子工程系硕博连读研究生,研究方向为语音信号处理.邮件:ozj@thsp.ee.tsinghua.edu.cn

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.

© 1995-2004 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.