鲁棒性话者辨识中的一种改进的马尔科夫模型

刘 鸣,戴蓓倩,李 辉,陆 伟,李霄寒
(中国科学技术大学电子科学与技术系,安徽合肥 230026)

摘 要: 为了提高话者识别系统的噪声鲁棒性,本文对 CHMM 进行了改进,将每帧特征参数之间的差分参数来对应状态之间的转移,从而使帧间信息在模型中得到了体现. 利用改进后的 CHMM 模型对不同的特征参数携带的信息进行信息融合. 使得在强噪环境下,鲁棒性好的特征参数起主导作用,而在噪声比较小的环境下,精细度高的特征参数起主导作用. 实验证明,这种改进的马尔可夫模型明显提高语音识别系统的鲁棒性能,这种技术具有良好的发展和应用前景.

关键词: 连续隐马尔可夫模型; 信息融合; 鲁棒性; 话者识别
中图分类号: TN912.34 **文献标识码**: A **文章编号**: 0372-2112 (2002) 01-0046-03

An Improved HMM for Robust Speaker Recognition

LIU Ming, DAI Bei-qian, LI Hui, LU Wei, LI Xiao-han
(Dept. of Electronic Science & Technology, University of Science & Technology of China, Hefei, Anhui 230026, China)

Abstract: A modified CHMM model is proposed to improve the robustness of the speaker recognition system. In the modified model, the difference of the speech feature parameter was the observed vectors of the state-transitions of CHMM. Because of the modification, the dependence between frames was used in the modified CHMM. By the fusion of different features, the robust feature will dominate under the noisy environment and the delicate feature will dominate under clean environment. The experiments indicate that the modified CHMM have effectively improved the robustness of the speaker recognition system.

Key words: CHMM; information fusion; robustness; speaker recognition

1 引言

提高语音识别系统的鲁棒性是语音识别技术走向实用的关键问题,因而如何提高系统的鲁棒性正成为语音识别的研究热点. 语音识别系统的鲁棒性主要包括对环境的鲁棒性和对说话人的鲁棒性两个方面,本文主要针对第一类鲁棒性进行了一定的研究. 语音识别系统的环境鲁棒性的实质是:当实际应用环境与系统的训练环境失配时,语音识别系统的性能会显著下降.

目前,鲁棒性语音识别通常是将多种特征参数结合起来使用,例如,将 MFCC 特征参数和它的差分型特征参数(ΔMFCC)构成一个大的特征矢量. 模型一般都使用各种类型的 HMM 模型,因为 HMM 模型能较好地刻画语音信号中的时序信息. 但是 HMM 也有它的不足之处,例如,它认为各帧矢量之间是独立同分布的,这就与实际情况不符,另外,在描述 HMM 模型的参数 $\lambda = (\pi, A, B)$ 中,A 矩阵是各状态之间的转移概率矩阵,它在 HMM 模型中被看成静态的,但我们认为随着模型在一个状态上停留时间的增加,模型应该更倾向于向后面的状态跳转而不是继续停留在这个状态上. 因此,A 矩阵应该随着计算帧数的增加而不断被修正,即它应该是动态的. 对此,本文提出了一种改进的 CHMM(连续隐马尔科夫模型),在这一改进的模型中,我们用特征参数的差分矢量来约束状态间的转移,即 A 修正矩阵,使得 A 矩阵具有动态特性. 从另一角度看,是利用状态和状态转移弧来综合特征参数和差分型特征参数的信息,使得他们之间的权重能随着噪声的变化而自适应地调整.

2 MFCC 和 ΔMFCC 的抗噪性比较

MFCC 参数是目前应用最为广泛的特征参数. 其特点是,在高信噪比的条件下,MFCC 特征参数具有很好的识别率,但在信噪比低的时候,识别性能很差. 而 ΔMFCC 参数则在低信噪比时,能有较好的识别率,但在高信噪比时识别率不如 MFCC. 目前鲁棒性话者识别中,一般是将 MFCC 和 ΔMFCC 两种参数构成一个大的特征参数,我们用 MFCC(18)和 ΔMFCC(18)构成一个大特征参数 MFCC + ΔMFCC (36),模型级采用 CHMM 来进行与文本有关的话者辨识实验. 结果如图 1 所示,可以看出,这种组合的特征参数的性能是组成它的两种参数的折中. 为了能获得更好的性能,我们给两种子特征参数加上不同的权重,来构成组合参数. 用两个不同的文本进行话者识

收稿日期:2000-08-23;修回日期:2001-04-23
基金项目:国家自然科学基金(No. 69872036)

别实验,在同一文本的话者实验系列中调整 ΔMFCC 参数与 MFCC 参数的比例,得到表 1 的结果.从中可以看出,在同一文本下,不同的比例的性能是不同的.而在不同文本的条件下,系统达到最佳性能的比例是不同的.(注:文本 1 和文本 2 识别环境信噪比为 10dB 和 15dB),$k = 1$ 可以看成次最佳的系数,以后的实验中均取 $k = 1$.

表 1　ΔMFCC 与 MFCC 比例对性能的影响

比例系数 k	0.2	0.4	0.6	0.8	1.0	1.4	1.8	2.4	3.8
文本 1	0.8241	0.8241	0.8241	0.8298	0.8298	0.8286	0.8286	0.8381	0.8381
文本 2	0.7845	0.7869	0.7855	0.7852	0.7821	0.7786	0.7762	0.7762	0.7798

3　利用改进的 CHMM 进行信息融合

为了充分发挥 MFCC 和 ΔMFCC 的特点,我们设计了一种改进的马尔科夫模型来综合这两种参数的信息。其主要思想是:由于 MFCC 在低噪声环境下性能更好,因此让它在低噪声

$$f(O|\mu,\Sigma)=\begin{cases}\dfrac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp(-(O-\mu)'\Sigma^{-1}(O-\mu)/2), & (O-\mu)'\Sigma^{-1}(O-\mu)\le k_1 n\\ \dfrac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp(-k_1 n), & k_1 n<(O-\mu)'\Sigma^{-1}(O-\mu)\le k_2 n\\ \dfrac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp(-k_2 n), & (O-\mu)'\Sigma^{-1}(O-\mu)>k_2 n\end{cases}\quad(1)$$

式中 n 为矢量的维数,$(O-\mu)'\Sigma^{-1}(O-u)<k_1 n$ 为一个超球,我们将这种概率密度函数叫做球域概率密度函数.k_1,k_2 为设定超球范围的参数,实验中我们取 $k_1=1$,$k_2=3$.在这种概率密度函数下,当参数矢量与均值差异较大时,函数给出一个缺省的输出,当参数矢量和均值的相似度合适时,才给出高斯型函数的输出.

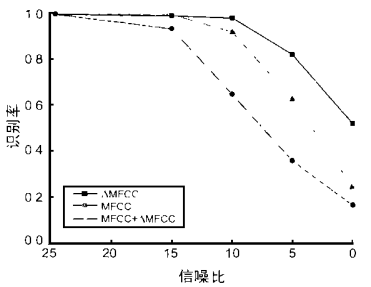

图 1　MFCC + DeltaMFCC 特征参数的性能

3.2　改进的 CHMM 模型(Delta HMM 模型)

我们知道,以 MFCC 为特征参数的 CHMM 模型中,每个状态都用一个混合的高斯密度函数来描述该状态输出的观察矢量的分布.由于 ΔMFCC 是通过 MFCC 进行差分得到的,我们可以认为 ΔMFCC 特征参数对应着状态之间的转移.因此,我们对 CHMM 模型进行了修改,将状态转移弧也看成一种随机过程,用相应的概率密度函数来和 ΔMFCC 特征参数联系.也就是说,用 ΔMFCC 来给出发生某个状态转移的置信度.这样当 MFCC 参数受到噪声干扰而偏离无噪情况下训练得到的均值较大时,由于超出了状态概率密度函数的有效区域(球域),状态的概率密度函数只给出一个缺省值,这时抗噪性好的 ΔMFCC 将起主导的作用,由它来确定当前的状态转移路径.这样,整个模型就能在强噪环境下获得接近或超过 ΔMFCC 在的环境下,发挥的作用大一些,而在强噪声环境下,MFCC 受的干扰大,让它作用的比重降低.对于 ΔMFCC 则相反.

3.1　对 CHMM 模型的概率函数的修正

CHMM 模型的一个不足是它的状态输出观察矢量为混合型高斯概率密度函数,为什么说这一点是它的不足呢?我们知道,在强噪声环境下,MFCC 的识别率变得很低。其原因是,在强噪声的干扰下 MFCC 的分布与无噪情况下有很大的差异,本来是该状态输出的特征矢量,但由于噪声的作用与本状态相差较远,此时密度函数将给出很低的输出,从而造成误判.所以我们认为应对状态输出密度函数进行修正,使得它只对足够相似的参数矢量给出密度函数的输出,而对相差较大的参数矢量则给一个缺省值,即密度函数不对这样的参数矢量给予评价(函数值的大小),而留给系统用其他类型的参数矢量来评价.因此我们定义了如下的概率密度函数:

相同环境下的识别性能.而在低噪环境下则由这两种参数共同确定模型的状态转移路径.这样在高信噪比时,MFCC 参数由于能和状态的密度函数较好地吻合而起主导的作用.

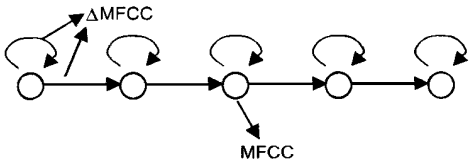

图 2　改进的 HMM 模型示意图

改进的 CHMM 如图 2 所示,每个圆圈代表 HMM 中的一个状态,每个带箭头的线段表示状态的转移.每个状态的观察矢量为 MFCC 参数,每条转移弧的观察矢量为 ΔMFCC 参数.

3.3　Delta HMM 模型的实现

模型的观察值序列记为 $O=O_1,O_2,\cdots,O_T$ 和由此序列得到的差分序列 $\Delta=\Delta_1,\Delta_2,\cdots,\Delta_{T-1}$;模型参数为 $\lambda=(\pi,A,B,D)$,其中 π 为初始概率矢量,A 为转移概率矩阵,B 为每个状态输出的观察值 O_t 的概率密度函数组成的函数系,D 为每个状态转移弧输出的观察值 Δ_t 的概率密度函数组成的函数系.由模型 λ 产生出 O 的概率密度为 $p(O/\lambda)$.对一个固定的状态序列 $S=q_1,q_2,\cdots,q_T$,有

$$p(O/S,\lambda)=\prod_{t=1}^{T}p(O_t/q_t,\lambda)=b_{q_1}(O_1)\,b_{q_2}(O_2)\cdots b_{q_T}(O_T)\quad(2)$$

其中 $b_{q_t}(O_t)=b_j(X)|q_t=\theta_j$,$X=O_t$,$1\le t\le T$,$1\le j\le N$ 为 q_t 状态的观察值的概率密度函数在观察矢量 O_t 上的输出.

而对给定 λ 产生状态序列 S 的概率密度为

$$p(S/\lambda)=\pi_{q_1}a_{q_1q_2}d_{q_1q_2}(\Delta_1)\,a_{q_2q_3}d_{q_2q_3}(\Delta_2)\cdots a_{q_{T-1}q_T}d_{q_{T-1}q_T}(\Delta_{T-1})\quad(3)$$

其中 π_{q_1} 为初始状态的概率,$a_{q_1q_2}$ 为状态 q_1 转移到状态 q_2 的

概率, $d_{q_1 q_2}(\Delta_1)$ 为 $q_1 q_2$ 这条状态转移弧的观察值的概率密度函数在观察矢量 Δ_1(由 O_1 和 O_2 差分得到)上的输出. 从公式(3),我们可以看出这种改进的本质,每个 $a_{q_t q_{t+1}}$ 都要和 $d_{q_t q_{t+1}}(\Delta_t)$ 相乘,实际计算中每帧的 A 矩阵中的 $a_{q_t q_{t+1}}$ 都要被 $d_{q_t q_{t+1}}(\Delta_t)$ 修正,从而给 A 矩阵赋予了动态特性.

改进模型的 Baum-Welch 重估

$$\xi_t(i,j) = P(O, q_t = \theta_i, q_{t+1} = \theta_j / \lambda) \tag{4}$$

可以推导出:

$$\xi_t(i,j) = \alpha_t(i)\, a_{ij} d_{ij}(\Delta_t)\, b_j(O_{t+1})\, \beta_{t+1}(j) \tag{5}$$

那么,时刻 t 时 Markov 链处于 θ_i 状态的概率为:

$$\xi_t(i) = P(O, q_t = \theta_i / \lambda) = \sum_{j=1}^{N} \xi_t(i,j) = \alpha_t(i)\beta_t(i) \tag{6}$$

因此, $\sum_{t=1}^{T-1} \xi_t(i)$ 表示从 θ_i 状态转移出去的次数的期望值,而 $\sum_{t=1}^{T-1} \xi_t(i,j)$ 表示从 θ_i 状态转移到 θ_j 状态的次数的期望值. 而每个状态均由 M 个概率密度函数混合而成,所以

$$b_j(O_t) = \sum_{k=1}^{M} c_{jk} N(O_t, \mu_{jk}, \Sigma_{jk}) \tag{7}$$

其中每个概率密度数所占比重为

$$\gamma_t(j,k) = \xi_t(j) \frac{c_{jk} N(O_t, \mu_{jk}, \Sigma_{jk})}{\sum_{m=1}^{M} c_{jm} N(O_t, \mu_{jk}, \Sigma_{jk})} \tag{8}$$

同理,每个状态转移弧的概率密度函数也是由多个概率密度函数混合而成

$$d_{ij}(\Delta_t) = \sum_{k=1}^{M} e_{ijk} N(\Delta_t, m_{ijk}, U_{ijk}) \tag{9}$$

其中每个概率密度数所占比重为

$$\rho_t(i,j,k) = \xi_t(i,j) \frac{e_{ijk} N(\Delta_t, m_{ijk}, U_{ijk})}{\sum_{m=1}^{M} e_{ijm} N(O_t, m_{ijk}, U_{ijk})} \tag{10}$$

DeltaHMM 的 Baum_Welch 重估公式:

$$\bar{\pi}_i = \xi_1(i) \quad \bar{a}_{ij} = \sum_{t=1}^{T} \xi_t(i,j) / \sum_{t=1}^{T} \xi_t(i) \quad \bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \xi_t(j)}$$

$$\bar{\mu}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)\, O_t}{\sum_{t=1}^{T} \gamma_t(j,k)} \quad \bar{\Sigma}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)\,(O_t - \bar{\mu}_{jk})(O_t - \bar{\mu}_{jk})^T}{\sum_{t=1}^{T} \gamma_t(j,k)}$$

$$\bar{e}_{ijk} = \frac{\sum_{t=1}^{T} \rho_t(i,j,k)}{\sum_{t=1}^{T} \xi_t(i,j)} \quad \bar{m}_{ijk} = \frac{\sum_{t=1}^{T} \rho_t(i,j,k)\, O_t}{\sum_{t=1}^{T} \rho_t(i,j,k)}$$

$$\bar{U}_{ijk} = \frac{\sum_{t=1}^{T} \rho_t(i,j,k)\,(\Delta_t - \bar{m}_{ijk})(\Delta_t - \bar{m}_{ijk})^T}{\sum_{t=1}^{T} \rho_t(i,j,k)} \tag{11}$$

4 利用 Delta HMM 进行信息融合的话者辨识实验

实验采用的语音库为 14 个男性话者的 14 * 100 次口令词发音,训练集为 20,测试集为 80. 对照的 CHMM 模型采用 5 状态从左到右无跳转马尔科夫链,每个状态采用 3 混合度的概率密度函数. DeltaHMM 模型也是采用 5 状态从左到右无跳转马尔科夫链,每个状态采用 3 混合度的概率密度函数,每条弧则用一个高斯概率密度函数来拟合. 模型训练是在无噪环境下进行的,识别则在不同强度加性噪声的干扰下完成.

表 2 DeltaHMM 和 CHMM 在 MFCC 和 ΔMFCC 的不同组合的抗噪性能的比较

特征参数	25dB	20dB	15dB	10dB	5dB	0dB
DeltaHMM + MFCC + ΔMFCC	0.998	0.998	0.998	0.9976	0.9262	0.5905
CHMM + ΔMFCC	0.996	0.996	0.99	0.981	0.8214	0.5214
CHMM + MFCC + ΔMFCC	0.998	0.998	0.9333	0.919	0.6262	0.2452
CHMM + MFCC	0.998	0.996	0.996	0.6476	0.3571	0.1643

注:以 MFCC(18) 和 ΔMFCC(18) 为特征参数的不同组合进行对比实验

实验结果如表 2 所示,可以看出,在强噪环境下 DeltaHMM 的识别性能优于单独用 CHMM + ΔMFCC 的系统,而在无噪环境下,DeltaHMM 也有与 CHMM + MFCC 相当的识别率. 但是,DeltaHMM 的计算量比 CHMM + MFCC 要大,实验中 CHMM + MFCC 作一次识别的时间为 1.357s,而 DeltaHMM 则需 3.217s.

5 结束语

如何提高系统的噪声鲁棒性一直是自动语音识别领域里的研究热点,由于噪声的复杂多变,没有通用的去噪方法,因此,我们提出了在模型级进行多种参数的融合的技术,较之传统方法进一步提高了系统的噪声鲁棒性,又不损失系统在无噪情况下的性能. 缺点是计算量比一般的 CHMM 要大.

参考文献:

[1] Kuo-Hwei You, Hsiao-Chuan Wang. Robust features derived from temporal trajectory filtering for speech recognition under the corruption of additive and convolution noises [A]. IEEE ICASSP, 98 [C]. 577 - 580.

[2] P Hanna, J Ming, F J Smith. Inter-frame dependence arising from preceding and succeeding frames-Application to speech recognition [J]. Speech Communication, 1999, 28(4): 301 - 312.

[3] R Schwartz, O Kimball, F Kubala, M Feng, Y Chow, C Barry, J Makhoul. Robust smoothing methods for discrete hidden Markov models [A]. Proc. ICASSP, 89 [C]: 548 - 551.

作者简介:

刘 鸣 男. 1976 年生于江西吉安. 中国科学技术大学电子科学与技术系,硕士研究生. 主要研究方向:语音信号处理,模式识别与人工智能.

戴蓓倩 女. 1941 年生于上海. 教授,博士生导师. 中国科学技术大学电子科学与技术系. 主要研究方向:语音信号处理,图象处理,模式识别与人工智能.