第 45 卷 第 1 期 厦门大学学报(自然科学版) Vol. 45 No. 1
2006 年 1 月 Journal of Xiamen University (Natural Science) Jan. 2006

基于 HMM 与神经网络的声学模型研究

林坤辉[1],息晓静[1],周昌乐[2]
(1. 厦门大学软件学院,2. 厦门大学信息科学与技术学院,福建 厦门 361005)

摘要:神经网络能依靠权值进行长时间记忆和知识存储,但是对输入模式的瞬时相应的记忆能力比较差;而隐马尔科夫模型的短时记忆的能力比较强,但是假定的前提又与实际情况不符. 因此,采用 HMM 和 ANN 的混合模型来取双方之长,并在这种混合模型的基础上,对神经网络从结构设计、训练、到训练后期的结构调整进行了全程的优化;应用隐节点剪枝算法,并利用广义的 Hebb 规则重新确定网络的参数. 实验表明,这种混合模型在语音识别中取得了良好的效果.

关键词:HMM;ANN;隐节点剪枝算法;广义 Hebb 算法
中图分类号:TP183;TN912.34 **文献标识码**:A **文章编号**:0438-0479(2006)01-0044-03

语音识别主要是让机器准确地识别出语音的内容. 利用隐马尔科夫链(HMM)与人工神经网络(ANN)各自的优点,通过 HMM/ANN 混合模型处理语音识别有综合的优势. 本文就其中的各主要技术展开论述.

1 HMM 的建模假设及对识别系统的影响

HMM 存在着很多和语音信号的实际情况不相符合的先验假设和训练方面的缺陷. 主要表现在:

(1) 一阶马尔可夫模型假设:马尔可夫链在 n 时刻处于状态 q_i^n 的概率只与 $n-1$ 时刻所处的状态 q_i^{n-1} 有关,而与其以前的状态及声学矢量序列无关.

(2) 观察矢量帧之间独立性假设,每一时刻 HMM 只能考虑当前帧语音. 这些假设使 HMM 对协同发音建模困难,同时 HMM 方法不同于人脑对语音的处理理解方式,其自适应能力、鲁棒性都不理想. 因此必须在探索人脑机理的基础上,寻求新的途径.

2 ANN 用于语音识别的优势

ANN 采用非线性处理单元来模拟人脑神经元,用处理单元之间的可变连接强度来模拟神经元的突触行为,构成了一个大规模并行的非线性系统[1]. 神经网络技术以其自适应性、并行性、非线性、鲁棒性和学习特性而被广泛应用于语音识别领域. 神经网络尤其是 MLP 之所以在语音识别领域有吸引力还在于:

(1) 具有基于误差 BP 的极强的学习能力;

(2) 实现输入输出信号间的复杂映射;

(3) 并行结构提供了高速度和高可靠性;

(4) 具有自适应、自组织及联想等特征,特别适合于语音识别中的分类问题.

由于 HMM 的时序性强、神经网络的多输入可以考虑帧间相关性和分类能力强等方面的综合优势,采用 HMM/ANN 混合模型用于语音识别很合适.

3 改进的混合模型

3.1 特征参数的提取

靠特征提取从语音信号中提取出对语音识别有用的信息,并去除对语音识别无关紧要的冗余信息,获得影响语音识别的重要信息.

目前多采用的是 LPCC 特征. LPCC 系数主要是模拟人的发声模型,未考虑人耳的听觉特性. MFCC 参数比 LPC 倒谱系数更符合人耳的听觉特性,在有信道噪声和频谱失真情况下,能产生更高的识别精度[2]. 研究者由心理学实验得到了类似耳蜗作用的一组滤波器组,即 Mel 频率滤波器组. Mel 频率可表示为:

$$f_{\mathrm{Mel}} = 2\,595\log(1 + f/700)$$

对频率轴的不均匀划分是 MFCC 特征区别于前述的普通特征的最重要的特点. 将频率按照上式变化到 Mel 域后,Mel 带通滤波器组的中心频率是按照 Mel 频率刻度均匀排列的[3]. Mel 倒谱系数计算如下:

(1) 经信号进行分帧、预加重和汉明窗处理,然后进行短时傅立叶变换并得到其频谱.

(2) 求出频谱平方,即能量谱,并用 M 个 Mel 带通滤波器进行滤波. 由于每一个频带中分量的作用在人耳中是叠加的,因此将每个滤波器频带内的能量进

收稿日期:2005-06-22
基金项目:厦门大学 985 二期信息创新平台项目资助
作者简介:林坤辉(1961-),男,副教授.

行叠加,这时第 k 个滤波器输出功率谱 $x(k)$.

(3) 将每个滤波器的输出取对数,得到相应频带的对数功率谱;并进行反离散余弦变换,得到 L 个 MFCC 系数,一般 L 取 12～16 个左右. MFCC 系数为:

$$C_n = \sum_{k}^{M} \log x(k) \cos[\pi(k - 0.5) n / M],$$

$$n = 1, 2, \cdots, L.$$

(4) 将直接得到的 MFCC 特征作为静态特征,再将该静态特征做一阶和二阶差分,得到动态特征.

3.2　网络中隐节点数目的优化

网络结构的改进主要体现在找到最优的隐节点的数目.确定隐节点的数目需要在网络的正则特性和学习能力之间取得平衡[4].隐节点的确定方式如下:

(1)用迭代自组织数据分析方法得到训练数据的聚类中心的数目,再为属于不同类的一对聚类中心分配一个隐节点.由此估计出一个对于训练和训练后的剪枝都合适的隐节点的数目 N.

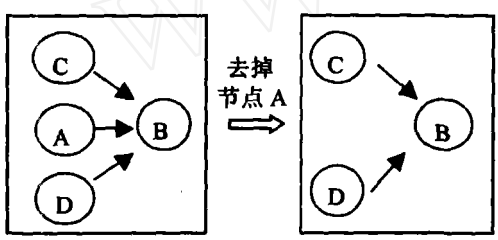

图 1　一种新的隐节点剪枝算法

Fig. 1　A novel algorithm for hidden nodes pruning

(2)训练 N 个节点的 MLP 网络.

(3)通过迭代去除网络中冗余隐节点.如图 1 所示,在移去节点 A 后,调整余下节点的权参数,使得节点 B 的净输入值在最小平方定义下近似保持不变,即对于训练集中所有的模式有:

$$\sum_{j \in \{A,C,D\}} w_{j,B} y_j(n) = \sum_{j \in \{C,D\}} (w_{j,B} + \delta_{j,B}) y_j(n),$$

$$\forall n \in \{1, 2, \cdots, N\},$$

其中 $w_{j,B}$ 代表节点 j 到 B 的权值,$\delta_{j,B}$ 代表节点 j 到 B 的残差,$y_j(n)$ 代表第 n 个节点的输出值,这就相当于一个线性方程:

$$\begin{pmatrix} y_C(1) & y_D(1) \\ M & M \\ y_C(N) & y_D(N) \end{pmatrix} \begin{pmatrix} \delta_{C,B} \\ \delta_{D,B} \end{pmatrix} = w_{A,B} \begin{pmatrix} y_A(1) \\ M \\ y_A(N) \end{pmatrix}.$$

用 LMS 迭代法可以得到这个线性方程在最小平方意义下的最优解 $(\delta^*_{C,B}, \delta^*_{D,B})$,为了保证输入输出关系,应该去掉使线性方程的残差达到最小的结点.因为用 LMS 算法解上述方程,残差是随着迭代次数单调减小的,所以只需要计算每个节点的初始残差就可以确定应该去掉哪个节点了.

3.3　改进网络权值参数的初始化

初始化问题是神经网络训练的一个很重要的问题,关系到训练结果的收敛[5,6].一般初始化的主要思想是通过考察网络的学习机制或是采用先验知识来尽可能优地进行网络权值初始化[7].我们用 Hebb 规则来进行初始化,取得了良好的效果.它的原理是:设隐节点是线性的,对于输入模式 $x_i, i = 1, 2, \cdots, k-1$,引起的输入节点到隐节点的权矩阵为:

$$V_{k-1} = [v_{1,k-1}, v_{2,k-1}, \cdots, v_{n_k,k-1}].$$

按照 Oja 准则,对于 x_k 更新权矩阵 V

$$V_k = V_{k-1} + \alpha_0 (I - V_{k-1} V_{k-1}^T) x_k x_k^T V_{k-1} =$$

$$V_{k-1} + \alpha_0 (x_k - V_{k-1} \bar{h}_{k-1}^T) \bar{h}_k^T,$$

式中:$\bar{h}_k = V_{k-1}^T x_k$,每个权向量 $v_{i,k}$ 由下式给出:

$$v_{i,k} = v_{i,k-1} + \alpha_0 \bar{h}_{i,k} (x_k - \sum_{i}^{n_k} \bar{h}_{i,k} v_{i,k-1}).$$

对于非线性广义 Hebb 学习规则来说,其中

$$V_k = V_{k-1} + \alpha_0 (x_k - V_{k-1} L_d(\bar{h}_k^T)) L_h(\bar{h}_k^T),$$

式中:$L_h(\bar{h}_k^T) = [L_h(\bar{h}_{1,k}^T), L_h(\bar{h}_{2,k}^T), \cdots, L_h(\bar{h}_{n_k,k}^T)]$ 是输出 $\bar{h}_k$ 的函数.每一个权向量 $v_{i,k}$ 可由下式给出:

$$v_{i,k} = v_{i,k-1} + \alpha_0 L_h(\bar{h}_{i,k}) (x_k - \sum_{i}^{n_k} L_d(\bar{h}_{i,k}) v_{i,k-1}),$$

终止准则:若 $n_k > n_i$,终止准则是基于误差 E_v 的减少,

$$E_v = \sum_{k}^{m} \| \bar{h}_k - V^T \bar{x}_k \|^2, \bar{x}_k = \sum_{i}^{n_k} L_d(\bar{h}_{i,k}) v_i.$$

若 $n_k > n_i$,终止准则应该是基于误差 E_v 的减少,

$$E_v = \sum_{k}^{m} \| \bar{x}_k - \sum_{i}^{n_k} L_d(\bar{h}_{i,k}) v_i \|^2, \bar{h}_k = V^T x_k.$$

学习函数 $L_d(\cdot)$、$L_h(\cdot)$ 的形式可以定义为:

$\varphi = \frac{d\rho^2(z)}{dz}$,其中 $\rho(\cdot)$ 是激励函数.

采用广义 Hebb 规则来初始化输入节点和隐层节点之间的连接权,然后采用监督训练算法初始化输出层的连接权,$v_i, i = 1, 2, \cdots, n_k$,具体过程如下:

(1) 用随机数来初始化 V;

(2) 选择确定 $L_d(\cdot)$,$L_h(\cdot)$,α_0,β;

(3) 设 $v = 1$;

(4) 对每一个 $k = 1, 2, \cdots m$,计算

$v_i, i = 1, 2, \cdots, n_k$ 和 $\bar{h}_k = V_{k-1}^T x_k$,则

$v_i, i = 1, 2, \cdots, n_k$ 和 $\bar{x}_k = \sum_{i}^{n_k} L_d(\bar{h}_{i,k}) v_i$;

(5) 用上一段的公式决定终止准则;

(6) 如果 $v > 1$,则

$E_v^{rel} = \frac{E_v^{old} - E_v}{E_v^{old}}$,否则令 $E_v^{old} = E_v$;

(7) 如果 $v = 1$ 或者 $E_v^{old} > E_v$,则令 $v = v + 1$,然后跳转到第 4 步循环.

我们随后采用监督训练算法初始化输出层的连接权:$w_i, i = 1,2,\cdots,n_0$,用样本对 $(y_k, h_k^{\exists})$, $k = 1,2,\cdots,m$ 初始化隐层和输出层之间的权值:

(i) 广义训练准则:

$$G(\lambda) = \lambda E + (1-\lambda)E' = \lambda \sum_{k=1}^{m}\sum_{i=1}^{n_0}\varphi_2(e_{i,k}) + (1-\lambda)\sum_{k=1}^{m}\sum_{i=1}^{n_0}\varphi_1(e_{i,k}).$$

式中 $\varphi_2(e_{i,k}) = \frac{1}{2}e_{i,k}^2$, $\lambda: 1.0 \rightarrow 0.0$, $\lambda = \lambda(E) = \exp(-\frac{\mu}{E^2})$,如果网络输出是二值的,而且取值 ± 1.0,则 $\varphi_1(e_{i,k}) = y_{i,k}(y_{i,k} - y_{i,k}^{\exists})$.

(ii) 基于梯度下降算法的权值更新为

$$w_{i,k} = w_{i,k-1} + \alpha\delta_{i,k}h_k^{\exists}.$$

(iii) 初始化输出层的连接权程序为

● Initialize W with random values

Select α, μ,

Set $\lambda = 1$, $E_w = 0$, and $v = 1$

● For each $k = 1,2,\cdots,m$;

Calculate $y_{i,k}^{\exists} = \sigma(w_i^T h_k^{\exists})$, $i = 1,2,\cdots,n_0$,

Evaluate $\delta_{i,k}$, $i = 1,2,\cdots,n_0$

Update w_i, $i = 1,2,\cdots,n_0$

Calculate $y_{i,k}^{\exists} = \sigma(w_i^T h_k^{\exists})$, $i = 1,2,\cdots,n_0$,

Set $E_w = E_w + \frac{1}{2}\sum_{i=1}^{n_0}(y_{i,k} - y_{i,k}^{\exists})^2$

Calculate $\lambda = \lambda(E) = \exp(-\frac{\mu}{E^2})$

If $v > 1$ then $E_w^{rel} = \frac{E_w^{old} - E_w}{E_w^{old}}$, $E_w^{old} = E_w$,

If $v = 1$ or $E_w^{old} > E_w$, then set $v = v + 1$ and goto 2

4 实验结果

实验证明:在许多识别任务上,改进的混合 HMM/ ANN 模型的识别性能比具有相同参数数目和输入特征的传统 HMM/ ANN 要好. 要实现相同的识别性能,HMM/ ANN 系统必须使用更多的参数和更复杂的模型结构. 对于同一个连续语音数据库来说,传统的 HMM/ ANN 模型的误识率为 7.5 %[4],而用本文提供的优化后的神经网络和隐马尔可夫链结合的混合模型,参数数目大致相同,整个训练过程用了 25 次迭代就收敛了,而且误识率降低到 4.1 %. 同时混合模型在非特定人识别和关键词检测问题上也有不俗的表现,充分显示了混合 HMM/ ANN 作为一种新的语音识别系统模型具有强大的生命力.

参考文献:

[1] 胡光锐,吴硕. 自组织特征映射神经网络用于语音识别的研究[J]. 应用科学学报,1997,15(1):55 - 60.

[2] 张欣研. 基于子带信息的鲁棒语音特征提取框架[J]. 中文信息学报,2002,16(1):19 - 24

[3] 韩纪庆,张磊,郑铁然. 语音信号处理[M]. 北京:清华大学出版社,2004:9.

[4] 张有为. 混合 HMM/ ANN 模型在汉语语音识别的应用(硕士学位论文)[D]. 广州:华南理工大学,2000.

[5] 宋叔飚. 神经网络在语音识别中的应用研究(硕士学位论文)[D]. 西安:西北工业大学,2002.

[6] 张卫清 周淑阁. 语音识别算法的研究(硕士学位论文)[D]. 南京:南京理工大学,2004.

[7] Tranzai Lee. New Feedback Method of Hybrid HMM/ ANN Methods for Continuous Speech Recognition[C]. USA: IEEE, 1998: 509 - 511.

An Acoustic Model Based on HMM/ ANN

LIN Kun-hui[1], XI Xiao-jing[1], ZHOU Chang-le[2]

(1. Software School, Xiamen Univ., 2. Information Science and Technique School, Xiamen Univ., Xiamen 361005, China)

Abstract: The Artificial Neural Network (ANN) can depend on weight values to store memory and knowledge for a long time. However it possesses a weak memory, not being suitable to store the instantaneous response to various input modes. The Hidden Markov Model (HMM) is better in instantaneous memory, but the presupposition precondition is not according with the real situation. So we design a hybrid HMM/ ANN model to overcome the flaws of using either of them. And basing on this model, we make a global optimization for ANN in structure design, training and structure adjustment in the later period of training. We propose an algorithm to prune hidden nodes in a trained neural network, and utilize the generalized Hebbian algorithm to reconfigure the parameters of the network. Some experiments show that the hybrid model has a good performance in speech recognition.

Key words: HMM; ANN; removing hidden nodes algorithm; generalized Hebbian algorithm