

# 语音识别关键技术研究

息晓静 林坤辉 周昌乐 蔡 骏  
(厦门大学软件学院, 厦门 361005)

**摘 要** 采用隐马尔可夫模型(HMM)进行语音声学建模是大词汇连续语音识别取得突破性进展最主要的原因之一, HMM本身依赖的某些不合理建模假设和不具有区分性的训练算法正在成为制约语音识别系统未来发展的瓶颈。神经网络依靠权能够进行长时间记忆和知识存储, 但对于输入模式的瞬时响应的记忆能力比较差。采用混合 HMM/ANN 模型对 HMM 的一些不尽合理的建模假设和训练算法进行了革新。混合模型用神经网络非参数概率模型代替高斯混合器(GM)计算 HMM 的状态所需要的观测概率。另外对神经网络的结构进行了优化, 取得了很好的效果。

**关键词** HMM ANN 非参数概率模型 BP

文章编号 1002-8331-(2006)11-0066-04 文献标识码 A 中图分类号 TP391.4

## Key Technology Research for Speech Recognition

Xi Xiaojing Lin Kunhui Zhou Changle Cai Jun  
(Software School, Xiamen University, Xiamen 361005)

**Abstract:** Because of the application of the Hidden Markov Model (HMM) in acoustic modeling, a significant breakthrough has been made in recognizing continuous speech with a large glossary. However, some unreasonable hypotheses for acoustic modeling and the unclassified training algorithm on which the HMM based form a bottleneck, restricting the further improvement in speech recognition. The Artificial Neural Network(ANN) techniques can be adopted as an alternative modeling paradigm. By means of the weight values of the network connections, neural networks can steadily store the knowledge acquired from the training process. But they possess a weak memory, not being suitable to store the instantaneous response to various input modes. To overcome the flaws of the HMM paradigm, we design a hybrid HMM/ANN model. In this hybrid model, the nonparametric probabilistic model (a BP neural network) is used to substitute the Gauss blender to calculate the observed probability which is necessary for computing the states of the HMM model. Besides, we optimize the structure of the network, and experiments show that the hybrid model has a good performance in speech recognition.

**Keywords:** HMM, ANN, Nonparametric Probabilistic Model, BP

### 1 引言

语音识别主要是让机器听懂人说的话, 即在各种情况下, 准确地识别出语音的内容, 从而根据其信息执行人的各种意图。HMM 在语音处理各个领域都获得了巨大的成功, 但是它本身有很大的限制。神经网络依靠权值信息能够进行长时间记忆和知识存储, 但是对于输入模式的瞬时响应的记忆能力较差, 很难有效地对时间变量建模。

鉴于上述 HMM 和 ANN 在处理语音识别方面各有优缺点, 所以引入了 HMM/ANN 混合模型。对于这种混合模型的研究得到了国外很多研究机构的重视, 如 Berkeley 的国际计算机学院(ICSI)、斯坦福研究院(SRI)、剑桥大学的 SVR 等机构。他们研究的语音识别系统的性能在某些方面已经明显优于传统的 HMM 系统。本文采用的利用 BP 网络代替高斯混合器来计算 HMM 的状态所需要的观测概率。该混合模型利用连续密度隐马尔可夫模型(CDHMM)来对短时语音分量建模, 同时利用 ANN 强大的区分能力来计算 CDHMM 的观察概率。另外对混

合模型进行优化, 提出了一种猜测初始隐节点数和对训练结束后的网络进行隐节点剪枝的算法, 并用广义的 Hebb 算法对网络的参数进行优化。

### 2 HMM 的建模假设及对识别系统的影响

随着对 HMM 技术的深入研究, 人们发现 HMM 还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, 它们正在成为语音识别系统的性能进一步提高的障碍。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  $q_n^i$  的概率只与  $n-1$  时刻所处的状态  $q_{n-1}^j$  有关, 而与  $n-1$  时刻以前的状态及声学矢量序列无关。这一假设对 HMM 产生的一个影响是 HMM 对协同发音的建模比较困难, 因为相邻状态对声学分布影响很大; 另一个影响是用指数递减分布描述时长分布, 而不是更精确的泊松分布。

(2) 观察矢量帧之间独立性假设, 每一时刻 HMM 只能考

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: 息晓静(1981-), 研究生, 主要研究方向: 语音识别。林坤辉(1961-), 副教授, 主要研究方向: 网络多媒体、数据库应用。周昌乐(1959-), 人工智能研究所所长, 教授, 博士生导师, 从事人工智能及其应用技术领域的研究工作。蔡骏(1966-), 厦门大学人工智能研究所副教授, 21994-2006, 主要从事计算机硬件和软件开发、计算机神经网络、遗传算法的研究。All rights reserved. <http://www.cnki.net>

虑当前帧语音,为了考虑前面语音帧的影响,HMM 必须将这些帧的信息吸收到当前帧。

(3) 观测概率的分布假设:通常假设 HMM 的状态观测概率分布是多维高斯分布或高斯分量的混合,对于连续密度的 HMM 模型,这一假设就导致了模型的不匹配,即统计模型的经验选择和真实密度之间的弱匹配。

(4) HMM 的前后向训练算法是在最大似然(ML)准则下导出的。由最大似然准则训练得到的声学模型的区分能力很弱,虽然用相互熵准则能够提高模型间的区分能力,但这一方法实施起来太复杂。

(5) HMM 是一种统计模型,模型参数的估计需要大量的训练数据,而在实际应用中总是存在参数估计精度与训练数据之间的矛盾,而且训练过程计算量大,耗时长,对连续密度 HMM 来说尤其如此。

正是由于 HMM 存在着这些不足,所以需要对其进行优化和改进,其中一个思路是引入新的模型来弥补 HMM 的缺陷。

### 3 神经网络的特点

神经网络的出发点是通过模拟大脑的机制,将包括听觉系统的生物神经系统的信息处理机制引入机器学习的研究中,使其具有学习和理解的能力。系统的所有计算都是由神经元单元完成的,而单元之间的连接权决定了网络对任意输入模式的计算响应。连接权的建立是通过训练算法进行的,具有传统的数字计算机系统即线性网络没有的优点,主要表现在:

(1) 自组织和自适应性神经网络可以从数据集中自适应地寻求答案,可以有效地用于学习数据集的内在联系。对于语音识别来说,它可用于解决非特定人语音识别等问题。

(2) 学习功能:神经网络通过学习能掌握输入和输出之间的任意映射关系,所以网络可以通过学习将语音模式映射成因素类别。

(3) 可推广性:神经网络不仅能记忆训练数据模式,还能学会相似的模式,可以从训练数据推广到新数据。在语音识别上可大大减少训练所需数据量。

(4) 非线性:神经网络可以提取系统输入之间复杂的相互作用关系,计算非线性函数,对输入进行复杂的变换,更符合现实世界的实际问题,如语音信号是一个高度非线性的过程。

(5) 鲁棒性:神经网络对物理损伤和带噪声的数据不敏感,而且带噪声的数据有助于网络形成更好的泛化能力。这一点对于语音识别特别有价值,因为经常在噪声环境下录入语音。

(6) 一致性:神经网络提供了一致的计算模式,可以很容易地融合各种约束条件,就使我们很容易使用倒谱和差分倒谱输入,还可以在听觉—视觉双模态语音识别系统中将声学 and 视觉因素结合起来。

(7) 并行性:神经网络在本质上是高度并行的,因此它非常适合在大规模并行机上实现,这就在根本上支持语音数据的高速处理。

由于上述的优点,神经网络在语音识别中得到了广泛应用,但是它也有很多的限制阻碍识别性能的提高。所以我们提出了 HMM 和 ANN 的混合模型发挥他们各自的优势。

### 4 HMM 和 ANN 的混合模型

语音信号是一种典型的动态模式序列,前后帧之间的时间

相关性非常强,所以要将神经网络应用于语音识别,必须解决好瞬时输出的记忆问题。因此考虑将神经网络与已经发展的方法 HMM 结合构成一种混合系统,在这种系统中,神经网络或者作为其前端进行预处理,或者作为后端进行后处理。

#### 4.1 HMM 和 ANN 的结合方式

ANN 和 HMM 可以以不同的方式进行结合,如 ANN 直接实现 HMM,两者在帧层面上的结合,语音层面上的结合,音段层面的结合和子层面上的结合等方式。本文介绍的并非用神经网络直接实现 HMM,而是探索一种混合模型,即能优化 HMM 模型,又能充分利用每一种技术的长处:HMM 的时间建模和 ANN 的声学建模,特别是用神经网络计算 HMM 状态的观测概率。用神经网络来估计观测概率,设计和训练神经网络的方法有很多,最简单的方式就是将一帧语音矢量直接映射成观察概率,网络的训练也是一帧接一帧地进行,这种方法就成为帧层面上的网络训练。

在音段层面方式中,神经网络的输入来自整个语音段,而不是一帧语音或固定长度的语音窗。这样就可以让网络更好地利用存在音段中的所有语音帧之间的相关性,而且还可以更容易地利用其它信息,比如时长。这个方法的缺点是必须先对语音进行分段,神经网络才能对分出的各个音段进行计算。在这些不同的结合形式中,实验结果表明,帧层面上的结合比较符合 HMM 和 ANN 两种方法的特点。

#### 4.2 混合 HMM/ANN 模型的特点

与传统的 HMM 相比,混合 HMM/ANN 模型不仅在理论上抛弃了 HMM 一系列不合理假设,而且在训练过程中自然地引入了 HMM 状态之间的区分机制。混合模型克服了最大似然准则(MLE)区分能力差的缺陷,而且在系统体系上又保持了 HMM 的框架,所以混合 HMM/ANN 模型是 HMM 和 ANN 两种模型的有机结合,具有明确的数学和物理意义。与传统的 HMM 相比,混合 HMM/ANN 模型还为融合多种语音特征矢量,考虑不同层次上的多种约束提供了一个非常方便的接口。

混合模型结构还充分利用了 ANN 的下述特性,克服了 HMM 的一系列缺陷和不足,主要特点如下:

(1) 混合模型可以自适应学习以适应语音数据的变化;可以不必拘泥于选取特殊的语音参数而对综合的输入模式进行训练和识别。

(2) 可以把人的听觉模型融合到 ANN 中,在 ANN 的输入端可以同时加载相邻数帧的语音特征矢量,因而和语音信号的实际情况更加符合。

(3) 分类神经网络的输出端可以逼近任意形式的概率分布函数,它能很好地掌握蕴含在训练数据中的概率分布特性,所以能够较真实地描述语音信号的概率分布特性。

(4) ANN 的训练着眼于描述类间的边界,而不是描述每一类内的分布,所以是一种区分性训练。在处理语音信号的时变特性方面,混合模型利用了 HMM 在这方面的特长。

所以混合模型在系统框架上保持了传统 HMM 的框架,而局部功能模块如:概率估计、上下文关联等则通过 ANN 实现,这样有机地将 HMM 和 ANN 的各自优势统一在一个系统之中,又弥补了各自的不足。

#### 4.3 混合模型的算法

在混合 HMM/ANN 中用神经网络来进行声学建模,而时域

建模依赖于传统的HMM。用神经网络进行声学建模一般有两种方法:

第一种方法是预测网络。预测网络的输入是相继的几帧语音特征矢量,输出下一帧语音的预测值,通过给每一个音子分配一个预测网络,就可以比较每个音素预测网络的预测误差,选取其中具有最小预测误差的网络与当前语音段匹配,体现了相邻语音帧之间的时域相关性。

第二种方法是分类网络。在分类网络中,输入还是相继的几帧语音矢量,但输出直接映射成HMM状态。分类网络可以用N个输出节点表示N个类,将输入映射成N类中的一种。分类网络具有简单直观、在本质上是区分性的、在设计上是模块化的、可以方便地组合成更大的系统、具有完善的数学解释等优点,所以可以很容易地集成到HMM的统计识别框架中。

在混合HMM/ANN模型方法中,采用分类网络来估计HMM状态的后验概率。传统的HMM/ANN模型是用神经分类网络代替了高斯概率密度分布混合器。因此可以用语法上下文相关的输入模式作为神经网络的输入,这样就考虑了语音矢量间的时间相关性。神经网络输出的后验概率密度为 $p(q|X_{t-d}^{hd})$ ,而HMM状态需要的是似然概率密度 $p(x_t|q)$ ,应用贝叶斯公式,我们就可以从后验概率密度 $p(q|X_{t-d}^{hd})$ 导出尺度化的似然概率密度: $\frac{p(x_t|q)}{p(x_t)} = \frac{p(q|x_t)}{p(q)}$ ,在进行识别时,因为 $p(x_t)$ 对所有的路径来说是一样的,所以尺度化的似然函数并不会改变识别的结果。因为分类网络极大地体现了混合模型的精髓,所以我们采用它来建立语音识别系统。在用混合HMM/ANN模型进行语音识别时,ANN计算的是HMM状态的尺度化观察概率。整个识别过程:分两步进行:(1)计算t时刻所有HMM状态的尺度化观察概率;(2)计算t时刻激活路径的路径积累概率,并根据路径积累概率进行剪枝,确定t+1时刻的激活路径。

对于每一帧语音向量,可以用后序遍历计算每个节点网络(BP)的输出向量 $g(x_t)$ 、 $g_{ji}(x_t)$ 和 $u_{i,j}(x_t)$ ,并计算节点的局部输出 $u_i(x_t)$ 如下:

$$u_i(x_t) = \sum_j g_{ji}(x_t) * u_{i,j}(x_t)$$

遍历到根节点,就得到了总体输出 $u(x_t)$ ,其中:

$$u(x_t) = \sum_i g(x_t) \sum_j g_{ji}(x_t) * u_{i,j}(x_t)$$

这个总体输出就是后验概率 $p(q|X_{t-c}^{hc})$ 的估计。用后验概率除以状态的先验概率就得到了尺度化的似然概率:

$$\frac{p(q|X_{t-c}^{hc})}{p(q)} = \frac{p(X_{t-c}^{hc}|q)}{p(X_{t-c}^{hc})}$$

在帧同步Viterbi路径搜索算法中,对于每一帧语音矢量 $x_t$ ,我们可以直接将尺度化观察概率累计到每条活跃路径的路径概率中:

$$a_{\lambda_m, q}(x_t) = \max_{i, \text{pre}(q_k)} \{ a_{\lambda_m, q}(x_{t-1}) + \log(a_{i,k}) + \log(p(q|x_t)) - \log(p(q)) \}$$

其中, $a_{\lambda_m, q}(x_t)$ 表示t时刻达到模型 $\lambda_m$ 的状态 $q_k$ 的最优路径积累概率, $\text{pre}(q_k)$ 表示状态的在搜索网络中的所有前趋节点集合, $a_{i,k}$ 表示从前趋节点i到状态k的转移概率,根据当前

活跃路径的积累概率的计算结构,我们可以剪枝掉一些积累概率小于某个阈值的活跃路径,得到t+1时刻的活跃路径,对于每一帧语音向量,我们都进行上面的步骤,直到到达发音语句结束。从结束节点的最优路径逐步向前回溯,就可以得到和发音语句相对应的HMM模型序列,这就是声学模型识别的首选结果。

#### 4.4 混合模型的优化

改进1提出一种新的隐节点剪枝算法。

优化模型考虑的首要问题是如何确定最优的隐节点个数,主要有两种方法:

(1) 隐节点增长算法:是在训练中不断增加隐节点数目来改进帧识别率。

(2) 隐节点剪枝算法:是在训练过程中去掉多余的隐节点和连接权。

经分析比较,我们提出了一种新的隐节点数目确定方法:

(1) 用迭代自组织数据分析方法得到训练数据的聚类中心数目,再为属于不同类的一对聚类中心分配一个隐节点。这样,隐节点就是对输入模式形成高维空间,在这个空间中输入节点更容易形成决策曲面。通过这样就估计了一个对于训练和训练后的剪枝都合适的隐节点的数目N。

(2) 训练具有N个隐节点的BP网络。

(3) 通过迭代去除网络中的冗余隐节点,然后在保持原有输入输出关系的前提下,调整剩下隐节点的权值,最后得到一个最优的网络隐节点个数,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:

$$\sum_{j \in \{A, C, D\}} w_{j,B} y_j(n) = \sum_{j \in \{C, D\}} (w_{j,B} + w_{j,B}) y_j(n) \quad \forall n \in \{1, 2, \dots, N\}$$

其中 $w_{j,B}$ 代表节点j到B的权值, $w_{j,B}$ 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:

$$\begin{bmatrix} y_C(1) & y_D(1) \\ y_C(N) & y_D(N) \end{bmatrix} \begin{bmatrix} c_{B,A} \\ d_{B,A} \end{bmatrix} = W_{A,B} \begin{bmatrix} y_A(1) \\ y_A(N) \end{bmatrix}$$

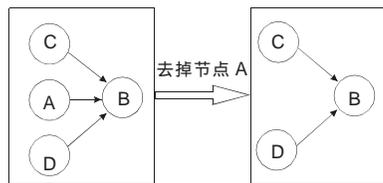


图1 一种新的隐节点剪枝算法

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 $(c_{B,A}, d_{B,A})$ ,我们为了保证输入输出关系,所以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,该算法的计算量非常小,而且冗余节点的检测简单易行。

改进2 网络权参数的初始化。

在开始训练网络之前,所有的权值都必须赋予一个初值,即初始化问题。它是神经网络训练的一个很重要的问题,直接关系到训练结果的收敛。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。对于 BP 神经网络, 权值初始化还没有一个简单有效的办法, 这里我们用神经网络来进行分类, 因此用 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  $x_i, i=1, 2, \dots, k-1$ , 引起的输入节点到隐节点的权矩阵为:

$$V_{k-1}=[v_{1,k-1}, v_{2,k-1}, v_{3,k-1}, \dots, v_{n_k,k-1}]$$

按照 Oja 准则, 对于  $x_k$  更新权矩阵  $V$ :

$$V_k=V_{k-1}+a_0(I-V_{k-1}V_{k-1}^T)x_kx_k^TV_{k-1}=V_{k-1}+a_0(x_k-V_{k-1}\bar{h}_{k-1})\bar{h}_{k-1}^T$$

式中:  $\bar{h}_k=V_{k-1}^Tx_k$ , 每一个权向量  $v_{i,k}$  可由下式给出:

$$v_{i,k}=v_{i,k-1}+a_0\bar{h}_{i,k}(x_k-\sum_{i=1}^{n_k}\bar{h}_{i,k}v_{i,k-1})$$

对于非线性广义 Hebb 学习规则来说, 其中:

$$V_k=V_{k-1}+a_0(x_k-V_{k-1}L_d(\bar{h}_k))L_h(\bar{h}_k)$$

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

$$v_{i,k}=v_{i,k-1}+a_0L_h(\bar{h}_{i,k})(x_k-\sum_{i=1}^{n_k}L_d(\bar{h}_{i,k})v_{i,k-1})$$

终止准则: 若  $n_k > n_1$ , 终止准则应该是基于误差  $E_v$  的减少:

$$E_v=\sum_{k=1}^m\|\bar{h}_k-V^T\bar{x}_k\|^2, \bar{x}_k=\sum_{i=1}^{n_k}L_d(\bar{h}_{i,k})v_i$$

若  $n_k < n_1$ , 终止准则应该是基于误差  $E_v$  的减少:

$$E_v=\sum_{k=1}^m\|\bar{x}_k-\sum_{i=1}^{n_k}L_d(\bar{h}_{i,k})v_i\|^2, \bar{h}_k=V^Tx_k$$

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

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

网络参数的初始化可以分成两部分, 一个是输入节点和隐层节点之间的连接权的初始化, 一个是隐层节点和输出层节点之间连接权的初始化。首先我们采用广义 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  $v_i, i=1, 2, \dots, n_k$ 。

具体过程如下:

1. 用随机数来初始化  $V$
2. 选择确定  $L_d(\cdot), L_h(\cdot), a_0, \beta$ ;
3. 设  $v=1$ ;
4. 对每一个  $k=1, 2, \dots, m$ , 计算  $v_i, i=1, 2, \dots, n_k$  和  $\bar{h}_k=V_{k-1}^Tx_k$ , 则

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

5. 用上一段的公式决定终止准则

$$6. \text{ 如果 } v>1, \text{ 则 } E_v^{\text{rel}}=\frac{E_v^{\text{dd}}-E_v}{E_v^{\text{dd}}}, \text{ 否则令 } E_v^{\text{dd}}=E_v;$$

7. 如果  $v=1$  或者  $E_v^{\text{dd}}>E_v$ , 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  $w_i, i=1, 2, \dots, n_0$ , 用样本对  $(y_k, h_k^3), k=1, 2, \dots, m$  初始化隐层和输出层之间的权值:

(1) 广义训练准则

$$G(\lambda)=\lambda E+(1-\lambda)E=$$

$$\lambda \sum_{k=1}^m \sum_{i=1}^{n_0} \phi_2(e_{i,k})+(1-\lambda) \sum_{k=1}^m \sum_{i=1}^{n_0} \phi_1(e_{i,k})$$

$$\phi_2(e_{i,k})=\frac{1}{2}e_{i,k}^2$$

$$\lambda: 1.0 > 0.0, \lambda=\lambda(E)=\exp\left(-\frac{\mu}{E^2}\right)$$

如果网络输出是二值的, 而且取值  $\pm 1.0$ , 则:

$$\phi_1(e_{i,k})=y_{i,k}(y_{i,k}-y_{i,k}^3)$$

(2) 基于梯度下降算法的权值更新

$$w_{i,k}=w_{i,k-1}+\alpha \frac{\partial}{\partial w_{i,k}} h_k^3$$

(3) 初始化输出层的连接权程序

Initialize  $W$  with random values Select  $\mu, \lambda$

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

For each  $k=1, 2, \dots, m$ ;

Calculate  $y_{i,k}^3=(w_i^T h_k^3), i=1, 2, \dots, n_0$

Evaluate  $E_{i,k}^0, i=1, 2, \dots, n_0$

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

Calculate  $y_{i,k}^3=(w_i^T h_k^3), i=1, 2, \dots, n_0$

Set  $E_w=E_w+\frac{1}{2} \sum_{i=1}^{n_0} (y_{i,k}-y_{i,k}^3)^2$

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

if  $v>1$  then  $E_w^{\text{rel}}=\frac{E_w^{\text{dd}}-E_w}{E_w^{\text{dd}}}$

$E_w^{\text{dd}}=E_w$

If  $v=1$  or  $E_w^{\text{dd}}>E_w$ ; then set  $v=v+1$  and go to 2

## 5 实验结果

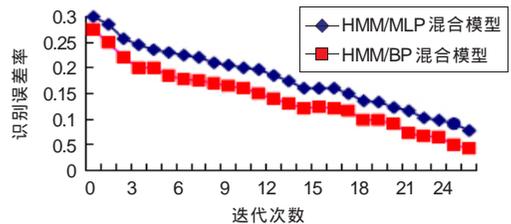


图 2 识别误差率随迭代次数的变化关系

实验证明: 在许多识别任务上, 混合 HMM/ANN 模型的识别性能要比具有相同参数数目和输入特征的传统 HMM 要好。要实现相同的识别性能, HMM 系统必须使用更多的参数和更复杂的模型结构。图 2 中显示的是误差率随着迭代次数变化的关系图, 以及 HMM 和 MLP 的混合模型与本文改进的 HMM 和 BP 网络混合模型误差率的比较。我们可以看到, 在对于同一个连续语音数据库来说, 用改进的 BP 网络输出作为后验概率估计的误差率为 7.5%, 而用本论文中提供的非参数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, 而且误差率降低到 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。(收稿日期: 2005 年 7 月)

## 6 结论

由于 IM 病毒不需要搜索目标主机即可进行传播, 和传统的需要扫描的蠕虫病毒相比, 它的危害更大。因此 IM 的安全和对 IM 蠕虫病毒的监控与防御显得越来越重要。本文首先分析了 IM 网络的拓扑结构并建立了病毒的传播模型, 考虑用户打开附件或者点击 URL 的概率及其变化, 并在不同拓扑的网络上进行了仿真, 发现 IM 病毒在无标度网络上更易于传播, 同时由于无标度网络的特点, 在无标度网络上也更利于我们监控和防御。同时, 我们还发现单独增强用户的安全意识能减少病毒的感染范围, 但不能减缓病毒的传播速度。

基于 IM 病毒的传播特征, 本文提出了监控和防御 IM 病毒的一条思路。通过对 IM 软件的修改, 使其能对已知的病毒免疫并能自动更新病毒库。通过监视延时消息队列, IM 软件能在不需要 IM 服务器的参与下就能监控病毒, 降低了服务器的负担。由于 IM 网络具有无标度网络的性质, 在网络中布置若干监控点, IM 服务器将其加入到度大节点的好友列表, 使其能在第一时间收到被感染节点发送的恶意消息, 从而能监控并遏制新病毒的爆发。本文提出的方法目前还是在原型阶段, 在实际的应用中需要根据实际情况, 做更具体和深入的细化工作。(收稿日期: 2005 年 11 月)

## 参考文献

1. iResearch. China Instant Messenger Research Report. [http://www.iresearch.com.cn/instant\\_messenger/](http://www.iresearch.com.cn/instant_messenger/), 2005
2. Wen WP et al. Research and development of Internet worms[J]. Journal of Software, 2004; 15(8): 1208-1219
3. Zou C C, W Gong, D Towsley. On the Performance of Internet Worm Scanning Strategies[R]. Technical Report, TR-03-CSE-07, University of Massachusetts, 2003
4. San Institute. The Twenty Most Critical Internet Security Vulnerabilities. <http://www.sans.org/top20>, 2005
5. Hindocha N, E Chien. Malicious threats and vulnerabilities in instant

messaging. Proceedings VB, 2003: 114-124

6. Gostev A. Malware Evolution. KasperskyLab, <http://www.viruslist.com/en/analysis?pubid=162454316>, 2005
7. WormBlog.com. Recent Spate of IM Worms. [http://www.wormblog.com/2005/03/\\_recent\\_spate\\_o.html](http://www.wormblog.com/2005/03/_recent_spate_o.html), 2005
8. Smith R D. Instant Messaging as a Scale-Free Network. eprint cond-mat/0206378, 2002
9. Williamson M M, A Parry, A Byde. Virus throttling for instant messaging[R]. Technical Report, HPL-2004-81, Chicago, 2004
10. Pastor-Satorras R, A Vespignani. Epidemic dynamics and endemic states in complex networks[J]. Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 2001; 63(6)
11. M E J Newman. The structure and function of complex networks[J]. SIAM Review, 2003; 45(2): 167
12. Albert R, A-L Barabási. Statistical mechanics of complex networks[J]. Reviews of Modern Physics, 2002; 74(1): 47
13. Erdős P, A Rényi. On the evolution of random graphs[J]. Publ Math Inst Hung Acad Sci, 1959; 5: 17-60
14. Watts D J, S H Strogatz. Collective dynamics of 'small-world' networks[J]. Nature, 1998; 393(6684): 440
15. Albert-László Barabási, R Albert. Emergence of scaling in random networks[J]. Science, 1999; 286: 509-512
16. Barabási A-L, R Albert, H Jeong. Mean-field theory for scale-free random networks[J]. Physica A, 1999; 272: 173-187
17. Zou C C, D Towsley, W Gong. Email Worm Modeling and Defense[C]. In: 13th International Conference on Computer Communications and Networks (ICCCN'04), Chicago, 2004
18. Petter H et al. Attack vulnerability of complex networks[J]. Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 2002; 65(5)
19. Albert R, H Jeong, A L Barabási. Error and attack tolerance of complex networks[J]. Nature, 2000; 406(6794): 378
20. Pastor-Satorras R, A Vespignani. Epidemic Spreading in Scale-Free Networks[J]. Physical Review Letters, 2001; 86(14): 3200

(上接 69 页)

## 参考文献

1. 张雄伟, 陈亮, 杨吉斌. 现代语音技术及应用[M]. 机械工业出版社, 2003: 219-222
2. 谢锦辉. 隐 Markov 模型 (HMM) 及其在语音处理中的应用[M]. 华中理工大学出版社, 1995: 103-113
3. 黄昌宁, 夏莹. 语音信号处理专论[M]. 清华大学出版社, 广西科学技术出版社, 1996
4. 庄越挺, 潘云鹤, 吴飞. 网上多媒体信息分析与检索[M]. 清华大学出版

社, 2002: 122-272

5. Lawrence R Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition[J]. Proceeding of the IEEE, 1989; 77(2)
6. L Rabiner, Biing-Hwang Juang. Fundamentals of Speech Recognition[M]. Prentice Hall, 1992
7. Austin L. Speech Recognition Using Segment Neural Nets[C]. In: Proc IEEE International Conference on Acoustics, Speech and Signal Processing
8. Lippmann R, Singer E. Hybrid Neural HMM Approaches Wordspotting. ICASSP

(上接 90 页)

3. 施法中. 计算机辅助几何设计与非均匀有理 B 样条 (CAGD&NURBS) [M]. 北京航空航天大学出版社, 1994
4. Fujichi Yoshimoto, Masamitsu Moriyama, Toshinobu Harada. Automatic Knot Placement by a Genetic Algorithm for Data Fitting with a Spline. Shape Modeling International '99, IEEE Computer Society Press, 1999: 162-169

5. Fujiichi Yoshimoto, Toshinobu Harada, Yoshihide Yoshimoto. Data fitting with a spline using a real-coded genetic algorithm[J]. Computer-Aided Design, 2003; 35: 751-760
6. 王小平, 曹立明. 遗传算法——理论、应用与软件实现[M]. 西安交通大学出版社, 2002
7. 蔡钧. 基于遗传算法的测量数据拟合[J]. 扬州大学学报 (自然科学版), 2002; 5(4): 59-62