第 20 卷 第 2 期 电 子 科 学 学 刊 Vol.20 No.2
1998 年 3 月 JOURNAL OF ELECTRONICS Mar. 1998

基于模糊神经网络的声母识别 [1]

梅 勇 王群生 徐秉铮
(华南理工大学无线电系 广州 510641)

摘 要 模板匹配法技术是汉语声母识别中较为成功的算法，但它的缺陷影响了其恢复错误、改善识别性能。神经网络 (NN) 和模糊系统的结合，保留了双方的优点，充分利用了模糊神经网良好的容错性能、计算性能、分类性能和决策性能。本文重点研究了两种基于模糊神经网的声母识别方案，通过对其结构、识别率和特点的分析，可看出模糊神经网的声母识别性能明显优于模板匹配法，是更适于语音识别的网络。

关键词 声母识别，模糊系统，神经网络

中图号 TN-052, TN912.3

1 引 言

目前在汉语语音识别中，较成熟的是以全音节为单位进行识别，且新技术不断涌现，如集成神经网络识别法 [1,2]。这一方案充分利用了汉语语音以音节为发音单位这一特点，受协同发音的影响较少。但是从识别的角度来看，数量偏多，计算量较大，而且音节之间的细微差别常被整个音节的相似性所淹没，不利于不认人的语音识别。

另一方面，我们注意到所有的汉语音节都可是由声母、韵母和声调所组合成。在汉语普通话里，只有 22 个声母，38 个韵母和 5 个声调 (包括零声调)。如果用声韵母作为一个音节的识别单位，所需样本数只为全音节的几分之一，且声韵母的切分较容易，同时，它对不认人连续语音识别也较适用。因此，将声、韵、调分开识别有机整合智能调整的方案，作为一个由认人向不认人语音识别的手段是较为理想的。

在声母实时识别方案中，传统的识别方案是：将声母识别分解为三个过程；先进行分类识别，随后再根据声调和韵母的识别情况对声母进行进一步的细分类，最后用模板匹配的方法对类内声母进行识别。上述的方案已在实际系统中应用，取得了较好的识别性能 [3]。尽管如此，传统的模板匹配法也存在一些缺点，如在将声母预分类时，每类的特性都与其它类别有所区别，但事实证明，无论如何划分，各类之间的特性仍具有一定的模糊性；另一方面，传统的模板匹配法不太适用于区分发音类似的声母；又如，因为语音识别知识隐含地包含在模板上，所以如何恢复错误、改善识别性能较困难。

目前，国内外的声母识别方法，除了上述的识别方案之外也有采用神经网络 [4] (如 MLP(多层感知器)) 作为识别方案的。但在这些方案中，神经网络的输出单元即为所识别的音素，即未与模糊系统相结合。这就是说，神经网络是作为音素的模板而存在的，所以不可避免地有着模板匹配法的类似缺点。

神经网络 (NN) 技术以其自适应性、并行性、非线性、鲁棒性和学习特性而倍受人们青睐，并被广泛应用于语音识别领域。不同的 NN 结构和算法在进行语音识别时都显示了实力，如

[1] 1996-10-17 收到，1997-08-29 定稿
国家和省自然科学基金资助项目

MLP， SOFM(自组织特征映射) 等。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) 它具有基于误差回传 (BP) 的极强的学习能力；(2) 可实现输入输出信号间的足够复杂的映射；(3) 可以通过并行处理结构提供高速度和高可靠性。而且，神经网络与模糊系统的结合，能取双方之长从而改进整个识别系统的性能。

本文将研究模糊神经网用于声母识别。

2 模糊神经网基础

总体说来，神经网络和模糊数学结合用于模式识别主要有两种方式：第一种方式是在已存在的模糊模型中，神经网络作为计算工具。比如说：用神经网络去构造隶属度函数；完成模糊逻辑 (如模糊并、交) 操作；为模糊控制器实现最优规则 [5]。第二种方式，是将模糊的观点引入神经网络，构造组合的模糊神经网。比如，分类网学习期间目标输出可以是模糊分类矢量，这样，神经网络就是模糊分类器。另外一个将模糊溶入神经网的方案是改变神经网节点的集成 / 传输函数，以便这些节点能根据到达的信息进行模糊聚类。

在模式识别中，最重要的神经网是分簇网 (clustering nets) 和分类网 (classifier nets)。在这些神经网中学习规则是一个重要因素。学习规则可以记作 $W_{t+1} = U(W_t)$，这里 $W_t = (W_{1t}, \cdots\cdots W_{Mt})$ 是网络权矢量在时刻 t 的取值。学习时，当网络输出不是所希望的输出值时，网络权值进行更新。当神经网络作为一个分类网时，它可以被认为是一个分类函数。我们可以选择不同的学习规则去匹配某一种特定的网络结构，其目的都是优化误差 (实际输出与所希望的输出之差) 函数。

目前，最通用的分类网是前馈神经网 (FFNN)。输入特征矢量送入网络，进入隐层，最后输出 c 个分量 $u_1, \cdots\cdots u_c, u \in R^c$。若学习算法采用误差回传 (BP) 算法，我们把此分类网叫作 FFBP(Feed Forward BP) 分类网。通常此网络的输出是模糊隶属度数值。

另外较常用的一种网络是 KCN 网 (Kohouen Clustering Net)。它的典型特征示意于图 1。图 1 中，$N_{\text{fcu}} = \{y \in R^c | y_k \in [0,1], \forall k\}$。KCN 网中，输入层直接与输出层相连。输出层中每个节点都有一个权矢量与之相连，$v = (v_1, \cdots\cdots v_c)$ 是未知的聚类中心矢量 ($v_i \in R^p, 1 \le i \le c$)。当输入矢量 x 输入网络时，计算 x 与 v_i 之间的距离。输出节点展开竞争，距离最小者为胜者。同时它和它的邻域权矢量将进行更新。

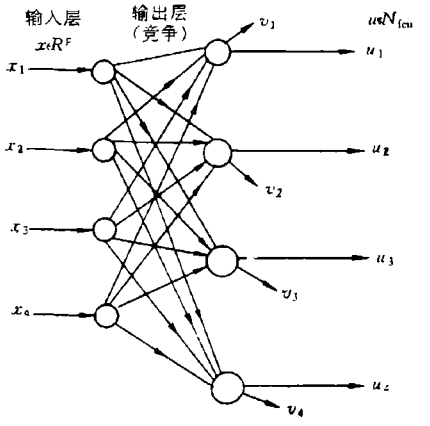

图 1 Kohonen 分簇网

KCN 网的最大局限性是，当学习模式较少时，网络的分类效果取决于输入模式的先后次序，且网络连接权向量的初始状态对网络的收敛性能有很大的影响。另外，KCN 网不具有生物神经网络的稳定性，即对学习模式的记忆将影响网络原有的模式记忆。

所以，Huntsberger 等将 KCN 网模糊化，引出模糊 KCN 网 (FKCN)[6]。其要点是将学习率 $\{\alpha_{ik,t}\}$ 变为模糊隶属度数值 $\{u_{ik,t}\}$，$u_{ik,t} = [D_{ik,t} / \sum_{j=1}^{c} D_{jk,t}]^{-2/(m-1)}$，其中 $D_{ik,t} = \|x_{k,t} -$

$v_{i,t}\|_A = \sqrt{(x_{k,t}-v_{i,t})^T A(x_{k,t}-v_{i,t})}$, $m=(m_0-\Delta mt)$, $m_0>1$。FKCN 网克服了 KCN 的缺陷，是一种较为理想的分簇网络。

3 用模糊神经网进行声母识别

以下研究两种典型的模糊神经网声母识别方案。

3.1 基于神经网络作声学特征检测器，模糊逻辑作决策的声母识别

3.1.1 基本原理 每个声母有它特定的声学特征，如声母 b，有 buz-bar 特性，说明了发声声母 (voiced sound) 的特性；有声功率的突然上升，说明了止声母 (stop consonant) 的特性。这些特性是声母与声母之间区别的音征 (cue)，所以叫做声学音征 (acoustic cues)。

本方案对声母的识别采用基于声母对区分规则的语音识别系统 [7]。

3.1.2 识别流图 首先进行韵母识别。然后是声母识别。图 2 示意出声母识别流图。流图中和下文将使用下述符号：C_i 为第 i 个声母候选者，$R(i,j,k)$ 为用来区分声母对 (C_i,C_j) 的第 k 个声学音征检测结果，$A(i,j,k)$ 为根据第 k 个声学音征检测结果判断为 C_i 的概率，$P(i,j)$ 为根据 K 个声学音征检测结果判断为 C_i 的概率，$S(i)$ 为通过累积 $P(i,j)$ 而得到的 C_i 的得分，$nC2$ 为声母对的数目。

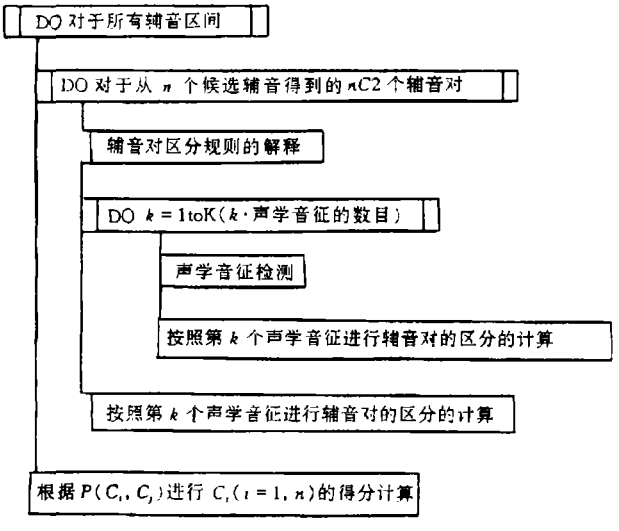

图 2 辅音识别流图

3.1.3 实验系统 在本实验系统中，先通过基于平稳部分检测进行分段和 VCV(Vowel Consonant Vowel) 模板匹配法，进行韵母识别和声韵母分开，得到候选声母，然后将这些声母通过图 3 所示的声母识别系统。

有三种类型的神经网络用来进行声学音征检测。

类型 1 通过输入几帧输入模式进行声学音征检测。这类网络有 80 个输入单元，24 个隐层单元 (第二层)，第三层有 8 个单元，输出层有 1 个单元，相邻层间互连。

类型 2 这类网络用来检测功率变化的局域特征。它有 7 个输入单元，第二层有 5 个单元，第三层有 3 个单元，输出层有 1 个单元，相邻层间互连。

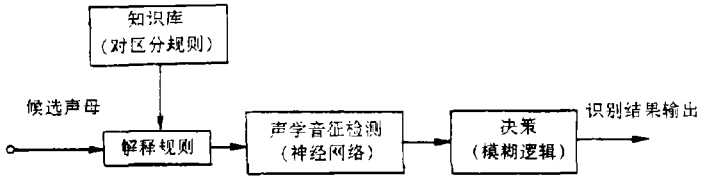

图 3　声母识别系统

类型 3　用来检测功率变化的全局特征。它的输入单元数目可变，第二层和第三层的单元数目亦可变。隐层单元仅与前一层的 5 个相邻单元相连接。

通过对典型的学习模式学习后，识别时，神经网络输出模糊隶属度数值。因此，第 k 个声学音征支持 (support) 声母 C_i 时，对于第 k 个声学特质得到的声母对区分结果为

$$A(i,j,k) = R(i,j,k), \tag{1}$$

$$A(i,j,k) = 1 - R(i,j,k). \tag{2}$$

总的对区分结果 $P(i,j)$ 可以从几个声学音征 $(A(i,j,k), k = 1,\cdots,m)$ 的声母对区分结果而得到。每个声学音征检测结果 $(R(i,j,k), k = 1,\cdots,m)$ 可认为是支持或否定某声母的证据。

$$P(i,j) = \bigcup_{k=1}^{m} A(i,j,k). \tag{3}$$

最后计算所有候选声母的最终得分。很明显，高分的声母最有可能是所要识别的声母。

$$S(i) = \bigcap_{i=1,i\neq j}^{m} P(i,j). \tag{4}$$

3. 1. 4 **实验结果**　识别实验由 2 位男性各发出 150 个城市名两次。神经网络的学习是利用第一次发音进行的。这两位发声人也为 VCV 模板发音一次。

神经网络的学习是通过误差回传 (BP) 算法进行学习的。学习时，选择明显具有某种声学音征的输入模式进行学习，此时学习结果应为 1；同时，选择明显不具有这种声学特质的输入模式输入，此时学习结果应为 0。

实验结果表明，本方案相对于传统的模板匹配法，识别率有较大提高。结果如下：

69.6%→74.6%　A 发音人，　第一次发音
77.2%→83.7%　A 发音人，　第二次发音
63.0%→71.4%　B 发音人，　第一次发音
64.5%→73.0%　B 发音人，　第二次发音

对于采用神经网络作为音素模板的系统来说，由于它实质上也是一种模板匹配法，所以它同模糊神经网的比较结果类似于以上结果。

3.2　**基于神经网络模拟生成控制规则，模糊微处理器进行分类的声母识别** [8]

从声母的时域特性看，虽然一个汉字的读音长度差异较大，但声母的长度却比较稳定，其它的一些特性，如过零率、相对能量等都有规律性，所以我们采用时域分析来对声母进行分类识别。我们可以根据汉语声母的发音特点，将声母分为几类，使每类的时域特性都与其它类有所区别。但实验表明，无论如何划分，各类之间的特性仍具有一定的重叠性，各类间没有非常严格的界限，即各类间具有一定的模糊性。这启发我们采用模糊分类的思想来实现声母分类识

别。本方案将以神经网络自学习来产生模糊分类规则，以专用的模糊逻辑处理芯片 (如三菱公司制成的模糊逻辑处理器 MELP740) 来进行声母分类识别。MELP740 支持规则数可达 1500 种以上，元函数种类达 30000 种以上。

3. 2. 1 识别方案　识别方案框图示意于图 4 。这里，选用三个参数作为模糊变量。它们是相关能量，平均过零率和声母音长，我们分别记作 RE 、 AZ 、 CL 。同时定义 PR 为预分类结果。(见图 5)

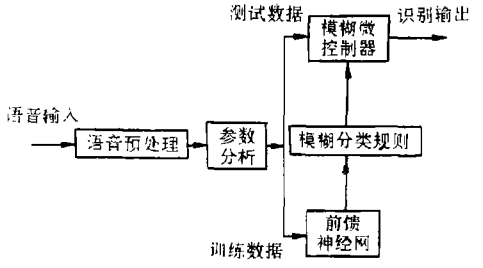

图 4 识别方案框图

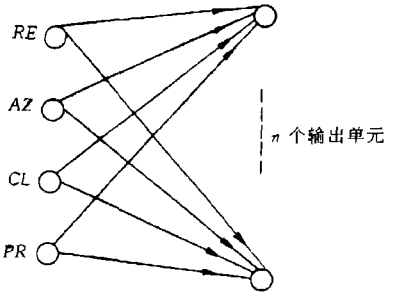

图 5 前馈神经网产生规则

本方案中神经网络采用前馈网络，应用自学习原理去估计规则。其结构如图 5 所示。

此网络输入为 RE 、 AZ 、 CL 、 PR 上的模糊集，输出为所属类别。输出单元数为 2000 ，每一单元代表一个虚规则。网络学习方法采用微分竞争学习。每次网络学习后，得到一个获胜神经元并按学习规则修正它的突触矢量，同时得到虚规则的频度值。把那些分布占大多数的虚规则视为有效规则，其余的舍去。在实验中，我们发现一些条件相同，但所判类别不同的规则，这是由于样本的离散性造成的，这种规则模糊性太强，将它们视为无效规则而舍去。

3. 2. 2 实验结果　采集不同发音人的声母样本数据约 2000 个作为神经网络的训练数据，得到自学习产生的规则和隶属度函数的中心值。将它们送入模糊处理器 MELP740 中，另外采集测试数据进行测试。得到的识别率与模板匹配法对比，可以看到，在模板匹配法判决正确率较高时，性能改善不明显，而在模板匹配法判决正确率较低时，性能有较明显改善。

4 结论和展望

模糊神经网为声母识别带来了新曙光。将模糊神经网引入声母识别领域，充分利用了模糊神经网的容错性能、计算性能、分类性能和决策性能。文中重点研究了两种基于模糊神经网的声母识别，并对其性能和特点进行了分析讨论。与传统识别手段相比，每种方法都从不同的方面，不同程度地提高了识别率。识别实验结果表明，模糊神经网的识别性能优于传统识别方法。此外，模糊神经网用于声母识别时，既可软件编程实现，又可硬、软件结合实现。

文中所述的网络目前仅在孤立词语音识别实验中进行了一些研究。因此，今后如何应用模糊神经网进行连续语音识别是一个研究方向。此外，从模糊神经网本身来讲，另一类较有发展前途的分类网是 ART 网络 [9] 。且一个大的发展趋势是将模糊联系 (fuzzy relations) 引入神经网络。可以预见，不久的将来这类网络会在语音识别中扮演重要角色。

参 考 文 献

[1] Tatsuo MATSUOKA. Syllable Recognition Using Integrated Neural Networks. in Proceedings of the IJCNN-89, Tokyo: 1989, June: 251–258.

[2] Matsuoka T, Hamada H, Nakatsu R. A Study on Integrated Neural Networks for Syllable Recognition. Proc. ASJ Fall Meeting, Tokyo: 1988, October: 2–16.

[3] 吴立忠. 汉语普通话辅音分类的研究：[硕士论文]. 华南理工大学无线电系，1986.

[4] Waibel A, *et al.* Phoneme Recognition: Neural Networks vs Hidden Markov Models. in Proc. of ICASSP88, New York: 1988, April: 107–110.

[5] Berenji H. A reiforcement larrg-based architecture for fuzzy logic control. Int'1, J. Approx. Reasoning, 1992, 6(2): 261–292.

[6] Huntsberger T, Ajjimarangsee P. Parallel self-organizing feature maps for unsupervised pattern recognition. Int'1. J. General Sys., 1990, 10(16): 261–292.

[7] Aikawa K, *et al.* Automatic Generation of Consonant Discrimination Rules. in Proc. of ICASSP86, Tokyo: 1986, April: 2755–2758.

[8] 诸　静，等. 模糊控制原理与应用. 北京：机械工业出版社，1995, 194–312.

[9] Carpenter G A, Grossberg S, Rosen D B. Fuzzy ART: Fast stable learning and categorization of analog pattern by an adaptive resonance system. Neural Networks, 1991, 4(6): 759–772.

CONSONANT RECOGNITION BASED ON FUZZY NEURAL NETWORK

Mei Yonng　　Wang Qunsheng　　Xu Bingzheng

(*South China University of Technology, Guangzhou* 510641)

Abstract　Conventional template matching technique is a successfully used algorithm in Chinese consonant recognition, yet its disadvantage limits its recovery of error,improvement of performance. The hybrid system based on the combination of neural network(NN) and fuzzy system maintains their advantages and makes full use of error tolerence performance, calculation performance, classification performance and decision performance for fuzzy neural network. In this paper, two kinds of cosonant recognition schemes based on fuzzy neural network are studied in detail. From the discussion of their structure,recognition rate and characteristics,it can be seen that recognition performance of fuzzy neural network is superior to template matching scheme and thus is more suitable for speech recognition.

Key words　Consonant recognition, Fuzzy system, Neural network

梅　勇：　男，1970 年生，博士生，从事神经网络的研究工作.

王群生：　男，1939 年生，教授，主要从事数字电视，HDTV 及图像通信等方面的教学和科研工作.