

小波网络和 RBF 网络的抗噪语音识别

侯雪梅

HOU Xue-mei

西安邮电学院 信息与控制系, 西安 710121

Department of Information & Control, Xi'an Institute of Post and Telecommunication, Xi'an 710121, China

E-mail: hxmqz@sina.com

HOU Xue-mei. Noise-robust speech recognition based on wavelet network and RBF network. Computer Engineering and Applications, 2009, 45(19): 150-152.

Abstract: To solve the problem that recognition rates of speech recognition systems decrease in the noisy environment presently, uses character possessing wavelet neural network which integrates the good time-field local property of wavelet transform, uses character possessing RBF neural network, which have best classification ability and recognize ability etc. This paper constructs a wavelet-RBF neural network structure using Morlet mother-wavelet as wavelet basis instead of activate function in the RBF network, adopts whole supervision algorithm and realizes a noise-robust speech recognition system based on wavelet network RBF network. The experiment results show that the system has better identify effect than RBF network especially has stronger Robust under noisy environment.

Key words: speech recognition; RBF neural network; wavelet neural network

摘要: 针对目前在噪音环境下语音识别系统性能较差的问题, 利用小波神经网络融合了小波变换良好的时频局域化性质和 RBF 神经网络具有最佳分类能力和辨识能力等特性。构建了一个用小波基替代 RBF 网络中激活函数的小波-RBF 神经网络结构, 并采用全监督训练算法, 实现了基于小波-RBF 网络的抗噪语音识别系统。实验结果表明该系统比 RBF 网络具有更好的识别效果, 尤其在噪声环境下, 具有更强的鲁棒性。

关键词: 语音识别; RBF 神经网络; 小波神经网络

DOI: 10.3778/j.issn.1002-8331.2009.19.046 **文章编号:** 1002-8331(2009)19-0150-03 **文献标识码:** A **中图分类号:** TP391

1 前言

语音识别实用化所遇到的难题之一是在噪声环境下获得与净音环境下相接近的识别结果。语音识别在实现过程中通常涉及多种因素, 需要同时考虑。由于语音信号的随机性, 以及对人类听觉机理了解甚浅, 因此, 目前在噪音环境下语音识别系统尚不能达到全部实用化, 语音识别实用化研究一直是业界的焦点。

目前应用于语音识别领域的识别网络主要有隐马尔可夫模型(HMM)和人工神经网络(ANN)。ANN 作为模拟人脑思维机制的工程模型, 其多维非线性映射能力泛化能力和并行信息处理能力已得到举世公认。其中 RBF 神经网络除了具有一般神经网络的优点, 还具有很强的分类决策能力, 它能将语音的动静态特性和听觉感知特性融合到网络特性之中^[1]。但是 RBF 网络中的激活函数不具有时频局部特性, 这使得网络在训练时由于出现冗余而降低收敛速度, 而小波神经网络融合了小波变换良好的时频局域化性质, 小波基对非平稳随机信号具有良好的时域局部特性和变焦能力。本文构建了一个用小波基替代 RBF 网络中激活函数的小波-RBF 神经网络模型作为识别网

络, 并采用符合人耳听觉特性的 Mel 频率与 LP 倒谱系数结合起来的 LP Mel 倒谱 (Linear Predictive Mel Cepstral Coefficients, LPMCC) 作为提取的语音特征参数, 在 VC 平台上实现了基于小波-RBF 网络的孤立词语音识别系统。在不同 SNR 下进行识别, 实验结果表明该系统较 RBF 网络抗噪性能较强, 识别效果好。

2 LPMCC 特征提取

LP 倒谱系数 (Linear Predictive Cepstral Coefficients, LPCC) 是最常用的特征参数。LPCC 是按实际频率尺度的倒谱系数, 而人耳所听到的声音的频率与实际频率并不成线性正比关系。通过实验研究发现, 基于人耳的听觉模型提取的特征参数更具有优于其他参数的鲁棒性。Mel 频带划分是对人耳听觉特性的一种工程化模拟, 人的听觉感知除了音调高低的感知外, 还有对响度的感知, 人对声音响度的感知与声音的频宽有关。Mel 频率尺度是将声音的频率非线性映射 (或称弯折) 到一个新的频率尺度, 可较充分地反映人耳对频率及幅度的非线性感知特性, 以及人耳在听到复杂声音时所表现的频率分析和谱合成特

基金项目: 西安邮电学院中青年教师项目科研基金 (No. 110-0417)。

作者简介: 侯雪梅 (1976-), 女, 助教, 主要研究方向: 语音识别。

收稿日期: 2008-04-16 **修回日期:** 2008-07-07

性。根据人类听觉系统对频率及幅度的感知实验结果,在此尺度下提取语音特征,更符合人耳的听觉特性^[2]。因此将常规的 LPC 进一步按符合人耳听觉特性的 Mel 尺度进行非线性变换,得到 LP Mel 倒谱系数(LPCMCC)作为特征参数^[3]。既考虑了声道激励,又兼顾人耳听觉,具有更强的抗噪性。

3 小波网络和 RBF 网络模型的设计

3.1 网络结构

小波神经网络是结合小波变换理论与人工神经网络思想而构造出来的一种“新型的神经网络模型”,融合了小波变换良好的时频局域化性质及神经网络的自学习功能。RBF 神经网络具有很强的分类决策能力,它可将语音的动静态特性和听觉感知特性融合到网络特性之中^[4]。构建了一个用小波基代替 RBF 网络中激活函数的多输入多输出的三层小波-RBF 网络结构,其拓扑结构如图 1 所示。

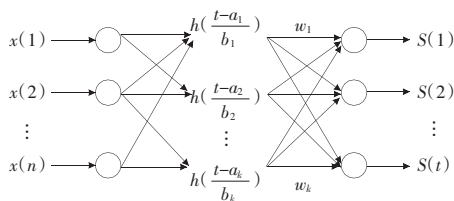


图 1 多输入多输出的小波-RBF 网络结构

输出函数 $S(t)$ 可用小波基函数进行如下拟合 $S(t) = \sum_{k=1}^k w_k h(\frac{t-a_k}{b_k})$, 式中 b_k 为尺度参数, a_k 为平移参数, k 为基函数个数。该网络中的 a_k 、 b_k 是不固定的, 仅 w_k 作为隐层与输出层之间的权值, $h(\frac{t-b_k}{a_k})$ 可看作输入节点的输出值。可以看出该小波-RBF 网络的表达式同 RBF 网络结构基本相同, 关键是隐层函数的差别。

网络的隐层节点数的选取没有统一的标准。隐层节点数太少, 网络的分类效果不好; 节点数太多, 会增加训练的复杂性。当训练样本词汇量不大时, 可用训练词汇数作为隐层节点数, 即网络中隐层节点数根据识别词汇量变化, 例如对 10 词汇的实验, 网络结构中隐节点定为 10 个; 50 词汇的实验, 隐节点数为 50。当然我们也可以采用自适应的方法, 开始将隐层节点数设为 1, 如果不能达到要求, 将隐层节点数加 1 直到满足要求, 这种方法可以找到最少隐层节点数来构成小波神经网络。在实验中均采用前一种方法确定隐节点数。输出层节点是待分类的模式类别总数, 即为待识别词汇数。如果对 40 词汇的识别, 那么相应地就有 40 种待分类模式, 输出层节点数等于 40。

3.2 小波基的选取

在 Hilbert 空间 $L_2(R)$ 中选取 $\Psi(x)$ 作为母小波, $\Psi(x)$ 应满足允许条件^[4]:

$$\int_{-\infty}^{\infty} \frac{\Psi(w)}{w} dw < +\infty \quad (1)$$

式中: $\Psi(w)$ 为 $\Psi(x)$ 的傅里叶变换。 $\Psi(x)$ 作伸缩和平移得到的函数族:

$$\Psi_j^k(x) = a_0^{-j/2} (\Psi(a_0^{-j} x - b_0^k)), j, k \in Z \quad (2)$$

适当选择 $a_0 > 0, b_0 > 0$, 当具有如下性质时:

$$Af^2 \leq \sum_{l,k} f, \Psi_{l,k}^2 \leq Bf^2, 0 < A < B < +\infty \quad (3)$$

便称 $\{\Psi_l^k; j, k \in Z\}$ 构成了 $L_2(R)$ 上的一个小波框架。同时, f 的小波基展开为:

$$f(x) = \sum_{i,k} C_{i,k}(f) \Psi_{i,k}(x) \quad (4)$$

其中: $C_{i,k}(f)$ 为待定的展开系数。从框架中选取有限个函数, 作为小波基函数, 得:

$$f(x) = \sum_{i,k} C_{i,k}(f) \Psi_{i,k}(x) \quad (5)$$

Z. Daubechies 已证明: 存在使式(5)以任意精度逼近 $f(x)$ 的小波基族。由此得小波神经网络结构, 其中 $\Psi_{i,k}(x)$ 便是隐层神经元激活函数^[4]。

由于 Morlet 母小波构成的小波基族满足 Z. Daubechies 框架理论。因为该小波是余弦调制的高斯波, 它的时频域同时分辨率高, 并且在许多实际问题中已经得到过很好的应用。小波变换的实质是一种不同参数空间之间通过小波基进行的积分变换。小波基是此变换的内核, 它的不同选择对变换起着关键的作用。本文采用 Morlet 母小波作为小波基函数。

3.3 网络训练

基于 Mooky 与 Darken^[5]算法的基本思想选用全监督训练算法。该算法的基本思路是网络的所有参数调整过程作为一个监督学习的过程, 同时加以调整, 以达到性能指标最小。

Morlet 母小波为:

$$h(t) = \cos(1.75t) \exp(\frac{-t^2}{2}) \quad (6)$$

网络参数 w_k, a_k 和 b_k 可以用最小方差能量函数进行优化。

定义:

$$E = \frac{1}{2} \sum_{i=1}^l [F(t_i) - S(t_i)]^2 \quad (7)$$

其中: l 是数据采样点个数, $F(t_i)$ 为实际信号值 $S(t_i)$ 为拟合信号值。

令:

$$t_1 = \frac{t - b_k}{a_k} \quad (8)$$

则对 w_k 的梯度为:

$$\frac{\partial E}{\partial w_k} = - \sum_{i=1}^l [F(t_i) - S(t_i)] \cos(1.75t_1) \exp(-\frac{1}{2}t_1^2) \quad (9)$$

对 a_k 的梯度为:

$$\frac{\partial E}{\partial a_k} = \frac{w_k (t_i - b_k) \exp(-\frac{1}{2}t_1^2)}{a_k^2} \times [1.75 \sin(1.75t_1) + t_1 \cos(1.75t_1)] \sum_{i=1}^l [F(t_i) - S(t_i)] \quad (10)$$

对 b_k 的梯度为:

$$\frac{\partial E}{\partial b_k} = \frac{w_k b_k \exp(-\frac{1}{2}t_1^2)}{a_k^2} \times [1.75 \sin(1.75t_1) + t_1 \cos(1.75t_1)] \sum_{i=1}^l [F(t_i) - S(t_i)] \quad (11)$$

训练前设定好网络初始值, 循环调节网络权值, 直到给定误差

E 小于预先给定的误差指标为止。

4 仿真实验及分析

用 VC 在 Windows 平台上实现了基于 LPMCC 特征和小波-RBF 网络的孤立词语音识别系统,图 2 给出了用小波-RBF 网络进行语音识别的原理框图,图中的实线表示语音识别的训练过程,虚线为语音识别的测试过程,即识别过程。

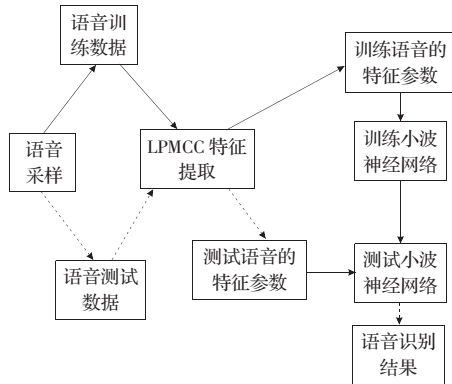


图 2 语音识别流程框图

语音信号采样率为 11.025 kHz, 帧长 $N=256$ 点, 帧移为 5 ms。由于每个词的语音波形文件长度各不相同, 本网络中小波神经网络的输入节点数是固定的, 所以为了处理方便, 将提取的特征参数进行时间归一化处理, 即每个单词的每一次发音最后得到统一的 1024 维的语音特征矢量序列, 作为小波网络的输入。实验共使用了 10 词、20 词、30 词、40 词、50 词, 共 16 个人在不同 SNR 下的语音数据文件(包括无噪音、15 dB、20 dB、25 dB、30 dB 的数据)。每人每个词发音 3 次, 实验采用 9 个人(某种 SNR)的发音训练成数据库, 用另外 7 个人(同一 SNR 下的)的发音进行识别, 得到不同 SNR 下和不同词汇量下的孤立词语音识别结果, 并与相同环境下的 RBF 网络的全监督训练算法结果进行比较。表 1 列出了在不同 SNR 和不同词汇量下的小波-RBF 神经网络的识别结果。

从表 1 中首先可以看到, 小波-RBF 用于语音识别得到了较好的识别率。比较两种网络的识别结果, 可以看到用小波基替代 RBF 网络中激活函数的小波-RBF 网络的识别率明显高于 RBF 网络。充分说明了结合了小波网络和 RBF 网络的识别网络抗噪性能会更好, 并且使用全监督训练算法对网络的性能提高有较大的作用, 使小波-RBF 具备了更强的分类能力和辨

表 1 基于 LPMCC 特征的识别算法的识别率比较 (%)

词汇量 (词)	训练方法	SNR				
		15 dB	20 dB	25 dB	30 dB	Clean dB
10	RBF	84.62	85.21	85.36	85.53	86.22
	小波-RBF	86.31	87.67	87.88	89.29	91.16
20	RBF	85.16	85.89	86.79	87.24	88.57
	小波-RBF	88.35	89.47	90.25	91.44	92.01
30	RBF	86.82	88.23	89.16	89.89	91.18
	小波-RBF	89.68	90.16	90.58	91.37	92.23
40	RBF	88.37	89.37	89.96	90.26	92.13
	小波-RBF	90.14	91.88	92.37	92.96	93.87
50	RBF	90.56	92.37	92.54	93.11	94.12
	小波-RBF	91.88	92.77	92.85	93.64	93.86

识能力。但缺点是训练速度较慢, 可在以后的研究中加以改进。

在实验中发现小波-RBF 网络对小波系数初值选取比较敏感, 小波参数初始化目前还缺少严格的理论指导。训练中也存在局部极值问题, 可考虑用选用正交小波函数或尺度函数来克服。此外, 采用不同的母小波和不同的误差能量函数, 有可能减小训练量, 加快网络收敛速度, 但这还需要进一步的实践和不断的探索。

5 结论

小波神经网络融合了小波变换良好的时频局域化性质, 小波基对非平稳随机信号具有良好的时域局部特性和变焦能力。RBF 神经网络具有很强的分类决策能力和自适应分辨性, 将小波基代替 RBF 网络中的径向基函数的小波-RBF 网络用于语音识别系统, 使系统具有更强的辨识能力和良好的容错能力, 具有良好的应用前景。

参考文献:

- [1] 鄢田云, 云霞. RBF 神经网络及其在基于输出的客观音质评价中的应用[J]. 电子学报, 2004, 32(8): 1282-1285.
- [2] 付强, 易克初. 一种采用余弦镶边临界带滤波器组的弯折谱失真测度[J]. 西安电子科技大学学报, 1999, 26(6): 823-827.
- [3] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2003.
- [4] 李弼程, 罗建书. 小波分析及其应用[M]. 北京: 电子工业出版社, 2003.
- [5] Han M, Xi Jian-hui. Efficient clustering of radial basis perception neural network for pattern recognition[J]. Pattern Recognition, 2004, 37(10): 2059-2067.

(上接 130 页)

- [11] 宫改云, 毛用才, 高新波, 等. 基于模糊 C-均值聚类的微阵列基因表达数据分析[J]. 西安电子科技大学学报: 自然科学版, 2004, 31(2): 291-295.
- [12] 易东, 杨梦苏, 李辉智, 等. 基因表达数据聚类分析结果的评价方法研究[J]. 中国卫生统计, 2002, 19(6): 332-335.
- [13] Wu Chung-chiang. Statistical analysis of parameters of yeast cell

cycle regulated genes[D]. Taipei: National Sun Yat-Sen University, 2004.

- [14] Spellman P T, Sherlock G, Zhang M Q, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization[J]. Molecular Biology of the Cell, 1998, 9(12): 3273-3297.