

基于过程神经网络的语音识别技术研究

田 丽, 刘英楠, 孟耀华

(黑龙江八一农垦大学信息技术学院, 黑龙江 大庆 163319)

摘 要:过程神经网络是一种基于过程神经元的新型神经网络,其输入及权值皆为时序函数。针对语音识别的特点,对过程神经网络在语音识别技术中的应用进行了探索性的研究,并与传统的BP神经网络、径向基函数网络进行了比较。仿真结果表明,采用过程神经网络进行的语音识别,识别性能得到了提高。

关键词:过程神经网络;语音识别;BP算法

中图分类号:TP183

文献标识码:A

Research of speech recognition based on process neural network

TIAN Li, LIU Ying-nan, MENG Yao-hua

(College of Information and Technology, Heilongjiang Bayi Agricultural University, Daqing 163319, China)

Abstract: Process Neural Networks (PNN) is a kind of new ANN based on Process Neuron, its inputs and weights are all functions associated with "time". Aim at the features of speech recognition, exploratory study the application of neural network on the process in speech recognition technology, and compares with the traditional BP neural network, radial basis function networks. Experiment results show that process neural network is used in speech recognition, recognition performance is promoted.

Key words: process neural network; speech recognition; BP algorithm

1 引 言

语音识别 SR (speech recognition) 是指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^[1]。现代语音识别技术以神经网络为主要发展趋势,进入20世纪90年代以来,神经网络已经成为语音识别的一条重要途径。由于神经网络反映了人脑功能的基本特征,故具有自组织性、自适应性和连续学习能力等特点,特别适合于解决象语音识别这类模拟人的认知过程和智能处理能力,难以用算法来描述而又有大量样本可供学习的问题。

过程神经网络是近几年出现的应用价值很强的新型智能网络。网络的训练学习是应用一系列输入矢量,通过已确定的算法逐步调整网络的权值,最终达到期望的目标。传统的BP(back propagation)算法的学习过程由信号的正向传播和误差的反向传播

组成。随着误差逆传播修正的不断进行,网络对输入模式响应的正确率不断上升。然而BP算法是一种梯度下降算法,梯度下降存在多极值问题,且BP网络学习训练速度慢,容易陷入局部最小或振荡。因此,参数的选择非常重要。为了克服传统神经网络BP算法存在的各种缺点,本文研究用过程神经网络结合改进的BP算法来训练网络进行语音识别^[2]。

2 过程神经元网络模型

2.1 过程神经元

过程神经元由多输入时变函数加权、时空二维聚合以及阈值激励输出等运算组成。聚合运算包括

基金项目:黑龙江省科技攻关项目(No. 2009G0085-00)资助。

作者简介:田 丽(1978-),女,讲师,主要从事通信工程专业方面的教学与科研工作,研究方向是图像及语音信号处理方向。
E-mail: tianli19781015@163.com

收稿日期:2010-05-07

对多维时变输入信号的空间加权聚合以及对时间过程效应的累积。若空间聚合算子取为加权和 Σ ,时间累计算子取为积分 $\int$,则过程神经元的结构如图1所示。

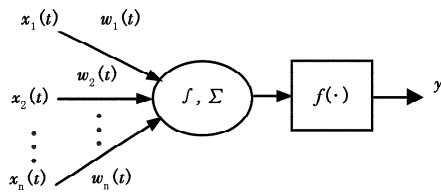


图1 过程神经元

由图1可知,过程神经元输入输出之间的映射关系为:

$$y = f\left(\int_0^T \sum_{i=1}^n w_i(t) x_i(t) dt - \theta\right) \quad (1)$$

式中, $x_1(t), x_2(t), \dots, x_n(t)$ ($t \in [0, T]$) 为过程神经元的时变输入函数; $w_1(t), w_2(t), \dots, w_n(t)$ 为连接权函数; $[0, T]$ 为时变信号过程输入区间; f 为过程神经元激励函数; y 为过程神经元输出^[3]。

2.2 过程神经网络模型

图2中,输入层有 n 个节点;中间隐层包含 m 个过程神经元节点,激励函数为 f ;输出层为1个非时变神经元节点。网络输入与输出之间的映射关系为:

$$y = \sum_{j=1}^m v_j f\left(\int_0^T \left(\sum_{i=1}^n w_{ij}(t) x_i(t)\right) dt - \theta_j\right) - \theta \quad (2)$$

式中, θ_j 为隐层过程神经元的阈值; θ 是输出神经元的阈值。

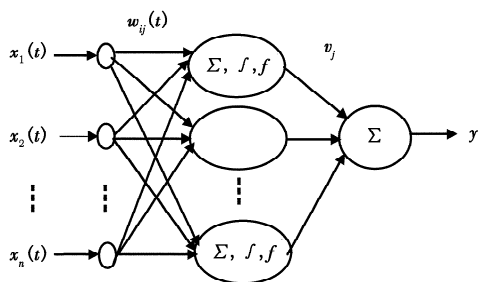


图2 含一个隐层的过程神经网络

为不失一般性,设网络输入函数、连接权函数均为 $[0, T]$ 上的连续时间函数。从数学意义上讲,在连续函数空间 $C[0, T]$ 中有多个基函数系存在,函数空间中的任意函数在给定的拟合精度下均能表示为该组基函数的有限项展开形式。在 $C[0, T]$ 空间中选择一组适当的函数基 $b_1(t), b_2(t), \dots, b_l(t), \dots$,将网络输入函数空间中的函数 $X(t) = (x_1(t), x_2(t), \dots, x_n(t))$ 在满足拟合精度的条件下进行有限项展开:

$$x_i(t) = \sum_{l=1}^L x_i^{(l)} b_l(t) \quad (3)$$

其中, L 为满足输入函数拟合精度要求的基函数的项数; $x_i^{(l)}$ 为 $x_i(t)$ 相对于 $b_l(t)$ 的展开系数。

将权函数 $w_{ij}(t)$ 表示为 $b_1(t), b_2(t), \dots, b_L(t)$ 的基函数展开形式:

$$w_{ij}(t) = \sum_{l=1}^L w_{ij}^{(l)} b_l(t) \quad (4)$$

式中, $w_{ij}^{(l)}$ 为 $w_{ij}(t)$ 相对于 $b_l(t)$ 的展开系数。把式(3)和式(4)代入式(2),并简化网络模型为线性输出,则含一个过程神经元隐层的权函数基展开过程神经网络可写为:

$$y = \sum_{j=1}^m v_j f\left(\int_0^T \left(\sum_{i=1}^n \left(\sum_{l=1}^L w_{ij}^{(l)} b_l(t)\right) x_i(t)\right) dt - \theta_j\right) \quad (5)$$

式中, $v_j, w_{ij}^{(l)}, \theta_j$ 为网络待定参数。

3 过程神经网络在语音识别中应用

采用过程神经网络进行语音识别的过程包括两个方面:一是网络学习过程,二是语音识别过程。网络学习过程是将已知语音信号作为学习样本,通过神经网络的自学习,最终得到一组连接权和偏置。而语音识别过程则是将待测语音信号作为网络输入,通过网络的联想得出识别结果。

3.1 神经网络的建模方法

这里先对训练语音信号进行了同态滤波处理,形成了语音识别的两种特征参数矢量,一种是LPC倒谱参数,第二种是Mel倒谱参数(MFCC)。将特征提取后的数据进行压缩,规整到特定帧数,以此作为神经网络的输入。具体分别采用了三种神经网络作为语音识别的模型,即BP网络、径向基函数网络以及过程神经网络^[4]。

(1) BP网络模型和径向基函数网络模型

这两类网络模型的结构类似,其有关参数的设置如下:

输入为4帧8维的特征矢量参数,故输入层有32个神经元;

输出层有10个神经元,分别对应于0~9这10个数字;

隐层的神经元个数选择通过仿真试验来确定。一般来说,可用以下几个公式作为参考:

$$\textcircled{1} k < \sum_{i=0}^n C \binom{n_1}{i}, \text{ 其中, } k \text{ 为样本数, } n \text{ 为输入神经元数。}$$

$$\textcircled{2} n_1 = \sqrt{m+n} + a, \text{ 其中, } m \text{ 为输入神经元数,}$$

a 为 1~10 之间的常数。

③ $n_1 = \log_2 n$, n 为输入神经元数。

(2) 过程神经网络模型参数的确定

过程神经网络是由若干个过程神经元按一定的结构关系组成的网络。这里首先建立一个三层前馈过程神经网络模型如图 2 所示,选择 10 个输入层节点,32 个隐层节点,10 个输出层节点,分别对应于 0~9 这 10 个数字,学习效率常数均为 0.1,最大学习次数 10000,学习精度 0.01,训练样本 5 个,测试样本 7 个,隐层激励函数取 Sigmoid 函数,输出层激励函数取线性函数。

3.2 神经网络的识别方法的性能分析

对于识别方法和软件而言,在保证一定识别率的条件下,自然是所需要的训练样本越少越好,所以说对样本个数的需求是判断语音识别方法和软件系统的一个基本指标^[5]。下面以特定人语音“3”为例,分别采用 BP 网络、径向基函数网络和过程神经网络进行了训练识别,实验结果如图 3 和表 1 所示。

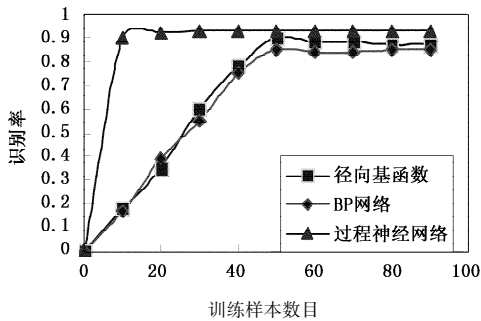


图3 训练样本数对识别率的影响

表 1 特定人语音“3”的识别结果对比表

神经网络模型	训练次数	训练时间/s	识别率/%
BP 神经网络	389	30	90
径向基函数网络	259	16	92
过程神经网络	218	12	95

从图 3 中可以看出:BP 网络和径向基函数网络对样本的需求量基本相当,不同的是径向基函数网络的识别率较 BP 网络略高,而过程神经网络在同等样本数的条件下,识别率明显高于其他两种。

4 结 论

采用神经网络进行的语音识别是目前研究的热点,本文针对语音识别的特点,将过程神经网络引入到语音识别的模型中,并与传统的 BP 神经网络和径向基函数网络进行比较和性能分析,仿真实验结果表明,采用过程神经网络进行的语音识别,训练时间变短,训练次数减少,识别率有所提高。

参考文献:

[1] 陈方,高升. 语音识别技术及发展[J]. 电信科学, 1996,10:23-25.

[2] 袁曾任. 人工神经网络及其应用[M]. 北京:清华大学出版社,1999:75-88.

[3] 何新贵,梁久祯,许少华. 过程神经网络的训练与应用[J]. 中国工程科学,2001,3(4):31-35.

[4] 闫雪梅,王晓华. 基于 PCA 和 BP 神经网络算法的车牌字符识别[J]. 激光与红外,2007,37(5):481-484.

[5] 谢秋菊,苏中滨. 电力语音充值系统呼叫处理过程的多线程实现[J]. 黑龙江八一农垦大学学报,2008,(1):46-48