

文章编号:1009-671X(2008)02-0017-04

一种基于优化小波神经网络的语音识别

陈立伟,宋宪晨,章东升,杨洪利

(哈尔滨工程大学 信息与通信工程学院,黑龙江 哈尔滨 150001)

摘 要:在以往的 BP 小波神经网络中,最常用的学习算法是 BP 算法,BP 算法实质上就是梯度下降法,是一种局部搜索算法,梯度下降法使得网络极易陷入局部最小值,从而使得网络训练结果不尽人意,搜索成功概率低.取代传统的梯度下降法,利用粒子群算法对小波神经网络中的参数进行优化.然后利用基于粒子群优化(PSO)的小波神经网络进行抗噪声语音识别实验,仿真结果表明,与 BP 网络相比,PSO 算法在迭代次数、函数逼近误差、网络性能方面均优于 BP 网络,系统的识别率也得到较大的提高.

关键词:粒子群优化;小波神经网络;语音识别;抗噪声

中图分类号:TN911.7 **文献标识码:**A

Speech recognition using an optimized wavelet neural network

CHEN Li-wei, SONG Xian-chen, ZHANG Dong-sheng, YANG Hong-li

(College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: The traditional BP wavelet neural networks usually adopt BP learning algorithm. BP algorithm is a gradient descent algorithm in essence, i. e. , a local search algorithm. The gradient descent algorithm is easy to fall into a local minimum, so the result of network training is unsatisfactory. Instead of the gradient descent algorithms, we use particle swarm optimization algorithm to train the parameters of the wavelet neural network. Then we use the PSO WNN in speech recognition in noisy environment. Compared with the BP network, the simulation results show that the iterative number, function approximation errors and the network performance are highly improved than BP network. The recognition rate are also raised.

Keywords: particle swarm optimization; wavelet neural network; speech recognition; speech recognition in noisy environment

目前,人工神经网络已经成为语音识别领域的一个研究热点,人工神经网络(ANN)作为模拟人脑思维机制的工程模型,其分类决策能力和对不确定性信息的描述能力已得到举世公认.

小波神经网络(wavelet neural network)是基于小波分析所构造的一种新的神经网络模型,它结合小波变换所具有的良好的高频域时间精度、低频域频率精度的性质和神经网络的自学习功能,因而具有较强的逼近、容错能力.

粒子群优化(particle swarm optimizer, PSO)算法是一类新兴的随机全局优化技术,因此是一类有着

潜在竞争力的神经网络学习算法^[1]. 群体智能优化算法的仿生基点是:群集动物通过群聚而有效地觅食和躲避追捕,这些群集动物所表现出的智能称为“群集智能”,它可概括表述为:一组相互之间可以进行直接通讯或间接通讯的主体,能够通过合作对问题进行分布式求解. 本文将粒子群优化算法应用于小波神经网络的设计,应用所设计的小波神经网络进行针对含噪汉语语音的识别实验,并与基于 BP 算法训练的小波神经网络进行识别结果比较,仿真实验结果证明了本文所提出的方法的有效性.

1 小波神经网络

小波神经网络是基于小波分析而构造的神经网络模型. 小波变换定义如下:

设 $\psi(t) \in L^2(R)$, ($L^2(R)$ 表示平方可积的实数空间, 即能量有限的信号空间), $\psi(t)$ 的傅里叶变换为 $\Psi(\omega)$. 当 $\Psi(\omega)$ 满足如下条件时,

$$C_{\Psi} = \int_R \frac{|\Psi(\omega)|}{|\omega|} d\omega < \infty, \quad (1)$$

称 $\psi(t)$ 为一个基本小波或母小波, 将母函数 $\psi(t)$ 经伸缩和平移后, 就可以得到一个小波序列:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right), \quad a, b \in R; a \neq 0. \quad (2)$$

式中: a 为伸缩因子, b 为平移因子.

对于任意的函数 $f(t) \in L^2(R)$, 其小波变换定义为

$$W_f(a, b) = \langle f, \psi_{a,b} \rangle = |a|^{-1/2} \int_R f(t) \psi\left(\frac{t-b}{a}\right) dt. \quad (3)$$

其逆变换(重构公式)为

$$f(t) = \frac{1}{C_{\psi}} \int_R \frac{1}{a^2} W_f(a, b) \psi\left(\frac{t-b}{a}\right) da db. \quad (4)$$

将式(4)离散化, 可以得到

$$f(t) = \sum_{i=1}^k \omega_i \psi\left(\frac{t-b_i}{a_i}\right). \quad (5)$$

式中: k 为小波基的个数. 式(5)可以理解为原信号函数由加权小波基进行线性叠加来实现. 一个 3 层的前向神经网络的输出表达式为

$$y = f\left(\sum \omega_i x_i - \theta_i\right). \quad (6)$$

式中: i 表示第 i 个神经元, $f(x)$ 为神经元激励函数, ω_i 为加权系数, x_i 为输入, θ_i 为偏移量.

对比式(5)和式(6), 可以看出, 构造一个 3 层的神经网络, 通过将所选取的小波基进行线性加权叠加来实现对非线性函数的逼近, 即式(5). 在一定的误差范围内, 总可以选择一组数量最少的小波基, 利用式(5)得到原始信号的拟和形式, 从而构造出小波神经网络. 小波神经网络的网络结构图如图 1 所示.

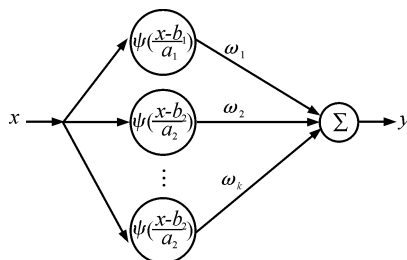


图 1 小波神经网络结构图

2 粒子群算法

与其他优化算法相似, 粒子群优化算法同样基于群体(粒子群)与适应度, 通过适应度将群体中的个体(粒子)移动到好的区域^[2]. 粒子代表问题的一个可能解, 它具有位置和速度 2 个特征, 每个粒子都被看做是 d 维空间中的一个没有体积和质量的微粒, 在搜索空间中以一定的速度飞行, 并根据对个体和集体飞行经验的综合分析来动态调整这个速度, 然后更新个体当前的位置. 算法首先初始化一群随机粒子, 然后通过迭代找到最优解. 在每一次迭代中, 粒子通过跟踪 2 个极值来更新自己: 一个是粒子本身所找到的最优解, 即记为个体极值; 另一个是整个粒子群目前找到的最优解, 即记为全局极值.

设 S 为 d 维欧氏空间的一个非空集合, 粒子群中第 i 个粒子在 S 中的位置记为 $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$, 粒子 i 的速度记为 $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$, 粒子在找到上述 2 个极值后, 就根据下面 2 个公式来更新自己的速度与位置^[3].

$$v_d^{k+1} = \omega v_d^k + C_1 \text{Rand}() (pbest_i^k - x_{id}^k) + c_2 \text{Rand}() (gbest_i^k - x_{id}^k), \quad (7)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^k. \quad (8)$$

式中: k 为迭代次数, ω 为惯性权重(inertia weight); 它使粒子保持运动的惯性, 使其有能力探索新的区域; c_1 和 c_2 为加速常数, 它们使每个粒子向 pbest 和 gbest 位置加速运动; Rand 为 $[-1, 1]$ 范围内变化的随机数. 此外粒子的速度 V_i 被一个最大速度 V_{\max} 限制. 如果当前粒子在某维的速度 v_{id} 超过该维的最大速度 $v_{\max d}$, 则该维的速度被限制为最大速度 $v_{\max d}$. V_{\max} 决定了粒子在解空间的搜索精度, 如果 V_{\max} 太大, 粒子会飞过最优解, 如果 V_{\max} 太小, 粒子则陷入局部最优解而无法进行全局搜索. 图 2 是粒子群优

化算法流程图.

公式(7)和(8)表示了粒子在求解空间中,由于相互影响导致了运动速度和位置的调整. 惯性权重 ω 、加速常数 c_1 和 c_2 和最大速度 V_{\max} 共同维护粒子对全局和局部搜索能力的平衡,因此算法的参数设置与算法系统性能之间有着密切的联系. 在实际应用中算法参数的设置与具体问题密切相关,目前比较常用的参数设置方法是针对具体求解问题,利用充分的实验来确定^[4-5].

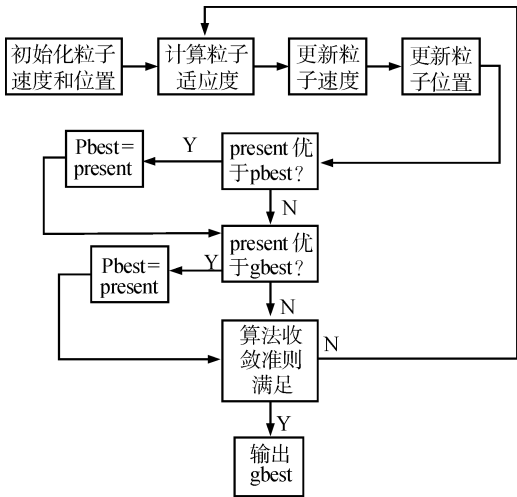


图2 粒子群优化算法流程图

3 应用粒子群算法来训练神经网络

在以往的BP小波神经网络中,最常用的学习算法是BP算法,BP算法实质上就是梯度下降法,是一种局部搜索算法,即利用梯度作为启发信息在局部范围沿最优方向搜索局部最优点. 然而,梯度下降法使得网络极易陷入局部最小值,从而使得网络训练结果不尽人意,搜索成功概率低,这使得BP网络在应用中存在着一定的局限性,如何提高网络的搜索成功率,选用新的算法替代梯度下降法是研究的重点.

基于梯度下降法的缺陷及上述粒子群算法的原理和优点,文中采用粒子群算法替代梯度下降法来对BP小波神经网络进行参数训练. 按照图1所示的网络结构建立小波神经网络,粒子数目初始化为100个. 每个粒子的位置向量为

$$\text{present}(i) = [\omega_1, \cdots, \omega_j, a_1, \cdots, a_j, b_1, \cdots, b_j],$$
$$i = 1, 2, \cdots, 100$$

式中; j 为隐含层神经元个数. 粒子的适应度定义为第 k 次迭代后网络实际输出 \hat{y} 和理想输出 y 间的误

差平方和:

$$J(k) = \sum_{m=1}^n c_m^k = \sum_{m=1}^n (y_m - \hat{y}_m^k)^2, k = 1, 2, \cdots, D.$$

(9)

式中: n 为非线性函数的采样点数, \hat{y}_m^k 表示对于第 k 次迭代第 m 个输入的网络实际输出, D 为最大迭代次数. 仿真步骤如下:

1) 初始化100个粒子的位置向量、速度向量 V 、目标误差值 $E_{\text{goal}} = 1.0$ 及最大迭代次数 T_{\max} , 其中每个粒子向量的元素随机产生. 设定各个粒子位置向量元素 $\omega_j, a_j, b_j, j = 1, 2, \cdots, j^*$ 的最大值和最小值,设定速度向量 $V_m, m = 1, 2, \cdots, 100$ 的最大值和最小值. 对于伸缩量 a_j 的初始化为:首先在区间 $[0, 1]$ 随机产生 j^* 个 a_j ,然后将位于 $[0, 0.5]$ 中的那些伸缩量 a_j 映射到区间 $[0, 1]$,将 $[0.5, 1]$ 中的 a_j 映射到 $[1, 5]$,这样做的原因在于 a_j 是伸缩量,若 a_j 都在 $[0, 1]$ 中则只起到压缩的作用.

2) 利用粒子群算法,式(7)和式(8)对每个粒子的位置向量 present 和速度 V 进行迭代更新,并且记录每个粒子的历史最优位置 pbest_i^k (pbest_i^k 表示第 i 个粒子第 k 次迭代的历史最优位置向量)和所有粒子中的全局最优位置向量 gbest^k . 另外还需记录对应于 pbest_i^k 和 gbest^k 的系统适应度值 $E_{\text{pbest}_i^k}$ 和 E_{gbest^k} .

3) 若 $E_{\text{gbest}^k} \leq E_{\text{goal}}$ 或者最大迭代次数达到了 T_{\max} , 训练结束. 否则转第2)步.

4) 利用 gbest (最终得到的全局最优值) 计算网络输出得到最终的拟合曲线.

仿真中,加速因子 $c_1 = c_2 = 1.494$, 惯性因子 ω 选为0.7.

将基于粒子群方法训练的小波神经网络和传统的BP小波网络进行了比较,误差目标为1.0,最大叠代次数为500次,隐含层为9个神经元,对2种算法分别进行1000次仿真,每次训练结束后记录其最终网络误差. 结果如图3和图4所示.

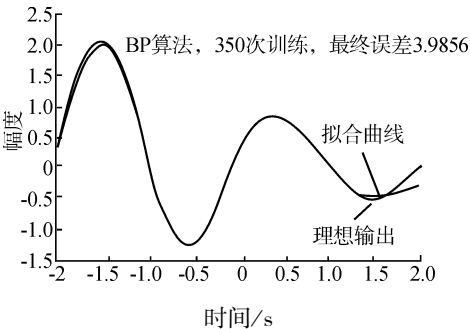


图3 BP小波神经网络的拟合结果

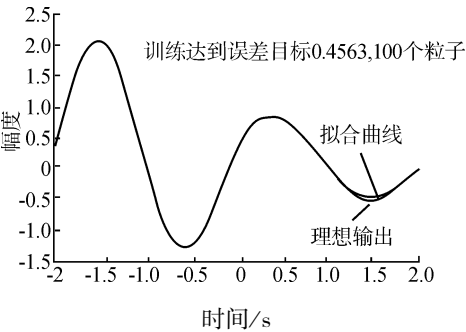


图4 粒子群方法训练的小波神经网络拟合结果

可以看到,粒子群算法的搜索成功率大于BP算法,也即其陷入局部极小值的概率小.而且,粒子群算法收敛速度快,且在粒子数目较大时能有效避免在BP网络中无法避免的局部极小值问题.当然,随着粒子数的增加,计算量也随之增大,在本文中,粒子数取为100是经过大量试验得到的经验值,能完成给定的误差目标.

4 仿真实验

实验数据是由男声数据组成,在实验室环境下录制,50个人的录音,每人300~450句话,将在此环境下得到的语音视为纯净语音,然后将语音数据分别混以不同信噪比的加性噪声,形成信噪比为5、10、15、20 dB的带噪信号.语音采样率为11.025 kHz,其中前45人的数据作为训练集,后5人的数据作为测试集.每个语音文件分帧长度为20 ms,帧移10 ms.语音特征参数为每帧语音的归一化能量及14维MFCC系数连同它们的一阶差分、二阶差分系数共同构成了45维的特征矢量.

本文的语音识别系统的声学层识别性能指标如下:假设 H 、 D 、 I 、 S 、 T 分别表示声学层的正确识别音节数、删除音节数、插入音节数、替代音节数、实际音节数,则

$$\text{正确识别率}(R) = \frac{H}{T} \times 100\%$$
$$\text{错误识别率}(W) = \frac{D + I + S}{T} \times 100\%$$

实验结果如表1所示.

表1 2种模型的识别率比较				%
	平均正确 识别率	平均错误 识别率	平均正确 识别率	平均错误 识别率
PSO-WNN	37.65	62.35	74.37	25.63
BP-WNN	34.26	65.74	70.18	29.82

由表1的结果可以看出,与BP小波神经网络相比较,基于粒子群优化的小波神经网络的识别率要更高.

5 结束语

将粒子群优化算法应用于小波神经网络的设计,应用所设计的小波神经网络进行噪声语音识别实验,并与基于BP算法训练的小波神经网络进行识别结果比较,仿真实验结果证明了所提出的方法的有效性.但是,粒子群算法也同样存在着局部极小值问题,如何减小陷入局部极小值的概率还需要进一步的对算法进行改进.

参考文献:

[1] SHI Y, EBERHART R C. A modified swarm optimizer [C]//IEEE International Conference on Evolutionary Computation. Anchorage, USA, 1998.

[2] CLERK M, KENNEDY J. The particle swarm-explosion, stability, and convergence in a multidimensional complex space [J]. IEEE Transaction on Evolutionary Computer, 2002, 6(1): 58-73.

[3] EBERHART R C, SHI Y. Particle swarm optimization: developments, applications and resources [C]//Proc Congress on Evolutionary Computation 2001, Piscataway, USA IEEE press, 2001.

[4] KENNEDY J, EBERHART R C. Particle swarm optimization [C]//IEEE International Conference on Neural Networks. [s. l], 1995.

[5] EBERHART R C, SHI Y. Comparing inertia weights and constriction factors in particle swarm optimization [C]//2000 Congress on Evolutionary Computation. La Jolla CA, USA, 2000.