

用遗传算法优化小波神经网络的结构

秦伟良¹⁾ 金龙²⁾ 李诗平³⁾

(1) 南京气象学院基础科学系, 南京 210044; 2) 江苏省气象局, 南京 210008; 3) 南京气象学院图书馆, 南京 210044)

摘要 用遗传算法确定小波神经网络中输入层单元数和隐含层单元数, 同时采用梯度法计算小波神经网络中的权系数、伸缩和平移系数, 从而达到优化小波神经网络的结构的目的。

关键词 遗传算法, 小波神经网络, 优化, 梯度法

分类号 O212

小波神经网络是小波分解和神经网络相结合而的产物, 它是一种具有输入层、隐含层和输出层的前馈神经网络^[1]。对于适当的小波函数和足够的隐层单元数, 该网络能够以任意精度逼近一个连续映射。但对给定映射, 如何决定隐单元数目, 还没有有效的方法。并且在网络学习时也遇到了巨大的困难^[2,3]。用小波网络实现映射逼近时, 确定隐单元数是至关重要的, 隐单元数目过多或过少将导致神经网络的学习能力不够或归纳能力很差。恰当地选择隐单元数, 则可提高网络对函数的逼近效果。遗传算法(Genetic Algorithm- GA)是由 J Holland 教授首先提出的, 它是模拟达尔文的遗传选择和自然淘汰的生物进化过程的计算模型。遗传算法采用同时处理群体中多个个体的方法, 即同时用适应度函数对搜索空间中的多个解进行评估, 并在此基础上进行遗传操作。而且适应度函数不受连续可微的约束, 仅要求该问题可计算, 因而遗传算法具有较好的全局搜索性能, 可以确定小波神经网络中输入层单元数和隐含层单元数。

1 小波神经网络

设基本小波为 (X) , 其相应的小波函数系为

$$n = \begin{cases} \overline{\det (D_j)} [D_j(X - t_j)] & t_j \in R^n \\ D_j = \text{diag} (1/s_j), & s_j \in R^+, j \in Z \end{cases} \quad (1)$$

由小波函数系在 $L^2(R^n)$ 中的稠密性^[1], 可构造如下形式的神经网络结构

$$G(X) = \sum_{i=1}^N w_i [D_i R_i(X - t_i)] + g \quad (2)$$

它是一个具有输入层、隐含层和输出层的前馈神经网络。其中输入层有 n 个单元, 相应于 $X = (x_1, x_2, \dots, x_n)$, 隐含层有 N 个单元, 相应于 N 个小波。设 θ 表示所有参数, $G(X)$ 表示(2)式给出的函数, 取目标函数为

$$C(\cdot) = \frac{1}{2} \sum_{k=1}^K [G(X_k) - y_k]^2 \quad (3)$$

其中 K 为训练数据总数, 对网络参数采用 BP 算法进行训练即可确定它们的值。但是输入单元数和隐单元数的多少影响着整个网络的逼近能力。一个好的结构应能圆满解决问题, 同时不允许冗余节点和冗余连接的存在。本文采用遗传算法对它们进行处理。

2 遗传算法

遗传算法(GA)的出现使神经网络的训练有了一个崭新的面貌, 目标函数对于输入单元数和隐单元数既不要求连续, 也不要求可微, 仅要求该问题可计算。而且它的搜索始终遍及整个解空间, 因此容易得到全局最优解。遗传算法中包含了 5 个基本要素: (1) 参数编码; (2) 初始群体的设定; (3) 适应度函数的设计; (4) 遗传操作设计; (5) 控制参数的设定。其中遗传操作算子包括选择(selection)、交叉(crossover)和变异(mutation)^[4,5]。

3 优化小波神经网络

用遗传算法来优化神经网络结构的步骤是 (1) 随机产生 M 个结构, 对每个结构编码, 每个编码个体对应一个结构; (2) 用许多不同的初始权值分布对个体的结构进行训练; (3) 根据训练的结果或其他策略确定每个个体的适应度; (4) 选择若干个适应度最大的个体, 直接继承给下一代; (5) 对当前一代群体进行交叉和变异等遗传操作, 以产生下一代群体; (6) 重复(2)~(5), 直到当前一代群体中的某个个体能够满足要求为止。

对于输入单元数和隐单元数, 采用二进制编码将它们表示成遗传空间的基因型串数据结构, 码长为 $2L$ 位, 其中高 L 位表示输入单元数, 低 L 位表示隐单元数。由于遗传算法的群体型操作需要, 我们为遗传操作准备一个由若干个初始解的初始群体, 群体规模取 M , 初始群体由随机方法产生。

对每个个体对应的小波网络中的权系数 w_i , 旋转和平移伸缩因子 R_i, D_i, t_i 的训练采用 BP 算法, 因为这些参数的梯度信息是已知的。

为了保证适应度函数的非负性, 取适应度函数为

$$f(\cdot) = \begin{cases} C_{\max} - C(\cdot) & , \text{当 } C(\cdot) < C_{\max} \\ 0 & , \text{其他} \end{cases}$$

其中 C_{\max} 为初始群体中能量函数的最大值, $C(\cdot)$ 为由(3)式确定的能量函数。

遗传操作的设计如下。

(1) 选择或复制操作的目的是为了从当前群体中选出优良的个体, 使它们有机会作为父代为下一代繁殖子孙。判断个体优良与否的标准就是其适应度值。我们采用和适应度值成比例的概率方法来进行选择。即第 j 个个体的适应度为 F_j , 则赋予其选择概率

$$P_{si} = F_i / \sum_{j=1}^M F_j$$

依此概率的大小选择 M 份复制到配对库, 以备配对繁殖。

(2) 交叉 首先对配对库中的个体进行随机配对, 在配对个体中随机设定两个交叉点, 依交叉概率 P_c 决定该配对个体是否进行交叉。即采用所谓的二点交叉, 一个二点交叉的例子如下所示。

个体 A 1101 0110 11 1101101011 新个体 A

个体 B 0010 1010 00 0010011000 新个体 B

即若个体 A 对应的输入单元数和隐单元数分别为 26、27, 个体 B 对应的输入单元数和隐单元数分别为 5、8, 则经交叉后得到的新个体 A 对应的输入单元数和隐单元数分别为 27、11, 新个体 B 对应的输入单元数和隐单元数分别为 4、24。

(3) 变异 随机地选择 0/1 串中的某些位以变异概率 P_m 随机地改变。设个体 A 的第 2 位和第 7 位变异, 得到的新个体为 C

个体 A 1101011011 个体 C 为 1001011111

在当前父代和子代中, 为了不使适应度最大的个体被淘汰, 我们采用父代中适应度最大的个体替代遗传操作后产生的个体中适应度最差的个体。

对以上步骤重复进行训练, 当群体适应度趋于稳定或能量函数小于某一给定值或已达到预定的进化代数, 则终止训练。

把以上方法用于 1958~1978 年的月平均温度的计算, 以建立预报方程

$$y = F(x_1, x_2, \dots, x_n)$$

取交叉概率 $P_c = 0.5$, 变异概率 $P_m = 0.05$, 群体规模 $M = 16$ 。每个个体的二进制编码长度为 10 位。

取基本小波函数为高斯导函数

$$(x) = -x e^{-x^2}$$

对每个个体对应的小波网络中的权系数 w_i , 旋转和平移伸缩因子 R_i, D_i, t_i 的训练按 BP 算法确定它们的值。

经几代遗传后, 各个体的适应度逐渐稳定下来, 可得小波数 N 取 6, 变量数 n 取 4, 此时能量函数即误差平方和为 $RE = 2.4123$ 。因而可达到较好的逼近效果。

对于神经网络的输入单元数和隐单元数的确定, 目前基本上是依赖于人的经验, 人们在设计一个神经网络时, 或者预先确定它的结构, 或者采用递增或递减的探测方法。递增法是从一个很小的网络结构开始, 在训练过程中, 根据特定问题的需要, 逐渐增加结构的各个部分, 直止找到能解决问题的结构为止。递减法正好相反。当网络结构较复杂, 单元数较大时, 这种方法难以找到全局最优解。而采用遗传算法, 利用其全局搜索的性能, 则较易得到满意的结构。

参 考 文 献

- 1 Zhang Qinghua, Albert Benveniste. Wavelet Networks. IEEE Trans. On Neural Networks, 1992, 3: 889 ~ 898
- 2 杜正春, 刘玉田, 夏道止. 前馈神经网络的一种有效学习算法. 电子学报, 1995, 23(8): 57 ~ 61
- 3 徐雷, 迟惠生. 静态前馈型网络的监督学习方法研究进展. 电子学报, 1992, 20(10): 106 ~ 113
- 4 陈国良, 王照法, 庄镇泉, 王东生. 遗传算法及其应用. 北京: 人民邮电出版社, 1996
- 5 刘勇, 康立山, 陈毓屏. 非数值并行算法(第二册)遗传算法. 北京: 科学出版社, 1995

OPTIMIZE THE STRUCTURE OF WAVELET NEURAL NETWORK WITH GENETIC ALGORITHM

Qin Weiliang

(Department of Basic Sciences, NIM, Nanjing 210044)

Jin Long

(Meteorological Bureau of Jiangsu Province, Nanjing 210008)

Li Shiping

(Library, NIM, Nanjing 210044)

Abstract The optimization to the structure of a wavelet neural network can be implemented if the genetic algorithm is used to determine both the number of elements in input level and that in hidden level while the gradient method is used to compute weighting, stretching and displacing coefficients.

Keywords genetic algorithm, wavelet neural network, optimization, gradient method