

# 神经网络模型的预测精度影响因素分析 及其优化

陈 果

(南京航空航天大学 民航学院 南京 210016)

**摘 要** 神经网络对于非线性模型的辨识和非平稳信号的预测,与传统预测模型相比具有较明显的优势,但是神经网络的结构对于信号预测或模型辨识的精度具有较大影响.本文针对广泛使用的 BP 神经网络预测模型,以太阳光子数据为例,分析了网络预测的拓扑结构(输入节点数、隐层节点数)及网络允许的训练误差 MSE (Mean of Squared Error) 对其预测能力的影响.发现最优网络模型对应于一定的拓扑结构,收敛于某个由 MSE 目标值决定的最优位置,该收敛位置并不是网络的全局最优点.在此基础上,利用遗传算法,对输入节点数、隐层节点数和 MSE 目标值进行了优化,得到了最优的网络预测模型.最后,用算例验证了本文对 BP 网络模型预测精度影响因素分析的正确性.

**关键词** 预测,神经网络,模型,影响因素,遗传算法,优化

**中图分类号** O329; F201

## Analysis of Influence Factors for Forecasting Precision of Artificial Neural Network Model and Its Optimizing

CHEN Guo

(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics,  
Nanjing 210016)

### ABSTRACT

In identifying of non-linear model and forecasting non-even signal, Artificial Neural Network (ANN) has obvious advantages over traditional forecasting models, but ANN structure has great influence on forecasting and identifying precision. In this paper BP neural network model which is used most widely is aimed at, the sunspot data is utilized, and the structure of ANN (including input layer node number and hidden layer node number) and the admitted training error MSE (Mean of Squared Error) are analyzed in order to make out how they affect the forecasting precision of ANN. It is found that the optimum network possesses a specific structure (including input layer node number and hidden layer node number) and converges on an optimum position which is direct related to MSE target value and is not the whole optimum position. On this basis, the Genetic Algorithm (GA) is utilized to optimize ANN model. The node number of input layer, the node number of hidden layer for 3-layer BP network, and MSE target value are optimized, then the optimum forecasting model of BP network is obtained. At last, an example shows that the analysis of influence factors for forecasting precision of ANN model is correct.

**Key Words** Forecasting, Neural Network, Model, Influence Factors, Genetic Algorithm, Optimizing

收稿日期:2004-02-03;修回日期:2004-11-18

**作者简介** 陈果,男,1972年生,博士,副教授,主要研究方向为航空发动机磨损状态监测与故障智能诊断、数据融合、神经网络与遗传算法、图像处理及模式识别等. E-mail: cgzyx@263.net.

## 1 引言

利用神经网络进行预测研究的基础是它能拟合任意的非线性函数并且具有一定的泛化能力. Lapedes<sup>[1]</sup>等人最早将神经网络应用于预测,用神经网络对由计算机产生的时间序列仿真数据进行了学习和预测. Weigend<sup>[2]</sup>等人利用神经网络研究了太阳黑子的年平均活动情况,将神经网络与回归分析方法作了比较,其研究结果表明神经网络预测优于统计预测. Chakraborty<sup>[3]</sup>等人进行了神经网络多变量时间序列预测研究,结果表明神经网络方法的预测能力比统计学方法好得多.

但是,神经网络应用到预测还有许多问题需要解决<sup>[4]</sup>,其中最为突出的问题就是,到目前为止,没有一个确定最合适的神经网络结构的标准方法,因此只能花大量的时间采用凑试法,从许多次实验中找出“最合适”的一种.然而,由于影响神经网络预测能力的参数很多,上述方法未必能找到真正意义上的最合适的预测模型结构.有鉴于此,本文针对最为常用的BP网络,对影响神经网络预测精度的因素进行了详细分析和研究,并在此基础上,应用遗传算法的全局搜索能力,获取最优的网络预测模型.

## 2 BP网络的预测方法

利用神经网络进行多变量时间序列的预测原理与单变量时间序列预测相同,多步预测可以由单步预测迭代而成,因此不失一般性,可以以单变量单步预测为例进行研究.设一个单变量时间序列  $\{x_1, x_2, \dots\}$ ,对它进行预测的前提是认为其未来值与其前面的  $m$  个值之间有着某种函数关系,可描述为

$$x_{n+k} = F(x_n, x_{n-1}, \dots, x_{n-m+1}). \quad (1)$$

利用神经网络来拟合这种函数关系  $F$ ,并用它来推导未来的值.进行时间序列预测的神经网络结构可以分为两种,一种是单步预测网络,一种是多步预测网络.单步预测网络输出个数为1个,一次可计算一步预测值;多步预测网络的输出个数不止一个而是多个( $k$ 个),每次可计算出  $k$  步的预测值.在预测过程中,可将得到的预测值作为下一步预测的输入来计算进一步的预测值,这样通过迭代来进行多步预测.

## 3 BP网络预测模型及对预测精度的影响因素

在BP网络的单变量单步预测模型中,输入层的节点数、隐层数和隐层节点数成为了BP网络预测模型中影响网络预测能力的最重要参数. Lippmann<sup>[5]</sup>和Cyberko<sup>[6]</sup>曾指出,有两个隐层,就可以解决任何形式的分类问题.后来R H Nielson<sup>[7]</sup>等人研究进一步指出:只有一个隐层的神经网络,只要隐层节点足够多,就可以以任意精度逼近一个非线性函数.由此可见,用于预测的BP网络只需要三层,且输出节点数可以根据预测的功能事先确定,所以影响BP网络预测能力的参数只剩下输入节点数和隐层节点数.因此本文选取输入节点数和隐层节点数来作为网络预测精度的影响因素来分析.三层的BP网络预测模型拓扑结构如图1所示.

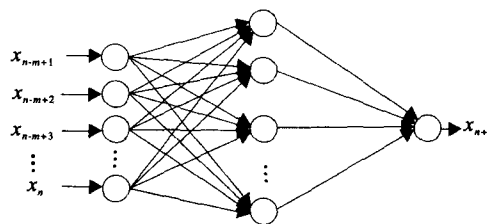


图1 BP网络预测模型结构

Fig.1 BP network structure of forecasting model

BP网络由于是按梯度下降法进行寻优,因此很难保证收敛到全局最小,同时,初始权值对收敛结果有很大的影响,即BP网络的训练具有不确定性.然而,训练不确定可以通过确定相同的初始权值来消除,此时,网络的收敛位置将直接受允许的训练误差目标——均方误差(Mean of Squared Error; MSE)决定,不同的MSE将使网络收敛到不同部位.当达到全局最小时,MSE将达到最小.通常BP算法容易陷入局部最小值,而使网络不容易收敛于全局最小.

事实上,网络训练精度越高,允许的训练误差MSE越小,只能说明网络对训练样本具有很高的拟合能力,而丝毫不能反映且不能保证网络具有很高的预测推广能力,相反,网络训练精度越高,将使模型更加复杂,反而降低了模型的适应能力和推广能力.因此本文认为BP神经网络预测模型的预测能力与收敛到全局最小有直接关系,不同的训练MSE目

标值,将决定网络收敛于部位,而收敛于不同位置的网络具有不同的预测能力.因此本文将MSE作为第三个影响因素进行分析.

## 4 时间序列预测精度评价函数的选取

### 4.1 神经网络预测精度评价函数

在实际应用中,对于实际测得的时间序列 $\{x_1, x_2, \dots\}$ ,可以利用其一部分数据(通常为前一半)来建模,而用另一部分数据(后一半)来对所建模型进行验证,如果预测值与实测值相差越少,显然模型越理想,理想情况是预测值与实测值相等,则达到完美预测.通常衡量预测值与实测值差别的变量采用平均相对变动值(Average Relative Variance, ARV)<sup>[8]</sup>,其定义为

$$ARV = \frac{\sum_{i=1}^N [x(i) - \hat{x}(i)]^2}{\sum_{i=1}^N [x(i) - \bar{x}(i)]^2}, \quad (2)$$

其中,  $N$  为比较数据个数,  $x(i)$  为实测数据值,  $\bar{x}$  为实测数据平均值,  $\hat{x}(i)$  为预测值. 显然, 平均相对变动值 ARV 越小, 也表明预测效果越好,  $ARV = 0$  表示达到了理想预测效果, 当  $ARV = 1$  时, 表明模型仅达到平均值的预测效果.

为了分析神经网络的泛化能力, 设训练点数为  $N_1$ , 测试点数  $N_2$ , 根据 ARV 的定义再定义三个函数, 即综合 ARV 值 ARV1、训练 ARV 值 ARV2、预测 ARV 值 ARV3.

$$\begin{aligned} ARV1 &= \frac{\sum_{i=1}^{N_1+N_2} [x(i) - \hat{x}(i)]^2}{\sum_{i=1}^{N_1+N_2} [x(i) - \bar{x}(i)]^2}, \\ ARV2 &= \frac{\sum_{i=1}^{N_1} [x(i) - \hat{x}(i)]^2}{\sum_{i=1}^{N_1} [x(i) - \bar{x}(i)]^2}, \\ ARV3 &= \frac{\sum_{i=1}^{N_2} [x(i) - \hat{x}(i)]^2}{\sum_{i=1}^{N_2} [x(i) - \bar{x}(i)]^2}. \end{aligned} \quad (3)$$

显然, ARV1 综合反映了网络对训练点和预测点的拟合程度, ARV2 仅反映了网络对训练点的拟合程度, ARV3 仅反映了网络对预测点的拟合程度.

### 4.2 评价神经网络预测精度的时间序列

为了分析神经网络结构对时间序列预测精度的影响, 需要选取标准的时间序列来进行实验. 国际上所采用的标准时间序列有很多, 其中太阳黑子数据是最具有代表性的数据之一. 本文选取 1700 至 1987 年的太阳黑子数据, 其中用前一半数据进行三层 BP 网络建模, 用后一半数据来对模型进行检验. 设网络输入节点数为  $N\_Input$ , 数据总个数为  $N_0$ , 则训练点数为  $N_1 = N_0/2 - N\_Input$ , 测试点数为  $N_2 = N_0/2$ .

## 5 输入层节点数和隐层节点数对网络预测精度的影响

下面研究分别给定不同 MSE 目标值时, 网络输入层和隐层节点数对网络预测能力的影响. 计算中, 输入层节点数为 5 ~ 45、隐层节点数为 10 ~ 50, 由于在 ARV1、ARV2、ARV3 中, ARV2 仅反映训练点的拟合程度, 不反映网络的预测能力, 所以分别以 ARV1 和 ARV3 来对网络的预测能力进行比较, 计算后将按 ARV1 或 ARV3 的值从小到大排序, 取前 5 个最优网络列入表中.

通过比较表 1 ~ 表 8, 可以得到以下结论:

(1) 网络的拓扑结构, 即输入层节点数和隐层节点数, 对其预测能力影响很大;

(2) 在较大的 MSE 目标值下(如 0.1 和 0.01), 网络训练不完善, 不仅不能对测试点进行较好地预测, 而且对训练点也不能进行较好地拟合;

(3) 在适当的 MSE 目标值下(如 0.001), 网络既能对训练点进行较好的拟合, 又能对测试点进行较好的预测;

(4) 在很小的 MSE 目标值下(如 0.0001), 网络对训练点进行更为准确的拟合, 但对测试点的预测能力却变得较差;

(5) ARV1 和 ARV3 均反映了网络的预测能力, 但 ARV1 还反映了网络对训练点的拟合程度, 信息更为全面, 因此, 用 ARV1 来衡量网络预测能力更为准确. 例如在 MSE 目标值为 0.01 时, 最优网络对应的 ARV1 为最小值 0.1313, ARV2 为最小值 0.1283, ARV3 为最小值 0.1372; 在 MSE 目标值为 0.001 时, ARV1 最小值为 0.1162, ARV2 为最小值 0.0275, ARV3 最小值为 0.1776. 显然在  $MSE = 0.001$  时, 网络收敛得更好, 综合预测能力更强.

表 1 按 ARV1 排序得到的前 5 个最优网络结构(MSE = 0.1)

Table 1 Five optimum network structures according to ARV1's order (MSE = 0.1)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	9	38	0.2544	0.3326	0.2105
2	9	26	0.2624	0.3193	0.2326
3	14	18	0.273	0.314	0.2529
4	5	20	0.2778	0.3171	0.2601
5	13	41	0.2877	0.3237	0.2711

表 2 按 ARV3 排序得到的前 5 个最优网络结构(MSE = 0.1)

Table 2 Five optimum network structures according to ARV3's order (MSE = 0.1)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	9	38	0.2544	0.3326	0.2105
2	9	26	0.2624	0.3193	0.2326
3	14	18	0.273	0.314	0.2529
4	5	20	0.2778	0.3171	0.2601
5	13	41	0.2877	0.3237	0.2711

表 3 按 ARV1 排序得到的前 5 个最优网络结构(MSE = 0.01)

Table 3 Five optimum network structures according to ARV1's order (MSE = 0.01)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	6	40	0.1313	0.1283	0.1372
2	12	23	0.1359	0.0937	0.1654
3	8	32	0.1441	0.1131	0.1684
4	22	43	0.1455	0.0853	0.1841
5	6	48	0.1459	0.1512	0.1467

表 4 按 ARV3 排序得到的前 5 个最优网络结构(MSE = 0.01)

Table 4 Five optimum network structures according to ARV3's order (MSE = 0.01)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	6	40	0.1313	0.1283	0.1372
2	6	48	0.1459	0.1512	0.1467
3	20	12	0.1591	0.1735	0.1531
4	5	47	0.1496	0.1432	0.1581
5	5	15	0.1481	0.1363	0.1601

表 5 按 ARV1 排序得到的前 5 个最优网络结构(MSE = 0.001)

Table 5 Five optimum network structures according to ARV1's order (MSE = 0.001)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	7	37	0.1162	0.0275	0.1776
2	6	37	0.1209	0.0277	0.1856
3	14	45	0.1292	0.0274	0.194
4	7	25	0.1371	0.028	0.2125
5	8	32	0.1376	0.0285	0.2125

表 6 按 ARV3 排序得到的前 5 个最优网络结构(MSE = 0.001)

Table 6 Five optimum network structures according to ARV3's order (MSE = 0.001)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	7	37	0.1162	0.0275	0.1776
2	6	37	0.1209	0.0277	0.1856
3	14	45	0.1292	0.0274	0.194
4	7	25	0.1371	0.028	0.2125
5	8	32	0.1376	0.0285	0.2125

表 7 按 ARV1 排序得到的前 5 个最优网络结构(MSE = 0.0001)

Table 7 Five optimum network structures according to ARV1's order (MSE = 0.0001)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	21	41	0.2052	0.0021	0.3312
2	19	29	0.2075	0.0027	0.335
3	21	22	0.2156	0.0028	0.3475
4	18	36	0.2203	0.0028	0.3558
5	26	37	0.2242	0.0023	0.3584

表 8 按 ARV3 排序得到的前 5 个最优网络结构(MSE = 0.0001)

Table 8 Five optimum network structures according to ARV3's order (MSE = 0.0001)

No	输入层 节点数	隐层 节点数	ARV1	ARV2	ARV3
1	21	41	0.2052	0.0021	0.3312
2	19	29	0.2075	0.0027	0.335
3	21	22	0.2156	0.0028	0.3475
4	18	36	0.2203	0.0028	0.3558
5	26	37	0.2242	0.0023	0.3584

## 6 网络训练的 MSE 目标值对其预测精度的影响

通过上述分析发现,与网络的拓扑结构(输入层节点数和隐层节点数)一样,网络训练的 MSE 目标值对网络的预测能力也有很大的影响.为了进一步研究其影响程度.下面以拓扑结构为输入层节点数为 7、隐层节点数为 37 的网络及拓扑结构为输入层节点数为 18、隐层节点数为 31 的网络在不同的 MSE 目标值下进行训练,研究 ARV1、ARV2 及 ARV3 函数随 MSE 目标值的变化情况.

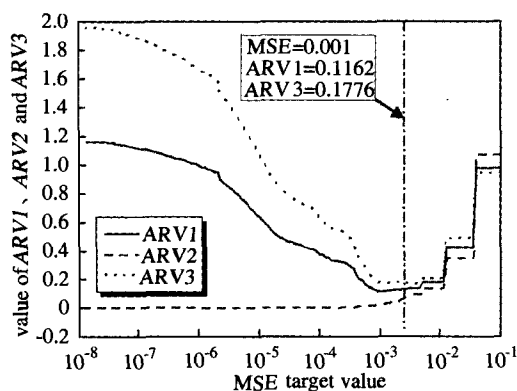


图 2 网络 1 的预测精度与 MSE 目标值的关系

Fig. 2 Relation between forecasting precision and MSE objective value of network 1

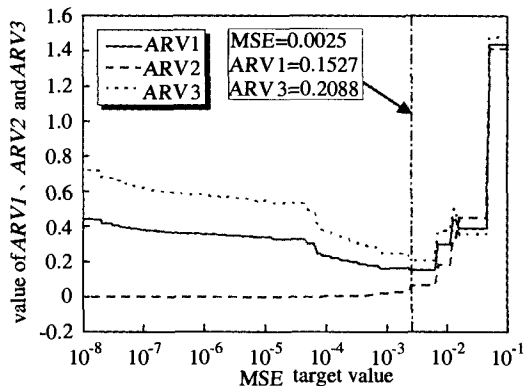


图 3 网络 2 的预测精度与 MSE 目标值的关系

Fig. 3 Relation between forecasting precision and MSE target value of network 2

图 2 和图 3 分别为网络权值相同时, MSE 目标值从  $10^{-1} \sim 10^{-8}$  变化时, ARV1、和 ARV2 和 ARV3 的变化情况.由图 2 和图 3 可以得出以下结论:

(1) 不同拓扑结构(输入层节点数和隐层节点数)的网络,达到其最优的预测能力的 MSE 目标值不一样.存在一个最优的网络,其拓扑结构和网络收敛部位均达到最优;

(2) 当具有某拓扑结构的网络,其收敛于某特定位置将具有最佳的预测能力.而网络允许的训练误差 MSE 目标值与网络的收敛位置直接相关,因此寻找最优网络,不仅要搜索出网络的拓扑结构,而且要搜索出网络的训练误差目标值 MSE.

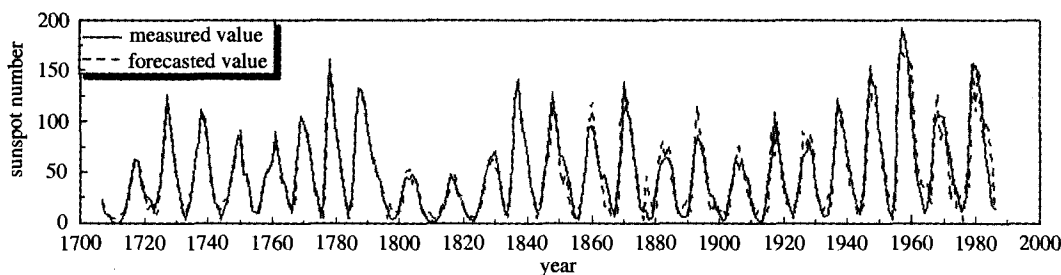


图 4 MSE 等于  $10^{-3}$  时网络的预测值与实测值的比较

Fig. 4 Comparison between forecasted value of network and measured value when MSE is equal to  $10^{-3}$

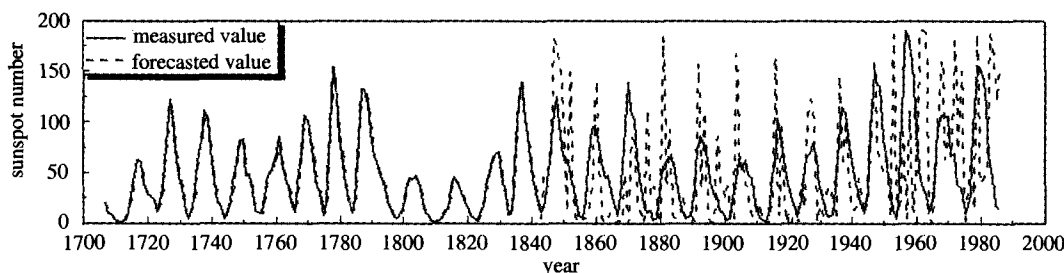


图 5 MSE 等于  $10^{-8}$  时网络的预测值与实测值的比较

Fig. 5 Comparison between forecasted value of network and measured value when MSE is equal to  $10^{-8}$

当网络收敛越逼近全局最小,其拟合训练点的能力越强,但其预测能力越差,从图 2 和图 3 不难看出此规律.图 4 和图 5 分别为网络(输入层节点数为 7 和隐层节点数为 37)在 MSE 目标值等于  $10^{-3}$  和等于  $10^{-8}$  所预测的太阳黑子数据与实测的太阳黑子数据的比较,从比较中可以看出, MSE 越小,网络收敛的越好,但是其预测值偏差却越大,显然, MSE 为  $10^{-3}$  时,网络的预测能力达到最佳的预测效果.

## 7 BP 网络预测模型优化的遗传算法构造

通过上述分析发现, BP 网络预测能力的影响参数为输入层节点数  $N$ 、中间隐层节点数  $M$  及 MSE 目标值,为了获取最佳预测性能的 BP 网络,需要得到最佳的  $N$ 、 $M$  和 MSE 目标值.显然这是一个优化问题,如果采取穷举的方式搜索最优值,计算量将十分巨大以至于无法实现.由于遗传算法<sup>[9]</sup>具有隐含的并行性和强大全局搜索能力,可以在很短的时间内搜索到全局最优点.因此本文利用遗传算法来进行 BP 网络的预测模型的结构优化.

本文遗传算法参数为:种群数  $n = 30$ ,基因串(染色体)采用二进制编码,交叉率和变异率分别为 0.50 和 0.05,进化代数数为 100 代.

$N$ 、 $M$  及 MSE 均用 5 位二进制编码,为了避免节点数为 0,规定解码后,对于输入层节点数加上 5

得到  $N$ ,而隐层节点数加上 10 以得到  $M$ ,对 MSE,通过计算  $MSE = 10^{-(1+MSE \times 5/32)}$  得到网络训练允许误差目标值.适应度函数取为  $f = 1/ARV1$ .

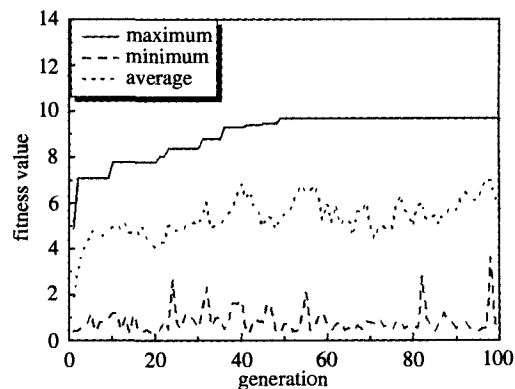


图 6 进化代数与适应度函数值的变化关系

Fig. 6 Relation between generation and fitness value

图 6 为适应度函数最大值、平均值、最小值随遗传代数的增加的变化规律.从图中可以看出,经过 50 代进化后,适应度函数达到了最大值,得到了最优的输入层节点数  $N$  和隐层节点数  $M$ .其优化结果为:最优的适应度值为 9.6896,平均相对变动值 ARV1 为 0.1032,输入层节点数  $N = 5$ ,隐层节点数  $M = 15$ ,  $MSE = 0.00392$ .图 7 为该网络结构对太阳黑子数据的预测值与实测值的比较.从图中可以看出其拟合程度达到了最佳.

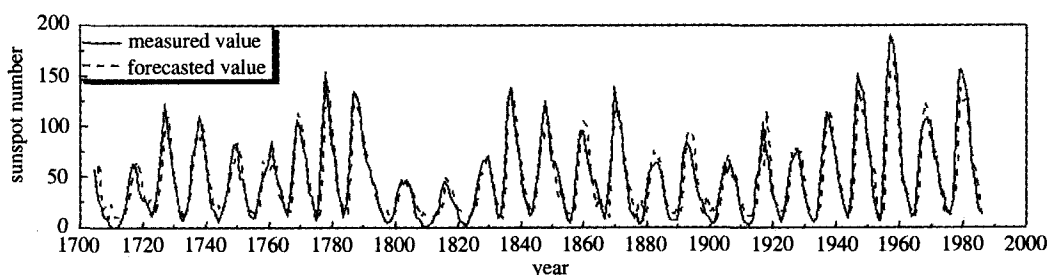


图 7 最优网络 ( $N = 5, M = 15$ ) 的预测值与实测值的比较

Fig. 7 Comparison between forecasted value of optimum network ( $N = 5, M = 15$ ) and measured value

## 8 结 论

(1)分析了神经网络进行非线性预测、多变量预测的优越性以及神经网络用于预测的缺点所在.

(2)提出了影响 BP 网络预测能力的三个重要

参数——输入层节点数、隐层节点数及网络训练允许误差 MSE 目标值.

(3)在一定允许训练误差的情况下,研究了输入层节点数和隐层节点数对网络预测精度的影响,发现存在一个最优的输入层节点数和隐层节点数,这样的网络结构具有较强的预测能力.

(4)在一定输入层节点数和隐层节点数的情况

下,研究了允许训练误差 MSE 目标值对网络预测精度的影响,发现具有最佳预测能力的网络模型并不收敛于全局最小,收敛于全局最小的网络模型往往只能对训练样本产生较好的拟合,但其预测精度将变得很差.

(5)本文用遗传算法构造了同时优化影响神经网络预测精度的参数(输入层节点数、隐层节点数及样本允许训练误差)的算法,得到了较优的网络预测模型.最后,用算例验证了本文分析结果的正确性.

### 参 考 文 献

- [1] Farber L A. Nonlinear Signal Processing Using Neural Network; Prediction and System Modeling. Technical Report, LA-UR--87-2662, Los Alamos National Laboratory, Los Alamos, USA, 1987
- [2] Weigend A B, *et al.* Predicting the Future: A Connectionist Approach. International Journal of Neural System, 1990, 1: 193-209
- [3] Chakraborty K, *et al.* Forecasting the Behavior of Multivariate Time Series Using Neural Networks. Neural Networks, 1992, 5(6): 961-970
- [4] Liu B, Hu D P. Some Application Research of Neural Networks in Forecasting. Journal of System Engineering, 1999, 14(4): 338-344 (in Chinese)  
(刘豹, 胡代平. 神经网络在预测中的一些应用研究. 系统工程学报, 1999, 14(4): 338-344)
- [5] Lippmann R P. An Introduction to Computing with Neural Nets. IEEE Acoustics, Speech and Signal Processing Magazine, 1987, 4(2): 4-22
- [6] Cyberko G. Approximations by Super-Positions of a Sigmoidal Function. Mathematics of Control, Signals and Systems, 1989, 2(2): 34-89
- [7] Zhang L M. Artificial Neural Network Model and Its Application. Shanghai, China: Fudan University Press, 1995, 1-92 (in Chinese)  
(张立明. 人工神经网络的模型及其应用. 上海: 复旦大学出版社, 1995, 1-92)
- [8] Cholewo T, Zurada J M. Sequential Network Construction for Time Series Prediction. In: Proc of the IEEE International Joint Conference on Neural Networks. Houston, USA, 1997, 2034-2039
- [9] Goldberg D. Genetic Algorithms in Search, Optimization and Machine Learning. Reading, USA: Addison-Wesley, 1989

\*\*\*\*\*

## 第六届中国 Rough 集与软计算学术研讨会 (CRSSC2006) 征 文 通 知

由中国人工智能学会粗糙集与软计算专业委员会和中国计算机学会人工智能与模式识别专业委员会主办、浙江师范大学承办的“第六届中国 Rough 集与软计算学术研讨会(CRSSC2006)”拟定于2006年10月30日至11月3日在浙江金华召开。

自2001年在重庆成功召开“第一届中国 Rough 集与软计算学术研讨会(CRSSC2001)”以来,我国每年的CRSSC系列研讨会在规模和质量上均呈良好的增长趋势,在此领域的研究工作发展很快。2003年成立了中国人工智能学会粗糙集与软计算专业委员会, Rough 集的研究队伍也更加壮大,研究成果在深度和广度上有了更大的发展。

现将有关征文事宜通知如下,欢迎各界人士踊跃投稿。

### 征文范围(但不限于)

- |                |             |             |              |
|----------------|-------------|-------------|--------------|
| • Rough 集理论及应用 | • 计算智能      | • 机器学习      | • 文字计算       |
| • Fuzzy 集理论及应用 | • 粒计算       | • 软计算及其应用   | • 演化计算       |
| • Petri 网      | • 软计算的逻辑基础  | • 非经典逻辑     | • 神经网络       |
| • 计算复杂性        | • 空间推理      | • 统计与概率推理   | • 智能 Agent   |
| • 多准则决策分析      | • 决策支持系统    | • 知识发现与数据挖掘 | • 多 Agent 技术 |
| • 近似推理与不确定性推理  | • 网络智能      | • 集成智能系统    | • 数据仓库       |
| • 模式识别与图像处理    | • 生物信息与生物计算 | • 认知信息学     |              |

(下转第 540 页)