

文章编号:1000-6788(2005)10-0095-06

BP神经网络网络结构优化问题的研究

唐万梅^{1,2}

(1. 内蒙古大学理工学院,内蒙古 呼和浩特 010021; 2. 重庆师范大学数学与计算机科学学院,重庆 400047)

摘要: 将灰色关联分析法应用到BP神经网络隐层结点数的确定中,实现了BP神经网络网络结构的优化,提高了网络的预测精度. 仿真结果表明这种方法非常有效,同时也对BP神经网络的进一步研究提供了一种全新的思路.

关键词: BP神经网络;灰色关联分析;网络结构的优化

中图分类号: TP18

文献标识码: A

The Study of the Optimal Structure of BP Neural Network

TANG Wan-mei^{1,2}

(1. College of Sciences and Technology, NeiMongol University, Hohhot 010021, China; 2. Dept of Mathematics and Computer Sciences, Chongqing Normal University, Chongqing 400047, China)

Abstract: The acting mechanism of the BP Neural Network's hidden units and the number's choice have been a difficult problem on the study of BP Neural Network. By using the grey correlation analysis on the hidden units number's determination of the BP Neural Network, this method realized the optimal of BP Neural Network, and also improve network's forecasting accuracy. Its simulated result shows that the method is effective. At the same time, it also afford a new thinking to us for a further study of the BP Neural Network.

Key words: BP neural network; grey correlation analysis; the optimal structure of the network

1 引言

对于BP网络的改进很多,主要从以下几个方面进行:学习算法的改进,网络结构的优化等.网络结构的改进又分对输入层结点数的确定,隐层结点数的确定^[1]等问题.对输入层结点数的确定,往往是根据样本指标的个数来确定输入层结点数的多少,但输入层结点过多,同样会出现网络训练时间过长、网络性能差等问题,针对此,文献[2]利用主成分分析法确定输入层结点数,可使问题得到一定的解决.然而,在实际应用中,当遇到样本数据很大,人们往往随机选取其中一部分作为学习样本,但这样很可能使样本集本身没有包含全部样本的特征,使预测的结果出现较大的误差.文献[3]利用模糊聚类法首先对样本进行分类,然后再从每一类中按一定比例选择学习样本,但在分类时,究竟应将样本数据分为几类往往随意性很大,显然,所分类别数不同,最后所选取的学习样本便有所不同,就有可能导致最后预测结果的精度差异.对此,可以利用统计分析中的对应分析结果,这样既可以减少输入层神经元的数目,达到降维的目的,同时又能从大样本集中确定出学习样本集,使得学习样本集尽可能包含全部样本的特征,从而提高问题的预测精度.关于这方面的内容将另文讨论.

隐层的设计问题是前馈神经网络(feedforward neural network, FNN)构建中最具挑战性的问题.隐单元直接影响着网络的容量、泛化能力、学习速度和输出性能.从网络容量和函数逼近的通用性考虑,隐单元数目越多越好;然而从网络的泛化能力来考虑,隐单元数目不宜过多,但隐层单元数目应该存在最优值^[4].本文拟利用灰色关联分析确定BP神经网络隐层结点数,从而达到优化网络结构的目的.实验结果表明这种改进方法取得了很好的效果.具有很强的泛化能力.

收稿日期:2004-10-12

资助项目:国家自然科学基金(10171118,10471159);教育部重点项目资助;重庆师范大学校级项目(05XL Y017)

作者简介:唐万梅(1965-),女,重庆,副教授,博士研究生,研究方向:计算机算法及区域规划. E-mail: cqtwm@163.com.

2 BP神经网络模型

在人工神经网络模型中,具有误差反向传播学习功能的多层前馈神经网络(BP网)是目前应用最为广泛且研究最为深入的神经网络^[5].BP网络模型具有结构严谨、工作状态稳定、可操作性强等特点,同时由于隐结点的引入,使得一个三层非线性网络可以以任意精度逼近任何连续函数,因而在模式识别、非线性映射、预测等许多领域得到了广泛的应用.

BP神经网络的研究已有很长的历史.BP神经网络是由简单神经单元构成的具有输入层、隐层和输出层的网络模型,具有很强的学习、联想和容错功能,能进行大规模并行信息处理,对非线性系统具有很强的模拟能力.在对神经网络的研究中,网络的泛化能力(Generalization capability)是衡量其网络性能好坏的重要标志,而从实际应用中得出的结论认为:网络的泛化能力依赖于网络结构(主要是隐层结点数,隐层数和隐结点的函数特性)和训练样本的特性^[6,7].

3 利用灰色关联分析优化网络结构

针对BP神经网络模型的研究,在网络结构的优化讨论中,隐层结点数的多少直接影响其网络的性能,但又不容易确定.若选取的隐结点数太多,神经元冗余现象加重,网络训练变得困难,但若选得太少,则又不能保证网络应有的精度要求.在实际应用中,一般凭经验选取,但这缺乏科学性.显然,在隐层的结点中,一部分神经单元对输出层结点的影响大,在网络中处于相对重要的位置,而另一部分结点可能对网络的输出影响较小,因而在网络中认为它们的作用相对较小,哪些对网络影响太小的隐结点,我们认为它们是冗余的,可考虑将它们删除.虽然神经网络神经元的冗余说明它具有良好的容错性,但冗余太多,网络性能势必会受到一定的影响,降低预测精度.因此,可调整网络结构,使其达到一个较为理想的状态.而BP网络隐层神经元的作用机理及其个数的选择已成为BP网络研究中的一个难点问题,本文拟在网络学习过程中,根据实际问题的具体要求,引入灰色关联分析法^[8]来调整网络结构,确定隐层结点数,达到优化网络结构的目的.记 $op_i = (op_i(1), op_i(2), \dots, op_i(N)) (i = 1, 2, \dots, n_2)$ 表示隐层第 i 个神经单元学习 N 个样本数据的输出; $y = (y_1, y_2, \dots, y_N)$ 表示输出层神经元对 N 个学习样本数据的输出.

3.1 灰色关联分析

灰色关联(简称灰关联)是指事物之间不确定性关联,或系统因子与行为因子之间的不确定性关联.而所谓的灰色关联分析是基于行为因子序列的微观或宏观接近,以分析确定因子间的影响程度或因子对主行为的贡献测度而进行的一种分析方法^[8].

3.2 灰色关联分析的基本步骤

3.2.1 求关联系数

首先利用灰色关联分析方法计算出BP网隐层结点的输出相对于整个网络输出的关联系数.当网络训练结束后,设对 N 个学习样本,整个网络的输出序列 $y = (y_1, y_2, \dots, y_N)$ 为参考序列(母序列),每一个隐结点的输出 $op_i = (op_i(1), op_i(2), \dots, op_i(N)) (i = 1, 2, \dots, n_2)$ 为比较序列(子序列),则 $op_i = (op_i(1), op_i(2), \dots, op_i(N)) (i = 1, 2, \dots, n_2)$ 与 $y = (y_1, y_2, \dots, y_N)$ 的关联系数为:

$$r_i(k) = \frac{\min_i \min_k |y_k - op_i(k)| + \max_i \max_k |y_k - op_i(k)|}{|y_k - op_i(k)| + \max_i \max_k |y_k - op_i(k)|}, \quad (1)$$

其中, $(0, +)$ 为分辨系数. 越小,分辨能力越大.一般取 $[0, 1]$, 更一般取 $= 0.5$.

3.2.2 关联度的计算

由于关联系数的计算得到的是各比较序列与参考序列在各点的关联系数值,结果较多,信息过于分散,不便于比较,因而有必要将每一比较序列各个时刻的关联系数集中体现在一个数值上,这个值即是关联度.

通常关联度的计算方法采用平均值法:

$$i = \frac{1}{N} \sum_{k=1}^N i(k), \tag{2}$$

$i(k) (k=1, 2, \dots, N)$ 为灰关联序列.

3.2.3 灰关联度排序

对比较序列与参考序列的灰关联度 $i_i (i=1, 2, \dots, n_2)$ 按从大到小进行排序, 即得灰关联序列.

若设灰关联序列为: $i_1 > i_2 > \dots > i_{n_2}$, 则表明第一个隐结点对网络输出的作用最大, 第二个隐结点次之, ..., 第 n_2 个隐结点对网络输出影响最小. 当隐结点对网络输出的影响小到一定程度时, 我们认为该隐结点在整个网络中的作用可忽略, 可将它从网络中删除. 这里, 对一个给定的 α , 取 $\alpha \in (0, 1)$, 对满足 $i_i < \alpha$ 的隐结点我们认为它们对网络输出的影响太小, 对它们实施删除操作从而优化网络结构, 提高网络的推广能力.

3.3 网络的学习训练过程

采用如下改进的 BP 学习算法训练网络, 优化网络结构.

首先对样本数据进行归一化处理, 使样本参数值在 $[-1, 1]$ 范围;

构造 BP 神经网络模型的初始结构: 一个输入层 (假设含有 n_1 个结点)、一个隐层 (含有 n_2 个结点) 和一个输出层 (含有 n_3 个结点, 该层的结点数由实际问题确定);

对网络权值和域值在 $[-1, 1]$ 范围内进行随机初始化操作;

让网络对确定的样本集 (共 N 个学习样本) 进行学习训练, 直至达到规定的学习步数或学习精度;

当网络达到稳定后, 计算网络的实际输出和隐层结点的输出, 记目标输出为 $y_i (i=1, 2, \dots, N)$, 记 $op_i(j) (i=1, 2, \dots, n_2; j=1, 2, \dots, N)$ 代表第 i 个隐结点学习第 j 个样本的输出;

对每一个隐结点的输出序列 $op_i = (op_i(1), op_i(2), \dots, op_i(N)) (i=1, 2, \dots, n_2)$, 计算其相对于输出层结点输出序列 $y = (y_1, y_2, \dots, y_N)$ 的灰关联度 i_i . 本文实际例子中输出层结点数为 1, 若输出层结点数不为 1, 则计算每一个隐结点的输出序列, 相对于输出层所有结点输出序列的期望输出序列 $\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_N)$ 的灰关联度 i_i ;

按 i_i 进行排序, 对凡是满足 $i_i < \alpha$ (α 是根据实际问题确定的域值, $\alpha \in (0, 1)$) 的隐结点实施删除操作, 然后转 , 否则结束算法.

4 实例建模

我们考虑将重庆市有完整检测记录的某钢铁集团公司某厂 3 号炉炼钢时的辅料数据及成品成分数据作为分析样本. 其中, 以添加的辅料数据作为输入, 共选择 12 个参数: 石灰 $x_1(t)$ 、轻烧 $x_2(t)$ 、萤石 $x_3(t)$ 、镁球 $x_4(t)$ 、溅渣球 $x_5(t)$ 、纯吹时间 $x_6(s)$ 、炼钢时间 $x_7(s)$ 、氧气耗量 $x_8(m^3)$ 、工作流量 $x_9(m^3/h)$ 、工作压力 $x_{10}(MPa)$ 、每炉氩耗 $x_{11}(m^3)$ 、每炉氮耗 $x_{12}(m^3)$; 选择成品成分中碳的含量作为输出: 碳 $y(\%)$.

利用 MATLAB^[9] 构造 BP 神经网络的初始结构: 输入层含 n_1 个神经元 ($n_1=12$), 输出层含 n_3 个神经元 ($n_3=1$), n_2 个隐含神经元 ($n_2=30$), 如图 1 所示.

4.1 样本数据的选取

本文共选取了 64 个样本数据, 但限于篇幅的原因, 这里只列出了部分样本数据如表 1.

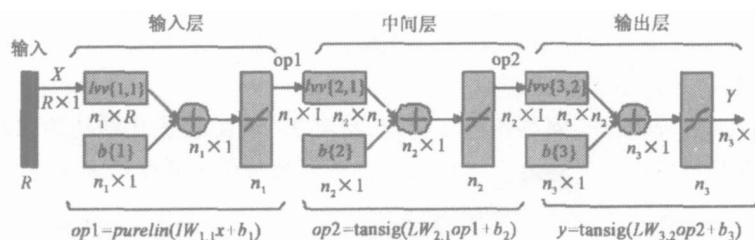


图 1 BP 神经网络模型

表1 训练及验证样本

| 序号 | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | x_{11} | x_{12} | y |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|------|
| 1 | 4.094 | 0.798 | 0.063 | 1.546 | 0.351 | 1007 | 1305 | 4682 | 17021 | 0.822 | 32.32 | 86.29 | 0.35 |
| 2 | 4.61 | 0.002 | 0.005 | 1.507 | 0.389 | 680.5 | 964.1 | 3172 | 16631 | 0.807 | 33.11 | 84.36 | 0.3 |
| 3 | 4.068 | 1.042 | 0.589 | 1.501 | 0.377 | 1020 | 1326 | 4739 | 16713 | 0.816 | 34.85 | 85.26 | 0.31 |
| 4 | 4.557 | 1.306 | 0.979 | 1.566 | 0.346 | 1125 | 1538 | 5302 | 16603 | 0.804 | 34.09 | 73.92 | 0.43 |
| 5 | 4.475 | 1.042 | 0.041 | 1.557 | 0.366 | 1063 | 1345 | 4995 | 16582 | 0.804 | 44.12 | 72.88 | 0.45 |
| 6 | 3.045 | 1.034 | 0.02 | 1.465 | 0.419 | 1063 | 1364 | 4865 | 16593 | 0.802 | 33.71 | 73.08 | 0.46 |
| 7 | 3.899 | 0.991 | 0.164 | 1.521 | 0.578 | 1080 | 1385 | 5062 | 16966 | 0.835 | 56.98 | 117.5 | 0.3 |
| 8 | 3.589 | 1.131 | 0.015 | 1.498 | 0.312 | 1060 | 1331 | 4994 | 16782 | 0.796 | 40.39 | 78.19 | 0.13 |
| 9 | 4.388 | 1.386 | 0.406 | 1.49 | 0.096 | 1137 | 1627 | 5334 | 16512 | 0.806 | 41.96 | 84.33 | 0.17 |
| 10 | 4.215 | 1.011 | 0.297 | 1.663 | 0.26 | 1060 | 1462 | 4995 | 16616 | 0.809 | 48.35 | 73.77 | 0.15 |
| 11 | 3.999 | 1.042 | 0.089 | 1.525 | 0.205 | 1111 | 1407 | 5353 | 16916 | 0.83 | 40.66 | 69.42 | 0.2 |
| 12 | 3.945 | 1.06 | 0.466 | 1.467 | 0.431 | 1087 | 1546 | 5241 | 17075 | 0.849 | 39.2 | 71.43 | 0.2 |
| 13 | 3.617 | 1.072 | 0.026 | 1.503 | 0.627 | 1091 | 4441 | 5129 | 16592 | 0.815 | 41.94 | 75.93 | 0.14 |
| 14 | 4.16 | 0.157 | 0.068 | 2.298 | 0.804 | 1186 | 1541 | 5546 | 16679 | 0.819 | 46.4 | 202.1 | 0.14 |
| 15 | 4.199 | 0.991 | 0.015 | 1.487 | 0.388 | 1068 | 1623 | 4961 | 11632 | 0.801 | 46.43 | 98.56 | 0.19 |
| 16 | 4.38 | 1.079 | 0.017 | 1.501 | 0.346 | 1070 | 1406 | 5002 | 16566 | 0.799 | 48.1 | 72.42 | 0.16 |
| 17 | 4.113 | 1.062 | 0.018 | 1.49 | 0.321 | 1120 | 1420 | 5256 | 17092 | 0.824 | 41.49 | 85.19 | 0.15 |
| 18 | 4.224 | 1.009 | 0.006 | 1.478 | 0.34 | 1034 | 1678 | 4881 | 16579 | 0.8 | 38.51 | 71.5 | 0.45 |
| 19 | 4.477 | 1.046 | 0.535 | 1.489 | 0.386 | 1067 | 1326 | 5000 | 16533 | 0.797 | 51.02 | 75.23 | 0.44 |
| 20 | 3.791 | 1.072 | 0.474 | 1.476 | 0.567 | 1069 | 1540 | 5023 | 16621 | 0.802 | 34.08 | 78.85 | 0.49 |

4.2 神经网络的学习训练

根据 3.3 节,用 MATLAB^[9]编程构造 BP 网络的初始结构:12-30-1.各层神经元的传递函数分别取 purelin 函数、tansig 函数和 tansig 函数,BP 网络的训练函数取 trainscg;设定网络的训练参数如下:

```
net.trainparam.show = 50;net.trainparam.lw = 0.05;net.trainparam.mc = 0.9;
net.trainparam.epochs = 30000;net.trainparam.goal = 1e-6;其余为系统默认值.
```

灰关联度计算时,选取的分辨系数 $\rho = 0.5$, $\sigma = 0.4$.

首先用函数 premmx 将样本数据归一化到 $[-1, 1]$ 范围,选取 60 个样本数据作为训练样本对网络进行学习,最终得到优化的 BP 神经网络结构为:12-14-1.

在程序的运行过程中,原 BP 神经网络的结构首先调整为:12-20-1,此时被删除的结点是 5,6,11,13,14,17,19,20,22,23 共 10 个,各隐结点对 60 个样本数据的输出序列相对于输出层结点的输出序列的灰色关联度如表 2.

在此基础上继续对网络进行训练学习,第二次删除的结点是:1,2,16,21,27,30 共 6 个,各隐结点相对于输出层结点输出序列的灰色关联度如表 3.

隐结点的删除操作进行两次后,网络达到了稳定,得到 BP 网的最优化结构为:12-14-1.此时各隐结点对 60 个样本数据的输出序列相对于输出层结点的输出序列的灰色关联度如表 4.

表 2 隐结点第一次调整时的相关关联度

| | | | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 关联度 | 0.6106 | 0.9342 | 0.9716 | 0.4221 | 0.3803 | 0.3628 | 0.7827 | 0.9594 | 0.9037 | 0.9621 |
| 序号 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 关联度 | 0.3426 | 0.9717 | 0.3799 | 0.3991 | 0.9767 | 0.9253 | 0.3894 | 0.9299 | 0.3693 | 0.3677 |
| 序号 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 关联度 | 0.6195 | 0.3591 | 0.3895 | 0.6184 | 0.6365 | 0.7469 | 0.7516 | 0.7060 | 0.4930 | 0.4305 |

表 3 隐结点第二次调整时的相关关联度

| | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|
| 序号 | 1 | 2 | 3 | 4 | 7 | 8 | 9 | 10 |
| 关联度 | 0.3530 | 0.3893 | 0.9619 | 0.5511 | 0.9187 | 0.9580 | 0.6538 | 0.9592 |
| 序号 | 12 | 15 | 16 | 18 | 21 | 24 | 25 | 26 |
| 关联度 | 0.4539 | 0.7386 | 0.3497 | 0.6679 | 0.3535 | 0.4569 | 0.8471 | 0.7242 |
| 序号 | 27 | 28 | 29 | 30 | | | | |
| 关联度 | 0.3622 | 0.5902 | 0.4004 | 0.3856 | | | | |

表 4 隐结点第三次调整时的相关关联度

| | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|
| 序号 | 3 | 4 | 7 | 8 | 9 | 10 | 12 | 15 |
| 关联度 | 0.5423 | 0.6708 | 0.7228 | 0.7663 | 0.6625 | 0.5809 | 0.5996 | 0.5846 |
| 序号 | 18 | 24 | 25 | 26 | 28 | 29 | | |
| 关联度 | 0.5449 | 0.5747 | 0.5734 | 0.6743 | 0.6382 | 0.6054 | | |

当网络稳定后,对训练样本及验证样本的实际输出和理想输出利用 MATLAB 的 postreg 函数进行了线性回归分析,得出了很好的结果,如图 1、2 所示。

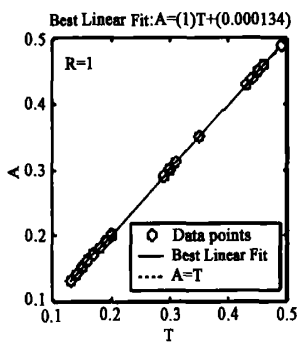


图 1 训练样本输出结果的线性回归分析

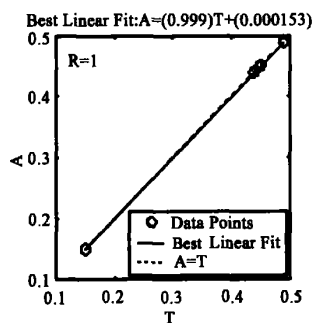


图 2 验证样本仿真结果的线性回归分析

利用灰关联分析优化后的 BP 网络,其预测结果及误差分析如表 5。从表 5 可知,利用灰色关联分析优化的 BP 模型的预测精度较高,最大相对误差为 0.3878%。且克服了利用传统 BP 网络学习算法训练网络时可能遇到的因隐层结点数估计不足,而导致网络性能下降(隐结点数过多)或网络训练时间过长,甚至可能不收敛(若隐结点数太少)的情况,网络的训练时间较传统 BP 网络学习算法运行时间大大缩短,提高了算法的效率。

表 5 优化 BP 模型预测结果及误差分析

| 样本号 | 理想输出 | 预测值 | 相对误差 (%) |
|-----|------|--------|----------|
| 61 | 0.15 | 0.1502 | 0.13333 |
| 62 | 0.45 | 0.45 | 0 |
| 63 | 0.44 | 0.4401 | 0.02273 |
| 64 | 0.49 | 0.4881 | - 0.3878 |

5 结论

神经网络有别于传统的数学模型,它基于的是并行运算的建模思想,为拟合未知的复杂系统提供了一条崭新的思路.神经网络的首要目的是保证网络泛化能力,即通过归纳已知样本的内在规律来推广到未知情况的预测,而网络的泛化能力依赖于网络结构(主要是隐结点数,隐层数和隐结点的函数特性)和训练样本的特性,本文将灰色关联分析应用到BP神经网络隐层结点数的确定中,从而对BP网络的结构进行优化,提高了网络的预测精度.仿真结果表明这种方法非常有效.同时也对BP神经网络的进一步研究提供了一种全新的思路.

致谢 感谢我的导师杨新民教授对我论文工作的指导.

参考文献:

- [1] 李晓峰,徐玖平,王嗣清,等. BP人工神经网络自适应学习算法的建立及其应用[J]. 系统工程理论与实践,2004,24(5):1-8.
Li Xiaofeng, Xu Jiuping, Wang Yinqing, et al. The establishment of self-adapting algorithm of BP neural network and its application [J]. Systems Engineering - Theory & Practice, 2004, 24(5):1-8.
- [2] 陈小前,罗世彬,王振国,等. BP神经网络应用中的前后处理过程研究[J]. 系统工程理论与实践,2002,22(1):65-70.
Chen Xiaoqian, Luo Shibin, Wang Zhenguo, et al. Research on preprocessing and postprocessing of the application of BP neural network[J]. Systems Engineering - Theory & Practice, 2002, 22(1):65-70.
- [3] 何勇,项利国. 基于模糊聚类的BP神经网络模型研究及应用[J]. 系统工程理论与实践,2004,24(2):79-82.
He Yong, Xiang Ligu. Study and application of BP neural network model based on fuzzy clustering[J]. Systems Engineering - Theory & Practice, 2004, 24(2):79-82.
- [4] 崔荣一,烘炳熔. 关于三层前馈神经网络隐层构建问题的研究[J]. 计算机研究与发展,2004,41(4):524-530.
Cui Rongyi, Hong Bingrong. On constructing a hidden layer for three-layered feedforward neural networks[J]. Journal of Computer Research and Development, 2004, 41(4):524-530.
- [5] Rumelhart D E, McClelland J L. Parallel Distributed Processing[M]. The MIT Press, Cambridge, 1986.
- [6] Atiya A, Ji C Y. How initial conditions affect generalization performance in large networks[J]. IEEE Trans Neural Networks, 1997, 8(2):448-451.
- [7] 江学军,唐焕文. 前馈神经网络泛化能力的系统分析[J]. 系统工程理论与实践,2000,20(8):36-40.
Jiang Xuejun, Tang Huanwen. System analysis for generalization of multilayer feedforward neural network[J]. Systems Engineering - Theory & Practice, 2000, 20(8):36-40.
- [8] 刘思峰,郭天榜,党耀国. 灰色系统理论及其应用[M]. 北京:科学出版社,1999.
Liou Sifeng, Guo Tianbang, Dang Yaoguo. Grey Systems Theory and Its Application [M]. Beijing: Science Press, 1999.
- [9] 许东,吴铮. 基于MATLAB 6. X的系统分析与设计——神经网络[M]. 西安:西安电子科技大学出版社,2002.
Xu Dong, Wu Zheng. Based on MATLAB 6. X System Analysis and Design—Neural Network[M]. Xian: Xidian University of Science & Technology Press, 2002.