

BP 神经网络性能与隐藏层结构的相关性探究

杨守建^{1,2}, 陈 愚^{1*}

(1. 宁波大学 信息科学与工程学院, 浙江 宁波 315211; 2. 同济大学 电子与信息工程学院, 上海 201804)

摘要: 利用 BP 神经网络对多个给定的复杂非线性系统控制进行定量研究, 着重讨论了 BP 神经网络因隐藏层层数和网络学习率之间差异从而引起对复杂非线性系统控制性能上的影响. 通过实验数据的对比分析发现, BP 神经网络隐藏层层数的递增与系统控制性能的提升并不成正相关性, 网络学习率的选取范围可控制在 0~2.0 之间, 具体参数因控制对象而异, 可采用分段调试和二分法运算以确定最佳网络学习率参数.

关键词: BP 神经网络; 隐藏层层数; 网络学习率; 多项式拟合

中图分类号: TP391.4

文献标志码: A

文章编号: 1001-5132 (2013) 01-0048-05

自 20 世纪 80 年代 BP 神经网络模型提出以来, 由于其相当强的非线性映射能力和灵活的网络结构等优点, 使其在模式匹配与分类、图像处理、智能控制、天气预测、市场估计等领域获得了广泛的应用. BP 神经网络之所以具有极强的信息处理能力, 其根本原因有两点: 第一采用误差反向传播算法; 第二来源于隐藏层神经元的学习. 但 BP 神经网络并非完美无瑕, 其性能与网络结构及网络学习率的相关性等问题尚待进一步研究^[1-5]. 笔者以多个复杂非线性系统为分析对象, 在讨论 BP 神经网络有关特性的基础上, 重点分别对 BP 神经网络隐藏层层数和网络学习率的不同从而导致对复杂非线性系统控制性能上的差异进行了定量研究与分析.

1 算法原理

1.1 BP 神经网络的拓扑模型

BP(Back-Propagation)神经网络是一种典型的多层前向神经网络, 层与层之间多采用全互联, 同层神经元之间无连接^[2,6]. 给定非线性系统的三层(1 个隐藏层)和四层(2 个隐藏层)BP 神经网络的拓扑结构如图 1 所示.

依次类推, 可以绘制出五层(3 个隐藏层)和六层(4 个隐藏层)BP 神经网络的拓扑结构图, 本文就不再一一详细绘制. BP 神经网络的每个神经元都是一个连续型的 S 型神经元. S 型神经元的数学模型有多种, 笔者采用的 S 型神经元的数学模型为双曲正切函数:

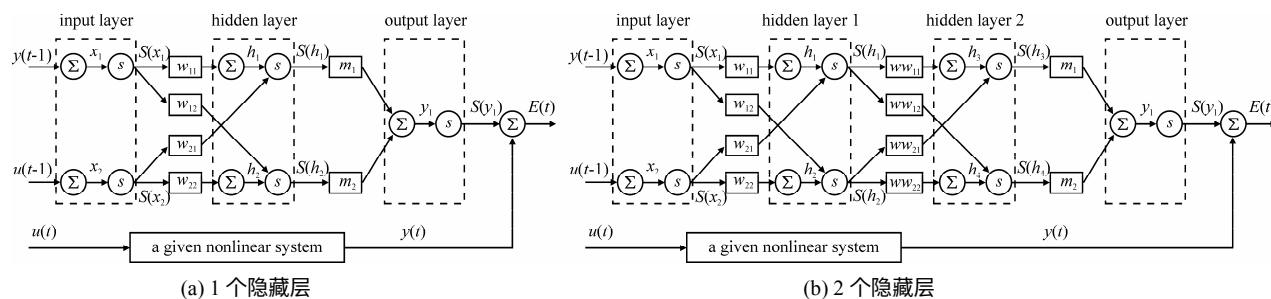


图 1 BP 网络的拓扑结构

收稿日期: 2012-09-08.

宁波大学学报(理工版) 网址: <http://nbdxlg.paperopen.com/>

基金项目: 宁波市自然科学基金 (2010A610109).

第一作者: 杨守建 (1989—), 男, 浙江长兴人, 在读硕士研究生, 主要研究方向: 电气自动化和人工智能. E-mail: 1233537@tongji.edu.cn

*通信作者: 陈 愚 (1962—), 男, 重庆人, 博士/副教授, 主要研究方向: 图像与视频处理、人工智能. E-mail: chenken@nbu.edu.cn

$$S(x) = K \left(\frac{2}{1 + e^{-dx}} - 1 \right) = K \left(\frac{1 - e^{-dx}}{1 + e^{-dx}} \right), \quad (1)$$

其中, K 为增益系数; d 为指数系数. 选择上述 S 型函数的要点在于: 改变 K 值可以相应地调整 S 型函数的输出动态范围 ($-K \sim K$), 调节 d 值可以改善网络的收敛速度, 使 BP 神经网络具有更大的灵活性. 本实验中选取 $K = 2.0, d = 1.0$.

1.2 非线性系统的选取

为充分考验 BP 神经网络的性能, 并使本文得出的结论更具有说服力, 因此选取多个复杂非线性系统进行研究, 现以其中之一作论证分析:

$$y(k) = ((0.8 + 0.5u^2(k-1))y(k-1) + u^3(k-1) + 0.5[0.5 - 0.5u^2(k-1)]e(k-1)) / (1 + u(k-1) + u^2(k-1) + e(k)), \quad (2)$$

其中, $u(k)$ 是一个独立、呈均匀分布的输入信号, 其平均值为零, 方差约为 0.1. $e(k)$ 表示高斯白噪声, 其平均值为零, 方差约为 0.04. 系统的输出不仅取决于系统前一刻的输入与输出, 同时还受当前与前一时刻随机产生的高斯白噪声影响.

1.3 BP 神经网络算法

BP 神经网络是一种前向网络, 其网络结构如图 1 所示为单向网络结构, 但其信息的传播却是双向的. 其学习过程分为两步^[7-9]: 首先是前向传播过程, 不断更新各层状态; 其次是误差逆传播过程, 将实际输出与期望输出的误差由输出层反向传播至输入层, 并修正相应权值.

以图 1(a)三层 BP 神经网络为例, 推导 BP 神经网络学习过程如下:

$$h_1 = S(x_1) \cdot w_{11} + S(x_2) \cdot w_{21}, \quad (3)$$

$$y_1 = S(h_1) \cdot m_1 + S(h_2) \cdot m_2, \quad (4)$$

$$E(t) = \frac{1}{2} [y(t) - S(y_1)]^2, \quad (5)$$

$$\frac{\partial S(x)}{\partial x} = K \cdot \left[\frac{2.0}{1.0 + \exp(-d \cdot x)} - 1.0 \right], \quad (6)$$

$$\begin{aligned} \Delta w_{11} = -C \cdot \frac{\partial E(t)}{\partial w_{11}} = -C \cdot \frac{\partial E(t)}{\partial S(y_1)} \cdot \frac{\partial S(y_1)}{\partial y_1} \cdot \frac{\partial y_1}{\partial h_1} \cdot \frac{\partial S(h_1)}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_{11}} = C \cdot [y(t) - S(y_1)] \cdot \frac{\partial S(y_1)}{\partial y_1} \cdot m_1 \cdot \frac{\partial S(h_1)}{\partial h_1} \cdot S(x_1), \end{aligned} \quad (7)$$

$$w_{11}(t+1) = w_{11}(t) + C \cdot [y(t) - S(y_1)] \cdot \frac{\partial S(y_1)}{\partial y_1}.$$

$$m_1 \cdot \frac{\partial S(h_1)}{\partial h_1} \cdot S(x_1). \quad (8)$$

同理可得:

$$\begin{aligned} w_{22}(t+1) = -C \cdot \frac{\partial E(t)}{\partial w_{22}} = w_{22}(t) + C \cdot [y(t) - S(y_1)] \cdot \frac{\partial S(y_1)}{\partial y_1} \cdot m_2 \cdot \frac{\partial S(h_2)}{\partial h_2} \cdot S(x_2), \end{aligned} \quad (9)$$

$$\begin{aligned} w_{12}(t+1) = -C \cdot \frac{\partial E(t)}{\partial w_{12}} = w_{12}(t) + C \cdot [y(t) - S(y_1)] \cdot \frac{\partial S(y_1)}{\partial y_1} \cdot m_2 \cdot \frac{\partial S(h_2)}{\partial h_2} \cdot S(x_1), \end{aligned} \quad (10)$$

$$\begin{aligned} w_{21}(t+1) = -C \cdot \frac{\partial E(t)}{\partial w_{21}} = w_{21}(t) + C \cdot [y(t) - S(y_1)] \cdot \frac{\partial S(y_1)}{\partial y_1} \cdot m_1 \cdot \frac{\partial S(h_1)}{\partial h_1} \cdot S(x_2), \end{aligned} \quad (11)$$

$$\begin{aligned} m_1(t+1) = -C \cdot \frac{\partial E(t)}{\partial m_1} = m_1(t) + C \cdot [y(t) - S(y_1)] \cdot \frac{\partial S(y_1)}{\partial y_1} \cdot S(h_1), \end{aligned} \quad (12)$$

$$\begin{aligned} m_2(t+1) = -C \cdot \frac{\partial E(t)}{\partial m_2} = m_2(t) + C \cdot [y(t) - S(y_1)] \cdot \frac{\partial S(y_1)}{\partial y_1} \cdot S(h_2), \end{aligned} \quad (13)$$

其中, C 为网络学习率, $x_1 = y(t-1)$, $x_2 = u(t-1)$. 从以上推导可以看出, BP 神经网络采用有导师学习机制, 当网络收敛, 学习过程结束, 输入信息(或其规律)和相应的权值便储存于网络之中. 网络联想回忆过程与学习过程类似, 但仅有前向传播过程. 在四至六层 BP 神经网络中, 每增加 1 个隐藏层, 相应层与层之间的权值便增加 4 个, 但神经网络的输出和权值的求取方法以及学习联想过程与三层 BP 神经网络一致, 此处不再详述.

2 仿真研究

2.1 隐藏层层数的研究

BP 神经网络之所以具有相当强的非线性映射能力和灵活的网络结构等优点, 是因为在输入层和输出层之间添加了隐藏层. BP 神经网络采用导师制的学习方式, 由隐藏层对输入信息的特征或规律进行提取, 结果表现为相应连接权值的“自调

整”过程,但隐藏层层数与网络性能的关系,具体层数如何选取目前尚无理论指导^[10-11]。笔者以(2)式所示的非线性系统为研究对象,研究了包含一至四层隐藏层的BP神经网络对该非线性系统的跟随性能。测试过程中的MSE数据见表1,学习与测试过程中误差 $E(t)=[y(t)-S(y_1)]^2/2$ 的1阶和3阶多项式拟合如图2和图3所示。

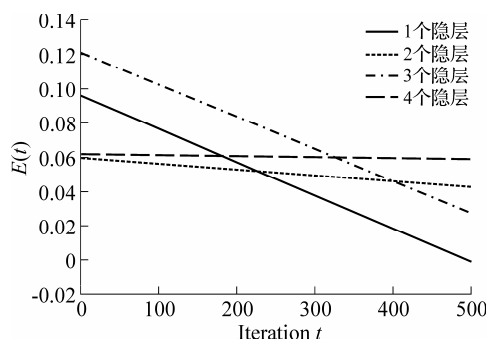
表1 测试过程的均方差

	1个隐层	2个隐层	3个隐层	4个隐层
测试1	0.062 4	0.105 3	0.065 7	0.113 8
测试2	0.067 5	0.108 5	0.068 7	0.116 4
测试3	0.069 2	0.100 6	0.069 2	0.108 3
测试4	0.065 5	0.101 3	0.067 6	0.109 9
测试5	0.065 4	0.120 9	0.068 8	0.131 1

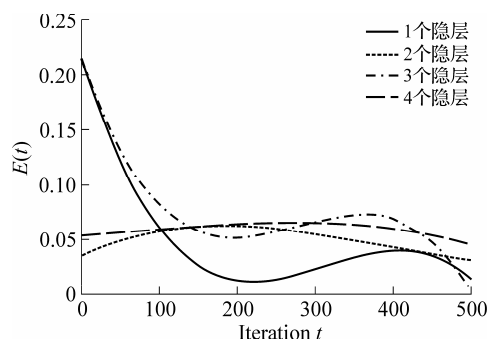
定义均方差为:

$$MSE = \frac{1}{M} \cdot \sum_{i=1}^M (y(t) - S(y_1))^2,$$

其中, M 为测试点数, 本实验取 $M = 500$; $y(t)$ 为期望输出, $S(y_1)$ 为实际输出, 输出范围均为-1.0~1.0。并根据实际测试结果定义: 当 $MSE < 0.070$ 且 $E(t) < 0.035$ 时的跟随效果为理想效果。

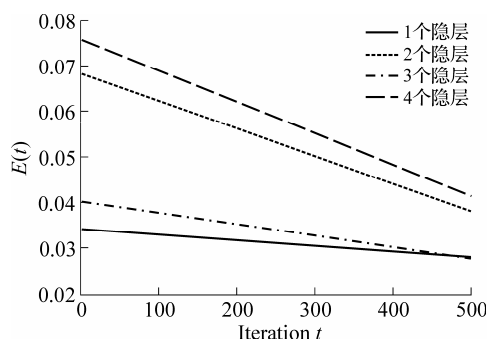


(a) 1阶多项式拟合

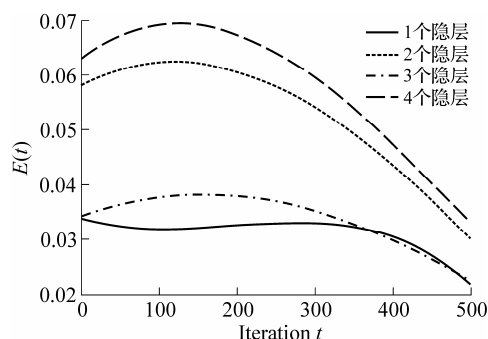


(b) 3阶多项式拟合

图2 学习过程中误差 $E(t)$ 多项式拟合



(a) 1阶多项式拟合



(b) 3阶多项式拟合

图3 测试过程中误差 $E(t)$ 多项式拟合

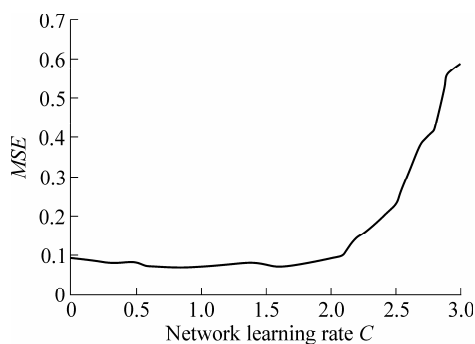
实验数据显示,三层(1个隐藏层)BP神经网络的跟随性能最好,然后依次为五层(3个隐藏层)、四层(2个隐藏层)、六层BP(4个隐藏层)神经网络。相对于四至六层BP神经网络,三层(1个隐藏层)BP神经网络的均方差MSE最小,网络收敛速度更快,网络稳定性更强。其次,三层(1个隐藏层)、五层(3个隐藏层)BP神经网络的跟随效果与四层(2个隐藏层)、六层(4个隐藏层)BP神经网络相比要明显好得多。由此可以得出以下两点结论:(1)BP神经网络隐藏层层数的递增与系统控制性能的提升并不成正相关性,亦不成负相关性。三层(1个隐藏层)BP神经网络的跟随性能最好,盲目增加隐藏层层数对改善BP神经网络性能收效甚微;(2)隐藏层层数为奇数的BP神经网络性能在一定程度上明显优于隐藏层层数为偶数的BP神经网络。对相同控制对象,隐藏层层数为奇数的BP神经网络其网络稳定性更好,输出收敛速度更快,输出误差更小。

2.2 网络学习率的研究

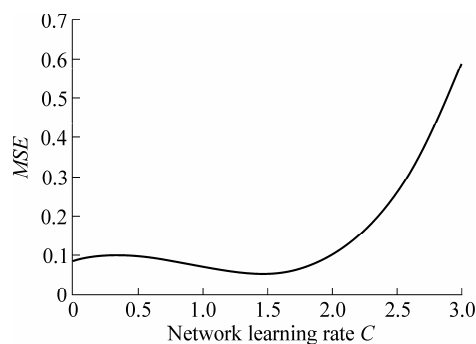
评价BP神经网络性能的好坏主要有两点:第一,网络的稳定性,即BP神经网络的权值是否收敛,输出是否有界;第二,网络的收敛速度。而BP

表 2 不同参数 C 测试下的均方差

学习率 C	均方差	学习率 C	均方差	学习率 C	均方差	学习率 C	均方差	学习率 C	均方差	学习率 C	均方差
0.000 05	0.094 4	0.4	0.083 3	1.3	0.082 2	2.2	0.145 9	5.0	0.610 0	200.0	1.422 4
0.000 5	0.094 3	0.5	0.093 1	1.4	0.083 9	2.4	0.163 9	10.0	0.636 8	500.0	1.570 8
0.005	0.094 3	0.6	0.073 5	1.5	0.079 6	2.4	0.198 1	15.0	0.644 5	800.0	1.679 3
0.001	0.094 2	0.7	0.072 6	1.6	0.071 6	2.5	0.225 9	20.0	0.642 3	1 000.0	1.548 2
0.05	0.107 8	0.8	0.072 1	1.7	0.074 1	2.6	0.309 4	25.0	0.651 8	2 000.0	网络振荡
0.01	0.093 0	0.9	0.068 6	1.8	0.081 5	2.7	0.386 5	30.0	0.650 8	5 000.0	网络振荡
0.1	0.091 3	1.0	0.073 6	1.9	0.085 4	2.8	0.417 3	40.0	0.704 1	10 000.0	网络振荡
0.2	0.087 9	1.1	0.074 4	2.0	0.092 2	2.9	0.561 2	50.0	0.704 6	50 000.0	网络振荡
0.3	0.080 5	1.2	0.076 3	2.1	0.100 4	3.0	0.589 4	100.0	0.694 3	100 000.0	网络振荡



(a) 未拟合



(b) 3 阶多项式拟合

图 4 不同参数 C 测试下的均方差

算法中主要影响收敛速度的关键因素之一便是网络学习率 C 。目前, 网络学习率 C 值的选取尚无理论指导, 需要凭设计者经验并通过实验选取^[12]。通过上述实验发现, 相比其他 BP 神经网络而言, 三层 BP 神经网络的控制性能最好。因此只选取三层 (1 个隐藏层) BP 神经网络和上述非线性系统为研究对象, 通过采取分段调试和二分法迭代的算法, 测试不同参数 C 下 BP 神经网络的输出与收敛状况。部分实验数据及其处理结果分别见表 2 和图 4, 每组测试点数为 2000 点。

实验结果表明, 网络学习率 C 选取过小或过大都不利于 BP 神经网络的收敛与稳定。 C 值取得过小, 网络学习速度较慢, 则须过较长时间才可能达到收敛状态; C 值取得过大, 又易于引起网络振荡, 使其根本难以收敛。由于之前定义了当 $MSE < 0.070$ 且 $E(t) < 0.035$ 时的效果为理想效果, 从表 2 中可以看出, 当 C 的范围为 0.6~1.1 或 1.6~1.7 之间时, 其跟随效果认为是理想的。

结合实际设计经验并考虑各种因素的影响,

从一定程度上来说, 网络学习率 C 的值选取圈定在 0.6~1.0 范围之间最为合适, 但这并不是绝对的。从更广的角度来看, 网络学习率 C 的值可以完全确定在 0 (取不到 0)~2.0 范围之内。然后采用分段调试的方法, 将范围 0~2.0 等分成若干区间段, 取各区间段中的任意 1 个参数进行调试, 从而近似确定出参数 C 最适值可能存在的区间段, 以缩小参数 C 的选取范围。再在上述选出的区间段中, 进一步将区间段细分, 进一步缩小取值范围, 进而找出参数 C 最适值所在的 1~2 个区间段, 如区间段 A 和区间段 B。然后对区间段 A 和区间段 B 采用二分法迭代运算, 直至求出适合于相应 BP 神经网络的一个或两个最佳网络学习率 C 。

3 结语

通过构造多个复杂的非线性系统, 在讨论 BP 神经网络有关特性的基础上, 重点分别对 BP 神经网络隐藏层数和网络学习率的不同从而导致对复杂非线性系统控制性能上的差异进行了定量研

究与分析. 实验结果表明, 三层 BP 神经网络的性能是最佳的, BP 神经网络隐藏层层数的递增与系统控制性能的提升并不成正相关性, 亦非负相关性. 且隐藏层层数为奇数的 BP 神经网络的性能在一定程度上明显优于隐藏层层数为偶数的 BP 神经网络. 其次, 网络学习率 C 值取得过小或过大都不利于 BP 神经网络的收敛与稳定, 实际应用中其最适值选取范围绝大多数情况下可以圈定在 0.6~1.0 范围之内. 保守情况下亦可在 0~2.0 的数值区间段内采用分段调试和二分法迭代的手段来确定 1~2 个最佳的网络学习率 C . 文章对 BP 神经网络在实际应用中关于隐藏层层数和网络学习的选取的问题上具有一定积极的指导意义.

参考文献:

- [1] 周开利, 康耀红. 神经网络模型及其 MATLAB 仿真[M]. 北京: 清华大学出版社, 2006:2.
- [2] 朱双东. 神经网络应用基础[M]. 沈阳: 东北大学出版社, 2000:3.
- [3] Wei C. Efficiency evaluation method of simulation system based on BP neural network[C]. Proceedings 2nd International Conference on Networking and Distributed Computing, 2011:41-44.
- [4] Peng L, Bao L, Du C. The evaluation of scientific reasoning ability based on BP neural network[J]. Communications in Computer and Information Science, 2011, 163:133-139.
- [5] Izeboudjen N, Bouridane A, Farah A, et al. Application of design reuse to artificial neural networks: case study of the back propagation algorithm[J]. Neural Computing and Applications, 2011, 21:1-14.
- [6] 张国翊, 胡铮. 改进 BP 神经网络模型及其稳定性分析[J]. 中南大学学报:自然科学版, 2011, 42(1):115-124.
- [7] 毛永毅, 周康磊. 基于 BP 神经网络的定位跟踪算法[J]. 计算机工程与应用, 2011, 47(20):238-240.
- [8] 车建国, 耿学和. 基于改进算法的 BP 神经网络在故障诊断中的应用[J]. 战术导弹控制技术, 2011, 28(1):45-47.
- [9] Li Y, Tian Y, Chen W. Multi-pattern recognition of sEMG based on improved BP neural network algorithm[C]. Proceedings of the 29th Chinese Control Conference, 2010:2867-2872.
- [10] Meng X, Meng X. Nonlinear system simulation based on the BP neural network[C]. Proceedings 3rd International Conference on Intelligent Networks and Intelligent Systems, 2010:334-337.
- [11] Dong C, Dong L, Yang M. The application of the BP neural network in the nonlinear optimization[J]. Advances in Intelligent and Soft Computing, 2011, 78:727-732.
- [12] 张华节, 李瑾. 学习速率连续变化的 BP 神经网络算法[J]. 四川兵工学报, 2011, 32(4):61-64.

Effect of Multi-hidden-layer on Performance of BP Neural Network

YANG Shou-jian^{1,2}, CHEN Ken^{1*}

(1.Faculty of Information Science and Technology, Ningbo University, Ningbo 315211, China;

2.School of Electronics and Information, Tongji University, Shanghai 201804, China)

Abstract: In this paper, quantitative study on a given complex nonlinear system is presented using BP neural network. The correlation is explored between the control performance on the given complex nonlinear system and the number of hidden layers and the difference of network learning rate. The analytical and test results well indicate that the improvement of control performance is not proportional to the increased number of hidden layers, and the optimal selection of neural learning rate is generally within 0-2. The specific value of learning rate depends on the specifically given system and the most suitable value range can be determined using sub-debugging method.

Key words: BP neural network; the number of hidden layers; network learning rate; curve fitting

(责任编辑 章践立)