

文章编号: 0253—987X (2002)01-0054-04

# 目标函数对神经网络性能影响的研究

杜海峰, 王孙安

(西安交通大学机械工程学院, 710049 西安)

**摘要:** 通过对常用目标函数的性能、网络的收敛性及其统计特性影响的系统分析, 给出了基于李雅普洛夫定理的有关收敛性的理论证明. 理论分析和相关实验表明, 神经网络采用选择性强的目标函数其逼近能力也强, 而选择合适的目标函数, 可以保证神经网络逼近是对函数的无偏估计.

**关键词:** 目标函数; 神经网络; 收敛性

**中图分类号:** TP273 **文献标识码:** A

## Effect of Object Function on the Performance of the Neural Network

Du Haifeng, Wang Sun'an

(School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract:** The capability of the general object functions is introduced. The effect of the object function on the statistical performance and astringency of the neural network (NN) is analyzed. Based on Lyapunov theorem, the astringency of the NN is proved under the general object function. The approximate speed and the statistical performance are explored under the different object function. The theoretica analysis and experiments indicate that the astringency can be improved through the better selectivity object function, and the appropriate object function can guarantee that NN approximates to the function completely.

**Keywords:** object function; neural network; astringency

### 1 常用目标函数的性能

目标函数应该具有选择性、有效性和实用性. 选择性是指在该目标函数下系统的最佳性能与非最佳性能有明显的区别; 有效性是要求目标函数必须是系统参数的函数, 这样目标函数才能反映系统的性能, 并且该目标函数存在极值; 实用性是指所选取的目标函数应该易于计算、分析和实验, 最好具有明确的物理意义<sup>[1]</sup>.

神经网络中常用的目标函数  $J$  主要是平方误差积分

$$\int_0^T e^2(t) dt^{[2]}$$

但是, 其他常用的函数还包括

$$\int_0^T te^2(t) dt, \int_0^T |e(t)| dt \text{ 和 } \int_0^T t |e(t)| dt$$

式中:  $t$  代表时间, 而  $\int_0^T e^2(t) dt$  的物理意义明确, 着重衡量大的误差, 较少考虑小误差, 其选择性不好. 对于初始误差较大的系统, 该指标有迅速减小误差

的倾向, 因此系统的响应迅速, 可能有振荡, 系统的相对稳定性较差;  $\int_0^T te^2(t)dt$  对起始误差考虑较少, 但考虑了稳态响应后的误差, 具有较好的选择性;  $\int_0^T |e(t)|dt$  的选择性不太好, 但是基于该准则设计的系统具有比较好的瞬态响应, 因此易于计算机实现;  $\int_0^T t|e(t)|dt$  同  $\int_0^T te^2(t)dt$  一样, 但它是  $\int_0^T |e(t)|dt$  的改进.

图 1 所示的是几种目标函数对同一系统在不同参数下的变化情况, 所研究的系统是

$$G(s) = \frac{1}{s(s+\zeta)} \tag{1}$$

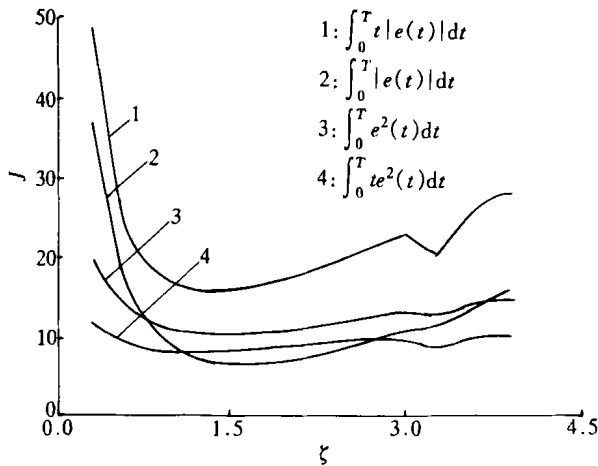


图 1 目标函数的性能曲线

采用不同的目标函数设计系统所得到的系统优化参数  $\zeta$  略有差别, 但大多在 1.4 附近, 只有采用  $\int_0^T |e(t)|dt$  时,  $\zeta$  约为 1.6. 另外,  $\int_0^T te^2(t)dt$ ,  $\int_0^T t|e(t)|dt$  的选择性比较强.

2 目标函数对神经网络收敛性的影响

多数神经网络都以误差及其相应的函数作为目标函数  $J$ , 而学习方法即联接权值的修改, 一般采用  $J$  的负梯度, 即

$$W(k+1) = W(k) - \eta \nabla J \tag{2}$$

式中:  $W$  是权值;  $k$  是学习步数;  $\eta$  是系数;  $\nabla J$  是目标函数的梯度. 显然,  $\nabla J$  将影响到网络的性能.

设  $y$  是网络的实际输出,  $y^d$  是期望输出,  $W$  是权系数, 取  $Y_{\max} = \max_k \left\| \frac{\partial y(k)}{\partial W} \right\|$ , 其中  $\| \cdot \|$  取欧氏

范数, 则关于  $\nabla J$  对网络学习收敛的影响, 定理 1 给出了充分条件.

定理 1 设目标函数  $J$  是误差  $e=y^d-y$  的函数, 且

$$\nabla J = \frac{\partial J}{\partial W} = ef(e) \left[ \frac{\partial e}{\partial} \right], f(e(k)) > 0, \text{ 则式(2)中的 } \eta \text{ 取值满足 } 0 < \eta < \frac{2}{f(e(k))Y_{\max}}, \text{ 学习算法式(2)一定收敛.}$$

证明 取离散李雅普洛夫函数为

$$V(k) = \frac{e^2(k)}{2} = \frac{(y^d(k) - y(k))^2}{2} \tag{3}$$

因为

$$\begin{aligned} \nabla V(k) &= V(k+1) - V(k) = \\ &= \frac{1}{2} (e^2(k+1) - e^2(k)) \end{aligned} \tag{4}$$

所以

$$\begin{aligned} \nabla V(k) &= (e(k+1) - e(k))(e(k) + \frac{1}{2}(e(k+1) - e(k))) = \\ &= \nabla e(k)(e(k) + \frac{1}{2} \nabla e(k)) \end{aligned} \tag{5}$$

$y(k)$  与权系数  $W=(W_1, W_2, \dots, W_n)$  之间是线性关系, 因此由一阶泰勒级数展开

$$\begin{aligned} \nabla e(k) &= e(k+1) - e(k) = \left[ \frac{\partial e(k)}{\partial W} \right]^T [W(k+1) - W(k)] = \\ &= \left[ \frac{\partial e(k)}{\partial W} \right]^T \nabla W \end{aligned} \tag{6}$$

又因为  $\nabla W = -\eta \nabla J, \nabla J = \frac{\partial J}{\partial W} = e(k)$

$$f(e(k)) \left[ \frac{\partial e(k)}{\partial W} \right], \frac{\partial e(k)}{\partial W} = -\frac{\partial y(k)}{\partial W}$$

$$\begin{aligned} \text{所以 } \Delta V(k) &= -\eta^2 e^2(k) f^2(e(k)) \left\| \frac{\partial y(k)}{\partial W} \right\|^2 + \\ &= \frac{1}{2} \eta^2 e^2(k) f^2(e(k)) \left\| \frac{\partial y(k)}{\partial W} \right\|^4 = \\ &= -\frac{1}{2} \eta^2 e^2(k) \left\| \frac{\partial y(k)}{\partial W} \right\|^2 (2f(e(k)) - \eta f^2(e(k)) \left\| \frac{\partial y(k)}{\partial W} \right\|^2) \end{aligned} \tag{7}$$

为了保证算法收敛, 必须

$$2f(e(k)) - \eta f^2(e(k)) \left\| \frac{\partial y(k)}{\partial W} \right\|^2 > 0 \tag{8}$$

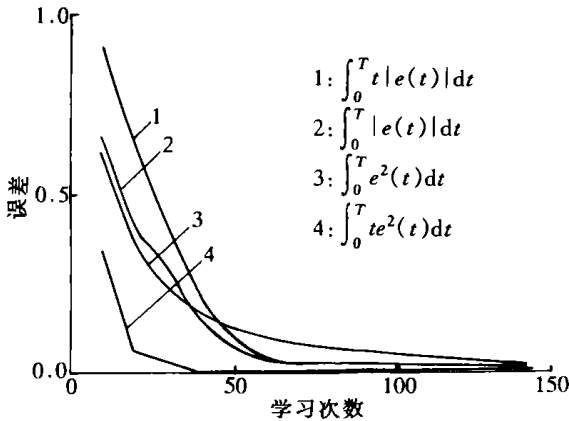
因为  $f(e(k)) > 0, Y_{\max} = \max_k \left\| \frac{\partial y(k)}{\partial W} \right\|$

所以当  $0 < \eta < \frac{2}{f(e(k))Y_{\max}}$  时,  $\nabla V(k) < 0$ , 算法收敛, 证毕.

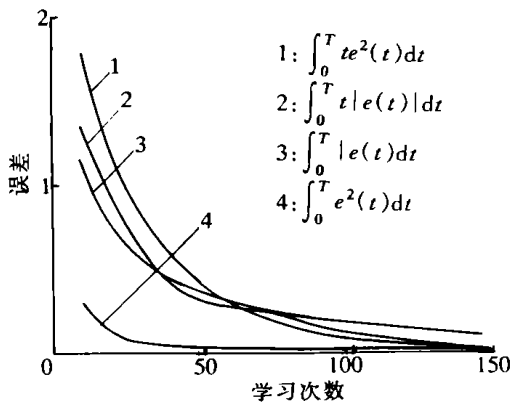
由于  $f(e(k))$  是由目标函数确定的, 因此在  $Y_{\max}$  确定后,  $\eta$  主要由目标函数确定. 因为上述 4 个

性能指标均满足定理 1 的条件, 所以能保证在该目标函数的指导下算法收敛. 特别当<sup>[3]</sup>  $J = \frac{1}{2} \sum_k e^2(k)$  时,  $f(e(k)) = 1$ , 则  $0 < \eta < \frac{2}{Y_{\max}}$ .

对于网络学习, 参数  $\eta$  的取值会影响到收敛速度. 采用式(2)的学习方法, 基于相同的网络结构, 分别采用上述 4 种目标函数, 对函数  $f_1(x) = \sin(\pi x)$  和  $f_2(x) = \sin(\pi x)e^{-x}$  ( $x \in [-1, +1]$ ) 逼近, 不同学习次数后, 误差变化结果如图 2 所示. 为了便于比较, 网络的初始权值均取为 0.5, 图中的误差均以误差平方和计算.



(a) 对  $f_1(x)$  的逼近情况



(b) 对  $f_2(x)$  的逼近情况

图 2 目标函数对网络逼近速度的影响

网络的逼近速度受目标函数选取的影响较大, 但是采用选择性强的目标函数的网络逼近能力强, 这是因为网络学习本质上就是在目标函数构成的搜索曲面上连续不断地调整权系数, 以求寻找到最小值的过程. 不同的目标函数构成的搜索曲面显然不一样, 选择性强的目标函数能将系统的最佳性能与非最佳性能明显区别, 构成的搜索曲面“陡”, 而选择性弱的目标函数构成的搜索曲面相对“光滑”, 因此

影响了网络的收敛速度, 所以通过提高目标函数的选择性, 可以加速网络收敛.

### 3 目标函数选取对网络学习过程统计性质的影响<sup>[4, 5]</sup>

未知概率分布  $P(X)$  的样本向量  $X$  是描述某物理现象的独立变量,  $D$  是在未知条件分布  $P(D|X)$  下的响应输出, 其中  $X = (X_1, X_2, \dots, X_n)$ ,  $D = (D_1, D_2, \dots, D_n)$ . 对于复杂系统, 很难知道  $D$  与  $X$  间的确切函数关系, 即模型未知, 从统计的角度可以把这种关系写为

$$D = g(X) + \epsilon$$

式中:  $\epsilon$  为随机变量, 描述了  $D$  与  $X$  关系中的未知部分. 上述模型是根据  $X$  描述  $D$  的数学模型, 而神经网络可以看作是实现这一目的的工具. 设网络的输出

$$Y = f(X, W) \quad (9)$$

此时, 函数  $g(X)$  定义为

$$g(X) = E(D | X) \quad (10)$$

即给定  $X$  后,  $D$  的期望称为  $D$  对  $X$  的回归. 设: ① 给定  $X$  后  $\epsilon$  的均值为 0, 即  $E(\epsilon | X) = 0$ ; ②  $\epsilon$  与  $g(X)$  不相关, 即  $E(\epsilon g(X)) = 0$ .

分别定义目标函数为  $J_1, J_2, J_3$ , 其中

$$J_1(W) = \frac{1}{2} (E(\epsilon | e))^2$$

$$J_2(W) = \frac{1}{2} E e^2 \text{ 和 } J_3(W) = \frac{1}{2} E (te^2)$$

当  $t \geq 1$  时, 由许瓦兹不等式可以得到

$$E(\epsilon | e) \leq (E | e^2 |)^{1/2} \leq (E | te^2 |)^{1/2} \quad (11)$$

因为

$$J_2 = 0.5 E(e^2) = 0.5 E((D - Y)^2) = 0.5 E((D - f(X, W))^2) \quad (12)$$

所以

$$J_2 = 0.5 E[(D - g(X) + g(X) - f(X, W))^2] = 0.5 E[(D - g(X))^2] + 0.5 E[(g(X) - f(X, W))^2] \quad (13)$$

由于上式中与网络有关的变量只有  $W$ , 因此欲使  $J_2$  最小, 以等价于使  $E[(g(X) - f(X, W))^2]$  极小. 显然, 在目标函数  $J_2$  下求出的  $W$ , 使  $f(X, W)$  成为在均方误差意义下对  $g(X)$  的最佳逼近. 由式(11)可知, 在目标函数  $J_3$  下求出的  $W$  使  $f(X, W)$  对  $g(X)$

(X)的逼近更精确. 另外

$$E[(g(X)-f(X,W))^2] = (E(f(X,W))-g(X))^2 + E(f(X,W)-E(f(X,W))^2)$$

(14)

式中:  $E(f(X,W)-g(X))$ 代表了  $f(X,W)$ 与  $g(X)$ 间的偏置;  $E(f(X,W)-E(f(X,W))^2)$ 代表了  $f(X,W)$ 本身的方差. 当  $f(X,W)=g(X)$ 时, 称  $f(X,W)$ 是  $g(X)$ 的无偏估计, 否则  $f(X,W)$ 是  $g(X)$ 的有偏估计. 由式(11)可知, 由于  $J_1$  的统计性能优于  $J_2$ , 因此能保证  $f(X,W)$ 是  $g(X)$ 的无偏估计, 而  $J_3$  不一定能保证.

4 结 论

通过上述讨论和相关实验研究, 可以得到以下结论.

(1)目标函数的选取对神经网络学习的收敛性

有较大影响, 但可以依据定理 1 保证算法收敛.

(2)神经网络采用选择性强的目标函数可以提高逼近速度和精度.

(3)选择适当的网络目标函数可以改善网络学习的统计性能, 从而保证网络是对函数的无偏估计.

参考文献:

[ 1 ] 绪方胜彦. 现代控制工程[ M ]. 卢伯英, 等译. 北京: 科学出版社, 1978. 226 ~ 228.

[ 2 ] 焦李成. 神经网络计算[ M ]. 西安: 西安电子科技大学出版社, 1993. 38 ~ 88.

[ 3 ] Chen Y C, Teng C C. A model reference control structure using fuzzy neural network[ J ]. Fuzzy Sets and System, 1995, 73(3): 291 ~ 312.

[ 4 ] Geman S. Neural networks and the bias/variance dilemma[ J ]. Neural Computation, 1992, (4): 1 ~ 58.

[ 5 ] 阎平凡. 人工神经网络的容量、学习与计算复杂性[ J ]. 电子学报, 1995, 23(4): 63 ~ 67.

(编辑 管咏梅)

(上接第 53 页)

4 结 论

本文采用激振实验的方法, 对纸张在空气中振动时的阻尼特性进行了研究, 提出了在纸张动态响应有限元的分析中, 阻尼矩阵的确定方法和详细步骤, 并对该方法引入的误差及减小误差的措施进行了探讨. 以该方法来确定阻尼矩阵, 对悬臂状态纸张的动态响应进行了有限元分析, 分析结果与实测结果相当吻合, 表明该方法确定的阻尼矩阵能够较真实地反映纸张在空气中振动的阻尼特性.

致谢 感谢日本理光公司中央研究所竹平修先生为论文所提供的实验数据.

参考文献:

[ 1 ] 马军星, 杨胜军, 何正嘉, 等. 办公设备用纸力学特性的非线性有限元分析[ J ]. 西安交通大学学报, 2000, 34(9): 67 ~ 71.

[ 2 ] 王勖成, 邵 敏. 有限单元法基本原理和数值方法[ M ]. 第 2 版. 北京: 清华大学出版社, 1997. 446 ~ 448.

[ 3 ] 杨胜军, 马军星, 何正嘉. 办公设备纸张自由振动阻尼特性研究[ J ]. 西安交通大学学报, 2000, 34(7): 48 ~ 51.

[ 4 ] 张云电, 陈光雄. 钢质薄壁缸套的模态分析[ J ]. 内燃机, 1985, 5: 22 ~ 24.

[ 5 ] 倪振华. 振动力学[ M ]. 西安: 西安交通大学出版社, 1989. 61 ~ 69.

(编辑 管咏梅)