

2 常规 BP 算法中学习率存在的问题^[2]

图4描述了 SSE 曲面与权值 W 与阈值 B 的关系。图5、图6分别为相应的 SSE 等高线与 W 、 B 的关系以及随学习次数 Epoch 的增加 SSE 不断减小的过程。图5清楚地表明, W 及 B 值从任意选择的初始值位置“+”, 沿着 SSE 等高线的垂直方向(也即 SSE 曲面的最陡下降方向)调整, 直到 SSE 的最小值位置“ $\times$ ”为止。(注: 为节省版面, 计算机仿真时打印的原图被缩小, 图5、图7中“ W 、 B 变化的轨迹垂直于 SSE 等高线”反应得不够清楚。)

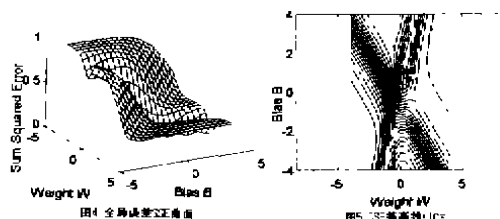


图4 全局误差SSE曲面

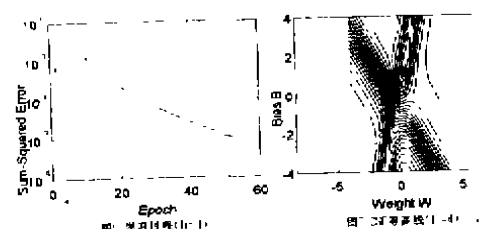


图5 学习过程(1-4)

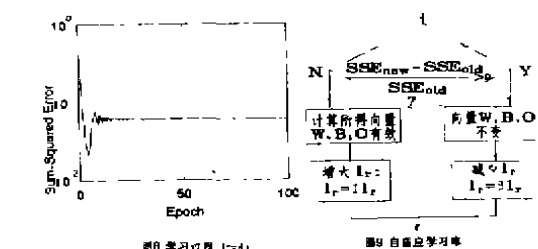


图6 学习过程(1-4)

图5和6是在 $l_r=1$ 情况下的仿真结果, 在其它参数完全相同的条件下, 若增大学习率, 取 $l_r=4$, 则其结果如图7与8所示。图7虽也能像图5那样沿误差 SSE 曲面的最陡下降方向接近“ $\times$ ”点, 但由于 l_r 取值过大, 出现在“ $\times$ ”点附近等幅振荡而不能收敛于该点的现象。图8也清楚地反应了这一点。

大量地仿真试验表明: 在 l_r 为常数的常规 BP 算

法中, 由于神经网络的非线性本质, 很难找到一种简单的确定 l_r 值的方法, 使其既能有较快的学习速度, 又能最终收敛, 为此作者对自适应学习率进行了研究。

3 自适应学习率^[3,4]

自适应学习旨在收敛的前提下, 尽量缩短学习过程。为此, 训练过程中, l_r 不可为常数, 各次 Epoch 所用的学习率, 应按图9进行自适应修改。其基本思想是: 在学习收敛时, 增大 l_r , 以缩短学习时间; 而当 l_r 偏大导致 SSE 不能收敛时, 及时减小 l_r , 直到收敛为止。图中系数 $\alpha > 1$, $0 < \beta < 1$, $r > 1$ 。本文取, $\alpha = 1.04$, $\beta = 0.7$ 和 $r = 1.05$ 。

下面给出自适应学习率在神经网络函数逼近中的应用结果。

图10中由“+”点构成的轨迹代表欲逼近的曲线或称期望的网络输出函数。

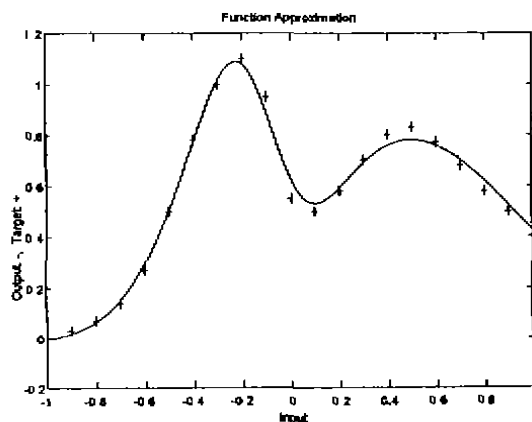
图10 函数逼近 ($SSE_{max}=0.02$)

图11(a), (b), 及(c)分别表示在网络结构及其参数完全相同情况下, 不同的学习率(恒值0.01、0.04及初始值为0.01的自适应学习率)对学习速度的影响。三个试验均是在 $SSE_{max}=0.02$ 下进行的。不难看出: 当 l_r 取值偏小(图11(a))时, 收敛速度太慢, 达到 $SSE_{max}=0.02$ 所需的学习次数 Epoch=14303; 反之, 当 l_r 值偏大(图11(b))时, 则导致 SSE 的大幅度振荡, 非但不能加速学习过程, 反而使其增长 (Epoch > 15000)。然而, 采用自适应学习率的图11(c), 给出了十分满意的结果: 达到 $SSE_{max}=0.02$ 所需的 Epoch=995, 学习速度提高14.4倍。

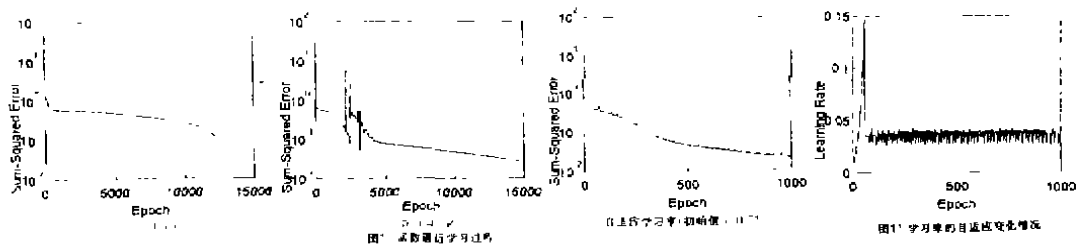


图11 自适应学习率过程

自适应学习率, 初始值: 0.1

图11 学习率的自适应变化过程

作者在各种不同初始条件下(含初始权值及初始 η 值)进行了大量仿真试验,均验证了自适应学习率在优化学习过程中的显著作用。限于篇幅,本文仅随机地给出一例。

图12给出其学习率自适应变化的情况,与图11(c)相对应。图10中的实线是经自适应学习后,神经网络的实际输出曲线。在学习速度得以显著提高的同时,满足了 $SSE_{max}=0.02$ 的要求。

4 结 论

本文以在神经网络中广泛应用的BP算法为例研究了自适应学习率在优化学习过程中的作用,在学习收敛时,能自动增大 η ,以便缩短学习时间;而当 η 过大导致SSE发散时,又能自动减小 η ,直到收敛

为止。此法用于神经网络函数逼近,并获得十分满意的结果。

参考文献

- [1] Jacek M. Zurada, Introduction to Artificial Neural System, West Publishing Company, 1993
- [2] Howard Demuth, Mark Beale, Neural Network Toolbox, The MATH WORKS Inc. 1993
- [3] Timothy Masters, Practical Neural Network Recipes in C++, Academic Press, Inc. 1993
- [4] John Hertz et al., Introduction to the Theory of Neural Computation, Addison-Wesley Publishing Company, 1991

(上接第78页)

参考文献

- [1] Hgeon H. et al., Concurrent Engineering, The Manufacturing Philosophy for The 90's, Computer Industry Engineering, 1991, 21 (4)
- [2] Robert M. et al., Organizing for Concurrent Engineering, Machine Design, 1992, 64(15)
- [3] G. Sohlenius, Concurrent Engineering, Annals of the CIRP, 1992, 41(2)
- [4] Petero G. et al., An Advice System for Concurrent Engineering, Int. Journal Computer Integrated Manufacturing, 1991, 4(2)
- [5] 查建中,郭伟等,机械产品的并行设计,机械设计,1993(4)
- [6] 牛占文,吴会林等,机械产品的并行设计及其实施策略的研究,第二届中国计算机集成制造系统(CIMS)学术会议论文集,1992,12
- [7] 查建中主编,《智能工程》,机械工业出版社,1992