BP 网络的最大误差学习算法

赵启林,卓家寿

(河海大学土木工程学院,南京 210098)

摘要 综合了标准 BP 算法与“批处理”BP 算法的各自特点,提出了一种新的 BP 网络的学习算法.该算法既具有“批处理”BP 算法收敛时迭代次数少的优点,又能克服“批处理”算法对大样本集进行学习时每次计算量较大且收敛时间长的缺点.该算法具有“批处理”算法同阶的迭代次数,但每次迭代所需计算工作量大约是“批处理”算法的样本几分之一.

关键词 BP 算法;BP 网络;批处理

中图分类号 O302　　**文献标识码** A　　**文章编号** 1000-1980(2000)01-0113-03

BP 网络是一种具有学习功能的人工神经网络,从结构上讲,三层 BP 网络是一个典型的 BP 网络,它被分为输入层 LA、隐含层 LB 和输出层 LC.其结构如图 1 所示.由图 1 可知,同层结点间没有任何联系,每一层结点的输出只影响下一层结点的输出.集中 LA 层含 m 个结点,对应的 BP 网络可以感知 m 个输入;LC 层含有 n 个结点,与 BP 网络的 n 中输出响应相对应;LB 层结点 u 根据需要来确定.每一个结点都具有单个神经元的结构,其单元特性通常是 sigmoid 型,因而具有高度非线性映射能力,当单元特性只具有线性时,那么网络只具有线性映射能力.令 LA 层结点 a_i 到 LB 层结点 b_r 间的权值为 W_{ir},LB 层结点 b_r 到 LC 层结点 c_j 间的权值为 $\boldsymbol{V}_{rj}$, T_r 为 LB 层结点的阈值,θ_j 为 LC层结点的阈值.针对 BP 网络,人们提出了误差反向传播(BP)算法.其基本思想是,沿着误差的负梯度方向不断修正网络中的权值与阈值,直到误差达到最小数值.其中最基本的 BP 算法有两种:一种是标准 BP 算法;另一种是“批处理”学习算法.BP 算法以下列准则函数作为目标函数:

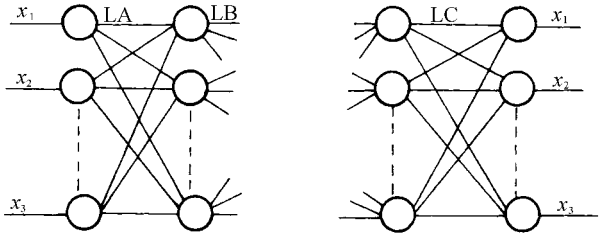

图 1　BP 网络

Fig.1　BP network

$$J(w) = \frac{1}{2}\sum_{j=1}^{n}(c_j - c_{js})^2 \qquad (1)$$

式中 c_{js}——输出结点 j 的希望输出;c_j——网络的实际输出.随机选取样本集中的样本一个一个地进行学习,要求沿着 $J(w)$ 的负梯度方向不断修正权值的数值,直到 $J(w)$ 达到最小值[1,2].“批处理”学习算法则是将所有 p 个样本学习完,以下面的准则函数作为目标函数:

$$J(w) = \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{n}(c_j^k - c_{js}^k)^2 \qquad (2)$$

即将样本集作为一个整体进行学习.要求沿着 $J(w)$ 的负梯度方向不断修正权值的数值,直到 $J(w)$ 达到最小值[2].其中标准 BP 算法学习过程中每次只利用到一个样本的信息,因而每次迭代的计算量较少,但是由于学习过程存在遗忘现象,收敛速度比较慢;“批处理”算法由于在整体上改进误差,因而学习过程迭代次数较少,但是当样本集较大时,存在每次迭代计算量较大而造成收敛速度慢的缺点[1].因而“批处理”学习算法主要适用于样本集较小的情况.本文在综合两种算法优点的基础上提出一种新的学习算法.它既能克服“批处理”学习时每次计算量较大的缺点,同时又能保持较少的迭代次数,从而在整体上提高了算法的效率.

收稿日期:1998-11-23

作者简介:赵启林(1972—),男,江苏淮安人,博士研究生,水工结构专业.

1 最大误差逆传播的(BE-BP)学习算法

BE-BP 算法的基本思想是:将所有样本学习后,并不以所有误差最小二乘累加作为目标函数,而选取误差最大的一个样本作为学习样本对权值进行修正,修正后再以修正后的权值作为起点对所有样本求得各自误差,选取最大误差的样本作为学习样本,直到达到误差要求.这样,每一次学习时都是用一个样本进行学习,避免了"批处理"多样本学习时每次计算量较大的缺点,同时选取误差最大的样本作为学习样本也在一定程度上保证误差沿整体减小的方向修正,避免了学习过程会产生像标准 BP 算法那样的振荡现象.具体算法如下:

a. 给 $W_{ir}, V_{rj}, T_r, \theta_j$ 随机赋一个较小的数值.

b. 对所有的模式对 $(A^{(k)}, C^{(k)})(k = 1, \cdots, p)$,进行下列操作:

依次正向计算输出数值:

$$b_r^k = f\left[\sum_{i=1}^{m} W_{ij}\alpha_i + T_r\right] \qquad r = 1, \cdots u;\ k = 1, \cdots, p \tag{3}$$

$$c_j^k = f\left[\sum_{r=1}^{u} V_{rj} b_r^k + \theta_j\right] \qquad j = 1, \cdots, n;\ k = 1, \cdots, p \tag{4}$$

对每一个 k 计算 $\qquad J(w) = \frac{1}{2}\sum_{j=1}^{n}(c_{js}^k - c_j^k)^2$

对所有的 k 比较 $J(w)$ 的大小,取 $J(w)$ 最大的那个样本对 $(A^{(m)}, C^{(m)})$ 作为学习样本.

c. 计算 LC 层结点输出与期望输出的误差,令

$$d_j = c_j^m(1 - c_j^m)(c_{js}^m - c_j^m) \tag{5}$$

d. 计算 LB 层结点反向分配误差,令

$$e_r = b_r^m(1 - b_r^m)\left[\sum_{j=1}^{n} V_{rj} d_j\right] \tag{6}$$

e. 按下式调整 V_{rj}, θ_j 的值:

$$V_{rj} = V_{rj} + \alpha b_r^m d_j \qquad \theta_j = \theta_j + \alpha d_j \qquad 0 < \alpha < 1 \tag{7}$$

f. 按下式调整 W_{ir}, T_r 的值:

$$W_{ir} = W_{ir} + \beta\alpha_i^m e_r \qquad T_r = T_r + \beta e_r \qquad 0 < \beta < 1 \tag{8}$$

g. 重复步骤 b ~ f,直到 $J(w)$ 变得足够小.

这种算法在 a ~ b 步实际上应用了"批处理"学习算法的思想,而从 c 步就转向标准 BP 算法.这样既从整体上对误差进行了修正,同时在主要计算步骤上采用了标准 BP 算法,减少计算量.标准 BP 算法,"批处理"算法和最大误差学习算法主要计算量对比如表 1 所示.

表 1 3 种算法主要计算量的对比

Table 1 Comparison of computation load of three algorithms

步骤	标准 BP 算法	"批处理"算法	最大误差学习算法
2	$mu + un$	$(mu + un)p$	$(mu + un)p$
2	2	$2p$	$2p$
4	n	np	np
5	$2nu + n$	$(2nu + n)p$	$2nu + n$
6	$2mu + u$	$(2mu + n)p$	$2mu + u$

2 计算机仿真结果

为了说明本算法的优点,本文对两个算例进行了对比研究.算例 1 考虑到用于矩阵求逆的 SCNN 中 BP 学习算法[3],对单位矩阵 $\boldsymbol{I}$ 进行迭代求解它的逆矩阵 $\boldsymbol{I}^{-1}$.我们选取修正系数 $\eta = 0.5$,起始权值矩阵 $\boldsymbol{I}^{-1}(O) = \begin{bmatrix} 1 & -10 & -1000 \\ 0 & 1 & 0 \\ 1 & -10 & 1 \end{bmatrix}$ 运用了 3 种算法后,计算机仿真对比结果见图 2.算例 2 以文献[4]中给出的 10 个样本作为学习样本集合,选取 3-10-1 网络结构,相同的修正系数($\eta = 0.5$)和起始权值,得到 3 种算法对比结果如图 3 所示.

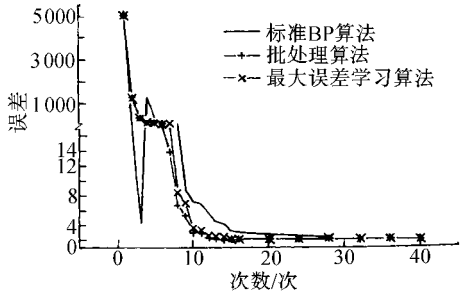

图 2　矩阵求逆迭代次数对比

Fig.2　Comparison of iteration times for matrix inversion

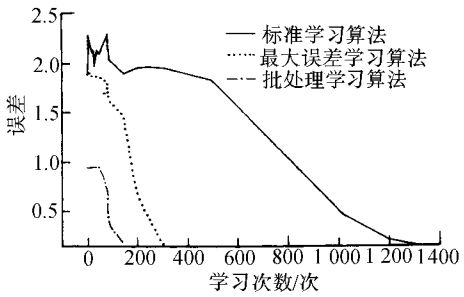

图 3　样本学习迭代次数对比

Fig.3　Comparison of iteration times for sample learning

3　结　　论

以上的分析与算例对比表明，BE-BP 算法一定程度上保留了“批处理”学习算法收敛较快的特点，而且算例中的 BE-BP 算法与“批处理”学习算法收敛速度基本处于同阶. 但由于进行修正时借鉴了标准 BP 算法的思想，用一个样本进行学习，因而可以克服“批处理”学习算法每一次学习时计算量较大的缺点，从而在整体上加快了收敛速度.

参考文献：

[1] 焦李成. 神经网络系统理论[M]. 西安：西安电子科技大学出版社，1990. 65 ~ 73.

[2] 杨行竣，郑君里. 人工神经网络[M]. 北京：北京高等教育出版社，1992. 23 ~ 51.

[3] 焦李成. 神经网络计算[M]. 西安：西安电子科技大学出版社，1990. 321 ~ 350.

[4] 史天运，王信义，张之进等. 神经网络与模糊故障诊断专家系统结合的应用研究[J]. 北京理工大学学报，1998，18(1)：81 ~ 86.

Largest Error Learning Algorithm of BP Network

ZHAO Qi-lin，ZHUO Jia-shou

(*College of Civil Engineering，Hohai Univ.，Nanjing* 210098，*China*)

Abstract：To synthesize the advantages of standard BP algorithm and“batch learning”BP algorithm，A new algorithm is put forword. It not only has the advantage of fewer iteration times like the“batch learning”algorithm，but also overcomes the disadvantages of much computational work and long learning time while learning big samples. The examples show that this algorithm has the same iteration times as the“batch learning”algorithm，but its computational work of every iteration is much less than that for the“batch learning”algorithm.

Key words：BP algorithm；BP neural network；batch learning