一种有效脱离平坦区的改进 BP 网络*)

刘 乐[1] 王洪国[1,2] 王 鑫[1] 王宝伟[3]
(山东师范大学管理与经济学院 济南 250014)[1] (山东省科学技术厅 济南 250011)[2]
(山东师范大学信息科学与工程学院 济南 250014)[3]

摘 要 本文提出了一种新型权值调整规则,当神经元输出值接近于 0 或 1 时,确保足够的权值调整幅度,解决了标准 BP(Back Propagation)网络在训练进入平坦区后难以摆脱,训练速度很慢的问题。并给出了利用单隐层 BP 网络逼近非线性函数的仿真实验,证实了新规则在训练速度上的优越性,得出新规则中控制因子的理想取值为 0.95。

关键词 平坦区,S 型函数,梯度下降

Modified BP Network in Breaking Away from Flat Areas Effectively

LIU Le[1] WANG Hong-guo[1,2] WANG Xin[1] WANG Bao-wei[3]
(Dept. of Management and Economy, Shandong Normal University, Jinan 250014, China)[1]
(Dept. of Science and Technology of Shandong Province, Jinan 250011, China)[2]
(Dept. of Information Science and Engineering, Shandong Normal University, Jinan 250014, China)[3]

Abstract This paper proposes a new rule to adjust the weight value, which ensures enough weight adjusted range when the output of neuron is close to 0 or 1. The new rule solves a difficult problem for standard BP network to break away from the flat area when it has entered, here its training speed is very slow. A experiment which uses a BPnetwork with single hidden layer to approximate a nonlinear function is presented in the paper. It verifies the superiority of the new rule in the training speed, and draws a conclusion that the ideal value of the control factor in the new rule is 0.95.

Keywords Flat area, S-type function, Gradient descent

1 引言

BP 网络是采用 BP 算法进行权值调整的多级前馈神经网络,具有很强的非线性映射能力、泛化能力和容错能力,被广泛应用于模式识别、函数逼近、图像处理、经济预测以及过程控制等领域。然而,标准 BP 网络在应用中暴露出不少严重的缺陷。如,训练次数多,收敛速度慢;易陷于局部极小而达不到全局最优;隐层层数和各隐层神经元个数的选取缺乏理论指导;BP 网络的学习和记忆具有不稳定性等等,大大限制了 BP 网络的推广应用。

针对上述缺陷,国内外许多学者已提出了很多有效的 BP 网络改进方案。例如,在权值调整公式中加入动量项[1],学习率自适应调节[2]等等。人们还先后利用变尺度法[3]、同伦算法[4]以及 LM 法[5]等优化算法对 BP 网络进行改进,并且尝试采用不同的有界可微函数(如混沌函数、小波函数等)作为激活函数,以提高 BP 网络的处理能力和收敛速度。此外,人们还将遗传算法[6]、模拟退火算法[7]与 BP 网络相结合,有效地克服了 BP 网络收敛速度慢,易于陷于局部极小的弱点。然而,众多 BP 网络改进方法中,有效地解决脱离平坦区问题的研究并不多。

本文在前人工作的基础上,提出了一种有效脱离平坦区的新型权值调整规则,目的在于提高 BP 网络的训练速度。

2 单隐层 BP 网络及其权值调整规则

BP 网络所采用的 BP 算法包括信号正向传播、误差反向传播两个过程,本质上是遵从误差梯度下降原则进行权值调整的学习算法,整个学习过程是一个非线性寻优过程。

2.1 单隐层 BP 网络

BP 网络是迄今应用最广泛的神经网络,其中尤其以单隐层 BP 网络应用最为普遍。由 Kolmogorov 定理可知,对于任意单隐层的 BP 网络,只要隐层神经元足够多,就能逼近任意非线性函数。

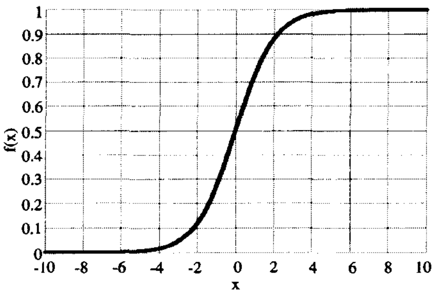

图 1 单极性 S 型函数

由于单隐层 BP 网络的重要地位以及单极性 S 型函数的特性,在 BP 网络中,神经元的激活函数大多采用典型的单极

*)基金项目:山东省自然科学基金(Q2006G03)。**刘 乐** 硕士研究生,主要研究方向为人工神经网络、数据挖掘;**王洪国** 教授,博士后,主要研究方向为组合优化算法、数据挖掘、电子政务等;**王 鑫** 硕士研究生,主要研究方向:数据挖掘、知识发现;**王宝伟** 硕士研究生,主要研究方向为遗传算法、人工神经网络。

性S型函数——Sigmoid函数，即：

$$f(x)=\frac{1}{1+e^{-x}} \tag{1}$$

如图1所示。

该函数具有以下特性：处处连续可导；单调递增；值域为(0,1)；

$$f'(x)=f(x)(1-f(x)) \tag{2}$$

单隐层BP网络包括三层：输入层、隐层和输出层，如图2所示。

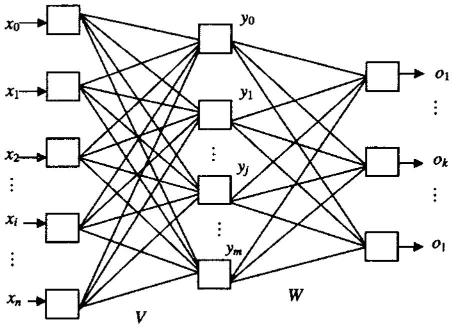

图2 单隐层BP网络

在图2所示的BP网络中，输入向量为$X=(x_1,x_2,\cdots,x_i,\cdots,x_n)^T$，$x_0=-1$是为隐层神经元引入阈值而设置的；隐层输出向量为$Y=(y_1,y_2,\cdots,y_j,\cdots,y_m)^T$，$y_0=-1$是为输出层神经元引入阈值而设置的；输出层输出向量为$O=(o_1,o_2,\cdots,o_k,\cdots,o_l)^T$；期望输出向量为$d=(d_1,d_2,\cdots,d_k,\cdots,d_l)^T$。输入层到隐层之间的权值矩阵用$V$表示，$V=(V_1,V_2,\cdots,V_j,\cdots,V_m)$，其中列向量$V_j$为隐层第$j$个神经元对应的权向量；隐层到输出层之间的权值矩阵用W表示，$W=(W_1,W_2,\cdots,W_k,\cdots,W_l)$，其中列向量$W_k$为输出层第$k$个神经元对应的权向量。除此之外，各层信号之间还存在以下数学关系：

对于输出层，有

$$o_k=f(net_k) \tag{3}$$

$$net_k=\sum_{j=0}^{m}w_{jk}y_j \quad k=1,2,\cdots,l \tag{4}$$

对于隐层，有

$$y_j=f(net_j) \tag{5}$$

$$net_j=\sum_{i=0}^{n}v_{ij}x_i \quad j=1,2,\cdots,m \tag{6}$$

2.2 单隐层BP网络的权值调整规则

我们知道，BP算法是按照误差梯度下降的原则调整权值的。依据文献[8]的理论，按照图2所示及其符号表达和相应数学关系，可以推导出神经元激活函数为Sigmoid函数的单隐层BP网络的权值调整规则为：

$$\begin{cases}\Delta w_{jk}=\eta\cdot(d_k-o_k)\cdot o_k\cdot(1-o_k)\cdot y_j\\ \Delta w_{ij}=\eta\cdot\sum_{k=1}^{l}(w_{jk}\cdot\delta_k)\cdot o_j\cdot(1-o_j)\cdot x_i\end{cases} \tag{7}$$

3 标准BP网络难以脱离平坦区

3.1 误差曲面上的平坦区

BP网络模型的误差度量一般为：

$$E=\frac{1}{2}\sum_{q=1}^{P}\sum_{k=1}^{l}(d_k^q-o_k^q)^2 \tag{8}$$

其中，P代表训练样本的总对数，q代表第q对训练样本，o_k^q和d_k^q分别表示对于第q对训练样本，输出层第k个神经元的实际输出值和理想输出值。经过仔细分析，BP网络的误差函数实际上是以各层权值、各神经元阈值和训练样本集对为自变量的非线性函数。一旦给定训练样本集，即可将误差函数想象为在多维空间中的一个形状复杂的曲面(误差曲面)，该曲面上每个点的“高度”对应于一个误差值，每个点的其他坐标值对应于各层权值、各神经元阈值。假设各层权值、各神经元阈值的总数目为M，则上面提到的多维空间实质上是$M+1$维空间，通常被称作M维权空间。为了直观描述误差曲面在权空间的起伏变化，简单起见，图3给出了在二维权空间中误差曲面的分布情况。

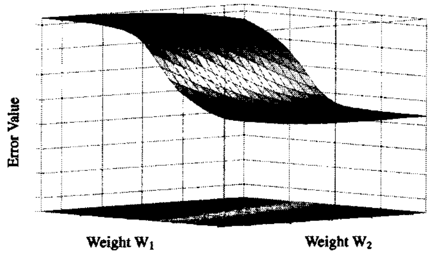

图3 二维权空间上的误差曲面

从图3中不难发现，在误差曲面上有些局部区域较为平坦，我们称其为平坦区。平坦区的数学特性在于，误差的梯度变化很小，如果权值调整量达不到一定程度，误差值难以显著下降。可见，一旦BP网络的训练过程陷入平坦区，而又不能及时有效地脱离它的话，必然会耗费大量的训练时间，这便是导致标准的BP网络模型训练速度慢的关键原因，同时也是努力提高标准BP网络收敛速度的重要突破口。

3.2 标准BP网络难以脱离平坦区的原因

在标准的BP网络模型中，存在这样一种异常现象：对于训练过程陷入平坦区的BP网络，即使理想输出与实际输出相差很大，权值的调整幅度也依然很小。欲探讨这一异常现象背后的原因，需要首先从权值调整规则入手。

我们以神经元激活函数为Sigmoid函数的单隐层BP网络为例，它的权值调整规则在1.2节(5)式已经给出。观察(5)式，不难发现，当o_k接近于0或1时，无论理想输出与实际输出相差多么大，权值w_{jk}的调整幅度都会大打折扣；当o_j接近于0或1时，无论$\sum_{k=1}^{l}(w_{jk}\cdot\delta_k)$的值多么大，都会严重影响它对增加权值$w_{ij}$调整幅度所起的作用。

可见，当隐层、输出层神经元的输出值接近于0或1时，训练进入平坦区的BP网络不得不花足够长的调整时间，以达到某个谷点。并且神经元的输出值接近于0或1可能性极大，原因在于Sigmoid函数具有饱和特性，即当净输入net的绝对值

$$|net|=|\sum_i(w_{ij}\cdot o_i)|>3$$

时，$f(net)$将处于接近于0或1的饱和区内，如图1所示。鉴于这种情况，我们可以在标准权值调整规则的基础上，人为加入一些控制条件和措施。

4 改进方案及其仿真实验

4.1 新的权值调整规则

基于以上分析，我们针对神经元激活函数为Sigmoid函数的单隐层BP网络提出一种新的权值调整规则，旨在提高它的收敛速度。

在新规则中，首先引入一个控制因子f，其中$f\in$

(0.5,1)。根据图 2,(5)式,新规则以数学公式可表达为:

$$\begin{cases}\Delta w_{jk}=\eta\cdot(d_k-o_k)\cdot o_k\cdot(1-o_k)\cdot y_j & o_k\in[1-f,f]\\ \Delta w_{jk}=\eta\cdot(d_k-o_k)\cdot f\cdot(1-f)\cdot y_j & o_k\notin[1-f,f]\\ \Delta w_{ij}=\eta\cdot\sum_{k=1}^{l}(w_{jk}\cdot\delta_k)\cdot o_j\cdot(1-o_j)\cdot x_i & o_j\in[1-f,f]\\ \Delta w_{ij}=\eta\cdot\sum_{k=1}^{l}(w_{jk}\cdot\delta_k)\cdot f\cdot(1-f)\cdot x_i & o_j\notin[1-f,f]\end{cases}\tag{9}$$

这样当隐层、输出层神经元的输出值接近于 0 或 1 时,依然保证拥有足够的权值调整幅度。从数学上看,新规则将 Sigmoid 函数的导函数进行了分段考虑,即:

$$f'(x)=\begin{cases}f(x)\cdot(1-f(x)) & f(x)\in[1-f,f]\\ f\cdot(1-f) & f(x)\notin[1-f,f]\end{cases}\tag{10}$$

可见,新规则在理论上大大减轻了隐层、输出层神经元输出值接近于 0 或 1 对权值调整幅度的不利影响。

可以看出,控制因子 f 的取值是新规则的关键问题。如果 f 取值过小,则会影响权值调整方向的正确性,限制梯度下降原则的优越性发挥;如果 f 取值过大,则新规则与标准规则十分接近,在性能上也相差不大。因此,并不是任取 $f\in(0.5,1)$,新规则在训练速度上都优越于标准规则。

下面将通过实验来验证新规则在收敛速度上的优越性,以及估计出控制因子的理想取值。

4.2 仿真实验

仿真实验依据标准 BP 网络的权值调整规则(5)式和本文提出的新规则(7)式,在 Visual C++ 6.0 环境下进行编程,以实现逼近一个非线性函数的训练过程。具体说来,本实验采用拓扑结构为 1-5-1 的 BP 网络来逼近一个非线性函数:

$$g(x)=x^2-2x+3 \quad x\in[0,2]\tag{11}$$

样本集中包含 20 对训练样本,整个训练过程,均采用单隐层的 BP 网络,学习率 $\eta=0.95$。通过与标准 BP 网络相关指标的对比,验证新的权值调整规则在训练速度上的优越性是本实验的基本思路。

在实验之前,需要对样本数据进行归一化,即通过变换处理将网络的输入、输出数据限制在[0,1]内。归一化常用以下变换式:

$$\bar{x}_i=\frac{x_i-x_{\min}}{x_{\max}-x_{\min}}\tag{12}$$

其中,x_i 代表输入、输出数据,$x_{\min}$ 和 $x_{\max}$ 分别代表数据变化范围的最小值和最大值。经过计算,该实验的归一化变换式为:

$$\bar{x}_i=\frac{x_i-0}{3-0}\tag{13}$$

通过实验可以得到:当误差终止条件是 $E=0.001$ 时,标准 BP 网络完成 100 次训练的平均迭代次数为 13130.39,迭代次数超过 18000 次的百分比为 11%,不足 8000 次的百分比为 0%;当误差终止条件是 $E=0.0005$ 时,标准 BP 网络完成 100 次训练的平均迭代次数为 40162.40,迭代次数超过 35000 次的百分比为 56%,不足 20000 次的百分比为 0%。

表 1 给出了在误差终止条件 $E=0.001$ 下,控制因子取不同值时,采用新型权值调整规则的 BP 网络完成 100 次训练的平均迭代次数,以及迭代次数超过 18000 次和不足 8000 次的百分比。

表 1 误差终止条件为 $E=0.001$ 时的仿真结果

控制因子	平均迭代次数	>18000 次百分比	<8000 次百分比
0.99	13498.79	12%	0%
0.98	11919.07	7%	0%
0.97	10182.01	0%	25%
0.96	9696.22	1%	29%
0.95	9565.48	0%	35%
0.94	9767.47	1%	25%
0.93	10813.13	1%	22%
0.92	12228.88	9%	9%
0.91	12780.57	6%	13%
0.90	13636.26	11%	3%
0.88	14563.37	17%	2%
0.85	16648.68	28%	0%

表 2 给出了在误差终止条件 $E=0.0005$ 下,控制因子取不同值时,采用新型权值调整规则的 BP 网络完成 100 次训练的平均迭代次数,以及迭代次数超过 35000 次和不足 20000 次的百分比。

表 2 误差终止条件为 $E=0.0005$ 时的仿真结果

控制因子	平均迭代次数	>35000 次百分比	<20000 次百分比
0.99	38174.66	44%	0%
0.98	32858.49	37%	0%
0.97	27109.07	13%	11%
0.96	25872.06	8%	21%
0.95	24828.47	6%	27%
0.94	28668.52	26%	15%
0.93	29279.50	23%	7%
0.92	30035.67	24%	1%
0.91	30642.17	18%	10%
0.90	32148.06	28%	10%
0.88	37290.25	40%	6%
0.85	43546.88	73%	0%

实验结果表明,当控制因子的取值在 0.91~0.98 范围内时,改进的 BP 网络在训练速度上明显优越于标准 BP 网络。并且,当控制因子的取值接近 0.95 时,改进 BP 网络的训练效率会比标准 BP 网络提高 30%左右。

结束语 标准 BP 网络在进入平坦区后难以尽快摆脱,是导致 BP 网络训练时间长,收敛速度慢的主要原因。本文提出的新权值调整规则旨在克服标准 BP 网络进入平坦区后难以有效脱离的弱点,引入了控制因子,并通过实验证实了新规则能够有效地提高收敛速度。然而,新规则并没有有效地克服 BP 网络易于陷于局部极小的弱点。因此,进一步提出全方位改进标准 BP 网络的方案将是未来研究的重点。

参 考 文 献

[1] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagation errors. Nature, 1986, 323:533-536

[2] Jacob R A. Increased rate of convergence through learning rate adaptation. Neural Networks, 1988(1):295-307

[3] 王正欧.一类有效的变尺度分层 BP 算法.天津大学学报,1996,29(3):364-369

[4] 高小榕,杨福生.采用同伦 BP 算法进行前向网络的训练.电子学报,1994,19(9):688-694

[5] Hagan M T, Menhaj M. Training feedforward networks with the marquardt algorithm. IEEE Tran on neural networks, 1994, 5(6):989-993

[6] 李建珍.基于遗传算法的人工神经网络学习算法.西北师范大学学报(自然科学版),2002,38(2):33-37

[7] Kirkpatrick S, Gelall C D, Vecch T M P. Optimization by simulated annealing. science, 1983, 220:671-680

[8] 韩力群.人工神经网络教程.第一版.北京:北京邮电大学出版社,2006

[9] 蒋宗礼.人工神经网络导论.第一版.北京:高等教育出版社,2001