

改进的快速BP算法

王彦春 余钦范 段云卿

摘 要 BP网络现在得到了广泛的应用。然而,对于学习样本较多,输入输出映射关系复杂的情况,学习速度较慢。通过分析网络误差曲面特征,提出了快速BP算法,它可以大幅度地提高学习速度。

关键词 误差曲面;BP网络;学习率

THE IMPROVED RAPID BP ALGORITHM

Wang Yanchun, Yu Qinfan, Duan Yunqing

(China University of Geosciences, Beijing 100083)

Abstract BP network has been extensively applied. Nevertheless, under the condition of numerous learning samples and complicated input-output mapping relationship, the learning speed is rather low. Based on analysing characteristics of error curved surface of the network, the authors have advanced the rapid BP algorithm, which can greatly raise the learning speed.

Key words error curved surface; BP network; learning rate

BP算法是目前应用最广泛且较成功的一种算法,它利用梯度搜索技术(gradient search technique)使代价函数(cost function)最小化。代价函数等于期望输出与实际输出的误差平方和。虽然BP算法能有效地解决很多问题,但对于输入输出的映射关系比较复杂、已知样本较多的情况,其学习速率慢。通过分析误差曲面的一般特征,提出了快速BP算法。对一些例子进行了试算,学习速度有大幅度的提高,取得了较好的成果。

1 BP算法及误差曲面的特征

BP算法是用来训练人工神经网络的。主要用BP算法训练前馈型人工神经网络,它由输入层、一个或多个隐蔽层(或称中间层)组成,每个节点只与邻层节点相连接,同一层的节点不相连。如果隐层节点足够多,一个三层人工神经网络可产生任意复杂的判别界面。

前馈型神经网络常用的激活函数是S型函数。在勘探地球物理中的模式识别问题大都采用S型函数。

这里先对下面所用符号作以说明。 W_{ijk} 是k层第i个节点与k+1层的第j个节点之间的连接权, I_{jk} 是第k层第j个节点的输入, O_{ik} 是第k层第i个节点的输出, $L_{(k)}$ 是第k层的节点数,在输出层以前各层分别多加一个节点,并使这些附加节点在每次迭代中输出均为1。这样就可以从数学上把阈值与其它项写成一个统一的公式,阈值的修改权值一样。

节点的输入、输出关系由激活函数给出

$$O_{jk} = f\left[\sum_{i=1}^{L_{(k-1)}} W_{ij(k-1)} O_{i(k-1)} - \theta_{jk}\right]$$

(1)

对于给定的已知样本 (X_p, Y_p) ($p=1, 2, \dots, P$)，也就是网络输入层及输出层的值均为已知，要寻找一个映射关系，它能够满足已知样本的输入、输出的对应关系。具体对于前馈型神经网络，就要确定节点间的连接权值，使网络满足已知的对应关系。

节点间权值的确定用BP算法的迭代方法。设已迭代到第 t 步，此时网络的权值向量为 $W_{ijk}(t)$ ，新的权值增量为

$$\Delta W_{ij(k-1)}(t+1) = \begin{cases} \eta O_{kj}(1 - O_{kj})(y_{pj} - O_{kj}) O_{i(k-1)} & (k \text{ 为输出层}) \\ \eta O_{kj}(1 - O_{kj}) \left(\sum_l \delta_{l(k-1)} - W_{il(k-1)} O_{i(k-1)} \right) & (k \text{ 为输出层}) \end{cases}$$

(2)

式中， η 为学习率。迭代到第 $t+1$ 步时，网络的权值为

$$W_{ijk}(t+1) = W_{ijk}(t) + \Delta W_{ijk}(t+1)$$

(3)

反复修改权值向量，直到在一定误差条件下，网络满足给定的已知样本的映射关系。

对于一组给定的训练数据集，用BP算法，对网络进行训练，网络可以一组特定的权值 W 实现一定精度的输入、输出映射。训练的目的在于希望产生最小误差的权值，可以以权值为变量描述误差的变化情况，用误差曲面反映出来，误差曲线一般为超曲面。现在用的BP算法速度较慢，通过考察误差曲面的特征来解释BP算法速度慢的原因。

根据前人的经验和我们的研究，一般情况下，误差曲面都比较复杂，有大量平缓和变化剧烈的区域，当训练样本较少时，可以观察到误差曲面呈台阶状，一般一个台阶对应一个训练样本，当训练样本较多时，台阶状便不明显。

现在常用的BP算法，其权值的调整量是采用误差梯度与学习率 η 的乘积， η 采用大于零的较小常数。在误差曲面变化非常平缓的地方，误差曲面的梯度便会很小，如果学习率 η 也较小，会导致权值增量 ΔW_{ijk} 也很小，其结果是迭代很多次仍不能摆脱这个平缓带。在误差曲面变化剧烈的地方，它的梯度较大，这时要求学习率很小，否则 ΔW_{ijk} 会很大，容易发生振荡现象而难以收敛。因此，在误差曲面的平缓区域，学习率要大一些，在误差曲面变化剧烈的区域，要求学习率很小。如果所用的学习率为常数，只有取一个较小的常数，这样，必然导致较慢的学习速度。

2 快速BP算法

根据误差曲面的特征，一些区域其学习率必须很小，而另外一些区域学习率相对较大。采用自

适应调节学习速率的误差反传算法对神经网络进行训练，必须在小学习速率的区域尽量采用小的学习速率，在学习速率允许相对较大的区域，在不导致振荡的前提下，尽量采用大的学习速率，具体实施方法如下。

采用“批学习”的方式，在每个迭代步上考察所有样本的误差平方和较上次迭代是否有所下降，再对学习率加以调整。

在每个迭代步上，考察期望输出和实际输出的误差平方和。设有P个已知样本矢量 $X_k(k=1,2,\dots,P)$ ，其相应的期望输出为具有m个分量的矢量

$$Y_k=(y_{k1},y_{k2},\dots,y_{km}) \quad (k=1,2,\dots,P)$$

设在第t次迭代的实际输出为

$$\hat{Y}(t) = (\hat{y}_{k1}(t), \hat{y}_{k2}(t), \dots, \hat{y}_{km}(t)) \quad (k = 1, 2, \dots, P)$$

则第t次迭代后误差平方和为

$$E(t) = \sum_{k=1}^P \sum_{j=1}^m (y_{kj} - \hat{y}_{kj}(t))^2 \quad (4)$$

迭代开始时，用一个较小的学习率因子，每迭代一步考察总误差是否下降。如果 $E(t) < E(t-1)$ ，则

$$\eta(t+1) = 1.05 \eta(t) \quad (5)$$

$$\Delta W_{ijk}(t+1) = \eta(t+1) \sum_{p=1}^P \delta_{pj(k+1)} O_{pi k} \quad (6)$$

$$W_{ijk}(t+1) = W_{ijk}(t) + \Delta W_{ijk}(t+1) \quad (7)$$

式中， $W_{ijk}(t)$ 为第t次迭代第k层的第i个节点到第k+1层第j个节点的连接权； $\Delta W_{ijk}(t+1)$ 为该连接权值在第t+1次迭代时的修正量； $\eta(t)$ 为第t次迭代所用的学习率因子； $O_{pi k}$ 为输入第p个样本时，第k

层第*i*个节点的输出，并有

$$\delta_{pj(k+1)} = - \frac{\partial E}{\partial I_{pj(k+1)}}$$

其中， $I_{pj(k+1)}$ 为输入第*p*个样本时，第*k+1*层第*j*个节点的输入。
如果 $E(t)>E(t-1)$ ，则

$$\eta(t+1)=0.6 \eta(t) \tag{8}$$

$$\Delta W_{ijk}(t+1)=0.3 [W_{ijk}(t)-W_{ijk}(t-1)] \tag{9}$$

$$W_{ijk}(t+1)=W_{ijk}(t-1)+ \Delta W_{ijk}(t+1) \tag{10}$$

当处于误差曲面的某一平缓区域时，每次迭代总误差都有微小的下降，学习率因子会呈指数上升，成为一个较大的数。当到误差曲面变化急剧的地方，如果发现迭代中误差反而上升，就迅速减小学习率因子，并用上一次迭代后的权作为下次更新权的基础。这种自适应调节权值的方法可以尽量采用大的学习速率，又不使学习过程发生振荡，采用这种学习方法，学习速度比常数学习速率的学习速度有明显地提高。从实际例证来看，速度一般会提高一个数量级。

现给出解决Parity - 3问题(表1)的例子，训练数据集如表1。采用单隐层网络，隐层节点数为3个。若误差最大容许限为0.001，用快速BP算法，仅迭代870次便达到误差要求，采用常数学习速率的算法，当学习速率为0.8时，迭代50000次时，最大误差仍为0.007。

表1 Parity - 3问题

训练样本序号	输 入	期望输出
1	000	1
2	001	0
3	010	0
4	011	1
5	100	0
6	101	1
7	110	1
8	111	0

图1为误差平方和随迭代次数变化的曲线，

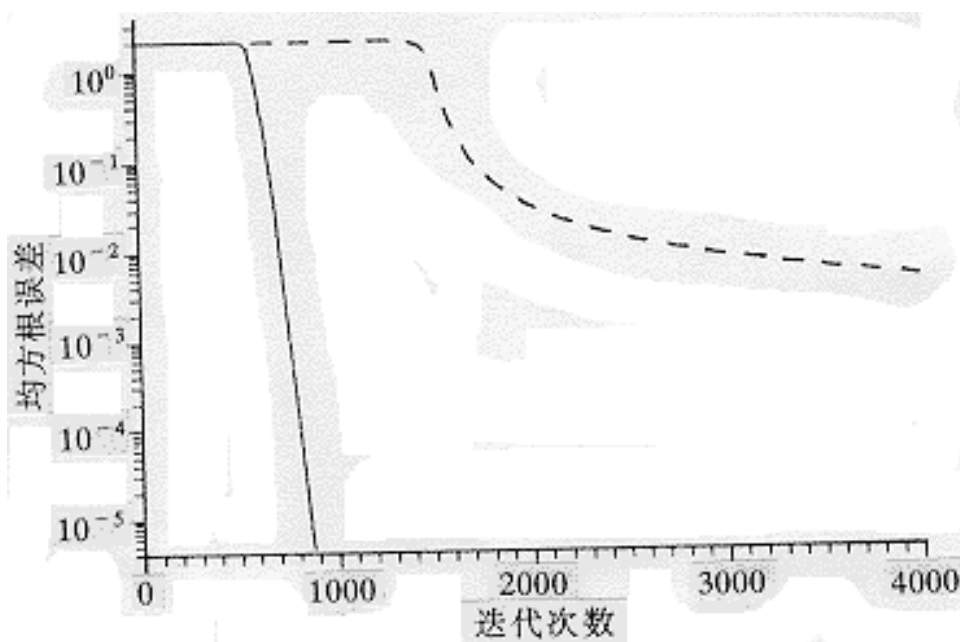


图1 误差平方和随迭代次数变化曲线
实线为新算法误差曲线；虚线为BP算法误差曲线

图2为学习速率随迭代次数变化的曲线，从图上可见，有时学习率可以很大。

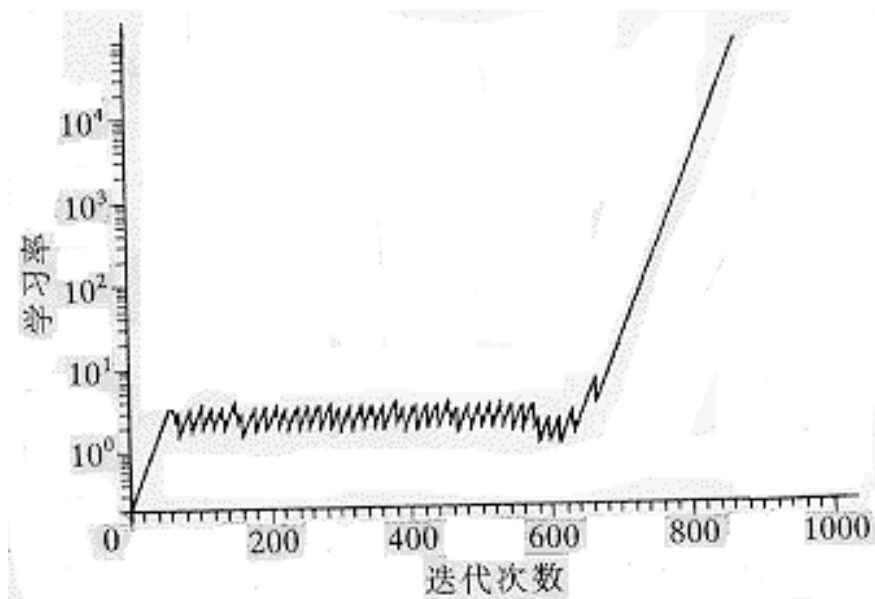


图2 学习率因子随迭代次数变化曲线

对文献〔5〕中的一组实验数据进行试算，表2为实验集合，文献作者对该数据集合分别用同伦算法和BP算法进行训练，结果如表3，想以此说明同伦算法比BP算法优越得多。用快速BP算法对该组数据进行训练，采用3层网络，隐层节点数选为2个，即为2-2-1网络，输出节点的最大容许误差选为0.01，仅用298次训练即已满足误差要求。

表2 信真实验数据集合

x_1	-1	0	1	0	1	0	1	-1	-1
x_2	1	0	-1	1	1	-1	0	0	-1
y	0	0	0	1	1	1	1	1	1

表3 模型2-2-1

输入	BP算法	同伦算法	期望输出
(0 , 0)	0.00365	0.00237	0
(1 , 0)	0.99699	0.98089	1
(1 , 1)	0.99856	0.98607	1
(0 , 1)	0.99697	0.95020	1
(1 , -1)	0.00363	0.00882	0
(0 , -1)	0.99670	0.98313	1
(-1 , -1)	0.99713	0.98552	1
(-1 , 0)	0.99675	0.98501	1
(-1 , 1)	0.00368	0.00312	0
误 差	9.00e?06	4.28e?06	
迭代次数	10535	2000	

图3为实际输出与期望输出的误差平方和随迭代次数变化曲线，

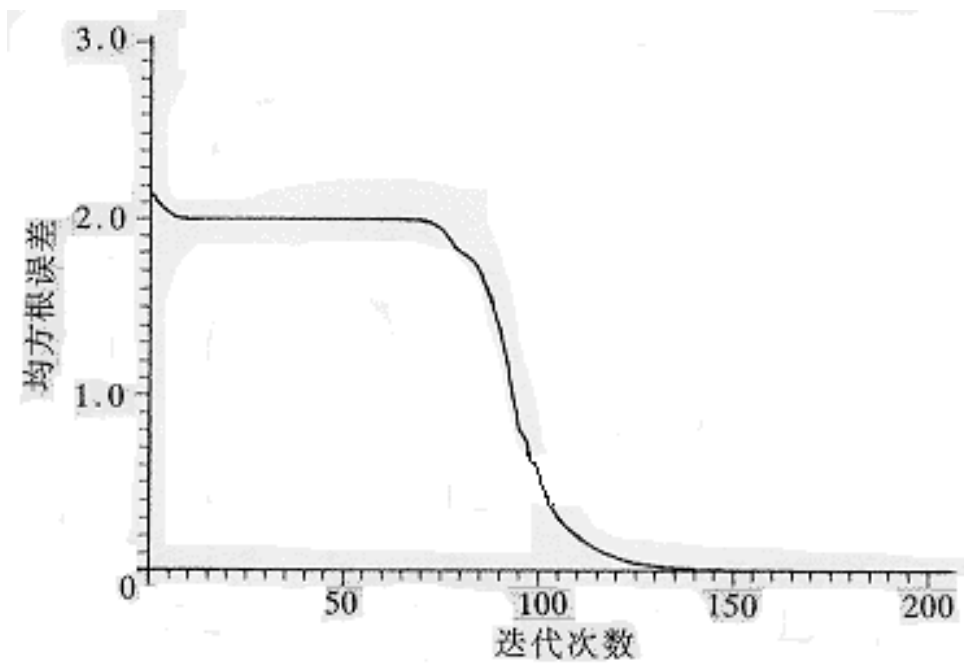


图3 误差平方和随迭代次数的变化曲线

图4为学习因子随迭代次数变化曲线。从此例可以看出快速BP算法较传统BP算法及同伦算法有很大的改进。

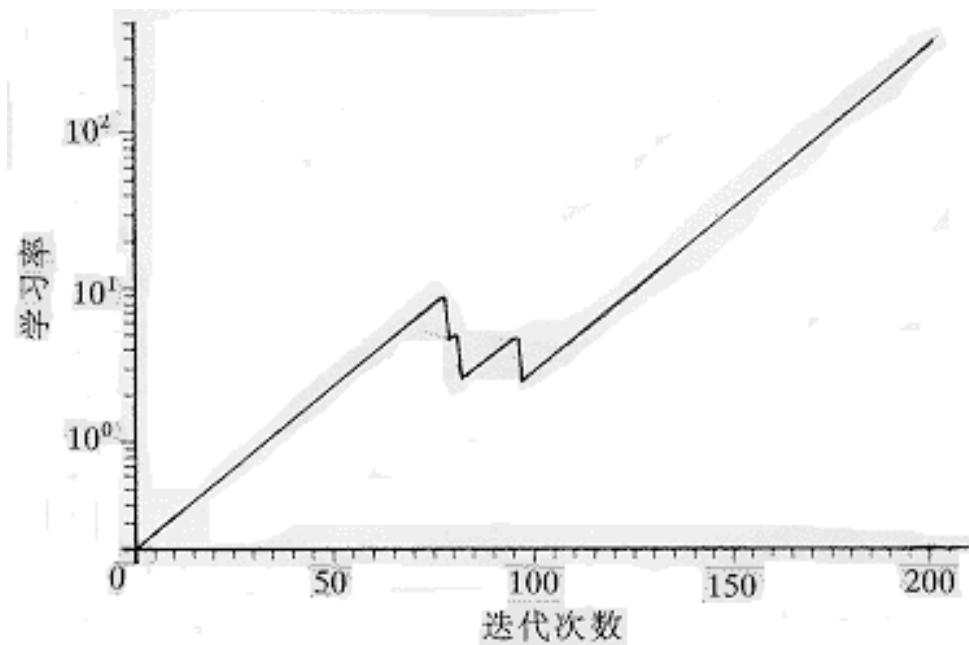


图4 学习率因子随迭代次数的变化曲线

在一些文献中，为了提高训练速度，在训练的初期，或者说实际输出与期望输出的误差相对较大时，采用较大的学习速率，在误差相对较小时采用相对较小的学习速率，是很难奏效的。这是因为权值修正量与误差有关，又与总误差梯度有关。由于现在大都采用S型函数，当误差很小时，误差的梯度也会很小；当误差大时，其误差梯度较大。所以，当误差较大时，采用较大的学习速率容易发生振荡现象，而当误差较小时，采用较小的学习速率，学习过程会非常缓慢。

3 结束语

对传统BP网络进行了深入研究，分析了误差曲面的特征，提出了自适应调节学习速率的误差反传算法，使学习速度有大幅度的提高，并对一些例子进行了试算，得出了满意的结果。

作者简介：王彦春，男，1959年生。1994年毕业于成都理工学院石油物探专业，获博士学位，1998年从中国地质大学(北京)博士后流动出站。现在中国地质大学(北京)物探系从事地震勘探的教学和科研工作，主要研究方向为折射波静校正、地震资料去噪、波阻抗反演、储层描述，发表论文20多篇。
作者单位：中国地质大学，北京 100083

4 参考文献

- [1]Moon W M,et al.Reservoir characterization using feedforward neuralnetworks.63rd Ann,Int. SEG Mtg.Expanded Abstracts,1993.258 ~ 263
- [2]Lippman R P.An introduction to computing with neural nets.IEEE ASSP Magazine,1987,4:4 ~ 22
- [3]Amari S I.Mathematical foundations of neurocomputing.Proceedings of the IEEE,1990,78(9)
- [4]McIormack M D.Neural computing in geophysics.In:The leading edge of exploration,1991.11 ~ 15
- [5]焦李成，保铮.同伦BP网络理论与算法.电子学报，1993，21(11)：1 ~ 6
- [6]包约翰.自适应模式识别与神经网络.北京：科学出版社，1992

1998年2月12日收稿，同年3月23日收修改稿。