

# 有隐层指导的多层前馈网络学习算法研究\*

朱江海<sup>1)</sup> 李梅<sup>2)</sup> 戚飞虎<sup>1)</sup>

(<sup>1)</sup>上海交通大学计算机科学与工程系, 上海, 200030;

(<sup>2)</sup>上海铁道大学计算中心, 上海, 200331)

TP15

**摘要** 采用基于知识的方法对隐层训练加以指导, 有效地改善了训练算法的效果, 并使隐层节点数可以明确地确定, 在增加新模式时, 训练不必完全重新开始, 使训练时间大为缩减.

**关键词** HDBP 算法, BP 算法, 神经网络.

隐层指导

## 引言

多层前馈神经网络已被广泛应用于各个领域, 如分类、函数逼近、模式识别、字符识别, 以及控制系统等等. 它的常用学习算法是 BP 算法, 但是, BP 算法有许多固有的缺点<sup>[1]</sup>, 如训练时需要给出隐层节点个数, 学习速度慢, 有可能收敛于局部极小点<sup>[2]</sup>, 训练有可能产生振荡等等. 文献[3]给出了对网络剪枝的方法; 文献[4]对 BP 网的隐层节点个数的确定方法进行了研究; 文献[5]对激励函数进行修改, 以改善 BP 算法的收敛稳定性; 文献[6]先构造一个决策树, 然后导出一个对应的多层网, 使训练过程较快.

本文提出了基于知识的有隐层指导的 HDBP (Hidden-layer Directed Backpropagation) 算法, 它能够解决隐层节点数的确定问题, 有效地提高算法的收敛速度和收敛稳定性, 不需要改变激励函数形式和目标函数形式, 也不需要复杂的方法事先构造一个决策树, 在应用中收到良好的效果.

## 1 有隐层指导多层前馈网络的 HDBP 学习算法

典型的 BP 算法收敛曲线如图 1 所示, 我们可以将其划分为 A、B、C 三个阶段, A 段为寻找搜索方向阶段, 在此阶段, 由于不断试探搜索方向, 当学习因子较大时, 容易产生振荡现象. 在确定搜索方向之后, 迭代过程在 B 段会有一个大的误差下降, 之后在 C 段, 误差值缓慢地下降. 在三个阶段中, 只有 B 段处于快速收敛状态, 而 A 段和 C 段收敛均很慢, 占用了绝大部分训练时间, 严重影响了整个学习过程的学习速度.

对于 C 段, 搜索方向已经确定, 收敛速度主要受到权值修改量的影响. 一些文献提到, 若学习因子过大, 当迭代过程接近误差函数曲面的谷底时, 迭代会在谷底两侧来回振荡. 但实验观察中则正好相反, 振荡往往发生在迭代刚开始的 A 段, 在迭代接近误差曲面谷底的 C 段中没有观察到振荡的现象. 实际上, 当迭代接近谷底时, 梯度模值已大大减小 (到达极小点时梯度模值为 0), 使得权值修改量  $\|\Delta W\|$  大大减小, 前进步长变得非常小, 发生振

\* 国家自然科学基金(69572026)资助项目  
稿件收到日期 1997-07-22, 修改稿收到日期 1998-03-05

荡的可能性已不存在,相反,迭代在搜索方向上得不到足够的前进,是造成C段误差下降缓慢的真正原因,要提高C段的学习速度,首先要将梯度模值归一化,只使用梯度方向,消除梯度模值减小带来的影响,然后,再采用优化方法调节步长<sup>[7-9]</sup>,使迭代在搜索方向上有适当的前进,我们在使用“基于知识的优化方法”<sup>[6]</sup>之后,C段的迭代与B段一样下降陡直,收敛非常迅速.

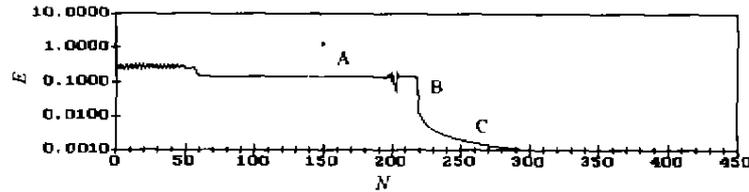


图1 BP算法收敛曲线

Fig.1 Convergence curve of BP algorithm

本文着重讨论迭代过程起始阶段A段的改进方法,也就是怎样在迭代开始的时候,尽快找到迭代计算的搜索方向.在这方面,冲量法及改进的冲量法<sup>[10-11]</sup>综合前面的搜索方向得到一个具有平均特性的搜索方向,文献<sup>[12]</sup>提出用隐层权值间的相互平滑以体现其代表的特征来达到确定搜索方向的目的.

在很多情况下,人们对领域知识已经有所总结,但传统的学习算法却无法使其方便地用于对神经网络训练的指导.常用的方法是改变神经元的激励函数,或者把领域知识加在目标函数中<sup>[13-14]</sup>,但这些方法对于不同的问题,要有不同的激励函数或目标函数,而且后者还要求加入的目标函数可导,并要显式给出其导数形式,使用很不方便.

研究BP算法的训练过程可以发现,BP算法的训练虽然有输出层教师信号的控制,但是隐层节点并没有明确的指导信号,训练后隐层节点的意义有其随意性.为此,本文以赋给隐层节点明确的意义为出发点,探讨解决上述BP算法缺点的方法.我们采用HDBP算法,使搜索方向从一开始就有基于已有领域先验知识的指导,迅速确定了搜索方向,有效减少了A段振荡.

### 1.1 HDBP算法

传统BP算法的训练过程如下.设输入 $P$ 个样本,则:

$$E = \sum_p E_p,$$

$$E_p = \sum_k (d_k - o_k)^2 / 2, p = 0, 1, \dots, P$$

$$W(n+1) = W(n) + \Delta W(n),$$

$$V(n+1) = V(n) + \Delta V(n),$$

$$\Delta W(n) = -\eta \frac{\partial E}{\partial W(n)},$$

$$\Delta V(n) = -\eta \frac{\partial E}{\partial V(n)};$$

其中: $\eta$ 为步长, $n=0, 1, 2, \dots$ 为迭代次数.

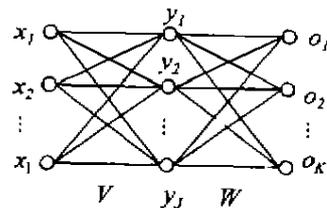


图2 多层前馈网络

Fig.2 Multilayer network

对于 3 层网络,第  $n$  次权值变化量为:

$$\Delta W_{kj} = \eta(d_k - O_k)f'_k(\text{net}_k)y_j, \quad \Delta V_n = \eta x_j f'(\text{net}_j) \sum_k (d_k - O_k) f'_k(\text{net}_k) W_{kj},$$

式中:  $\text{net}$  为神经元净输入.

HDBP 算法对每一个输入样本,除给出输出层教师指导信号  $d_i$  外,另外给出隐层节点输出教师指导信号  $d_{h_j}$ ,即在隐层的每个节点,给出一个教师信号.对于 3 层网络有隐层教师信号的 HDBP 算法训练过程如下:

$$\begin{aligned} E_{hd} &= \sum_j (d_{h_j} - y_j)^2 / 2, \\ y_j &= (1 - \eta_{hd})f(\text{net}_j) + \eta_{hd} \cdot d_{h_j}, \\ \Delta v_n &= -(1 - \eta_{hd})\eta \frac{\partial E}{\partial V_n} - \eta_{hd}\eta \frac{\partial E_{hd}}{\partial V_n}. \end{aligned}$$

则:

$$\Delta W_{kj} = \eta(d_k - O_k)f'_k(\text{net}_k)y_j,$$

$$\Delta v_n = \eta(1 - \eta_{hd})x_j f'(\text{net}_j) \sum_k (d_k - y_k) f'_k(\text{net}_k) W_{kj} + \eta \cdot \eta_{hd} x_j f'(\text{net}_j) (d_{h_j} - O_j),$$

式中:  $d_{h_j}$  为隐层教师信号;  $E_{hd}$  为隐层指导的误差函数,  $\eta_{hd}$  为控制因子,其值为  $0 \sim 1$ .

当  $\eta_{hd}$  为 0 时, HDBP 学习算法退化为标准的 BP 算法; 当  $\eta_{hd}$  为 1 时, 为两个单层网的级连; 当  $0 < \eta_{hd} < 1$  时, 可使隐节点的笔划特征在输出层误差的调整下更加适应大量各种各样的同一个字的不同写法的笔划, 更加具有弹性, 并有更好的适应性及抗噪性.

## 1.2 隐层指导信号的加入

HDBP 算法对每一个输入样本,除给出输出层教师指导信号  $d_i$  外,还给出隐层节点输出教师指导信号  $d_{h_j}$ . 此信号可通过对已有领域知识的描述分析得到. 对于 XOR 问题, 可以描述为“ $x_1 > x_2$  相同时为真, 相异时为假”, 即“ $x_1 > x_2$  或  $x_1 < x_2$  时为假”. 因此可以将隐节点 1 定义为“ $x_1 > x_2$ ”, 将隐节点 2 定义为“ $x_1 < x_2$ ”(见表 1). 需要注意的是, 在描述领域知识时, 隐层节点的意义应当能够用  $x_i$  的线性组合来表示. 对字符识别, 隐层节点表示的就是笔划特征. 训练时, 隐层也要提供教师信号. 输入层输入字符点阵, 经过训练, 在隐层得到笔划特征, 在输出层得到字符识别结果. 例如, 对于数字 ‘0’ - ‘9’, 通过分析, 我们可以指定一组笔划形状, 如表 2 中的  $d_{h_1} \sim d_{h_1}$ . 当  $d_{h_j}$  为 1 时, 表示隐层输出为 1; 为 0 时, 表示隐层输出为 0; 其余的表示隐层不加指导信号. 在这组笔画下, ‘0’ 可以由  $d_{h_1}$ 、 $d_{h_2}$  来表示, ‘3’ 可以由  $d_{h_1}$ 、 $d_{h_2}$ 、 $d_{h_3}$  来表示.

易混淆模式是造成识别率下降的主要原因之一, 可以增设隐层节点, 使其代表易混淆模式之间的区别特征, 如表 2 中的  $d_{h_{10}}$  代表了 ‘3’ 和 ‘8’ 之间的区别特征, 而  $d_{h_{11}}$  代表了 ‘5’ 和 ‘6’ 之间的区别特征. 这样, 经过训练之后, 将有效降低易混淆模式的错误率.

为了有效利用 HDBP 算法的优势, 训练样本中除了有一般的输入训练样本集之外, 还应当增加隐层教师模板样本集. 对于字符识别即为笔划字形模板. 对这些输入样本, 在输出层不规定其输出类别, 也不影响输出层的误差计算.

对 HDBP 网, 隐节点数将由特征数(如笔划形状数)而定, 因此解决了隐节点数确定问题. 对于特定的问题, 有时领域知识尚不足以完全描述问题, 此时可以增设  $1 \sim 2$  个冗余隐层

节点,由输出层指导信号的误差反向传播作用来修改其权值大小.

表 1 XOR 问题隐层指导信号  
Table 1 Hidden-layer directing signal of XOR problem

$x_1$	$x_2$	$d_{h1}(x_1 > x_2)$	$d_{h2}(x_1 < x_2)$	$\overline{D_{h3}(x_1 < x_2 \text{ XOR } x_1 > x_2)}$
0	0	0	0	0
1	1	0	0	0
0	1	0	1	1
1	0	1	0	1

表 2 数字 '0' - '9' 的隐层指导信号  
Table 2 Hidden-layer directing signal of Digits '0' - '9'

$\lambda$ (输入图案)	$d_{h1}$	$d_{h2}$	$d_{h3}$	$d_{h4}$	$d_{h5}$	$d_{h6}$	$d_{h7}$	$d_{h8}$	$d_{h9}$	$d_{h10}$	$d_{h11}$
'0'	F	F	0	0	0	0	0	0	0	0	0
'1'	0	0	0	0	0	0	0	0	0	0	0
'2'	1	1	1	1	0	0	0	0	0	0	0
'3'	1	1	1	0	1	0	0	1	0	0	0
'4'	0	1	1	0	1	0	1	1	0	0	0
'5'	0	1	1	0	0	1	1	1	0	0	0
'6'	0	1	1	0	0	1	1	1	0	0	F
'7'	0	1	1	0	0	1	1	1	0	0	F
'8'	1	1	1	0	0	1	1	1	0	F	F
'9'	1	1	1	0	0	1	1	1	0	F	0

## 2 实验结果分析

### 2.1 XOR 问题

以 XOR 问题为例,取三层前馈神经网络,输入层、隐层和输出层的神经元个数为 2、2、1,初始权值取  $(-0.1, 0.1)$  之间的随机值,激励函数为单极 Sigmoid 函数.

BP 算法和 HDBP 算法求解 XOR 问题误差收敛曲线如图 3 所示,其中 BP 算法  $\eta = 20.0$ ,HDBP 算法  $\eta = 30.0, \eta_{hd} = 0.8$ .从图 3 可以看出 BP 算法在  $\eta = 20$  时出现振荡,即迭代

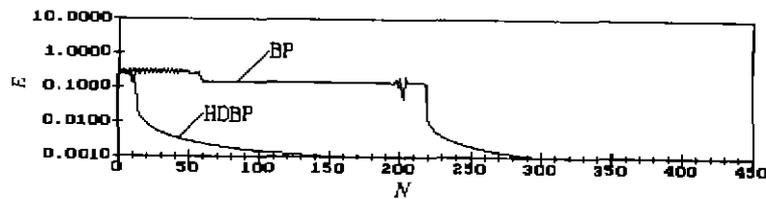


图 3 XOR 问题收敛曲线  
Fig. 3 Convergence curves of XOR problem

计算在谷底两边来回跳跃, 产生了振荡, 而 HDBP 算法有效地增加了收敛稳定性, 减少了振荡的发生。

表 3 为两种算法在不同步长下随机设置权值初值后的 10 组运行结果的平均值, 括号内为均方差值。在 A 段, 改进算法的收敛速度大大提高, 是 BP 算法的 7.4~17.9 倍。BP 算法的迭代次数均方差为 7.1~114.2, 而改进算法的仅为 0.5~5.6, 算法收敛稳定性大大提高。

表 3 XOR 问题  
Table 3 XOR problem

$\eta$		3.0	5.0	8.0	10.0	12.0	15.0	18.0	20.0	30.0	40.0
BP	$A(E=X-0.1)$	348.8 (42.2)	254.0 (114.2)	204.3 (25.5)	232.5 (52.2)	147.5 (29.1)	113.3 (7.1)	116.8 (31.4)	106.4 (61.3)	不收敛	不收敛
	$B(E=0.1-0.01)$	93.0	50.0	25.0	12.3	7.5	5.7	2.2	2.2		
	$C(E=0.01-0.001)$	795.7	470.8	285.7	212.2	172.3	151.8	94.8	86.2		
HDBP	$A(E=X-0.1)$	47.0 (5.0)	27.5 (5.0)	15.8 (3.3)	13.0 (1.6)	12.8 (1.5)	9.4 (0.5)	12.2 (2.0)	11.0 (2.3)	18.8 (4.1)	38.5 (5.6)
	$B(E=0.1-0.01)$	120.0	75.0	42.0	35.4	25.3	20.6	13.8	11.6	1.6	0.5
	$C(E=0.01-0.001)$	821.0	576.8	302.6	259.6	202.3	176.8	132.8	128.0	60.6	19.5

## 2.2 字符识别

以数字 '0'-'9' 为例, 取三层前馈网络, 输入层为 256 个神经元, 隐层如前所述取为 13 个(11 个有指导, 2 个无指导)神经元, 输出层为 10 个神经元。表 2 为笔划形状训练样本图。

表 4 给出了 BP 算法与 HDBP 算法在不同步长下随机设置权值初值后的 10 组运行结果的平均值, 括号内为均方差值, 其中  $\eta_{hd}=0.8$ 。

表 4 数字 '0'-'9'  
Table 4 Digits '0'-'9'

$\eta$		1.2	1.5	2.0	2.5	3.0	4.0	5.0	6.0	8.0	9.0	10.0
BP	$A(E=X-0.1)$	120.0 (2.0)	117.0 (1.0)	95.7 (9.5)	125.0 (14.0)	251.0 (44.0)	不收敛	不收敛	不收敛	不收敛	不收敛	不收敛
	$B(E=0.1-0.01)$	532.5	494.0	371.0	292.0	258.0						
HDBP	$A(E=X-0.1)$	82.0 (5.0)	76.0 (4.0)	55.5 (5.5)	46.0 (4.0)	54.5 (6.5)	42.5 (3.5)	37.0 (1.0)	50.5 (8.5)	42.0 (12.0)	112.7 (50.1)	103.0 (30.0)
	$B(E=0.1-0.01)$	446.5	392.0	274.5	240.5	232.5	157.5	110.5	107.5	58.5	17.3	12.0

从表 4 可知, 在 A 段改进算法的收敛速度大大提高了, 迭代次数受步长的影响明显得到改善。在很大的步长范围内, HDBP 都可以稳定地收敛。

图 4 给出了 13 个隐节点与 256 个输入层节点连接权的灰度图表示, 其中  $y_{11} \sim y_{13}$  隐节点有指导信号,  $y_{12}, y_{14}$  为两个冗余节点。从图 4 可见, BP 算法由于隐层节点意义的随意性,

笔划信息被瓜分后无意义地存放着,造成权值图杂乱无章,陡峭不平,这大大影响了识别稳定性和鲁棒性<sup>[12]</sup>;而 HDBP 权值图光滑平坦,使其识别鲁棒性得到了大大提高.另外其笔划意义明显,说明了领域先验知识在训练过程中起到了很好的指导作用.

与平滑法<sup>[12]</sup>相比较,HDBP 算法有更好的性质.平滑法是将一个隐层节点到输入层的相邻连接权平滑,使得权值三维曲面平滑,更具特征性.由于笔划具有连续线段性质,故平滑法可保持类似于笔划的特征,但它并没有从根本上解决隐层节点特征的随意性,而有隐层节点指导的 HDBP 算法则很好地解决了这个问题.

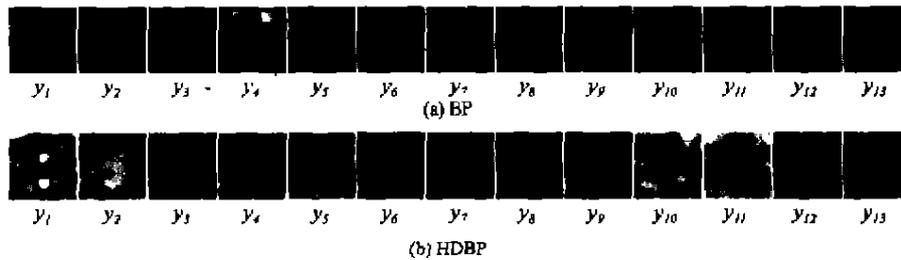


图 4 训练后的权值灰度图 (a)BP, (b)HDBP  
Fig. 4 Image of weights after training (a)BP, (b)HDBP

在神经网络字符识别中,相似字符的区分错误是造成识别错误的重要原因,文献<sup>[12]</sup>中采用正负面样本训练的方法来加以改善,但没有加入必要的人工分析指导.采用 HDBP 算法之后,对字符混淆的产生原因加以分析之后的领域知识可以方便地加入到对训练过程的指导中去,可使相似字符的识别错误率大大下降.

### 3 结语

传统的 BP 网训练时只对输出层有教师信号,对隐层不加指导,把 BP 网当成了黑盒,只管由输入得到输出,不管中间的学习,不教授学习方法,实际上 BP 网是处于盲目自学状态.训练后隐层的节点所代表的特征具有随意性,各种特征无意义散放在隐层中,使得鲁棒性不好<sup>[12]</sup>.而 HDBP 在训练时,隐层也有人工分析后提供的基于领域先验知识的指导,使准确率和鲁棒性均得到提高.另外传统的 BP 网络增加新的样本后,因为以前训练时随意得到的隐节点特征要重新调整分配,所以要针对所有的样本重新进行训练是其一个重要特征,而 HDBP 算法中每个隐层节点的输出有其特定的意义,增加新的待识别模式后,原有隐层节点输出意义不变.根据对新的模式进行分析之后,确定其利用原有的隐层节点意义,以及增加新的隐层节点.由于原有隐层节点的意义没有发生变化,训练不必完全重新开始,而是从原有隐层节点的权值开始训练,原有隐层节点的所有权值信息都作为训练初始值加以利用,这就大大减小了反复训练时的代价.

总之,本方法较成功地解决了 BP 算法隐层无指导信号带来的弊病,使得人工分析得出的领域先验知识能够容易地加入到神经网络的训练过程中,减少了训练的代价,提高了识别率.尤其在分析神经网络的出错情况之后对网络重新训练时,此方法有着很重要的用途.同时,HDBP 算法的计算量仅有很小的增加,因此是一种有效的改进算法.

## REFERENCES

- 1 BAO Li-Wei. *Pattern Recognition and Artificial Intelligence*(鲍立威. 模式识别与人工智能), 1995, 8: 1-5
- 2 SUN De-Bao, GAO Chao. *Information and Control*(孙德保, 高超. 信息与控制), 1995, 24(5): 283~287
- 3 EHUD D. KARNIN. *IEEE Trans. Neural Networks*, 1990, 1(2): 239~242
- 4 Shih-Chi Huang, Yih-Fang Huang. *IEEE Trans. Neural Networks*, 1991, 2(1): 47~55
- 5 Minoru Fukume, Sigeru Omatu. *IEEE Trans. Neural Networks*, 1991, 2(5): 535~538
- 6 Richard P. Brent. *IEEE Trans. Neural Networks*, 1991, 2(3): 346~354
- 7 LI Yian-Bin, QI Fei-Hu, et al. *Radio Engineering*(李艳斌, 戚尺虎, 等. 无线电工程), 1995, 25: 6
- 8 ZHU Jiang-Hai, QI Fei-Hu. *J. Infrared and Millimeter Waves*(朱江海, 戚飞虎. 红外与毫米波学报), 1997, 16(5): 357~363
- 9 SUN Yao-Ru. *Microcomputer*(孙杏如. 微型计算机), 1995, 15: 38
- 10 J. J. Mulawka, B. K. Verma. *Microcomputer Applications*, 1994, 13(2): 85~88
- 11 Y. C. Lin, F. Y. Wu. *Microcomputer Applications*, 1994, 13(1): 37~41
- 12 WANG J, JEAN J. *Pattern Recognition*, 1993, 26(1): 175~187
- 13 Brown R H, Ruchti T L. *First IEEE Conference on Control Application*, 1992, 1: 571
- 14 HU Tie-Song, CAI De-Suo, et al. *C<sup>2</sup>N<sup>2</sup>-96*, (胡铁松, 蔡德所, 等. 1996 中国神经网络学术大会), 1996, 1: 243~246

### A NEW LEARNING ALGORITHM FOR MLP WITH KNOWLEDGE DIRECTED HIDDEN-LAYER\*

ZHU Jiang-Hai<sup>1)</sup> LI Mei<sup>2)</sup> QI Fei-Hu<sup>1)</sup>

<sup>1)</sup>Department of Computer Science & Engineering, Shanghai Jiaotong University, Shanghai 200030, China;

<sup>2)</sup>Computing Center, Shanghai Tiedao University, Shanghai 200331, China

**Abstract** A new algorithm to deal with the shortcuts of traditional backpropagation algorithm was proposed. In the proposed algorithm, the hidden layer is directed by knowledge. The training effects are then greatly improved and the number of hidden neurons can be easily determined. The training need not be restarted entirely when adding new patterns, so that the training time will be saved obviously.

**Key words** HDBP algorithm, BP algorithm, neural network.

---

\* The project supported by the National Natural Science Foundation of China  
Received 1997-07-22, revised 1998-03-05