

## BP 算法分析与改进

贾丽会, 张修如

(中南大学 信息科学与工程学院, 湖南 长沙 410075)

**摘要:**在人工神经网络中, BP 神经网络是一种应用广泛的多层前馈神经网络。分析了 BP 算法的基本原理, 指出了 BP 算法具有收敛速度慢、易陷入局部极小点等缺陷以及这些缺陷产生的根源。针对这些缺陷, 通过在标准 BP 算法中引入变步长法、加动量项法、遗传算法、模拟退火算法等几种方法来优化 BP 算法。实验结果表明, 这些方法有效地提高了 BP 算法的收敛性, 避免陷入局部最小点。

**关键词:**BP 算法; 收敛速度; 局部极小点; 遗传算法; 模拟退火算法

**中图分类号:**TP183

**文献标识码:**A

**文章编号:**1673-629X(2006)10-0101-03

## Analysis and Improvements of BP Algorithm

JIA Li-hui, ZHANG Xiu-ru

(College of Information Science &amp; Engineering, Central South University, Changsha 410075, China)

**Abstract:** Back-propagation neural network is an extensively applied multi-layer feedforward neural network in artificial neural network. Basic principle of BP algorithm is analyzed firstly. Then some defects such as slow convergence rate and getting into local minimum in BP algorithm are pointed out, and the root of the defects is presented. Finally, in view of these limitations, several methods such as genetic algorithm and simulated annealing algorithm etc. are led to optimize BP algorithm. Experiment results show that these methods increase efficiently the convergence performance of BP algorithm and avoid local minimum.

**Key words:** BP algorithm; convergence rate; local minimum; genetic algorithm; simulated annealing

## 0 引言

由于 BP (back propagation) 神经网络结构简单、算法成熟且具有很强的非线性映射能力, 故它在模式识别、故障诊断、医学工程等各个领域得到了广泛的应用。虽然理论上 BP 网络能逼近任意的非线性函数, 但由于 BP 算法的误差函数曲面高度复杂, 并且按梯度下降法调整权值, 使得网络在学习过程中出现了收敛速度慢、易陷入局部极小点等缺点<sup>[1,2]</sup>, 从而限制了 BP 神经网络的应用。文中针对这些缺陷, 介绍了几种改进方法来优化 BP 算法。

## 1 BP 神经网络与学习算法

BP 神经网络含有输入层、输出层以及处于输入输出层之间的隐含层。隐含层又包括单层或多层。图 1 是一个典型的 3 层 BP 神经网络, 只有一层隐含层。

设输入为  $x = (x_1, x_2, \dots, x_n)$ ; 隐层有  $r$  个神经元; 输出层有  $m$  个神经元, 输出为  $y = (y_1, y_2, \dots, y_m)$ ; 输入层到隐层的权为  $w_{ij}$ , 阈值为  $\theta_j$ ; 隐层到输出层的权为  $w_{jk}$ ,

阈值为  $\theta_k$ 。于是隐层各神经元的输出为:

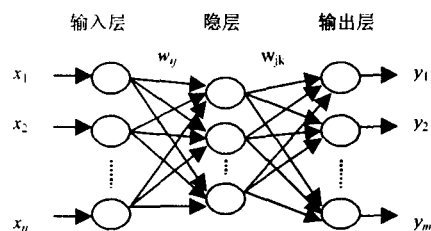


图 1 典型的 3 层 BP 神经网络

$$x'_j = f\left(\sum_{i=1}^n w_{ij}x_i - \theta_j\right), \quad j = 1, 2, \dots, r$$

输出层各神经元的输出为:

$$y_k = f\left(\sum_{j=1}^r w_{jk}x'_j - \theta_k\right), \quad k = 1, 2, \dots, m$$

BP 算法也称误差反向传播算法 (Error Back-propagation Algorithm), 是一类有指导学习算法, 用于 BP 网的权值和阈值学习。BP 算法实质是求均方误差函数的最小值问题。这种算法采用非线性规划中的最速下降法, 按误差函数的负梯度方向修改权系数。设有  $p$  个学习样本矢量, 对应的期望输出为  $d^{(1)}, d^{(2)}, \dots, d^{(p)}$ , 学习是通过误差校正权值, 使各  $y^{(p)}$  接近  $d^{(p)}$ 。通常为简化推导, 把各计算节点的阈值并入到权矢量。当第  $p$  个样本输入网络, 并产生输出时, 均方误差应为  $m$  个输出单元误差平方之和, 即

收稿日期: 2005-12-06

**作者简介:**贾丽会 (1976-), 女, 湖北钟祥人, 硕士研究生, 讲师, 研究方向为图形图像处理技术、信号处理、模式识别; 张修如, 副教授, 研究方向为信息系统、GIS、图形图像处理技术、模式识别。

$$E^{(p)} = \frac{1}{2} \sum_{l=1}^m (d_l^{(p)} - y_l^{(p)})^2$$

当所有样本都输入一次后,总误差为:

$$E_A = \sum_{p=1}^P E^{(p)} = \frac{1}{2} \sum_{p=1}^P \sum_{l=1}^m (d_l^{(p)} - y_l^{(p)})^2$$

设  $w_{sp}$  为网络中的一个连接权值,则根据梯度下降法,权值修正量应为:

$$\Delta w_{sp} = -\eta \frac{\partial E}{\partial w_{sp}}$$

其中,若是批量学习,则  $E = E_A$ ;若是增量学习,则  $E = E^{(p)}$ 。采用梯度法,通过每次调整的增量  $\Delta w_{ij}$  和  $\Delta w_{jk}$ ,使总的误差向减小的方向变化,直到满足一定的条件。这里以批量学习为例,通过求解,各层权值第  $n_0 + 1$  次与第  $n_0$  次的关系为:

$$w_{jk}(n_0 + 1) = w_{jk}(n_0) + \eta \sum_{p=1}^P \delta_{jk}^{(p)} x_j^{(p)}$$

$$w_{ij}(n_0 + 1) = w_{ij}(n_0) + \eta \sum_{p=1}^P \delta_{ij}^{(p)} x_i^{(p)}$$

式中,  $\delta_{jk}^{(p)} = (d_j^{(p)} - y_j^{(p)}) y_j^{(p)} (1 - y_j^{(p)})$

$$\delta_{ij}^{(p)} = \sum_{k=0}^m \delta_{jk}^{(p)} w_{jk} x_j^{(p)} (1 - x_j^{(p)})$$

显然,BP 网络学习分两个阶段:由前向后正向计算各隐层和输出层的输出,由后向前误差反向传播,用于权值修正。

## 2 BP 算法存在的不足及原因分析

BP 算法是目前应用最广泛的神经网络学习算法,但人们在使用过程中发现 BP 算法也存在着训练时间长、收敛速度慢、易陷入局部最小点等缺陷。

### 2.1 收敛速度慢

在 BP 算法中,对于比较简单的问题也要经过几千次才能达到充分的学习。BP 算法实质是求均方误差函数的最小值问题。为保证算法的收敛性,学习率必须小于某一上限。这就决定了 BP 算法的收敛速度不可能很快。另外,在 BP 网络的训练过程中,有时候会进入误差曲面的平坦区。这时,误差虽然也在减少,但在一定时间内的减少极度缓慢,从而使训练次数大大增加,影响了收敛速度。导致 BP 算法收敛速度慢的原因很多,譬如:学习步长、初始权值、目标函数等的选取都会对收敛速度产生影响。Fahlman 等人<sup>[3]</sup>认为,导致 BP 算法收敛速度慢的主要原因有两个:学习步长问题和变动目标问题,并提出了著名的神经网络构造方法:级联相关(Cascade - Correlation, CC)算法,解决了传统 BP 算法收敛速度慢的缺点。

### 2.2 易陷入局部最小点

BP 算法本质上采用的是梯度下降法,训练是从某一起点沿误差函数的斜面逐渐达到误差的最小值。误差曲面的多极小点会使训练陷入局部极小。实质上,误差函数为一个凹凸不平的多维空间的超曲面,存在非常多的“低谷”。因而在 BP 神经网络进行训练时,可能陷入某一小谷区(局部最小点),在这一小谷区里,网络的输出与期望

输出之间的误差在某一个值的左右徘徊,但是无法再减少,这时就表明网络的训练陷入了局部极小值而不是全局最小值。

平坦区、局部最小值与全局最小值的二维示意图如图 2 所示。

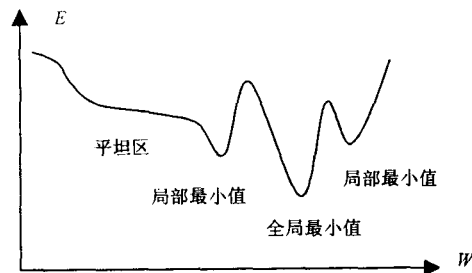


图 2 误差平坦区、局部最小值与全局最小值二维示意图

以上两个问题都是 BP 算法的固有缺陷,其根源在于按误差梯度下降的权值调整原则是每一步求解都取局部最优,即所谓“贪心”(Greedy)算法的原则。

## 3 BP 算法的几种改进方法

### 3.1 变步长法

BP 算法的有效性和收敛性在很大程度上取决于学习步长  $\eta$  的值。采用一般的固定长梯度下降法求解时,起码可能导致两个主要问题:局部极小解;收敛速度慢。所以,一般要求是,当训练到误差曲面的平坦区时,梯度较小,为加快收敛应使  $\eta$  增大;当训练到深而窄的误差曲面时,应使  $\eta$  减小,以免因步长过大而出现振荡现象。为加快收敛,应使  $\eta$  合理化,可采用变步长算法。变步长算法的基本思想是,先设一个初始步长  $\eta$ ,若一次迭代后误差增大,则将步长乘以  $\beta (< 1)$ ,沿原方向重新计算该迭代点;若一次迭代后误差减小,则将步长乘以  $\alpha (> 1)$ ,计算下一个迭代点,以缩短学习时间。

### 3.2 加动量项法

为了加速 BP 算法的收敛,根据 Rumelhart 等人的建议,可考虑在权值调整算式中加入动量项<sup>[4]</sup>,即

$$\Delta w(n+1) = -\eta \frac{\partial E}{\partial w(n)} + \alpha \Delta w(n)$$

式中,  $\alpha$  为动量因子,一般取 0.1~0.8。这时权值修正量加上了有关上一时刻权值修改方向的记忆,加速了网络的收敛。加动量项法的具体原理:若相邻两次迭代点处的梯度方向是一致的,引入动量项可使权值的调整量增大,从而加速收敛;若相邻两次迭代点处的梯度方向相反,引入动量项可使权值的调整量减小,避免了来回振荡,加快了收敛。

### 3.3 串连法

BP 算法的收敛速度主要取决于输入-输出模式间非线性映射的复杂程度。显然,这种非线性映射关系越复杂,收敛时间越长。因此,对那些高度复杂的非线性问题,用两个串连的 BP 网络代替一个 BP 网络,能够有效地缩

短训练时间。

### 3.4 利用遗传算法优化 BP 算法

BP 算法的优点是寻优具有精确性,但它易陷入局部极小、收敛速度慢,而遗传算法(Genetic Algorithm, GA)是基于自然选择和遗传规律的全局优化搜索算法,具有很强的宏观搜索能力和寻优全局性。因此,在 BP 神经网络理论中引入遗传算法的思想,则可以很好地达到全局寻优和快速高效的目的。用 GA 优化神经网络主要包括 3 个方面:优化网络结构、优化权系数及同时优化网络结构和学习规则等<sup>[5]</sup>。文中主要介绍在 BP 网络结构固定的前提下,先用 GA 优化 BP 网络的初始权值,然后用 BP 算法精确地搜索出权值最优解的实现方法,其具体步骤如下:

(1) 初始化种群  $p$ , 包括种群规模、交叉概率、突变概率和随机产生的个体。种群中的每个个体代表一个网络所有初始权值(包括阈值)编码的集合。

(2) 将种群中的每个个体解码成网络的各个权值,然后运行网络,计算各个个体的适应度函数(评价函数)并将其排序。

(3) 对当代种群,按交叉概率和突变概率分别进行交叉操作和变异操作产生新个体。

(4) 计算新个体的适应度函数值并根据值的大小将新个体插入到种群  $p$  中。

(5) 根据评价函数的顺序,仅选择一半具有较高适应度的个体,形成下一代种群。如果找到了满意的个体,或达到最大进化代数,则转(6);否则转(3)。

(6) 将满意的个体进行解码,从而得到通过 GA 优化后的 BP 网络的初始权值。

(7) 按照 BP 学习算法不断进行网络权值的调整,若 BP 网络达到指定精度( $E < \epsilon$ ),或达到最大学习次数,则停止学习并输出此时的权值;否则继续学习训练。

遗传算法具有全局寻优的特点,而 BP 算法具有精确训练的特点,将遗传算法和 BP 算法相结合可以具有寻优的全局性和精确性。

### 3.5 采用模拟退火算法改进 BP 算法

模拟退火算法(SA, simulated annealing)<sup>[6]</sup>是一种著名的全局寻优算法,具有描述简单、应用灵活并极少受初始条件限制等优点。在神经网络中引入模拟退火的思想可以避免算法陷入局部最小。实际上,把模拟退火算法应用于神经网络时,网络的可调节权值等效于金属中的粒子,而网络的输出误差(代价函数)则等效于金属的能量状态。模拟退火算法在神经网络中的概念与应用的原理:为了使网络收敛到低温下的平衡态,可在开始时给定一个较高温度,然后用模拟退火技术逐渐降温,最后系统以相当高的概率收敛到网络的最小能量状态。用模拟退火算法改进的 BP 算法也称为加“噪声”的 BP 算法,模拟退火算法和 BP 算法混合学习策略具体实现步骤如下:

(1) 随机产生初始权值作为初始状态  $s$ , 并设置初始高温为  $T$ 。

(2) 利用 BP 算法求权值  $W$ , 达到状态  $a$ 。

(3) 在所有的权上加一个噪声,获得新状态  $a'$  (在状态  $a$  的领域内)。

(4) 计算  $\Delta E = E_{a'} - E_a$ 。

(5)  $\Delta E \leq 0$  时,接受新状态,转到步骤(6);  $\Delta E > 0$  时,若  $\exp(-\Delta E/T) > \text{random}(0,1)$ , 接受新状态(爬坡),转到步骤(6);否则转到步骤(3),再随机改变状态。

(6) 从新状态起,用 BP 算法继续进行权值修正,直到稳定平衡状态  $b$ 。

(7) 若  $E_b \geq E_a$ , 搜索失败,转到步骤(3);若  $E_b < E_a$ , 转到步骤(8)。

(8) 若  $E_b < \epsilon$ , 停止;否则  $k = k + 1$ ,  $T = T/(1+k)$ , 转到步骤(3),从状态  $b$  继续模拟退火。

很显然,采用模拟退火算法改进 BP 算法,可以使原来一味追求网络误差单调下降(即只会“下坡”)的情况增加了“爬坡”能力,使网络能跳出局部极小点而收敛于全局极小点。利用模拟退火算法虽然可以有效地达到全局能量最小状态,但收敛速度很慢,往往和上面提到的方法结合使用。

## 4 仿真实验与结果

异或问题是典型的线性不可分问题,常用于网络算法的验证,因此文中用它进行仿真实验。异或问题比较简单,没必要采用串连法。这里, BP 神经网络采用 2-2-1 的结构,输入模式为  $\{0,0\}, \{0,1\}, \{1,0\}, \{1,1\}$ ; 对应的输出模式为  $\{0\}, \{1\}, \{1\}, \{0\}$ 。网络所有的初始权值和阈值在  $[-20,20]$  范围内随机取值,分别用标准 BP 算法和改进的算法进行实验,实验结果如表 1 所示。

表 1 异或问题的实验结果对照表

学习算法	迭代数	误差 $E$
标准 BP 算法( $\eta = 0.1$ )	3500	0.03
变步长法	50	0.03
加动量项法	1000	0.01
遗传算法	100	$1.2\exp(-6)$
模拟退火算法 & BP 算法	1000	0.0401
变步长法 & 模拟退火算法	100	$3.9\exp(-6)$

从实验得出:标准 BP 算法收敛速度最慢,并且当学习率很小的时候还容易陷入局部最小点,达不到要求的误差精度;变步长法和加动量项法加快了网络的收敛;遗传算法不仅加快了网络的速度,并且使网络收敛于全局最小;模拟退火算法和 BP 算法相结合虽然收敛速度很慢,但能使网络避免陷入局部最小;在实验中尝试将变步长法和模拟退火算法结合使用,网络的收敛速度得到很大的提高,并且使网络收敛于全局最小。

## 5 结束语

在分析标准 BP 算法的基础上,指出了 BP 算法存在收敛速度慢、易陷入局部极小点等缺陷以及这些缺陷产生的原因。为了优化 BP 算法,文中提出了五种改进方法:

(下转第 107 页)

成一个梯度隶属度函数。如图 3 所示。

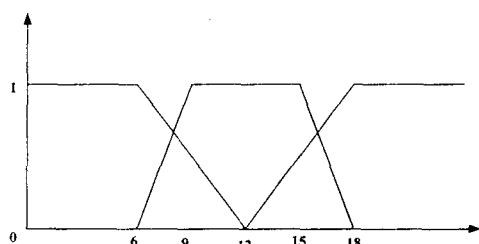


图 3 汉字的隶属度函数

对汉字的垂直直方图,其左右结构也用以上梯度隶属度函数(见图 3)表示。于是可以得到汉字的谷点位置的隶书度的信息,文中用一个 6 维的特征向量表示。

提取汉字轮廓结构的模糊统计特征最后得到一个 17 维向量:

$$F = [F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}, F_{13}, F_{14}, F_{15}, F_{16}, F_{17}]$$

其中  $F_1$  表示手写体汉字的轮廓特征信息;  $F_2 \sim F_{11}$  为汉字的结构特征信息;  $F_{12} \sim F_{17}$  为谷点的模糊特征信息。

#### 4 识别试验

文中以金连文实验室的手写字体为实验样本,每个汉字有样本 40 个,其中 30 个作为训练样本,10 个该汉字样本加上其他 10 个汉字样本共 20 个样本作为待识别样本,实验选取了 10 个汉字共 400 个样本。特征提取前先对样本预处理,其中归一化采用插值法,把样本归一化为 2424 点阵。去噪采用中值滤波,特征提取结合小波网格、笔画密度特征和直方图的模糊统计特征方法提取汉字的特征,并加入了部分冗余维特征。最后形成了一个 140 维特征。

采用 Matlab 的支持向量机工具箱:OSU-SVM3.00。分别选取高斯径向基核和多项式核作为核函数进行训练识别,其识别结果如表 1 所示。

#### 5 结论

从识别结果可以看出,支持向量机对较规则手写字符有比较好的识别结果,对于小样本识别问题,用支持向量

机训练识别是可行的,文中所采用的笔划密度特征和直方图统计特征及模糊统计特征方法对手写汉字的特征提取及数据融合是科学有效的。另外对于手写字符撇捺方向的变形,可能会导致识别率的下降,如何解决此问题还有待进一步的研究。

表 1 基于支持向量机方法的手写汉字识别结果

字符	识别样本个数	多项式核识别率	多项式核误识率	径向基核识别率	径向基核误识率
大	20	0.90	0.10	1.00	0.00
吹	20	1.00	0.00	1.00	0.00
处	20	0.90	0.10	0.85	0.15
川	20	1.00	0.00	1.00	0.00
计	20	0.90	0.10	0.95	0.05
崔	20	0.95	0.05	0.95	0.05
呆	20	1.00	0.00	1.00	0.00
寸	20	1.00	0.00	0.95	0.00
从	20	0.80	0.20	0.85	0.15
括	20	1.00	0.00	0.90	0.10

#### 参考文献:

- [1] 张中. 汉字识别技术综述[J]. 语言文字应用, 1997(2): 77-86.
- [2] Vapnik V N. The Nature of Statistical of Learning Theory [M]. New York Spring, 1995.
- [3] 柳回春, 马树元. 支持向量机的研究现状[J]. 中国图像图形学报, 2002(6): 619-623.
- [4] 王晓光, 王群. 用于车牌字符识别的 SVM 算法[J]. 现代电子技术, 2004(8): 8-10.
- [5] 张学工. 关于统计学习与支持向量机理论[J]. 自动化学报, 2000(1): 32-42.
- [6] 金连文, 彭秀兰, 尹俊勋. 一种手写体汉字特征提取新方法—小波变换及弹性网络技术的应用[J]. 中国图像图形学报, 1998(7): 499-552.
- [7] 王志红. 小波和神经网络模式识别技术及其在车牌识别中的运用[D]. 合肥: 合肥工业大学, 2003. 30-55.
- [8] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2002. 284-325.

(上接第 103 页)

变步长法、加动量项法、串连法、利用遗传算法优化 BP 算法、采用模拟退火算法改进 BP 算法。其中, 前三种方法在加快网络收敛速度方面比较显著, 后两种方法主要是避免网络学习陷入局部极小点从而能够收敛到全局最小点。在实际应用中表明, 将以上几种方法结合使用, 网络的学习效果将更理想。

#### 参考文献:

- [1] Weymace N, Martens J P. A fast and robust learning algorithm for feedforward neural networks[J]. Neural Networks, 1991, 4(3): 363-369.

- [2] Chan L W, Fallside F. An adaptive training algorithm for back propagation networks[J]. Computer Speech and Language, 1987, 2: 205-218.
- [3] Fahlman S C, Lebiere C. The Cascade Correlation Learning Architecture[J]. Advance in neural information processing systems, 1990, 2: 524-532.
- [4] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagation errors[J]. Nature, 1986, 323: 533-536.
- [5] 阎平凡, 张长水. 人工神经网络与模拟进化计算[M]. 北京: 清华大学出版社, 2001.
- [6] Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by simulated annealing[J]. Science, 1983, 220: 671-680.