

文章编号:1009-3087(2000)02-0105-05

## 人工神经网络 BP 算法的改进及其应用

李晓峰, 刘光中

(四川大学 管理科学与工程系, 成都 610065)

**摘 要:**对传统的 BP 算法进行了改进,提出了 BP 神经网络动态全参数自调整学习算法,又将其编制成计算机程序,使得隐层节点和学习速率的选取全部动态实现,减少了人为因素的干预,改善了学习速率和网络的适应能力。计算结果表明:BP 神经网络动态全参数自调整算法较传统的方法优越,训练后的神经网络模型不仅能准确地拟合训练值,而且能较精确地预测未来趋势。

**关键词:**人工神经网络;BP 算法;自调整;自组织方法**中图法分类号:**0224;029**文献标识码:**A

## The Improvement of BP Algorithm and Its Application

LI Xiao-feng, LIU Guang-zhong

(Dept. of Management Sci. and Eng., Sichuan Univ., Chengdu 610065, China)

**Abstract:** In this paper, BP algorithm of artificial neural network is improved, the self-adjusted algorithm of all parameters has been proposed for the back-propagation learning, and programmed a C language procedure. It can make the selection of hidden layer units and rate of studying easily in the course of training, reduce artificial influence and improve the adaptive ability of rate of studying and neural network. Our conclusion shows that the self-adjusted BP algorithm of all parameters is superior to the statistical modeling approach, the model of artificial neural network in training can not only exactly imitate training valuation but also make prediction accurately.

**Key words:** artificial neural network; BP algorithm; self-adjusted; group method of data handling(GMDH)

近年来,国际上掀起了一股人工神经网络研究、开发应用的热潮。人工神经网络的理论的应用已渗透到各个领域,并在智能控制、模式识别、计算机视觉、自适应滤波和信号处理、非线性优化、自动目标识别、生物医学工程等方面取得了显著成效。BP 神经网络模型是人工神经网络的重要模型之一。通常, BP 算法是通过一些学习规则来调整神经元之间的连接权值,在学习过程中,学习规则以及网络的拓扑结构不变。

然而,一个神经网络的信息处理功能不仅取决

于神经元之间的连接强度,而且与网络的拓扑结构(神经元的连接方式)、神经元的输入输出特性和神经元的阈值有关。因而,神经网络模型要加强自身的适应和学习能力,应该知道如何合理地自组织网络的拓扑结构,改变神经元的激活特性以及在必要时调整网络的学习参数等。本文基于此,在前人研究的基础上<sup>[7]</sup>,对传统的 BP 算法进行了改进,提出了 BP 神经网络动态全参数自调整学习算法,不仅加快了网络的收敛速度,而且可以优化网络的拓扑结构,从而增强了 BP 神经网络的适应能力。

**收稿日期:**1999-06-17**作者简介:**李晓峰(1972-),男,助教,数量经济。

## 1 BP 人工神经网络模型的构造和评述

### 1.1 BP 神经网络概言

BP 神经网络,即误差反向传播神经网络,是神经网络模型中应用最广泛的一种。它由输入层、隐含层和输出层构成。假设 BP 神经网络每层有  $N$  个节点,作用函数为非线性的 Sigmoid 型函数,一般采用  $f(x) = 1/(1 + e^{-x})$ ,学习集包括  $M$  个样本模式  $(X_p, Y_p)$ 。对第  $P$  个学习样本( $P = 1, 2, \dots, M$ ),节点  $j$  的输总和记为  $net_{pj}$ ,输出记为  $O_{pj}$ ,则:

$$net_{pj} = \sum_{i=0}^N W_{ji} O_{pi} \quad O_{pj} = f(net_{pj})$$

如果任意设置网络初始权值,那么对每个输入样本  $P$ ,网络输出与期望输出( $d_{pj}$ )间的误差为

$$E = \sum E_p = (\sum_j (d_{pj} - O_{pj})^2)/2$$

BP 网络的权值修正公式为

$$W_{ji} = W_{ji}(t) + \eta \delta_{pj} O_{pi}$$

$$\delta_{pj} = \begin{cases} f'(net_{pj})(d_{pj} - O_{pj}), & \text{对于输出节点} \\ f'(net_{pj}) \sum_k \delta_{pk} W_{kj}, & \text{对于输入节点} \end{cases}$$

上式中引入学习速率  $\eta$ ,是为了加快网络的收敛速度。通常权值修正公式中还需加一个惯性参数  $\alpha$ ,从而有:

$$W_{ji} = W_{ji}(t) + \eta \delta_{pj} O_{pi} + \alpha (W_{ji}(t) - W_{ji}(t-1))$$

式中,  $\alpha$  为一常数项,它决定上一次的权值对本次权值的影响。其具体算法步骤详见文献[1]。

### 1.2 传统 BP 网络的评述

传统 BP 网络模型把一组样本的输入/输出问题变为一个非线性优化问题,使用了优化中的最普通的梯度下降算法,对问题的识别具有很强的功能,对于复杂的非线性模型仿真从理论上来说其误差可以达到任意小的程度。但它仍然存在一些缺陷:

1) 传统的 BP 网络既然是一个非线性优化问题,这就不可避免地存在局部极小问题。网络的极值通过沿局部改善的方向一小步一小步进行修正,力图达到使误差函数  $E$  最小化的全局解,但实际上常得到的是局部最优解。

2) 学习过程中,下降慢,学习速度缓,易出现一个长时间的误差平坦区,即出现平台。

3) 网络结构选择不当,网络过大,在训练中效率不高,而且还会由于过拟合造成网络性能脆弱,容错性下降,浮点溢出,而太小的网络可能根本不收敛。在实际应用中,网络结构人为性较大,缺乏规则指导。

## 2 BP 人工神经网络模型的改进

### 2.1 BP 人工神经网络结构的自我调整

在 BP 人工神经网络拓扑结构中,输入节点与输出节点是由问题的本身决定的,关键在于隐层的层数与隐节点的数目。对于隐层的层数,许多学者作了理论上的研究。Lippmann<sup>[2]</sup> 和 Cyberko<sup>[3]</sup> 曾指出,有两个隐层,就可以解决任何形式的分类问题;后来 Robert Hecht Nielson<sup>[5]</sup> 等人研究进一步指出:只有一个隐层的神经网络,只要隐节点足够多,就可以以任意精度逼近一个非线性函数。相对来说,隐节点数的选取很困难。隐节点少了,学习过程不可能收敛;隐节点多了,网络性能下降,节点冗余。为了找到合适的隐节点数,最好的办法是在网络的学习过程中,根据环境要求,自组织地学习、调整自己的结构,最后得到一个大小合适的神经网络模型。

设  $O_{pi}$  是隐节点  $i$  在学习第  $p$  个样本时的输出,  $O_{pj}$  是隐节点  $j$  在学习第  $p$  个样本时的输出,  $N$  为学习样本总数,

$$\bar{O}_i = \frac{1}{N} \sum_{p=1}^N O_{pi} \quad \bar{O}_j = \frac{1}{N} \sum_{p=1}^N O_{pj}$$

$$x_p = O_{pi} - \frac{1}{N} \sum_{p=1}^N O_{pi} = O_{pi} - \bar{O}_i$$

$$y_p = O_{pj} - \frac{1}{N} \sum_{p=1}^N O_{pj} = O_{pj} - \bar{O}_j$$

则两序列  $O_{pi}$  与  $O_{pj}$  ( $p = 1, \dots, N$ ) 的相关系数为:

$$\rho_{ij} = \frac{\sum_{p=1}^N x_p y_p}{\sqrt{\sum_{p=1}^N x_p^2} \cdot \sqrt{\sum_{p=1}^N y_p^2}}$$

显而易见:  $|\rho_{ij}| \leq 1$ 。

$\rho_{ij}$  愈接近  $\pm 1$ , 表示两序列  $O_{pi}$  与  $O_{pj}$  的线性相关程度愈大,互相回归的离散度愈小;反之,则两序列的线性相关程度愈小,互相回归的离散度愈大。

定义 1° 同层隐节点  $i$  和  $j$  的相关系数

$$\rho_{ij} = \frac{\sum_{p=1}^N x_p y_p}{\sqrt{\sum_{p=1}^N x_p^2} \cdot \sqrt{\sum_{p=1}^N y_p^2}}$$

$\rho_{ij}$  说明隐节点  $i$  和  $j$  的相关程度,  $\rho_{ij}$  过大,说明节点  $i$  和  $j$  功能重复,需要压缩合并。

定义 2° 样本发散度  $S_i$

$$S_i = \frac{1}{N} \sum_{p=1}^N O_{pi}^2 - \bar{O}_i^2$$

$S_i$  过小,说明隐节点  $i$  的输出值变化很少,它对网络的训练没有起到什么作用,则可以删除。根据以上定义,提出如下动态合并与删减规则。

规则1 若  $|\rho_{ij}| \geq C_1$  且  $S_i, S_j \geq C_2$  则同层隐节点  $i$  和  $j$  可以合二为一。其中  $C_1$  和  $C_2$  为规定的下限值,一般  $C_1$  取  $0.8 \sim 0.9$ ,  $C_2$  取  $0.001 \sim 0.01$ 。

规则2 若  $S_i < C_2$ ,则节点  $i$  的作用如同阈值,可与阈值合并,即节点  $i$  被删除。

## 2.2 神经网络学习参数的自适应学习

在传统的 BP 神经网络模型中,对学习参数  $\eta$  的选取一般都是根据建模者的经验选取一个值。但事实上  $\eta$  的选取对算法的成败有着重要影响。BP 神经网络模型实际上是一个多元函数的优化问题,即以连结权系数为变量,误差函数最小为目标的优化问题。当求出对连结权的梯度后,对连结权进行修正时,学习速率  $\eta$  实际上是一个沿负梯度方向的步长问题,步长过大会使误差函数发生振荡,步长过小,收敛过慢。并且在不同的点,可选步长也不一样。总之,不存在原 BP 算法中所提到的固定学习速率。

对于学习参数  $\eta$  的选取,我们可以引入学习参数自适应算法加以确定。其基本思想是:当  $w_{jk}$  远离稳定点(学习要达到的目标点)时,  $\eta$  取较大值,而当其逼近稳定点 ( $E_1 \rightarrow 0$ ) 时,  $\eta$  取较小值。具体作法为:

$$\eta(t+1) = \eta(t) \cdot E_1(t)/E_1(t-1)$$

其中,  $E_1 = \sum_{p=1}^N E_p/N$ ,  $N$  为学习样本容量。 $E_1$  称为全局平均误差。

该方法也有缺陷,对局部极小值往往无法“逃避”。针对 BP 算法中存在局部极小这一问题,本文采用统计学上随机抽取的思想。在算法中产生大量的随机初始点,逐次进行迭代最终选优。这样,误差函数  $E$  能取得全局最优解的概率将大大提高。

## 3 动态全参数自调整学习算法

综合 1、2 部分,我们得到结构参数和学习参数自调整的算法,其具体步骤如下:

1) 根据实际问题,设定一个较大规模的网络,并对网络初始化;

2) 输入学习样本,使样本参数变化为  $[0, 1]$  区间;

3) 在  $[-1, 1]$  之间随机产生数值赋给初始权矩阵;

4) 按改进的 BP 法训练网络;

5) 判断迭代步数是否超过规定步数或学习精度达到要求否,是,转入 6); 否,返回 4),继续学习;

6) 计算隐节点间的相关参数及发散度。先按规则 2,进行节点删除;若规则 2 不满足,按规则 1,进行节点合并;若两规则都满足,则只按规则 2 删除;若两规则都不满足,则不进行结点合并、删除。若无节点合并、删除,则转到步骤 7); 否则,返回 3);

7) 学习精度是否达到要求或迭代参数是否超过规定步数,是,算法终止;否,返回 3)。

## 4 应用实例及与自组织方法的比较

能源是国民经济发展的基础。它是一种极其宝贵的自然资源。随着社会的发展,人们生活质量的提高,对能量的需求量日益加大。能源消费量不仅与一个国家的经济结构、科学水平、生产模式等因素有关,而且与地理条件、人民生活习惯、国家的发展政策等因素有关。因此,能源的消费量与影响它的因素之间存在着复杂的关系。表 1 为我国能源消费量及相关因素一览表。

这些因素中影响能源消费量的主要有三个,即国民生产总值指数(1978 年 = 100)、广义节能率和能源产品出厂价格指数(上年 = 100)。能源消费量与国民生产总值关系密切,它们之间一般具有复杂的非线性关系。这里的广义节能率包含了由于科技进步、产业结构调整等而综合产生的节能效果。根据市场经济规律,能源产品出厂价格的高低也影响能源消费量。它们之间亦具有复杂的非线性关系。为计算方便,这里的能源产品出厂价格指数仅综合了煤炭和石油两种产品的出厂价格指数。我们利用动态全参数自调整 BP(三层)学习算法来预测能源消费量。神经网络的输入变量取三个:国民生产总值、广义节能率和能源产品出厂价格综合指数,即输入层节点数取 3。隐层节点数应适当取大些,这里取 10。输出层的节点数为输出向量的分量数,这里为 1,即能源消费量。

表 1 样本值

Tab.1 The date of Chinese energy resources consumption

序号	年份	国民生产 总值指数 (1989 年 = 100)	广义节 能率 %	煤炭在能源 消耗中的 比例 %	煤炭产品出 厂价格指数 (上年 = 100)	石油在能源 消耗中的 比例 %	石油产品出 厂价格指数 (上年 = 100)	能源产品出厂 价格综合指数 (上年 = 100)	能源消费量 实际值/mt of SCE
1	1978	100.0	4.79	70.7	100.0	22.7	100.0	100.0	571.44
2	1979	107.6	4.88	71.3	113.4	21.8	100.6	110.4	585.88
3	1980	116.0	4.53	72.2	106.4	20.7	102.1	105.4	602.75
4	1981	121.02	5.65	72.7	102.6	20.0	99.3	101.9	594.47
5	1982	131.8	3.09	73.7	101.9	18.9	100.5	101.6	626.46
6	1983	145.4	4.39	74.02	101.5	18.8	106.3	102.4	626.46
7	1984	170.9	8.67	75.3	102.6	17.4	112.0	104.4	709.04
8	1985	193.7	4.59	75.8	117.6	17.1	107.2	115.7	766.82
9	1986	210.2	2.84	75.8	96.8	17.2	104.6	98.2	808.50
10	1987	234.1	3.80	76.2	102.8	17.0	104.0	103.0	866.32
11	1988	260.5	3.56	76.2	110.6	17.0	106.8	109.9	928.97
12	1989	271.2	0.13	76.0	112.2	17.1	108.4	111.5	929.97
13	1990	281.8	2.00	76.2	106.2	16.6	107.1	106.4	987.03
14	1991	308.1	3.89	76.1	113.1	17.1	118.8	114.1	1037.83
15	1992	351.7	7.86	75.7	116.1	17.5	115.3	115.9	1091.7
16	1993	399.3	9.80	72.8	139.7	19.6	171.3	146.4	1117.68
17	1994	445.6	1.59	75.0	122.2	17.4	148.7	127.2	1193.73
18	1995	498.4	2.00	73.4	112.1	17.3	145.7	117.2	1249.11
19	1996	569.8	1.53	72.6	107.3	17.9	140.8	111.4	1268.27
20	1997	627.6	1.60	73.7	102.7	17.5	120.3	109.6	1298.69
21	1998	688.1	1.22	72.8	101.4	18.1	118.	108.1	1311.10

注:表中资料源自《中国统计年鉴》(1987~1997 年数据)及成都市经济信息中心(1998 年数据)

#### 4.1 动态全参数自调整 BP 神经网络预测模型

连接权值  $W, V$  矩阵。

##### 4.1.1 建立动态全参数自调整 BP 神经网络模型

由表 1, 我们得到 21 个样本输入输出数据组。将 1978~1994 年间的数作为网络训练样本, 1995~1998 年间的数作为检验预测结果的样本。取  $\eta = 0.15, \alpha = 0.075, \epsilon = 0.0002, C_1 = 0.9, C_2 = 0.001$ 。网络经过若干次训练后, 得到最优化的 BP 神经网络结构: 3-2-1 (输入层节点为 3 个, 隐层节点为 2 个, 输出层 1 个节点)。同时, 又得到了模型的最优

在网络训练过程中, 原 BP 神经网络结构由 3-10-1 调整到了 3-2-1。在此中间, 隐节点的合并, 删除共进行了两次。表 2 为第一次隐节点调整时的相关系数及发散度, 从中可看出节点 0、2、7、9 合并, 1、6、5、8 合并, 节点 3 被删除, 剩下 4、8、9 三节点。表 3 为第二次隐节点调整时的相关系数及发散度, 从中显而易得: 节点 4 与 9 合并, 最终剩下了 4、8 两节点。

表 2 隐节点第一次调整时的相关系数及发散度

Tab.2 The coefficients and the divergent degrees, adjusted for the first time

$i$	$j$										
	0	1	2	3	4	5	6	7	8	9	$S_i$
0	1.000	0.678	0.971	0.077	0.687	0.229	0.484	0.999	0.029	0.926	0.0561
1	0.678	1.000	0.485	0.678	0.066	0.868	0.968	0.688	0.713	0.351	0.0118
2	0.971	0.485	1.000	0.3113	0.839	0.004	0.267	0.967	0.265	0.988	0.2059
3	0.077	0.678	0.311	1.000	0.774	0.943	0.823	0.064	0.997	0.446	0.0005
4	0.687	0.066	0.839	0.779	1.000	0.547	0.299	0.677	0.745	0.890	0.0311
5	0.229	0.868	0.004	0.943	0.547	1.000	0.962	0.244	0.962	0.152	0.0490
6	0.484	0.968	0.267	0.823	0.299	0.962	1.000	0.498	0.854	0.121	0.0213
7	0.999	0.688	0.967	0.064	0.677	0.244	0.498	1.000	0.015	0.920	0.1364
8	0.029	0.713	0.265	0.997	0.745	0.962	0.854	0.015	1.000	0.404	0.0141
9	0.926	0.351	0.988	0.446	0.890	0.152	0.121	0.920	0.404	1.000	0.0026

表3 隐节点第二次调整时的相关系数及发散度

Tab.3 The coefficients and the divergent degrees, adjusted for the second time

i	j			
	4	8	9	$S_i$
4	1.000	0.601	0.996	0.0204
8	0.601	1.000	0.5293	0.0036
9	0.996	0.5293	1.000	0.0083

由以上 BP 神经网络的训练过程可知,网络的结构通过自身的自组织学习,调整出了自己的结构,最后得出了一个大小合适的优化模型结构,免去了人为因素的干预。

经过 BP 神经网络的动态全参数自调整学习过,我们得到的模型的连接权矩阵如下:

$$W = \begin{bmatrix} -0.6966 & -0.3256 & 0.2342 & -0.4221 \\ 0.1532 & 0.8888 & -0.2309 & 0.2885 \end{bmatrix}$$

$$V = [-1.65564 \quad -1.54538 \quad -0.48602]$$

表4 BP神经网络动态全参数自调整算法与 GMDH 方法预测结果及检验

Tab.4 The forecast values, achieved by the improvement of BP algorithm and GMDH

年份		1995	1996	1997	1998
能源消费量实际值		1249.11	1268.27	1298.69	1311.10
BP 动态全参数自调整算法	预测值	1229.87	1255.18	1265.31	1286.51
	相对误差	-1.54%	-1.03%	-2.56%	-1.88%
GMDH 方法	预测值	1214.54	1287.92	1340.67	1411.42
	相对误差	-2.76%	1.56%	3.24%	7.65%

从表4显而易见,采用 BP 神经网络动态全参数自调整算法所得的预测精度较 GMDH 方法高,其最大相对误差不超过 2.56%,预测结果比较可靠。

## 5 结束语

从减少人为因素干预的思想出发,改进了 BP 算法,提高了神经网络的学习和适应能力,并对其在经济领域的预测作了应用上的研究,得到了较好的结果。

## 参考文献:

- [1] Kandel E R and Schwartz J. Principles of neural science[M]. Elsevier, 1985. 1~49.
- [2] Lippmann R P. An Introduction to computing with neural nets [J]. IEEE ASSP Magazine, 1987, April: 12~35.
- [3] Cyberko G. Approximations by superpositions of a sigmoidal

## 4.1.2 模型的应用——预测

用上面的模型,我们对 1995~1998 年我国能源消费量进行预测。得到预测结果如表 4 所示。

## 4.2 与自组织方法(GMDH)的对比

自组织理论与方法<sup>[4]</sup>自七十年代初诞生至今已近三十年的时间了,正逐步趋于成熟。在对一些复杂的经济问题的研究过程中,取得了令人满意的效果。为了和 BP 神经网络的动态全参数自调整算法的预测效果比较,我们同样采用搜集的样本数据,应用 GMDH 方法对 1995~1998 年我国的能源消费量这一经济指标进行预测,得到的模型方程为:

$SCEY = 0.00589 + 0.12290x_1 - 0.00945x_2 + 0.47953x_3$   
其中 SCEY 表示能源消费量实际值,  $x_1$  表示国民生产总值指数,  $x_2$  表示广义节能率,  $x_3$  表示能源产品出厂价格综合指数。

由此模型中得到的预测结果见表 4 所示。

function[J]. Math Control Singnal System, 1989. 34~89.

- [4] Hema Rao and Alexey G I. Inductive leaning algorithms for complex system modeling[M]. CRC Press, Inc, 1994.
- [5] 张立明. 人工神经网络的模型及其应用[M]. 上海: 复旦大学出版社, 1995. (3): 1~92.
- [6] 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. 30~42.
- [7] 雷鸣, 伊申明, 杨叔子. 神经网络自适应学习研究[J]. 系统工程与电子技术, 1994, (3): 29~36.
- [8] 李卓立. 实用经济计量学[M]. 北京: 清华大学出版社, 1983. 10~32.
- [9] 彭建良, 李新建, 王斌, 等. 能源消费量模拟分析和预测的神经网络方法[J]. 系统理论与实践, 1998, (7): 76~84.
- [10] 罗小增, 王仁华. 学习常数自适应的反向传播算法[A]. 中国神经网络首届学术大会论文集[C]. 北京: 科技出版社, 1990. 60~65.

(编辑 杨 蓓)