

# 一种新型双权值人工神经网络的数据拟合研究

曹 宇, 赵星涛

(中国科学院半导体研究所神经网络组, 北京 100083)

摘 要: 在本文中提出了一种针对新型双权值神经网络的数据拟合算法. 采用这种新型网络结构和算法, 可以克服传统的通用前馈网络中 BP 算法易陷入局部极小的问题. 通过实验比较证明在相同的网络规模下, 采用这种新型网络结构和算法可以取得比径向基(RBF)网络更高的拟合精度和更少的迭代次数.

关键词: 双权值; 神经网络; 数据拟合

中图分类号: TP39114 文献标识码: A 文章编号: 0372-2112 (2004) 10-1672-03

## Data Fitting Based on a New Double Weights Neural Network

CAO Yu, ZHAO Xingtao

(Lab of Artificial Neural Networks, Institute of Semiconductors, CAS, Beijing 100083, China)

Abstract: We construct a new method in data fitting fields. Usually, in traditional BP neural network model, data fitting may become trapped at a local minimum. By using the new Double Weights Model, this algorithm can give the Direction Weight, also the Central Weight at the same time. Experimental results show that this algorithm can get more accurate fitting effects and use less generations to calculate, compared with the RBF (Radial Basis Functions) while using the same environment and equal network scale. Data fitting on it should be a new method to modern industry applications in data control and analyses and so on.

Key words: neural network; double weights; data fitting

### 1 引言

使用神经网络进行数据拟合是人工神经网络的一个重要应用. 使用神经网络进行模式识别, 一般也是转化为一个数据拟合的问题而对网络进行训练. 传统的通用前馈网络一般是采用 BP 算法进行训练学习. 但是 BP 算法容易陷入局部极小的问题而使网络的学习效果达不到预定的要求. 径向基函数(RBF)网络是一种仿照生物的神经细胞的局部感受特性而设计的一种网络模型. 一个 RBF 网络中的神经元是一个具有局部特性的非线性函数. 理论分析证明, 只要有足够多的 RBF 神经元, 一个 RBF 网络可以任意精度的逼近一个非线性函数.

文献[1]中提出了一种新型双权值人工神经网络模型. 在这种新型的网络结构下, 传统的通用前馈网络和径向基函数(RBF)网络都是它的一种特例. 文献[1, 2]中提出了神经网络多维空间复杂几何形体覆盖方法, 为人工神经网络在模式识别方面的应用开辟了一个新的研究方向. 同时在文献[1]中指出采用具有封闭超曲面特性的高阶神经元在模式识别应用中的优良特性. 一个 RBF 神经元就是一个具有封闭超曲面特性的各项同性的 2 阶神经元. RBF 网络在数据拟合方面比传统的通用前馈网络有着更优良的特性. 这就证实了具有封闭超曲面特性的高阶神经元在数据拟合应用中的广阔前景.

由于传统的通用前馈网络和径向基函数(RBF)网络都是双权值人工神经网络的一种特例, 因此关于这两种网络的拟合算法只是调整了双权值人工神经网络模型的部分参数. 在本文中提出了一种针对这种新型双权值神经网络的数据拟合算法. 它采用的是一种具有封闭超曲面特性的 2 阶各项异性高斯型神经元. 采用这种新型网络结构和算法, 可以克服传统的通用前馈网络中 BP 算法易陷入局部极小的问题, 并且在相同的网络规模下能够取得比径向基(RBF)网络更高的拟合精度和更少的迭代次数.

### 2 双权值神经网络

在神经网络连中, 每一个神经元完成了输入矢量与一个权值矢量的计算作用. 在文献[1]中提出了一种新型双权值人工神经网络模型, 每一个双权值神经元可以完成输入矢量与两个权值矢量的计算作用. 双权值神经元的基本计算式为:

$$Y = f \left[ \sum_{j=0}^M \left( \frac{W_j (X_j - W_{c_j})}{|W_j (X_j - W_{c_j})|} \right)^S |W_j (X_j - W_{c_j})|^{p-H} \right] \quad (1)$$

其中 Y 为神经元的输出, f 为神经元激励函数, H 为神经元的阈值,  $W_j$  为由第 j 个输入端接至神经元的方向权值,  $W_{c_j}$  为由第 j 个输入端接至神经元的核心权值,  $X_j$  为第 j 个输入端(正值), M 为输入空间维数, S 为决定单项正负号方法的参数, S = 0 时单项符号永为正, S = 1 时单项的符号与  $W_j$  的符号相

同,  $p$  为幂参数.

这种网络模型是一种通用的神经网络模型. 如果核心权值  $W_{cj}$  全为 0,  $S = 1, p = 1$ , 则式 (1) 就是一个经典的神经网络数学模型; 如果核心权值  $W_{cj}$  全为 1,  $S = 0, p = 2$ , 则式 (1) 是一个径向基函数 (RBF) 网络的神经元. 如果设定  $S = 0$ , 则该基本计算式是一个高维空间中的封闭超曲面神经元. 从高维空间的角度来考虑, BP 网络中的神经元代表一个超平面, RBF 神经元代表一个超球面, 而双权值神经元在高维空间中可以构造出形状各异的封闭超曲面. 以三维空间中的图示为例, 双权值神经元所构造出的封闭超曲面可如图 1 所示.

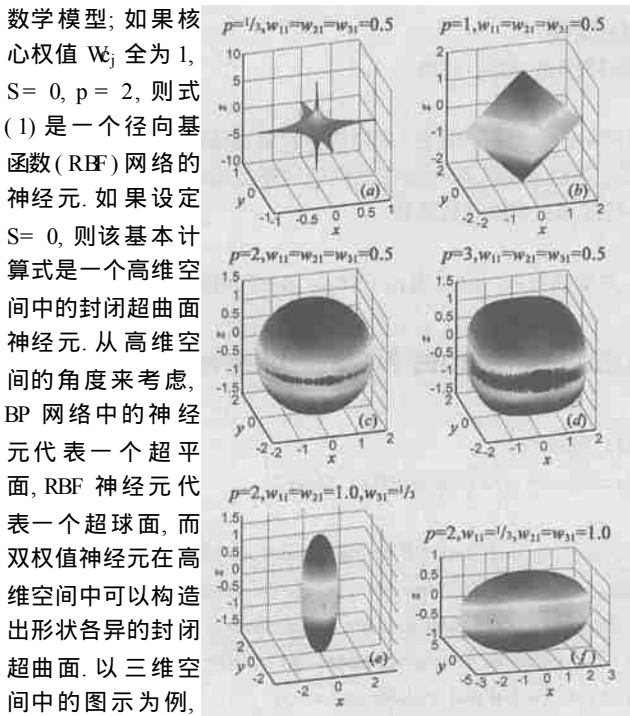


图 1 双权值神经元在三维空间中所构造出的封闭超曲面示例

### 3 双权值神经网络的数据拟合

在高维空间中要拟合的曲面一般是由多个峰谷组合而成的, 对于每一个峰(谷)可以采用一个各向异性的高斯曲面去近似它. 原始样本减去该高斯曲面在相应各点的值之后得到一组新的样本. 这个拟合后的误差曲面仍然是由许多峰谷组合而成的, 可以再用另一个各向异性的高斯曲面去拟合其中的一个峰(谷). 这样通过不断的拟合误差曲面的峰(谷), 直到总体误差小于一定阈值为止, 这就是本算法的基本思想.

在式 (1) 中取  $S = 0, p = 2, f(x) = e^{-x}, H = 0$  得

$$Y = A \exp \left[ - \sum_{j=0}^M W_j^2 (X_j - W_{cj})^2 \right] \quad (2)$$

的形式. 可以看出这种特殊的神经元代表了一个各向异性的高斯曲面. 这种神经元同 RBF 神经网络的形式一样具有局部特性, 但比后者多了一组可以调节各个方向的拉伸长度的方向权值  $W_j$ . 为了计算简单, 将式 (2) 改写成为

$$Y = A \exp \left[ - \frac{1}{M} \sum_{j=0}^M W_j^2 (X_j - W_{cj})^2 \right] \quad (3)$$

这样算法最终拟合出的神经网络的计算式为

$$f(\vec{X}) = \sum_{k=1}^K A_k \exp \left[ - \frac{1}{M} \sum_{j=0}^M W_{jk}^2 (X_j - W_{cj})^2 \right] \quad (4)$$

在本算法中采用了神经网络的构造方法, 每次采用最速下降法确定一个神经元的参数 (即一个高斯曲面), 然后对每个样本点的函数值减去拟合出的高斯曲面在该点处的函数

值, 得到一组新的样本, 对新的样本再通过不断添加新的神经元来构造出符合误差要求的神经网络来. 设对第  $k$  个神经元拟合前的样本数据为  $\{(X_i, y_i^{(k)})\}_{i=1}^N$ , 拟合后的样本数据为  $\{(X_i, y_i^{(k+1)})\}_{i=1}^N$ , 该神经元的核心权值为  $W_{cj}$ , 方向权值为  $W_{jk}$ , 幅值为  $A_k$ , 则第  $k$  个神经元所代表的高斯函数  $f^{(k)}(X)$  为:

$$f^{(k)}(X) = A_k \exp \left[ - \frac{1}{M} \sum_{j=0}^M W_{jk}^2 (X_j - W_{cj})^2 \right] \quad (5)$$

有

$$y_i^{(k+1)} = y_i^{(k)} - f^{(k)}(X_i), \quad i = 1, \dots, N \quad (6)$$

该神经元拟合后的均方差为

$$E^{(k)} = \frac{1}{N} \sum_{i=1}^N (y_i^{(k+1)})^2 = \frac{1}{N} \sum_{i=1}^N (y_i^{(k)} - f^{(k)}(X_i))^2 = \frac{1}{N} \sum_{i=1}^N \left\{ y_i^{(k)} - A_k \exp \left[ - \frac{1}{M} \sum_{j=0}^M W_{jk}^2 (X_{ji} - W_{cj})^2 \right] \right\}^2 \quad (7)$$

该均方差对神经元的核心权值  $W_{cj}$ , 方向权值  $W_{jk}$ , 幅值  $A_k$  分别求偏导:

$$\frac{\partial E^{(k)}}{\partial W_{cj}} = \frac{2}{N} \sum_{i=1}^N \left\{ (y_i^{(k)} - f^{(k)}(X_i)) f^{(k)}(X_i) \left[ \frac{2}{M} W_{jk}^2 (X_{ji} - W_{cj}) \right] \right\} \quad (8)$$

$$\frac{\partial E^{(k)}}{\partial W_{jk}} = \frac{2}{N} \sum_{i=1}^N \left\{ (y_i^{(k)} - f^{(k)}(X_i)) f^{(k)}(X_i) \left[ - \frac{2}{M} W_{jk} (X_{ji} - W_{cj})^2 \right] \right\} \quad (9)$$

$$\frac{\partial E^{(k)}}{\partial A_k} = \frac{2}{N} \sum_{i=1}^N \left\{ (y_i^{(k)} - f^{(k)}(X_i)) \exp \left[ - \frac{1}{M} \sum_{j=0}^M W_{jk}^2 (X_{ji} - W_{cj})^2 \right] \right\} \quad (10)$$

给定初始核心权值  $W_{cj}^{(0)}$ 、方向权值  $W_{jk}^{(0)}$ 、幅值  $A_k^{(0)}$ , 分别加一个正比于梯度负值的变化量进行迭代, 即:

$$W_{jk}^{(m+1)} = W_{jk}^{(m)} - L_W \frac{\partial E^{(k)}}{\partial W_{jk}} \Big|_{W_{jk} = W_{jk}^{(m)}} \quad (11)$$

$$W_{cj}^{(m+1)} = W_{cj}^{(m)} - L_W \frac{\partial E^{(k)}}{\partial W_{cj}} \Big|_{W_{cj} = W_{cj}^{(m)}} \quad (12)$$

$$A_k^{(m+1)} = A_k^{(m)} - L_A \frac{\partial E^{(k)}}{\partial A_k} \Big|_{A_k = A_k^{(m)}} \quad (13)$$

逐步求得  $E^{(k)}$  的极小值点.

在每次拟合一个神经元时, 优先拟合样本点中幅值最大的峰(或谷). 设样本点中幅值最大的点为  $n$ . 设定样本点  $n$  的横坐标  $X_n$  作为初始核心权值  $W_{cj}^{(0)}$ , 该点的纵坐标  $y_n$  作为初始幅值  $A_k^{(0)}$ . 设该峰(或谷)中包含的最大的超球的半径为  $r$ , 则取初始方向权值  $W_{jk}^{(0)} = \sqrt{M \ln 100} / r$ . 即初始化该神经元为一个各向同性的高斯曲面, 该高斯曲面的中心为  $X_n$ , 高度为  $y_n$ , 在距离中心点  $r$  处该高斯函数衰减为  $\frac{1}{100} y_n$ .

在传统的神经网络数据拟合中一般是事先定好网络规模, 然后调节所有神经元的权值及阈值. 这样的算法一般是采用最速下降法求误差曲面的最小值点, 但训练过程非常容易陷入局部极小点, 并且很难跳出这个局部极小值点. 采用模拟退火的方法, 理论上在经过足够多的迭代后网络一定能够收敛到全局最小值点, 但在实际使用中往往运算量太大, 网络训练速度非常慢. 而在本文提出的算法中, 局部极小值点带来的

问题并不严重,即使当前神经元在迭代过程中陷入了局部极小值点,也可以通过增加新的神经元来补偿单次迭代中陷入局部极小值所带来的影响,因此在整个算法的运行过程中不存在不收敛的情况.随着神经元的个数的增加,误差逐步减小,只要有足够多的神经元就一定能够将拟合均方误差降低到所需要的水平.

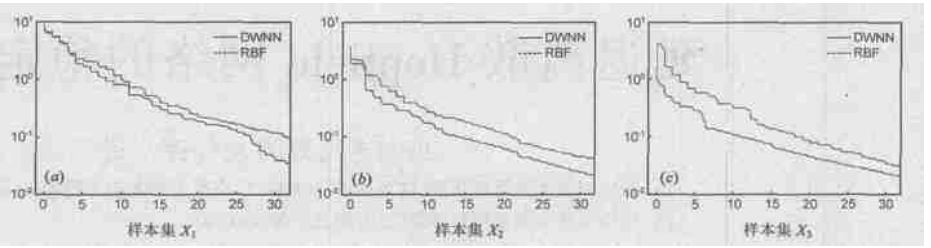


图 2 均方误差的变化曲线  
率  $L_A = 1, L_W = 3, L_{W_c} = 1$ . 在 4 中所述算法中设定  $W_{0k} = W_{1k} = \dots = W_{Mk} = W_k$ , 用来作为 RBF 网络的数据拟合算法. 图 2 为拟合过程中均方误差的变化曲线, 其中横坐标代表神经元的标号. 在  $(k-1, k]$  间的曲线为第  $k$  个神经元拟合过程中的均方误差的局部变化曲线. 表 1 中列出了对于三个样本集  $X_1, X_2$  和  $X_3$ , 分别采用 DWNN 与 RBF 网络各取 32 个神经元进行拟合后的均方误差以及总的迭代次数.

#### 4 算法流程

- (1) 令  $k = 1$ . 取本次拟合前的样本数据  $y_i^{(1)} = y_i, i = 1, 2, \dots, N$ .
- (2) 初始化核心权值  $W_{jk}^{(0)}$ 、方向权值  $W_{jk}^{(0)}$  和幅值  $A_k^{(0)}$ .  
设样本中  $y_i^{(k)}$  的绝对值最大的点下标为  $n$ , 则设置  $W_{jk}^{(0)} = X_{jn}^{(k)}, A_k^{(0)} = y_n^{(k)}$ . 令  $K_1 = \min_{\substack{\text{所有} \\ \text{使得 } y_i^{(k)} < 0}} X_i - X_n +, K_2 = \max_{\substack{\text{所有} \\ \text{使得 } y_i^{(k)} < 0}} X_i - X_n +$ , 取  $W_{jk}^0 = \frac{1}{2}(K_1 + K_2)$ .
- (3) 利用(11)、(12)、(13)式进行迭代, 搜索均方误差  $E^{(k)}$  的极小值点, 保存该点处的最优核心权值  $W_{jk}$ 、方向权值  $W_{jk}$  和幅值  $A_k$ .
- (4) 利用(6)式计算拟合后的样本数据  $\{(X_i, y_i^{(k+1)})\}_{i=1}^N$ . 利用式(7)计算该神经元拟合后的均方误差  $E^{(k+1)}$ . 若  $E^{(k+1)} >$  预定阈值, 则令  $k = k + 1$ , 返回(2); 否则算法结束.

#### 5 实验结果

构造以下三个样本集:

- (1) 二维指数衰减余弦函数  $f_1(x_1, x_2) = e^{-\frac{(x_1^2 + x_2^2)}{32}} \cos \sqrt{x_1^2 + x_2^2}$ . 对  $f_1(x_1, x_2)$  随机选取  $(x_1, x_2) \in (-3P, 3P) \times (-3P, 3P)$  中 100 个样本点, 构造样本集  $X_1$ .
- (2) 多维空间中的多峰谷函数的拟合. 在 10 维空间中随机设定 6 个峰谷, 采用各向异性的高斯函数叠加的方法构造一个最大值为 1 的函数  $f_2(x) = f_2(x_1, \dots, x_{10})$ . 在区域  $x \in R^{(10, -3P, 3P)}$  中随机选取  $f_2(x)$  的 100 个样本点, 构造样本集  $X_2$ .
- (3) 与(2)相同, 构造一个 10 空间中的最大值为 1 的 6 峰谷函数  $f_3(x)$ . 构造带噪声函数  $f_3(x) = f_3(x) + n(x)$ , 其中  $n(x)$  是一个均值为 0, 方差为  $R^2 = 0.01^2$  的白高斯噪声. 在区域  $x \in R^{(10, -3P, 3P)}$  中随机选取  $f_3(x)$  的 100 个样本点, 构造样本集  $X_3$ .

表 1 DWNN 与 RBF 拟合结果比较

样本集	网络类型	32 个神经元的拟合均方误差	总迭代次数
$X_1$	DWNN	0.0359	20415
	RBF	0.0992	23448
$X_2$	DWNN	0.0206	11757
	RBF	0.0402	20542
$X_3$	DWNN	0.0205	5617
	RBF	0.0312	16033

#### 6 讨论

在本文中提出了一种针对新型双权值神经网络的数据拟合算法. 采用这种新型网络结构和算法, 可以克服传统的通用前馈网络中 BP 算法易陷入局部极小的问题. 通过实验比较证明在相同的网络规模下, 采用这种新型网络结构和算法可以取得比径向基(RBF)网络更高的拟合精度和更少的迭代次数.

#### 参考文献:

- [1] 王守觉, 李兆洲, 陈向东, 王柏南. 通用神经网络硬件中神经元基本数学模型的讨论[J]. 电子学报, 2001, 29(5): 577- 580.
- [2] 王守觉, 王柏南. 人工神经网络的多维空间几何分析及其理论[J]. 电子学报, 2002, 30(1): 1- 4.
- [3] 王守觉, 等. 通用前馈网络及排序学习前向掩蔽模型在模式识别中的应用[J]. 电子学报, 1998: 26(8): 1- 6.
- [4] Wang Shoujue, etc. The sequential learning ahead masking (SLAM) model of neural networks for pattern classification[A]. Proceedings of ICIS 98, RTP, North Carolina, USA: October 2328, 1998, IV. 199- 202.
- [5] Wang Shoujue. Priority ordered neural networks with better similarity to human knowledge representation [J]. Chinese Journal of Electronics, Jan. 1999, 8 (1): 1- 4.
- [6] Wang Shoujue, etc. Priority ordered architecture of neural networks [A]. # 58 Session: 4. 2, Proceedings of IJCNN 99[C]. Washington, DC, USA. (Electronic Version), July 10- 16, 1999.
- [7] Wang Shoujue, etc. Direction2basis2function neural networks [A]. # 89 Session: 5. 4, Proceedings of IJCNN 99[C]. Washington, DC, USA. (Electronic Version), July 10- 16, 1999.

#### 作者简介:

曹 宇 男, 1976 年出生, 1999 年毕业于清华大学物理系, 同时进入中科院半导体研究所攻读工学硕士学位, 主攻电路与系统专业神经网络方向, 主要课题是双权值神经网络在数据拟合上的应用.  
赵星涛 男, 1975 年出生, 1997 年毕业于山东大学电子工程系, 获工学学士学位, 于 1999 年在中国科学院半导体研究所师从王守觉院士攻读硕士、博士学位, 主要从事人工神经网络模型与算法方面的研究.

采用 4 中所述算法进行数据拟合实验(其中设定学习速