

文章编号:1001-9081(2008)08-2077-04

覆盖框架与主方向神经网络

殷维栋

(中国科学院 半导体研究所,北京 100083)

(wdyin1004@semi.ac.cn)

摘要:分析了 BP、RBF 和 ARTMAP 等人工神经网络在实现非线性映射方面的共同之处,基于 RBF 等网络对于人脑功能方面的模拟和仿生模式识别的思想,总结出一种处理这类问题的基本框架。该框架的特点是将问题分解为样本覆盖问题和基于模型的映射拟合问题。在利用该框架研究某个函数集在连续函数空间中的稠密性的基础上,提出了一种新的人工神经网络模型——主方向神经网络(PDNN)。通过与 BP 网络和 RBF 网络在函数拟合和混沌时间序列预测方面的对比实验,发现 PDNN 具有非常良好的逼近性能和鲁棒性能。

关键词:RBF 神经网络;仿生模式识别;覆盖框架;马氏距离;主方向

中图分类号: TP183 文献标志码:A

Covering framework and primary direction neural network

YIN Wei-dong

(Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China)

Abstract: Some widely used artificial neural networks such as BP, RBF and ARTMAP were discussed on the common points they shared when solving nonlinear regression problems. A framework, proposed accordingly, suggested that such problems should be implemented in two steps: sample covering and model-based map regression. This idea stems from two aspects, i.e. simulation of human brain's information processing and Biometric Pattern Recognition (BPR). As an instance, a new network named as Primary Direction Neural Network (PDNN) was proposed on the ground of the density of certain function set in continuous function space. It is used to solve two kinds of problems, function regression and stochastic time series prediction. Compared with BP and RBF neural network, PDNN has preferable approximation and robust performance.

Key words: RBF neural network; biometric pattern recognition; covering framework; Mahalanobis distance; primary direction

0 引言

机器学习研究的问题是怎样让机器进行学习,从而具备智能。人工神经网络是实现机器学习强有力的手段,原因在于人工神经网络研究的基本出发点是模仿生物神经网络。人工神经网络由大量神经元互连而成,单个神经元的功能可以非常简单,但是它们之间复杂连接而成的网络却具有极强的非线性映射、分布式存储、并行处理和容错能力,这些是实现智能的基础。当前得到广泛应用的网络有 BP 网络、RBF 网络、Hopfield 网络、SOM 网络、LVQ 网络、ART 系列的网络 (ART1^[1]、ART2^[2]、ART3^[3]、ARTMAP^[4]、FAM^[5] 等)。国内也有一批学者在神经网络模型方面做出了贡献,比如王守觉院士提出的基于高维空间几何分析理论的神经网络模型^[6-8]。

通过分析 RBF 等网络,借鉴仿生模式识别和高维空间复杂形体覆盖的思想,我们认为,用神经网络作非线性映射可以分两个阶段考虑:1) 样本所在区域的覆盖表示;2) 基于这一表示的映射建模。

第一个阶段依赖于样本点在高维空间中的几何分布,一些常用的网络,如 BP 网络以超平面划分的形式表示样本所在的区域,RBF 网络以超球面表示,FAM 则是采用一些矩形来表示^[9],仿生模式识别以超香肠表示^[8,10]。

第二阶段是映射建模,不同的神经网络具有不同的建模方法,如 BP 网络采用基于 Sigmoid 基函数的线性组合来构建映射模型,而 RBF 网络则是基于径向基函数的线性组合来构建模型。

基于上面的分析,我们总结得到处理这类问题的统一框架,称之为覆盖框架,如图 1 所示。

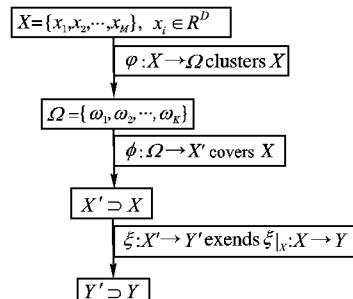


图 1 覆盖框架的示意图

这里作一些必要的说明:映射 φ 由样本数据的几何分布决定,得到该映射的过程是无监督的,称映射 φ 为样本的自然聚类。映射 ϕ 是建立在自然聚类基础上的覆盖映射,实现映射 ϕ 的实质就是对样本数据集所在的几何区域进行表示,我们称映射 ϕ 为表示映射。接下来需要确立输入空间到输出空间的映射关系。映射 $\xi: X' \rightarrow Y'$ 是神经网络的映射模型,不同

神经网络的这个模型假定是不同的。 $\xi: X' \rightarrow Y'$ 应该在 X' 的子集 X , 即样本集上满足映射要求, 我们依据 ξ 在数据集上的限制 $\xi|_X: X \rightarrow Y$ 确定模型参数, 从而确定了映射 ξ 。神经网络的泛化能力体现在 $X' \supset X$ 上, ξ 能够对训练样本外的数据实现映射。

1 理论知识

1.1 一类函数集的稠密性

设 $C(R^s)$ 是 R^s 上的连续函数全体构成集合, 它关于数乘和加法运算构成一个线性空间, 定义连续函数间的乘法运算后构成一个代数^[18]。为简化符号起见, 该代数依然记为 $C(R^s)$ 。

考虑 $C(R^s)$ 的一个子集 $G = \{x \mapsto \|x - c\|^r : c \in R^s\}$, 这里 $r \geq 1$ 是一个常数。

定义 G 的非负张成集合如下:

$$H = \left\{ \sum_{i=1}^m \lambda_i g_i : m \in N, \lambda_i \geq 0, g_i \in G, i = 1, 2, \dots, m \right\} \subset C(R^s)$$

设函数 $\varphi(x) = \exp(-x), x \geq 0$, 考虑下列函数集:

$$F = \{x \mapsto \varphi \circ h(x) : h \in H\} \subset C(R^s)$$

我们希望 F 成为 $C(R^s)$ 中的基本集^[16], 事实上确实如此。

稠密定理 $F = \{x \mapsto \varphi \circ h(x) : h \in H\} \subset C(R^s)$ 是 $C(R^s)$ 中的基本集。

证明 设 $A = sp(F) = \{\sum_{j=1}^n \gamma_j \varphi \circ h_j : n \in N, \gamma_j \in R, h_j \in H, j = 1, 2, \dots, n\}$

首先说明 $A = sp(F)$ 是 $C(R^s)$ 中的一个子代数, 事实上, $\forall h_1, h_2 \in H, \forall x \in R^s$

$$((\varphi \circ h_1) \cdot (\varphi \circ h_2))(x) = (\varphi \circ h_1(x)) \cdot$$

$$(\varphi \circ h_2(x)) = \varphi \circ (h_1 + h_2)(x)$$

$$\text{i.e. } (\varphi \circ h_1) \cdot (\varphi \circ h_2) = \varphi \circ (h_1 + h_2)$$

$$h_1 + h_2 \in H, \text{ which follows } \varphi \circ (h_1 + h_2) \in F$$

其次注意到 $h = 0 \Rightarrow \varphi \circ h \equiv 1$, 即 $A = sp(F)$ 包含常数函数。

此外, 代数 A 中的元素可以区分 R^s 中的点, 事实上,

$$\forall x_1, x_2 \in R^s, x_1 \neq x_2, h(x) = \|x - x_1\|^r$$

$$h \in H, \varphi \circ h \in A, \text{ 但 } \varphi \circ h(x_1) \neq \varphi \circ h(x_2)$$

$$(\text{Notice } \varphi \circ h(x_1) = 1, \varphi \circ h(x_2) =$$

$$\exp(-\|x_2 - x_1\|^r) < 1)$$

由 Stone-Weierstrass 定理^[16-17], F 是 $C(R^s)$ 中的基本集。证毕。

这个函数类的稠密性为我们设计的网络提供了理论基础, 从上面函数集的形式很容易构造出我们的网络。

1.2 Mahalanobis 距离

距离是基于变量之间的互相关的, 是一种决定未知样本与一个已知样本集相似程度的十分有用的方法。与欧氏距离不同的是, 它考虑数据集的相关性并且是尺度不变的, 也就是不依赖于测量的尺度, 这一特点是十分有意义的。

两点 x, y 之间关于某协方差矩阵为 Σ 的样本集的 Mahalanobis 距离^[14-15] 定义为:

$$d(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$$

通常 Σ 不是单位矩阵, 特别当其为单位矩阵时, Mahalanobis 距离退化成通常意义上的欧氏距离^[14]。注意, 点 x 到样本中心的距离是 $d(x, \mu) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$, 显然 s 维空间中的马氏等距离面 $d(x, \mu) = \text{const}$ 是一个 $s - 1$ 维的超椭球面。 Σ 的特征值对应的特征向量方向称为主方向, 最大特征值对应的特征向量方向称为第一主方向。

1.3 乘幂法

在算法的实现中遇到的一个问题是如何计算协方差矩阵的最大特征值和相应的特征向量。既可以用奇异值分解 (Singular Value Decomposition, SVD) 的办法, 也可以用矩阵计算中提供的一种迭代算法, 称为乘幂法, 迭代格式如下^[11]:

$$y_k = Au_k, \mu_k = \xi_j^{(k)}, u_{k+1} = y_k / \mu_k, k := k + 1$$

这里 $\xi_j^{(k)}$ 是 y_k 的模最大分量。

结果表明, 只要 u_1 不太糟糕, 序列 $\{\mu_k\}_{k=1}^\infty$ 收敛到最大特征值 λ_1 , 序列 $\{u_k\}_{k=1}^\infty$ 收敛到相应的特征向量 v_1 。

2 主方向神经网络

2.1 模型与结构

本文提出一种基于覆盖框架的神经网络模型, 基本思想是利用各自然聚类第一主方向 (最大方差方向) 上的几个特殊点作为该聚类的代表, 以这些代表点构建的某种几何形体实现对该类样本点的覆盖。

根据稠密定理有:

$$\forall f \in C(R^s), \forall \varepsilon > 0, \exists n \in N, w_j \in R, k_j \in N, \lambda_j^l \geq 0, c_j^l \in R^s$$

使得:

$$\forall x \in R^s, \left| f(x) - \sum_{j=1}^n w_j \exp\left(-\sum_{l=1}^{k_j} \lambda_j^l \|x - c_j^l\|^r\right) \right| < \varepsilon$$

仿照这一形式, 为构建一个可用的网络, 我们要求神经网络实现如下的映射:

$$\Psi(x) = \sum_{j=1}^K w_j \exp\left(-\lambda_j \sum_{l=1}^3 \|x - c_j^l\|^r\right) + w_0$$

可以看出, 这是一个 3 层前馈式网络, 称之为主方向神经网络 (Principal Direction Neural Network, PDNN)。这里, 我们赋予各参量以实际意义: 自然聚类数目 K 是隐含层神经元的数目, w_j 是隐含层第 j 神经元到输出层的连接权值, w_0 为输出层偏置, λ_j 为与第 j 聚类有关的参数, $c_j^l (l = 1, 2, 3)$ 是第 j 聚类的代表点, $r \geq 1$ 是一个可调参数, 它可以控制基本覆盖体的形状。隐含层的激励函数是 $\varphi(x) = \exp(-x) (x \geq 0)$ 。

可以看到, 这种神经网络的结构与经典的径向基函数神经网络是有共同之处的, 而不同之处在于, 径向基函数神经网络应用一个代表点表示一个聚类, 用其超球邻域覆盖样本, 主方向神经网络利用第一主方向上 3 个点表示一个聚类, 用形体 $\sum_{l=1}^3 \|x - c_j^l\|^r \leq const$ 来覆盖第 j 聚类的样本点。当 $r = 2$ 时, 容易证明这是一个超球, 同于径向基函数神经网络的覆盖形体。

2.2 训练算法

依据覆盖框架, 主方向神经网络的训练算法分为两个关键阶段: 第一阶段是实现自然聚类; 第二阶段是监督学习, 获得网络需要存储的各连接权值和输出偏置。

2.2.1 自然聚类

为了更精细地刻画样本数据的分布特征,采用基于马氏距离的 K 均值聚类。设有 K 类,类 j 来自总体 X_j , D 维随机向量 X_j 均值向量为 μ_j , 协方差矩阵为 Σ_j , 根据最大后验准则, 判决函数为^[15]:

$$d_j(z) = \frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j) + \frac{1}{2}\ln(\det(\Sigma_j)) - \ln P(\omega_j) + \frac{1}{2}D\ln(2\pi); j = 1, 2, \dots, K$$

判决方法为:

$$z \in \omega_{j_0} \Leftrightarrow d_{j_0}(z) < d_j(z), \forall j \neq j_0; j \in \{1, 2, \dots, K\}$$

算法 1

- 1) 输入数据集 X , 输入预聚类数 K , 容许误差界 ε 。
- 2) 采用一定的算法,选取初始中心 μ_j , 并设定初始协方差矩阵和初始分布 $\Sigma_j = I, P(\omega_j) = 1/K (j = 1, 2, \dots, K)$ 。
- 3) 对每一个观测 X_i , 计算 $d_j(X_i)$, $\text{index}(i) = \arg \min_j d_j(X_i)$ 。
- 4) 判断是否有空的聚类,若存在空聚类,则剔除该聚类,重新设置初始分布,转向 3)。
- 5) 按 index 重新计算各聚类中心向量和协方差矩阵以及分布。
- 6) 判别新的中心向量与原来中心向量的距离是否都小于容许误差界 ε 。若是,则输出聚类的相关信息,算法结束;否则更新各聚类中心向量和协方差矩阵以及分布,然后转向 3)。

初始中心的选择对于算法的收敛速度和聚类结果有重要影响。Shimon Cohen 和 Nathan Intrator 介绍了一种简便有效的选取初始中心的算法^[12]:

算法 2

- 1) 随机选择一个向量,作为第一个中心。
- 2) 当聚类数小于 K 时,对剩余数据集中的每一个向量,计算与它们距离最近的中心。选取与对应中心距离最大的数据作为下一个中心。

直观上,我们可以选取训练数据中距离最远的两个点的连线上等间距的 K 个点作为初始中心。在算法的实现中我们发现,该方法更为简便可行。

2.2.2 监督学习

根据映射模型 $\Psi(x) = \sum_{j=1}^K w_j \exp(-\lambda_j \sum_{l=1}^3 \|x - c_j^l\|') + w_0$, 需要确定的量有 $c_j^l, \lambda_j, w_j, w_0$ 。各聚类的 3 个代表点 c_j^l 由无监督学习来确定,选取第一主方向 p_j 上与中心 μ_j 成对称分布的两个点,连同 μ_j 一起作为第 j 聚类的 3 个代表点,用方程描述为 $c_j^l = \mu_j + 2(l-2)\sigma_j p_j (l = 1, 2, 3; j = 1, 2, \dots, K)$ 这里 $\sigma_j > 0 (j = 1, 2, \dots, K)$ 为各类样本第一主方向上的标准差, σ_j^2 是 Σ_j 的最大特征值, p_j 为相应的特征向量,我们只对其方向感兴趣,故将其长度定为 1。 (σ_j^2, p_j) 可以用乘幂法^[11]确定,也可以用 Σ_j 的奇异值分解来确定。

参数 λ_j 可以和 w_j 一起由一个最优化问题确定,但该问题不是凸优化问题,求解比较困难,而且存在陷入局部极小的困难。为简化求解,给出确定 λ_j 的经验方法: $\lambda_j = (\frac{\tau}{\sigma_j})^r, j = 1, 2, \dots, K$, 这里 $\tau \geq 0$ 为可调参数,这样就只需要调整好参数 $\tau > 0$, 就可以确定 λ_j 。 τ 越大,基函数越尖锐; τ 越小,基函数

越平坦。

连接权值 w_j 和输出偏置 w_0 由监督学习来确定。设有 M 个训练样本 $(x_i, d_i) \in R^d \times R (i = 1, 2, \dots, M)$ 那么导出如下含有 $K+1$ 个未知数的线性方程组:

$$\sum_{j=1}^K w_j \exp(-\lambda_j \sum_{l=1}^3 \|x_i - c_j^l\|') + w_0 = d_i; i = 1, 2, \dots, M$$

当 $M > K+1$ 时,这是一个超定的线性方程组,通过伪逆法可以求其最小二乘意义下的解。

至此,完成网络的全部训练过程。使用网络时只需要将测试向量代入映射方程中求取映射函数值即可。

3 数值实验

使用该网络解决一些具有广泛意义的问题,例如函数拟合、时间序列预测等。在解决同一问题上采用一些其他常用神经网络如 BP 网络和 RBF 网络,并比较所得的结果。

对 BP 网络和 RBF 网络,直接使用 Matlab 7.0 的神经网络工具箱中的有关函数,而 PDNN 则是用自己编写 Matlab 程序代码实现。本次实验中 PDNN 的可调参数作如下说明: τ 取 $0.01 \sim 0.02$, r 取 $1.2 \sim 1.6$, 而 K 的典型值取为训练数据量的 $1/6 \sim 1/3$ 。

3.1 函数回归实验

函数拟合问题有非常强的应用背景,这类问题是这样描述的:给定一组数据点,要求回归出一个尽可能通过这些点的函数。神经网络强大的非线性逼近能力使其非常适合求解这类问题。

实验将对 4 个数据集进行拟合实验。

数据集 1 是简单的 1D 正弦函数: $y = \sin(12x); 0 \leq x \leq 1$ 。在 $0 \sim 1$ 随机抽样 200 个作为训练集,另随机抽样 200 个作为测试集。

数据集 2 是一个比较复杂的 1D 非线性函数,函数表达式如下:

$$y = 0.4\sin(x^2) + 0.32\sin(e^{x^2}) + 0.48\cos(e^{x^2})$$

其中, $-5 \leq x \leq 5$ 。

图 2 表示该函数的曲线。在 $-5 \sim 5$ 随机抽样 200 个作为训练集,另随机抽样 200 个作为测试集。

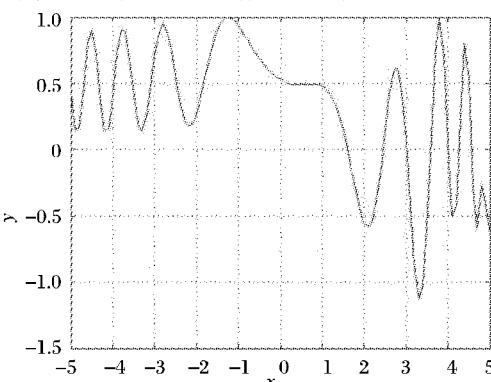


图 2 数据集 2 的函数曲线

数据集 3 是简单的 2D 正弦函数:

$$y = \sin(x_1/4)\sin(x_2/2); 0 \leq x_1 \leq 10, -5 \leq x_2 \leq 5$$

在该范围内随机抽取 400 个样本作为训练集,另外随机抽取 400 个点作为测试集。

数据集 4 是复杂的 2D 非线性函数:

$$y = 0.1 \times [\sin(x_1^2) + 0.8\sin(e^{x_1^2}) + 1.2\cos(e^{x_1^2})] \times [\sin(x_2^2) + 0.8\sin(e^{x_2^2}) + 1.2\cos(e^{x_2^2})]$$

$$1.2 \cos(e^{x_2/1.5})]$$

其中, $-5 \leq x_1 \leq 5$, $-5 \leq x_2 \leq 5$ 。

它的曲面如图 3 所示。这个曲面是十分复杂的,因此,可以将其作为检验神经网络逼近性能的一个基准。在该范围内随机抽取 400 个样本作为训练集,另外随机抽取 400 个点作为测试集。

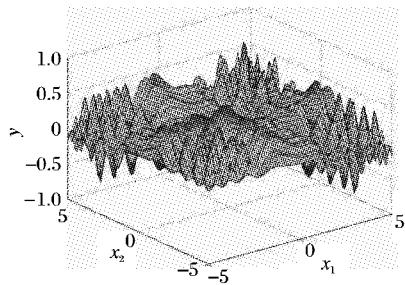


图 3 数据集 4 的函数平面

表 1 是 3 种神经网络算法在 4 个数据集上测试的均方误差(Mean Square Error, MSE)。

表 1 各数据集上测试的均方误差

神经网络	数据集			
	1	2	3	4
BP 网络	3.2000e-3	0.1221	5.8840e-5	0.0257
RBF 网络	4.2357e-4	0.0100	6.5796e-5	0.0442
PDNN	2.4330e-7	5.4653e-3	3.3365e-5	0.0236

分析表 1 不难得出,在函数回归方面,PDNN 的表现优于 BP 网络和 RBF 网络,而 RBF 网络优于 BP 网络。可以说,PDNN 具有良好的非线性逼近能力。

3.2 时间序列预测实验

对 Logistic 映射、Henon 映射这两个时间序列进行预测,这是两个著名的混沌时间序列。Logistic 映射由 $X_{n+1} = 4X_n(1-X_n)$ 给出,初值为 0.1。产生 300 个训练样本,前 280 个用作训练,后 20 个用作测试。而 Henon 映射由 $Y_t = 1 - 1.4Y_{t-1}^2 + 0.3Y_{t-2}$ 给出,初值取为 [0,0],产生 300 个训练样本,前 280 个用作训练,后 20 个用作测试^[13]。

本实验的另一个目的是研究 PDNN 算法对噪声的敏感程度。为此设计了一组对比试验:将 Logistic 和 Henon 时间序列分别作 3 种处理:不加噪声,加 $N(0,0.001)$ 的高斯噪声和加 $N(0,0.01)$ 的高斯噪声,然后对这 3 种情形分别用 3 种网络来进行预测。实验中我们得到了预测的均方误差,结果分别列于表 2、3 中。

表 2 Logistic 序列预测结果

神经网络	不同情况		
	无噪声情况	加入 $N(0,0.001)$ 噪声	加入 $N(0,0.01)$ 噪声
BP	4.5546e-008	0.0071	0.0517
RBF	8.8948e-017	0.0073	1.5648
PDNN	4.4295e-009	0.0075	0.0573

表 3 Henon 预测结果

神经网络	不同情况		
	无噪声情况	加入 $N(0,0.001)$ 噪声	加入 $N(0,0.01)$ 噪声
BP	7.1156e-007	0.0047	0.0441
RBF	2.7109e-013	0.0039	0.0252
PDNN	4.8061e-005	0.0038	0.0282

从表 2、3 可以看出,在混沌时间序列预测方面:

- 1) 在无噪声情况下,RBF 网络的表现很好,优于 PDNN 和 BP 网络。
- 2) PDNN 对噪声的敏感程度与 BP 网络相当,它们都没有 RBF 网络那么敏感于噪声。
- 3) 总体上说,PDNN 的预测性能略好于 BP 网络。

4 结语

本文提出了覆盖框架的概念和思想,在此框架上,从理论分析到数值实验,对主方向神经网络作了较为详尽的说明。该网络既继承了经典的 RBF 网络的 3 层前馈式结构,也吸收了高维仿生信息学的思想。对训练样本的细致描述使主方向神经网络在映射逼近方面性能良好,选取适当的可调参数可以使性能进一步提高。通过实验还发现,主方向神经网络对噪声的敏感度较低,从而具备一定的鲁棒性。这样一些优点使得主方向神经网络可以作为一个可选的算法应用于实际问题中。

当然该网络也存在一定问题,比如可调参数如何选取,隐藏层的神经元数目确定等。由于问题的复杂性,目前没有解析的表达式,因而本文描述的可调参数 τ, r 是经验选取的。类比于在 RBF 网络上的研究方法^[12],可以考虑将可调参数和神经网络的连接权值及偏置一起,由监督学习一并得到,这样就免去了调节参数的乏味劳动。但是这么做会引起如 BP 网络一样的问题,如收敛过慢且易陷入性能函数的局部极小。合理解决这一矛盾具有重要意义,这将是我们下一阶段的研究任务。

参考文献:

- [1] CAPENTER G A, GROSSBERG S. A massive parallel architecture for a self-organizing neural pattern recognition machine[J]. Computer Vision, Graphics and Image Processing, 1987, 37(1): 54–115.
- [2] CAPENTER G A, GROSSBERG S. ART2: Self-organization of stable category recognition codes for analog input patterns[J]. Applied Optics, 1987, 26(23): 4019–4930.
- [3] CAPENTER G A, GROSSBERG S. ART3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures[J]. Neural Networks, 1990, 3(2): 129–152.
- [4] CAPENTER G A, GROSSBERG S. ARTMAP: Supervised real-time learning and classification of non-stationary data by a self-organizing neural network[J]. Neural Networks, 1991, 4(5): 169–181.
- [5] CAPENTER G A, GROSSBERG S, MARKZON N, et al. Fuzzy ARTMAP: An adaptive resonance architecture for incremental learning of analog maps[C]// Proceedings of the International Joint Conference on Neural Networks. [S. l.]: IEEE, 1992: 309–314.
- [6] 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.
- [7] WANG SHOU-JUE, SHI JING-PU, CHEN CHUAN, et al. Direction basis function neural networks [C]// IJCNN'99. Washington, DC, IEEE Press, 1999: 1251–2171.
- [8] WANG SHOU-JUE. Biometric pattern recognition [J]. Acta Electronica Sinica, 2002, 30(10): 1417–1420.

(下转第 2083 页)

表1 XOR问题的实验结果对照表

算法	误差为0.01的迭代步数	迭代5000步的误差
标准BP算法	在10000次内无法实现	0.0253
改进算法1	5371	0.0158
改进算法2	4519	0.0096

实验2 六重XOR分类器问题。在六重XOR分类器问题中,输入变量数为6,输出变量数为1,共有 $2^6=64$ 个输入模式。如果输入全为0或为偶数个1时,对应的希望输出是1,否则希望输出为0。采用6-11-1的结构,初始权值在[-1,1]内随机取值,学习率 $\eta=0.7$,停止训练的误差精度设置为0.001,最大迭代步数为10000。并同时记录迭代2000步的误差。进行30次重复BP网络训练,取其平均值,实验结果如表2所示。

表2 六重XOR分类器问题的实验结果对照表

算法	误差为0.001的迭代步数	迭代2000步的误差
标准BP算法	6374	0.0174
改进算法1	3296	0.0083
改进算法2	2058	0.0012

实验3 函数逼近问题。这里考察单输入单输出的非线性函数:

$$y = \frac{2x}{1+x^2}; \quad x \in [1, 10]$$

的逼近问题。取10组数据作为样本,采用1-4-1的网络结构,初始权值在[-1,1]内随机取值,学习率 $\eta=0.9$,停止训练的误差精度设置为0.01,最大迭代步数为10000。分别用标准BP算法和改进算法逼近上述连续函数,逼近效果如图3和图4。

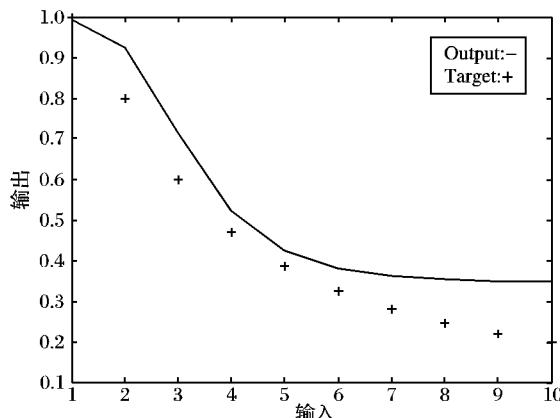


图3 标准BP算法的函数逼近效果

(上接第2080页)

- [9] CAPENTER G A, GROSSBERG S, MARKUZON N, et al. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps[J]. IEEE Transactions on Neural Networks, 1992, 3(5): 698–713.
- [10] 王守觉. 仿生模式识别(拓扑模式识别)——一种模式识别新概念的理论和应用[J]. 电子学报, 2002, 30(10): 1417–1420.
- [11] 徐树方. 矩阵计算中的理论与方法[M]. 北京: 北京大学出版社, 1995: 238–240.
- [12] COHEN S, INTRATOR N. Global optimization of RBF networks [EB/OL]. [2007-08-23]. <http://www.mat.univie.ac.at/~neum/glopt/mss/Cohl01.pdf>.

从实验结果可以看出,两种改进算法比标准BP算法在收敛速度和逼近效果上都有很大的改善,原因在于它们能够放大误差信号,使在不引起误差较大震荡的条件下,尽量加快权值调节,在一定程度上解决了饱和问题。

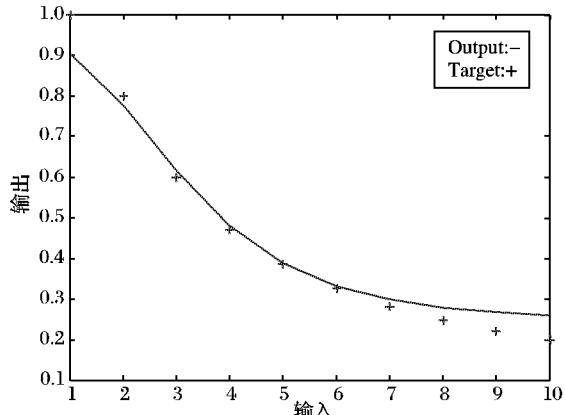


图4 改进算法的函数逼近效果

4 结语

由于饱和区域是导致BP算法收敛速度慢的主要原因之一,本文提出了一种基于误差信号放大的自适应BP算法,并分析了改进算法的收敛性。仿真结果表明,改进的算法在不增加计算量和存储量的前提下,收敛速度和误差精度都要优于标准BP算法。

参考文献:

- [1] HAGAN M T, DEMUTH H B, BEAL M. Neural network design [M]. 北京: 机械工业出版社, 2002.
- [2] 飞思科技产品研发中心. 神经网络理论与Matlab 7实现[M]. 北京: 电子工业出版社, 2005.
- [3] NG S C, CHEUNG C C, LEUNG S H. Magnified gradient function with deterministic weight modification in adaptive learning [J]. IEEE Transactions on Neural Networks, 2004, 15(6): 1411–1423.
- [4] 胡上尉, 刘琼荪, 刘佳璐, 等. 基于修改误差函数新的BP学习算法[J]. 系统仿真学报, 2007, 19(19): 4591–4593.
- [5] 武妍, 王守觉. 一种新的快速收敛的反向传播算法[J]. 同济大学学报: 自然科学版, 2004, 32(8): 1092–1095.
- [6] TAKASE H, KITA H, HAYASHI T. Partially weight minimization approach for fault tolerant multi-layer neural networks[C]// Proceedings of the 2002 International Joint Conference on Neural Networks, Honolulu, Hawaii, USA: ACM Press, 2002: 1092–1096.
- [7] 杨博, 王亚东, 苏小红. 一种基于误差放大的快速BP学习算法[J]. 计算机研究与发展, 2004, 41(5): 774–779.

- [13] 吴金培, 孙德山. 现代数据分析[M]. 北京: 机械工业出版社, 2006: 146–148.
- [14] 蔡元龙. 模式识别[M]. 西安: 西北电讯工程学院出版社, 1986: 19.
- [15] FORSYTH D A, PONCE J. 计算机视觉: 一种现代方法[M]. 北京: 清华大学出版社, 2004: 499–500.
- [16] CHENEY W, LIGHT W. 逼近论教程(英文版)[M]. 北京: 机械工业出版社, 2004: 167.
- [17] RUDIN W. Functional Analysis [M]. New York: McGraw-Hill, 1973: 115.
- [18] 杜珣. 现代数学引论[M]. 北京: 北京大学出版社, 1996: 55.