第 24 卷 第 5 期 电子与信息学报 Vol.24 No.5
2002 年 5 月 JOURNAL OF ELECTRONICS AND INFORMATION TECHNOLOGY May 2002

基于正交校正共轭梯度法的快速神经网络学习算法研究 [1]

郑建国　　刘　芳　　焦李成

(西安电子科技大学雷达信号处理重点实验室　西安　710071)

摘　要　前馈神经网络由于具有理论上逼近任意非线性连续映射的能力，因而非常适合于非线性系统建模及构成自适应控制. 为了提高前馈神经网络的权的学习效率及稳定性，该文提出一种基于正交校正共轭梯度优化方法的快速神经网络学习算法. 通过与其它学习算法 (如：BP 算法、变尺度法、用差商近似代替导数的 Powell 法等) 的比较，经仿真试验表明，本算法是一种高效、快速的学习算法.

关键词　神经网络，学习算法，正交校正共轭梯度

中图号　TN-052

1 引　言

近年来，人工神经网络 (ANN) 正广泛地被应用于模式识别、信号处理与图像处理、人工智能及控制等领域. 在各类神经网络模型中应用最广泛的一类是前馈神经网络，用于训练前馈网络的最常用学习方法是 BP 算法. 然而，在实际应用中发现 BP 算法存在一些不足：主要是收敛速度很慢，往往收敛于局部极小点，数值稳定性差，学习率、动量项系数和初始权值等参数难以调整，为此文献 [1] 提出了牛顿方法的学习算法，文献 [2] 提出了基于共轭梯度的学习算法，文献 [3,4] 各自提出了基于变尺度方法的学习算法，文献 [5] 提出了用差商近似代替导数 Powell 的学习算法. 本文研究了基于正交校正共轭梯度优化方法的快速神经网络学习算法，经仿真试验表明，是一种高效、快速的学习算法.

2 前馈神经网络及学习算法

2.1 前馈神经网络

前馈神经网络由输入层、隐层和输出层组成. 令 $U=[u_1,u_2,\cdots,u_m]^T$ 为网络的输入向量，$Y=[y_1,y_2,\cdots,y_n]^T$ 为网络的输出向量，$X=[x_1,x_2,\cdots,x_N]^T$ 为网络的权及阈值的全体所组成的向量. 给定 P 组输入输出训练样本 $\{(u^{(p)},t^{(p)})|p=1,2,\cdots,P\}$，定义网络的误差指标函数为

$$E(X)=\frac{1}{2P}\sum_{p=1}^{P}\sum_{j=1}^{n}(y_j^{(p)}-t_j^{(p)})^2 \tag{1}$$

把 (1) 式改写为

$$\min F(X)=\sum_{i=1}^{M}f_i^2(X) \tag{2}$$

式中

$$f_i(X)=y_j^{(p)}-t_j^{(p)},\quad M=n\times P,\quad p=[i/n]+1,\quad j=\mathrm{rem}(i/n) \tag{3}$$

[1] 2000-11-03 收到，2001-07-04 定稿

国家自然科学基金 (No. 60073053) 和国家 “863” 计划项目 (批准号：863-306-ZT06-06-1) 资助

然后应用比较成熟的非线性最小二乘法。

2.2 Levenberg-Marquardt 算法 (简称 LM 法)

(1) 给定初始点 $X^{(0)}$ ，精度 $\varepsilon_0, k=0, t_k=10^4$ 。

(2) 对 $i=1,2,\cdots,M$ 求 $f_i(X^{(k)})$ ，得向量 $f(X^{(k)})=[f_1(X^{(k)},\cdots,f_M(X^{(k)})]^T$ ，对 $i=1,2.\cdots,M$ ； $j=1,2,\cdots,N$ 求 $J_{ij}(X^{(k)})=\partial f_i(X^{(k)})/\partial X_j$ ，得 Jacobi 矩阵 $J(X^{(k)})=[J_{ij}(X^{(k)})]$ 。

(3) 解线性方程组 $[J(X^{(k)})^TJ(X^{(k)})+t_kI]d^{(k)}=-J(X^{(k)})^Tf(X^{(k)})$ 求出搜索方向 $d^{(k)}$ 。在解线性方程组时，如果发现矩阵 $[J(X^{(k)})^TJ(X^{(k)})+t_kI]$ 的秩不是 N ，则不解方程组，直接取 $d^{(k)}$ 为负梯度方向，即 $d^{(k)}=-(1/2)\nabla F(X^{(k)})=-J(X^{(k)})^Tf(X^{(k)})$ 。

(4) 直接搜索， $X^{(k+1)}=X^{(k)}=\lambda d^{(k)}$ ，其中 λ_k 满足

$$F(X^{(k)}+\lambda_k d^{(k)})=\min F(X^{(k)}+\lambda d^{(k)})$$

(5) 若 $\|X^{(k+1)}-X^{(k)}\|<\varepsilon$ ，则得到解 $X_{\rm opt}$ ，停止计算；否则转向 (6) 。

(6) 若 $F(X^{(k+1)}<F(X^{(k)})$ ，则令 $t_k=t_k/2$ ， $k=k+1$, 转向 (2) ；否则 $t_k=2t_k$ ，转向 (3)

2.3 用差商近似代替导数的 Powell 法 [5]

该算法与 Levenberg-Marquardt 算法大致相同，只是用点 $X^{(k)}$ 处的差商近似代替导数．因为有时导数不易求出．

2.4 基于正交校正共轭梯度优化方法的快速神经网络学习算法

一般的神经网络采用 BP 算法、牛顿学习算法、基于变尺度方法的学习算法、用差商近似代替导数 Powell 的学习算法等等。但经实验表明 [1-5] ，这些算法收敛速度慢，学习率不易选取，初始训练权值时，稳定性较差，训练的权值只能是局部最优．本文提出的一种基于正交校正共轭梯度优化方法的快速神经网络学习算法，可解决上述问题。

设 $f(X)$ 对 X 的负梯度方向为

$$Z^K=-\nabla f(X)=\sum_{K=1}^{N}(\hat{y}_K-Y_K)\frac{\partial Y_K}{\partial X} \tag{4}$$

网络的权向量修正公式为

$$X^{K+1}=X^K+\lambda_K Z^K \tag{5}$$

其中最佳步长 λ_K 由下式求得

$$\min_{\lambda\geq 0} f(X^K+\lambda Z^K)=f(X^K+\lambda_K Z^K) \tag{6}$$

该算法的具体步骤如下：

(1) 给定初始权值 $X^{(0)}$ ，精度 $\varepsilon_0, k=0$ ， $f_k=f(x^k), g_k=\nabla f(x^k), z_k=-g_k$ 。

(2) $\min_{\lambda\geq 0} f(x^k+\lambda z^k)=f(x^k+\lambda_k z^k)$, 令 $x^{k+1}+\lambda_k z^k$

(3) $g_{k+1}=\nabla f(x^{k+1})$, $z^{k+1}=-g_{k+1}+\beta_k z^k$

其中 $\hat{\beta}_k=(g_{k+1}^Tg_{k+1}/(g_k^Tg_k)$ ， $\gamma_k=(g_k^Tg_{k+1})/(g_{k+1}^Tg_{k+1})$ 。

(a) 若 $|\gamma_k|<\varepsilon$ ， (取 $\varepsilon=0.175$), 则令 $\beta_k=\hat{\beta}_k$

(b) 若 $|\gamma_k|\geq\varepsilon$ ，则令 $\beta_k=\hat{\beta}_k(1-\gamma_k)$

(4) 若满足收敛条件 $\|x^{k+1}-x^k\|<\delta_1$ ， $\|f_{k+1}-f_k\|<\delta_2$, 则停止；否则令 $k=k+1$ 转 (2) 。

3 仿真研究

例 1　函数 y=0.6924exp(-0.7843x)sin(8.2346x+18.4321)，21 输入样本 (−1 到 1 上以 0.1 为间隔的 21 个数)．BP 网络结构为 1-5-1．分别用 BP 算法、DFP 法、LM 法、Powell 法、正交校正共轭梯度法进行了 10 次仿真，每次都是不同的随机的权值，终止误差为 10^{-4}，结果见表 1．

例 2　Rosenbrock 函数，$\min f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ 的仿真结果 (初始值给不同的随机数) 如表 2 所示．

例 3　对于一个多元函数 (初始值给不同的随机数)，求它的极小值 $\min f(x) = 10(x_1 + x_2 - 5)^2 + (x_1 - x_2)^2 + (x_3 - 3)^2 + (x_4 - 4)^2 + (x_5 - 5)^2$ 的仿真结果如表 3 所示．

表 1　各种算法分析比较

算法	迭代次数	不稳定次数	终止误差
BP	606, 448, 693, 735, 1461, 1921, 710, 739	2	10^{-2}
DFP	641, 952, 887, 1509, 2320, 2130, 2157, 3258, 5849, 6311	0	10^{-4}
LM	781, 230, 66, 114, 177, 243, 174, 774	2	10^{-4}
Powell	317, 293, 255, 311, 375, 387, 452, 479, 839, 1363	0	10^{-4}
基于正交校正共轭梯度的神经网络算法	42, 105, 37, 62, 161, 87, 664, 189, 275, 153	0	10^{-4}

表 2　正交校正共轭梯度法仿真结果

初始点 (x_1^0, x_2^0)	迭代次数	最优解 $(x_1, x_2)^*$	$\min f(x_1, x_2)^*$	终止误差
(−7.1588, 8.9284)	31	(1.0000,1.0000)	1.9495×10^{-11}	10^{-5}
(−2.7310, 2.5477)	12	(1.0000,1.0000)	9.2155×10^{-13}	10^{-5}
(−8.6560, 2.3235)	9	(1.0000,1.0000)	1.4956×10^{-12}	10^{-5}
(2.3189, 2.3931)	12	(1.0000,1.0000)	1.8835×10^{-14}	10^{-5}
(−8.4387, 1.7390)	8	(1.0000,1.0000)	7.8037×10^{-13}	10^{-5}
(−1.1975, 0.3813)	22	(1.0000,1.0000)	2.3491×10^{-16}	10^{-5}
(−9.3424, 2.6445)	11	(1.0000,1.0000)	2.1296×10^{-13}	10^{-5}
(−2.3788, 6.4583)	18	(1.0000,1.0000)	9.7419×10^{-13}	10^{-5}
(−9.6689, 6.6493)	46	(1.0000,1.0000)	4.0781×10^{-13}	10^{-5}
(−4.3017, 8.9032)	22	(1.0000,1.0000)	4.8172×10^{-14}	10^{-5}

4 结　论

通过例 1 可以看出，BP 算法收敛速度慢且不稳定，精度难以提高；DFP 法的收敛速度太慢 (迭代次数几千次)、LM 法的不稳定、Powell 法虽然稳定，但收敛速度太慢 (迭代次数从 200~1300 次)，正交校正共轭梯度法的稳定性和收敛速度均优于以上算法．例 2 是对于典型的 Rosenbrock 函数 ($\min f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$) 的仿真，仿真结果为迭代次数均小于 50 次，且精度为 10^{-5}，表明了该算法的快速收敛性．例 3 是对一个一般的五元函数的 $\min f(x) = 10(x_1 + x_2 - 5)^2 + (x_1 - x_2)^2 + (x_3 - 3)^2 + (x_4 - 4)^2 + (x_5 - 5)^2$ 的仿真，仿真结果为迭代次数均少于 15 次，且精度为 10^{-5}，再次表明了该算法的有效性．

表 3 正交校正共轭梯度法仿真结果

初始点 $(x_1^0, x_2^0, x_3^0, x_4^0, x_5^0)$	迭代次数	最优解 $(x_1, x_2, x_3, x_4, x_5)$	min $f(x_1, x_2, x_3, x_4, x_5)$	终止误差
(15.0873, 69.7898, 37.8373, 86.0012, 5.3655)	7	(2.5, 2.5, 3, 4, 5)	3.0896×10^{-12}	10^{-5}
(593.5629, 495.5524, 899.7692, 821.6292, 644.9104)	12	(2.5, 2.5, 3, 4, 5)	4.0963×10^{-12}	10^{-5}
(817.9743, 660.2276, 341.9706, 289.7259, 341.1936)	10	(2.5, 2.5, 3, 4, 5)	1.7735×10^{-11}	10^{-5}
(370.4136, 702.7399, 546.5712, 444.8802, 694.5672)	8	(2.5, 2.5, 3, 4, 5)	1.6541×10^{-12}	10^{-5}
(534.0790, 727.1132, 309.2902, 838.4960, 568.0725)	10	(2.5, 2.5, 3, 4, 5)	1.5877×10^{-14}	10^{-5}
(621.3101, 794.8211, 956.8434, 522.5903, 880.1422)	14	(2.5, 2.5, 3, 4, 5)	5.1523×10^{-13}	10^{-5}
(172.9561, 979.7469, 271.4473, 252.3293, 875.7419)	10	(2.5, 2.5, 3, 4, 5)	9.3293×10^{-12}	10^{-5}
(737.3060, 136.5187, 11.7567, 893.8980, 199.1381)	12	(2.5, 2.5, 3, 4, 5)	1.4198×10^{-11}	10^{-5}
(298.7230, 661.4426, 284.4086, 469.2243, 64.7811)	12	(2.5, 2.5, 3, 4, 5)	1.6228×10^{-12}	10^{-5}
(988.3349, 582.7917, 423.4963, 515.5118, 333.9515)	13	(2.5, 2.5, 3, 4, 5)	1.1665×10^{-12}	10^{-5}

参 考 文 献

[1] 徐嗣鑫，戴友元，前向神经网络的一种快速学习算法及其应用，控制与决策， 1993, 8(4), 284-288.
[2] 王耀南，童调生，蔡自兴，基于神经元网络的智能 PID 控制及应用，信息与控制， 1994, 23(3), 185-189.
[3] 张星昌，前馈神经网络的新学习算法研究及其应用，控制与决策， 1997, 12(3), 213-216.
[4] 王耀南，基于变尺度优化方法的快速神经网络学习算法研究，系统仿真学报， 1997, 9(1), 34-39.
[5] 徐春晖，徐向东，前馈神经网络新学习算法的研究，清华大学学报 (自然科学版)1999, 39(3), 1-3.

STUDY OF FAST LEARNING ALGORITHM FOR NEURAL NETWORKS BASE ON CGM-OC APPROACH

Zheng Jianguo Liu Fang Jiao Licheng

(Key Lab. for Radar Signal Processing, Xidian University, Xi'an 710071, China)

Abstract Because the feedforward neural network has an ability of approach to arbitrary nonlinear mapping, it can be used effectively in the modeling and controlling of nonlinear system. In order to improve the learning efficiency and stability of feedforward neural network, a fast learning algorithm for neural networks base on CGM-OC approach is presented. Compared with other learning methods such as BP method, Broyden Flecher Goldfarl Shanno method, Powell method etc., simulation results show that the proposed method is an efficient and fast methed.

Key words Neural network, Learning algorithm, CGM-OC approach

郑建国： 男， 1962 年生，副教授，博士生，主要研究方向：神经网络，智能信息处理，数据挖掘与知识发现.
刘 芳： 女， 1963 年生，副教授，硕士生导师，主要研究方向：人工智能，数据挖掘.
焦李成： 男， 1959 年生，教授，博士生导师， IEEE 高级会员，主要研究方向：神经网络，智能信息处理，非线性科学等.