文章编号:1000 - 582X(2004)04 - 0118 - 04

基于 GaussNewton-NL2SOL 法的前馈神经网络及应用*

徐　晋
(上海交通大学 管理学院,上海　200030)

摘　要:目前基于高斯牛顿法及其衍生算法的前馈神经网络虽然可以达到局部二阶收敛速度,但只对小残量或零残量问题有效,对大残量问题则收敛很慢甚至不收敛。为了实时解决神经网络学习过程中可能遇到的小残量问题和大残量问题,引入 NL2SOL 优化算法,并与 GaussNewton 法相结合,构建基于 GaussNewton-NL2SOL 法的前馈神经网络。仿真实例表明,该神经网络较好地解决了残量问题,具有良好的收敛性和稳定性。

关键词:前馈神经网络; GaussNewton 法;NL2SOL 法;残量问题;收敛性;稳定性
中图分类号: TP18　　**文献标识码**: A

前馈神经网络(Feed - forward Neural Network)在神经网络理论研究与非线性系统建模及控制中占据非常重要的地位[1]。目前的研究重点主要集中在网络权值学习算法[2]、误差函数[3]、网络结构[4]以及相关的收敛性和稳定性等方面。基于优化理论可以给出许多权值学习算法,包括应用较为广泛的最速下降法,以及近年来提出的基于高斯牛顿法及其衍生算法的权值学习算法。许多文献对基于高斯牛顿法的前馈神经网络进行研究与改进,并取得良好成果[5-6]。值得商榷的是,有些应用和研究不加证明地假定基于高斯牛顿法及其衍生算法的权值学习算法具有全局二阶收敛性。事实上,无论是基于高斯牛顿法还是其衍生算法的权值学习算法,在解决小残量或零残量问题时具有局部二阶收敛速度,而在解决神经网络学习过程中可能出现的大残量问题时,则可能收敛很慢或不收敛而导致网络性能不稳定。这主要因为在高斯牛顿法及其衍生算法中没有充分利用 Hesse 矩阵中的二阶项信息[7]。

笔者尝试在前馈神经网络的权值学习算法中引入 NL2SOL 法[8],用以解决神经网络学习过程中可能遇到的大残量问题,并与 GaussNewton 法相结合,构建基于 GaussNewton - NL2SOL 法的前馈神经网络,最后给出仿真实例。对比实验表明,该神经网络较好地解决了残量问题,具有良好的收敛性和稳定性。

1　前馈神经网络及算法迭代步

1.1　前馈神经网络

一般而言,前馈神经网络由输入层,隐含层和输出层组成。在同层神经元节点之间没有互连;在相临层之间,神经元节点全互连,连接通路上存在权 W。对于给定输入 X,输出 $Y = F(X, W)$,并通过权值的调整,使 Y 与期望值 $\hat{Y}$ 之间的误差符合要求。

设前馈神经网络共有 L 层,$L \geqslant D$。其中,第 $l = 1$ 层为输入层,$l = L$ 层为输出层,其余为隐含层。并设第 l 层共有 n_l 个神经元节点;第 l 层第 i 个神经元与第 $l + 1$ 层第 j 个神经元之间的连接权值为 $W_{ji}(l)$;对第 l 层第 i 个神经元,其输出值为 $y_i(l)$,阈值为 $\theta_i(l)$,目标输出值为 $\hat{y}_i(l)$,误差为 $\delta_i(l)$,输入为 $u_i(l) = \sum_{j=1}^{n_l - 1} y_j(l - 1) W_{ji}(l - 1) + \theta_i(l)$;定义第 l 层第 i 个神经元的节点函数为 $f_{il}(\cdot)$,则其输出值为 $y_i(l) = f_{il}(u_i(l))$。

由于阈值的学习算法与权值学习算法完全一样,故不失一般性地设阈值 $\theta_i(l)$ 为零或定值。给定 P 组训

---

* 收稿日期:2003 - 09 - 15
基金项目:国家自然科学基金资助项目(70172001)
作者简介:徐晋(1976 -),男,江苏涟水人,上海交通大学管理学院博士研究生,香港理工大学访问学者,研究方向:风险投资、复杂性理论。

练样本，定义网络输出误差函数为：$E(W)=\frac{1}{2}\sum_{p=1}^{P}[r_p(W)]^2$，其中 $r_p(W)=Y_p-\hat{Y}_p$。然后就可以根据不同的学习算法训练网络权值，以最小化网络输出误差，即：

$$\min_{W\in R^N}E(W)=\frac{1}{2}\sum_{p=1}^{P}[r_p(W)]^2 \qquad (1)$$

其中，样本数 P 应该不小于权 W 的维数 N，即 $P\geq N$；同时考虑到在一般情况下，前馈神经网络的神经元函数为简单的非线性函数（如 Sigmoid 函数 $f(x)=\frac{1}{1+e^{-x}}$），因此 $r(W)$ 一般为非线性函数。这样，问题(1)就转化为非线性最小二乘问题。

1.2　GaussNewton 迭代步

令 $\boldsymbol{J}(W)$ 为 $r(W)$ 的 Jacobi 矩阵，则 $E(W)$ 的梯度为 $g(W)=J(W)^Tr(W)$，$E(W)$ 的 Hesse 矩阵为 $\boldsymbol{G}(W)=J(W)^T+S(W)$，其中 $S(W)=\sum_{p=1}^{P}r_p(W)\nabla^2r_i(W)$。从而得到网络误差函数 $E(W)$ 的二次型为：

$$\begin{aligned}e_K(W)=&E(W(K))+g(W(K))^T(W-W(K))+\\&\frac{1}{2}(W-W(K))^TG(W(K))(W-W(K))=\\&\frac{1}{2}r(W(K))^tr(W(K))+\\&[J(W(K))^tr(W(K))]^T(W-W(K))+\\&\frac{1}{2}(W-W(K))^t[J(W(K))^TJ(W(K))+\\&S(W(K))](W-W(K))\end{aligned} \qquad (2)$$

其中 $W(K)$ 表示第次迭代后的神经网络权值。

考虑到(2)中 Hesse 矩阵 $\boldsymbol{G}(W(K))$ 中的二阶项 $S(W(K))$ 很难计算，为了简化计算令，$S(W(K))=0$，进而就可以得到解问题(1)的 GaussNewton 法：

$$\begin{aligned}W(K+1)=&W(K)-\\&[J(W(K))^TJ(W(K))]^{-1}J(W(K))^Tr(W(K))=\\&W(K)+s_K\end{aligned} \qquad (3)$$

其中

$s_K=-[J(W(K))^TJ(W(K))]^{-1}J(W(K))r(W(K))$。

这样，基于 GaussNewton 法的神经网络权值学习算法可归纳为：

算法1　(GaussNewton 迭代步)

第 K 次迭代：

1)求解：$[J(W(K))^TJ(W(K))]s=-J(W(K))^Tr(W(K))$，得 s_K；

2)令 $W(K+1)=W(K)+s_K$。

1.3　NL2SOL 迭代步

显然，从 $W(W(K))=\sum_{p=1}^{P}r_p(W(K))\nabla^2r_i(W(K))$ 可知，只有当 $r_p(W(K))$ 接近于零或 $r_p(W)$ 接近线性函数从而 $\nabla^2r_i(W)$ 接近于零时，$S(W(K))$ 才可以忽略为零。对于这类问题，通常成为小残量问题，否则称为大残量问题[7]。高斯牛顿法及其衍生的阻尼高斯牛顿法、Levenberg - Marquard*t* 法等，是解决非线性最小二乘问题的基本方法，这些方法对处理小残量或零残量问题非常有效，对不是很严重的大残量问题有较慢的局部收敛速度，而对于大残量问题往往不收敛或收敛速度很慢。其根本原因在于没有充分利用 Hesse 矩阵中的二阶信息项 $S(W)$。为了充分利用二阶信息项，可以构造 $S(W)$ 的割线近似。同时，考虑到分量 $r_p(W)$ 有时比分量 $\nabla^2r_p(W)$ 变化的更快，故在每次迭代中选取一个因子乘以 B_K 作为调比策略，这就得到解决问题(1)的 NL2SOL 法[8](Denis, Gay 和 Welsch, 1981)。

令 $v_K=J(W(K+1))^Tr(W(K+1))-j(W(K))^Tr(W(K))$，$z_K[J(W(K+1))-J(W(K))]^Tr(W(K+1))$，设 $B(K)$ 为 $S(W(K))$ 的割线近似，则其校正公式为：

$$B(K+1)=B(K)+\frac{(z_K-B(K)s_K)\nu_K^T+\nu_K(z_K-B(K)s_K)^T}{s_K^Tv_K}-\frac{s_K^T(z_K-B(K)s_K)\nu_K\nu_K^T}{(s_K^T\nu_K)^2)}$$

基于 NL2SOL 法的神经网络权值学习算法可归纳为：

算法2(*NL2SOL* 迭代步)

第 K 次迭代：

1)求解：$[J(W(K))^TJ(W(K))+B(K)]s=-J(W(K))^Tr(W(K))$，得 s_K；

2)令 $W(K+1)=W(K)+s_K$；

3)求解 $\mu_K=\min\{\frac{s_K^T\nu_K}{s_K^TB_Ks_K},1\}$，并令 $B(K)=\mu_KB(K)$；

4)校正 $B(K)$ 产生 $B(K+1)$。

在算法2中，$B(K)$ 的初始值 $B(1)$ 一般可以取单位矩阵 $\boldsymbol{I}$。对于大残量问题，NL2SOL 法具有明显的优越性；对小残量问题，则不如 GaussNewton 法。

2　基于 GaussNewton - NL2SOL 法的神经网络学习算法

对前馈神经网络而言，神经元节点一般取非线性

函数,而且根据非线性系统建模及控制的不同情况选取不同的神经元函数,这都将影响 $r_p(W)$ 接近线性函数的程度;同时,由于神网络初始化权值 $W(1)$ 为随机确定,$r_p(W(1))$ 不一定接近于零;在网络优化学习过程中也不一定能保证 $r_p(W(K))$ 接近于零。这些情况说明,在前馈神经网络权值学习过程中,$S(W(K))$ 不应该轻易忽略,它可能在小残量与大残量间摆动。因此,无论是采取 GaussNewton 法,还是 *NL2SOL* 法,都可能存在收敛速度变慢或不收敛以致网络性能不稳定的情况。

为了解决前馈神经网络学习过程中遇到的残量问题,笔者尝试在前馈神经网络的权值学习算法中综合考虑 GaussNewton 法与 NL2SOL 法,即根据每次迭代的结果判断属于大残量问题还是小残量问题,进而选择不同的迭代步。其基本的判断准则为:$E(W(K+1))\leqslant\eta E(W(K))$,一般取 $\eta=0.75$。满足该准则,则认为是小残量问题,采取 GaussNewton 迭代步;否则就认为是大残量问题,采取 NL2SOL 迭代步。

根据以上准则,构建了基于 GaussNewton-NL2SOL 法的前馈神经网络权值学习算法,其具体算法描述如下:

算法 3(基于 GaussNewton-NL2SOL 法的权值学习算法)

1)$K=1$,初始化权值 $W(1)$,设定神经网络误差精度 ε,判断准则系数 η,最大迭代次数 N,并令 $E(W(0))=\varepsilon/\eta$;

2)给定学习样本:输入 X_p,期望输出 $\hat{Y}_p,p=1,2,\cdots,P$;

3)输入 X_p,得到网络实际输出 $Y_p,p=1,2,\cdots,P$;

4) 计算系统误差

$$E(W(K))=\frac{1}{2}\sum_{p=1}^{P}[r_p(W(K))]^2,$$其中 $r_p(W)=Y_p-\hat{Y}_p,p=1,2,\cdots,P$;

5)若 $E(W(K))\leqslant\varepsilon$,输出结果,结束;

6)若 $E(W(K))\leqslant\eta E(W(K-1))$,则采取 GaussNewton 迭代步,得 $W(K+1)$;

7)若 $E(WK))>\eta E(W(K-1))$,则采取 NL2SOL 迭代步,得 $W(K+1)$;

8)$K=K+1$,若 $K\leqslant N$,返回3);若 $K>N$,不收敛,输出结果,结束。

3 仿真实验分析

3.1 实验描述

仿真对象:$f(x)=\exp(\cos(9x+1)\sin(0.5x+1)+0.5\sin(x))$,如图 1 所示。输入样本:$X=[0:0.05:5]$,共有 101 个输入样本点。期望输出:$Y=f(x),x\in X$。网络结构:1-10-1(输入层-隐含层-输出层)。误差范围:$\varepsilon\leqslant10^{-3}$。作对比的三种神经网络算法:1)基于最速下降法的传统 BP 神经网络权值学习算法;2)基于 GaussNewton 法的神经网络权值学习算法;3)基于 GaussNewton-NL2SOL 法的神经网络权值学习算法。

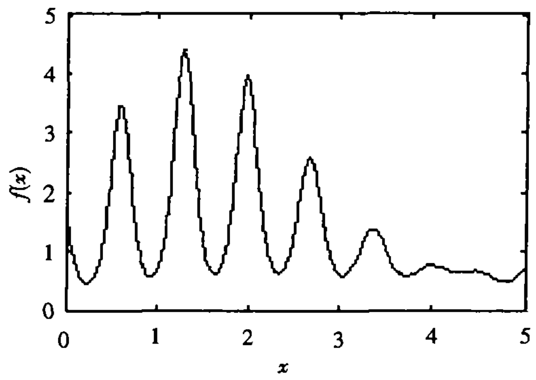

图 1 函数 $f(x)$ 的图象

比较分析的内容:1)收敛性,即网络输出达到指定误差范围的情况下,网络训练所需要的迭代步数;2)稳定性,即 10 次训练中各个算法收敛的情况。

3.2 实验结果

由于神经网络初始权值的选取对其网络性能表现包括收敛性等都有一定的影响,所以为了说明基于 GaussNewton-NL2SOL 法的网络权值学习算法的有效性,笔者连续做 10 组对比实验,结果见表 1。

表 1 网络收敛迭代步数对比

算法	1	2	3	4	5	6	7	8	9	10	平均
传统 BP 算法	1 1542	N	10 439	12 945	12 243	8 815	1 327	1 599	13 645	N	9 096
GaussNewton 算法	1 105	1 329	N	764	697	859	N	N	1 068	649	924
GaussNewton - NL2SOL 算法	615	425	327	392	340	814	363	606	318	463	357

1)收敛性:在神经网络系统达到稳定的情况下,传统 BP 算法的平均迭代步数为 9 096 次;基于 GaussNewton 法的平均迭代步数为 924 次;基于 GaussNewton-NL2SOL 法的权值学习算法平均迭代步数为 357

次。可以看出，基于 GaussNewton-NL2SOL 法的权值学习算法具有较快的收敛速度。

2）稳定性：从表 1 中可以看出，在 10 次仿真训练中，传统 BP 算法与 GaussNewton 法均有 2 次不收敛，而基于 GaussNewton-NL2SOL 法的神经网络权值学习算法则全部收敛。显然，基于 GaussNewton-NL2SOL 法的前馈神经网络具有良好的稳定性。

4　结束语

笔者首次将 NL2SOL 法与 GaussNewton 法同时引入前馈神经网络的权值学习算法中，成功地解决了在网络权值优化训练过程中存在的大残量与小残量问题。仿真实例表明，所提出的基于 GaussNewton-NL2SOL 法的前馈神经网络具有良好的收敛性和稳定性，有一定的推广价值。

参考文献：

[1]　CHEN S, BILLING S A. Neural Network for Nonlinear Dynamic System Modeling and Identification[J]. Int Journal of Control, 1992,56(2):319 - 346.

[2]　MALIK M, AMIR F A. The Early Restart Algorithm[J]. Neural Computation, MIT, 2000,(12):1 303 - 1 312.

[3]　CHANG K S, ABEL GHATTAR. A Universal Neural Net with Guaranteed Convergence to Zero System Error[J]. IEEE Trans. Signal Processing, 1992, 40(12):3 022 - 3 031.

[4]　何述东，瞿坦，黄献青，等. 多层前向神经网络结构的研究进展[J]. 控制理论与应用, 1998,(4):313 - 318.

[5]　李一波，黄小原，吴志红. 修正高斯模型神经网络的色谱重叠峰解析[J]. 计算机与应用化学, 2001,(9), 484 - 488.

[6]　徐春晖，徐向东. 前馈神经网络新学习算法研究[J]. 清华大学学报(自然科学版), 1999,(3):1 - 3.

[7]　袁亚湘，孙文瑜. 最优化理论与方法[M]. 北京：科学出版社, 1997.

[8]　DENNIS J E, GAY D M, WELSH R E. An Adaptive Nonlinear Lieast-squares Algorithm[J]. ACM Transaction on Math. Software, 1981,(7):348 - 368.

Feed-forward Neural Network Based on GaussNewton-NL2SOL Algorithm and Its Application

XU Jin

(School of Management, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: Though the feed forward neural network based on GaussNewton algorithm and its derivation will converge with order two, it is only effective toward little residual problem. In order to solve the little and large residual problems at the same time, NL2SOL algorithm is introduced and combined with the GaussNewton algorithm so as to form a feed forward neural network based on GaussNewton - NL2SOL algorithm. The application shows that this neural network can solve the residual problem properly and the convergence and stability of it performs well.

Key words: feed-forward neural network; gauss newton algorithm; NL2SOL algorithm; residual problem; convergence; stability

（编辑　吕赛英）