一种神经网络稳健估计方法的推广性研究 [1]

刘光远 廖晓峰 虞厥邦 邱玉辉 *
(电子科技大学光电子技术系 成都 610054)
*(西南师范大学计算机系 重庆 400715)

摘 要 本文根据统计学的稳健性原理，将柯西 (Cauchy) 函数作为新的神经网络目标函数．在网络参数相同的前提下，利用传统的均方目标函数和新的柯西目标函数对 BP 网络分别进行训练后，加入小噪声及异常值 (Outliers) 干扰对该网络进行测试．结果表明，具有稳健性目标函数的网络不但有更快的收敛速度，而且对异常值有更好的抵抗能力．

关键词 稳健误差估计器， BP 算法，目标函数，异常值
中图号 TN-052

1 引 言

神经网络研究成果也越来越多地应用于工程实际中， BP 网络就是其中典型的一例．在实际工程应用上，训练网络的原始数据中常包含有大量的小噪声信息，甚至包含有远大于真实数据的异常值．如何利用这样的数据使学习后的网络其映射关系基本上不受这些噪声特别是异常值干扰的影响，就是一个值得研究的问题．传统的前馈型 BP 网络的目标函数为均方函数，它假定所估计的误差呈正态分布．统计学理论分析的结果已经表明，该目标函数对异常值的抵抗力很弱 [1,2] ．针对这一情况，统计学家们自 60 年代以来开展了稳健估计 (Robust Estimation) 的研究及应用．稳健性 (Robustness) 指的是统计方法的一种属性，它一般包含两个方面的内容：一是指当实际问题的模型与理论模型有一些不大的差别时，方法的性能应受较小的影响；二是指当样本中包含少量的“异常值”时，方法的性能应受不大的影响．基于上述思想，国内外学者开展了将稳健估计方法应用于神经网络的研究工作 [3−6] ．但认真研读这些文献后，发现其共同的做法都是：依照统计学方式提出某种稳健性误差目标函数后，在数据中加入小的和异常的数据对网络进行训练，考察此稳健估计方法与原均方方法的差异．本文并不依照这种思路，而是采取先对网络进行两种方法 (稳健及非稳健) 的训练 (以 BP 网络为例)，其训练数据集相同且无噪声污染，网络的初始参数 (权及阈值) 及学习参数亦相同，再分别加入有小噪声带异常值的数据集对网络的性能进行测试，考察网络的稳健性．使用这种思路是基于这样的考虑，即网络在训练后能否适应更坏的情况？仿真实验给出了这个问题的答案．

2 稳健估计原理

神经网络具有输入输出矢量的强映射作用．当输入模式为 $\boldsymbol{X}_p$ 时，经学习后的网络映射结果为 $\boldsymbol{Y}_p$:

$$\boldsymbol{Y}_p=\boldsymbol{W}_P\boldsymbol{X}_p+\boldsymbol{E}_p,\quad p=1,2,\cdots,N, \tag{1}$$

式中 $\boldsymbol{W}_P$ 为联接权矩阵矢量， $\boldsymbol{Y}_p$ 为映射结果， $\boldsymbol{E}_p$ 为目标函数．在均方估计意义下，网络的学习过程，实际上是寻找一个合适的 $\boldsymbol{W}_P$ ，使下式成立：

$$\boldsymbol{E}_M=\min\frac{1}{N}\sum_{p=1}^{N}\boldsymbol{E}_p=\min\frac{1}{N}\sum_{p=1}^{N}(\boldsymbol{Y}_p-\boldsymbol{W}_P\boldsymbol{X}_p)^2=\min\frac{1}{N}\sum_{p=1}^{N}(\boldsymbol{r}_p)^2. \tag{2}$$

---

[1] 1997-04-22 收到， 1998-06-05 定稿
国家教委博士点基金和重庆市科委资助课题

如前所述，(2) 式及由其决定的估计受少数“异常值”的影响过大[7]。为了减少这种异常值的影响，根据文献 [3], 定义新的目标函数 (标量式) 为

$$E_C = \log(1 + r_p^2/2). \tag{3}$$

当极小化 (3) 式时，为何可以使网络对异常值不敏感呢？可通过目标函数的影响函数[3]来说明这一点。

我们都知道，在 BP 网络中误差的逆传播依下两式对联接权分布产生影响：

对输出层单元 k:

$$\delta_{pk} = \rho(r_p)O_{pk}(1 - O_{pk}); \tag{4}$$

对隐含层单元 j:

$$\delta_{pj} = O_{pj}(1 - O_{pj})\delta_{pk}W_{kj}. \tag{5}$$

从以上二式可以看出，当激励函数确定后，影响误差修正的唯一因素便是函数 $\rho(r_p)$, 它是目标函数关于误差函数变量 r_p 的一阶偏导数，即

$$\rho(r_p) = \partial E/\partial r_p, \tag{6}$$

称之为影响函数 (influence function)。图 1 给出了 (2)、(3) 两式的影响函数曲线。曲线 1 为均方估计的影响函数，曲线 2 为柯西估计的影响函数。显然，在 $|r_p| \to 0$ 的近邻内，两曲线无显著差异，但当 $|r_p|$ 迅速增大时，曲线 1 呈线性无界增长，而曲线 2 却为有界性增长，且当 $|r_p| \to \infty$ 时，只有曲线 2 趋于有界值 0 ($\lim_{r\to\infty}\rho_c(r_p) = \lim_{r\to\infty} r_p/(1 + r_p^2/2) = 0$)。这就很好地说明了柯西函数可抵御大的异常值的影响，而均方函数则无能为力。由此判断出柯西目标函数具有稳健性。

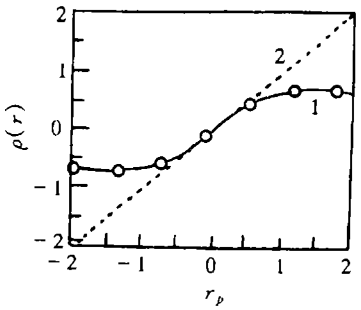

图 1　影响函数曲线图

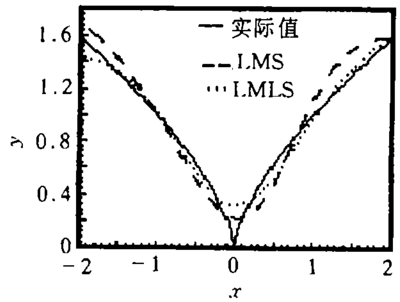

图 2　两种方法的训练结果

3　计算机仿真及讨论

为叙述方便，将 (2) 式的方法称作 LMS(least mean squares) 方法，而 (3) 式的方法称作 LMLS(least mean log squares) 方法。为了比较这两种方法的性能，每一个可控制的参数 (比如网络结构、初始权、学习率、学习算法) 都准确地保持为相同，唯一区别是误差计算不一样。

实验中，选取一个 1-10-1 的三层结构 BP 网络，其中隐层的激励函数为 Sigmoid, 而输出层则为 $f(x) = x$。学习算法为 BP 算法，学习率为 $\alpha = 0.2$, 冲量项系数 $\beta = 0.6$。由函数

$$y = x^{2/3}, \quad x \in [-2, 2] \tag{7}$$

产生 500 个训练模式对．测试集有两个：一是将产生的 500 个高斯小噪声 $G \sim N(0, 0.1)$ 加入独立变量 x 中构成；二是将掺有 5 个异常值 $H \sim N(100, 0.1)$ 的子集加入前面的小噪声测试数据集构成．

对 LMS 及 LMLS 两种方法学习时都进行了各 3 万次的迭代，结果示于图 2 中．直观分析易知，在相同的迭代次数下， LMLS 方法训练效果好于 LMS 方法．图 3 是小噪声数据集 (图中小黑点) 的测试结果，两种方法经回归分析后所得结果与图 2 结果基本相同，说明 LMS 方法对小噪声数据有抵抗能力．

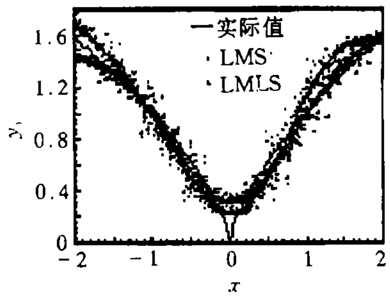

图 3　小噪声数据测试结果

表 1 异常值的测试值

点位	LMS	LMLS
50	2.07285	1.44031
150	1.53036	1.32085
250	2.97285	1.44046
350	1.53036	1.32085
450	2.97285	1.44063

图 4 是带异常值数据集的测试结果，图中有显著标记的点 (小黑圆点和小方块) 为异常值测试大小 (详见表 1) ．须提醒的是，所加异常值为均值 100, 方差等于 0.1 的正态型随机数据．

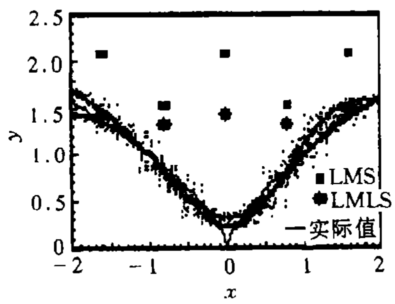

图 4　有异常值数据测试结果

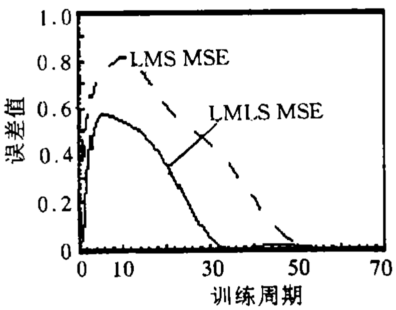

图 5　网络收敛曲线

均方根值 (RMS) 常常被用来定量说明网络的性能，其定义为

$$\mathrm{RMS} = \sqrt{\frac{\sum_{p=1}^{N}(y_p - \overline{y}_p)^2}{N}}. \tag{8}$$

这里的 y_p 是函数 $y = x^{2/3}$ 在 x_p 点的实际值， $\overline{y}_p$ 则为网络训练后的测试值．两种方法的均方根值列于表 2．表 2 说明，在两次测试中， LMLS 方法的均方根值不但比 LMS 方法的均方根值小，而且在测试异常值数据的情况下其值也只有微小的增加．

图 5 是两种方法训练同样次数下的收敛过程曲线图．显然， LMLS 方法下的目标函数较之 LMS 方法下的目标函数有更快的收敛速度．横坐标以 500 个训练模式均完成一次迭代为记录．

表 2 两种方法的均方根值

方法	LMS	LMLS
小噪声测试	0.2424	0.1683
异常值测试	0.2482	0.1689

4 结 论

将网络用不同的目标函数进行一次性训练，并用近似于实际情况的带噪声及有异常值的数据进行测试，所得结果说明了具有稳健性的目标函数有更好的推广应用性能。将此方法应用于 XOR 问题及奇偶码校验等问题也取得了好的效果 (限于篇幅，此处未加报道) . 因此，我们认为这一研究对神经网络的实际应用有着重要的指导意义。

参 考 文 献

[1] Huber P J. Robust estimation of a location parameter. The Annals of Mathematical Statistics, 1964, 35: 73–101.

[2] Andrews D A. Robust method for mutiple linear regression. Technometrics, 1974, 16(): 523–531.

[3] Liano K. Robust error measure for supervised neural network learning with outliers. IEEE Trans. on Neural Networks, 1996, NN-7(1): 244–250.

[4] Chen D S, Jain R C. A robust back propagation learning algorithm for function appoximation. IEEE Trans. on Neural Networks, 1994, NN-5(3): 467–469.

[5] Oja E, Wang L. Robust fitting by nonlinear neural units. Neural Networks, 1996, 9(3): 435–444.

[6] Humpert B K. Improving back propagation with a new error function. Neural Nerworks, 1994, 7(8): 1191–1192.

[7] 陈希孺，王松桂. 近代实用回归分析. 南宁：广西人民出版社，1984: 301–321.

[8] 廖晓峰，刘光远，虞厥邦. 几种误差估计器的稳健 BP: 理论与算法. 信号处理，1997, 13(3): 235–240.

GENERALIZATION STUDY FOR A ROBUST ESTIMATION METHOD OF NEURAL NETS

Liu Guangyuan Liao Xiaofeng Yu Juebang Qiu Yuhui*

(Department of Optoelectronic Technology, UEST of China, Chengdu 610054)

*(*Department of Computer Science, Southwest China Normal University, Chongqing* 400715)

Abstract In this paper, the Cauchy function is taken as a new target function of neural network accordings to the robustness theorem of statistics. Under the same network parameter conditions the BP net is trained using both mean squares and Cauchy target function first, then the net is tested by data sets including small Gaussian noises and outliers separately. Simulation results indicate that the network has both faster convergence speed and better performance against outliers after learning with robust target function.

Key words Robust estimator, BP nets, Target function, Outliers

刘光远： 男，1961 年生，博士，副教授，主要从事电磁测量技术和神经网络的理论与应用研究.
廖晓峰： 男，1964 年生，博士，副教授，主要兴趣是神经网络定性理论和应用研究.
虞厥邦： 男，1932 年生，教授，博士生导师，主要从事非线性网络理论和神经网络理论及应用研究.
邱玉辉： 男，1938 年生，教授，主要从事人工智能和神经网络的理论研究.