一种基于鲁棒估计的极限学习机方法*

胡义函[a]，张小刚[a]，陈　华[b]，李晶辉[a]

（湖南大学 a. 电气与信息工程学院；b. 信息科学与工程学院，长沙 410082）

摘　要： 极限学习机（ELM）是一种单隐层前馈神经网络（single-hidden layer feedforward neural networks，SLFNs），它相较于传统神经网络算法来说结构简单，具有较快的学习速度和良好的泛化性能等优点。ELM 的输出权值是由最小二乘法（least square，LE）计算得出，然而经典的 LS 估计的抗差能力较差，容易夸大离群点和噪声的影响，从而造成训练出的参数模型不准确甚至得到完全错误的结果。为了解决此问题，提出一种基于 M 估计的采用加权最小二乘方法来取代最小二乘法计算输出权值的鲁棒极限学习机算法（RBELM），通过对多个数据集进行回归和分类分析实验，结果表明，该方法能够有效降低异常值的影响，具有良好的抗差能力。

关键词： 极限学习机；稳健估计；鲁棒极限学习机；M 估计；神经网络

中图分类号： TP18　　**文献标志码：** A　　**文章编号：** 1001-3695(2012)08-2926-05

doi:10.3969/j.issn.1001-3695.2012.08.033

Extreme learning machine on robust estimation

HU Yi-han[a], ZHANG Xiao-gang[a], CHEN Hua[b], LI Jing-hui[a]

(*a. College of Electrical & Information Engineering, b. School of Information Science & Engineering, Hunan University, Changsha* 410082, *China*)

Abstract: Extreme learning machine(ELM) is a kind of single-hidden layer feedforword neural networks(SLFNs). Comparing with traditional neural network algorithms, it is simpler in structure, with higher learning speed, and good generalization performance. The output-weight of ELM was calculated by LSE(least square estimation) method. However, LSE lack of robustness, the result would be seriously damaged when there were outliers in the training data. In order to solve this problem, this paper derived a novel approach based on M-estimators of extreme learning machine called RBELM. Simulation results indicate that the RBELM proposed can significantly robust against data noise and outliers.

Key words: extreme learning machine(ELM); robust estimation; robust extreme learning machine(RELM); M-estimator; neural networks

0　引言

稳健估计（robust estimation）是在粗差不可避免的情况下，选择适当的估计方法使未知量估值尽可能减免粗差的影响，得出正常模式下的最佳估值。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：a）当实际问题的模型与理论模型差别不大时，方法的性能应受较小的影响；b）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。文献[1]对 BP 算法提出了一个误差估计器，有一定的抗差作用，但可能导致过调或振荡；文献[2]提出一种稳健 BP 算法，利用一个自适应可调的目标函数去降低噪声对训练的影响，然而这些工作并未解决 BP 对初始权值的敏感以及速度慢的问题。文献[3]在 BP 中引入模拟退火概念（ARBP 算法）来抑制拟合过程中的离群点；文献[4]提出一种稳健 RBF 算法，该方法加入 S 型激活函数以及具有稳健性能的目标函数，使之具有抗差能力。这两种方法都引入了稳健型目标函数，并利用分界点去识别离群点，但分界点的设置问题却难以得到解决。模糊神经网络是将模糊系统作为神经网络部分。文献[5]提出一种鲁棒模糊神经聚类算法，能有效地减小非线性系统的噪声和离群点的影响。上述改进的神经网络方法[1~5]一定程度上提高了学习的稳健性，但是网络训练速度慢，设置参数过多，方法结构复杂，仍然是其所存在的重要问题。

极限学习机（ELM）是 Huang 等人[6~8]在 2005 年提出的一种新型单隐层前馈神经网络，研究已证实 ELM 具有与神经网络（NN）相同的全局逼近性质，其网络输入权值及隐层神经元偏移量是随机生成，只需要设置隐层节点数，由 LS 法得出输出权重即可产生唯一最优解。ELM 在保证网络具有良好的泛化性能的同时，极大地提高了学习速度，同时避免了由于梯度下降算法产生的许多问题，如局部极小、迭代次数过多、学习时间长、性能指标以及学习率等[9]。ELM 算法一经提出后，其改进算法的研究和现场应用受到广泛关注[10]。

传统 ELM 的输出权重的计算是直接由最小二乘估计方法得出的。从网络学习的稳健性考虑，ELM 仍有两个尚待解决的固有缺陷：a）其隐层部分参数的选取为随机确定，这会造成 ELM 网络的训练结果有较大的不稳定性[11]，为了解决这一问题，文献[12]利用 PSO 对随机选取的参数进行优化以提升学

收稿日期： 2012-01-29；**修回日期：** 2012-02-29　　**基金项目：** 国家自然科学基金资助项目（60874096，61174050）

作者简介： 胡义函（1987-），男，硕士研究生，主要研究方向为模式识别（hu_yihan@qq.com）；张小刚（1972-），男，教授，博士，主要研究方向为工业窑炉过程控制、模式识别；陈华（1973-），女，讲师，博士，主要研究方向为图像处理、模式识别；李晶辉（1985-），男，硕士研究生，主要研究方向为模式识别.

习器精度。文献[13,14]则采用多个分类器集成的方法如 Adaboost 理论来提升算法性能;b)如果训练数据存在共线性或粗差干扰,输出层权值的最小二乘估计结果会很差。文献[15]提出贝叶斯极限学习机(Bayesian extreme learning machine, BELM),试图利用一些先验知识来弥补学习过程粗差干扰问题;文献[16]提出一种基于最小二乘辨识中的零估计来解决训练数据中的共线性问题,但回归参数 k 的最优值依赖于模型未知参数 β 和误差的方差 δ^2,使得 $\bar{\beta}(k)$ 的估计实质成为一个非线性估计,丧失了最小二乘的快速性。上述改进的 ELM 学习方法虽然提高了学习的稳定性,但或因计算复杂度较高而丧失了极限学习的快速性,或需要一些先验知识较难获取。

本文结合线性系统的稳健估计理论,将一种鲁棒估计引入 ELM 网络,即鲁棒极限学习机(RBELM),采用加权最小二乘方法计算网络输出权值,充分利用了 ELM 的快速性,并通过 SinC 函数加异常值逼近及基准数据回归的方针实验和实际分类数据测试,验证了本文方法具有良好的抗粗差干扰的能力。

1　极限学习机介绍

对于 N 个不同的学习样本 (x_i,t_i),其中 $x_i=[x_{i1},x_{i2},\cdots,x_{in}]^{\mathrm{T}}\in\mathrm{R}^n$,$t_i=[t_{i1},t_{i2},\cdots,t_{im}]^{\mathrm{T}}\in\mathrm{R}^m$,具有 L 个隐层节点和激活函数 $g(x)$ 的单隐层前向网络,$L\leqslant N$,可描述为下述线性方程组形式:

$$H\beta=T \tag{1}$$

其中:$H=\begin{bmatrix} g(a_1\cdot x_1+b_1) & \cdots & g(a_L\cdot x_1+b_L)\\ \vdots & & \vdots \\ g(a_1\cdot x_N+b_1) & \cdots & g(a_L\cdot x_N+b_L)\end{bmatrix}_{NxL}$,$\beta=[\beta_1,\beta_2,\cdots,\beta_L]^{\mathrm{T}}\in\mathrm{R}^{Lxd}$ 为输出层权值向量;$T=[t_1,t_2,\cdots,t_N]^{\mathrm{T}}\in\mathrm{R}^N$ 为样本输出向量;H 为神经网络的隐层输出矩阵,它的第 i 列表示的是第 i 个隐层神经元关于输入 $x_1,x_2,\cdots,x_N$ 的输出。

文献[17]已经证明,如果激活函数 $g(x)$ 无限可微,那么随机制定的输入权值、隐层节点偏移值均不需要调节。对于已经固定的输入权值 a_i 和隐层节点偏移值 b_i,训练一个单隐层前馈神经网络实际上就是找到线性系统 $H\beta=T$ 的关于 $\hat{\beta}$ 的最小二乘解。如果隐层节点数 L 等于训练样本数 N 时,即 $L=N$,那么当输入权值 a_i 和隐层节点偏移值 b_i 被随机确定时,H 是个可逆的方阵,并且单隐层前馈神经网络能够零误差逼近训练样本。然而大多数情况下,隐层节点数目是远远小于训练样本个数的,即 $L<<N$,则 H 是一个 $N\times L$ 的矩阵,此时就需要求 H 的伪逆,即

$$\hat{\beta}=H^{+}T=(H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}T \tag{2}$$

由于 ELM 输出权值的估计方法是最小二乘法,而最小二乘法是对每个观察数据均给予相同的权重,因此,当数据集中存在离群点时,该方法将夸大这些奇异值的影响,从而导致整个系统估计偏差变大,甚至得到完全错误的结果。

为了克服上述缺点,本文把一种鲁棒估计引入 ELM 输出权值的估计中,提出一种鲁棒极限学习机(RBELM)。

2　鲁棒极限学习机

2.1　稳健回归理论

稳健回归的目的是减少大量随机误差及少量奇异值对待估参数的影响。当数据误差为正态分布时,稳健回归的估计精度与 LSE 相当;但当数据误差为非正态分布时,其估计精度将比 LSE 要好很多。这种对误差项分布的稳健特性,能有效排除奇异值的干扰[18]。

最小二乘估计(least square estimation, LSE)是现行线性回归中应用最为广泛的方法,该方法简单实用,能在正态假定下应用统计检验理论。然而最小二乘法存在一定的局限性:a)由于最小二乘法对每个观察数据均给予相同的权重,因此只适用于实验数据相关性较高的情况,不适用于数据离散并有奇异值存在的情况;b)由于最小二乘法求解残差平方和来达到最小求解回归系数,容易夸大实验数据中奇异值的影响,统计误差增大,说明最小二乘估计方法缺乏稳健性[19~21]。

此处采用的稳健估计为 M 估计,它是在经典极大似然估计基础上的推广,对线性回归模型:

$$Y=X\beta+e \tag{3}$$

最小二乘法使残差平方和 $Q=\sum_{i=1}^{N}(Y_i-\sum_{j=1}^{p}x_{ij}\beta_j)^2=\sum_{i=1}^{N}e_i^2$ 达到最小。由该式可以得出,异常值对于系统估计的影响很大。M 估计稳健回归基本思想是采用迭代加权最小二乘估计回归系数,根据回归残差大小来确定各样本的权重,以达到稳健的目的。优化目标函数为

$$Q=\sum_{i=1}^{N}\rho(e_i)=\sum_{i=1}^{N}\rho(Y_i-\sum_{j=1}^{p}x_{ij}\beta_j) \tag{4}$$

其中:ρ 为影响函数,令 $\psi=\rho'$ 为 ρ 的导函数,目标函数对参数 β 求偏导,得

$$\sum_{i=1}^{N}\psi(Y_i-\sum_{j=1}^{p}x_{ij}\beta_j)x_{ij}=0 \tag{5}$$

令 $w(e_i)=\psi(e_i)/e_i$,即 $\psi(e_i)=w(e_i)\cdot e_i$,则式(5)可转换为 $\sum_{i=1}^{N}w_iX_ie_i=0$,即

$$X^{\mathrm{T}}We=0 \tag{6}$$

将式(3)代入式(6)中可得

$$X^{\mathrm{T}}WY-X^{\mathrm{T}}WX\beta=0 \tag{7}$$

即

$$\beta=(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}WY \tag{8}$$

此时就变成一个加权最小二乘估计的问题。为减少异常点作用,对不同的点施加不同的权重,即对残差小的点给予较大的权重,而对残差较大的点给予较小的权重。根据残差大小确定权重,并据此建立加权的最小二乘估计。反复迭代以改进权重系数,直至权重系数的改变小于一定的允许误差,这样就降低了异常值的影响,提高了抗差性[22]。

2.2　鲁棒极限学习机

基于上述稳健回归的优点,可以将加权最小二乘引入到 ELM 输出权重的学习中,降低在估计输出权重时由于异常值造成的影响,提高算法稳健性。

在稳健估计中影响函数众多,可以根据实际需要选择各种不同的影响函数来代替残差平方和,如采用 Huber[19] 研究的影响函数。

$$\rho(x)=\begin{cases}x^2/2 & |x|\leqslant k\\ k|x|-k^2/2 & |x|>k\end{cases} \tag{9}$$

其中:k 为调和常数,默认取 $k=1.345$。

这时关于残差的目标函数为

$$Q(\beta)=\sum_{i=1}^{N}\rho(e_i)=\sum_{i=1}^{N}\rho(T_i-H_i\beta) \tag{10}$$

其中:H 是一个 $N\times L$ 的矩阵;β 是一个 $L\times 1$ 的矩阵;N 是

样本数;e_i 是残差,则目标函数对参数β 求偏导,并令偏导数等于零,得到

$$\sum_{i=1}^{N}\psi(\hat{T}_i - H_i\beta)H_i = 0 \tag{11}$$

其中:$\psi(x)$为 $\rho(x)$的导函数。

$$\psi(x) = \begin{cases} -k & x < -k \\ x & |x| \leqslant k \\ k & x > k \end{cases} \tag{12}$$

为了使 M 估计具有稳健性,在权重函数中引入一个稳健的尺度估计 s 使残差标准化,即 e_i/s,s 一般取值为 Hampel 提出的绝对离差中位数 MAD(median absolute deviation)除以一个常数 0.6745,得到标准化残差 $u_i = e_i/s = 0.6\ 745e_i/\text{med}(|e_i|)$,其中,med 代表中位数计算。

于是由式(11)可得

$$\sum_{i=1}^{N}\psi(T_i - H_i\beta)H_i = \sum_{i=1}^{N}\psi(u_i)H_i = \sum_{i=1}^{N}\frac{\psi(u_i)}{u_i}\cdot u_i H_i = \sum_{i=1}^{N}W_i\cdot u_i\cdot H_i = 0$$

其中:$W_i = \dfrac{\psi(u_i)}{u_i}$为序号为 i 的样本权重,则β 的稳健估计可用迭代加权最小二乘求解。由式(8)得

$$\hat{\beta} = (H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}WT \tag{13}$$

于是,RELM 的算法步骤为:

给定数据集 $Z = \{(x_i,t_i)\,|\,x_i \in \mathrm{R}^n, t_i \in \mathrm{R}^m, i = 1,2,\cdots,N\}$,激活函数 $g(x)$,隐层节点数 L。

a)随机选取输入连接权值 a_i 和隐层节点偏移值 b_i。

b)计算隐层节点输出矩阵 H。

c)选取 LS 估计的 $\hat{\beta}^{(0)} = H^{+}T = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}T$ 为迭代初值,求得初始残差 e。

d)标准化残差得 u,由 $W_i = \dfrac{\psi(u_i)}{u_i}$求得每个样本的初始权重。

e)利用加权最小二乘法 $\hat{\beta} = (H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}WT$ 求得的 $\hat{\beta}^{(1)}$ 代替 c)中的 $\hat{\beta}^{(0)}$,求得新的残差 e,从而获得新的权重。

f)返回步骤 d),依次类推计算稳健估计值 $\hat{\beta}$。当相邻两步的回归系数的差的绝对值的最大值小于设定的标准误差时,迭代结束,即 $\max(|\hat{\beta}^{(i)} - \hat{\beta}^{(i-1)}|) < \varepsilon$。

3 实验

3.1 回归分析

3.1.1 SinC 函数

SinC 函数的表达式为

$$y(x) = \begin{cases} \sin x/x & x \neq 0 \\ 1 & x = 0 \end{cases}$$

数据是在(-10,10)内随机产生,训练集与测试集均为 5 000组数据,并在[-0.2,0.2]内添加随机噪声,而测试集无噪声。测试结果如表 1 所示。

表 1 算法在 SinC 上的比较

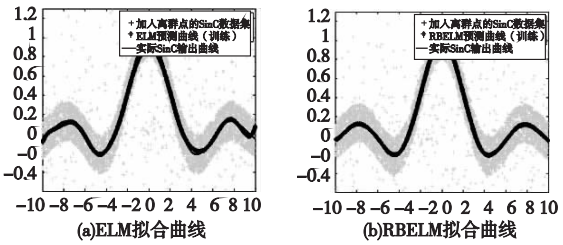
算法	时间/s		RMSE		Dev		隐节点个数
	训练	测试	训练	测试	训练	测试	
ELM	0.031 6	0.025 0	0.115 2	0.009 7	0.003 7	0.002 8	20
RELM	0.079 4	0.025 0	**0.112 9**	**0.007 0**	0.016 2	**0.001 6**	20

由表 1 可得出,由于 RBELM 中加权迭代的影响,其训练时间比 ELM 稍长。但从 RMSE 上可看出,RBELM 是 0.0070 而 ELM 是 0.009 7,所以均方根上 RBELM 比 ELM 要略好。

为验证 RBELM 的抗差能力,将 SinC 训练数据集的 5000 组数据随机选取 500 个取值范围[-2,+2]的离群噪声点。测试结果如表 2 所示。图 1 和 2 为加离群点情况下的 SinC 测试集和训练集的拟合情况。由表 2、图 1 和 2 可以得出,在有离群点情况下的 ELM 拟合偏差得相当厉害,受离群点影响很大;而 RBELM 却拟合效果很好,说明其受离群点的影响很小。由图可知,RBELM 的输出对训练和测试真实数据的拟合效果均优于 ELM。

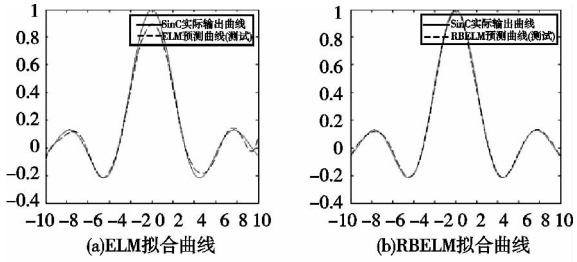
表 2 算法在有 10% 离群点的 SinC 上的比较

算法	时间/s		RMSE		Dev		隐藏层节点数
	训练	测试	训练	测试	训练	测试	
ELM	0.031 6	0.025 0	0.259 3	0.039 9	0.007 1	0.004 3	20
RBELM	0.093 8	0.025 0	**0.257 0**	**0.008 4**	0.037 5	**0.001 7**	20



(a)ELM拟合曲线 (b)RBELM拟合曲线

图1 加离群点情况下的 SinC训练集的拟合情况



(a)ELM拟合曲线 (b)RBELM拟合曲线

图2 加离群点情况下的 SinC测试集的拟合情况

3.1.2 实际回归问题

由于 RBELM 采用的是基于 M 估计的迭代加权最小二乘估计,所以针对它的鲁棒性对回归问题进行测试。对于七组实际回归数据集,数据集信息如表 3 所示。首先对原数据集进行测试,比较 ELM 和 RBELM 的算法性能,然后对数据进行加噪处理,再比较 ELM 和 RBELM 的算法性能。

表 3 回归数据集属性

数据集	样本数目		属性数目	
	训练样本	测试样本	连续属性	标称属性
Abalone	2 000	2 177	7	1
Servo	80	87	6	0
Bank	4 500	3 692	8	0
Censor	10 000	12 784	8	0
Computer compact	4 000	4 192	8	0
elevator	4 000	5 517	6	0
Machine CPU	100	109	6	0

添加噪声的方法如下:如对 UCI 数据集 Abalone,该数据集有 4 177 个样本,每个样本中具有 7 个连续属性,1 个标称属性。随机选取 10% 或者 20% 的数据,按照(data × (1 + randn))进行处理。具体结果如表 4 所示。

由实验结果表明,在回归问题上,RBELM 在训练时间上要大于 ELM,但在有噪声添加的回归数据集上,RBELM 精度要优于 ELM。通过大量的实验,RBELM 在大多数有噪声的数据

集中测试的 RMSE 要小于 ELM,说明其具有良好的泛化性能以及较强的抗差能力。

表 4　实际回归问题在 RBELM 和 ELM 的性能比较

数据集	训练 RMSE		测试 RMSE		时间		隐藏层节点数
	ELM	RBELM	ELM	RBELM	ELM	RBELM	
abalone	0.076 6	0.076 6	0.076 4	**0.075 9**	0.016 9	0.040 6	25
+10% 噪声	0.092 9	0.094 1	0.081 1	**0.079 9**	0.017 5	0.048 4	25
+20% 噪声	0.098 7	0.099	0.085 7	0.085 7	0.015 9	0.050 3	25
Servo	0.056 8	0.059 6	0.119 5	**0.102 4**	0.003 7	0.006 3	30
+10% 噪声	0.128	0.189 2	0.165 7	**0.140 9**	0.003 7	0.007 5	30
+20% 噪声	0.117 2	0.141 8	0.201 7	**0.151**	0.003 7	0.01	30
bank	0.038 7	0.038 7	0.041 8	0.041 8	1.240 6	2.941	220
+10% 噪声	0.057 5	0.141 7	0.046 5	**0.042**	1.376 6	3.043 7	220
+20% 噪声	0.080 3	0.629 8	0.053 1	**0.043 1**	1.357 8	3.002	220
censor	0.06	0.061	0.076 4	**0.071 7**	2.015 6	4.210 7	160
+10% 噪声	0.078 5	0.079 7	0.079 6	**0.078**	2.103 5	4.231 1	160
+20% 噪声	0.082 8	0.084 6	0.191 7	**0.168 5**	2.103 5	4.231 5	160
Computer compact	0.035 3	0.035 6	0.043 7	0.043 6	0.914 1	1.998 4	125
+10% 噪声	0.062 2	0.063 4	0.077 2	**0.072 7**	0.919 5	2.000 6	125
+20% 噪声	0.091 1	0.094 7	0.081 8	**0.075**	0.917 8	2.366	125
elevator	0.050 6	0.050 8	0.054 5	0.054 3	0.618 8	1.153 1	125
+10% 噪声	0.054 2	0.054 3	0.053 7	0.053 5	0.729 7	1.518 8	125
+20% 噪声	0.055 9	0.056	0.054	0.053 8	0.725	1.509 4	125
Machine CPU	0.028	0.028 1	0.081 4	**0.078 5**	6.25E-04	0.003 7	10
+10% 噪声	0.050 6	0.068	0.073 4	**0.063 5**	0.001 9	0.006 6	10
+20% 噪声	0.04	0.051 3	0.090 8	**0.082 2**	0.010 6	0.004 1	10

3.2　分类问题

在 3.1 节中证明了 RBELM 在回归问题上的性能,本节对 RBELM 的分类性能进行测试。数据集属性如表 5 所示,测试结果如表 6 所示。对于分类问题的加噪方式与 3.1.2 节回归问题相同。其中,Dsnl 数据集为从回转窑火焰图像提取的数据,其余五组均为 UCI 数据库中的实际分类数据集。从表 6 可以得出,RBELM 相较于 ELM 在分类问题上同样具有很好的泛化能力以及鲁棒性。

表 5　分类数据集属性

数据集	样本数目		属性数目	
	训练样本	测试样本	属性	类别数目
House-vote	300	135	16	2
Kr-vskp	2 140	1 056	36	2
Banana	7 000	2 800	2	2
Letter	7 000	3 000	16	26
Wine	126	52	13	3
Haberman	200	106	3	2
Dsnl	2 000	750	4	3

3.3　调和参数 k 的影响

在 RBELM 中,影响函数的调和参数 k 的大小,从函数的几何意义来说可以理解为影响函数对训练样本残差的容忍程度。例如 tarwar 权重函数(图 3),如果 k 取值越大,那么对残差较大的样本所赋予的权重值就越大;如果 k 取值越小,那么对残差较大的样本赋予权重值越小;而当 $k\to\infty$ 时,则 RBELM 将退化为普通 ELM,相当于对每个样本赋予相同的权重。因此,如果参数 k 能取得最优,将很大程度地提高 RBELM 的精确度。

为了获取较好的分类精度,本文所选取的影响函数中,调和参数 k 是误差估计中的关键参数,其受到样本数目、待估样本分布、影响函数形式的影响,影响函数的选取原则上是让单个样本的立方体随样本数量的增大缓慢减小,达到估计密度函数的收敛,但由于实验样本有限,多采用经验确定。

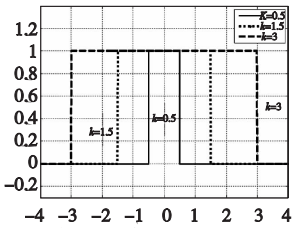


图3　权重函数tarwar在不同调和参数下的情况

表 6　ELM 和 RBELM 在分类问题上的比较

数据库	训练精度		测试精度		训练时间		隐节点数目
	ELM	RBELM	ELM	RBELM	ELM	RBELM	
Vote	0.959 3	0.954 1	0.958 5	**0.964 9**	0.005	0.013 8	30
+10% 噪声	0.958 2	0.953 1	0.929 2	**0.954 8**	0.005	0.013 8	30
+20% 噪声	0.957 7	0.941 8	0.909	**0.955 7**	0.005	0.013 8	30
Kr-vskp	0.952 7	0.952 8	0.946 1	**0.946 7**	0.314 1	0.632	150
+10% 噪声	0.904 2	0.921 1	0.915 8	**0.930 7**	0.333 6	0.964 1	150
+20% 噪声	0.854 5	0.865	0.888 7	**0.901 5**	0.333 6	0.964 1	150
Banana	0.897 4	0.899 5	0.901 3	**0.905 7**	0.333 6	0.725 8	60
+10% 噪声	0.829 1	0.848 4	0.841	**0.874 4**	0.323 4	0.917 2	60
+20% 噪声	0.793 8	0.822 8	0.835 9	**0.861 8**	0.323 4	0.917 2	60
Letter	0.884 3	0.868 5	0.839 6	0.825 1	2.418 8	6.434 4	300
+10% 噪声	0.663 1	0.673	0.660 4	**0.700 4**	2.403 1	6.434 4	300
+20% 噪声	0.575 5	0.574 4	0.615 7	**0.632 0**	2.403 1	6.434 4	300
Wine	0.995 4	0.992 4	0.986 9	0.986 2	0.003 1	0.01	20
+10% 噪声	0.971 9	0.971 6	0.95	**0.981 5**	0.004 7	0.011 9	20
+20% 噪声	0.932 2	0.936 7	0.948 1	**0.963 1**	0.005	0.013 4	20
Haberman	0.756 1	0.758 2	0.825 3	**0.831 1**	0.003 1	0.004 4	20
+10% 噪声	0.807 0	0.804 3	0.746 6	**0.753 4**	0.003 1	0.006 9	20
+20% 噪声	0.788 9	0.777 7	0.707 2	**0.721 5**	0.002 8	0.003 4	20
Dsnl	0.843 6	0.843 3	0.841 4	0.840 2	0.021 9	0.012 0	30
+10% 噪声	0.784 7	0.788 3	0.798 0	**0.802 0**	0.023 1	0.099 7	30
+20% 噪声	0.724 4	0.769 3	0.742 1	**0.788 3**	0.023 1	0.099 7	30

本实验中选取 Dsnl(+20% 噪声)和 Haberman(+10% 噪声)数据集进行测试,调和参数变化范围为[0.01,5],步长为 0.2,分别使用 tarwar、cauchy、welsch、andrews 四个影响函数进行测试,测试结果如图 4 和 5 所示。其中,y 轴表示分类精度,x 轴表示影响参数。由图 4 可知,Dsnl 数据集在四个不同影响函数下的分类精度均在 79.5% 附近达到了峰值,明显要高于表 5 中数据 dsnl 的 RBELM 测试精度 78.33%。在图 5 中,Haberman 数据集在四个不同影响函数下 RBELM 的分类精度最优值均达到了 78% ~79%,测试结果明显高于表 5 中 RBELM 的测试精度 75.34%。该实验证明了调和参数 k 对实验结果有着重要的影响。由于目前多采用经验和测试方法确定其值,所以关于 k 的选取方法仍然需要进一步的研究。

4　结束语

本文通过分析传统 ELM 算法在稳健性上的不足,提出了一种基于 M 估计的利用加权最小二乘方法取代最小二乘计算网络输出权值的鲁棒极限学习机,即 RBELM 算法。该算法具有良好的泛化性能,其不仅延续了 ELM 的快速学习的特点,并且对于存在有明显噪声的数据集,在赋予残差权重时根据其每

个样本的残差大小分配权值,这样就降低了异常值的影响,从而达到降噪的目的。为了验证该算法的有效性,本文针对回归和分类问题做了大量的实验,实验数据表明,RBELM 相较于传统 ELM 来说具有很好的泛化能力以及鲁棒性,能够对噪声有一定的抗干扰能力。当然,对于该算法还有待进一步研究之处,如隐层节点数能否通过自适应确定、影响函数的选取等。本文最后提出并验证了调和参数 k 对实验精度的影响,但关于 k 的取值目前仍然是多采用经验和测试方法确定,所以这些问题将是笔者下一步工作的重点。

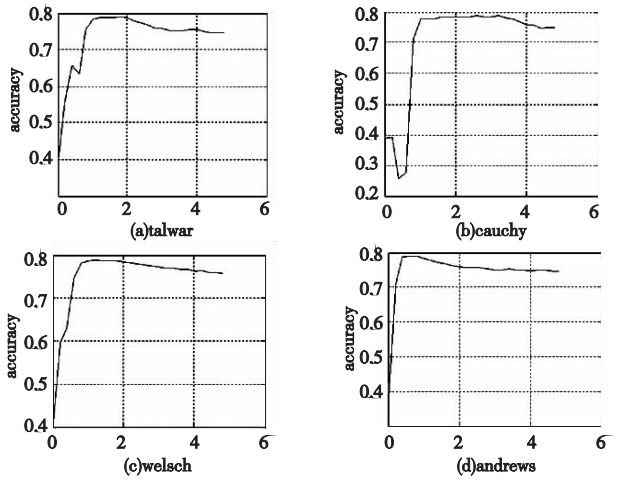


图4 Dsnl在四种影响函数下的测试曲线

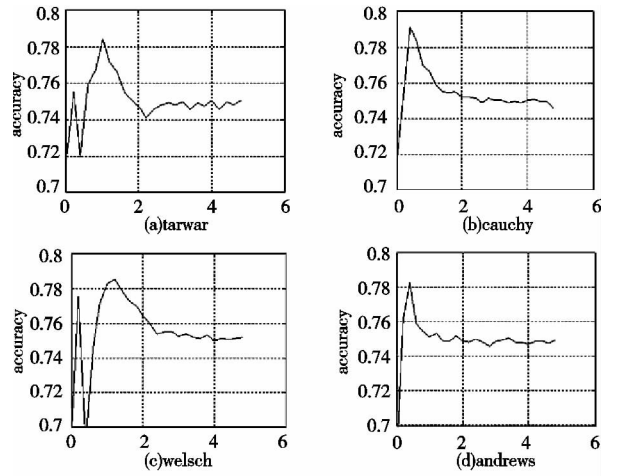


图5 Haberman数据集在四种影响函数下的测试曲线

参考文献:

[1] BAUM E, WILCZEK F. Supervised learning of probability distributions by neural networks[C]//Proc of Neural Information Processing Systems. 1987: 52-61.

[2] CHEN D S, JAIN R C. A robust back propagation learning algorithm for function approximation[J]. IEEE Trans on Neural Network, 1994,5(3): 467-479.

[3] CHUNG C C, SU Shun-feng, HSIAO C C. The annealing robust backpropagation (ARBP) learning algorithm[J]. IEEE Trans on Neural Networks,2000,11(5):1067-1077.

[4] LEE C C, CHUNAG P C, TSAI J R,*et al*. Robust radial basis function neural networks[J]. IEEE Trans on Systems Man, and Cybernetics,Part B,1999,29(6): 674-685.

[5] SHIEH H L, YANG Y K, CHANG Po-lun,*et al*. Robust neural-fuzzy method for function approximation[J]. Expert System with Application,2009,36(3): 6903-6913.

[6] HUANG Guang-bin, ZHU Qin-yu, SIEW C K. Extreme learning machine: a new learning scheme of feedforward neural networks[C]// Proc of IEEE International Joint Conference on Neural Networks. 2004:985-990.

[7] HUANG Guang-bin, ZHU Qin-yu, SIEW C K. Extreme learning machine: theory and application[J]. Neurocomputing, 2006, 70 (1-3): 489-501.

[8] HUANG Guang-bin, CHEN Lei, SIEW C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. IEEE Trans on Neural Networks, 2006, 17 (4): 879-892.

[9] ZHU Qin-yu, QIN A K,QIN A K,*et al*. Evolutionary extreme learning machine[J]. Pattern Recognition,2005,38(10):1759-1763.

[10] HUANG Guang-bin, WANG Dian-hui,LAN Yuan. Extreme learning machines: a survey[J]. International Journal of Machine Leaning and Cybernetics,2011,2(2): 107-122.

[11] SURESH S, BABU R V, KIM H J. No-Reference image quality assessment using modified extreme learning machine classifier[J]. Applied Soft Computing,2009,9(2): 541-552.

[12] SARASWATHI S, SUNDARAM S, SUNDARARAJAN N,*et al*. ICGA-PSO-ELM approach for accurate multiclass cancer classification resulting in reduced gene sets in which genes encoding secreted proteins are highly represented[J]. IEEE/ACM Trans on Computational Biology and Bioinformatics,2011,8(2): 452-463.

[13] TIAN Hui-xin, MAO Zhi-zhong. An ensemble ELM based on modified adaBoost. RT algorithm for predicting the temperature of molten steel in ladle furnace[J]. IEEE Trans on Automation Science and Engineering,2010,7(1): 73-80.

[14] HEESWIJK M, MICHE Y,LINDH-KNUUTILA T,*et al*. Adaptive ensemble models of extreme learning machines for time series prediction [C]//Proc of the 19th International Conference on Artificial Neural Networks. 2009: 305-314.

[15] SORIA-OLIVAS E,GÓMEZ-SANCHIS J,MARTIN J D,*et al* BELM: Bayesian extreme learning machine[J]. IEEE Trans on Neural Networks,2011,22(3): 505-509.

[16] 邓万宇,郑庆华,陈琳. 神经网络极速学习方法研究[J]. 计算机学报,2010,33(2): 279-287.

[17] HUANG Guang-bin. Learning capability and storage capacity of two-hidden-layer feedforward networks[J]. IEEE Trans on Neural Networks,2003,14(2): 274-281.

[18] LAWSON C, KEATS J B, MONTGOMERY D C. Comparison of robust and least-squares regression in computer-generated probability plots[J]. IEEE Trans on Reliability,1997,46(1):108-115.

[19] HUBER P J. Robust statistics[M]. New York;Wiley,1981.

[20] 吴耀华. 线性模型中 M 估计的渐进性质[J]. 应用数学学报, 1994,7(3):401-410.

[21] ANDREWS D F. A robust method for multiple linear regression[J]. Technometrics,1974,16(4):523-531.

[22] STREET J O,CARROLL R J,RUPPERT D. A note on computing robust regression estimates via iteratively reweighted least squares [J]. The American Statistician,1988,42(2):152-154.