

基于广义最大相关熵准则的宽度学习系统

赵海全 陆 鑫

(西南交通大学电气工程学院, 四川成都 610031)

摘 要: 宽度学习系统(broad learning system, BLS)是近几年提出的一种新型判别学习方法,具有结构简单,训练快速的特点,在各种回归和分类问题上得到广泛应用。然而标准的 BLS 是在最小均方误差(MMSE)准则下推导出来的,对异常值的存在十分敏感,这无疑降低了系统的准确性。为了提高 BLS 的鲁棒性,有学者提出了最大相关熵准则(MCC)的 BLS(C-BLS)。相对于最小均方误差准则,最大相关熵准则包含了更多的高阶误差信息,所以 C-BLS 对异常值具有良好的鲁棒性。但考虑到相关熵中默认的核函数固定为高斯核,这并不适用于绝大多数情况。本文中引入了以广义高斯密度(GGD)函数作为核函数的广义相关熵,并将广义最大相关熵准则(GMCC)应用于 BLS,提出了新的鲁棒算法(GC-BLS)。相较于高斯核函数,广义高斯密度函数更为灵活,高斯核可以看作它的一个特例,在选取适当参数时,GC-BLS 将退化为 C-BLS,这使得新算法至少能获得与 C-BLS 算法相当的性能。实验中以均方根误差作为标准,在回归数据集与时间序列数据集上对新算法进行检验,在绝大多数情况下,GC-BLS 都能取得相较于其他算法更小的均方根误差。实验表明,该算法是非常稳定的。仿真结果验证了理论上的期望,并验证了新算法的性能。

关键词: 宽度学习系统(BLS); 最大相关熵准则(MCC); 广义最大相关熵准则(GMCC); 回归任务

中图分类号: TP183

文献标识码: A

DOI: 10. 16798/j. issn. 1003-0530. 2023. 11. 005

引用格式: 赵海全,陆鑫. 基于广义最大相关熵准则的宽度学习系统[J]. 信号处理, 2023, 39(11): 1957-1963. DOI: 10. 16798/j. issn. 1003-0530. 2023. 11. 005.

Reference format: ZHAO Haiquan, LU Xin. Broad learning system based on generalized maximum correntropy criterion [J]. Journal of Signal Processing, 2023, 39(11): 1957-1963. DOI: 10. 16798/j. issn. 1003-0530. 2023. 11. 005.

Broad Learning System Based on Generalized Maximum Correntropy Criterion

ZHAO Haiquan LU Xin

(School of Electrical Engineering, Southwest Jiaotong University, Chengdu, Sichuan 610031, China)

Abstract: Broad learning system (BLS) is a new kind of discriminative learning method proposed in recent years. It has the characteristics of simple structure and fast training, and has been widely used in various regression and classification problems. However, the standard BLS is derived under the Minimum Mean Square Error (MMSE) criterion, which is very sensitive to the existence of outliers, which undoubtedly reduces the accuracy of the system. In order to improve the robustness of BLS, some scholars have proposed BLS with maximum correntropy criterion (MCC) (C-BLS). Compared with the minimum mean square error criterion, the maximum correntropy criterion contains more high-order error information, so C-BLS has good robustness to outliers. However, considering that the default kernel function in correntropy is fixed as Gaussian kernel, this is not applicable in the vast majority of cases. In this paper, the generalized correntropy with generalized

Gaussian density (GGD) function as the kernel function is introduced, and the generalized maximum correntropy criterion (GMCC) is applied to BLS, and a new robust algorithm (GC-BLS) is proposed. GC-BLS can be regarded as a special case of GC-BLS. When appropriate parameters are selected, GC-BLS will degenerate to C-BLS, which makes the new algorithm at least obtain the same performance as C-BLS algorithm. In the experiment, the root mean square error is used as the standard to test the new algorithm on regression data sets and time series data sets. In most cases, GC-BLS can achieve smaller root mean square error than other algorithms. Experiments show that the algorithm is very stable. Simulation results validate the theoretical expectations and demonstrate the performance of the new algorithm.

Key words: broad learning system (BLS); maximum correntropy criterion (MCC); generalized maximum correntropy criterion (GMCC); regression tasks

1 引言

近年来神经网络迅速发展,展现了强大的特征提取和非线性逼近能力。但当数据量较大时,由于层数增加,神经网络在逐层反向更新权重时,会遇到梯度消失,梯度爆炸,训练速度慢等问题。对此,Chen等人提出了宽度学习系统(broad learning system, BLS)^[1]。BLS具有训练速度快,结构简单的特点,已成功地应用于人脸识别^[2]、文本分类^[3]、时间序列预测^[4]等领域。但标准BLS使用最小均方误差(minimum mean square error, MMSE)准则训练网络输出权值,MMSE准则在处理含有如离群值等非线性噪声^[5]的数据时会降低BLS的性能。对此,众多学者对标准BLS进行了改进。Jin提出将L1范数与不同的正则化项相结合,生成鲁棒的BLS(RBLS)^[6]。由于L1范数对异常值的敏感性较低,BLS的鲁棒性得到了显著提高。同样,Zheng提出了基于最大相关熵准则(maximum correntropy criterion, MCC)的BLS(C-BLS)^[7],通过设置适当的核宽度,削弱异常值对系统的影响。Guo通过融合不同的M-estimator代价函数对训练样本进行逆向加权计算提出了具有鲁棒性的RBLS^[8],以减轻离群值误差带来的不利影响。

受Zheng的启发,我们采用基于信息论学习^[9-10](information theory learning, ITL)的广义最大相关熵准则(generalized maximum correntropy criterion, GMCC)^[11]来训练BLS的输出权值,从而进一步提高BLS的性能。主要贡献总结如下:

我们提出了一种基于广义最大相关熵准则的BLS(GC-BLS)。通过设置合理的参数,可以减轻离群值误差产生的不利影响,提高BLS对于离群值的抗干扰能力。为了检验新算法的性能,在各种回归

数据集与时间序列数据集上进行了实验,并得到了一些令人满意的结果。

2 基于广义最大相关熵准则的宽度学习系统

2.1 宽度学习系统

BLS是在RVFLNN(random vector functional link neural network)基础上改进而来。对于训练数据 X ,使用特征映射 $f(XW_{ei} + \beta_{ei})$ 将数据投影,生成映射特征 Z_i 。再将前 n 组映射特征表示为 $Z_n = [Z_1 \cdots Z_n]$ 。通过非线性激活函数处理 Z_n 即 $\xi(Z_n W_{hj} + \beta_{hj})$ 生成增强节点 H_j ,其中 $W_{ei}, \beta_{ei}, W_{hj}$ 和 β_{hj} 都是随机生成的。将 m 组增强节点拼接为 $H_m = [H_1 \cdots H_m]$,合并 Z_n 与 H_m 得到最终的输入数据 $A = [Z_n | H_m]$ 。BLS的输出为 $Y = AW$,其中 W 为隐藏层到输出层的权重, A 我们称之为状态矩阵。BLS的目标函数为:

$$\arg \min_W (\|AW - Y\|_2^2 + \lambda \|W\|_2^2) \quad (1)$$

其中: Y 表示 X 的标签; $\|AW - Y\|_2^2$ 用于控制训练误差最小化; $\lambda \|W\|_2^2$ 用于防止模型过拟合; λ 是正则化系数。容易求得:

$$W = (A^T A + \lambda I)^{-1} A^T Y \quad (2)$$

2.2 广义最大相关熵准则

在相关熵中,高斯核函数并不总是最好的选择,基于广义高斯密度(generalized Gaussian density, GGD)函数,Chen等人提出了广义相关熵^[11]。给定两个随机变量 X, Y ,其广义相关熵可以表示为:

$$V_{\alpha, \beta}(X, Y) = E[G_{\alpha, \beta}(X - Y)] \quad (3)$$

其中广义高斯密度函数 $G_{\alpha, \beta}(e)$ 定义为:

$$G_{\alpha, \beta}(e) = \gamma_{\alpha, \beta} \exp(-\lambda |e|^\alpha) \quad (4)$$

其中 $\alpha > 0$ 为形状参数, $\lambda = |\beta|^\alpha > 0$ 为核参数, $\Gamma(\cdot)$ 表

示伽马函数, $\gamma_{\alpha,\beta} = \alpha/2\beta\Gamma(1/\alpha)$ 表示归一化常数。当 $\alpha=2, \beta=\sqrt{2}\sigma$ 时, 广义相关熵就退化为相关熵。对于广义最大相关熵准则, 有以下两点需要注意:

(1) 在广义相关熵中, 核函数不一定为高斯核, $G_{\alpha,\beta}(X-Y)$ 正定的条件是 $0 < \alpha \leq 2$, 此处广义相关熵的值可以用有限长样本 $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$ 的平均值来近似表示

$$V_{\alpha,\beta}(X, Y) = \frac{1}{N} \sum_{i=1}^N G_{\alpha,\beta}(\mathbf{x}_i - \mathbf{y}_i) \quad (5)$$

广义相关熵具备了相关熵的大多数性质, 部分已经被证明^[12-14]。具体如下所示:

性质 1: $V_{\alpha,\beta}(X, Y)$ 是对称的, 由此可以得出 $V_{\alpha,\beta}(X, Y) = V_{\alpha,\beta}(Y, X)$ 。

性质 2: $V_{\alpha,\beta}(X, Y)$ 是正的且有界, 其范围是 $0 < V_{\alpha,\beta}(X, Y) \leq \gamma_{\alpha,\beta}$, 只有满足 $X=Y$ 时, $V_{\alpha,\beta}(X, Y)$ 才取到最大值。

性质 3: 广义相关熵包含了误差 $|X - Y|$ 的更高阶特性:

$$V_{\alpha,\beta}(X, Y) = \gamma_{\alpha,\beta} \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} E[|X - Y|^{\alpha n}] \quad (6)$$

(2) 当广义相关熵的核参数足够小时, 可以得到 $V_{\alpha,\beta}(X, Y) \approx \gamma_{\alpha,\beta}(1 - \lambda_w E[|X - Y|^\alpha])$ 。

性质 4: 假设样本 $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$ 是从联合概率密度函数 $p_{X,Y}(\mathbf{x}, \mathbf{y})$ 中提取出来的, 设 $\hat{p}_e(\cdot)$ 为关于样本 $\{\mathbf{e}_i = \mathbf{x}_i - \mathbf{y}_i\}_{i=1}^N$ 概率密度函数的 Parzen 估计。其广义高斯密度函数 $G_{\alpha,\beta}$ 为 Parzen 窗的核函数。 $\hat{V}_{\alpha,\beta}(X, Y)$ 为 $\hat{p}_e(\cdot)$ 在零点的估计值, 表示为:

$$\hat{V}_{\alpha,\beta}(X, Y) = \hat{p}_e(0) \quad (7)$$

其中 $\hat{p}_e(\varepsilon) = \frac{1}{N} \sum_{i=1}^N G_{\alpha,\beta}(\varepsilon - \mathbf{e}_i)$ 。

性质 5: 当 $0 < \alpha \leq 2$ 时, 广义相关熵是映射到特征空间数据的二维统计量。

与相关熵类似, 广义相关熵也可以作为相关问题中的优化准则。在系统的实际输出 $\hat{Y}$ 和理想输出 Y 之间误差最小化的基础上, 使得代价函数 J_{GMCC} 的值达到最大, 即

$$\arg \max_{\mathbf{W}} J_{\text{GMCC}} = \frac{1}{N} \sum_{i=1}^N G_{\alpha,\beta}(\hat{Y} - Y) \quad (8)$$

其中 $\hat{Y} = X\mathbf{W}$, 这就是广义最大相关熵准则。

2.3 GC-BLS 的训练算法

与标准的 BLS 相同, 此处状态矩阵 A 由特征节

点与增强节点合并生成。因此, 将 BLS 和 GMCC 相结合的优化模型可以表示为:

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}} \sum_{i=1}^N G_{\alpha,\beta}(\mathbf{e}_i) - \frac{\lambda}{2} \|\mathbf{W}\|_2^2 \quad (9)$$

其中 $\mathbf{e}_i = A_i \mathbf{W} - Y_i$, 通过求解 $J(\mathbf{W})$ 对 $\mathbf{W}$ 的梯度:

$$\frac{\partial J}{\partial \mathbf{W}} = -\sum_{i=1}^N \lambda_w \alpha G_{\alpha,\beta}(\mathbf{e}_i) |\mathbf{e}_i|^{\alpha-2} A_i^T \mathbf{e}_i - \lambda \mathbf{W} \quad (10)$$

其中 λ_w 为广义相关熵中的核参数, 与正则化参数 λ 区分开; 将式(10)矩阵化可以得到:

$$\frac{\partial J}{\partial \mathbf{W}} = -\gamma_{\alpha,\beta} \lambda_w \alpha A^T \Lambda_w (A\mathbf{W} - Y) - \lambda \mathbf{W} \quad (11)$$

其中 Λ_w 为 GMCC 算子, 表示为:

$$\Lambda_w = \begin{pmatrix} G_{\alpha,\beta}(\mathbf{e}_1) |\mathbf{e}_1|^{\alpha-2} & & \\ & \ddots & \\ & & G_{\alpha,\beta}(\mathbf{e}_N) |\mathbf{e}_N|^{\alpha-2} \end{pmatrix} \quad (12)$$

令式(11)等于零, $\mathbf{W}$ 的解可以写成式(13)形式:

$$\mathbf{W} = (A^T \Lambda_w A + \lambda_{\text{GMCC}} I)^{-1} A^T \Lambda_w Y \quad (13)$$

为了保持公式的简洁, 在式(13)中将众多参数进行整合, 其中 $\lambda_{\text{GMCC}} = \lambda/\gamma_{\alpha,\beta} \lambda_w \alpha$, 显然, Λ_w 是 $\mathbf{W}$ 的函数, 因此式(13)实际上是一个不动点方程, 可以将其表示为:

$$\mathbf{W} = f(\mathbf{W}) \quad (14)$$

$$f(\mathbf{W}) = (A^T \Lambda_w A + \lambda_{\text{GMCC}} I)^{-1} A^T \Lambda_w Y \quad (15)$$

具体算法流程如表 1 所示:

表 1 GC-BLS 算法流程

Tab. 1 GC-BLS algorithmic flow

GC-BLS 算法流程
1. 输入: 训练样本 $\{X, Y\}$ 。
2. 参数设置: 网络节点参数 k, n, q, m , 正则化参数 λ , 核参数 λ_w , 终止公差 ε , 最大循环次数 T 。
3. 初始化: 设置初始权重 $\mathbf{W}_0$, 构建输入矩阵 A 。
4. 循环运算: $t = 1, \dots, T$, 根据式(12)计算加权矩阵 Λ_w ; 根据式(16)更新权重 $\mathbf{W}$, 直到 $\ \mathbf{W}_{t+1} - \mathbf{W}_t\ _2^2 < \varepsilon$ 。
5. 输出: 权重矩阵 $\mathbf{W}$ 。
6. 结束。

表 1 中, k 表示每个特征映射组中特征节点个数, n 表示特征映射组数, q 表示每个增强节点组中增强节点个数, m 表示增强节点组数, 参考定点迭代方法^[14-17], 通过式(16)求解 $\mathbf{W}$:

$$\mathbf{W}_{t+1} = f(\mathbf{W}_t) \quad (16)$$

其中, $\mathbf{W}_t$ 表示迭代 t 时的解。设 ε 表示终止公差。

停止准则可设置为 $|\mathbf{W}_{t+1} - \mathbf{W}_t|_2^2 < \varepsilon$,表示当权重更新趋势趋向于稳定时,停止迭代。

3 实验分析

本节中,以均方根误差(root mean square error, RMSE)为衡量标准评估各算法在回归任务上的性能。为了消除数据尺度的影响,下面所有数据集的输入和输出属性都被归一化到(0,1)的范围。所有的实验结果都是在配备 Intel(R) Core(TM) i5-6200U CPU 2.30 GHz 和 8 GB RAM 的机器上使用 pycharm 获得的。

3.1 GC-BLS在时间序列数据集上的表现

本节中使用麦基-格拉斯时间序列预测和月平均总“太阳黑子数”预测进行实验,实验中给两者 30% 训练集标签上添加范围在 $y_{\min}$ 到 $y_{\max}$ 的离群值噪声,以验证算法性能。

麦基-格拉斯时间序列显示了混沌动力学的特征,可以用式(17)^[18]所示的时滞微分方程来描述:

$$\frac{dx(t)}{dt} = -bx(t) + \frac{ax(t-\tau)}{1+x(t-\tau)^{10}} \quad (17)$$

其中,本文中设置 $a=0.1$; $b=0.2$; $\tau=30$ 。正如文献[17]中所建议的,采用前7点来预测当前的一点。因此 t 时刻的输入和输出的形式分别为 $x_t = [x(t-7), x(t-6) \cdots x(t-1)]$ 和 $y_t = x(t)$ 。

“太阳黑子数量”是一个实时系列数据集,记录了1749年1月至2017年12月每月平均太阳黑子的数量。实验中的嵌入维数和延迟时间分别设置为4和1。

数据集参数如表2所示。由表3中的实验结果可知,与其他算法相比,新算法的性能要明显优于其他算法。

表2 时间序列数据集
Tab. 2 Time series data set

数据集	训练集	测试集	(d, τ)
Mackey-glass	770	330	(7,1)
Sunspot	1974	846	(4,1)

3.2 GC-BLS在回归数据集上的表现

本节在UCI(university of california irvine)基准数据集上进行实验。数据集参数如表4所示。由于这些数据是现实世界的真实数据,已经包含误差,

表3 各算法时间序列数据集上的表现

Tab. 3 Performance of each algorithm on time series data set

数据集	算法	参数组合	训练时间/s	测试结果
Mackey-glass	BLS	(5,1,160)	0.0209~0.0033	0.1234~0.0007
	RBLS(huber)	(13,15,180)	0.4229~0.0080	0.1306~0.0009
	C-BLS	(5,1,140, 2 ¹)	0.0635~0.0054	0.1233~0.0009
	GC-BLS	(1,15,100,3.5, 1.5)	1.6482~0.0349	0.1083~0.0013
Sun-spot	BLS	(1,1,100)	0.0871~0.0080	0.0877~0.0019
	RBLS(huber)	(1,1,100)	0.5037~0.0430	0.1002~0.0084
	C-BLS	(3,1,100, 2 ³)	0.3564~0.0205	0.0891~0.0002
	GC-BLS	(1,1,100,3,1)	2.3786~0.4281	0.0834~0.0001

表4 回归数据集

Tab. 4 Regression data set

数据集	属性数	训练集	测试集
Abalone	8	997	427
Bodyfat	14	177	75
Cancer	30	399	171
Winequality white	11	780	333
Heart disease cleveland	13	213	91
Slump	7	73	31
yacht hydrodynamics	6	216	92
bike sharing	14	511	221
combined cycle power plant	4	1210	520

不额外添加人工噪声。

首先采用网格搜索法确定各算法的最优参数。对于所有算法的网络结构参数, m 固定为1,参数组合 $\{k, n, q\}$ 的搜索范围分别设置为 $\{1, 3, 5, 7, 9, 11, 13, 15\}$, $\{1, 3, 5, 7, 9, 11, 13, 15\}$, $\{100, 120, 140, 160, 180, 200\}$ 。GC-BLS的参数组合为 $\{k, n, q, \alpha, \lambda_w\}$,形状参数 α 的从候选集 $\{1, 1.5, 2, 2.5, 3, 3.5, 4\}$ 中选择;核参数 λ_w 从候选集 $\{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$ 中选择。C-BLS的参数组合为 $\{k, n, q, \sigma\}$,参考文献[7],核参数 σ 从候选集 $\{2^{-3}, 2^{-1}, 2^0, 2^1, 2^3\}$ 中选择。

所有算法均采用统一的正则化系数,正则化系数 λ 固定为 10^{-5} 。在各自最佳参数的条件下实验,每种算法独立运行 20 次,以运行结果的平均值作为最终实验结果。由表 5 中的数据可知,GC-BLS 与标准 BLS 和其他 BLS 鲁棒算法相比,虽然 GC-BLS 的训练时间略高于其他算法,但总能获得更小的测试 RMSE,表现出很强的鲁棒性。新算法的性能要明显优于其他算法。

表 5 各算法在回归数据集上的表现
Tab. 5 Performance of each algorithm on regression data set

数据集	算法	参数组合	训练时间/s	测试结果
Abalone	BLS	(3,1,140)	0.0279~0.0032	0.0868~0.0005
	RBLS(huber)	(1,15,200)	0.2889~0.0256	0.0865~0.0014
	C-BLS	(3,1,140, 2 ³)	0.0781~0.0025	0.0863~0.0008
	GC-BLS	(5,1,140,4,2)	2.9757~0.1052	0.0686~0.0002
Bodyfat	BLS	(1,15,140)	0.0668~0.0012	0.1239~0.0015
	RBLS(huber)	(7,1,140)	0.0339~0.0025	0.1227~0.0056
	C-BLS	(1,9,140, 2 ⁻³)	0.0469~0.0003	0.1220~0.0009
	GC-BLS	(11,1,100,4,1)	0.0469~0.0011	0.1061~0.0010
Cancer	BLS	(15,1,140)	0.0060~0.0001	0.1368~0.0021
	RBLS(huber)	(5,3,200)	0.1587~0.0035	0.1511~0.0019
	C-BLS	(15,9,200, 2 ⁻³)	0.4845~0.0007	0.1329~0.0003
	GC-BLS	(7,11,100,1,1.5)	0.1037~0.0014	0.1316~0.0008
Winequality white	BLS	(13,1,120)	0.0229~0.0012	0.0912~0.0005
	RBLS(huber)	(15,13,200)	0.3411~0.0002	0.0947~0.0013
	C-BLS	(13,1,120, 2 ⁰)	0.1147~0.0005	0.0913~0.0006
	GC-BLS	(9,15,100,1,3)	0.4338~0.0014	0.0910~0.0002
Heart disease cleveland	BLS	(9,11,200)	0.0715~0.0012	0.2071~0.0020
	RBLS(huber)	(13,15,200)	0.1099~0.0125	0.2195~0.0035
	C-BLS	(13,13,100, 2 ³)	0.0469~0.0032	0.2067~0.0014
	GC-BLS	(13,13,140,3.5,3)	1.5921~0.0458	0.2052~0.0009
Slump	BLS	(13,9,180)	0.0608~0.0002	0.1446~0.0011
	RBLS(huber)	(15,13,200)	0.1556~0.00155	0.1895~0.0025
	C-BLS	(5,15,120,2 ⁻¹)	0.0808~0.0003	0.1417~0.0013
	GC-BLS	(5,15,12,1.5,3.5)	0.1017~0.0028	0.1412~0.0020
yacht hydrodynamics	BLS	(13,9,160)	0.0359~0.0011	0.1046~0.0009
	RBLS(huber)	(15,15,140)	0.0781~0.0022	0.1464~0.0017
	C-BLS	(13,9,160, 2 ⁰)	0.0691~0.0010	0.1044~0.0005
	GC-BLS	(13,9,160,2,4)	0.1855~0.0015	0.1038~0.0007
bike sharing	BLS	(9,7,200)	0.0519~0.0008	0.0989~0.0005
	RBLS(huber)	(13,15,120)	0.7919~0.0029	0.0979~0.0012
	C-BLS	(7,3,140, 2 ⁻³)	0.0691~0.0030	0.0841~0.0009
	GC-BLS	(3,3,120,1,3.5)	1.3155~0.0055	0.0835~0.0015
combined cycle power plant	BLS	(7,7,100)	0.0403~0.0047	0.0829~0.0021
	RBLS(huber)	(15,15,200)	1.9830~0.0125	0.0830~0.0026
	C-BLS	(1,11,140, 2 ⁻³)	0.9280~0.0108	0.0798~0.0008
	GC-BLS	(15,7,140,1,3.5)	2.9573~0.0526	0.0796~0.0017

3.3 GC-BLS的鲁棒性分析

与传统 BLS 的权重求解方式(2)相比,新算法式(13)加入了一个加权对角矩阵 $\mathbf{A}_w$,这也是新算法鲁棒性的由来,其对角元素由核参数 λ_w 以及实际输出 $\hat{y}$ 和理想输出 y 之间误差控制。可以验证,当 λ_w 趋向于 0 时,式(13)将退化为标准 BLS 的权重求解方式(2);当 $\alpha=2, \beta=\sqrt{2}\sigma$ 时,GMCC 准则退化为 MCC 准则,这说明,GC-BLS 至少可以实现与 C-BLS 和 BLS 相当的性能。当第 i 个样本受到离群值的污染时, $\hat{y}_i$ 和 y_i 之间通常会有很大的差异,记为 $v_i = |\hat{y}_i - y_i|$ 。正常情况下,通过设置合理的 λ_w 值,当 v_i 越大,说明该值偏离程度越大,在 λ_w 值的作用下,相应的对角元素越小,从而该离群值在训练过程中将被赋予较小的权重,这使得离群值对训练过程没有很大的影响。

4 结论

为了解决 BLS 对离群值噪声敏感的问题,参考 Zheng 提出的 C-BLS 算法,本文采用更具有灵活性的广义最大相关熵准则(GMCC)提出了新的鲁棒 BLS (GC-BLS)。GC-BLS 以 BLS 模型为基础,通过引入 GMCC 得到新的代价函数,并采用不动点迭代方法对学习样本进行逆向加权计算,从而抑制离群值样本带来的不良影响。在 9 个回归数据集和 2 个时间序列数据集上的实验表明,当样本包含离群值时,新算法能够取得较 BLS 与其他改进算法更优的学习性能。目前新算法涉及到的参数数量较多,利用网格搜索法选择最佳参数非常耗时,如何简化算法参数,或者在训练过程中尝试寻找某些参数的最优值,从而进一步提高系统效率是下一步的研究工作。

参考文献

- [1] CHEN C L P, LIU Zhulin. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(1): 10-24.
- [2] LI Ping, SHENG Bin, CHEN C L P. Face sketch synthesis using regularized broad learning system [J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(10): 5346-5360.
- [3] DU Jie, VONG C M, PHILIP CHEN C L. Novel efficient RNN and LSTM-like architectures: Recurrent and gated broad learning systems and their applications for text classification [J]. IEEE Transactions on Cybernetics, 2021, 51(3): 1586-1597.
- [4] HAN Min, LI Weijie, FENG Shoubo, et al. Maximum information exploitation using broad learning system for large-scale chaotic time-series prediction [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(6): 2320-2329.
- [5] 朱迎莹,赵海全.非线性主动噪声控制的随机傅里叶特征-核滤波算法[J].信号处理,2020,36(6): 984-990.
ZHU Yingying, ZHAO Haiquan. Random Fourier feature kernel filtered-x LMS algorithm for nonlinear active noise control system [J]. Journal of Signal Processing, 2020, 36(6): 984-990. (in Chinese)
- [6] JIN Junwei, PHILIP CHEN C L. Regularized robust broad learning system for uncertain data modeling [J]. Neurocomputing, 2018, 322: 58-69.
- [7] ZHENG Yunfei, CHEN Badong, WANG Shiyuan, et al. Broad learning system based on maximum correntropy criterion [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(7): 3083-3097.
- [8] 郭威,徐涛.基于 M-estimator 的鲁棒宽度学习系统[J].控制与决策,2023,38(4): 1039-1046.
GUO Wei, XU Tao. M-estimator-based robust broad learning system [J]. Control and Decision, 2023, 38(4): 1039-1046. (in Chinese)
- [9] 赵海全,陈奕达.广义最大总体相关熵自适应滤波算法[J].信号处理,2021,37(8): 1378-1383.
ZHAO Haiquan, CHEN Yida. Generalized maximum total correntropy adaptive filtering algorithm [J]. Journal of Signal Processing, 2021, 37(8): 1378-1383. (in Chinese)
- [10] 宋普查,赵海全.基于最大混合相关熵准则的主动噪声控制算法[J].信号处理,2020,36(6): 942-947.
SONG Pucha, ZHAO Haiquan. Active noise control algorithm based on maximum mixture correntropy criterion [J]. Journal of Signal Processing, 2020, 36(6): 942-947. (in Chinese)
- [11] CHEN Badong, XING Lei, ZHAO Haiquan, et al. Generalized correntropy for robust adaptive filtering [J]. IEEE Transactions on Signal Processing, 2016, 64(13): 3376-3387.
- [12] ERDOGMUS D, PRINCIPE J C. An error-entropy mini-

- mization algorithm for supervised training of nonlinear adaptive systems [J]. IEEE Transactions on Signal Processing, 2002, 50(7): 1780-1786.
- [13] CHEN Badong, ZHU Yu, HU Jinchun. Mean-square convergence analysis of ADALINE training with minimum error entropy criterion [J]. IEEE Transactions on Neural Networks, 2010, 21(7): 1168-1179.
- [14] ALGHAMDI M A, PETRUŞEL A, SHAHZAD N. Correction: A fixed point theorem for cyclic generalized contractions in metric spaces[J]. Fixed Point Theory and Applications 2012, 2012(1): 122-122.
- [15] CHEN Badong, WANG Jianji, ZHAO Haiquan, et al. Convergence of a fixed-point algorithm under maximum correntropy criterion[J]. IEEE Signal Processing Letters, 2015, 22(10): 1723-1727.
- [16] HERAVI A R, ABED HODTANI G. A new robust fixed-point algorithm and its convergence analysis [J]. Journal of Fixed Point Theory and Applications, 2017, 19(4): 3191-3215.
- [17] CHEN Badong, XING Lei, WANG Xin, et al. Robust learning with kernel mean p-power error loss [J]. IEEE Transactions on Cybernetics, 2018, 48(7): 2101-2113.
- [18] GUO Yu, WANG Fei, CHEN Badong, et al. Robust echo state networks based on correntropy induced loss function[J]. Neurocomputing, 2017, 267: 295-303.

作者简介



赵海全 男, 1974年生, 河南人。西南交通大学电气工程学院教授, 博士生导师, 主要研究方向为统计信号处理、人工智能算法等。

E-mail: hqzhao@swjtu.edu.cn



陆 鑫 男, 1997年生, 江苏人。西南交通大学电气工程学院研究生, 研究方向为宽度学习系统、自适应滤波。

E-mail: 17695794976@163.com