

# 基于截断最大相关熵准则损失函数的鲁棒极限学习机

王晶\*, 陈志祥

绍兴文理学院数理信息学院, 浙江 绍兴

收稿日期: 2023年6月25日; 录用日期: 2023年7月19日; 发布日期: 2023年7月26日

## 摘要

极限学习机(Extreme Learning Machine, ELM)在前馈神经网络中具有高效快速的优势。然而ELM在面对异常值时较敏感, 鲁棒性较差。本文对最大相关熵损失函数进行了截断, 将最大相关熵损失函数进一步改进为能够限定最大损失为常数的截断最大相关熵损失函数, 构建基于截断最大相关熵准则损失函数的鲁棒ELM模型, 以此来抑制噪声和异常值对模型的影响。采用迭代重加权算法对模型进行求解。最后在数据集上验证所提出的模型的有效性。实验结果表明, 该模型在噪声条件下提高了对噪声和离群值的鲁棒性。

## 关键词

最大相关熵, 损失函数, 截断, 极限学习机, 鲁棒性

# Robust Extreme Learning Machine Based on Truncated Maximum Correlation Entropy Criterion Loss Function

Jing Wang\*, Zhixiang Chen

School of Mathematical Information, Shaoxing University, Shaoxing Zhejiang

Received: Jun. 25<sup>th</sup>, 2023; accepted: Jul. 19<sup>th</sup>, 2023; published: Jul. 26<sup>th</sup>, 2023

## Abstract

Extreme Learning Machine (ELM) has the advantage of being efficient and fast in feed forward

\*通讯作者。

neural networks. However, ELM is sensitive to outliers and has poor robustness. In this paper, the maximum correlation entropy loss function is truncated, and the maximum correlation entropy loss function is further improved to a truncated maximum correlation entropy loss function that can limit the maximum loss as a constant. A robust ELM model based on the truncated maximum correlation entropy criterion loss function is constructed to suppress the influence of noise and outliers on the model. The iterative reweighting algorithm was used to solve the model. Finally, the validity of the proposed model is verified on the data set. Experimental results show that the robustness of the model to noise and outliers is improved under noisy conditions.

## Keywords

Maximum Correlation Entropy, Loss Function, Truncation, Extreme Learning Machine, Robustness

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

极限学习机(Extreme Learning Machine, ELM)是由黄等人在文献[1]中提出的一种简单高效的单隐层前馈神经网络算法。ELM 随机选取网络的内权和隐层偏置,采用均方误差(Mean Squared Error, MSE)作为损失函数,并通过求解线性方程组得到唯一的解析解,因此克服了传统神经网络的一些不足,已被广泛应用于医学[2]、生物学[3]、故障诊断[4]、图像分类[5]、荷电状态估计[6]等领域。

由于 ELM 采用 MSE 作为损失函数,当样本中存在噪声和离群值时,离群值会放大对参数估计的影响,从而导致模型的敏感性和鲁棒性较差。因此,极限学习机的许多变体都致力于提高对极限学习机鲁棒性的研究。对于残差较大的异常点,通常采用迭代修正的方式来减少异常值的影响[7],以获得鲁棒的极限学习机模型。当数据集中添加了不同程度的高斯噪声时,可以采用自编码器对输入数据进行重构,将隐层输出值关于输入的雅克比矩阵的 F 范数引入目标函数中,以提取更鲁棒的抽象特征[8]。

当然,还能从损失函数的角度出发,对损失函数进行变体以提高极限学习机的鲁棒性能。文献[9]提出了一种基于指数 Laplace 损失函数回归估计鲁棒极限学习机,由于指数 Laplace 损失函数具有非负、对称、有界等特点,能够有效地提高预测精度。文献[10]建立了基于 1-范数损失函数和 2-范数正则化项的鲁棒 ELM 模型,可以处理高维数据,且对不平衡的数据集也能产生良好的结果。文献[11]提出了四种不同的损失函数(1-范数、Huber、bisquare 和 Welsch),采用了 1-范数和 2-范数正则化项防止过拟合,提出了统一的鲁棒正则化 ELM。文献[12]提出了基于正则化相关熵准则的 ELM,采用半二次规划求解所建模型,通过在人工数据集上进行仿真实验,实验结果表明正则化相关熵准则 ELM 优于传统 ELM 和正则化 ELM。文献[13]提出了相关熵融合极限学习机,将核映射与系数加权相结合,用相关熵损失函数替代均方误差损失函数,推导出相关熵循环更新公式,增强了它的鲁棒性。文献[14]提出了基于最大相关熵准则的多尺度高斯核极限学习机,用最大相关熵准则代替多尺度核极限学习机中传统的最小均方误差准则构造目标函数,推导出基于最大相关熵准则的多尺度高斯核极限学习机,进而提高其鲁棒性。文献[15]提出了一种基于 pinball 损失函数和 L1 范数正则化的 ELM 模型,以实现鲁棒性和稀疏性的提高。

受上述研究的启发,本文在最大相关熵准则损失函数的基础上,采用截断思想,提出了截断最大相关熵损失函数,并将其引入到 ELM 框架中,建立了能够降低噪声和异常值影响的基于截断最大相关熵

准则的鲁棒极限学习机 (Truncated Maximum Correntropy Criterion Extreme Learning Machine, TMCCELM), 实验结果表明当数据中异常值过多且残差较大时, 本文 TMCCELM 的结果优于对比的几种鲁棒极限学习机算法, 具有较好的鲁棒性和较高的预测精度。

## 2. ELM 模型

给定  $N$  个训练样本的数据集  $\{(x_i, y_i)\}_{i=1}^N$  ( $i=1, 2, \dots, N$ ),  $x_i \in R^n, y_i \in R$ , 具有  $L$  个隐藏节点和激活函数  $g_i(x)$  的数学表达式为:

$$f(x) = \beta_i G(\omega_i x_i + b_i), i=1, 2, \dots, N \quad (1)$$

其中  $\omega_i$  是输入层和第  $i$  个隐藏层神经元之间的输入权值,  $\beta_i$  是连接第  $i$  个隐藏层神经元与输出层之间的输出权值,  $\omega_i \cdot x_i$  表示  $\omega_i$  与  $x_i$  的内积, 并且输出节点被选择为线性的, ELM 网络中隐含层的映射生成的中间矩阵为:

$$H[\omega_1, \dots, \omega_L, x_1, \dots, x_N, b_1, \dots, b_L] = \begin{bmatrix} G(\omega_1 \cdot x_1 + b_1) & \cdots & G(\omega_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ G(\omega_1 \cdot x_N + b_1) & \cdots & G(\omega_L \cdot x_N + b_L) \end{bmatrix}$$

其中  $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T$  表示输出层和隐层之间的权值矩阵,  $y = [y_1, y_2, \dots, y_N]^T$  表示训练样本期望输出矩阵, 式(1)可以写成如下形式:

$$H\beta = y$$

$H\beta = y$  的最小范数二乘解  $\hat{\beta}$ :

$$\hat{\beta} = H^+ y$$

其中  $H^+$  表示隐层输出矩阵  $H$  的 Moore-Penrose 广义逆。

## 3. T-MCC-ELM 模型

### 3.1. 截断最大相关熵准则损失函数

相关熵是一种用来描述两组随机变量相似性的度量, 已被广泛应用于信号处理和鲁棒学习领域[16]。对于给定任意两个随机变量  $A, B$ , 它们的相关熵可以定义为:

$$V_\sigma(A, B) = E[K_\sigma(A, B)],$$

其中  $K_\sigma(\cdot)$  是满足 Mercer 理论的核函数  $K_\sigma = \exp\left(-\frac{x^2}{2\sigma^2}\right)$ ;  $\sigma$  表示核函数宽度;  $E$  表示数学期望。在实践中, 联合概率密度通常是未知的, 并且只有有限数量的数据可用, 这导致了相关熵有以下的样本估计:

$$V_\sigma = \frac{1}{M} \sum_{i=1}^M K_\sigma(A_i, B_i),$$

其中  $K_\sigma$  是高斯核函数, 对于变量  $A = [a_1, a_2, \dots, a_N]$  和变量  $B = [b_1, b_2, \dots, b_N]$ ,  $A$  和  $B$  的差值可以表示为  $S = A - B = (s_1, s_2, \dots, s_N)$ , 其中  $s_i = a_i - b_i$ 。上式相关熵的最大值称为最大相关熵(Maximum Correntropy Criterion, MCC), 可以表示为[17]:

$$\max \frac{1}{M} \sum_{i=1}^M K_\sigma(s_i) = \max \frac{1}{M} \sum_{i=1}^M \exp\left(-\frac{s_i^2}{2\sigma^2}\right).$$

在 MCC 下, 可以将问题转换为实际值  $y$  与预测值  $f$  之间的相似性问题, 即

$$\max V_{\sigma}(y, f) = \exp\left(-\frac{s^2}{2\sigma^2}\right).$$

其对偶问题为:

$$\min 1 - \exp\left(-\frac{s^2}{2\sigma^2}\right).$$

综上所述, 在最大相关熵损失函数基础上, 提出了截断最大相关熵损失函数:

$$l_{\theta}(s_i) = \begin{cases} 1 - \exp\left(-\frac{s^2}{2\sigma^2}\right), & 1 - \exp\left(-\frac{s^2}{2\sigma^2}\right) \leq \theta, \\ \theta, & 1 - \exp\left(-\frac{s^2}{2\sigma^2}\right) > \theta. \end{cases}$$

其中  $\theta > 0$  为预先给定的异常值的最大惩罚,  $l_{\theta}(s_i)$  的上界表明当误差超过某定值时, 损失函数  $l_{\theta}(s_i)$  值为常数, 能够抑制异常值的影响。

### 3.2. 基于截断最大相关熵准则损失函数的鲁棒极限学习机

基于截断最大相关熵准则损失函数, 本文建立具有抑制异常值影响的鲁棒 ELM 模型, 其对应的优化模型具有如下形式:

$$\begin{aligned} \min & \frac{1}{2} \|\beta\|_2^2 + c \sum_{i=1}^N l_{\theta}(s_i), \\ \text{s.t.} & h(x_i)\beta = y_i - s_i, i=1, 2, \dots, N. \end{aligned} \quad (2)$$

其中第一部分是为了防止过拟合引入的  $L_2$  正则化函数; 第二部分是 T-MCC 函数,  $C$  表示正则化参数。为了解决(1)的二次规划问题, 这里引入权重矩阵的  $w(s_i)$ , 构造如下拉格朗日函数:

$$l(\beta, s_i, \alpha) = \frac{1}{2} \|\beta\|_2^2 + c \sum_{i=1}^N l_{\theta}(e_i) - \sum_{i=1}^N \alpha_i (h(x_i)\beta - y_i + s_i) \quad (3)$$

其中  $\alpha_i$  表示每个样本所对应的拉格朗日乘法法中的参数变量, 对式(2)中每个参数求偏导数, 令偏导数为 0, 可得:

$$\begin{aligned} \frac{\partial L}{\partial \beta} = 0 & \Rightarrow \beta = \sum_{i=1}^N \alpha_i h(x_i), \\ \frac{\partial L}{\partial s_i} = 0 & \Rightarrow \alpha_i = C \frac{\partial l_{\theta}(s_i)}{\partial s_i} = c w_i s_i, \\ \frac{\partial L}{\partial \alpha_i} = 0 & \Rightarrow h(x_i)\beta - y_i + s_i = 0 \Rightarrow H\beta - y + s = 0, \end{aligned} \quad (4)$$

其中

$$\begin{aligned} \alpha &= [\alpha_1, \alpha_2, \dots, \alpha_N]^T, \\ w_i &= \frac{\partial l_{\theta}(s_i)}{s_i} = \begin{cases} \frac{1}{\sigma^2} \exp\left(-\frac{s^2}{2\sigma^2}\right), & \left[1 - \exp\left(-\frac{s}{2\sigma^2}\right)\right]^2 \geq \theta, \\ 0, & \left[1 - \exp\left(-\frac{s}{2\sigma^2}\right)\right]^2 < \theta, \end{cases} \\ s &= [s_1, s_2, \dots, s_N]^T. \end{aligned} \quad (5)$$

由(4)可得隐含层输出权重  $\beta$ , 其最优解  $\beta$ :

$$\beta = \begin{cases} \left( H^T W H + \frac{I}{C} \right)^{-1} H^T W y, N \geq L \\ H^T \left( W H H^T + \frac{I}{C} \right)^{-1} W y, N < L \end{cases}$$

其中对角矩阵  $W = \text{diag}\{w_1, w_2, \dots, w_N\}$ 。

#### 4. 迭代重加权算法及 T-MCC-ELM 算法的设计步骤

为了得到  $\beta$  的最优权重, 同时兼顾模型的训练速度, 本文采用了迭代重加权算法。单个样本的权重可表示为  $W(s) = (\partial L(s) / \partial s) / s$ , 即

$$W(s) = \frac{1}{\sigma^2} \exp\left(-\frac{s^2}{2\sigma^2}\right). \quad (6)$$

因此,  $N$  个样本的权重可表示为:

$$W(s) = \text{diag}(w(s_1), w(s_2), \dots, w(s_N)).$$

T-MCC-ELM 算法设计

输入: 训练集  $\{(x_i, y_i)\}_{i=1}^N$ , 正则化参数  $C$ , 隐含层节点  $L$ , 最大迭代次数  $k_{\max}$  和迭代停止条件  $\rho$ ,  $\rho > 0$ , 计算隐藏层输出矩阵  $H$ , 令  $W^{(0)} = I$  和  $k = 1$ 。

输出:  $\beta$ 。

步骤 1: 根据下式, 计算  $\beta^{(k)}$

$$\beta = \begin{cases} \left( H^T W H + \frac{I}{C} \right)^{-1} H^T W y, N \geq L, \\ H^T \left( W H H^T + \frac{I}{C} \right)^{-1} W y, N < L, \end{cases}$$

由表达式 (1), 可得  $f^k(x)$ ;

步骤 2:  $s_i^{(k)} = y_i - f^{(k)}(x_i)$  设置对角矩阵  $W^{(k)}$ , 其中  $W_i^{(k)}$  为对角元素;

步骤 3: 令  $k = k + 1$ , 计算  $\beta^{(k+1)}$ ;

步骤 4: 若  $k < k_{\max}$ , 且  $\|\beta^{(k+1)} - \beta^k\| > \rho$ , 则转步骤 5;

步骤 5: 由式 (1) 得  $f^{(k+1)}(x)$ , 令  $k = k + 1$ , 转步骤 2。

#### 5. 实验

下面的实验设计旨在验证 TMCCELm 在不同异常值存在情况下的鲁棒性能和计算效率。本文给出了 TMCCELm 和 ELM、RELM、WRELM、RCCELm 算法的实验结果, 并通过统计测试比较了这些算法的性能。本实验的数据集分为训练集和测试集, 选择 sigmoid 函数  $g(x) = 1/(1 + \exp(-x))$  作为激活函数, TMCCELm 的最大迭代次数设置为 20, 且在 Matlab9.10(R2021a)的环境下进行。

人工数据集采用回归问题中广泛使用的 sinc 函数作为目标函数的近似, 定义如下:

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

对于每个数据点  $(x, y)$ , 用于训练的  $x$  以统一的步长在  $[-1, 1]$  中选取, 用于测试的  $x$  是从区间  $[-10, 10]$  中选取。为了全面评估异常值的鲁棒性, 我们从  $[-10, 10]$  中随机选取扰动值添加在  $y$  的测试点上, 且服从均匀分布, 值得注意的是训练数据不包含异常值。为了揭示 TMCCELM 算法的鲁棒性, 在不同水平异常点(包括 0%、10%、20%、……、80%)数据集上分别进行了实验, 在具有不同异常值水平的噪声环境下, 对比了几个改进的极限学习机的鲁棒性。

为了更清楚地比较这些算法的性能, 下面给出 5 种算法在异常值 10%、30%、50%、80% 下的 sinc 函数的回归图像, 实验结果如下:

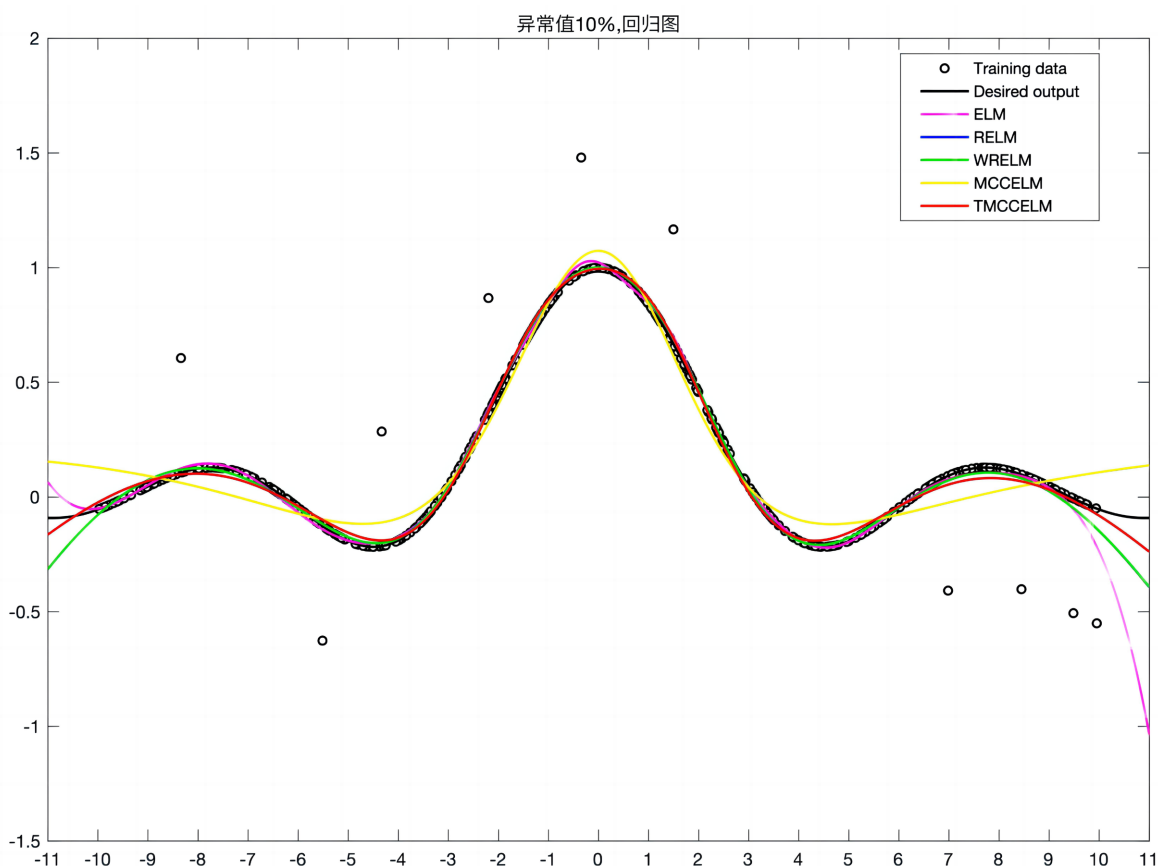


Figure 1. Sinc regression image of outlier 10%

图 1. 异常值 10% 的 sinc 回归图像

从实验结果我们可以看出, ELM、RELM、WRELM 在曲线两端出现波动, 这是由于异常值的出现引入了额外的噪声, 从而干扰了 ELM 的训练过程, 导致在曲线两端拟合时出现波动。RCCELM 模型在峰值处有较小的波动, 表明最大相关熵损失函数在异常值存在的情况下表现相对于另外四种算法较感, 通过观察图像两端发现, TMCCELM 算法在异常值 10% 下的整体拟合效果较好。

当异常值增加至 30% 时, MCCELM 和 ELM 均出现了较大的波动, ELM 在右侧端点处对异常值的敏感程度较为明显。RELM、WRELM 在端点处对异常值的敏感度增加, 由于 TMCCELM 通过采用截断思想对模型进行了改进, 这种改进使 TMCCELM 在处理包含异常值的数据时更具有鲁棒性, 尤其体现在曲线两端拟合时减少了波动。

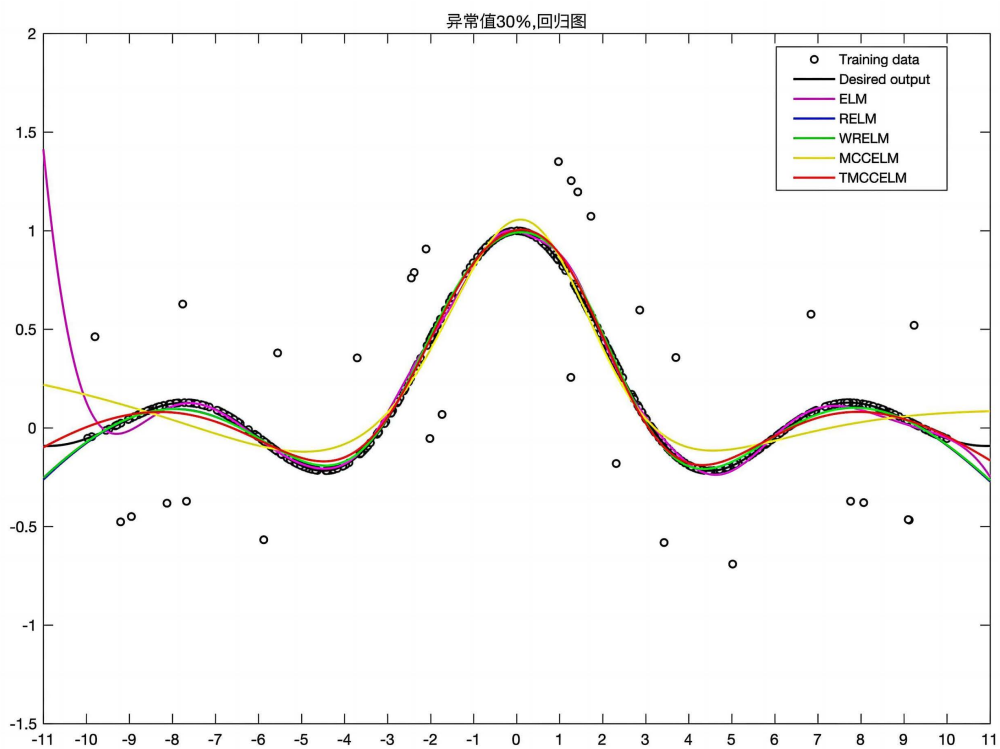


Figure 2. Sinc regression image of outlier 30%  
图 2. 异常值 30%的 sinc 回归图像

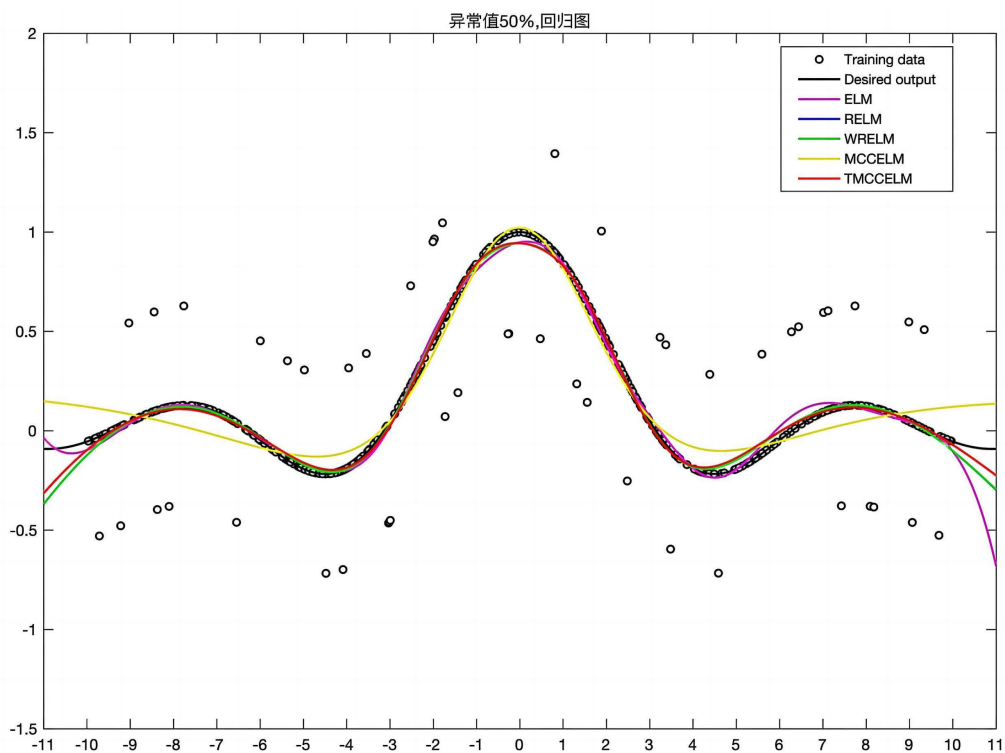


Figure 3. Sinc regression image of outlier 50%  
图 3. 异常值 50%的 sinc 回归图像



当异常值增加至 50% 时, RELM 和 WRELM 由于引入了正则化和加权处理, 仍然在端点处出现波动, 并且在曲线其他部分并不是完全拟合, ELM 在左侧端点处出现较大的波动, 表现其在异常值 50% 的情形下对异常值的干扰更为敏感, TMCCELM 在左侧端点处出现轻微波动, 在曲线其余部分达到了完美拟合的效果。

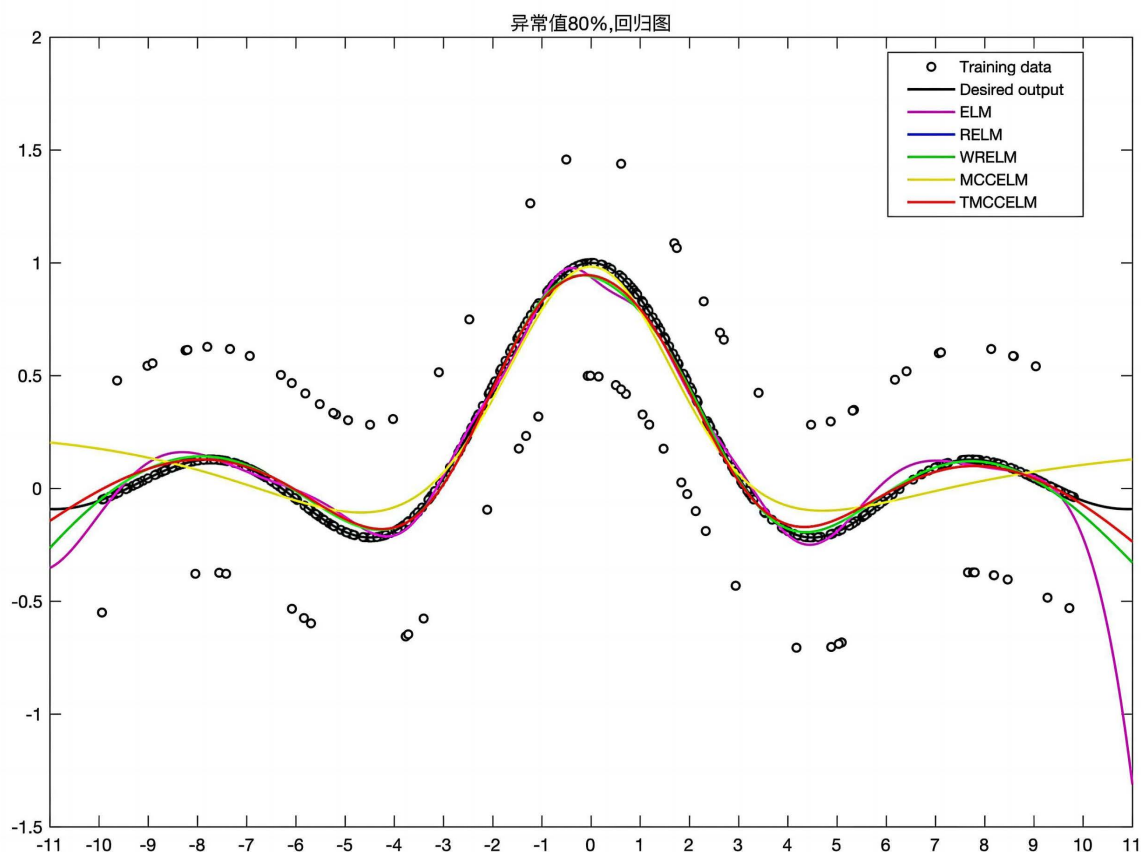


Figure 4. Sinc regression image of outlier 80%

图 4. 异常值 80% 的 sinc 回归图像

当异常值达到 80% 时, 五种模型出现了不同程度的波动情况, 其中 MCCELM 表现较为明显, ELM、RELM、WRELM、MCCELM 曲线部分偏离原始曲线, 朝向异常点, 而 TMCCELM 的曲线始终最接近原始曲线, 证明 TMCCELM 对异常值不敏感, 从而验证了该模型的可行性以及对待噪声和异常值的鲁棒性能。

以上图 1~4 分别显示了五种不同的算法在受不同数量的异常值影响下的测试数据的性能。从这些图中, 可以得到以下结果:

- 1) ELM 和 MCCELM 导出的曲线被推向异常值的一侧;
- 2) 在异常值较大的情况下, MCCELM 有偏离 sinc 曲线的趋势;
- 3) TMCCELM 在这些算法中取得了最佳性能, 相对于 ELM、RELM、MCC-ELM 和 WRELM, 其对异常值更鲁棒。

为了进一步评估这五种模型的鲁棒性能, 接下来给出这五种算法在不同异常值下的训练 RMSE 和测试 RMSE, 如表 1~4 所示:



**Table 1.** RMSE of different algorithms with 10% outliers**表 1.** 不同算法在 10%异常值下的 RMSE

算法	训练 RMSE	测试 RMSE
ELM	0.3356	0.0698
RELM	0.3349	0.0709
WRELM	0.2989	0.0633
MCCELM	0.0016	0.0005
TMCCELM	0.0006	0.0001

从训练 RMSE 来看, TMCCELM 的表现最佳为 0.0006, 其次是 RCCELM, WRELM 通过加权和正则化处理, 相对于 ELM、RELM 表现较好。从测试 RMSE 来看, TRCCELM 和 MCCELM 相对于另外三个模型表现较好, 表明 TMCCELM 和 MCCELM 在测试集上能够获得更准确的预测结果。

**Table 2.** RMSE of different algorithms under 20% outliers**表 2.** 不同算法在 20%异常值下的 RMSE

算法	训练 RMSE	测试 RMSE
ELM	0.3585	0.0982
RELM	0.3577	0.0997
WRELM	0.3554	0.0991
MCCELM	0.0017	0.0006
TMCCELM	0.0006	0.0002

在异常值 20%的情况下, TMCCELM 和 MCCELM 在处理异常值时表现较好, 它们的测试 RMSE 远低于另外三种算法。与异常值 10%相比, TMCCELM 的训练 RMSE 不变, 测试 RMSE 增加了 0.0001。

**Table 3.** RMSE of different algorithms under 50% outliers**表 3.** 不同算法在 50%异常值下的 RMSE

算法	训练 RMSE	测试 RMSE
ELM	0.1553	0.0546
RELM	0.1564	0.0586
WRELM	0.1562	0.0587
MCCELM	0.0008	0.0937
TMCCELM	0.0003	0.0381

在异常值 50%的情形下, TMCCELM 算法的测试 RMSE 较于异常值 10%和 20%的情形, 有略微增加, 但相对于另外四种算法仍具有较低的误差, 表明在异常值 50%的情形下, TMCCELM 模型对异常值更具有鲁棒性。

**Table 4.** RMSE of different algorithms under 80% outliers**表 4.** 不同算法在 80%异常值下的 RMSE

算法	训练 RMSE	测试 RMSE
ELM	0.1117	0.1965
RELM	0.0721	0.1974
WRELM	0.0719	0.1948
MCCELM	0.0922	0.0010
TMCCELM	0.0517	0.0003

当异常值增加至 80% 的情形下, TMCCELM 仍具有较低的 RMSE, 其值为 0.0517, 其次是 WRELM 和 RELM, 分别为 0.0719 和 0.0721。ELM 和 MCCELM 的训练 RMSE 和测试 RMSE 均较高, 因此, 在较多异常值的情形下, ELM 和 MCCELM 对异常值更为敏感, 鲁棒性能较差。

通过比较不同模型在异常值下的拟合效果以及这五种算法在包含噪声和异常值的数据的鲁棒性能, 表明了 TMCCELM 在对待异常值时的鲁棒性能较强, 为了更好的验证该模型的优越性, 表 5 给出其在不同异常值下的训练时间。

**Table 5.** The training time of 5 algorithms with different outliers  
**表 5.** 不同异常值下 5 种算法的训练时间

算法	不同异常值的训练时间			
	10%	20%	50%	80%
ELM	0.0066	0.0053	0.0031	0.0011
RELM	0.0021	0.0016	0.0045	0.0011
WRELM	0.0061	0.0052	0.0066	0.0020
MCCELM	0.0050	0.0027	0.0069	0.0023
TMCCELM	0.0021	0.0064	0.0046	0.0014

通过比较不同异常值下的训练效果, 由表 5 可以得出, 在异常值 10% 时, TMCCELM 的训练时间为 0.0021, 在异常值 20% 时, 训练时间增加了 0.004, 随着异常值的增加, 训练时间只产生微量变化, 当异常值增加至 80% 时, 训练时间不足一秒, 因此, 针对截断最大相关熵损失函数, 本文给出了一个合理且有效的提升极限学习机鲁棒性的模型。

## 6. 总结

基于截断最大相关熵准则损失函数的鲁棒极限学习机采用截断最大相关熵作为损失函数, 与传统的极限学习机相比, 在对损失函数的选择上作出了改进, 由于传统 ELM 采用最小二乘损失函数, 对异常值表现较为敏感, 而 TMCCELM 考虑采用截断思想, 对最大相关熵损失函数进行了截断, 将异常值的影响限制在可接受的范围内, 使得模型在面对噪声和异常值时具有更强的鲁棒性和稳定性。

## 参考文献

- [1] Huang, G.-B., Zhu, Q.-Y. and Siew, C.-K. (2005) Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. 2004 *IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, Budapest, 25-29 July 2004, 985-990.
- [2] 张志洁, 侯睿. 基于极限学习机的脑卒中患病风险预测模型研究[J]. 电脑编程技巧与维护, 2022(6): 54-55, 113.
- [3] Zhang, Y., Wang, Y., Zhou, G., et al. (2018) Multi-Kernel Extreme Learning Machine for EEG Classification in Brain-Computer Interfaces. *Expert Systems with Applications*, **96**, 302-310. <https://doi.org/10.1016/j.eswa.2017.12.015>
- [4] Chyzyhyk, D., Savio, A. and Graña, M. (2015) Computer Aided Diagnosis of Schizophrenia on Resting State fMRI Data by Ensembles of ELM. *Neural Network*, **68**, 23-33. <https://doi.org/10.1016/j.neunet.2015.04.002>
- [5] 宋雯琦. 基于鲁棒性极端学习机的高光谱特征与图像分类研究[D]: [硕士学位论文]. 大连: 辽宁师范大学, 2020.
- [6] 王桥, 叶敏, 魏孟, 等. 基于 ELM 和 MCCKF 的锂离子电池 SOC 估计[J]. 工程科学学报, 2023, 45(6): 995-1002.
- [7] 吕新伟, 鲁淑霞. 迭代修正鲁棒极限学习机[J]. 计算机应用, 2023, 43(5): 1342-1348.
- [8] 陈剑挺, 吴志国, 叶贞成, 等. 基于收缩极限学习机的故障诊断鲁棒方法[J]. 计算机工程与设计, 2020, 41(1): 208-213.
- [9] 王快妮, 曹进德, 刘庆山. 基于指数 Laplace 损失函数的回归估计鲁棒超限学习机[J]. 应用数学和力学, 2019,

- 40(11): 1169-1178.
- [10] Horata, P., Chiewchanwattana, S. and Sunat, K. (2013) Regularized Extreme Learning Machine. *Neurocomputing*, **102**, 31-44. <https://doi.org/10.1016/j.neucom.2011.12.045>
  - [11] Chen, K., Lv, Q., Lu, Y. and Dou, Y. (2016) Robust Regularized Extreme Learning Machine for Regression Using Iteratively Reweighted Least Squares. *Neurocomputing*, **230**, 345-358. <https://doi.org/10.1016/j.neucom.2016.12.029>
  - [12] Xing, H.-J. and Wang, X.-M. (2013) Training Extreme Learning Machine via Regularized Correntropy Criterion. *Neural Computing and Applications*, **23**, 1977-1986. <https://doi.org/10.1007/s00521-012-1184-y>
  - [13] 刘彬, 刘静, 吴超, 杨有恒. 空间金字塔与局部感受野相结合的相关熵极限学习机[J]. 电子与信息学报, 2021, 43(8): 2343-2351.
  - [14] 刘兆伦, 武尤, 王卫涛, 等. 基于最大相关熵准则的多尺度高斯核极端学习机[J]. 计量学报, 2021, 42(5): 658-667. <https://doi.org/10.3969/j.issn.1000-1158.2021.05.18>
  - [15] 陈聪. L1 正则化与 pinball 损失函数的极限学习机[J]. 信息技术与信息化, 2023(3): 37-40.
  - [16] Lei, M.D., Wu, B., Yang, W.Y., Li, P., Xu, J.H. and Yang, Y.J. (2023) Double Extended Kalman Filter Algorithm Based on Weighted Multi-Innovation and Weighted Maximum Correlation Entropy Criterion for Co-Estimation of Battery SOC and Capacity. *ACS Omega*, **8**, 15564-15585. <https://doi.org/10.1021/acsomega.3c00918>
  - [17] Liu, J., Yang, G.B., Zhou, N., Qin, K.Y., Chen, B.D., Wu, Y.H. and Choi, K.-S. (2023) Event-Triggered Consensus Control Based on Maximum Correntropy Criterion for Discrete-Time Multi-Agent Systems. *Neurocomputing*, **545**, Article ID: 126323. <https://doi.org/10.1016/j.neucom.2023.126323>