[文章编号] 1003－4684(2026)04-0079-05

基于窗宽优化的鲁棒核密度估计器

彭　博

(湖北工业大学计算机学院，湖北 武汉 430068)

[摘　要] 核密度估计器作为一种非参数估计统计模型，在对数据进行概率密度函数估计时，其结果易受到异常点和窗口参数选择的影响。为了提升核密度在异常点条件下的估计准确性，提出了一种基于窗宽优化的鲁棒核密度估计器。该优化的估计器基于两个阶段来实现。在第一阶段中，首先通过引入莱维飞行策略来实现杂交水稻优化算法的寻优性能提升。该改进的杂交水稻优化算法将被用于优化无偏交叉验证目标函数，实现搜索最优的窗口值。在第二阶段中，将 M 稳健估计引入希尔伯特空间映射后的均值计算过程，降低了异常点在概率密度函数计算过程中的权重，排除异常点的影响。通过上述两个阶段的优化，最终达成异常点条件下概率密度函数的精准估计目标。实验结果表明，当数据中包含有异常值时，优化后的鲁棒核密度估计器比传统的核密度估计器具有更高的估计准确性。

[关键词] 稳健估计；鲁棒核密度估计；杂交水稻优化算法；莱维飞行；希尔伯特空间

[中图分类号] TP391　　[文献标识码] A

分布估计属于统计推断范畴中重要的研究方向，其目的是从一组观测数据中估计出随机变量的概率分布模型。连续值数据通常涉及复杂的概率密度函数估计。由于非参数估计在概率密度函数计算过程中无需对基本分布形式做出假设，它被广泛地应用于各类统计学习应用中[1]。核密度估计作为其中典型代表，近年来受到了广泛的关注 。核密度估计器的窗口宽度的选择是一个关键因素，决定了估计的精度和计算的复杂度。采用过小窗宽，估计结果可能会受到噪声的影响，稳定性较差。当窗宽过大，计算量较小，可能导致估计的精度较低[2]。无偏交叉验证法是一种常见的带宽求解方法[3]，但该方法的目标函数计算复杂度较高，且容易落入局部最优。因此，需要一种优化性能较好的算法来解决搜索最优窗宽的问题。

在数据科学领域，异常点的出现可能是由于多种原因，包括但不限于数据收集过程中的错误、测量误差、自然现象中的异常事件、或者是数据分布的自然属性。这些异常数据通常会影响正常数据中所包含的有效信息，甚至会产生错误的影响[4]。异常值也会影响学习模型的归纳偏好[5]，降低模型的泛化能力。异常值检测是常用方法，可以排除这些离群点的影响[6]。然而，异常点检测一般会被单独作为一种数据预处理的方法，且根据检测结果也难以全面剔除异常点[7-8]。如何提升核估计的鲁棒性，排除异常点的影响成为了亟待解决的问题。本研究针对数据包含异常样本的情况，构建优化的鲁棒核密度估计器，实现分布函数的精准估计；通过引入改进的杂交水稻优化和 M-估计对传统的核密度估计器进行改进，选择最优化的窗宽，降低异常点在估计过程中的权重，最终实现鲁棒概率密度估计目的。

1　核密度估计异常点问题概述

如图 1 所示，当数据集中存在异常值数据时，蓝色的原始点来自 $\mathcal{L}_1$ 的分布函数，橙色五边形数据点来自 $\mathcal{L}_2$ 分布。如何在离群点未知的条件下，不单独使用离群点检测的方法来实现 $\mathcal{L}_1$ 的分布精准估计是本研究的重点。

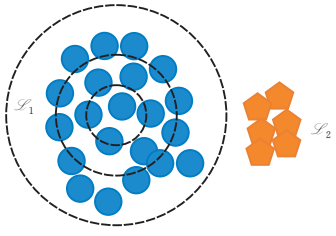

图 1　异常点对分布估计的影响

在解决此类问题的方法中，有参数估计方法[9]假设样本来自于某个分布簇，未知的只是总体分布具体参数值。密度函数估计可以被视为求解分布簇

[收稿日期] 2024－05－08

[第一作者] 彭　博(1999－)，男，湖北孝感人，湖北工业大学硕士研究生，研究方向为数据挖掘。

参数的过程。然而,在求解图 1 所展示的问题时,常见的有参数估计已经很难对 $\mathscr{L}_1$ 做出有效分布模型假设。当 $\mathscr{L}_2$ 分布存在且比较复杂情况下,完成对 $\mathscr{L}_1$ 的有效假设变得几乎不可能。作为非参数估计的核密度估计器,从样本本身出发,它无需总体假设。针对上述问题的求解具有极大优势,且适用性广泛,容易理解[10-11]。这种估计方式能够显著减少在实际应用中的限制条件,且不必考虑 $\mathscr{L}_2$ 的具体函数形式。但是常见的核密度估计器对所有的样本采用平均权重的方式进行处理,即便在优化窗口宽度的条件下,也很难实现 $\mathscr{L}_1$ 的精准估计。要改进传统的核密度估计,一方面需要考虑最优的窗宽[12],另一方面需要考虑样本的权重。本研究通过提出一种优化鲁棒核密度估计方法来实现二者并解决该问题。

2 鲁棒核密度估计器

本研究采用了鲁棒核密度估计(robust kernel density estimation,RKDE)[13],该方法通过结合核密度估计器与 M-估计来实现密度计算过程中的鲁棒性。半正定核的 KDE 过程可以被视为再生核希尔伯特空间[14]中样本均值的内积。由于原空间中存在异常值,它们映射后也会对样本均值造成较大的影响。通过引入 M-估计来提高计算过程中均值的准确性[15]。为了优化样本权重,RKDE 采用核化迭代重加权最小二乘法(kernelized iteratively Reweighted least squares, KIRWLS)。KIRWLS 是一种有效的迭代计算样本权重的算法,它具有收敛性[13]。假设样本 $X=\{x_1,\cdots,x_j,\cdots,x_n\}$ 表示包含未知离群点的 n 个 d 维样本,传统的核密度估计可依据公式(1)来完成估计概率密度 $\hat{f}(x)$ 的计算:

$$\hat{f}(x)=\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{x-x_j}{h}\right) \tag{1}$$

其中,$K(u)$ 表示核函数;$\int_{-\infty}^{+\infty}K(u)du=1$ 且 $K(u)\geqslant 0$;h 表示窗口宽度,高斯核为:

$$K(x,x')=\left(\frac{1}{\sqrt{2\pi}h}\right)^{d}\exp\left(-\frac{\|x-x'\|^2}{2h^2}\right) \tag{2}$$

公式(2)中的窗口宽度参数 h 的参数选择,对拟合效果有着巨大的影响,而公式(2)中 $1/n$ 则表示了均权化处理。

2.1 多元 M-估计量

Huber 首先提出了一种称为 M-估计量的稳健估计方法[16],它是极大似然估计在异常点条件下的一种推广。M-估计量源自最大似然估计理论,通过选择合适的 ρ 函数来替换似然函数实现鲁棒估计:

$$\hat{\theta}=\underset{\theta}{\operatorname{argmax}}\sum_{i=1}^{n}\ln f(x_i;\theta)=\underset{\theta}{\operatorname{argmin}}\sum_{i=1}^{n}\rho(\|x_i-\theta\|) \tag{3}$$

公式(3)中,若 ρ 可导,M-估计量 $\hat{\theta}$ 也表示为:

$$\sum_{i=1}^{n}\psi(x_i-\hat{\theta})=0 \tag{4}$$

公式(4)中,通过选择合适的损失函数 ψ,可达到稳健的性能。常见的稳健估计函数包括 Hampel 损失函数。其定义如公式(5)所示:

$$\psi(x)=\begin{cases}|x|,0\leqslant|x|<\gamma_1\\ \gamma_1,\gamma_1\leqslant|x|<\gamma_2\\ \gamma_1\dfrac{\gamma_3-|x|}{\gamma_3-\gamma_2},\gamma_2\leqslant|x|<\gamma_3\\ 0,\gamma_3\leqslant|x|\end{cases} \tag{5}$$

γ_1,γ_2,γ_3 表示分段参数。

2.2 希尔伯特空间下的核密度估计及核化迭代重加权算法

假设核函数满足 $K(\cdot,\cdot)>0$、$\int K(x,\cdot)dx=1$ 和 $K(x,x')$,并且还满足核函数是半正定的。高斯核显然满足上述性质。令 $\Phi(x)=K(\cdot,x)$ 表示特征映射,$\mathscr{H}$ 为样本希尔伯特空间映射集合 $\{\Phi(x):x\in R^d\}$。在再生核希尔伯特空间中,$\mathscr{H}$ 的再生性质,对于所有的 $g\in H$ 和 $x\in R^d$,$g(x)=\langle\Phi(x),g\rangle_H$,取 $g=K(\cdot,x')$,可以得到 $K(x,x')=\langle\Phi(x),\Phi(x')\rangle$。高斯核计算可以视为 Φ 变换后的内积。高斯核密度估计可以表示为 $\hat{f}_{KDE}(\cdot)=\frac{1}{n}\sum_{i=1}^{n}\Phi(X_i)$,也就是希尔伯特空间中的均值,故该问题的核心为最小化 $\min_{g\in H}\sum_{i=1}^{n}\|\Phi(X_i)-g\|_H^2$,它可以通过使用 M-估计来获取优化的均值。高斯核密度估计可由公式(6),而变权的高斯核密度估计可通过公式(7)来计算:

$$\hat{f}_{KDE}(x)=\frac{1}{n}\sum_{i=1}^{n}\langle\Phi(x),\Phi(X_i)\rangle=\left\langle\Phi(x),\frac{1}{n}\sum_{i=1}^{n}\Phi(X_i)\right\rangle \tag{6}$$

由于 $\frac{1}{n}\sum_{i=1}^{n}\Phi(X_i)$ 易受异常值影响。因此,将式(4)的稳健均值估计量来代替该均值。由此,稳健核密度估计器的加权形式可表示为公式:

$$\hat{f}_{KDE}(x)=\sum_{i=1}^{n}w_iK_h(x,x_i) \tag{7}$$

通过公式(7)对所有样本进行加权处理,每个样本的权重大于零,且所有样本的权重和为 1。在经过迭代优化之后,异常点的权重采用低值加权处理实现鲁棒性。为了得到最优的权重分配,样本的投影均值用稳健的均值来代替。采用 M-估计准则来构建如下的目标函数:

$$\hat{\eta}=\underset{\eta\in H}{\operatorname{argmin}}\sum_{i=1}^{n}\psi(\|\Phi(X_i)-\eta\|) \tag{8}$$

公式(8)中 ψ 表示损失函数，$\| \Phi(X_i) - \eta^{(k)} \|$ 值可通过公式(9)来获得(k 表示第 k 次迭代)：

$$\begin{aligned}\| \Phi(X_i) - \eta^{(k)} \|^2 &= \langle \Phi(X_i) - \eta^{(k)}, \Phi(X_i) - \eta^{(k)} \rangle \\ &= \langle \Phi(X_i), \Phi(X_i) \rangle - 2\langle \Phi(X_i), \eta^{(k)} \rangle + \langle \eta^{(k)}, \eta^{(k)} \rangle\end{aligned} \tag{9}$$

又 $\eta^{(k)} = \sum_{j=1}^{n} w_j^{(k-1)} \Phi(X_j)$，同时，因存在公式(10)、公式(11)和公式(12)：

$$\langle \Phi(X_i), \Phi(X_i) \rangle = K(X_i, X_i) \tag{10}$$

$$\langle \Phi(X_i), \eta^{(k)} \rangle = \sum_{j=1}^{n} w_j^{(k-1)} K(X_i, X_j) \tag{11}$$

$$\langle \eta^{(k)}, \eta^{(k)} \rangle = \sum_{j=1}^{n} \sum_{i=1}^{n} w_j^{(k-1)} w_i^{(k-1)} K(X_j, X_i) \tag{12}$$

可以得到公式(14)：

$$\hat{f}_{KDE}(x) = \langle \Phi(x), \hat{\eta} \rangle = \langle \Phi(x), \sum_{i=1}^{n} w_i^{(*)} \Phi(x_i) \rangle \tag{13}$$

$$\hat{f}_{\mathrm{KDE}}(x) = \sum_{i=1}^{n} w_i^{(*)} \langle \Phi(x), \Phi(X_i) \rangle = \sum_{i=1}^{n} w_i^{(*)} K(x, X_i) \tag{14}$$

其中，w 为最终收敛迭代权重。

3　鲁棒核密度估计器的窗宽优化

本节将介绍改进的杂交水稻优化算法，通过将无偏交叉验证作为杂交水稻优化算法的目标函数来实现搜索最优化窗口宽度。

3.1　核密度估计的窗口宽度影响

如图 2 至图 4 分别给出了不同窗口宽度的拟合效果比较。结果显示窗口宽度参数 h 对估计器的性能有很大的影响。当窗口宽度 h 选取 0.3 时拟合效果较好。在窗口宽度 h 选取 0.15 时，核密度估计器对于细节过于敏感，导致过拟合。在窗口宽度 h 取 0.6 时，结果明显欠拟合，无法显示有效信息。

选择一个合适的窗口宽度 h 可以最小化拟合误差，使核密度估计器所拟合的结果能够更加接近真实的分布。

3.2　无偏交叉验证目标函数

采用积分平方误差计算核密度估计的窗口宽度[17]，其结果可以用公式(15)表达。

$$ISE(h) = R(\hat{f}(x)) - 2\int \hat{f}(x) f(x) dx + R(f(x)) \tag{15}$$

其中，$R(\psi(x)) = \int \cdots \int \psi(x)^2 dx$ 被称为 $\psi(\cdot)$ 的粗糙度，而 $R(f(x))$ 可以忽略，因为对于 h 来说是一个常数，UCV 的表达式可以写成公式(16)的形式

$$UCV(h) = R(\hat{f}) - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}(x_i) \tag{16}$$

其中，

$$\hat{f}_{-i}(x_i) = \frac{1}{(n-1)h_1 \cdots h_d} \sum_{j \neq i} \left\{ \prod_{k=1}^{d} \mathrm{K}\left(\frac{x_{ik} - x_{jk}}{h_k} \right) \right\} \tag{17}$$

$$R(\hat{f}) = \frac{1}{(2\sqrt{\pi})^d n h_1 \cdots h_d} + \frac{1}{(2\sqrt{\pi})^d n^2 h_1 \cdots h_d} \times \sum_{i=1}^{n} \sum_{j \neq i} \exp\left\{ -\frac{1}{4} \sum_{k=1}^{d} \Delta_{ijk}^2 \right\} \tag{18}$$

将公式(18)和公式(17)带入公式(16)中，可以得到 $UCV(h)$ 最终的表达式为公式(19)

$$UCV(h) = \frac{1}{(2\sqrt{\pi})^d n h_1 \cdots h_d} + \frac{1}{(2\sqrt{\pi})^d n^2 h_1 \cdots h_d} \times \sum_{i=1}^{n} \sum_{j \neq i} \left[\exp\left(-\frac{1}{4} \sum_{k=1}^{d} \Delta_{ijk}^2 \right) - (2 \times 2^{d/2}) \exp\left(-\frac{1}{2} \sum_{k=1}^{d} \Delta_{ijk}^2 \right) \right] \tag{19}$$

其中，d 表示维度，$\Delta_{ijk} = (x_{ik} - x_{jk})/h_k$。可以通过将公式(19)作为杂交水稻优化算法的目标函数来获取鲁棒核密度估计器的最优化窗口宽度。

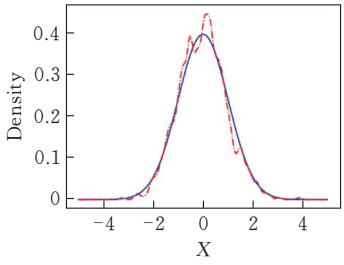

图 2　窗口宽度 h =0.15

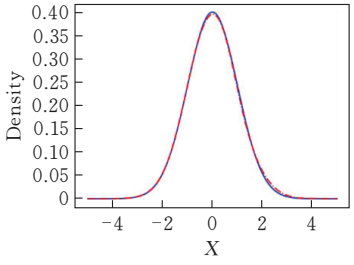

图 3　窗口宽度 h =0.3

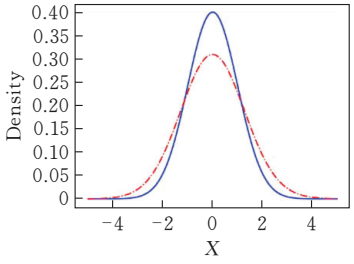

图 4　窗口宽度 h =0.6

3.3　改进的杂交水稻优化算法

受到三系法育种启发，叶志伟教授提出了杂交水稻优化算法[18]。采用杂交水稻优化算法作为优化窗口宽度的算法。在初始阶段，将种子根据适应度值由高到低分为三个部分，每个部分包含三分之一的种子，三个部分分别是保持系、恢复系和不育系。其中，保持系种子具有最优的适应度值，其作用

是给不育系种子更新时提供较优位置的参考，适应度值最差的种子称为不育系，最后将种群中剩余的种子标记为恢复系。进行分类之后，不育系通过杂交的方式对自身进行更新，通过对比新生成种子的适应度值来判断是否要进行替换，通过替换的方式保留具有较优适应度值的种子。杂交操作可以通过公式(20)表示：

$$X_{new}=\frac{r_1\cdot X_{sr}+r_2\cdot X_{br}}{r_1+r_2} \tag{20}$$

X_{new} 表示新生种子，X_{br} 表示随机选取的保持系种子，X_{sr} 表示随机选取的不育系种子，r_1 和 r_2 表示随机数，取值范围为 $[-1,1]$。恢复系种子通过自交的方式对自身进行更新。在自交过程中，算法将遍历恢复系中所有种子，产生拥有新适应度值的种子，通过替换的方式保留具有较优适应度值的种子。恢复系种子自交可以用公式(21)表示为：

$$X_{\text{new}}=r_3(X_{\text{best}}-X_j)+X_i \tag{21}$$

其中，X_{new} 表示新生种子，X_i 表示当前被遍历的恢复系种子，X_j 表示从恢复系种群中随机选取的恢复系种子，且 $i\neq j$，X_{best} 表示全局最优种子的位置，r_3 表示随机数，取值范围为 $[0,1]$。判断是否进行替换操作由新种子 X_{new} 和与旧种子 X_{old} 的适应度值比较决定，替换操作的公式(22)为：

$$X_{\text{new}}^{t+1}=\begin{cases}X_{\text{new}}^{t},g(X_{\text{new}}^{t})>g(X_{\text{old}}^{t})\\X_{\text{old}}^{t},g(X_{\text{old}}^{t})>g(X_{\text{new}}^{t})\end{cases} \tag{22}$$

当恢复系种子在多次迭代中保持不变，表示可能陷入局部最优解，故采用重置方式进行更新。需要设定自交上线次数 T。在重置操作后更新种子的位置和适应度值，并且将 T 归零。重置操作可以用公式(23)表示：

$$X_{new}=X_i+r_4(R_{\max}-R_{\min})+R_{\min} \tag{23}$$

其中，X_{new} 代表新生成的种子，而原种子将被其替换，X_i 代表当前种子，$R_{\max}$ 和 $R_{\min}$ 表示搜索范围的上下界，r_4 的取值范围在[0,1]之间。在杂交水稻优化算法中，由于保持系种子自身不会更新位置，考虑到可能面临陷入局部最优的情况，通过引入莱维飞行策略对算法进行改进，能够有效提高算法性能。随机性在优化过程中至关重要，它是一种随机游走行为[14]。随机游走通常由一系列随机步骤组成，可以用公式(24)表示：

$$S_N=\sum_{i=1}^{N}X_i=X_1+X_2+\cdots+X_N \tag{24}$$

其中，S_N 代表游走的整体；X_i 表示每一段自由方向的移动。莱维飞行具有多种模型[19]，Mantegna 于 1994 年提出了一种稳定随机莱维飞行模型，其模型如公式(25)所示：

$$s=\frac{u}{|v|^{1/\beta}} \tag{25}$$

其中，s 表示莱维飞行的移动路径；β 的取值范围在 [0,2] 之间；u 和 v 表示高斯分布的随机数，具体见公式(26)和公式(27)：

$$u\sim N(0,\sigma_u^2) \tag{26}$$

$$v\sim N(0,\sigma_v^2) \tag{27}$$

其中，

$$\sigma_u=\left\{\frac{\Gamma(1+\beta)\cdot\sin(\pi\beta/2)}{\Gamma[(1+\beta)/2]\,\beta\cdot 2^{(\beta-1)/2}}\right\}^{1/\beta} \tag{28}$$

$$\sigma_v=1 \tag{29}$$

在公式(28)中，$\Gamma(x)$ 表示伽马函数，$\Gamma(x)$ 的表达如公式(30)：

$$\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt \tag{30}$$

通过公式(30)可得到随机游走计算方式为：

$$X_{new}=X_i+\nu*s \tag{31}$$

其中 ν 表示随机游走的缩放因子，通过公式(31)对保持系种子进行更新，避免陷入局部最优的情况。

4 优化鲁棒核密度估计器实验

在本节实验中，选择多种分布数据进行模拟对比试验。表 1 给出了数据生成的参数说明。KDE 和 RKDE 的窗口宽度值通过经典的拇指法法则获取。基于 ORKDE 的窗口宽度值由改良的杂交水稻优化算法获取。实验中，采用了多种分布数据和异常点的参数设置，如表 1 所示。

表 1 分布函数生成参数

分布函数	一维分布	二维分布
高斯分布	μ1:0 样本个数:600 δ1:0.5 异常点: 样本个数:20 μ2:1 δ2:3	μ1:[0,0]样本个数:600 δ1:[[3,1.2], [1.2,3]] 异常点:样本个数:100 μ2:[0.5,0.5] δ2:[[3,1.5],[1.5,3]]
T 分布	L1:0 DF1=5 S1:2.1 样本个数:600 异常点: 样本个数:20 μ2:0.5 δ2:3	L1:[0,0] DF1=5 S1:[[2,0.1],[0.1,2]] 样本个数 600 异常点:样本个数 20 μ2:[0.5,0.5] δ2:[[3,1.5],[1.5,3]]
柯西分布	L1:0 样本个数:600 S1:2.1 DF1=1 异常点: 样本个数:20 μ2:0.5 δ2:3	L1:[0,0]样本个数:600 S1:[[2,0.1], [0.1,2]] DF1=1 异常点:样本个数:20 μ2:[0.5,0.5] δ2:[[3,1.5],[1.5,3]]

续表 1　分布函数生成参数

分布函数	一维分布	二维分布
拉普拉斯	L1:0 样本个数:600 S1:2.2 异常点: 样本个数:20 $\mu2$:1　$\delta2$:3	L1:[0,0]样本个数:600 S1:[[2,0.1],[0.1,2]] 异常点: 样本个数:20 $\mu2$:[0.5,0.5] $\delta2$:[[3,1.5],[1.5,3]]

接下来,将对基于这些设置的数据进行 MISE (mean integrated squared error)对比实验。表 2 和表 3 将分别展示实验结果。分析表 2 和表 3 中的数据,可以看到 ORKDE 具有最好的 MISE 值,这一优势在各种分布数据和参数设置下都得到了体现。这说明 ORKDE 在处理异常点时能够更好地保持数据分布的稳定性和准确性。这为后续研究提供了重要的参考依据。实验结果表明,通过窗口宽度优化和异常点权重优化后,ORKDE 的 MISE 值在上述拟合测试中为最优。

表 2　一维分布的 MISE 对比($\times10^{-3}$)

分布函数	一维数据分布			
	高斯分布	T 分布	柯西分布	拉普拉斯分布
KDE	6.16	4.57	9.87	3.44
RKDE	4.76	4.36	9.33	2.37
ORKDE	4.36	3.16	8.25	2.28

表 3　二维分布的 MISE 对比($\times10^{-3}$)

分布函数	二维数据分布			
	高斯分布	T 分布	柯西分布	拉普拉斯分布
KDE	5.31	6.58	4.42	9.31
RKDE	4.27	5.56	3.43	9.15
ORKDE	3.57	4.51	2.32	7.94

5　结论

本研究提出了一种基于窗口优化的鲁棒核密度估计器。鲁棒核密度估计被引入来完成连续型数据的稳健估计,有效地抑制了异常点数据在拟合过程中的权重,减少了异常点数据对希尔伯特空间映射的均值影响,使数据拟合结果更加准确。同时,改良的杂交水稻优化算法被用于获取最优的窗口宽度值。实验结果表明在包含异常点的概率密度计算问题中,本研究所提出的方法能够有效提升核密度估计器的准确性。

[参 考 文 献]

[1] 杨筱菡.污染分布的非参数估计[J].同济大学学报,2001,29(6):700-702.

[2] 陈佳琪,何玉林,黄哲学,等. 基于统计感知策略的高斯混合模型求解方法[J]. 数据采集与处理, 2023, 38(3):525-538.

[3] 王迎光,李瑞尧. 极端海况 Meta-ocean 轮廓线的双变量核密度估计[J]. 船舶力学,2021,25(12):1581-1593.

[4] 刘财辉,刘地金.离群点检测的邻近性方法综述[J].计算机工程与应用,2022,58(21):1-12.

[5] Shu X, Ye Y. Knowledge Discovery: Methods from data mining and machine learning[J]. Social Science Research, 2023, 110:102817.

[6] 于万国,何玉林,覃芸霖. 基于观测点机制的异常点检测算法[J]. 深圳大学学报,2022,39(3):355-362.

[7] Su S, Xiao L, Ruan L, et al. An efficient density-based local outlier detection approach for scattered data [J]. IEEE Access, 2018, 7:1006-1020.

[8] Wang H, Bah M J, Hammad M. Progress in outlier detection techniques: A survey[J]. IEEE Access, 2019, 7:107964-108000.

[9] 陈健,杜兰,廖磊瑶. 基于参数化统计模型的雷达 HRRP 目标识别方法综述[J]. 雷达学报,2022,11(6):1020-1047.

[10] Chacón J E, Duong T. Multivariate kernel smoothing and its applications [M]. Boca Raton: Chapman & Hall/CRC, 2018.

[11] Scott D W. Multivariate density estimation: theory, practice, and visualization [M]. Hoboken: John Wiley & Sons, 2015.

[12] Knuth K H. Optimal data-based binning for histograms and histogram-based probability density models [J]. Digital Signal Processing, 2019,95:102581.

[13] Kim J S, Scott C D. Robust kernel density estimation [J]. The Journal of Machine Learning Research, 2012, 13 (1):2529-2565.

[14] Jian W X, Paiva A R C, Park I, et al. A reproducing kernel hilbert space framework for information-theoretic learning[J]. IEEE Transactions on Signal Proces, 2008, 56(12):5891-5902.

[15] Humbert P, Le Bars B, Minvielle L. Robust kernel density estimation with median-of-means principle[C]. In: International Conference on Machine Learning. 2022:9444-9465.

[16] 祝金甫,汤伟,冯兴东. 线性回归 M 估计量的 Wild-Bootstrap 方法研究[J]. 统计研究, 2015(08):99-103.

[17] Sain S R, Baggerly K A, Scott D W. Cross-validation of multivariate densities[J]. Journal of the American Statistical Association, 1994, 89(427):807-817.

[18] 秦泽青,叶志伟,王泽松,等. 一种多策略改进遗传算法的医学图像配准方法[J]. 小型微型计算机系统,2021,42(12):2600-2606.

[19] Jia H, Lang C. Salp swarm algorithm with crossover scheme and Levy flight for global optimization[J]. Journal of Intelligent and Fuzzy Systems, 2021, 40(5): 9277-9288.

(下转第 133 页)

tual information of the feature map using multiple scales of null convolution with different sampling rates. Second, the ECA is combined with the C3 module so that the ECA has the C3 attribute, which enables channel feature enhancement of the input feature map and greatly reduces the number of parameters. The experimental results show that, on the Tsinghua Tencent 100K dataset and the Traffic Camera Object Detection dataset, the ACE Yolov5 model proposed in this paper outperforms other models in terms of detection performance. Compared with Yolov5, the mAP50 improved by 5.1% and 2.2%, respectively.
Keywords: small object detection; Yolov5; context; attentional mechanisms

[责任编校: 闫　品]

(上接第 83 页)

A Robust Kernel Density Estimator Based on Window Width Optimization

PENG Bo

(*School of Computer Science, Hubei Univ. of Tech., Wuhan* 430068, *China*)

Abstract: The kernel density estimator, as a non-parametric estimation statistical model, is susceptible to the influence of outlier and window parameter selection when estimating the probability density function of data. In order to improve the accuracy of kernel density estimation under outlier conditions, this study proposes a robust kernel density estimator based on window width optimization. The optimized estimator is implemented in two stages. In the first stage, this study introduces the Levi flight strategy to improve the optimization performance of hybrid rice optimization algorithms. The improved hybrid rice optimization algorithm will be used to optimize the unbiased cross validation objective function and achieve the search for the optimal window value. In the second stage, this study introduces M-robust estimation into the mean calculation process of Hilbert space mapping, reducing the weight of outliers in the probability density function calculation process and eliminating the influence of outliers. Through the optimization of the above two stages, the goal of accurately estimating the probability density function under outlier conditions is ultimately achieved. The experimental results show that when the data contains outliers, the optimized robust kernel density estimator has higher estimation accuracy than traditional kernel density estimators.
Keywords: robust estimation; kernel density estimation; hybrid rice optimization algorithm; Lévy fight; Hibert space

[责任编校: 闫　品]

(上接第 120 页)

Study on Anti-collapse Performance of CFST Frame with Single-sided Bolt End Plates Connection

LI Yutong[1], WANG Donghua[1,2], WANG Haoyu[1], MA Jun[1]

(1 *School of Civil Engineering, Anhui Jianzhu Univ., Hefei*, 230601, *China*;
2 *Prefabricated Building Research Institute of Anhui province, Hefei* 230601, *China*)

Abstract: Based on two typical collapse test models, the accuracy of anti-collapse analysis was verified. A simplified analysis model of half-span beams was proposed to study anti-collapse performance of CFST frame substructures with blind bolt end plates connection. The failure mode, load-displacement curve and resistance mechanism were obtained, and the influence of key parameters is analyzed by numerical simulation. The results show that the structure resistance mechanism is mainly beam mechanism, transition mechanism and catenary mechanism. The ultimate bearing capacity and corresponding displacement of extended end plates are 40.42% and 24.2% higher than flush end plates. The ultimate bearing capacity and corresponding displacement of square CFST substructure are 15.4% and 26.24% higher than circle CFST substructure. The anti-collapse performance of the structure is affected by the key parameters such as bolt diameter, end plate thickness and column thickness. The improvement of bolt diameter for anti-collapse performance and ultimate displacement is most significant.
Keywords: blind bolt; end plate connection; CFST; numerical simulation; anti-collapse performance

[责任编校: 裴　琴]