

文章编号:1671-6833(2023)01-0024-07

# 混沌时间序列的核自适应滤波预测算法

刘 强, 王世元, 黄雪微, 王代丽

(西南大学 电子信息工程学院, 重庆 400715)

**摘 要:** 在实际环境中, 混沌时间序列常包含大量的噪声和异常值。由于这些干扰因素, 基于二阶相似性度量的核自适应滤波器在混沌时间序列预测中的预测性能显著下降。基于上述问题, 提出了一种鲁棒混沌时间序列的核自适应滤波预测算法。所提算法基于广义对数核损失函数的非线性相似性度量, 有效地提高了核自适应滤波器在脉冲噪声环境中的鲁棒性, 与此同时, 该算法采用自适应  $K$ -Means 采样的稀疏 Nyström 非线性映射方法, 预先固定了算法的网络尺寸, 从而降低核自适应滤波算法的计算复杂度。在所提算法中, 使用递归更新方式, 使算法具备较快的收敛速度。最后对滤波算法进行 Mackey-Glass 混沌时间序列的预测仿真。仿真结果表明: 作为一种新的鲁棒  $K$ -Means 采样的 Nyström 递归最小广义对数核损失预测方法, 与稀疏化核自适应滤波算法相比, 该算法在脉冲噪声中具备更好的鲁棒性; 与其他典型鲁棒核自适应滤波预测算法相比, 该算法具备更快的收敛速度和更高的滤波精度。

**关键词:** 混沌时间序列预测; 核自适应滤波; 广义对数核损失函数; Nyström 映射; 递归更新

**中图分类号:** TN911.7

**文献标志码:** A

**doi:** 10.13705/j.issn.1671-6833.2023.01.001

混沌现象<sup>[1]</sup>是由不附加随机因素的确定性非线性动力学系统产生而来。实际应用中, 混沌信号往往受到不同程度的噪声干扰, 经典的去噪预测方法主要有 3 类: 局部预测法<sup>[2]</sup>、全局预测法<sup>[3]</sup>、自适应预测法<sup>[4]</sup>。其中, 局部预测法利用局部数据特征拟合非线性函数, 提高了处理速度; 全局预测法利用全部数据的特征来表征非线性函数, 对整个混沌信号的噪声实现有效抑制; 自适应预测法自适应地追踪混沌的运动轨迹, 通过简单配置参数确保了低数据量时仍然有较好预测精度。良好的去噪性能以及简单的参数配置使自适应预测法得到广泛应用。

在自适应滤波预测中, 数据空间中的二阶相似性度量是高斯噪声环境下性能优越的测量方法, 如基于均方误差 (mean square error, MSE)<sup>[5]</sup> 的仿射投影<sup>[6]</sup> 和最小二乘法<sup>[7]</sup>。但是采用二阶度量的自适应预测方法在脉冲噪声中的去噪性能显著下降。为解决这些问题, 非二阶相似度量方法被提出, 如平均  $p$  阶误差 (mean  $p$ -power error, MPE) 准则<sup>[8]</sup>, 但一般情况下使用 MPE 准

则仍不能在脉冲条件下提供理想的滤波性能。近几年, 基于可再生核希尔伯特空间<sup>[5]</sup> 的非线性度量方法成为研究热点。作为信息理论<sup>[9]</sup> 的代表, 相关熵损失 (correntropic loss, C-loss)<sup>[10]</sup> 和广义相关熵损失 (generalized correntropic loss, GC-loss)<sup>[11-12]</sup> 利用数据的高阶统计量, 成功地衍生出一些在脉冲噪声情况下鲁棒的自适应滤波预测方法<sup>[4]</sup>。然而, 熵损失性能曲面的非凸性导致其收敛性能较差。为提高熵损失性能, 核风险敏感损失 (kernel risk sensitive loss, KRS-loss)<sup>[13]</sup> 因具有更好的性能曲面而被提出, 然而 KRS-loss 在脉冲噪声中并不具鲁棒性。进一步, 广义对数损失 (generalized logarithmic loss, GL-loss)<sup>[14]</sup> 被提出, 用来抑制较大异常值的干扰。值得注意的是, 在一些环境下, GL-loss 比 C-loss 和 KRS-loss 具有更好的性能。

然而, 采用上述度量的线性滤波器在复杂非线性混沌时间序列预测中都存在预测性能下降的问题。核方法<sup>[5]</sup> 能够有效地解决非线性问题, 如极限学习机<sup>[15]</sup> 和核自适应滤波器 (kernel adaptive filter,

收稿日期: 2022-07-22; 修订日期: 2022-09-08

基金项目: 国家自然科学基金资助项目 (62071391); 重庆市自然科学基金资助项目 (cstc2020jcyj-msxmX0234); 中央高校基本科研业务费 (2020jd001)

通信作者: 王世元 (1980—), 男, 重庆人, 西南大学教授, 博士, 博士生导师, 主要从事信号处理研究, E-mail: wsy@swu.edu.cn。

引用本文: 刘强, 王世元, 黄雪微, 等. 混沌时间序列的核自适应滤波预测算法[J]. 郑州大学学报(工学版), 2023, 44(1): 24-30. (LIU Q, WANG S Y, HUANG X W, et al. Kernel adaptive filtering prediction algorithm of chaotic time series [J]. Journal of Zhengzhou university (engineering science), 2023, 44(1): 24-30.)

KAF)<sup>[5]</sup>。核方法的核心在于将原始空间中的内积操作转换到高维空间中,使用“核技巧”替换内积操作,从而降低计算复杂度。典型的核自适应算法包括核最小均方(kernel least mean squares, KLMS)算法<sup>[16-17]</sup>和核递归最小二乘(kernel recursive least squares, KRLS)算法<sup>[18]</sup>等。KRLS 算法考虑先前的多个误差,与基于随机梯度下降的 KLMS 算法相比,该算法具有更快的收敛速度和更高的滤波精度。

值得注意的是,在 KAF 算法中随着输入的不断增加,算法网络尺寸会随之线性增长,最终导致巨大的计算和存储开销。为抑制网络尺寸的增长,一些数据稀疏处理方法被提出来,比如稀疏化<sup>[5]</sup>和量化策略<sup>[17]</sup>。但是,使用这些稀疏处理方法仍然存在网络尺寸次线性增长问题。相比之下,采用自适应 K-Means 采样<sup>[19]</sup>的 Nyström 映射<sup>[20]</sup>通过提前固定网络尺寸,在时间和空间复杂度上都表现出了更大优势。

近年来,脉冲噪声下的鲁棒性滤波算法日益成为一重要研究领域。本文提出一种新的广义对数核损失(generalized logarithmic kernel loss, GLKL),目的在于提高 GL-loss 在某些非线性系统中的性能。为解决核方法带来的网络尺寸线性增长问题,将 GLKL 与 K-Means 采样的 Nyström 映射和递归更新方式相结合,推导出一种新的鲁棒算法,即 NRMGLKL-KM 算法。该算法与稀疏化核自适应滤波算法相比具备更高的鲁棒性,与其他典型鲁棒核自适应滤波预测算法相比,该算法具备更快的收敛速度和更高的滤波精度。

## 1 背景

### 1.1 广义对数核损失的定义

为构建鲁棒性预测方法,GL-loss<sup>[14]</sup>作为两个随机变量  $X$  和  $Y$  的相似性度量被定义为

$$L(X, Y) = \frac{1}{\alpha} E[\ln(1 + \alpha |X - Y|^p)]。 \quad (1)$$

式中: $\alpha > 0, p > 0$ 。式(1)在处理异常值时具有较强的鲁棒性,已经广泛应用于鲁棒性算法研究。然而,在处理复杂的非线性系统问题时,基于式(1)的线性滤波算法不具备较强的非线性滤波性能。

利用核方法,GLKL 函数通过提取数据的不同阶统计量,有效提升了在脉冲噪声下非线性问题的学习性能,GLKL 定义如下:

$$\begin{aligned} L_p(X, Y) &= \frac{1}{\alpha} E[\ln(1 + \alpha \|\varphi(X) - \varphi(Y)\|_{\mathbf{F}}^p)] \\ &= \frac{1}{\alpha} E[\ln(1 + \alpha(1 - \kappa_\sigma(X - Y))^{p/2})] \\ &= \frac{1}{\alpha} \int \ln(1 + \alpha(1 - \kappa_\sigma(X - Y))^{p/2}) \\ &\quad dF_{X,Y}(X, Y)。 \end{aligned} \quad (2)$$

式中: $\|\cdot\|_{\mathbf{F}}$ 为特征空间  $\mathbf{F}$  域的欧几里得范数; $F_{X,Y}(X, Y)$ 为联合密度函数; $\varphi(\cdot)$ 表示非线性映射函数。核宽为  $\sigma$  的高斯核  $\kappa_\sigma(X, Y) = \exp\left(-\frac{(X-Y)^2}{2\sigma^2}\right)$ 。通常情况下, $F_{X,Y}(X, Y)$ 无法获知,故采用样本集  $\{x_i, y_i\}_{i=1}^N$  对 GLKL 进行估计,经验 GLKL 定义为

$$\hat{L}_p(\mathbf{x}, \mathbf{y}) = \frac{1}{N\alpha} \sum_{i=1}^N \ln(1 + \alpha(1 - \kappa_\sigma(x_i - y_i))^{p/2})。 \quad (3)$$

经验 GLKL 可以被描述为  $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$  和  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$  之间的广义“距离”。

### 1.2 相关性质

关于 GLKL 的一些重要性质如下所示。

**性质 1:**  $L_p(X, Y)$  是对称、正定且有界的。

证明:由于  $\kappa_\sigma(X, Y) = \kappa_\sigma(Y, X)$ ,故有  $L_p(X, Y) = L_p(Y, X)$ ,即  $L_p(X, Y)$  是对称的。当  $0 < \kappa_\sigma(X, Y) \leq 1$ ,  $L_p(X, Y)$  是正定且有界的,即  $0 < L_p(X, Y) \leq 1/\alpha \cdot \ln(1 + \alpha)$ ,同时当且仅当  $X = Y$  时可以求出它的最大值。

**性质 2:** 当  $\sigma$  足够大时,有

$$L_p(X, Y) \approx (2\sigma^2)^{-p/2} E[|X - Y|^p]。 \quad (4)$$

证明:由  $\lim_{x \rightarrow 0} e^x \approx 1 + x$  和  $\lim_{x \rightarrow 0} \ln(1 + x) \approx x$ ,当  $\sigma$  足够大时有

$$\begin{aligned} L_p(X, Y) &= \frac{1}{\alpha} E[\ln(1 + \alpha(1 - \kappa_\sigma(X - Y))^{p/2})] \\ &= \frac{1}{\alpha} E[\ln(1 + \alpha(1 - (2\sigma^2)^{-p/2} |X - Y|^p))] \\ &\approx (2\sigma^2)^{-p/2} E[|X - Y|^p]。 \end{aligned} \quad (5)$$

由性质 2 可知,GLKL 可以捕获误差的  $p$  阶绝对值。当  $\sigma = 1$  时,GLKL 可以视为 MPE 准则。

**性质 3:** 定义  $\mathbf{e} = \mathbf{x} - \mathbf{y} = [e_1, e_2, \dots, e_i]^T$ ,其中: $e_i = x_i - y_i, i = 1, 2, \dots, N$ 。当满足  $\|\mathbf{e}\|_\infty = \max |e(i)| < \sigma$  且  $p \geq 2$  时,  $\hat{L}_p(\mathbf{x}, \mathbf{y})$  在任何点处都是关于  $\mathbf{e}$  的凸函数。

证明:基于误差向量  $\mathbf{e}$ ,  $\hat{L}_p(\mathbf{x}, \mathbf{y})$  的黑塞矩阵可以表示为如下形式:

$$\mathbf{H}_{\hat{L}_p(x,y)}(\mathbf{e}) \left[ \frac{\partial^2 \hat{L}_p(\mathbf{x}, \mathbf{y})}{\partial e_i \partial e_j} \right] = \text{diag}[\rho_1, \rho_2, \dots, \rho_N]. \quad (6)$$

$$\text{式中: } \rho_i = \theta_i(\eta + \mu); \theta_i = \frac{p(1-\kappa_\sigma)^{\frac{p-4}{2}} \kappa_\sigma}{2N\sigma^4(1+\alpha(1-\kappa_\sigma)^{p/2})^2} > 0; \kappa_\sigma$$

为  $\kappa_\sigma(X, Y)$  的简写;  $\mu = \alpha(1-\kappa_\sigma)^{\frac{p}{2}}(\sigma^2 - e_i^2) + \alpha(1-\kappa_\sigma)\sigma^2$ ;  $\eta = (1-\kappa_\sigma)\sigma^2 + \left(\frac{p}{2}\kappa_\sigma - 1\right)e_i^2$ 。当  $\max |e_i| < \sigma$ ,  $i=1, 2, \dots, N$  且  $p \geq 2$  时, 可知  $\mu \geq 0$  且  $\eta \geq (1-\kappa_\sigma)\sigma^2 + (\kappa_\sigma - 1)e_i^2 \geq 0$ , 故  $\rho_i \geq 0$ 。可推出  $\mathbf{H}_{\hat{L}_p(x,y)}(\mathbf{e})$  中的每个元素都是非负的, 即 GLKL 具有凸性质。

设置  $\sigma$  和  $\alpha$  的值均为 1, 图 1 比较了在不同参数  $p$  下, GL-loss 和 GLKL 的损失函数随误差变化的情况。从图 1 可以得出: ①当误差趋近于 0 时,  $p$  越大, GLKL 曲线就越平滑, 即 GLKL 对稳态误差有更好的平滑性; ②与 GL-loss 相比, 在误差较大的情况下, GLKL 的损失函数更平滑, 即 GLKL 对大的异常值不敏感。

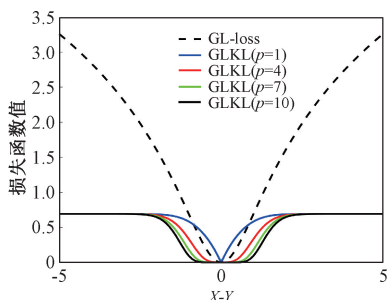


图 1 GL-loss 与不同的  $p$  下 GLKL 的比较

Figure 1 Comparison of GL-loss and GLKL with different  $p$

## 2 算法推导

### 2.1 Nyström 映射

在处理非线性问题时, 利用核方法将输入向量  $\mathbf{u}$  映射到高维特征空间, 可得  $\varphi(\mathbf{u}) = [\sqrt{\lambda_1}\phi_1(\mathbf{u}), \dots, \sqrt{\lambda_i}\phi_i(\mathbf{u}), \dots]^T$ , 其中,  $\varphi(\mathbf{u})$  为映射输入,  $\lambda_i$  和  $\phi_i(i=1, 2, \dots)$  分别定义为 Mercer 核函数的特征值和特征向量。由于  $\varphi(\mathbf{u})$  的维度会无限增长, 最终会导致巨大的计算量。

Nyström 映射作为一种稀疏化方法, 可以显著减少计算量。基于独立同分布(i. i. d.)输入数据集  $\mathbf{X} = \{\mathbf{u}_i\}_{i=1}^N$ , 可以得到 Gram 矩阵  $\mathbf{G} \in \mathbf{R}^{N \times N}$ , 其中  $G_{ij} = \kappa(\mathbf{u}_i, \mathbf{u}_j)$ ,  $i$  和  $j$  均为大于 0 且小于等于  $N$  的整数。为了去近似矩阵  $\mathbf{G}$ , 从数据集  $\mathbf{X}$  中随机选取  $m$  个样本  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ , 同样可以得到秩为  $m$  的 Gram 矩

阵  $\mathbf{W} \in \mathbf{R}^{m \times m}$ , 其中  $W_{ij} = \kappa(\mathbf{v}_i, \mathbf{v}_j)$ ,  $i$  和  $j$  均为大于 0 且小于等于  $m$  的整数。由  $m$  个样本和所有的输入数据可得到矩阵  $\mathbf{K} \in \mathbf{R}^{N \times m}$ , 其中  $K_{ij} = \kappa(\mathbf{u}_i, \mathbf{v}_j)$ ,  $0 < i \leq N$ ,  $0 < j \leq m$ 。通过  $\mathbf{K}$  和  $\mathbf{W}$  可得到  $\mathbf{G}$  的近似矩阵:

$$\mathbf{G} \approx \hat{\mathbf{K}} = \mathbf{K}\mathbf{W}^\dagger \mathbf{K}^\top = \mathbf{K}\mathbf{V}\mathbf{D}^{-1} \mathbf{V}^\top \mathbf{K}^\top = (\mathbf{K}\mathbf{V}\mathbf{D}^{-\frac{1}{2}})(\mathbf{K}\mathbf{V}\mathbf{D}^{-\frac{1}{2}})^\top. \quad (7)$$

式中:  $\mathbf{V} = [\lambda_1, \lambda_2, \dots, \lambda_m]$  和  $\mathbf{D} = \text{diag}[\zeta_1, \zeta_2, \dots, \zeta_m]$  分别为矩阵  $\mathbf{W}$  特征向量及其对应的特征值所组成的矩阵。故  $\mathbf{G}$  中的元素可以近似表示为

$$\begin{aligned} \kappa(\mathbf{u}_i, \mathbf{u}_j) &= \kappa(i, j) \approx \hat{\mathbf{K}}(i, j) \\ &= \mathbf{K}(i, \cdot) \mathbf{V} \mathbf{D}^{-\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} \mathbf{V}^\top \mathbf{K}(j, \cdot)^\top \\ &= (\mathbf{D}^{-\frac{1}{2}} \mathbf{V}^\top \mathbf{K}(i, \cdot)^\top)^\top \mathbf{D}^{-\frac{1}{2}} \mathbf{V}^\top \mathbf{K}(j, \cdot)^\top \\ &= \mathbf{z}(\mathbf{u}_i)^\top \mathbf{z}(\mathbf{u}_j). \end{aligned} \quad (8)$$

式中:  $\mathbf{z}(\mathbf{u}_i)$  为  $\mathbf{u}_i$  映射到高维特征空间的近似函数。从式(8)中可以看出,  $\varphi(\mathbf{u})$  作为一个无限维的隐式输入, 实际上可以用低维输入  $\mathbf{z}(\mathbf{u})$  来近似表示。

### 2.2 基于 $K$ -Means 采样的 Nyström 映射

Nyström 映射中, 核矩阵近似的精度主要取决于样本点的选择。 $K$ -Means 方法作为一种经典的聚类方法, 能够抽取与原始数据的分布基本一致的数据, 有效地提高 Nyström 近似精度。 $K$ -Means 采样是将给定的样本集分成  $k$  个簇, 并将每个簇的质心作为样本。

在线应用中, 给定一个大小为  $L$  且服从 i. i. d. 的采样集  $\hat{\mathbf{Z}} = \{\hat{\mathbf{x}}_i\}_{i=1}^L$ 。当  $0 < k < L$  时,  $K$ -Means 采样的基本步骤如下。

步骤 1 从采样集  $\hat{\mathbf{Z}}$  中随机选择  $k$  个样本作为  $k$  个簇的初始质心:  $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ 。

步骤 2 将每个数据分类到距离最近的一个簇中, 即

$$\begin{aligned} \hat{\mathbf{Z}} &= [\underbrace{\hat{\mathbf{x}}_{1,1}, \dots, \hat{\mathbf{x}}_{1,j}, \dots, \hat{\mathbf{x}}_{1,|c_1|}}_{|c_1|}, \underbrace{\hat{\mathbf{x}}_{2,1}, \dots, \hat{\mathbf{x}}_{2,j}, \dots, \hat{\mathbf{x}}_{2,|c_2|}}_{|c_2|}, \\ &\dots, \underbrace{\hat{\mathbf{x}}_{i,1}, \dots, \hat{\mathbf{x}}_{i,j}, \dots, \hat{\mathbf{x}}_{i,|c_i|}}_{|c_i|}, \underbrace{\hat{\mathbf{x}}_{k,1}, \dots, \hat{\mathbf{x}}_{k,j}, \dots, \hat{\mathbf{x}}_{k,|c_k|}}_{|c_k|}]。 \end{aligned} \quad (9)$$

式中:  $\hat{\mathbf{x}}_{i,j}$  为第  $i$  个簇 ( $i=1, 2, \dots, k, j=1, 2, \dots, |c_i|$ ) 中的第  $j$  个数;  $|c_i|$  为第  $i$  个簇的长度并且满足  $\sum_{i=1}^k |c_i| = L$ 。

步骤 3 更新每个质心  $\mathbf{c}_i = \frac{1}{|c_i|} \sum_{\hat{\mathbf{x}}_i \in c_i} \hat{\mathbf{x}}_i, 1 \leq i \leq k$ 。

步骤 4 重复步骤 2 和步骤 3, 直到满足收敛条件。

一般地,当前质心与上一次更新所获得质心之间的欧氏距离小于 0.001 时,终止聚类过程。由于本文研究重点为在线学习,故此采样过程只会发生在初始化阶段。基于采样所获得的样本,本文提出以下鲁棒  $K$ -Means 采样的 Nyström 递归最小广义对数核损失算法,即 NRMGLKL-KM。

### 2.3 基于 $K$ -Means 采样的 Nyström 递归最小广义对数核损失算法

为使输入数据自相关矩阵可逆,引入一个范数惩罚项,广义对数核损失如下:

$$J_i = \frac{1}{\alpha} \sum_{j=1}^i \beta^{i-j} \ln(1 + \alpha(1 - M_j)^{p/2}) + \beta^i \gamma \|\mathbf{Q}_i^z\|^2. \quad (10)$$

式中:  $M_j = \exp\left(-\frac{e_j^2}{2\sigma^2}\right)$  为非线性函数;  $e_j = d_j - (\mathbf{Q}_i^z)^T \cdot \mathbf{z}(u_j)$  为估计误差;  $\mathbf{Q}_i^z \in \mathbb{R}^m$  为  $m$  维权重向量。为简洁起见,  $\mathbf{z}(u_j)$  简写为  $\mathbf{z}_j$ 。

关于权重向量  $\mathbf{Q}_i^z$  的一阶偏导表示为

$$\nabla J_i = -\frac{p}{2\sigma^2} \sum_{j=1}^i \beta^{i-j} \theta_j \mathbf{z}_j e_j + \beta^i \gamma \mathbf{Q}_i^z - \frac{p}{2\sigma^2} \sum_{j=1}^i \beta^{i-j} \theta_j \mathbf{z}_j d_j + \frac{p}{2\sigma^2} \sum_{j=1}^i \beta^{i-j} \theta_j N_j \mathbf{Q}_i^z + \beta^i \gamma \mathbf{Q}_i^z. \quad (11)$$

式中:  $N_j = \mathbf{z}_j \mathbf{z}_j^T$ ,  $\theta_j = (1 - M_j)^{(p-2)/2} M_j \cdot (1 + \alpha(1 - M_j)^{p/2})^{-1}$ 。

定义:

$$\begin{cases} \mathbf{d}_i = [d_1, d_2, \dots, d_i]^T; \\ \mathbf{Y}_i = \text{diag}[\beta^{i-1} \theta_1, \beta^{i-2} \theta_2, \dots, \theta_i]; \\ \mathbf{Z}_i = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_i]. \end{cases} \quad (12)$$

令式(11)等于 0,得

$$\mathbf{Q}_i^z = \mathbf{P}_i \mathbf{\Psi}_i. \quad (13)$$

式中:  $\mathbf{P}_i = \mathbf{R}_i^{-1}$ ;  $\mathbf{R}_i = \sum_{j=1}^i \beta^{i-j} \theta_j N_j + \beta^i \gamma 2\sigma^2 / p \mathbf{I}$ ;  $\mathbf{I}$  为单位矩阵;  $\mathbf{\Psi}_i = \mathbf{Z}_i \mathbf{Y}_i \mathbf{d}_i = \sum_{j=1}^i \beta^{i-j} \theta_j \mathbf{z}_j d_j$ 。

基于指数衰减的数据窗口,  $\mathbf{R}_i$  和  $\mathbf{\Psi}_i$  的在线更新形式写为

$$\begin{cases} \mathbf{R}_i = \beta \mathbf{R}_{i-1} + \mathbf{N}_i; \\ \mathbf{\Psi}_i = \beta \mathbf{\Psi}_{i-1} + \mathbf{z}_i d_i. \end{cases} \quad (14)$$

进一步,根据矩阵求逆引理,利用递归形式来计算式(13)中的  $\mathbf{P}_i$ ,即

$$\begin{aligned} \mathbf{P}_i &= \beta^{-1} \left( \frac{\theta_i \mathbf{P}_{i-1}}{\beta + \theta_i \mathbf{z}_i^T \mathbf{P}_{i-1} \mathbf{z}_i} \right) \\ &= \beta^{-1} \left( \mathbf{P}_{i-1} - \frac{\beta^{-1} \theta_i \mathbf{P}_{i-1} \mathbf{N}_i \mathbf{P}_{i-1}}{\beta + \theta_i \mathbf{z}_i^T \mathbf{P}_{i-1} \mathbf{z}_i} \right). \end{aligned} \quad (15)$$

将式(15)代入式(13),权重更新可表示为

$$\begin{aligned} \mathbf{Q}_i^z &= \mathbf{Q}_{i-1}^z + \mathbf{Q}_i^z e_i = \mathbf{Q}_{i-1}^z + \frac{\theta_i \beta^{-1} \mathbf{P}_{i-1} \mathbf{z}_i e_i}{1 + \theta_i \beta^{-1} \mathbf{z}_i^T \mathbf{P}_{i-1} \mathbf{z}_i} \\ &= \mathbf{Q}_{i-1}^z + \mathbf{P}_i \mathbf{z}_i e_i. \end{aligned} \quad (16)$$

为简洁起见,令式(16)中  $\mathbf{Q}_i = \theta_i \beta^{-1} \mathbf{P}_{i-1} \mathbf{z}_i / q_i$  为增益向量,其中,  $q_i = 1 + \theta_i \beta^{-1} \mathbf{z}_i^T \mathbf{P}_{i-1} \mathbf{z}_i$ ;  $e_i = d_i - (\mathbf{Q}_{i-1}^z)^T \mathbf{z}_i$  为估计误差。

从式(10)~(16)中发现, NRMGLKL-KM 算法具有较好的灵活性,其中可调参数  $m$  直接影响滤波性能。 $m$  取值越大,算法滤波精度越高,但计算复杂度也会随之增加。因此,一个合适的  $m$  可以在滤波精度和计算复杂度之间进行很好的折中。NRMGLKL-KM 算法总结为①初始化参数  $\sigma, \alpha, \gamma, \beta, \mathbf{Q}_0^z = \mathbf{0}, \mathbf{P}_0 = \mathbf{0}$ ; ②当输入输出序列对存在时,进入循环体:计算  $u_i$  的高维映射  $\mathbf{z}_i = \mathbf{D}^{-\frac{1}{2}} \cdot \mathbf{V}^T \mathbf{K}(i, \cdot)^T$ , 计算误差  $e_i = d_i - (\mathbf{Q}_{i-1}^z)^T \mathbf{z}_i$ , 更新权重向量  $\mathbf{Q}_i^z = \mathbf{Q}_{i-1}^z + \mathbf{P}_i \mathbf{z}_i e_i$ , 更新矩阵  $\mathbf{P}_i = \beta^{-1} \left( \mathbf{P}_{i-1} - \frac{\beta^{-1} \theta_i \mathbf{P}_{i-1} \mathbf{N}_i \mathbf{P}_{i-1}}{\beta + \theta_i \mathbf{z}_i^T \mathbf{P}_{i-1} \mathbf{z}_i} \right)$ 。

### 3 仿真

本实验通过 Mackey-Glass (MG) 混沌时间序列预测来验证算法的预测能力。混沌时间序列的频谱分布状态十分丰富且能反映混沌系统中信号的激励特征,由以下非线性延迟微分方程获得:

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

式中:  $a = 0.2, b = 0.1, \tau = 30$ 。为实现数据离散化,采样周期设为 6 s。式(17)中,时延参数  $\tau$  越大,系统混沌程度越明显。从混沌系统中提取序列  $[x_{i-7}, x_{i-6}, \dots, x_{i-1}]^T$  用于估计期望输出  $d_i = x_i$ 。结合实际问题,混沌序列中需引入噪声干扰。

脉冲噪声环境由两个独立的噪声混合产生,表示为

$$v_i = (1 - b_i) v_{x,i} + b_i v_{y,i}. \quad (18)$$

式中:  $b_i$  由 Bernoulli 随机过程组成,其概率分布为  $P(b_i = 1) = c, P(b_i = 0) = 1 - c$  且发生概率  $c \in [0, 1]$ ;  $v_{x,i}$  为零均值且方差为  $\sigma_v^2 = 0.01$  的高斯白噪声,概率  $P\{x = 1\} = P\{x = -1\} = 0.5$ ;  $v_{y,i}$  采用参数向量为  $[0.8, 0, 1, 0]$  的  $\alpha$ -stable 噪声,用来表示异常值。

将 2 000 个受混合脉冲噪声干扰的数据作为训练集,用以验证算法的降噪能力,200 个无噪声数据作为测试集,额外的 200 个用作 Nyström 映射。基于 Monte Carlo (MC) 方法,仿真结果都取 50 次 MC 的平均值。作为性能衡量标准的方法之一,以 dB



为单位的均方误差  $mse = 10\lg(\sum_{n=1}^N (d_i - \hat{d}_i)^2 / N)$  被用于评估算法之间滤波精度的差异,其中  $N$  为测试数据的长度,  $d_i$  为期望输出,  $\hat{d}_i$  为估计量。

讨论样本数  $m$  对 NRMGLKL-KM 算法的性能影响,其中  $m$  的取值为  $[10, 100]$ 。图 2 展示了不同维度  $m$  对应的 NRMGLKL-KM 算法的稳态误差和平均消耗时间,其中稳态误差取最后 100 次迭代的  $mse$  均值,其他自由参数设置为  $\gamma = 0.1, \beta = 1, \alpha = 1, \sigma = 1, p = 10$ 。从图 2 中可以看出:①NRMGLKL-KM 算法的平均消耗时间随着  $m$  增大而增加;②在初始阶段随着  $m$  的增加,稳态误差下降比较明显,当  $m > 50$  时,稳态误差的值几乎无变化。因此,为了平衡滤波精度和时间消耗,在本例中  $m$  的取值为 50。

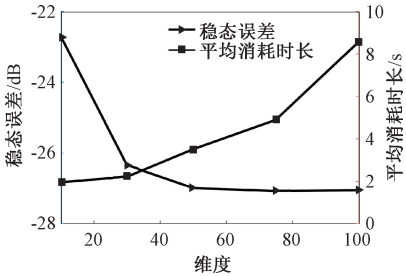


图 2 在不同维度下 NRMGLKL-KM 算法的稳态误差和平均消耗时长

Figure 2 Steady-state error and averaged consumed time of the NRMGLKL-KM with different dimension

讨论核宽  $\sigma$  对 NRMGLKL-KM 算法的滤波精度的影响。 $\sigma$  在区间  $[0.4, 2.0]$  以 0.2 为间距取值,其他自由参数分别设置为  $m = 50, p = 10, \alpha = 1.5, \gamma = 0.1, \beta = 1$ 。不同  $\sigma$  值对应的 NRMGLKL-KM 算法的稳态误差如图 3 所示,从图 3 中可以发现,当  $\sigma = 1.0$  时获得最佳滤波精度,在  $\sigma$  过大或者过小时均会导致滤波精度下降。

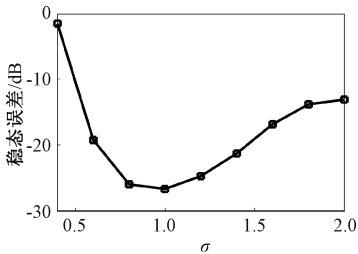


图 3 参数  $\sigma$  对 NRMGLKL-KM 性能的影响  
Figure 3 Influence of parameter  $\sigma$  on the performance of NRMGLKL-KM

讨论参数  $p$  和  $\alpha$  对 NRMGLKL-KM 算法性能的影响,其中  $p$  在区间  $[2, 16]$  以 2 为间距取值,  $\alpha$  在区间  $[0.5, 4.0]$  以 0.5 为间距取值。不同  $p$  和  $\alpha$  对应 NRMGLKL-KM 算法的稳态误差如图 4 所示,其他自

由参数为  $m = 50, \sigma = 1, \gamma = 0.1, \beta = 1$ 。从图 4 中可以发现,随着参数  $p$  的增大,稳态误差先下降后升高,并且在  $p = 10$  的附近达到最优;  $\alpha$  在一定程度上对所提算法的影响较低。因此,为了达到理想的性能,在本例中将 NRMGLKL-KM 算法的参数设置为  $m = 50, \sigma = 1, \alpha = 2, p = 10, \gamma = 0.1, \beta = 1$ 。

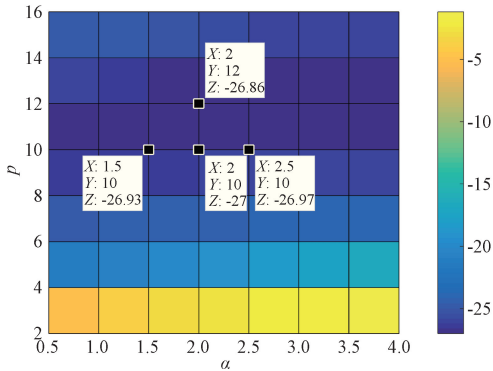


图 4 NRMGLKL-KM 稳态误差随  $p$  和  $\alpha$  的变化  
Figure 4 Steady-state error of NRMGLKL-KM versus  $p$  and  $\alpha$

为验证 NRMGLKL-KM 算法在鲁棒性和收敛性方面的优势,将 NRMGLKL-KM 算法与一些典型递归算法进行比较,如: NysKRLS-KM<sup>[18]</sup>, KRMC<sup>[21]</sup>, NysKRMC<sup>[20]</sup>, KRMC-NC<sup>[21]</sup>。为达到理想的精度,对所有算法均设置了最优参数,其中各算法中  $\gamma, \beta$  以及  $m$  的取值均为  $\gamma = 0.1, \beta = 1, m = 50$ 。仿真结果如图 5 所示,所有算法的仿真结果数据和计算复杂度列于表 1。从图 5 可以看出,与其他典型算法相比,所提算法在脉冲噪声中具有较高的鲁棒性,同时还具备更快的收敛速度和更高的滤波精度。表 1 中  $i$  为当前迭代次数,  $n$  为 KRMC-NC 在当前迭代中的字典大小。因为  $i$  和  $n$  的值最终会远大于  $m$ ,所以相比于 KRMC 和 KRMC-NC 算法,所提算法具有更小的计算复杂度。同时从表 1 中可以发现,所提算法能够在更短的时间内达到比 KRMC 算法更高的滤波精度。

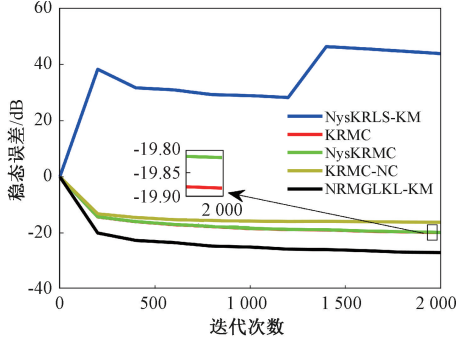


图 5 MG 时间序列下不同算法的稳态误差比较  
Figure 5 Comparison of the steady-state error of different algorithms in MG time series prediction

表 1 MG 时间序列下不同算法的仿真结果及计算复杂度

算法	字典大小	消耗时长/s	稳态误差/dB	加法	乘法
NysKRLS-KM	50	3.542 1	—	$5m^2+(2d-2)m+1$	$5m^2+(d+6)m+9$
KRMC( $\sigma=3.5$ )	2 000	43.331 0	-19.982 1	$2i^2+(2d+1)i+2d$	$2i^2+(d+9)i+d$
NysKRMC( $\sigma=3.5$ )	50	2.941 3	-19.843 8	$5m^2+(2d-2)m+1$	$5m^2+(d+6)m+12$
KRMC-NC( $\tau_1=0.01,\tau_2=0.1,\sigma=3.5$ )	676.7	4.302 5	-16.630 8	$2n^2+(4d-3)n+2d$	$2n^2+(3d+8)n+d$
NRMGLKL-KM( $\alpha=1.5,p=10,\sigma=1$ )	50	3.501 3	-26.986 6	$5m^2+(2d-2)m+5$	$5m^2+(d+6)m+22$

4 结论

本文将原始空间中的广义对数损失映射到再生核希尔伯特空间,提出了广义对数核损失准则(GLKL),并给出了该准则的若干性质。GLKL 准则在混沌时间序列预测中能有效地消除脉冲干扰。同时,在提出的 GLKL 准则中引入了  $K$ -Means 采样的 Nyström 映射,从而推导出一种新的基于  $K$ -Means 采样的 Nyström 递归最小 GLKL 预测方法,即 NRMGLKL-KM 算法。通过实验仿真确定自由参数对算法性能的影响,选择最优参数以提高滤波精度。除此之外,在脉冲噪声环境下进行混沌时间序列预测仿真。仿真结果表明:与稀疏化核自适应滤波算法相比,所提算法在脉冲噪声中具备更高的鲁棒性,同时,相比于他典型鲁棒核自适应滤波预测算法相比,该算法具备更快的收敛速度和更高的滤波精度。

参考文献:

[1] 牛莹,张勋才. 基于 Duffing 映射与遗传操作的图像加密方法[J]. 郑州大学学报(工学版), 2019, 40(4): 61-67.

NIU Y, ZHANG X C. Image encryption algorithm based on Duffing map and genetic operators[J]. Journal of Zhengzhou university (engineering science), 2019, 40(4): 61-67.

[2] 吴腾,张志利,赵军阳,等. 一种局部最佳阈值预测的自适应角点检测方法[J]. 计算机工程, 2018, 44(3): 270-274.

WU T, ZHANG Z L, ZHAO J Y, et al. An adaptive corner detection method of local optimal threshold prediction[J]. Computer engineering, 2018, 44(3): 270-274.

[3] 任天赐,黄向生,丁伟利,等. 全局双边网络的语义分割算法[J]. 计算机科学, 2020, 47(增刊1): 161-165.

REN T C, HUANG X S, DING W L, et al. Global bilateral segmentation network for segmantic segmentation[J]. Computer science, 2020, 47(S1): 161-165.

[4] 王伟然,闫景昊,杨冠军,等. 基于船舶磁悬浮减振装置的自适应预测控制[J]. 水下无人系统学报,

2021, 29(4): 383-390.

WANG W R, YAN J H, YANG G J, et al. Adaptive predictive control based on ship magnetic suspension vibration reduction device[J]. Journal of unmanned undersea systems, 2021, 29(4): 383-390.

[5] LIU W F, PRNCEPE J C, HAYKIN S. Kernel adaptive filtering[M]. Hoboken, USA: John Wiley & Sons, Inc., 2010.

[6] SHIN H C, SAYED A H. Mean-square performance of a family of affine projection algorithms[J]. IEEE transactions on signal processing, 2004, 52(1): 90-102.

[7] BHOTTO M Z A, ANTONIOU A. New improved recursive least-squares adaptive-filtering algorithms[J]. IEEE transactions on circuits and systems I: regular papers, 2013, 60(6): 1548-1558.

[8] PEI S C, TSENG C C. Least mean  $p$ -power error criterion for adaptive FIR filter[J]. IEEE journal on selected areas in communications, 1994, 12(9): 1540-1547.

[9] KURIAN N C, PATEL K, GEORGE N V. Robust active noise control: an information theoretic learning approach[J]. Applied acoustics, 2017, 117: 180-184.

[10] SYED M N, PARDALOS P M, PRINCIPE J C. On the optimization properties of the correntropic loss function in data analysis[J]. Optimization letters, 2014, 8(3): 823-839.

[11] ZHANG T, WANG S Y. Nyström kernel algorithm under generalized maximum correntropy criterion[J]. IEEE signal processing letters, 2020, 27: 1535-1539.

[12] CHEN B D, XING L, ZHAO H Q, et al. Generalized correntropy for robust adaptive filtering[J]. IEEE transactions on signal processing, 2016, 64(13): 3376-3387.

[13] CHEN B D, XING L, XU B, et al. Kernel risk-sensitive loss: definition, properties and application to robust adaptive filtering[J]. IEEE transactions on signal processing, 2017, 65(11): 2888-2901.

[14] WANG S Y, WANG W Y, XIONG K, et al. Logarithmic hyperbolic cosine adaptive filter and its performance analysis[J]. IEEE transactions on systems, man, and cybernetics: systems, 2021, 51(4): 2512-2524.

[15] 陈炳煌,缪希仁,江灏,等. 融合粒子群与极限学习

机的输电杆塔灾害分类方法[J]. 郑州大学学报(工学版), 2021, 42(4): 77-83.

CHEN B H, MIAO X R, JIANG H, et al. A method for disaster status classification of transmission line towers by integrating particle swarm optimization and extreme learning machine[J]. Journal of Zhengzhou university (engineering science), 2021, 42(4): 77-83.

[16] LIU W F, POKHAREL P P, PRINCIPE J C. The kernel least-mean-square algorithm [J]. IEEE transactions on signal processing, 2008, 56(2): 543-554.

[17] ZHANG Q Q, WANG S Y. Quantised kernel least mean square algorithm with a learning vector strategy[J]. Electronics letters, 2020, 56(21): 1146-1147.

[18] ZHANG T, WANG S Y, HUANG X W, et al. Kernel recursive least squares algorithm based on the Nyström method with  $k$ -means sampling[J]. IEEE signal processing letters, 2020, 27: 361-365.

[19] 胡燕, 朱晓瑛, 马刚. 基于  $K$ -Means 和时间匹配的位置预测模型[J]. 郑州大学学报(工学版), 2017, 38(2): 17-20.

HU Y, ZHU X Y, MA G. Location prediction model based on  $K$ -means algorithm and time matching[J]. Journal of Zhengzhou university (engineering science), 2017, 38(2): 17-20.

[20] WANG S Y, DANG L J, QIAN G B, et al. Kernel recursive maximum correntropy with Nyström approximation [J]. Neurocomputing, 2019, 329: 424-432.

[21] WU Z Z, SHI J H, ZHANG X, et al. Kernel recursive maximum correntropy[J]. Signal processing, 2015, 117: 11-16.

Kernel Adaptive Filtering Prediction Algorithm of Chaotic Time Series

LIU Qiang, WANG Shiyuan, HUANG Xuewei, WANG Daili

(College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China)

**Abstract:** In practical environment, chaotic time series often contain a lot of noise and outliers. Because of these interference factors, the prediction performance of the kernel adaptive filter based on the second-order similarity measure could decrease significantly in chaotic time series prediction. Based on the above problems, a robust kernel adaptive filter prediction algorithm for chaotic time series was proposed. The proposed algorithm based on the nonlinear similarity measure of the generalized logarithmic kernel loss function, could improve the robustness of the kernel adaptive filter in impulsive noise environment effectively. At the same time, the algorithm adopted adaptive  $K$ -Means sampling sparse Nyström nonlinear mapping method, which could fix the network size of the algorithm in advance, and thus reduce the computational complexity of the kernel adaptive filter algorithm. Through the recursive updating method, the proposed algorithm had faster convergence speed. Finally, the Mackey-Glass chaotic time series prediction simulation was carried out for the adaptive filter algorithms. The simulation results showed that, as a new robust  $K$ -Means sampling Nyström recursive minimum generalized logarithmic kernel loss prediction method, compared with sparse kernel adaptive filter algorithm, this algorithm had better robustness in impulsive noise. Compared with other typical robust kernel adaptive filter prediction algorithms, this algorithm had faster convergence rate and higher filtering accuracy.

**Keywords:** prediction of chaotic time series; kernel adaptive filtering; generalized logarithmic kernel loss; Nyström mapping; recursive update