

# 基于核自适应滤波器的时间序列在线预测研究综述

韩敏<sup>1</sup> 马俊珠<sup>1</sup> 任伟杰<sup>1</sup> 钟凯<sup>1</sup>

**摘要** 核自适应滤波器 (Kernel adaptive filter, KAF) 是时间序列在线预测的重点研究领域之一, 本文对核自适应滤波器的最新进展及未来研究方向进行了分析和总结. 基于核自适应滤波器的时间序列在线预测方法, 能较好地解决预测、跟踪问题. 本文首先概述了三类核自适应滤波器的基本模型, 包括核最小均方算法、核递归最小二乘算法和核仿射投影算法 (Kernel affine projection algorithm, KAPA). 在此基础上, 从核自适应滤波器在线预测的内容和机理入手, 综述基于核自适应滤波器的时间序列在线预测方法. 最后, 本文将介绍这一领域潜在的研究方向和发展趋势, 并展望未来的挑战.

**关键词** 核自适应滤波器, 时间序列在线预测, 核最小均方, 核递归最小二乘, 核仿射投影算法

**引用格式** 韩敏, 马俊珠, 任伟杰, 钟凯. 基于核自适应滤波器的时间序列在线预测研究综述. 自动化学报, 2021, 47(4): 730-746

**DOI** 10.16383/j.aas.c190051

## A Survey of Time Series Online Prediction Based on Kernel Adaptive Filters

HAN Min<sup>1</sup> MA Jun-Zhu<sup>1</sup> REN Wei-Jie<sup>1</sup> ZHONG Kai<sup>1</sup>

**Abstract** Kernel adaptive filter (KAF) is one of the significant research fields of time series online prediction. In this paper, the latest development and future research directions of kernel adaptive filter are analyzed and summarized. The time series online prediction algorithms based on kernel adaptive filter can better solve the problem of prediction and tracking. This paper first generalizes the basic models of three types of kernel adaptive filters, including kernel least mean square, kernel recursive least squares and kernel affine projection algorithm (KAPA). On this basis, starting from the content and mechanism of online prediction of kernel adaptive filter, the time online prediction methods for kernel adaptive filter are reviewed. Finally, this paper will introduce the potential research direction and development tendency in this field, and look forward to the challenges ahead.

**Key words** Kernel adaptive filter (KAF), time series online prediction, kernel least mean squares (KLMS), kernel recursive least squares (KRLS), kernel affine projection algorithm (KAPA)

**Citation** Han Min, Ma Jun-Zhu, Ren Wei-Jie, Zhong Kai. A survey of time series online prediction based on kernel adaptive filters. *Acta Automatica Sinica*, 2021, 47(4): 730-746

气象、水文、金融、医学以及工程等领域, 其数据多以时间序列的形式进行存储. 时间序列是某种变量或统计指标的数值随时间排序所形成的序列集合, 其中既含有全部变量的历史信息, 又含有参与系统动态演化的变量信息, 例如在气象数据中, 温度、湿度、风速、风向、气压、降雨量、日照时数、紫外线强度的变化等. 从时间序列数据的历史信息中预测未来的变化, 以提前做好防护措施, 对保障人类生命财产安全具有重要意义<sup>[1-6]</sup>.

近年来, 随着对时间序列研究的深入化, 传统

的线性自适应滤波算法<sup>[7]</sup>, 例如最小均方 (Least mean squares, LMS) 算法、递归最小二乘 (Recursive least squares, RLS) 算法等都不满足预测要求. LMS 收敛过程缓慢, 步长与收敛速度、失调之间也存在矛盾. RLS 具有计算复杂度较高、所需存储量较大的缺陷. 同时, 它们的应用实时性也较差, 在处理非线性、非平稳、高复杂性等问题时效果并不理想. 所以对非线性、非平稳、复杂系统的研究已经引起了许多国内外专家学者的广泛关注, 并取得了一定的研究成果.

对于上述问题, 出现的建模方法主要包括: 人工神经网络 (Artificial neural networks, ANN)<sup>[8-10]</sup>、支持向量机 (Support vector machines, SVM)<sup>[11]</sup>、核自适应滤波器 (Kernel adaptive filter, KAF)<sup>[12-13]</sup> 以及一些其他非线性方法. ANN 和 SVM 都属于离线方法, 不满足在线实时性要求. 而 KAF 是指线性自适应滤波器在核空间 (Kernel space, KS) 的

收稿日期 2019-01-21 录用日期 2019-09-24

Manuscript received January 21, 2019; accepted September 24, 2019

国家自然科学基金 (61773087) 资助

Supported by National Natural Science Foundation of China (61773087)

本文责任编辑 莫红

Recommended by Associate Editor MO Hong

1. 大连理工大学电子信息与电气工程学部 大连 116023

1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023

扩展<sup>[14-15]</sup>, 该过程满足在线预测要求, 对复杂时间序列预测具有实时输出的特点. 其中“核技巧”操作不需被显式地知道训练样本在特征空间内的映射<sup>[16]</sup>, 它直接通过核函数计算来完成特征空间的内积计算, 整个计算过程被简化, 该映射过程如图 1.

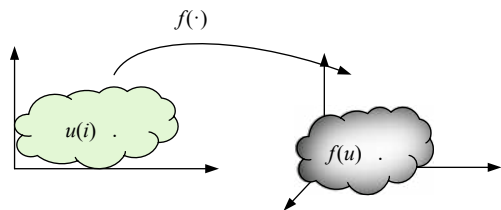


图 1 从输入空间到特征空间的非线性映射  $f(\cdot)$   
Fig.1 Nonlinear mapping  $f(\cdot)$  from input space to feature space

目前, 国内外学者使用最多的是高斯核函数. 核函数是先将输入数据映射到高维空间, 再在高维空间进行线性操作. 该过程对解决非线性问题具有明显的优势<sup>[17-18]</sup>. 随着对核函数<sup>[15-18]</sup>研究的深入, 学者们相继提出了应用于时间序列的计算复杂度较低、跟踪时变能力较强的 KAF 在线预测算法.

近年来, 由于时间序列组成的非线性系统越来越复杂, 简单的离线预测已经不能满足用户的要求. 对于上述问题, 多种时间序列在线预测方法相继出现. 目前, 时间序列在线预测方法主要分为以下几类: 重新建模方法、动态神经网络方法、在线支持向量回归方法、递归贝叶斯估计方法和 KAF 方法.

KAF 使用核方法实现非线性传递函数. 在 KAF 中, 信号被映射到高维线性特征空间, 并且非线性函数被近似为核的总和, 其域是特征空间. 如果这是在再生核希尔伯特空间 (Reproducing kernel Hilbert space, RKHS) 中完成的, 则核方法可以是非线性函数的通用逼近器. 内核方法具有凸损函数的优点, 同时没有局部最小值<sup>[13]</sup>. 近年来, 随着环境的不断复杂化, 其预测、跟踪问题也日趋复杂化. 在线预测作为一种实时预测、跟踪时变的有效方法, 在时间序列预测方面展现出较好的能效结果, 并满足用户的实时监测与控制的要求<sup>[19-21]</sup>. 基于 KAF 的在线预测方法在许多实际应用领域中都已展现出较好的预测性能, 其中主要包括三类: 核最小均方 (Kernel least mean squares, KLMS)<sup>[19]</sup> 算法、核递归最小二乘 (Kernel recursive least squares, KRLS)<sup>[20]</sup> 算法和核仿射投影算法 (Kernel affine projection algorithm, KAPA)<sup>[21]</sup>. 在此基础上, 许多学者也提出相应的改进算法, 根据 KAF 的发展进程, KAF 在线预测模型的研究进展具体如图 2 所示, 它详细地概括了文章的整体研究脉络.

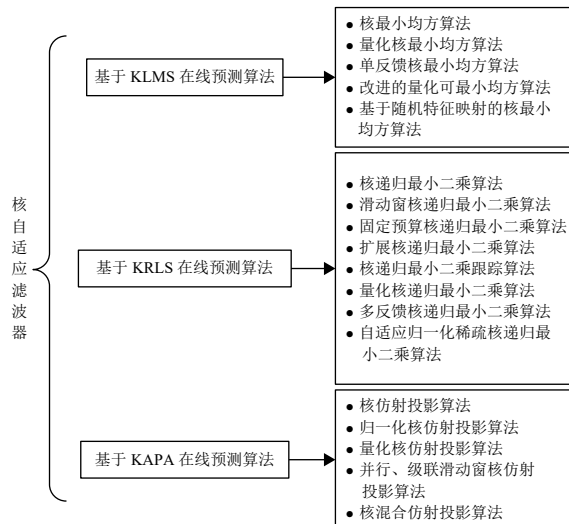


图 2 KAF 方法分类框图

Fig.2 Classification diagram of the KAF method

然而, KAF 方法在时间序列在线预测过程中还存在以下的问题: 随着新样本的加入, 系统所占用的内存不断增加 (主要表现为核矩阵维度的增加), 计算复杂度会随样本量的增加而增长. 针对该问题, 学者们对以上三类 KAF 方法进行了相应的改进<sup>[22-36]</sup>. 在一定程度上, 在线预测帮助研究者实现了对非线性、非平稳、复杂时间序列系统的实时预测, 对进一步将要发生的变化做出更科学、合理的决策.

本文的贡献主要集中在: 总结基于 KAF 的时间序列在线预测研究进展, 重点介绍基于 KAF 的时间序列在线预测方法, 主要包括 KLMS、KRLS 和 KAPA 三类, 在本文还介绍了这一研究领域的研究趋势和发展延伸, 并展望未来的挑战, 即充分理解基于 KAF 的时间序列在线预测的发展趋势, 以供进一步研究.

## 1 基于 KLMS 在线预测算法

对于 KLMS 中存在的计算复杂度和核矩阵的增长问题, 文献 [22] 提出一种量化核最小均方 (Quantization kernel least mean square, QKLMS) 算法, 其量化操作限制了核函数维度的无限增长, 基本思想是通过量化操作压缩输入数据, 而不像其他稀疏化方法那样直接丢弃冗余数据. 文献 [23] 提出了单反馈核最小均方 (Single feedback kernel least mean square, SF-KLMS) 算法, 它使用单个延迟输出以循环方式更新权重, 过去信息的使用也显著加快了收敛速度. 文献 [24] 提出了改进的量化核最小均方 (Modification quantization kernel least mean square, M-QKLMS) 算法, 它在使用预测误差的同时还用梯度下降方法来更新滤波器系

数. 不同于 QKLMS 只考虑预测误差, M-QKLMS 采用新的训练数据和预测误差来调整字典中最接近中心的系数. 文献 [25] 提出了一种基于随机特征网络的 KLMS 算法 (Kernel least mean square base on random feature networks, KLMS-RFN), 它与理论上隐式地将输入映射到无限维空间高斯核相反, 随机特征映射是将样本输入到相对低维的特征空间, 并使变换后的样本与使用移位不变内核的特征空间中近似等效.

KAF 作为在线预测模型, 已经在时间序列预测领域展现出较好的性能<sup>[18]</sup>. 以下将详细说明基于 KLMS 及其改进方法的时间序列在线预测机理.

### 1.1 核最小均方算法

目前, 对于复杂难预测环境, LMS 与核方法结合的应用方式随之出现. 2008 年, Liu 等将 LMS 嵌入 KS 提出 KLMS<sup>[19]</sup>. KLMS 是指在 RKHS 中进行的自适应过程, 它结合核技巧和 LMS 来实现样本到样本的更新<sup>[37]</sup>.

对于一个有限的训练数据集, KLMS 可通过未知映射将输入样本由低维的原始空间映射到高维特征空间, 再在高维特征空间执行简单的线性计算. 其中涉及步进参数的预设置, 主要有两种方法实现: 1) 根据经验设置步进参数; 2) 随着新样本的到来, 自适应更新权重  $\mathbf{w}(i)$  的同时实时改变步进参数.

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \mu \mathbf{e}(i) \boldsymbol{\varphi}(i) \quad (1)$$

其中,  $\mu$  是步进参数.  $\mathbf{e}(i)$  表示预测误差,  $\boldsymbol{\varphi}(i)$  表示在特征空间映射后的输入样本.

$$\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{w}^T(i-1) \boldsymbol{\varphi}(i) \quad (2)$$

其中,  $\mathbf{d}(i) = \boldsymbol{\varphi}^T(i) \mathbf{w}(i)$ .

上述过程在增加较少运算量的情况下加快了模型收敛速度, 可较快跟踪到自适应参数, 减少训练序列的预处理时间, 从而提高自适应参数的利用效率. 通过以上过程, 得到 KLMS 模型的预测输出  $\mathbf{y}(i)$ .

$$\mathbf{y}(i) = \sum_{j=1}^i \boldsymbol{\alpha}_j(i) k(\mathbf{u}(i), \mathbf{u}(j)) \quad (3)$$

其中,  $\boldsymbol{\alpha}_j(i) = \eta \mathbf{e}(j)$ ,  $\eta$  表示学习率.  $k(\cdot, \cdot)$  表示核函数, 该计算过程也被称为核技巧.

总的来说, KAF 在时间序列预测中是一个记忆强化操作. 同时 KAF 还属于在线操作范畴, 它通过使用之前的样本和预测误差来逐步构造出 KAF 的输出.

随着 KLMS 不断应用于时间序列预测, 其预测也出现了许多亟待解决的问题<sup>[38]</sup>. 随着预测样本的不断增多, 预测过程中出现计算复杂度较大的问题,

即随着新样本的不断到来, 核矩阵维数不断增大, 模型内存消耗也不断增大, 从而出现预测时间长、效率低、精度低等问题. 对于上述问题, 第 1.2 节 ~ 第 1.5 节对给出的多种解决方案进行详细说明.

### 1.2 量化核最小均方算法

2012 年 Chen 等在 KLMS 的基础上提出矢量量化 (Vector quantization, VQ) 操作, 得到量化核最小均方算法. VQ 方法不同于近似依赖准则 (Approximate linear dependence, ALD)<sup>[20]</sup>, 惊奇准则 (Surprise criterion, SC), 固定预算准则 (Fixed budget, FB)<sup>[39-40]</sup> 和新奇准则 (Novel criterion, NC) 等稀疏方法. VQ 作为稀疏<sup>[41-42]</sup> 的一种替代方法, 其基本思想是矢量量化并以此压缩输入或特征空间, 用以遏制自适应滤波器中高斯核结构的生长. QKLMS 中权重  $\boldsymbol{\Omega}(i)$  的更新过程如下所示:

$$\boldsymbol{\Omega}(i) = \boldsymbol{\Omega}(i-1) + \eta \mathbf{e}(i) Q[\boldsymbol{\varphi}(i)] \quad (4)$$

其中,  $Q[\cdot]$  是量化算子. 与稀疏方法不同, VQ 操作是使用“冗余”数据来更新最接近中心的系数, 该过程提高了数据的利用效率<sup>[22]</sup>.

QKLMS 的预测输出  $\mathbf{y}(i)$  如下式所示:

$$\mathbf{y}(i) = \sum_{j=1}^{\text{size}(\mathbf{C}(i-1))} \boldsymbol{\alpha}_j(i) k(\mathbf{C}_j(i-1), \mathbf{u}(i)) \quad (5)$$

其中,  $\mathbf{C}_j(\cdot)$  表示字典  $\mathbf{C}$  中的第  $j$  个元素.

QKLMS 已经应用于短时混沌时间序列预测<sup>[22]</sup>. 同时 VQ 利用“冗余”数据来更新最接近的中心系数, 以产生更紧凑的网络, 最终在时间序列在线预测上得到更高的预测精度和预测效率.

### 1.3 单反馈核最小均方算法

前述的 KLMS 和 QKLMS 都是前馈网络, 2015 年 Zhao 等首次提出具有单反馈结构的单反馈核最小均方 (SF-KLMS) 算法. 在 SF-KLMS 中只有一个延迟输出被用于以循环方式更新权重, 过去信息的使用显著地加快了模型的收敛速度. SF-KLMS 具有更紧凑、更高效的循环结构.

在 SF-KLMS 中包含两个权重更新过程: 1) 前馈权重  $\mathbf{w}(i)$ ; 2) 反馈权重  $\gamma(i)$ .

$$\mathbf{w}(i+1) = \mathbf{w}(i) + \frac{\eta_w(i)}{\rho_w(i)} \mathbf{e}(i) \mathbf{k}(i) + \frac{\eta_w(i)}{\rho_w(i)} \mathbf{e}(i) \beta_w(i) \gamma(i) \mathbf{D}_w(i-1) \quad (6)$$

$$\gamma(i+1) = \gamma(i) + \frac{\eta_\gamma(i)}{\rho_\gamma(i)} \mathbf{e}(i) \mathbf{y}(i-1) + \frac{\eta_\gamma(i)}{\rho_\gamma(i)} \mathbf{e}(i) \beta_\gamma(i) \gamma(i) \mathbf{D}_\gamma(i-1) \quad (7)$$



其中,  $\eta_w$  和  $\eta_\gamma$  是学习率,  $\rho_w$  和  $\rho_\gamma$  是归一化参数,  $\beta_w$  和  $\beta_\gamma$  是调整最新梯度信息的参数,  $\mathbf{D}_w(i) = \mathbf{k}(i) + \beta_w(i)\gamma(i)\mathbf{D}_w(i-1)$  是当前梯度信息,  $D_\gamma(i) = y(i-1) + \beta_\gamma(i)\gamma(i)D_\gamma(i-1)$ .

SF-KLMS 具有以下优势: 1) SF-KLMS 利用延迟输出的先前梯度信息  $\mathbf{D}_w(i-1)$  来更新当前梯度  $\mathbf{D}_w(i)$ ; 2) 与无反馈的 KAF (如 KLMS) 和含多反馈的 KAF (如 LRKOL) 相比, SF-KLMS 具有更紧凑的结构, 更快的收敛速度和更好的预测性能.

根据上述分析, 得到 SF-KLMS 模型的输出  $\mathbf{y}(i)$ .

$$\mathbf{y}(i) = \sum_{l=1}^m k(\mathbf{u}(i), \mathbf{u}(C_l)) \mathbf{w}_l + \gamma(i) \mathbf{y}(i-1) \quad (8)$$

其中,  $\mathbf{u}(C_l)$  表示字典的第  $l$  个信号.

在短时时间序列预测中, 调节因子  $\beta_w$  被设置成较小值, 而  $\beta_\gamma$  可被设置为相对较大的值用以提高反馈的效果<sup>[43]</sup>.

#### 1.4 改进的量化核最小均方算法

KAF 在线预测算法已经在解决复杂系统的预测、跟踪方面展现出强大的能力. 改进的量化核最小均方 (M-QKLMS) 算法于 2016 年被 Zheng 等提出, 该方法是在 QKLMS 方法的基础上进行的改进<sup>[24]</sup>. M-QKLMS 采用梯度下降方法来更新核自适应滤波器系数, 不同于 QKLMS 只考虑预测误差, 该方法还采用新训练数据与预测误差来共同调整字典最接近中心的系数.

计算新样本与现存字典之间的距离.

$$\text{dis}(\mathbf{u}(i), D(i-1)) = \min_{1 \leq l \leq D(i-1)} \|\mathbf{u}(i) - D_l(i-1)\| > \gamma \quad (9)$$

其中,  $\gamma$  是量化参数. 如果  $\text{dis}(\mathbf{u}(i), D(i-1)) \leq \gamma$ , 那么权重矩阵为  $\mathbf{w}_{l^*}(i) = \mathbf{w}_{l^*}(i-1) + \eta\kappa(D_{l^*}(i-1), \mathbf{u}(i))e(i)$ . 否则, 权重矩阵为  $\mathbf{w}(i) = [\mathbf{w}(i-1), \eta e(i)]$ .

根据上述判断结果, 得到输出  $\mathbf{y}(i)$ .

$$\mathbf{y}(i) = \sum_{j=1}^i \mathbf{w}_j(i-1) k(\mathbf{D}_j(i-1), \mathbf{u}(i)) \quad (10)$$

M-QKLMS 先执行 VQ 操作, 再利用梯度下降方法与预测误差共同作用, 从而有效地降低网络结构复杂度. 该过程提高了历史信息的利用效率<sup>[24]</sup>. 总之, M-QKLMS 利用新训练数据隐藏的信息, 达到了更好的预测精度.

#### 1.5 基于随机特征网络的核最小均方算法

为了在非静态环境中构建在线 KAF, Liu 等于

2018 年提出了基于随机特征网络的核最小均方算法<sup>[25]</sup>. KLMS-RFN 与将输入映射到高维特征空间的高斯核相比, 随机特征映射变换是将样本输入到相对低维的特征空间, 其中变换后的样本近似等于在特征空间中使用一个移位不变核, 如下所示:

$$\mathbf{k}(\mathbf{x}_1, \mathbf{x}_2) = \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2) \rangle \approx \varphi^T(\mathbf{x}_1)\varphi(\mathbf{x}_2) \quad (11)$$

其中,  $\varphi(\mathbf{x}) : \mathbf{R}^m \rightarrow \mathbf{R}^M$ , 表示特征映射后的输入矩阵.

然后, KLMS-RFN 模型的预测过程可总结为:

1) 计算随机傅立叶特征向量  $\varphi(\mathbf{x}(n))$

$$\varphi(\mathbf{x}(n)) = \begin{bmatrix} \cos(\mathbf{w}_1^T \mathbf{x}_n) \\ \vdots \\ \cos(\mathbf{w}_M^T \mathbf{x}_n) \\ \sin(\mathbf{w}_1^T \mathbf{x}_n) \\ \vdots \\ \sin(\mathbf{w}_M^T \mathbf{x}_n) \end{bmatrix}^T \quad (12)$$

2) 得到自适应滤波器的输出  $\mathbf{y}(n)$

$$\mathbf{y}(n) = \mathbf{w}^T(n-1)\varphi(\mathbf{x}(n)) \quad (13)$$

3) 得到权重更新矩阵  $\mathbf{w}(n)$

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \eta e(n)\varphi(\mathbf{x}(n)) \quad (14)$$

其中,  $e(n) = d(n) - y(n)$  表示预测误差.

KLMS-RFN 不需要存储重要样本, 可节省大量存储成本. 当所选样本较多时, KLMS-RFN 的计算复杂度明显低于核矩阵的计算复杂度. 此外, KLMS-RFN 还可以约束不断增长的权重网络的结构, 同时更新样本来跟踪输入的特征变化. 该模型已经在时间序列得到应用, 明显的提高了模型在非平稳环境下的预测性能<sup>[25]</sup>.

#### 1.6 基于 KLMS 在线预测的总结分析

通过对 KLMS 及其改进模型的说明, 就降低核矩阵维数、提高历史信息的利用效率方面来说, 可以看出 KAF 在时间序列在线预测领域已得到广泛的应用<sup>[44-46]</sup>. 在本节, 主要讨论基于 KLMS 在时间序列在线预测中出现的核矩阵计算量大, “冗余”信息利用效率低等问题的解决方法. 针对上述问题, 通过 VQ 来提高历史信息的利用效率, 并在此基础上采用预测误差来自适应更新滤波器结构, 降低核矩阵的维数.

KLMS 是在高维特征空间中推导的, 它使用随机梯度方法解决最小二乘问题, 从 Hadamard 意义上讲, KLMS 不需要明确的正则化. KLMS 及相

应的改进模型已经在复杂时间序列中得到广泛的应用<sup>[19, 22, 24, 47]</sup>. KAF 在一定程度上缩短了预测时间, 简化了核矩阵的计算步骤, 降低了计算复杂度, 也通过使用“冗余”样本, 提高了信息的利用效率, 提高了时间序列在线预测精度.

## 2 基于 KRLS 在线预测算法

对于 KRLS 建模中出现的在线样本量和核矩阵维数的限制问题, 文献 [26] 提出滑动窗口核递归最小二乘方法 (Sliding window kernel recursive least squares, SW-KRLS), 它只选取固定长度的序列样本进行计算, 从而降低核矩阵维度. 文献 [27] 提出扩展核递归最小二乘 (Extended kernel recursive least squares, EX-KRLS) 算法, 该方法可以在 KS 内表达状态空间模型, 还可以利用状态空间模型动态表示在时变环境条件下的时间序列的变化趋势. 文献 [28] 提出固定预算核递归最小二乘 (Fixed budget kernel recursive least squares, FB-KRLS) 算法, 它只选取固定数量样本, 且保留相关性较强的数据样本进行计算, 抛弃相关性弱的数据样本, 该过程不仅降低了核矩阵的维度, 而且提高了预测精度. 文献 [29] 提出核递归最小二乘跟踪 (Kernel recursive least squares with tracker, KRLS-T) 算法, 包括遗忘因子原则和数值稳定的方式, 它固定了内存量和每一个时间步长的计算, 限制了核矩阵的维数, 并以自然的方式结合正则化, 从而跟踪每次预测的时变特性. 与文献 [22] 类似, 文献 [30] 提出了量化核递归最小二乘 (Quantization kernel recursive least squares, QKRLS) 算法, 其基本思想也是通过量化操作压缩输入数据, 限制核矩阵维数, 该方法将“冗余”数据利用起来, 用以提高预测精度. 文献 [31] 提出了多反馈核递归最小二乘 (Multi-feedback kernel recursive least squares, MF-KRLS) 算法, 它是利用多个先前输出来递归更新结构参数, 具有跟踪时变的特性. 文献 [32] 提出自适应归一化稀疏核递归最小二乘 (Adaptive normalized sparse kernel recursive least squares, ANS-KRLS) 算法, 它结合一致性准则、ALD 和动态自适应调整来限制核矩阵维数, 同时提高了预测精度.

目前, KRLS 已经广泛应用于数据挖掘、机器学习等领域<sup>[20]</sup>. KRLS 将输入样本通过 Mercer 核映射到高维特征空间, 然后在高维特征空间中利用矩阵反演引理执行线性回归, 从而在线解决复杂时间序列预测问题<sup>[48-57]</sup>. 以下将详细说明基于 KRLS 及其改进方法的时间序列在线预测机理.

### 2.1 核递归最小二乘算法

KRLS 利用矩阵反演引理简化核矩阵的求逆过程, 目的在于简化计算, 并降低存储要求、提高预测效率. 目前, KRLS 已经在复杂时间序列预测中展现出较好的预测性能<sup>[20]</sup>, 它能够以在线的形式实时处理样本序列.

KRLS 模型的 ALD 阈值是根据经验获得, 太大或太小都会降低预测精度<sup>[20]</sup>. 当样本进入 KRLS 模型后, 计算该样本与当前字典在特征空间的线性扩展距离  $\delta(i)$ , 如果大于  $\nu$ , 那么该样本满足要求被加入字典进行预测. 反之, 如果小于  $\nu$ , 那么该数据样本被移除, 这保证了核矩阵维数的有限性.

$$\delta(i) = \left\| \sum_{k=1}^{i-1} \alpha(k) \varphi(\mathbf{x}(k)) - \varphi(\mathbf{x}(i)) \right\|^2 > \nu \quad (15)$$

其中,  $\nu$  是设置的 ALD 阈值.

根据以上过程, 得到权重矩阵  $\mathbf{w}(i)$ .

$$\mathbf{w}(i) = \left[ \lambda \mathbf{I} + \Phi(i)^T \Phi(i) \right]^{-1} \Phi(i) \mathbf{d}(i) = \Phi(i) \times \left[ \lambda \mathbf{I} + \Phi(i)^T \Phi(i) \right]^{-1} \mathbf{d}(i) \quad (16)$$

其中,  $\lambda$  是正则化参数,  $\Phi(i)$  表示特征空间映射后的输入矩阵,  $\mathbf{I}$  表示单位矩阵,  $\mathbf{d}(i)$  表示期望输出. 该过程使用矩阵反演引理.

再将  $\mathbf{w}(i)$  代入输出模型, 得到输出结果  $f(\mathbf{u}_*)$ .

$$f(\mathbf{u}_*) = \sum_{j=1}^i \alpha_j(i) k(\mathbf{u}(j), \mathbf{u}_*) \quad (17)$$

根据上述过程对预测结果进行分析<sup>[48-51]</sup>, 得到复杂系统下一步的发展趋势及实时信息, 最终做出更科学的决策.

与第 1 节的 KLMS 相比, 就收敛速度而言, KRLS 模型比 KLMS 模型小一个数量级, 但是这种性能的改善是以增加计算复杂度为代价的. 所以, 如何改善预测模型的计算复杂度? 第 2.2 节 ~ 第 2.8 节对给出的多种改进方法进行详细说明.

### 2.2 滑动窗口核递归最小二乘算法

滑动窗口核递归最小二乘方法<sup>[26]</sup> 是 Van Vaerenbergh 等于 2006 年提出. SW-KRLS 是结合滑动窗和传统 L2 范数正则化得到的. SW-KRLS 中滑动窗对核矩阵维数的要求是固定而非限制, 是始终保持字典的维度不变. L2 范数正则化则是改进了模型的泛化能力. 该方法多用于无记忆非线性系统, 并在时变环境、跟踪突变问题上体现出较好的预测性能. 在 SW 方法中, 选定窗口  $n$ , 得到正则化核矩

阵  $\mathbf{K}_n$ .

$$\mathbf{K}_n = \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T + c\mathbf{I} \quad (18)$$

其中,  $c$  是正则化参数,  $\mathbf{I}$  表示单位矩阵, 观测矩阵被表示为  $\mathbf{X}_n = [X_n, X_{n-1}, \dots, X_{n-N+1}]^T$ . 当处于在线环境时, 核矩阵的维数取决于观测值  $N$ .

如何保持核矩阵维度固定不变? 对于新到来样本, 核矩阵添加新行和新列通过式 (20) 实现, 核矩阵移除最旧行和列可通过式 (19) 实现.

$$\mathbf{K}^{-1} = \begin{bmatrix} \mathbf{A}^{-1}(\mathbf{I} + \mathbf{b}\mathbf{b}^T\mathbf{A}^{-1}\mathbf{H}g) & -\mathbf{A}^{-1}\mathbf{b}g \\ -(\mathbf{A}^{-1}\mathbf{b})^T g & g \end{bmatrix} \quad (19)$$

$$\mathbf{D}^{-1} = \mathbf{G} - \frac{\mathbf{f}\mathbf{f}^T}{e} \quad (20)$$

其中,  $\mathbf{A}$  是非奇异矩阵,  $g = (d - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b})^{-1}$ ,  $\mathbf{K}$  是添加新行和列后的矩阵,  $\mathbf{D}$  是移除最旧行和列后的矩阵,  $\mathbf{b}^T\mathbf{f} + dg = 1$ .

根据以上过程, 得到系数矩阵  $\alpha_n$ .

$$\alpha_n = \mathbf{K}_n^{-1}\mathbf{y}_n \quad (21)$$

其中,  $\mathbf{y}(n) = [y_n, y_{n-1}, \dots, y_{n-N+1}]$  表示观测向量,  $\mathbf{K}_n^{-1} = (\mathbf{K} + c\mathbf{I})$ .

SW-KRLS 采用正则化解决了过拟合问题, 并结合滑动窗口 (Sliding window, SW) 和矩阵反演引理来解决计算复杂度的限制问题, 不仅节省了内存空间, 简化了计算, 也提高了预测精度<sup>[26]</sup>.

### 2.3 扩展核递归最小二乘算法

2009 年 Liu 等又提出了在 RKHS 中的扩展递归最小二乘算法的核化形式, 即扩展核递归最小二乘算法<sup>[27]</sup>. EX-KRLS 在 RKHS 中首次实现了一般线性状态模型.

EX-KRLS 使用显式状态转移和卡尔曼滤波器来学习内核回归权向量, 也通过样本数据完成在线学习<sup>[27]</sup>. 当  $\mathbf{A} = \alpha\mathbf{I}$  时, 得到的状态转换模型 (22) 与观测模型 (23).

$$\mathbf{x}(i+1) = \alpha\mathbf{x}(i) + \mathbf{n}(i) \quad (22)$$

$$\mathbf{d}(i) = \boldsymbol{\varphi}^T(i)\mathbf{x}(i) + \mathbf{v}(i) \quad (23)$$

其中,  $\boldsymbol{\varphi}(i) \in \mathbf{F}$  代替  $\mathbf{u}(i) \in \mathbf{U}$ ,  $\alpha$  是一个升降参数, 既可放大也可衰减.

EX-KRLS 具有明确的状态转移过程, 且可以提高 KRLS 的跟踪能力<sup>[27]</sup>, 其跟踪模型如式 (24).

$$\mathbf{a}(i) = \begin{bmatrix} \mathbf{a}(i-1) - \mathbf{z}(i)\mathbf{r}^{-1}(i)e(i) \\ \mathbf{r}^{-1}(i)e(i) \end{bmatrix} \quad (24)$$

其中,  $\mathbf{r}(i) = \lambda + \boldsymbol{\varphi}^T(i)\boldsymbol{\varphi}(i) - \mathbf{z}^T(i)\mathbf{k}(i)$ ,  $\lambda$  是用以控制初始状态向量的正则化参数,  $\mathbf{z}(i)$  表示非奇异矩阵,  $\mathbf{k}(i)$  表示核矩阵,  $e(i)$  是预测误差,  $\mathbf{a}(i)$  是系数矩阵, 它们都是根据样本点不断地自适应更新. 所以, EX-KRLS 适合于慢衰落、状态变化小的非线性系统建模.

### 2.4 固定预算核递归最小二乘算法

对于计算复杂度较大问题, 固定预算核递归最小二乘方法于 2010 年被 Van Vaerenbergh 等提出<sup>[28]</sup>. 其中 FB 准则在限制内存的同时, 不仅减少预测时间, 还提高历史信息的利用效率.

FB 准则是 Dekel 等在 2006 年提出的. 在预测时间序列样本时, 随着新样本的加入, 直到字典达到预设阈值  $M$  之前, FB 准则不起作用. 当样本达到  $M+1$  时, 核矩阵如式 (25), FB 起作用, 抛弃最不重要、不相关的数据如式 (26), 期间字典大小始终为  $M$ , 即当新样本满足条件时, 就删除最不相关的数据. 反之, 新样本如果不满足条件, 那么新样本就不被允许加入字典, 即字典保持原来的状态不变. 该过程不仅限制了输入数据的维数, 也提高了数据利用效率.

$$\hat{\mathbf{K}}_n^{-1} = \begin{bmatrix} \mathbf{K}_{n-1}^{-1} + \mathbf{g}\mathbf{e}\mathbf{e}^T & -\mathbf{g}\mathbf{e} \\ -\mathbf{g}\mathbf{e}^T & \mathbf{g} \end{bmatrix} \quad (25)$$

$$\hat{\mathbf{K}}_n^{-1} = \begin{bmatrix} \mathbf{e} & \mathbf{f}^T \\ \mathbf{f} & \mathbf{G} \end{bmatrix} \Rightarrow \mathbf{K}_n^{-1} = \mathbf{G} - \frac{\mathbf{f}\mathbf{f}^T}{e} \quad (26)$$

其中,  $\hat{\mathbf{K}}_n$  表示添加新行和列的核矩阵,  $\mathbf{e} = \mathbf{K}_{n-1}^{-1}\mathbf{b}$ ,  $\mathbf{g} = (d - \mathbf{b}^T\mathbf{e})^{-1}$ ,  $\mathbf{b}^T\mathbf{f} + dg = 1$ ,  $\mathbf{K}_n$  表示移除任意行和列的核矩阵.

为保持核矩阵维度不变, FB 准则中的抛弃计算如下所示:

$$d(\mathbf{x}_i, \mathbf{y}_i) = \frac{|\alpha_i|}{\left[\hat{\mathbf{K}}_n^{-1}\right]_{i,i}} \quad (27)$$

根据上述抛弃准则, FB-KRLS 模型可以有效地跟踪时变序列的信息<sup>[28]</sup>. FB-KRLS 主要应用于非线性系统识别问题. 它不仅能够保持内存不变, 抛弃最不重要的样本点, 提高信息利用效率, 而且其标签更新过程也表现出模型的跟踪时变特性, 更拓宽了 FB-KRLS 的应用范围<sup>[28]</sup>.

### 2.5 核递归最小二乘跟踪算法

不同于 SW-KRLS、EX-KRLS 和 FB-KRLS, 核递归最小二乘跟踪<sup>[29]</sup> 是引入贝叶斯框架, 在统一现有的 KRLS 理论下, 提出的“遗忘”的概念, 即通过明确方式处理复杂的时间序列数据. 为保证最优



性能, 需要确定核参数、正则化参数, 以及在非平稳环境中的遗忘因子.

由于 KRLS-T 等价于特定时空协方差的高斯过程 (Gaussian processing, GP) 回归模型, 所以可通过 GP 超参数估计来确定 KRLS-T 的所有参数. 采用潜在函数边缘化的方法, 用超参数表示观测数据的概率<sup>[29]</sup>.

$$\ln p(y|\delta, \sigma, \varepsilon) = -\frac{1}{2} \mathbf{y}^T (\mathbf{\Lambda} \circ \mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} - \frac{1}{2} |\mathbf{\Lambda} \circ \mathbf{K} + \sigma^2 \mathbf{I}| - \frac{n}{2} \ln(2\pi) \quad (28)$$

其中,  $\circ$  表示 Hadamard 乘积,  $\mathbf{\Lambda}$  是一个对角矩阵, 它的第  $j$  个对角元素值为  $\lambda^{\frac{|j|}{2}}$ ,  $\lambda \in (0, 1]$ .

在 KRLS-T 中, 需要存储和更新三个变量: 1) 字典中与基对应的当前预测均值  $\boldsymbol{\mu}_t$ ; 2) 表示先验值的核逆矩阵  $\mathbf{Q}_t = \mathbf{K}_t^{-1}$ ; 3) 相应的预测方差  $\Sigma_t$  它可以捕获后验协方差.

$$\boldsymbol{\mu}_t = \frac{y_t k(\mathbf{x}_t, \mathbf{x}_t)}{\sigma_n^2 + k(\mathbf{x}_t, \mathbf{x}_t)} n$$

$$\mathbf{K}_t^{-1} = \mathbf{Q}_t = k(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon \delta_{ij}$$

$$\mathbf{h}_{t+1} = \sum_t \mathbf{q}_{t+1}$$

其中,  $\mathbf{q}_{t+1} = \mathbf{Q}_t \mathbf{k}_{t+1}$ ,  $\varepsilon$  是遗忘因子且  $\varepsilon \in (0, 1)$ ,  $\sigma^2$  是超参数,  $\delta$  是正则化参数.

因为 KRLS-T 不能提供对过去时刻的预测, 仅保留与当前时刻相关的小部分样本, 所以对短期时间序列预测有更好的预测效果. 同时这样的预测机制极大降低了计算成本<sup>[29]</sup>.

## 2.6 量化核递归最小二乘算法

Chen 等于 2013 年又将 VQ 应用到 KRLS 模型, 它是以递归形式被推导得到量化核递归最小二乘算法<sup>[30]</sup>. QKRLS 也是将输入空间量化分配, 其基本思想也是使用较小数据集来表示整个输入数据的信息, 从而提高预测精度<sup>[23, 30]</sup>.

VQ 具有多种优势: 1) 计算简单; 2) 量化的大小比较容易选择; 3) 冗余数据可以很容易的被利用, 用以提高预测性能.

首先在线 VQ 需要设置量化尺码  $\varepsilon$ , 然后计算  $\mathbf{u}(i)$  和  $\mathbf{C}(i-1)$  之间的距离.

$$\text{dis}(\mathbf{u}(i), \mathbf{C}(i-1)) = \|\mathbf{u}(i) - \mathbf{C}_{j^*}(i-1)\| \quad (29)$$

其中,  $j^* = \arg \min_{1 \leq j \leq |\mathbf{C}(i-1)|} \|\mathbf{u}(i) - \mathbf{C}_j(i-1)\|$ ,  $\mathbf{C}_j(i-1)$  表示  $\mathbf{C}(i-1)$  的第  $j$  个元素.

当新样本  $\mathbf{u}(i)$  进入预测模型时, 如果  $\text{dis}(\mathbf{u}(i),$

$\mathbf{C}(i-1))$  不大于  $\varepsilon$ , 此时的量化码本保持不变, 用式 (30) 更新权重系数矩阵. 如果  $\text{dis}(\mathbf{u}(i), \mathbf{C}(i-1))$  小于  $\varepsilon$ , 那么新样本被加入到量化码本  $\mathbf{C}(\cdot)$  中, 用式 (31) 更新权重系数.

$$\tilde{\boldsymbol{\alpha}}^*(i) = \tilde{\boldsymbol{\alpha}}^*(i-1) + \frac{\mathbf{P}_{j^*}(i-1) [\mathbf{y}(i) - \tilde{\mathbf{K}}_{j^*}^T(i-1) \tilde{\boldsymbol{\alpha}}^*(i-1)]}{1 + \tilde{\mathbf{K}}_{j^*}^T(i-1) \mathbf{P}_{j^*}(i-1)} \quad (30)$$

$$\tilde{\boldsymbol{\alpha}}^*(i) = \begin{bmatrix} \tilde{\boldsymbol{\alpha}}^*(i-1) - \mathbf{z}_\Lambda(i) \mathbf{r}^{-1}(i) \mathbf{e}(i) \\ \mathbf{r}^{-1}(i) \mathbf{e}(i) \end{bmatrix} \quad (31)$$

其中,  $\mathbf{z}_\Lambda(i) = \mathbf{P}(i-1) \mathbf{\Lambda}(i-1) \mathbf{h}(i)$ ,  $\mathbf{e}(i) = \mathbf{y}(i) - \mathbf{h}^T(i) \tilde{\boldsymbol{\alpha}}^*(i-1)$ ,  $\mathbf{z}(i) = \mathbf{P}^T(i-1) \mathbf{h}(i)$ ,  $\mathbf{r}(i) = \gamma + \boldsymbol{\kappa}(\mathbf{u}(i), \mathbf{u}(i)) - \mathbf{h}^T(i) \mathbf{z}_\Lambda(i)$ .

QKRLS 在预测时间序列时, 除了能够保持期望性能, 在产生紧凑模型方面也非常有效<sup>[30]</sup>. 计算复杂度和预测精度之间的折衷也可以通过改变量化码本的大小进行简单、有效地控制.

## 2.7 多反馈核递归最小二乘算法

近年来, 大多数 KAF 都是无反馈结构的预测模型<sup>[26-30]</sup>. 在第 1.3 节, SF-KLMS 首次将反馈结构引入 KAF 模型. 随后 Wang 等在 2017 年提出了具有反馈结构的多反馈核递归最小二乘算法<sup>[31]</sup>. MF-KRLS 是将延迟输出与最近的输入结合起来估计最近输出的过程.

在 MF-KRLS 结构中, 它的前馈系数 (Feed forward coefficient, FFC) 和反馈系数 (Feedback coefficient, FBC) 利用最速下降法进行更新. 同时 MF-KRLS 采用多个先前的输出以递归的形式更新结构参数. FFC 和 FBC 都采用最速下降法进行在线更新.

$$\boldsymbol{\alpha}(i+1) = \begin{bmatrix} \boldsymbol{\alpha}(i) \\ 0 \end{bmatrix} + \frac{\eta_\alpha(i)}{\rho_\alpha(i)} \mathbf{D}_\alpha(i) \mathbf{e}(i) \quad (32)$$

$$\boldsymbol{\lambda}(i+1) = \begin{bmatrix} \boldsymbol{\lambda}(i) \\ 0 \end{bmatrix} + \frac{\eta_\lambda(i)}{\rho_\lambda(i)} \mathbf{D}_\lambda(i) \mathbf{e}(i) \quad (33)$$

其中,  $\mu_\alpha(i) = \eta_\alpha(i) / \rho_\alpha(i)$  表示 FFC 的步进参数,  $\mu_\lambda(i) = \eta_\lambda(i) / \rho_\lambda(i)$  表示 FBC 的步进参数.  $\mathbf{D}_\alpha(i) = \partial \mathbf{y}(i) / \partial \boldsymbol{\alpha}(i)$ ,  $\mathbf{D}_\lambda(i) = \partial \mathbf{y}(i) / \partial \boldsymbol{\lambda}(i)$ .

根据上述分析, 在时变环境中, MF-KRLS 具有较好的自适应更新调整能力, 在估计精度和收敛速率上也得到改善<sup>[31]</sup>.

将 FFC 和 FBC 代入输出模型得到  $\mathbf{y}(i)$ .

$$\mathbf{y}(i) = \sum_{j=1}^i k(\mathbf{u}(i), \mathbf{u}(C_j)) \boldsymbol{\alpha}(j) + \lambda(j) \mathbf{y}(i-1) \quad (34)$$

在 MF-KRLS 的时间序列在线预测应用中, 其前馈和反馈结构的有效结合不仅提高了 KRLS 的滤波精度, 还加快了收敛速度<sup>[31]</sup>.

## 2.8 自适应归一化稀疏核递归最小二乘算法

随着在线稀疏方法的不断发展, Han 等于 2017 年提出自适应归一化稀疏核递归最小二乘方法<sup>[32]</sup>, 该方法已经在多元混沌时间序列在线预测方面得到较好的预测性能.

ANS-KRLS 是结合动态调整策略和一致性准则 (Coherence criterion, CC), 并基于 KRLS 得到的改进模型. 在 ANS-KRLS 中, 一致性准则不仅可以减少核矩阵维度, 还可以降低模型的计算复杂度. 动态调整策略在时变环境中具有自适应调整权值的能力.

首先, 动态调整过程权重系数  $\tilde{\boldsymbol{\alpha}}(i)$  的更新过程如下所示:

$$\tilde{\boldsymbol{\alpha}}(i) = \tilde{\boldsymbol{\alpha}}(i-1) + \frac{\mathbf{q}(i) \tilde{\mathbf{K}}^{-1}(i) \left( \mathbf{y}(i) - \tilde{\mathbf{k}}_{i-1}^T(\mathbf{x}(i)) \tilde{\boldsymbol{\alpha}}(i-1) \right)}{\varepsilon + \left\| \tilde{\mathbf{k}}_{i-1}^T(\mathbf{x}(i)) \right\|^2} \quad (35)$$

其中,  $\mathbf{q}(i) = \frac{\mathbf{q}(i) \boldsymbol{\alpha}(i)}{1 + \boldsymbol{\alpha}^T(i) \mathbf{q}(i) \boldsymbol{\alpha}(i)}$ ,  $\mathbf{q}(i)$  是正则自相关矩阵,  $\varepsilon$  用以避免误调的发生.

其次, 一致性稀疏准则如下式所示:

$$\mu = \max_{j=1,2,\dots,i-1} |k(\mathbf{x}(i), \mathbf{x}(j))| \quad (36)$$

其中, 当  $\max_{j=1,2,\dots,i-1} |k(\mathbf{x}(i), \mathbf{x}(j))| \leq \mu_0$ ,  $\mathbf{x}(j)$  被加入字典. 否则, 字典保持不变, 且  $\mu \in (0, 1)$ .

当新样本到来时, 如果  $\mathbf{x}(j)$  不满足式 (15) 且式 (36) 不大于  $\mu_0$  时, 新样本  $\mathbf{x}(j)$  被加入字典, 权重系数根据式 (37) 进行更新. 如果  $\mathbf{x}(j)$  不满足式 (15), 或式 (36) 大于  $\mu_0$  时, 字典大小保持不变, 此时根据式 (35) 更新权重系数.

$$\tilde{\boldsymbol{\alpha}}(i) = \frac{1}{\delta(i)} \begin{bmatrix} \tilde{\boldsymbol{\alpha}}(i-1) - \tilde{\boldsymbol{\alpha}}(i) \left( \mathbf{y}(i) - \tilde{\mathbf{k}}_{i-1}^T(\mathbf{x}(i)) \tilde{\boldsymbol{\alpha}}(i-1) \right) \\ \left( \mathbf{y}(i) - \tilde{\mathbf{k}}_{i-1}^T(\mathbf{x}(i)) \tilde{\boldsymbol{\alpha}}(i-1) \right) \end{bmatrix} \quad (37)$$

其中,  $\delta(i)$  表示 ALD 阈值.

总之, ANS-KRLS 在时间序列在线预测可通过

上述稀疏过程降低计算复杂度. ANS-KRLS 还具有在线操作和在时变环境中自适应调整的能力. 所以, ANS-KRLS 在提高预测精度和效率方面更具优势<sup>[32]</sup>.

## 2.9 基于 KRLS 在线预测的总结分析

通过以上对 KRLS 及其改进模型的介绍, 就基于 KRLS 时间序列在线预测方法中改进的 KAF 模型而言, SW-KRLS、FB-KRLS、QKRLS 和 ANS-KRLS 是以不同类型的“稀疏”方式来简化核矩阵维数, 从而降低计算复杂度、提高预测精度. EX-KRLS、KRLS-T 除了降低计算复杂度和提升预测精度外, 在跟踪时变方面也展现出优越的性能. 而 MF-KRLS 是以添加反馈结构的形式来提高模型的预测精度.

KRLS 改进方法的自相关矩阵在后期运算时会出现秩亏的现象, 引入正则化可以解决秩亏的问题. 和第 1 节的 KLMS 模型相比, KRLS 的不同之处在于需要明确的正则化. 但该正则化过程也存在以下问题: 随着数据样本的增加, 正则化的效率会开始缓慢地减弱.

从以上分析可知, 基于 KAF 的时间序列在线预测在降低计算复杂度、提高预测精度和效率, 以及在跟踪时变方面都展现出较好的优势<sup>[58-60]</sup>.

## 3 基于 KAPA 在线预测算法

对于 KAPA 建模中出现的计算复杂度和核矩阵维数限制问题, 文献 [33] 提出了归一化复合核仿射投影算法 (Normalized composite kernel affine projection algorithm, NC-KAPA). 文献 [34] 提出并行滑动窗口核仿射投影算法 (Parallel sliding window kernel affine projection algorithm, PSW-KAPA) 和级联滑动窗口核仿射投影算法 (Cascade sliding window kernel affine projection algorithm, CSW-KAPA). 文献 [35] 首次将多核引入 APA 中, 提出混合核仿射投影算法 (Kernel hybrid affine projection algorithm, KHAPA), 它将最大相关熵准则与均方误差策略结合起来改进 KAPA. 文献 [36] 提出了量化核仿射投影算法 (Quantization kernel affine projection algorithm, QKAPA), 它将矢量量化操作应用于 KAPA 来提高预测精度.

KAPA 是 Liu 等在 2008 年提出的, 它是由核技巧与 APA (Affine projection algorithm)<sup>[61]</sup> 结合产生的非线性扩展方法<sup>[21, 62]</sup>. KAPA 使用随机梯度法解决最小二乘问题, 该非线性扩展已经在时间序列在线预测领域得到广泛应用<sup>[21]</sup>. 以下将详细说明



基于 KAPA 及其改进方法的时间序列在线预测机理.

### 3.1 核仿射投影算法

KAPA 通过类似 KLMS 的方法将 APA 扩展到 RKHS 中, 除了样本数量以外, 还存在另外两个自由度: 1) 代价函数的正则化; 2) Newton 更新方法. 文献 [21] 通过输入相关矩阵的特征值扩展来解决梯度下降法收敛速度慢的问题.

根据不同的参数设置得到不同 KAPA, 从而组成 KAPA 族.

1) 根据梯度下降法得到 KAPA-1 的权重矩阵  $\mathbf{w}_1(i)$ .

$$\mathbf{w}_1(i) = \mathbf{w}(i-1) + \eta \Phi(i) [\mathbf{d}(i) - \Phi^T(i) \mathbf{w}(i-1)] \quad (38)$$

其中,  $\eta$  表示学习率,  $d(i)$  表示期望输出.

2) 根据 Newton 法得到 KAPA-2 的权重矩阵  $\mathbf{w}_2(i)$ .

$$\mathbf{w}_2(i) = \mathbf{w}(i-1) + \eta \Phi(i) [\Phi^T(i) \Phi(i) + \varepsilon \mathbf{I}]^{-1} \times [\mathbf{d}(i) - \Phi^T(i) \mathbf{w}(i-1)] \quad (39)$$

其中,  $\Phi(i) = [\varphi(i-K+1), \dots, \varphi(i)]$ ,  $\varepsilon$  是正则化参数.

3) 再给定正则化代价函数, 根据随机梯度下降法可得到 KAPA-3 的权重矩阵  $\mathbf{w}_3(i)$ .

$$\mathbf{w}_3(i) = (1 - \lambda\eta) \mathbf{w}(i-1) + \eta \Phi(i) [\mathbf{d}(i) - \Phi^T(i) \mathbf{w}(i-1)] \quad (40)$$

4) 相应地, 再根据 Newton 法得到 KAPA-4 的权重矩阵  $\mathbf{w}_4(i)$ .

$$\mathbf{w}_4(i) = (1 - \eta) \mathbf{w}(i-1) + \eta \Phi(i) [\Phi^T(i) \Phi(i) + \lambda \mathbf{I}]^{-1} \mathbf{d}(i) \quad (41)$$

其中,  $\lambda$  也是正则化参数, 当传统风险最小化病态时, 通常是约束解的范数  $\min_w E[\mathbf{d} - \mathbf{w}^T \boldsymbol{\varphi}(i)]^2 + \lambda \|\mathbf{w}\|^2$ , 得到式 (40). 与 KAPA-1 不同, KAPA-3 添加了尺度因子  $(1 - \lambda\eta)$  用于对先前数据强加“遗忘”机制, 并使该部分数据的影响呈指数下降趋势.

上述的 4 种变换中, 前三种都需要误差信息来更新网络, 计算量较大. 而 KAPA-4 不需要, 因为它只需对一个阶的矩阵求逆. 总之, KAPA 算法的性能已经在时间序列、信道均衡方面得到体现<sup>[21, 62]</sup>.

### 3.2 归一化复合核仿射投影算法

近年来, KLMS 和 KRLS 及其改进模型的研究都集中在单核形式. Ogunfunmi 等在 2011 年首次

将复合核应用于 APA 模型, 得到归一化复合核仿射投影算法<sup>[33]</sup>. 该算法适用于复杂数据的实时非线性建模.

在 NC-KAPA 中, 使用复杂核函数  $\kappa(\mathbf{z}, \mathbf{w})$ .

$$\kappa(\mathbf{z}, \mathbf{w}) = \exp\left(-\sigma \sum_{i=1}^N (\mathbf{z}_i - \mathbf{w}_i^*)^2\right) \quad (42)$$

其中,  $\mathbf{z}, \mathbf{w} \in C^N$ ,  $\sigma$  是影响转换映射的核参数.

定义  $\mathbf{X}_{ap}(i) = [\Phi(\mathbf{z}(i-K+1)) \cdots \Phi(\mathbf{z}(i))]$ , 得到权重矩阵  $\mathbf{w}(i)$ .

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \mu \mathbf{X}_{ap}(i) \times [\mathbf{X}_{ap}^T(i) \mathbf{X}_{ap}(i) + \varepsilon \mathbf{I}]^{-1} \mathbf{e}_{ap}^*(i)$$

其中,  $\boldsymbol{\lambda}_{ap}(i) = (\mathbf{X}_{ap}^T(i) \mathbf{X}_{ap}(i))^{-1} \mathbf{e}_{ap}^*(i)$ ,  $\varepsilon$  正平滑因子,  $\mathbf{e}_{ap} = (\mathbf{X}_{ap}^T(i) \mathbf{X}_{ap}(i) \boldsymbol{\lambda}_{ap}(i))^*$ ,  $\mu$  的选择需考虑稳定性、收敛速度和失调.

NC-KAPA 算法在理论上推导并扩展了复杂 RKHS 的 Wirtinger 演算, 最终还得到 APA 的梯度 MSE 代价函数. 该算法更适用于对收敛速率要求较高的高维复杂的混沌时间序列的在线预测<sup>[33]</sup>.

### 3.3 并行、级联滑动窗口核仿射投影算法

由上节 Ogunfunmi 等提出的 NC-KAPA<sup>[33]</sup> 可知, KAF 的发展已经由单核研究逐渐过渡到多核研究<sup>[63]</sup>. 2012 年 Gil-Cacho 等提出并行和级联滑动窗口核仿射投影算法<sup>[34]</sup>. 该类算法由线性核和高斯核的加权和组成. 机理是分离线性和非线性子问题.

PSW-KAPA 的并行配置在 KAPA-3 中可直接应用得到核矩阵  $\mathbf{k}_{wsk}$ .

$$\mathbf{k}_{wsk}(i, j) = \alpha \mathbf{k}_L(i, j) + \beta \mathbf{k}_G(i, j) \quad (43)$$

其中,  $\mathbf{k}(i, j) = \exp(-\kappa \|\mathbf{X}(i) - \mathbf{X}(j)\|^2)$ ,  $\kappa$  是高斯核参数,  $\alpha < 1$ ,  $\beta = 1 - \alpha$ .

CSW-KAPA 中的级联结构由以下 2 个步骤组成: 1) 独立执行标准线性归一化最小均方算法 (Normalized least mean square, NLMS), 存储滤波器输出. 2) 将存储的 NLMS 输出用作 SW-KAPA 的输入. 得到基于 PSW-KAPA 和 CSW-KAPA 的预测输出  $\hat{\mathbf{y}}(i)$ .

$$\hat{\mathbf{y}}(i) = \hat{\mathbf{h}}^T(i) \mathbf{X}(i) \quad (44)$$

其中,  $\hat{\mathbf{h}}(i)$  是 NLMS 的权重向量.

与 KAPA 相比, PSW-KAPA 和 CSW-KAPA 降低了计算复杂度. 与线性 NLMS 相比, 它提高了预测性能. PSW-KAPA 和 CSW-KAPA 中核函数的分离、加权和过程都在滑动窗口方法中执行. 当

对其施加不同的“遗忘”机制时, 它又转化为更灵活的正则化形式, 进一步提高时间序列的预测性能<sup>[34]</sup>.

### 3.4 核混合仿射投影算法

2014年 Yang 等提出了核混合仿射投影算法<sup>[35]</sup>. KHAPA 将混合准则 (Hybrid criteria, HC) 应用于 APA 中, 并将得到的 HAPA 映射到 RKHS. HC 是均方误差和最大相关熵准则 (Maximum correntropy criterion, MCC) 的混合.

HC 是从最小裁剪平方 (Least trimmed squares, LTS) 估计的角度避免异常值的过度影响. 在 LTS 中, 数据通过排序的方式被分为两类: 正常数据集和异常值数据集, 且异常值数据被完全丢弃. HC 不是单纯地丢弃这些数据, 而是对大数据采用鲁棒 MCC 准则. KHAPA 就是利用上述准则来进一步提高预测性能<sup>[35]</sup>.

HC 准则不仅避免了异常值的过度影响, 还充分利用了异常值, 从而获得更好的性能. 定义 HC 如式 (45), 并将 HC 应用于 KAPA 模型, 用  $f_i$  表示它的学习映射.

$$W_{HC} = \arg \min J_{HC} = \arg \min \left( \sum_{i=1}^M e_{j(i)}^2 - \sum_{i=M+1}^N e \exp \left( -\frac{e_{j(i)}^2}{2\sigma^2} \right) \right) \quad (45)$$

$$f_i = f_{i-1} + \eta K_i \xi_i \quad (46)$$

其中,  $K_i = [\kappa(x_{i-L+j(1)}, \cdot), \dots, \kappa(x_{i-L+j(L)}, \cdot)]$ ,  $\xi_i = \left[ e_{i-L+j(i)}, e_{i-L+j(i+1)} \exp \left( -\frac{e_{i-L+j(i+1)}^2}{2\sigma^2} \right) \right]$  ( $i = 1, \dots, M$ ),  $\eta$  是学习率.

HC 充分利用了数据, 避免了异常值数据集的不恰当影响. 所以 KHAPA 在时间序列预测过程中提高了数据的利用效率, 同时也得到了更准确的预测结果<sup>[35]</sup>.

### 3.5 量化核仿射投影算法

2015年 Ren 等提出量化核仿射投影算法<sup>[36]</sup>. QKAPA 和 QKLMS、QKRLS 算法相似, QKAPA 是在 KAPA 模型上加入 VQ 用以约束网络, 并进一步减小计算负担和内存占用.

在 QKAPA 中, 计算  $dis(x_i, C(n, i-1))$  如式 (47), 如果它小于  $\varepsilon_X$ , 此时  $x_i$  是“冗余”的, 然后字典保持不变, 权重系数根据式 (48) 更新. 否则,  $x_i$  不是“冗余”的, 加入字典, 此时权重系数根据式 (49) 更新.

$$dis(x_i, C(n, i-1)) = \min (\|x_i - C_j(n, i-1)\|) \quad (47)$$

其中,  $1 \leq j \leq size(C(n, i-1))$ .

$$\alpha_{j^*}(n, i) = \alpha_{j^*}(n, i-1) + \eta e(n, i) \quad (48)$$

其中,  $j^* = \arg \min_{1 \leq j \leq size(C(n, i-1))} \|x_i - C_j(n, i-1)\|$ ,  $e(n, i) = d_n - f(n, i)$ .

$$\alpha(n, i) = \{\alpha(n, i-1), \eta e(n, i)\} \quad (49)$$

根据上述更新过程, 得到 QKAPA 模型的预测输出  $f(n, i)$ .

$$f(n, i) = \sum_{j=1}^{size(C(n, i-1))} \alpha_j(n, i-1) \kappa(x_i, C_j(n, i-1)) \quad (50)$$

其中,  $\kappa(\cdot, \cdot)$  表示核函数.

由于时间序列都是复杂的, 且包含未知噪声. QKAPA 模型在预测中不仅减小了梯度噪声, 而且约束了网络规模, 提高了“冗余”数据的利用效率, 进一步减小了计算负担和存储占用, 明显加快了运算速度, 提高了预测性能<sup>[36]</sup>.

### 3.6 基于 KAPA 在线预测的总结分析

通过以上对 KAPA 及其改进模型的说明, 就基于 KAPA 的改进模型来说, PSW-KAPA、CSW-KAPA<sup>[34]</sup> 和 NC-KAPA<sup>[33]</sup> 以投影次梯度的方式简化了预测模型. QKAPA<sup>[36]</sup> 和 KAPHA<sup>[35]</sup> 分别通过 VQ 和 HC 稀疏方式降低了计算复杂度, 并提升了预测精度. KAPA 及其改进模型都在 RKHS 中应用随机梯度法解决最小二乘问题. 同 KLMS 或 KRLS 相比, KAPA 能提供灵活的在线非线性滤波器计算方法. 根据实际预测需求, 还可在应用性能与计算复杂度之间选择平衡点, 以达到最优结果.

KAPA 的通用形式是基于自适应投影次梯度法被推导. 通过投影映射测度完成分类, 再通过正交投影实现稀疏化, 还可以通过斜投影解决在线存储需求和自适应跟踪能力<sup>[32]</sup>. 此外, KAPA 也为径向基函数 (如 NN)、正则化提供了理论认识基础.

在性能方面, KAPA 介于 KLMS 和 KRLS 之间, KAPA 的性能可以通过窗口长度  $K$  调整, 同时窗口长度也控制计算复杂度. 而且大部分时间序列都呈现非线性、非平稳、复杂的特点, 通常的离线算法不能满足在线实时性要求. 所以基于 KAF 的时间序列在线预测方法在降低计算复杂度、提高预测精度和效率方面都展现出独特的优势<sup>[32-36]</sup>.

总的来说, 基于 KAF 的 3 类在线预测方法已经在多元混沌时间序列中得到较好的预测性能<sup>[12-13]</sup>.

其中,模型中部分参数的设置也起着关键的作用,比如步进参数  $\mu \in (0, 1)$ , 学习率  $\eta \in (0, 1)$  等, 它们的选取对模型的预测效果有较大的影响. 一般情况下, 步进参数和学习率都会选取  $(0.5, 1)$  之间的值, 这样可以得到较好的性能结果<sup>[19-36]</sup>. 同样的, 正则化参数  $\lambda$ , 调整参数  $\beta$  等参数的选取也会影响实验结果<sup>[29]</sup>, 比如正则化参数  $\lambda$  可以根据高斯过程回归的标准超参数估计来进行选取<sup>[17, 29]</sup>, 此时得到的  $\lambda$  更适合模型的后续预测. 所以, 参数的恰当选择可以有效地反映模型的某些性能<sup>[41-42, 61, 64]</sup>, 这里不再赘述.

#### 4 基于 KAF 在线预测的应用

目前, 动态神经网络 (Dynamic neural network, DNN)、支持向量回归 (Support vector regression, SVR)、KAF 等方法已应用于复杂时间序列预测<sup>[8-10]</sup>. DNN 与 SVR 都是离线操作, 前者在更新过程中涉及大量计算<sup>[65]</sup>, 后者适用于小样本预测与跟踪<sup>[13]</sup>. 而 KAF 不仅具有较低的计算复杂度和简单的递归更新形式, 且满足在线预测的要求, 因此受到许多学者的青睐<sup>[11-13]</sup>. Frieb 和 Harrison 根据“核化”思想提出 ADALINE, 该方法基于所有的训练数据并采用确定性梯度法得到<sup>[65]</sup>. Kivinen 等提出 NORMA, 该方法直接对正则化代价函数进行微分获取随机梯度<sup>[17]</sup>. Engel 等使用矩阵求逆引理研究了 KRLS<sup>[20]</sup>. Liu 等研究了具有正则化特性的 KLMS<sup>[19]</sup>. Liu 和 Principe, Ogunfunmi 和 Paul, Ren 和 Yu 等从不同角度研究了 KAPA<sup>[21, 33-36, 61]</sup>.

根据 KLMS、KRLS 和 KAPA 三类模型在时间序列中的预测效率、预测精度、收敛速度以及相应的特点, 三类典型 KAF 方法在时间序列在线预测中的性能对比结果如表 1 所示. 由表 1 可知, 与 KLMS 和 KAPA 相比, KRLS 的预测效率较低, 但在收敛速度和预测精度方面更具优势. 所以, 我们可以得到以下结论: 1) 在实际应用中, 如果将预测效率放在首位而不追求收敛速度, 那么 KLMS 模型可以得到较好的结果; 2) 如果以提高目标预测精度为目的, 同时要求较快的收敛速度, KRLS 模型无

表 1 不同 KAF 方法的时间序列在线预测特性对比结果  
Table 1 Comparison of online prediction characteristics of time series of different KAF methods

算法类型	预测效率	预测精度	收敛速度	特点
KLMS <sup>[19]</sup>	较高	较低	较慢	泛化能力和正则化特性
KRLS <sup>[20]</sup>	较低	较高	较快	白化处理, 收敛速度较快
KAPA <sup>[21]</sup>	中	中	中	考虑多个样本, 降低梯度噪声

疑成为最佳选择; 3) 如果预测目标中包含多种噪声, KAPA 模型更具优势.

目前, 大多数时间序列都表现出非线性、非平稳、复杂难预测的特点, 所以对其进行合理的预测具有十分重要的意义. 例如, 从气象数据的历史信息预测未来天气变化<sup>[1-2]</sup>. 从水文数据的历史信息预测水资源的变化趋势, 并作出合理的控制及调配<sup>[3-4]</sup>. 根据医学上的不同症状序列进行提前预测并做好预防, 降低发病几率等<sup>[5-6]</sup>. 复杂系统中包含着许多相关的有用信息, 而且随着时间的推移, 其中包含的信息量也在不断增加. 随着信息技术的发展, 时间序列的数据量和应用也随之快速增长.

如何实现复杂时间序列的实时预测, 并跟踪其将要发生的变化, 是诸多研究领域非常关注的问题. 文献 [55-56] 在 LSSVM 的基础上添加过滤窗、迭代误差补偿和组合核函数的形式对复杂时间序列进行预测. 文献 [66] 采用卡尔曼滤波在高频金融时间序列进行预测. 文献 [57] 采用自适应遗传算法和基于大脑情感学习模型预测时间序列. 对于含噪声和数据缺失的时间序列来说, 文献 [58-60] 给出了很好地解决方案. 但是, 传统的数据预测方法集中于建模检验预测、自适应预测等, 难以实时预测数据将要发生的变化. 现有的真实数据本身较复杂, 在实际应用中会出现计算时间长、效率低、存储空间占用量大、预测精度低、跟踪能力差等问题.

因此, 通常需要对复杂时间序列进行稀疏选择来降低核矩阵维度, 然后再进行更为有效地预测与跟踪. 综上所述, 用 KAF 模型对复杂难预测系统进行在线预测、跟踪具有十分重要的意义<sup>[52-54]</sup>.

KAF 通过分析现有时间序列的有效信息, 能较好地反映当前时间序列的潜在特征. 然后再根据所得特征跟踪时变信息, 并进行相关分析总结, 以提醒相应部门对于将要到来的各种变化做出科学的决策<sup>[67-71]</sup>.

KAF 是人工智能领域中常用于时滞时变预测的技术, 它采用滤波的形式解决预测中出现的问题<sup>[37-38]</sup>, 并能够有效地增强预测模型的泛化性能与自适应跟踪能力. 总的来说, KAF 在处理非线性系统时, 随着新样本的不断到来, 模型不断地迭代更新. 就已提出的 KAF 模型来说, 不同的在线预测模型对应不同的计算复杂度. 原始核矩阵的计算的复杂度为  $O(L^3)$ , 当采用 ALD 准则时, 其计算复杂度为  $O(L^2)$ . 对于不同的稀疏方法得到的 KAF 模型, 其复杂度也不同. 表 2 给出了三类 KAF 模型在每次迭代过程涉及的计算复杂度 (其中,  $L$  表示样本长度,  $K$  表示字典大小). 从表 2 可以看出: 首先, 表述的 8 种在线稀疏方法都在不同程度上降低了核



表 2 每次迭代过程涉及的计算复杂度比较  
Table 2 Comparison of computational complexity involved in each iteration

核自适应滤波器类型	在线稀疏类型	计算复杂度	
KLMS <sup>[19]</sup>	VQ	更新 $\alpha(i)$	$O(L^2)$
		在线VQ	$O(L)$
	SF	更新 $\omega(i)$	$O(L)$
		更新 $e(i)$	$O(L)$
KRLS <sup>[20]</sup>	VQ	更新 $\alpha(i)$	$O(L)$
		在线VQ	$O(L)$
	ALD	更新 $\alpha(i)$	$O(L^2)$ ( $O(L^2)$ 假如字典改变)
		更新ALD	$O(L^2)$
更新 $P(i)$		$O(L^2)$	
KRLS <sup>[20]</sup>	SW	更新 $\alpha(i)$	$O(K^2)$ ( $O(K)$ 假如字典改变)
		更新 $P(i)$	$O(K^2)$
		更新 $D(i)$	$O(K^2)$
	FB	更新 $\alpha(i)$	$O(K^2)$ ( $O(K)$ 假如字典改变)
		更新 $P(i)$	$O(K^2)$
		更新 $\hat{K}_n(i)$	$O(K^2)$
MF	更新 $\alpha(i)$	$O(L)$	
	更新 $e(i)$	$O(L)$	
	更新 $D(i)$	$O(L^2)$	
KAPA <sup>[21]</sup>	CC	更新 $\alpha(i)$	$O(K^2)$ ( $O(K)$ 假如字典改变)
		更新 $e(i)$	$O(K^2)$
		更新 $D(i)$	$O(K^2)$
	VQ	更新 $\alpha(i)$	$O(L)$
		在线VQ	$O(L)$
	HC	更新 $P(i)$	$O(L^2)$
更新 $e(i)$		$O(L)$	
更新 $\zeta(i)$		$O(L)$	

矩阵的维度, 即降低了模型的计算复杂度. 其次, 由于 SW、FB 和 CC 准则在预测过程中都需要固定字典的大小 ( $K$  远小于  $L$ ), 所以它们在降低核矩阵维度、模型内存占用和计算复杂度上都表现出较大优势. 最后, SF、MF 和 HC 不仅在降低模型计算复杂度上出较大贡献, 而且在对有效利用历史信息上也具优势 (主要是反馈信息的有效利用).

由于 KAF 具有自适应更新能力, 因此常用于复杂时间序列预测. 基于 KLMS 的随机行为分析, 如文献 [72] 在 LMS 的基础上采用核在线技术提出顺序优化方法, 并采用随机梯度法顺序更新权重, 分析网络流量序列的行为特征, 从而实现网络流量

序列的在线预测. 基于 KRLS 的在线异常检测, 如文献 [55] 通过迭代误差补偿弥补 SVM 的误差特性, 从而减少冗余信息的残差干扰, 最终达到异常检测的效果. 基于 KAPA 的在线时间序列分类等<sup>[56]</sup>, 如文献 [36] 在 KAPA 的基础上添加可重构的 FPGA 加速器, 该过程减小了梯度噪声, 在处理数据路径和处理单元数量方面具有突出的优势. 总之, KAF 通用的非线性建模能力对复杂时间序列预测更具优势.

所以, 根据 KAF 的特殊思想, 许多学者已经将 KAF 应用于时间在线序列预测<sup>[72-77]</sup>. 由于环境变化不断加剧, 时间序列包含的噪声也不断复杂化, 传统的线性模型并不适应非线性、复杂系统的剧烈变化, 所以 KAF 在线模型在数据挖掘处理噪声方面也展现出较好的结果<sup>[67]</sup>. 文献 [72] 是以在线形式将 KAF 应用于含噪声的时间序列, 过滤冗余噪声信息, 从而实现平滑太阳黑子数量时间序列的预测. KAF 还通过最小平均绝对第三损失函数来滤除噪声, 从而得到更好的预测性能. 文献 [74] 在时间序列上执行在线学习, 使用遗憾最小化技术解决在噪声条件下的最小假设问题, 为执行有效地在线学习过程剔除干扰项. 文献 [75] 结合奇异鲁棒策略和 KRLS 来解决非线性系统识别问题. 文献 [76] 采用序贯在线学习方法预测水文时间序列 (包括气温、水位、风速、总发电量等), 由于水文数据包含噪声, 所以该方法在滤除噪声方面具有更好的特性. 文献 [77] 通过“核技巧”以循环在线训练的方式来执行学习过程, 该过程在循环损失函数中涉及自动更新权重的正则化项, 在提升泛化能力的同时执行在线学习. 文献 [78] 将 KAF 以递归的形式应用于脑电和自适应天线阵列过程, 该过程排除了脑电和天线阵列的噪声信号. 文献 [79] 将 VQ 和 MC 准则应用于 KAF 模型, 它在处理异常值和脉冲噪声方面更具特色. 文献 [80, 82] 基于贝叶斯框架建立“遗忘”机制, 它能很好地处理带噪的脑电信号, 衰落和突变信道的跟踪.

此外, 文献 [83] 推导和分析了用于二元和多元分类、回归和预测的算法, 它们的更新步骤都是基于简单约束优化问题的分析解决方案. 文献 [84-85] 提出采用矩阵分解或负矩阵分解的降维方式来实现在线稀疏学习, 它已经广泛应用于图像处理 and 模式识别. 文献 [86] 提出子梯度方法动态地结合早期迭代中观察到的数据几何知识, 以执行更具信息性的基于梯度的在线学习过程. 文献 [87] 介绍了在线凸优化, 在线分类, 在线批量转换以及在线预测过程的约束限制问题. 文献 [88-89] 提出预算方法实现

对核的约束,从而推导出以预算随机梯度下降和近似逼近的方式来适应大规模 SVM 预测.文献 [90] 提出一个用于大规模在线分类任务的高效且可扩展的在线学习算法库.文献 [91] 讨论了在线预测和批处理在使用有限计算资源的过程中渐近地提供最佳的泛化性能的有效方法.文献 [92] 介绍了基于 LS 回归的收敛率问题,该方法通过对初始条件和 Hessian 矩阵的假设进行分析,并最终实现平均加速正则梯度下降.文献 [93] 研究了与凸损失函数相关的 RKHS 中的在线学习算法.

就预期的超额泛化误差而言,它们可以作为基于内核的批量学习算法快速收敛.从而实现对学习序列规范的预期值和在线学习算法的精确误差分解的敏锐估计.

文献 [94] 提出 KNLMS 方法,它使用自适应步长加速学习,并在收敛速度和稳态之间提供合适的折中.文献 [95] 提出具有非线性主路径的有源噪声控制系统的 KxLMS 方法,主要用于改进在参考噪声与多个窄带信号和附加高斯白噪声混合时的预测性能.文献 [96] 研究了在 RKHS 中没有正则化条件时在线梯度下降法的收敛性,以收敛期望中的超泛化误差.文献 [97] 提出了改进的在线随机梯度下降法,由于特征稀疏性的异质性,该方法用于解决收敛速度慢和方差大的问题.文献 [98] 建立了高斯核参数的自适应更新方法并应用于 KAF,该方法的更新规则与 L1 正则化一同作用来避免预测中的过拟合和维数高问题.文献 [99] 提出尺度自适应在线学习方法,通过评估核函数和系数向量之间的内积实现 KAF.文献 [100] 提出了基于 L2 空间迭代投影的新型在线非线性函数学习、估计方法.文献 [101] 提出了一种由节点网络进行非线性函数分布式学习的自适应方案,基于多个 RKHS 的笛卡尔积度量,证明了方法的全面收敛性.

综上所述, KAF 在时间序列预测领域已经得到了广泛的应用<sup>[96-101]</sup>.由 KAF 的改进模型可以看出,其计算复杂度被明显降低,内存空间的占用也明显减小,相应的预测效率和预测精度也显著提高.同时,对于含噪声和缺失值的时间序列还具有特殊的滤除噪声和补差性能<sup>[58-59]</sup>.

## 5 结论与展望

KAF 的研究在近年来受到了越来越广泛的关注,诸多学者对 KAF 学习机理和应用进行了全面的研究. KLMS、KRLS 和 KAPA 奠定了 LMS、RLS 和 APA 在核空间应用的基础.近年来对 KAF 的研究已经取得了一定的进展,并广泛应用于时间序列在线预测,如气象领域的温度、湿度、风速、风向、日

照时数、气压等实时性预测,水文领域的水位、水温、流量、凝结度等预测,金融领域的股票联动性和股指期货套期保值预测,医疗领域的心电、脑电数据异常检测,交通领域的车流量、密度预测等.本文首先对 KAF 的学习方法和发展现状进行了详细综述,探讨了 KLMS、KRLS 和 KAPA 的基本原理,及其改进模型在提高时间序列在线预测精度和预测效率,跟踪时变特性和如何利用有效数据信息等方面的机理,同时将 KLMS、KRLS 和 KAPA 作为 KAF 在线预测的研究基础,分析在线稀疏策略在具有时变特性的非线性系统上的预测精度和预测效率的提高,接着结合其在线稀疏过程对计算复杂度进行分析,最后总结 KAF 模型在时间序列预测、数据挖掘噪声处理、性能补差、性能估计等方面的实际应用.然而随着研究的不断深入,仍存在着一些值得研究和关注的问题:

1) KAF 在线预测方法中包含 KLMS、KRLS、KAPA 等多种模型,对于不同的模型所对应的预测领域存在较大差异.例如, KLMS 模型多用于非线性系统建模,即在非线性时间序列预测、信道均衡领域获得较好的性能. KRLS 模型多用于慢衰落和状态变化很小的非线性时间序列跟踪领域. KAPA 模型多用于时间序列预测、非线性信道均衡和非线性滤除噪声等领域.所以,不同的模型可能对应不同领域及不同的问题,这是当前研究的一个重点.同时如何对预测模型进行正则化,提高模型的泛化能力也是一个研究重点.

2) KAF 在线预测方法对复杂时间序列预测具有更好的跟踪性能,通过建模过程中的动态更新机制,使在线预测过程能够实时反映核矩阵权重更新和系数矩阵更新,最终得到的预测模型具有自适应实时更新能力.然而 KAF 在预测样本时,随着新样本的不断加入,核矩阵维数不断增加,其计算复杂度也随之增大(线性关系),同时内存消耗(占用量)也增大,相应的降低了计算效率.对于上述问题,有学者已经提出了 ALD、NC、SC、HC 等多种在线稀疏方法来解决此类问题.因此,在预测模型中如何对输入样本进行在线稀疏已经成为继续深入研究 KAF 的方向之一.

3) 在时间序列在线预测过程中,随着 ALD、NC、SC、HC 等多种在线稀疏方法加入预测模型,一些“冗余”的数据被剔除,在这个过程中也会伴随数据利用效率的降低,即剔除的数据可能包含有效的历史信息,如何判断稀疏后的样本序列是否包含有效信息也是需要研究的重点.所以,在加入在线稀疏方法时如何保持稀疏后数据的“有效性”也是

KAF 在线预测未来研究的重要方向之一。

4) 目前, 随着单核自适应滤波器研究的深入, 多核自适应滤波器也不断地发展. 大多数多核算法是将两个相同类型的核函数应用于预测模型, 或者将两种不同的核函数的加权和应用于预测模型, 此类方法已经取得了较好的预测结果. 由于核函数表示的是一种将输入 (数据) 空间非线性映射到高维 (隐藏) 特征空间的过程, 该过程是隐式的、未知的. 所以在选用单核函数或多核函数的过程中, 不同的选择得到的预测结果大相径庭, 其预测性能也存在较大差异. 所以, 如何在复杂的非线性预测系统中选择合适的核函数及其参数, 也是 KAF 在线预测的研究热点之一。

## References

- Gan M, Chen C L P, Li H X, Chen L. Gradient radial basis function based varying-coefficient autoregressive model for nonlinear and nonstationary time series. *IEEE Signal Processing Letters*, 2015, **22**(7): 809–812
- Schamberg G, Ba D, Coleman T P. A modularized efficient framework for non-Markov time series estimation. *IEEE Transactions on Signal Processing*, 2018, **66**(12): 3140–3154
- Gonzalez J P, San Roque A M S M, Perez E A. Forecasting functional time series with a new Hilbertian ARMAX model: Application to electricity price forecasting. *IEEE Transactions on Power Systems*, 2018, **33**(1): 545–556
- Rahman M M, Islam M, Murase K, Yao X. Layered ensemble architecture for time series forecasting. *IEEE Transactions on Cybernetics*, 2016, **46**(1): 270–283
- Zhu J, Chen N, Peng W W. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Transactions on Industrial Electronics*, 2019, **66**(4): 3208–3216
- Chang F J, Chen P A, Lu Y R, Huang E, Chang K Y. Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *Journal of Hydrology*, 2014, **517**: 836–846
- Vapnik V N. *The Nature of Statistical Learning Theory*. New York: Springer, 2000.
- Girosi F, Jones M, Poggio T. Regularization theory and neural networks architectures. *Neural Computation*, 1995, **7**(2): 219–269
- Platt J. A resource-allocating network for function interpolation. *Neural Computation*, 1991, **3**(2): 213–225
- Huang G B, Saratchandran P, Sundararajan N. A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation. *IEEE Transactions on Neural Networks*, 2005, **16**(1): 57–67
- Suykens J A K, Vandewalle J. Least squares support vector machine classifiers. *Neural Processing Letters*, 1999, **9**(3): 293–300
- Scholkopf B, Smola A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: MIT Press, 2001.
- Liu W F, Principe J C, Haykin S. *Kernel Adaptive Filtering: A Comprehensive Introduction*. Hoboken: John Wiley & Sons, 2010.
- Scholkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1998, **10**(5): 1299–1319
- Feng D Z, Bao Z, Jiao L C. Total least mean squares algorithm. *IEEE Transactions on Signal Processing*, 1998, **46**(8): 2122–2130
- Henon M. A two-dimensional mapping with a strange attractor. *The Theory of Chaotic Attractors*. New York, NY: Springer, 1976. 94–102
- Kivinen J, Smola A J, Williamson R C. Online learning with kernels. *IEEE Transactions on Signal Processing*, 2004, **52**(8): 2165–2176
- Richard C, Bermudez J C M, Honeine P. Online prediction of time series data with kernels. *IEEE Transactions on Signal Processing*, 2009, **57**(3): 1058–1067
- Liu W F, Pokharel P P, Principe J C. The kernel least-mean-square algorithm. *IEEE Transactions on Signal Processing*, 2008, **56**(2): 543–554
- Engel Y, Mannor S, Meir R. The kernel recursive least-squares algorithm. *IEEE Transactions on Signal Processing*, 2004, **52**(8): 2275–2285
- Liu W F, Principe J C. Kernel affine projection algorithms. *EURASIP Journal on Advances in Signal Processing*, 2008, **2008**(1): 784292
- Chen B D, Zhao S L, Zhu P P, Principe J C. Quantized kernel least mean square algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(1): 22–32
- Zhao J, Liao X F, Wang S Y, Tse C K. Kernel least mean square with single feedback. *IEEE Signal Processing Letters*, 2015, **22**(7): 953–957
- Zheng Y F, Wang S Y, Feng J C, Tse C K. A modified quantized kernel least mean square algorithm for prediction of chaotic time series. *Digital Signal Processing*, 2016, **48**: 130–136
- Liu Y Q, Sun C, Jiang S D. A kernel least mean square algorithm based on randomized feature networks. *Applied Sciences*, 2018, **8**(3): 458
- Van Vaerenbergh S, Via J, Santamaria I. A sliding-window kernel RLS algorithm and its application to nonlinear channel identification. In: Proceedings of 2006 IEEE International Conference on Acoustics Speech and Signal Processing. Toulouse, France: IEEE, 2006. 789–792
- Liu W F, Park I, Wang Y W, Principe J C. Extended kernel recursive least squares algorithm. *IEEE Transactions on Signal Processing*, 2009, **57**(10): 3801–3814
- Van Vaerenbergh S, Santamaria I, Liu W F, Principe J C. Fixed-budget kernel recursive least-squares. In: Proceedings of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, TX, USA: IEEE, 2010. 1882–1885
- Van Vaerenbergh S, Lazaro-Gredilla M, Santamaria I. Kernel recursive least-squares tracker for time-varying regression. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(8): 1313–1326
- Chen B D, Zhao S L, Zhu P P, Principe J C. Quantized kernel recursive least squares algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, **24**(9): 1484–1491
- Wang S Y, Wang W, Duan S K, Wang L D. Kernel recursive least squares with multiple feedback and its convergence analysis. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2017, **66**(10): 1237–1241



- 32 Zhang S H, Han M, Wang J, Wang D. Multivariate time series online prediction based on adaptive normalized sparse kernel recursive least squares algorithm. In: Proceedings of the 7th International Conference on Information Science and Technology. Da Nang, Vietnam: IEEE, 2017. 38–44
- 33 Ogunfunmi T, Paul T. On the complex kernel-based adaptive filter. In: Proceedings of 2011 IEEE International Symposium of Circuits and Systems. Rio de Janeiro, Brazil: IEEE, 2011. 1263–1266
- 34 Gil-Cacho J M, Van Waterschoot T, Moonen M, Jensen S H. Nonlinear acoustic echo cancellation based on a parallel-cascade kernel affine projection algorithm. In: Proceedings of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. Kyoto, Japan: IEEE, 2012. 33–36
- 35 Yang X H, Qu H, Zhao J, Chen B D. Hybrid affine projection algorithm. In: Proceedings of the 13th International Conference on Control Automation Robotics & Vision. Singapore, Singapore: IEEE, 2014. 964–968
- 36 Ren X W, Yu Q H, Chen B D, Zheng N N, Ren P J. A reconfigurable parallel FPGA accelerator for the kernel affine projection algorithm. In: Proceedings of 2015 IEEE International Conference on Digital Signal Processing. Singapore, Singapore: IEEE, 2015. 906–910
- 37 Singh A, Ahuja N, Moulin P. Online learning with kernels: Overcoming the growing sum problem. In: Proceedings of 2012 IEEE International Workshop on Machine Learning for Signal Processing. Santander, Spain: IEEE, 2012. 1–6
- 38 Haykin S. *Adaptive Filter Theory*. Dorling Kindersley, India: Pearson Education, 2008.
- 39 Zhao S L, Chen B D, Zhu P P, Principe J C. Fixed budget quantized kernel least-mean-square algorithm. *Signal Processing*, 2013, **93**(9): 2759–2770
- 40 Rzepka D. Fixed-budget kernel least mean squares. In: Proceedings of the 17th IEEE International Conference on Emerging Technologies & Factory Automation. Krakow, Poland: IEEE, 2012. 1–4
- 41 Fan H J, Song Q. A linear recurrent kernel online learning algorithm with sparse updates. *Neural Networks*, 2014, **50**: 142–153
- 42 Liu W f, Park I, Principe J C. An information theoretic approach of designing sparse kernel adaptive filters. *IEEE Transactions on Neural Networks*, 2009, **20**(12): 1950–1961
- 43 Ma W T, Duan J D, Man W S, Zhao H Q, Chen B D. Robust kernel adaptive filters based on mean p-power error for noisy chaotic time series prediction. *Engineering Applications of Artificial Intelligence*, 2017, **58**: 101–110
- 44 Chen B D, Xing L, Zhao H Q, Zheng N N, Principe J C. Generalized correntropy for robust adaptive filtering. *IEEE Transactions on Signal Processing*, 2016, **64**(13): 3376–3387
- 45 Chen B D, Wang J J, Zhao H Q, Zheng N N, Principe J C. Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Processing Letters*, 2015, **22**(10): 1723–1727
- 46 Chen B D, Zhao S L, Zhu P P, Principe J C. Mean square convergence analysis for kernel least mean square algorithm. *Signal Processing*, 2012, **92**(11): 2624–2632
- 47 Gao W, Chen J, Richard C, Huang J G. Online dictionary learning for kernel LMS. *IEEE Transactions on Signal Processing*, 2014, **62**(11): 2765–2777
- 48 Han M, Wang Y N. Nonlinear time series online prediction using reservoir Kalman filter. In: Proceedings of 2009 International Joint Conference on Neural Networks. Atlanta, USA: IEEE, 2009. 1090–1094
- 49 Zhou H W, Huang J Q, Lu F, Thiyaalingam J, Kirubarajan T. Echo state kernel recursive least squares algorithm for machine condition prediction. *Mechanical Systems and Signal Processing*, 2018, **111**: 68–86
- 50 Wang X Y, Han M. Multivariate time series prediction based on multiple kernel extreme learning machine. In: Proceedings of 2014 International Joint Conference on Neural Networks. Beijing, China: IEEE, 2014. 198–201
- 51 Xu M L, Zhang S H, Han M. Multivariate time series modeling and prediction based on reservoir independent components. In: Proceedings of the 6th International Conference on Intelligent Control and Information Processing (ICICIP). Wuhan, China: IEEE, 2015. 325–330
- 52 Han M, Zhang S H, Xu M L, Qiu T, Wang N. Multivariate chaotic time series online prediction based on improved kernel recursive least squares algorithm. *IEEE Transactions on Cybernetics*, 2019, **49**(4): 1160–1172
- 53 Han M, Ren W J, Xu M L, Qiu T. Nonuniform state space reconstruction for multivariate chaotic time series. *IEEE Transactions on Cybernetics*, 2019, **49**(5): 1885–1895
- 54 Han Min, Ren Wei-Jie, Xu Mei-Ling. An improved echo state network via  $L_1$ -norm regularization. *Acta Automatica Sinica*, 2014, **40**(11): 2428–2435  
(韩敏, 任伟杰, 许美玲. 一种基于  $L_1$  范数正则化的回声状态网络. 自动化学报, 2014, **40**(11): 2428–2435)
- 55 Tang Zhou-Jin, Ren Feng, Peng Tao, Wang Wen-Bo. A least square support vector machine prediction algorithm for chaotic time series based on the iterative error correction. *Acta Physica Sinica*, 2014, **63**(5): 050505  
(唐舟进, 任峰, 彭涛, 王文博. 基于迭代误差补偿的混沌时间序列最小二乘支持向量机预测算法. 物理学报, 2014, **63**(5): 050505)
- 56 Liu Chang, Lang Jin. Wind power prediction method using hybrid kernel LSSVM with batch feature. *Acta Automatica Sinica*, 2020, **46**(6): 1264–1273  
(刘畅, 郎劲. 基于混核 LSSVM 的批特征风功率预测方法. 自动化学报, 2020, **46**(6): 1264–1273)
- 57 Mei Ying, Tan Guan-Zheng, Liu Zhen-Tao, Wu He. Chaotic time series prediction based on brain emotional learning model and self-adaptive genetic algorithm. *Acta Physica Sinica*, 2018, **67**(8): 080502  
(梅英, 谭冠政, 刘振焱, 武鹤. 基于大脑情感学习模型和自适应遗传算法的混沌时间序列预测. 物理学报, 2018, **67**(8): 080502)
- 58 Yang Zhen-Ming, Yue Ji-Guang, Wang Xiao-Bao, Xiao Yun-Shi. Noisy time series prediction using independent component analysis. *Control and Decision*, 2013, **28**(4): 501–505  
(杨臻明, 岳继光, 王晓保, 萧蕴诗. 基于独立成分分析的含噪声时间序列预测. 控制与决策, 2013, **28**(4): 501–505)
- 59 Song Xiao-Xiang, Guo Yan, Li Ning, Yu Dong-Ping. Missing data prediction algorithm based on sparse Bayesian learning in coevolving time series. *Computer Science*, 2019, **46**(7): 217–223  
(宋晓祥, 郭艳, 李宁, 余东平. 基于稀疏贝叶斯学习的协同进化时间序列缺失数据预测算法. 计算机科学, 2019, **46**(7): 217–223)
- 60 Song Xiao-Xiang, Guo Yan, Li Ning, Wang Meng. Missing data prediction based on compressive sensing in time series. *Computer Science*, 2019, **46**(6): 35–40  
(宋晓祥, 郭艳, 李宁, 王萌. 基于压缩感知的时间序列缺失数据预测算法. 计算机科学, 2019, **46**(6): 35–40)
- 61 Paul T K, Ogunfunmi T. On the convergence behavior of the affine projection algorithm for adaptive filters. *IEEE Transactions*

- on *Circuits and Systems I: Regular Papers*, 2011, **58**(8): 1813–1826
- 62 Slavakis K, Theodoridis S, Yamada I. Online kernel-based classification using adaptive projection algorithms. *IEEE Transactions on Signal Processing*, 2008, **56**(7): 2781–2796
- 63 Yukawa M. Multikernel adaptive filtering. *IEEE Transactions on Signal Processing*, 2012, **60**(9): 4672–4682
- 64 Zhang Fan. Simulation study of stochastic approximation LMS algorithm of adaline neural network. *Electronic Design Engineering*, 2009, **17**(9): 88–90  
(张帆. Adaline 神经网络随机逼近 LMS 算法的仿真研究. 电子设计工程, 2009, **17**(9): 88–90)
- 65 Han M, Zhong K, Qiu T, Han B. Interval type-2 fuzzy neural networks for chaotic time series prediction: A concise overview. *IEEE Transactions on Cybernetics*, 2019, **49**(7): 2720–2731
- 66 Li Hai-Lin, Liang Ye. Co-movement research of stock time series based on dynamic time warping. *Journal of Data Acquisition & Processing*, 2016, **31**(1): 117–129  
(李海林, 梁叶. 基于动态时间弯曲的股票时间序列联动性研究. 数据采集与处理, 2016, **31**(1): 117–129)
- 67 Li H L. Piecewise aggregate representations and lower-bound distance functions for multivariate time series. *Physica A: Statistical Mechanics and its Applications*, 2015, **427**: 10–25
- 68 Han Min, Wang Ya-Nan. Multivariate time series online predictor with Kalman filter trained reservoir. *Acta Automatica Sinica*, 2010, **36**(1): 169–173  
(韩敏, 王亚楠. 基于 Kalman 滤波的储备池多元时间序列在线预报器. 自动化学报, 2010, **36**(1): 169–173)
- 69 Aghabozorgi S, Teh Y W. Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, 2014, **41**(4): 1301–1314
- 70 Jeon S, Hong B, Chang V. Pattern graph tracking-based stock price prediction using big data. *Future Generation Computer Systems*, 2018, **80**: 171–187
- 71 Wan Xiao-Ji, Li Hai-Lin. Financial multivariate time series data analysis based on feature representation. *Statistics and Decision*, 2015, (23): 151–155  
(万校基, 李海林. 基于特征表示的金融多元时间序列数据分析. 统计与决策, 2015, (23): 151–155)
- 72 Lu L, Zhao H Q, Chen B D. Time series prediction using kernel adaptive filter with least mean absolute third loss function. *Nonlinear Dynamics*, 2017, **90**(2): 999–1013
- 73 Takizawa M A, Yukawa M. Efficient dictionary-refining kernel adaptive filter with fundamental insights. *IEEE Transactions on Signal Processing*, 2016, **64**(16): 4337–4350
- 74 Anava O, Hazan E, Mannor S, Shamir O. Online learning for time series prediction. In: Proceedings of the 26th Annual Conference on Learning Theory. Princeton University, USA: JMLR.org, 2013. 172–184
- 75 Santos J D A, Barreto G A. An outlier-robust kernel RLS algorithm for nonlinear system identification. *Nonlinear Dynamics*, 2017, **90**(3): 1707–1726
- 76 Hong Y S T. Dynamic nonlinear state-space model with a neural network via improved sequential learning algorithm for an online real-time hydrological modeling. *Journal of Hydrology*, 2012, **468-469**: 11–21
- 77 Chen B D, Liang J L, Zheng N N, Principe J C. Kernel least mean square with adaptive kernel size. *Neurocomputing*, 2016, **191**: 95–106
- 78 Song Q. Time series prediction based on online learning. In: Proceedings of the 14th IEEE International Conference on Machine Learning and Applications. Miami, USA: IEEE, 2015. 857–864
- 79 Tuia D, Munoz-Mari J, Rojo-Alvarez J L, Martinez-Ramon M, Camps-Valls G. Explicit recursive and adaptive filtering in reproducing kernel Hilbert spaces. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(7): 1413–1419
- 80 Wang S Y, Zheng Y F, Duan S K, Wang L D, Tan H T. Quantized kernel maximum correntropy and its mean square convergence analysis. *Digital Signal Processing*, 2017, **63**: 164–176
- 81 Constantin I, Lengelle R. Performance analysis of kernel adaptive filters based on RLS algorithm. In: Proceedings of the 25th International Conference on Microelectronics. Beirut, Lebanon: IEEE, 2013. 1–4
- 82 Van Vaerenbergh S, Santamaria I, Lazaro-Gredilla M. Estimation of the forgetting factor in kernel recursive least squares. In: Proceedings of 2012 IEEE International Workshop on Machine Learning for Signal Processing. Santander, Spain: IEEE, 2012. 1–6
- 83 Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y. Online passive-aggressive algorithms. *The Journal of Machine Learning Research*, 2006, **7**: 551–585
- 84 Mairal J, Bach F, Ponce J, Sapiro G. Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 2010, **11**: 19–60
- 85 Guan N Y, Tao D C, Luo Z G, Yuan B. Online nonnegative matrix factorization with robust stochastic approximation. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(7): 1087–1099
- 86 Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research*, 2011, **12**: 2121–2159
- 87 Zhao P L, Hoi S C H, Jin R. Double updating online learning. *The Journal of Machine Learning Research*, 2011, **12**: 1587–1615
- 88 Wang Z, Crammer K, Vucetic S. Breaking the curse of kernelization: Budgeted stochastic gradient descent for large-scale SVM training. *The Journal of Machine Learning Research*, 2012, **13**(1): 3103–3131
- 89 Le T, Nguyen T D, Nguyen V, Phung D. Approximation vector machines for large-scale online learning. *The Journal of Machine Learning Research*, 2017, **18**(1): 3962–4016
- 90 Hoi S C H, Wang J L, Zhao P L. LIBOL: A library for online learning algorithms. *The Journal of Machine Learning Research*, 2014, **15**(1): 495–499
- 91 Bottou L, Le Cun Y. Large-scale online learning. In: Proceedings of the 16th International Conference on Neural Information Processing Systems. Vancouver, BC, Canada, 2003. 217–224
- 92 Dieuleveut A, Flammarion N, Bach F. Harder, better, faster, stronger convergence rates for least-squares regression. *The Journal of Machine Learning Research*, 2017, **18**(101): 1–51
- 93 Lin J H, Zhou D X. Online learning algorithms can converge comparably fast as batch learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(6): 2367–2378
- 94 Flores A, de Lamare R C. Set-membership adaptive kernel NLMS algorithms: Design and analysis. *Signal Processing*, 2019, **154**: 1–14
- 95 Liu Y Q, Sun C, Jiang S D. Kernel filtered-x LMS algorithm for active noise control system with nonlinear primary path. *Circuits, Systems, and Signal Processing*, 2018, **37**(12): 5576–5594
- 96 Lei Y W, Shi L, Guo Z C. Convergence of unregularized online learning algorithms. *The Journal of Machine Learning Research*,

2017, **18**(171): 1–3

- 97 Ma Y T, Zheng T. Stabilized sparse online learning for sparse data. *The Journal of Machine Learning Research*, 2017, **18**(131): 1–36
- 98 Wada T, Fukumori K, Tanaka T. Dictionary learning for gaussian kernel adaptive filtering with variablekernel center and width. In: Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada: IEEE, 2018. 2766–2770
- 99 Wang S Y, Dang L J, Chen B D, Ling C X, Wang L D, Duan S K. Kernel online learning algorithm with scale adaptation. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2018, **65**(11): 1788–1792
- 100 Ohnishi M, Yukawa M. Online nonlinear estimation via iterative  $L^2$ -space projections: Reproducing kernel of subspace. *IEEE Transactions on Signal Processing*, 2018, **66**(15): 4050–4064
- 101 Shin B S, Yukawa M, Cavalcante R L G, Dekorsy A. Distributed adaptive learning with multiple kernels in diffusion networks. *IEEE Transactions on Signal Processing*, 2018, **66**(21): 5505–5519



**韩 敏** 大连理工大学电子信息与电气工程学部教授. 主要研究方向为模式识别, 复杂系统建模及时间序列预测. 本文通信作者.

E-mail: minhan@dlut.edu.cn

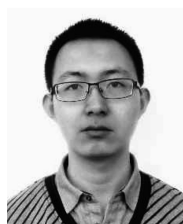
(**HAN Min** Professor at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. Her research interest covers pattern recognition, modeling of complex system, and time series prediction. Corresponding author of this paper.)



**马俊珠** 大连理工大学电子信息与电气工程学部硕士研究生. 主要研究方向为时间序列在线建模, 预测.

E-mail: majunzhu@mail.dlut.edu.cn

(**MA Jun-Zhu** Master student at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. Her research interest covers time series online modeling and prediction.)



**任伟杰** 大连理工大学电子信息与电气工程学部博士研究生. 主要研究方向为时间序列分析和特征选择.

E-mail: renweijie@mail.dlut.edu.cn

(**REN Wei-Jie** Ph.D. candidate at the Faculty of Electronic Information and Electrical Engineering,

Dalian University of Technology. His research interest covers time series analysis and feature selection.)



**钟 凯** 大连理工大学电子信息与电气工程学部博士研究生. 主要研究方向为工业过程监控, 故障诊断.

E-mail: zhongkai0402@mail.dlut.edu.cn

(**ZHONG Kai** Ph.D. candidate at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. His research interest covers industrial process monitoring and fault diagnosis.)