



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于核共轭梯度演化模糊系统的混沌时间序列在线预测

胡磊, 韩敏

引用本文:

胡磊, 韩敏. 基于核共轭梯度演化模糊系统的混沌时间序列在线预测[J]. *控制与决策*, 2024, 39(9): 3099–3107.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0573>

您可能感兴趣的其他文章

Articles you may be interested in

[基于KPCA和G-G聚类的多元时间序列模糊分段](#)

Fuzzy segmentation of multivariate time series with KPCA and G-G clustering

控制与决策. 2021, 36(1): 115–124 <https://doi.org/10.13195/j.kzyjc.2019.0849>

[基于矩阵的双论域模糊概率粗糙集增量更新算法](#)

Incremental updating of fuzzy probability rough sets over two universes based on matrix method

控制与决策. 2021, 36(3): 553–564 <https://doi.org/10.13195/j.kzyjc.2019.0692>

[一种基于池计算的宽度学习系统](#)

A broad learning system based on reservoir computing

控制与决策. 2021, 36(9): 2203–2210 <https://doi.org/10.13195/j.kzyjc.2019.1729>

[一种基于深度学习的时间序列预测方法](#)

A time series prediction method based on deep learning

控制与决策. 2021, 36(3): 645–652 <https://doi.org/10.13195/j.kzyjc.2019.0809>

[基于正态云模型的状态转移算法求解多目标柔性作业车间调度问题](#)

State transition algorithm based on normal cloud model for solving multi-objective flexible job shop scheduling problem

控制与决策. 2021, 36(5): 1181–1190 <https://doi.org/10.13195/j.kzyjc.2019.1233>

基于核共轭梯度演化模糊系统的混沌时间序列在线预测

胡磊¹, 韩敏^{2†}

(1. 大连理工大学 电子信息与电气工程学部, 辽宁 大连 116024;

2. 大连理工大学 教育部工业设备智能控制与优化重点实验室, 辽宁 大连 116024)

摘要: 为了增强演化模糊系统在混沌时间序列中的在线预测性能, 提出一种基于核共轭梯度的演化模糊系统模型. 该模型通过基于结构演化的前件部分, 挖掘并捕获隐藏在时间序列中的模糊规则. 同时, 基于参数更新的后件部分有机结合稀疏学习策略和核共轭梯度算法, 以降低计算复杂度和提高模型收敛性能. 基于核共轭梯度的演化模糊系统不仅剔除了样本中的冗余信息, 而且有效平衡了模型的预测精度和预测效率. 实验结果表明, 所提模型在标杆和真实混沌时间序列的在线预测任务中具有有效性.

关键词: 混沌时间序列; 在线预测; 演化模糊系统; 模糊规则; 稀疏学习策略; 核共轭梯度

中图分类号: TP183 文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0573

引用格式: 胡磊, 韩敏. 基于核共轭梯度演化模糊系统的混沌时间序列在线预测[J]. 控制与决策, 2024, 39(9): 3099-3107.

Online prediction of chaotic time series based on kernel conjugate gradient evolving fuzzy system

HU Lei¹, HAN Min^{2†}

(1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China; 2. Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian 116024, China)

Abstract: To enhance the online prediction performance of evolutionary fuzzy systems in chaotic time series, an evolving fuzzy system based on kernel conjugate gradient (EFS-KCG) is proposed. The model explores and captures the fuzzy rules hidden in the time series through a structural evolution-based antecedent part. Meanwhile, the consequent part based on parameter update organically combines sparse learning strategy and kernel conjugate gradient algorithm to reduce the computational complexity and improve the convergence performance of the model. The EFS-KCG not only eliminates the redundant information in the samples, but also effectively balances the prediction accuracy and efficiency of the model. Experiment results show that the proposed model is effective in online prediction tasks for both benchmark and real chaotic time series.

Keywords: chaotic time series; online prediction; evolving fuzzy system; fuzzy rules; sparse learning strategy; kernel conjugate gradient

0 引言

当今世界正处于科技快速发展的大数据时代, 科技的进步改变了数据采集的方式, 促进了大数据的发展^[1]. 数据通常在高速情况下产生, 以在线的方式进行传输, 并且呈现出复杂的结构特性^[2-3]. 例如, 天文、水文和交通等领域的复杂系统存在混沌特性, 通过观测具备混沌特性的系统获得的数据属于混沌时间序列, 表现出非线性、非平稳和复杂难预测的特征^[4]. 离线预测模型因为采用批处理方式训练数据, 无法满足数据的实时性. 因此, 针对混沌时间序列开发相应

的在线预测模型成为一个热门且具有挑战性的问题. 核自适应滤波器(kernel adaptive filter, KAF)^[5]因为结构简单且可以迭代更新, 在目前的在线预测领域占据了很大的份额, 但处理复杂混沌时间序列时鲁棒性差、预测性能较弱^[6]. 为了更好地研究在线预测问题, 学者们提出了演化模糊系统(evolutionary fuzzy system, EFS)^[7]作为解决这一问题的新导向. EFS是一种计算智能方法, 可提取混沌时间序列中的隐藏信息, 同时调整其功能和结构, 克服KAF鲁棒性弱的问题^[8]. EFS具有以下特点: 1) 在未知时间序列中能通

收稿日期: 2023-04-27; 录用日期: 2023-08-16.

基金项目: 国家自然科学基金项目(62173063).

†通讯作者. E-mail: minhan@dlut.edu.cn.

过学习自适应更新模型内部结构;2)可跟随时间序列动态变化;3)即使是处理复杂混沌时间序列,也有良好的预测性能。

EFS的基本思想是先对样本进行模糊聚类,在不破坏数据结构的情况下实现稀疏性,然后再进行预测。值得注意的是,EFS在在线学习方面的有效性已经被许多学者通过研究验证和解释^[9]。EFS模型最早可以追溯到Angelov等^[10]提出的进化Takagi-Sugeno (evolving Takagi-Sugeno, ETS)模型,通过结合有监督学习和无监督学习递归更新模型结构和参数,并使用加权递归最小二乘(weighted recursive least square, wRLS)算法计算模型输出。随后,Lemos等^[11]提出了基于EFS改进的多变量高斯演化模糊建模系统(multivariable Gaussian evolving fuzzy modeling system, eMG),前件部分使用多元高斯隶属度函数以保存输入变量之间的交互信息,但其需要生成大量的规则才可以得到精确输出,这可能会导致模型结构冗余。为增强模型的非线性处理能力,Vieira等^[12]结合KAF和EFS,提出了一种增强的基于核递归最小二乘的演化参与式学习(enhanced evolving participatory learning with kernel recursive least square, ePL-KRLS+)模型。该模型采用核递归最小二乘(kernel recursive least square, KRLS)算法作为后件部分学习器,通过核函数将非线性数据映射为高维特征空间即重构核希尔伯特空间(reproducing kernel Hilbert space, RKHS)中的线性数据,然后采用线性方法处理,实现高效逼近非线性系统的目标。近期,Alves等^[13]提出了一种基于核递归最小二乘与距离相关的演化参与式学习(evolutionary participatory learning with kernel recursive least square and distance correlation, ePL-KRLS-DISCO)模型。相比传统EFS模型集成不同规则的局部输出作为模型全局输出,ePL-KRLS-DISCO模型根据最相容规则计算全局输出,不仅能降低计算成本,还能避免相容性差的规则对模型最终输出的干扰。然而,ePL-KRLS-DISCO模型的后件部分采用KRLS算法进行参数更新和计算输出,虽具有良好的非线性能力和收敛速度,但计算成本较高。

为平衡预测精度和预测效率,本文提出一种基于核共轭梯度的演化模糊系统(evolutionary fuzzy system based on kernel conjugate gradient, EFS-KCG)模型。EFS-KCG模型在前件部分的规则生成阶段引入余弦相似度,增加了新样本与当前规则的相关性分析。采用核共轭梯度(kernel conjugate gradient, KCG)算法作为后件部分的学习器,在不降低收敛速度的情况下有效降低计算复杂度。同时,模型通过基于矢量

投影的稀疏学习策略剔除冗余样本,避免样本中的冗余信息降低预测精度。最终,所提EFS-KCG模型可以在混沌时间序列中实现在线单步预测和多步预测。

1 演化模糊系统

EFS模型包括前件部分和后件部分,具体可表示为

$$R_i : \underbrace{\text{if } (x_1 \text{ is } \mathcal{A}_{i,1}) \text{ and } \dots \text{ and } (x_m \text{ is } \mathcal{A}_{i,m})}_{\text{前件部分}}, \\ \text{then } \underbrace{y_i = a_{i,0} + a_{i,1}x_1 + \dots + a_{i,m}x_m}_{\text{后件部分}}. \quad (1)$$

其中: $i = 1, 2, \dots, L$, L 是模糊规则数, R_i 是第*i*个模糊规则, $\mathcal{A}_{i,l}$ 是 R_i 对应的模糊集; $\mathbf{x} = [x_1, x_2, \dots, x_m]^T \in \mathbb{R}^m$ 为输入变量, m 为输入变量维度, $a_{i,l}$ ($l = 0, 1, \dots, m$)表示后件参数; y_i 是根据 R_i 计算得到的局部输出,也叫作第*i*个线性系统的输出。

前件部分是模糊规则形成阶段,模糊规则对应模糊集可看作线性子系统。后件部分会根据模糊规则计算每个线性子系统的输出,是EFS模型重要组成部分。各线性子系统的输出是局部输出,耦合各局部输出可得模型全局输出,且线性子系统对模型全局输出的贡献水平与每条规则的触发程度成正比。

2 基于核共轭梯度的演化模糊系统

本节首先介绍稀疏学习策略和核共轭梯度算法,然后提出基于核共轭梯度的演化模糊系统模型,最后对模型性能进行分析。

2.1 稀疏学习策略

在线学习时,数据样本 $\{(\mathbf{x}(n), y(n))\}_{n=1}^N$ (N 为样本数)是按照顺序依次抵达的,随着时间的推移,需要耗费的计算资源越来越多。为了解决这一问题,学者们提出了一系列稀疏化方法^[14],包括近似线性相关、角度准则、相干准则和惊奇准则等。这些方法的基本思想是丢弃携带冗余信息的样本,保留携带重要信息的样本并纳入样本字典。本文采用矢量投影方法^[15],通过限制新样本和当前样本字典成员之间的余弦值以实现样本稀疏学习。该策略选择包含更多有效信息的样本来保证样本字典结构更加紧凑,从而降低模型的计算复杂度。新样本 $\mathbf{x}(n)$ 与当前样本字典 $D(n-1)$ 中最相关成员的余弦值计算方法为

$$\cos = \frac{\langle \varphi(\mathbf{x}(n)), \varphi(D_{j^*}(n-1)) \rangle}{\|\varphi(\mathbf{x}(n))\|_2 \|\varphi(D_{j^*}(n-1))\|_2}. \quad (2)$$

其中: $j = 1, 2, \dots, B$, B 为样本字典成员; $j^* = \arg \max_{1 \leq j \leq B} \cos(\varphi(\mathbf{x}(n)), \varphi(D_j(n-1)))$, n 是当前时间步, $\varphi(\cdot)$ 是核函数映射, $D_j(n-1)$ 是 $D(n-1)$ 中第*j*个成员; $\langle \cdot, \cdot \rangle$ 和 $\|\cdot\|_2$ 分别表示内积运算和 L_2 范数。

基于矢量投影的稀疏学习策略通过判断新样本与样本字典中最相关成员的余弦值是否超过阈值,分析新样本与当前样本字典是否存在关联.也就是说,判断新样本是否具有与样本字典相同的流形信息.因此,基于矢量投影的稀疏学习策略相当于维持低维流形信息的特征选择过程.

2.2 核共轭梯度算法

核共轭梯度(kernel conjugate gradient, KCG)方法由Zhang等^[16]提出,是核技巧成功应用于机器学习和信号处理中的非线性问题的重要体现.在文献[16]中,Zhang等对在线模式下的KCG算法进行了深入的研究.共轭梯度(conjugate gradient, CG)算法^[17]是数值优化和信号处理的一种常用方法,本质上是具有迭代性质的,可以采取降秩处理技术来平衡收敛速度和计算复杂度,还对过拟合问题具有良好的鲁棒性.KCG算法先将CG方法的解向量表示为输入向量的线性组合形式,再通过核技巧进行求解.KCG算法的收敛速度与KRLS算法相当,但是计算成本仅为KRLS算法的四分之一^[16].而且,相比KRLS算法,KCG算法的预设参数较少.假设第 $n-1$ 个时间步的样本字典为 $D(n-1) = \{d_1, d_2, \dots, d_B\}$ (B 为字典成员数),则目标函数 $J(\cdot)$ 为

$$J(n-1) = \min \|Y - G^{n-1}\theta^{n-1}\|_2^2. \quad (3)$$

其中: Y 是样本输出, θ^{n-1} 是权值向量, G^{n-1} 是样本字典对应的格拉姆矩阵,其表达式为

$$G^{n-1} = \begin{bmatrix} \kappa(d_1, d_1) & \dots & \kappa(d_1, d_B) \\ \vdots & \ddots & \vdots \\ \kappa(d_B, d_1) & \dots & \kappa(d_B, d_B) \end{bmatrix}. \quad (4)$$

$\kappa(\cdot, \cdot)$ 为高斯核函数,其计算方法为

$$\kappa(\mathbf{x}(n_1), \mathbf{x}(n_2)) = \exp\left(-\frac{\|\mathbf{x}(n_1) - \mathbf{x}(n_2)\|_2^2}{2\sigma_2^2}\right), \quad (5)$$

σ_2 为高斯核带宽.当样本字典 $D(n-1)$ 在容纳新样本 $(\mathbf{x}(n), y(n))$ 时, G^{n-1} 的更新方法为

$$G^n = \begin{bmatrix} G^{n-1} & \mathbf{g}^n \\ (\mathbf{g}^n)^T & q^n \end{bmatrix} \in \mathbb{R}^{(B+1) \times (B+1)}. \quad (6)$$

其中: $\mathbf{g}^n = [\kappa(d_1, \mathbf{x}(n)), \dots, \kappa(d_B, \mathbf{x}(n))]^T$, $q^n = \kappa(\mathbf{x}(n), \mathbf{x}(n))$.KCG的核心在于更新 θ^n .一般情况下,两次迭代的KCG算法在解决非线性问题时具有良好的性能^[18-19],因此本文采用的KCG指代两次迭代下的KCG算法.基于KCG方法, θ^n 的更新过程为

$$\begin{aligned} \mathbf{r}_0 &= [e(n), y(n) - \mathbf{g}^n \theta^{n-1}]^T, \quad \mathbf{p}_1 = G^n \mathbf{r}_0, \\ \alpha_1 &= \langle \mathbf{p}_1, \mathbf{r}_0 \rangle / \langle \mathbf{p}_1, \mathbf{p}_1 \rangle, \quad \mathbf{r}_1 = \mathbf{r}_0 - \alpha_1 \mathbf{p}_1, \\ \mathbf{s}_1 &= G^n \mathbf{r}_1, \quad \beta_1 = \langle \mathbf{s}_1, \mathbf{r}_1 \rangle / \langle \mathbf{p}_1, \mathbf{r}_0 \rangle, \\ \mathbf{p}_2 &= \beta_1 \mathbf{p}_1 + \mathbf{s}_1, \quad \alpha_2 = \langle \mathbf{s}_1, \mathbf{r}_1 \rangle / \langle \mathbf{p}_2, \mathbf{p}_2 \rangle, \\ \theta^n &= [\theta^{n-1}; 0] + (\alpha_1 + \alpha_2 \beta_1) \mathbf{r}_0 + \alpha_2 \mathbf{r}_1, \\ e(n) &= [\mathbf{r}_1 - \alpha_2 \mathbf{p}_2]^T. \end{aligned} \quad (7)$$

其中: $e(n)$ 是算法的训练误差, α_1 、 α_2 和 β_1 均是过渡参数, \mathbf{r}_0 、 \mathbf{r}_1 和 \mathbf{s}_1 都是残差向量, \mathbf{p}_1 和 \mathbf{p}_2 是搜索方向.

2.3 基于核共轭梯度的演化模糊系统模型

本节首先概述EFS-KCG模型框架,然后介绍基于结构演化的前件部分和基于参数更新的后件部分.

2.3.1 模型的结构框架

如图1所示,EFS-KCG模型主要包括两个部分:前件部分和后件部分.

$$R_i : \underbrace{\text{if } \mathbf{x}(n) \text{ is } \mathcal{A}_i}_{\text{前件部分}}, \text{ then } \underbrace{\hat{y}(n) = f_i(\mathbf{x}(n), \theta_i^n)}_{\text{后件部分}}. \quad (8)$$

其中: $i = 1, 2, \dots, L$, L 是模糊规则数; R_i 是第 i 个模糊规则, \mathcal{A}_i 是 R_i 的模糊集, θ_i^n 是 R_i 的后件参数; n 是当前时间步, $\mathbf{x}(n) = [x_1, x_2, \dots, x_m]^T \in \mathbb{R}^m$ 是输入变量, m 是输入变量维度, $\hat{y}(n)$ 是EFS-KCG模型输出.

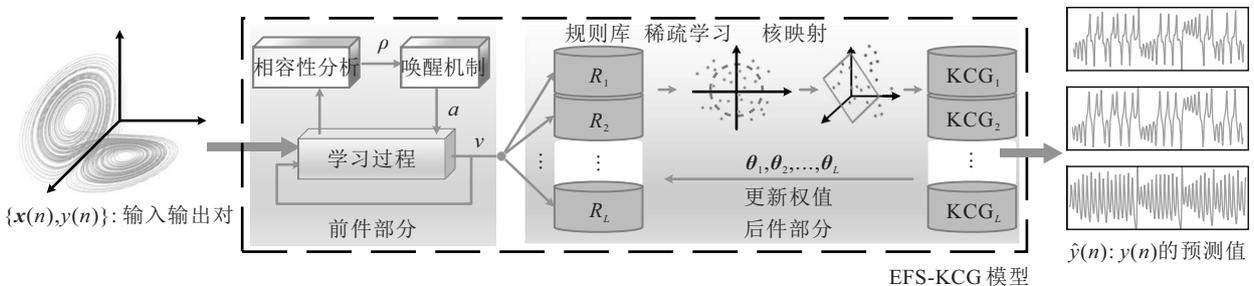


图1 EFS-KCG模型的学习机制

2.3.2 基于结构演化的前件部分

模型在前件部分将输入空间聚类成由不同规则控制的模糊集.在图1中,新输入 $\mathbf{x}(n) \in \mathbb{R}^m$ 进入模型以后,模型会根据规则库中已有规则对其进行判别

并且划分到相应的模糊集.

1) 规则的生成和合并.根据第1个输入 $\mathbf{x}(1)$ 初始化得到第1个规则 R_1 的中心,即 $\mathbf{v}_1(1) = \mathbf{x}(1)$.后续新输入 $\{\mathbf{x}(n)\}_{n=2}^N$ (N 为样本数)进入模型后,模型会

将其纳入到最近规则 R_i 的模糊集中,并根据下式更新其规则中心 $\mathbf{v}_i(n)$:

$$\mathbf{v}_i(n) = \mathbf{v}_i(n-1) + G_i(n)(\mathbf{x}(n) - \mathbf{v}_i(n-1)). \quad (9)$$

其中: $G_i(n)$ 是判别系数,其计算方法为

$$G_i(n) = \alpha(\rho_i(n))^{(1-a_i(n))}, \quad (10)$$

式中 α 为学习率, $\rho_i(n)$ 为相容性系数, $a_i(n)$ 为唤醒系数. $\rho_i(n)$ 可衡量新输入与最近模糊集间的相容程度,其计算方法如下:

$$\rho_i(n) = \left(1 - \frac{\|\mathbf{x}(n) - \mathbf{v}_i(n-1)\|_2}{m}\right) \left(\frac{\zeta + 1}{2}\right). \quad (11)$$

其中: ζ 是 $\mathbf{x}(n)$ 和 $\mathbf{v}_i(n-1)$ 的余弦相似度,计算方法为

$$\zeta = \frac{\mathbf{x}^T(n)\mathbf{v}_i(n-1)}{\|\mathbf{x}(n)\|_2\|\mathbf{v}_i(n-1)\|_2}. \quad (12)$$

唤醒系数 $a_i(n)$ 可作为 $\rho_i(n)$ 的补充,通过创建新规则更好地跟踪时间序列动态变化. 如果 $a_i(n)$ 不超过阈值 τ , 则不更新规则库,并更新 $a_i(n)$ 如下:

$$a_i(n) = a_i(n-1) + \beta(1 - \rho_i(n) - a_i(n-1)), \quad (13)$$

其中 β 为感知率. β 越接近 1, 模型感知相容性系数能力越强; β 越接近 0, 模型感知相容性系数能力越弱. 如果 $a_i(n)$ 超过阈值 τ , 则激活唤醒机制, 建立新规则 R_{i+1} 并初始化新规则的中心 $\mathbf{v}_{i+1}(n) = \mathbf{x}(n)$.

2) 规则的修剪. 为避免因为拟合噪声数据和模型结构冗余导致计算成本高,模型对覆盖率低且针对噪声创建的规则进行修剪^[20-21]. 根据下式计算各规则效用度量,当 R_i 的效用度量 $\eta_i(n)$ 低于阈值 ε 时,修剪该规则^[10]:

$$\eta_i(n) = \sum_{l=1}^n \delta_i(l) / (n - I(i)). \quad (14)$$

其中: n 为当前时间步; $I(i)$ 为第 i 个规则被创建时的时间步; $\delta_i(l)$ 表示第 l 个时间步时第 i 个规则的归一化激活水平,其计算公式为

$$\delta_i(l) = \tau_i / \sum_{i=1}^{L(n)} \tau_i. \quad (15)$$

$L(n)$ 是当前时间步时模糊规则数; τ_i 是第 i 个规则的激活水平,可通过下式计算:

$$\tau_i = \Omega_{i,1} \times \Omega_{i,2} \times \dots \times \Omega_{i,m}. \quad (16)$$

式(16)中 $\Omega_{i,j}$ 为输入中各特征元素对于相对应规则中心元素的高斯隶属度函数,其计算公式如下:

$$\Omega_{i,j} = \exp\left(\frac{-\|\mathbf{x}_j(n) - \mathbf{v}_{i,j}(n)\|_2^2}{2\sigma_1^2}\right). \quad (17)$$

其中: $j = 1, 2, \dots, m$, $\mathbf{x}_j(n)$ 是 $\mathbf{x}(n)$ 中第 j 个元素, $\mathbf{v}_{i,j}(n)$ 是 $\mathbf{v}_i(n)$ 中第 j 个元素, σ_1 是传播系数. 前面已经对前件部分的具体内容进行了介绍,接下来将详细介绍后件部分的主要任务.

2.3.3 基于参数更新的后件部分

模型的后件部分使用基于矢量投影的稀疏学习策略和 KCG 算法计算模型输出. 后件部分先对新样本进行稀疏学习,然后采用 KCG 算法计算模型输出.

1) 稀疏学习策略. 每个规则的模糊集都会通过样本字典对新样本进行筛选. 如果样本字典容纳所有新输入,则计算成本会显著增加. 因此,该模型使用 2.1 节介绍的基于矢量投影的稀疏学习策略对输入进行筛选,其计算方法如式(2)所示. 令 $D_i(n) = [d_{i,1}, \dots, d_{i,B_i}] \in \mathbb{R}^{B_i \times m}$ 表示第 n 个时间步第 i 个规则的样本字典(其中 B_i 是 $D_i(n)$ 中成员数). 模型的稀疏学习过程和样本字典的更新机制可表示为

$$\begin{aligned} \text{if } \cos > \epsilon \text{ then } D_i(n) &= D_i(n-1) \cup \mathbf{x}(n), \\ \text{else } D_i(n) &= D_i(n-1), \end{aligned} \quad (18)$$

其中 ϵ 是矢量投影的阈值.

2) 在线 KCG 方法. 当新样本 $\mathbf{x}(n)$ 完成稀疏学习并纳入样本字典中,即 $D_i(n) = [D_i(n-1) \cup \mathbf{x}(n)]$ 时,根据式(6)和(7)更新样本字典对应的格拉姆矩阵 \mathbf{G}_i^n 和权值向量 $\boldsymbol{\theta}_i^n$. 基于 KCG 方法,模型输出的计算方法如下:

$$\hat{y}(n) = \mathbf{g}_i^n \boldsymbol{\theta}_i^n, \quad (19)$$

其中 $\mathbf{g}_i^n = [\kappa(d_{i,1}, \mathbf{x}(n)), \dots, \kappa(d_{i,B_i}, \mathbf{x}(n))]^T$. 当创建新规则 R_{i+1} 时,新规则对应的权值向量 $\boldsymbol{\theta}_{i+1}^n$ 可通过下式进行初始化:

$$\boldsymbol{\theta}_{i+1}^n = (\kappa(\mathbf{x}(n), \mathbf{x}(n)) + \lambda)^{-1} y(n). \quad (20)$$

其中: $\lambda \in [0, 1]$ 是正则化参数, $y(n)$ 是样本真实输出.

2.4 模型的性能分析

本节从收敛性和计算复杂度两个方面对模型性能进行分析.

2.4.1 收敛性分析

本文提出的 EFS-KCG 模型利用 KCG 算法对后件部分进行参数更新, KCG 算法本质上采用 CG 算法优化目标函数,而不是基于梯度的算法. 基于 CG 算法^[17,19], 权值向量 \mathbf{w}_n 和方向向量 \mathbf{p}_n 的更新方法为

$$\begin{cases} \mathbf{w}_{n+1} = \mathbf{w}_n + \alpha_n \mathbf{p}_n, \\ \mathbf{p}_{n+1} = -\mathbf{g}_{n+1} + \beta_n \mathbf{p}_n. \end{cases} \quad (21)$$

其中: α_n 和 β_n 是迭代更新的过渡参数, \mathbf{g}_n 是梯度向量(负残差向量). 更新 \mathbf{p}_n 是寻找模型损失函数最小值点的关键. 根据 CG 方法的下降条件^[17], 得到

$$\mathbf{g}_{n+1}^T \mathbf{p}_{n+1} = -\|\mathbf{p}_{n+1}\|_2^2 \leq 0. \quad (22)$$

根据 Zoutendijk 准则^[22], 因为迭代搜索方法满足式(21)和(22), 所以 \mathbf{g}_n 和 \mathbf{p}_n 之间的关系可表示为

$$\sum_{n=1}^{\infty} \frac{(\mathbf{g}_n^T \mathbf{p}_n)^2}{\|\mathbf{p}_n\|_2^2} < +\infty. \quad (23)$$

命题1 根据式(21), 模型最终收敛到损失函数最小值点, 即 $\liminf_{n \rightarrow \infty} \|g_n\| = 0$.

证明 通过反证法证明, 假设命题1是假命题, 即 $\liminf_{n \rightarrow \infty} \|g_n\| \geq \xi$, 其中 ξ 是任意小的正常数. 基于式(21), 两边同时乘以 g_n^T 后得到

$$g_n^T p_{n+1} = -g_n^T g_{n+1} + g_n^T \beta_n p_n. \quad (24)$$

基于CG方法的正交条件^[17] $g_n^T g_{n+1} = 0$, 可将 β_n 转化为

$$g_n^T p_{n+1} = g_n^T \beta_n p_n \Rightarrow \beta_n = \frac{g_n^T p_{n+1}}{g_n^T p_n}. \quad (25)$$

根据CG方法的性质有^[17] $g_n = R w_n$, 结合式(21)后得到

$$\begin{cases} g_{n+1} - g_n = R(w_{n+1} - w_n) \\ w_{n+1} = w_n + \alpha_n p_n \end{cases} \Rightarrow \begin{cases} g_{n+1} - g_n = \alpha_n R p_n, \end{cases} \quad (26)$$

其中 R 是与输入相关的对称正定矩阵. 在式(26)两边同时乘以 p_{n+1}^T 后得到

$$p_{n+1}^T g_{n+1} - p_{n+1}^T g_n = \alpha_n p_{n+1}^T R p_n. \quad (27)$$

根据CG方法的共轭条件^[17] $p_{n+1}^T R p_n = 0$, 将其代入到式(27)后有

$$p_{n+1}^T g_{n+1} = p_{n+1}^T g_n \Rightarrow g_{n+1}^T p_{n+1} = g_n^T p_{n+1}. \quad (28)$$

将式(28)代入(25), 可得

$$\beta_n = \frac{g_n^T p_{n+1}}{g_n^T p_n} = \frac{g_{n+1}^T p_{n+1}}{g_n^T p_n}. \quad (29)$$

对式(21)的两边求取 L_2 范数后可得到

$$\begin{aligned} \|p_{n+1}\|_2^2 &= \\ \beta_n^2 \|p_n\|_2^2 - 2g_{n+1}^T(\beta_n p_n) + \|g_{n+1}\|_2^2 &= \\ \beta_n^2 \|p_n\|_2^2 - 2g_{n+1}^T(p_{n+1} + g_{n+1}) + \|g_{n+1}\|_2^2 &= \\ \beta_n^2 \|p_n\|_2^2 - 2g_{n+1}^T p_{n+1} - \|g_{n+1}\|_2^2. \end{aligned} \quad (30)$$

将式(29)代入(30)后得到

$$\begin{aligned} \|p_{n+1}\|_2^2 &= \\ \left(\frac{g_{n+1}^T p_{n+1}}{g_n^T p_n} \right)^2 \|p_n\|_2^2 - 2g_{n+1}^T p_{n+1} - \|g_{n+1}\|_2^2, \end{aligned} \quad (31)$$

然后, 式(31)等号两边同时除以 $(g_{n+1}^T p_{n+1})^2$ 得到

$$\frac{\|p_{n+1}\|_2^2}{(g_{n+1}^T p_{n+1})^2} \leq \frac{\|p_n\|_2^2}{(g_n^T p_n)^2} + \frac{1}{\|g_{n+1}\|_2^2}. \quad (32)$$

根据式(32)的递归关系, 可推导出

$$\frac{(g_n^T p_n)^2}{\|p_n\|_2^2} = \left(\sum_{i=1}^n \frac{1}{\|g_i\|_2^2} \right)^{-1} \leq \frac{\xi^2}{n}. \quad (33)$$

因为 $\sum_{n=1}^{\infty} \frac{\xi^2}{n} = +\infty$, 可进一步推出

$$\sum_{n=1}^{\infty} \frac{(g_n^T p_n)^2}{\|p_n\|_2^2} \geq +\infty. \quad (34)$$

这与式(23)冲突. 因此, 假设“命题1为假命题”是不合理的. 根据命题1, EFS-KCG模型具有收敛性. \square

2.4.2 计算复杂度分析

EFS-KCG模型本质上是先模糊聚类再预测建模, 虽然会比KAF模型存储更多信息, 但训练阶段的计算成本高于KAF模型, 而测试阶段的计算成本具体依赖于数据规模、数据维度和数据质量. EFS-KCG模型的总体计算复杂度需要综合生成的模糊规则数 L 和输入变量维度 m 进行分析. EFS-KCG模型的前件部分主要起到对数据样本模糊聚类的作用, 因此前件部分的计算复杂度为 $O(Lm)$, 与一般EFS模型相差不大. EFS-KCG模型的后件部分采用KCG算法, 承担参数更新和计算输出的功能, 需要根据每个规则的样本字典具体分析. 在后件部分中, B_i ($i = 1, 2, \dots, L$) 是第 i 个规则下字典成员数, G_i^n 的计算复杂度为 $O(B_i m)$, 且还需考虑核函数的计算复杂度, 其余部分中加法和乘法的计算复杂度分别为 $O(2B_i^2 + 8B_i)$ 和 $O(2B_i^2 + 10B_i)$. 值得注意的是, 当样本不满足稀疏化条件时, KCG算法不消耗额外的计算成本. 同时, EFS-KCG模型不仅采用最相容规则计算输出, 而且引入了稀疏学习策略控制样本字典规模, 可以有效降低模型计算复杂度.

3 仿真实验

本节选用Lorenz时间序列和太阳黑子时间序列验证所提EFS-KCG模型的在线预测性能. 选择3种EFS模型(eMG^[11]、ePL-KRLS^[12]、ePL-KRLS-DISCO^[13])和两种KAF模型(KCG^[16]、DASKRLS^[15])作为对比模型. 选择均方根误差(root mean square error, RMSE)和平均绝对误差(mean absolute error, MAE)作为预测精度的评价指标, 计算方法如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (y(n) - \hat{y}(n))^2}, \quad (35)$$

$$MAE = \frac{1}{N} \sum_{n=1}^N |y(n) - \hat{y}(n)|. \quad (36)$$

其中: $y(n)$ 是真实值, $\hat{y}(n)$ 是预测值, N 是样本数. 选择测试时间作为预测效率的评价指标, 并且通过最终生成的规则数判断模型的复杂程度.

3.1 Lorenz时间序列

Lorenz系统是一个经典的非线性动力学系统, 由Lorenz于1963年首次提出^[23-24], 用于描述对流运动的不稳定性. Lorenz系统具有混沌理论的“蝴蝶效应”, 其奇异吸引子和观测到的时间序列如图2所示.

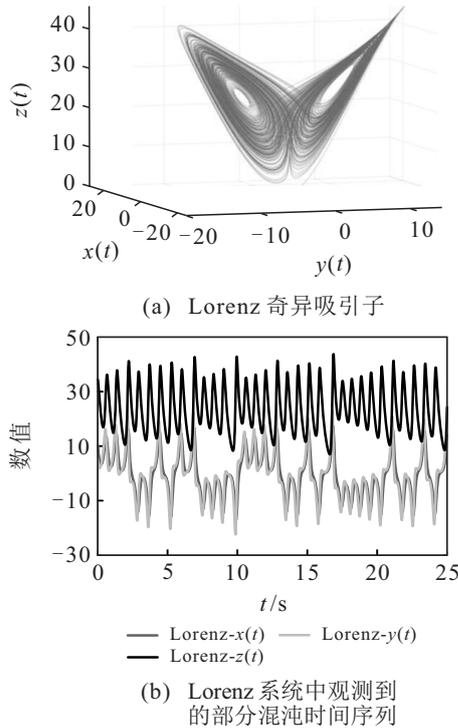


图2 Lorenz 奇异吸引子及相应的混沌时间序列

Lorenz 系统方程可表示为

$$\begin{cases} dx/dt = -ax + yz, \\ dy/dt = b(z - y), \\ dz/dt = -xy + cy - z. \end{cases} \quad (37)$$

其中 a, b, c 是参数. 基于 Lorenz 系统方程, 设定初始条件为 $(x_0, y_0, z_0) = (1, 1, 1)$ 和步长为 0.01, 采用四阶龙格-库塔方法生成 Lorenz 时间序列. 通过 C-C 方法选择合适的延迟时间和嵌入维数后进行相空间重构, 选择 $Lorenz-x(t)$ 作为预测目标, 并划分时间序列样本进行训练和测试. 表 1 为 Lorenz 数据集参数设置.

表1 Lorenz 时间序列: 仿真数据集参数设置

方程参数			延迟时间			嵌入维数			样本数	
a	b	c	$x(t)$	$y(t)$	$z(t)$	$x(t)$	$y(t)$	$z(t)$	训练	测试
10	28	8/3	8	7	8	6	6	6	1800	600

根据网格搜索法对不同模型的参数进行寻优, 得到最终的模型参数如表 2 所示.

图 3 是 EFS-KCG 模型的 10 步预测结果. 由图 3 发现, EFS-KCG 可以很好地拟合真实值曲线, 几乎完全重合, 预测误差在零误差附近且控制在 $[-0.5, 0.5]$. 随着迭代次数增加, 模型会修剪效用度量较低的规则, 最终规则数稳定在 7. 这样可有效降低计算复杂度, 保证模型高效运行. 为检验 EFS-KCG 的优越性, 整理不同模型多步预测结果见表 3.

表2 Lorenz 时间序列: 不同模型的参数设置

模型	参数
KCG ^[17]	$v_0 = 0.95, \sigma = 0.6$
DASKRLS ^[16]	$\alpha = 2, \omega = 10^{-5}, \varepsilon = 0.98, \beta = 0.1, c = 1, \sigma_2 = 10, \sigma_2 = 0.6$
eMG ^[12]	$\alpha = 0.001, \lambda = 10^{-8}, \omega = 10$
ePL-KRLS+ ^[13]	$\beta = \tau = 0.04, \gamma = 0.2, \alpha = 0.001, \lambda = 10^{-8}, r = 0.4, \varepsilon = 0.1$
ePL-KRLS-DISCO ^[14]	$\beta = \tau = 0.04, \alpha = 0.001, \lambda = 10^{-8}, \sigma = 0.6, \varepsilon = 0.1, \omega = 1$
EFS-KCG	$\beta = \tau = 0.04, \alpha = 0.001, \lambda = 10^{-8}, \sigma_1 = \sigma_2 = 0.6, \varepsilon = 0.1, \epsilon = 0.01$

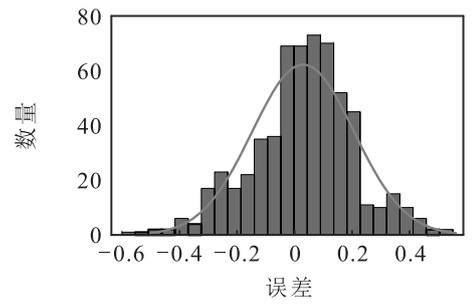
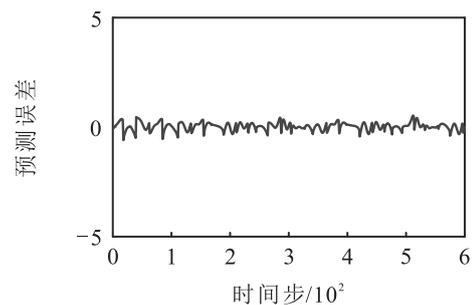
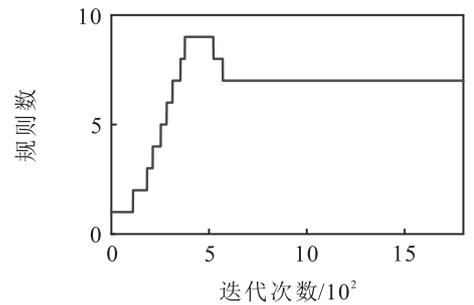
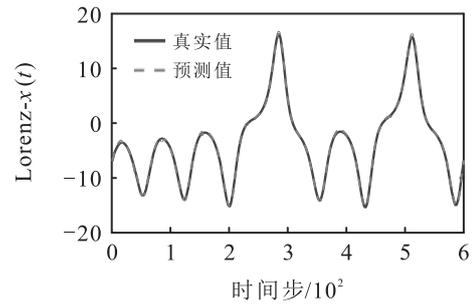


图3 Lorenz 时间序列: EFS-KCG 模型的 10 步预测结果

表3 Lorenz时间序列: 不同模型的多步预测结果

模型	5步预测				10步预测			
	RMSE	MAE	测试时间/s	规则数	RMSE	MAE	测试时间/s	规则数
KCG ^[17]	0.42	0.36	1.13	—	1.42	1.53	2.91	—
DASKRLS ^[16]	0.31	0.71	2.64	—	1.75	1.25	4.03	—
eMG ^[12]	0.47	0.66	4.36	17	1.37	0.17	5.52	16
ePL-KRLS+ ^[13]	0.73	0.55	6.10	2	1.86	1.32	5.94	2
ePL-KRLS-DISCO ^[14]	0.95	0.49	3.23	8	1.04	0.91	6.66	9
EFS-KCG	0.18	0.14	2.47	7	0.90	0.12	3.85	7

由表3可知,相比其他模型,EFS-KCG的预测误差最小,预测精度最优,但学习到的规则数高于ePL-KRLS+。这说明:EFS-KCG通过增加余弦相似性度量优化了规则唤醒机制,从而挖掘出更多重要规则。同时,EFS-KCG没有与eMG一样生成太多规则数造成结构冗余。相比ePL-KRLS-DISCO,EFS-KCG因稀疏学习和KCG方法减少了运行时间,提高了预测效率。

值得注意的是,因为Lorenz混沌时间序列变化比较缓慢,所以KCG模型有足够的进行高效建模,预测效率高于所提模型,仅在Lorenz混沌时间序列上进行实验不能很好地评估EFS-KCG模型的在线性能。因此,为进一步检验所提模型的在线性能,选取真实数据集进行仿真实验。

3.2 太阳黑子时间序列

因为太阳黑子数具有混沌特性且与黄河流域径流量相关^[25],本节选取1749年1月至2022年3月的月平均太阳黑子数和黄河月径流量作为特征变量对太阳黑子数进行预测。选择延迟时间和嵌入维数分别为[2, 4]和[4, 1],对太阳黑子-黄河径流量时间序列进行相空间重构,并划分成2400个训练样本和800个测试样本。不同模型参数设置如表4。

图4是EFS-KCG模型在太阳黑子时间序列中5步和10步的预测结果。由图4发现,EFS-KCG模型能

表4 太阳黑子时间序列: 不同模型的参数设置

模型	参数
KCG ^[17]	$v_0 = 0.95, \sigma = 0.6$
DASKRLS ^[16]	$\alpha = 2, \omega = 10^{-4}, \varepsilon = 0.98, \beta = 0.15, c = 1, \sigma_2 = 30, \sigma_2 = 1$
eMG ^[12]	$\alpha = 0.001, \lambda = 10^{-6}, \omega = 10$
ePL-KRLS+ ^[13]	$\beta = \tau = 0.01, \gamma = 0.2, \alpha = 0.001, \lambda = 10^{-6}, r = 0.4, \varepsilon = 0.08$
ePL-KRLS-DISCO ^[14]	$\beta = \tau = 0.01, \alpha = 0.001, \lambda = 10^{-6}, \sigma = 0.6, \varepsilon = 0.08, \omega = 1$
EFS-KCG	$\beta = \tau = 0.01, \alpha = 0.001, \lambda = 10^{-6}, \sigma_1 = \sigma_2 = 10, \varepsilon = 0.08, \epsilon = 0.015$

学习出太阳黑子时间序列的内在演化规律,最终学习到的规则数稳定在11。虽然预测曲线有一定的滞后性,但基本能跟踪真实曲线的动态变化。5步和10步预测误差均在零误差附近波动且符合正态分布,5步预测误差控制在[-30, 30]。

为检验EFS-KCG模型的优越性,将其预测结果与其他模型对比。表5中,EFS-KCG模型预测精度最高且测试时间最短。这说明所提模型在处理复杂时间序列时相比对比模型具有良好的鲁棒性。因此,所提模型可有效平衡预测精度和效率,在太阳黑子时间序列中进行多步在线预测具有有效性。

表5 太阳黑子时间序列: 不同模型的多步预测结果

模型	5步预测				10步预测			
	RMSE	MAE	测试时间/s	规则数	RMSE	MAE	测试时间/s	规则数
KCG ^[17]	24.58	13.88	8.01	—	30.45	20.83	12.88	—
DASKRLS ^[16]	25.51	15.09	19.92	—	29.96	19.94	23.81	—
eMG ^[12]	16.08	10.95	6.58	19	24.57	18.02	6.66	19
ePL-KRLS+ ^[13]	32.17	26.09	6.25	3	63.84	48.62	7.96	3
ePL-KRLS-DISCO ^[14]	20.39	15.20	6.62	3	38.88	27.37	6.67	3
EFS-KCG	9.52	7.20	3.47	11	18.71	14.40	5.31	11

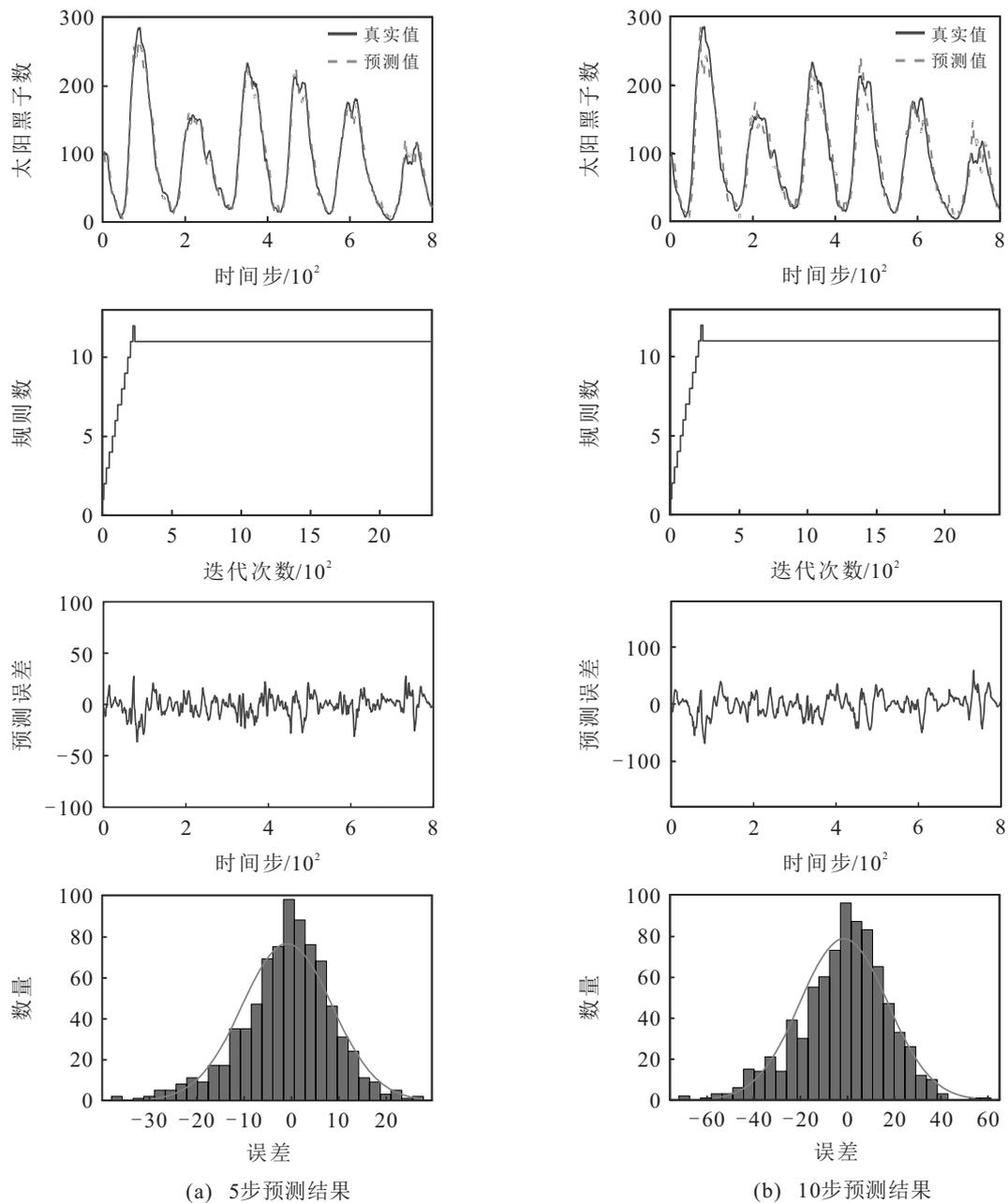


图4 太阳黑子时间序列: EFS-KCG模型的5步和10步预测结果

4 结论

本文提出了一种基于核共轭梯度的演化模糊系统模型进行混沌时间序列的在线预测. 该模型通过前件部分挖掘隐藏在复杂时间序列中的潜在演化规则, 在后件部分有机结合稀疏学习策略和核共轭梯度方法, 不仅有效消除样本中冗余信息和降低模型计算复杂度, 而且提高了模型的收敛性能和预测性能. 本文在Lorenz时间序列和太阳黑子时间序列上进行仿真实验, 验证了所提模型的有效性和优越性.

参考文献(References)

- [1] Fan J Q, Han F, Liu H. Challenges of big data analysis[J]. National Science Review, 2014, 1(2): 293-314.
- [2] 王姝, 魏楠, 孟思彤, 等. 一种基于时间动态因果图
- [3] 张淑美, 王福利, 王姝, 等. 基于数据特性分析的多变量过程监测[J]. 东北大学学报: 自然科学版, 2017, 38(5): 609-613.
(Zhang S M, Wang F L, Wang S, et al. Multivariate process monitoring based on the characteristic analysis of the data[J]. Journal of Northeastern University: Natural Science, 2017, 38(5): 609-613.)
- [4] 那晓栋, 王嘉宁, 刘墨燃, 等. 基于层次化可塑性回声状态网络的混沌时间序列预测[J]. 控制与决策, 2023, 38(1): 133-142.
- [5] 复杂工业过程故障预测方法[J]. 控制与决策, 2024, 39(7): 2242-2250.
(Wang S, Wei N, Meng S T, et al. A fault prediction method for complex industrial process based on time dynamic causality diagram[J]. Control and Decision, 2024, 39(7): 2242-2250.)

- (Na X D, Wang J N, Liu M R, et al. Hierarchical plasticity echo state network for chaotic time series prediction[J]. *Control and Decision*, 2023, 38(1): 133-142.)
- [5] Liu W F, Príncipe J C, Haykin S. Kernel adaptive filtering: A comprehensive introduction[M]. Hoboken: John Wiley & Sons, 2010.
- [6] 韩敏, 马俊珠, 任伟杰, 等. 基于核自适应滤波器的时间序列在线预测研究综述[J]. *自动化学报*, 2021, 47(4): 730-746.
(Han M, Ma J Z, Ren W J, et al. A survey of time series online prediction based on kernel adaptive filters[J]. *Acta Automatica Sinica*, 2021, 47(4): 730-746.)
- [7] Pratama M, Lu J, Lughofer E, et al. An incremental learning of concept drifts using evolving type-2 recurrent fuzzy neural networks[J]. *IEEE Transactions on Fuzzy Systems*, 2017, 25(5): 1175-1192.
- [8] Ge D J, Zeng X J. Learning data streams online—An evolving fuzzy system approach with self-learning/adaptive thresholds[J]. *Information Sciences*, 2020, 507: 172-184.
- [9] Škrjanc I, Iglesias J A, Sanchis A, et al. Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A Survey[J]. *Information Sciences*, 2019, 490: 344-368.
- [10] Angelov P P, Filev D P. An approach to online identification of Takagi-Sugeno fuzzy models[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2004, 34(1): 484-498.
- [11] Lemos A, Caminhas W, Gomide F. Multivariable Gaussian evolving fuzzy modeling system[J]. *IEEE Transactions on Fuzzy Systems*, 2011, 19(1): 91-104.
- [12] Vieira R, Maciel L, Ballini R, et al. Stock market price forecasting using a kernel participatory learning fuzzy model[C]. *North American Fuzzy Information Processing Society Annual Conference*. Cham: Springer, 2018: 361-373.
- [13] Alves K S T R, Pestana de Aguiar E. A novel rule-based evolving Fuzzy System applied to the thermal modeling of power transformers[J]. *Applied Soft Computing*, 2021, 112: 107764.
- [14] Tan F X, Guan X P. Research progress on intelligent system's learning, optimization, and control—Part II: Online sparse kernel adaptive algorithm[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, 50(12): 5369-5385.
- [15] Zhong K, Ma J Z, Han M. Online prediction of noisy time series: Dynamic adaptive sparse kernel recursive least squares from sparse and adaptive tracking perspective[J]. *Engineering Applications of Artificial Intelligence*, 2020, 91: 103547.
- [16] Zhang M, Wang X J, Chen X M, et al. The kernel conjugate gradient algorithms[J]. *IEEE Transactions on Signal Processing*, 2018, 66(16): 4377-4387.
- [17] Chang P S, Willson A N. Analysis of conjugate gradient algorithms for adaptive filtering[J]. *IEEE Transactions on Signal Processing*, 2000, 48(2): 409-418.
- [18] 齐乐天, 王世元, 沈明琳, 等. 基于Nyström柯西核共轭梯度算法的混沌时间序列预测[J]. *物理学报*, 2022, 71(10): 392-401.
(Qi L T, Wang S Y, Shen M L, et al. Prediction of chaotic time series based on Nyström Cauchy kernel conjugate gradient algorithm[J]. *Acta Physica Sinica*, 2022, 71(10): 392-401.)
- [19] Zhang H N, Yang B, Wang L, et al. General cauchy conjugate gradient algorithms based on multiple random Fourier features[J]. *IEEE Transactions on Signal Processing*, 2021, 69: 1859-1873.
- [20] Angelov P. Evolving Takagi-Sugeno fuzzy systems from streaming data (eTS+)[J]. *Evolving Intelligent Systems: Methodology and Applications*, DOI: 10.1002/9780470569962.CH2.
- [21] Angelov P. Fuzzily connected multimodel systems evolving autonomously from data streams[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2011, 41(4): 898-910.
- [22] Zoutendijk G. Nonlinear programming, computational methods[J]. *Integer & Nonlinear Programming*, 1970: 37-86.
- [23] Lorenz E N. Deterministic nonperiodic flow[J]. *Journal of the Atmospheric Sciences*, 1963, 20(2): 130-141.
- [24] Li D C, Han M, Wang J. Chaotic time series prediction based on a novel robust echo state network[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(5): 787-799.
- [25] 李春晖, 杨志峰. 基于Morlet小波的太阳活动与黄河流域径流量关系分析[J]. *水资源与水工程学报*, 2004, 15(3): 1-4.
(Li C H, Yang Z F. Relationship between Sun activities and natural runoff in the Yellow River Basin based on morlet wavelet[J]. *Journal of Water Resources and Water Engineering*, 2004, 15(3): 1-4.)

作者简介

胡磊(1997—), 男, 博士生, 从事多元混沌时间序列在线预测的研究, E-mail: hl666888@mail.dlut.edu.cn;

韩敏(1959—), 女, 教授, 博士生导师, 从事复杂工业系统建模与控制、智能技术及优化算法等研究, E-mail: minhan@dlut.edu.cn.