

带FIR滤波的非线性滑动平均动态软测量模型

孙文心¹, 马君霞¹, 熊伟丽^{1, 2†}

(1. 江南大学 物联网工程学院, 江苏 无锡 214122; 2. 江南大学 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122)

摘要: 非线性滑动平均(NMA)模型能有效描述工业过程的动态特性, 是一种典型的动态软测量模型. 而受限于模型复杂度, NMA模型的输入时序边界相对较窄, 难以适应带有大滞后或强测量噪声的动态工业过程. 针对该问题, 本文将NMA模型与结构简单、输入时序边界宽的FIR滤波器相结合, 构造一种非线性、强抗干扰的软测量建模策略. 并设计层白化结构来避免二者间的参数耦合现象, 采用Adam算法进行同步优化, 提高模型的预测精度及训练效率. 最后, 利用数值仿真和硫回收过程建模实验, 验证所提模型的预测精度以及模型设计的合理性.

关键词: 动态软测量; NMA模型; FIR滤波; 参数解耦

引用格式: 孙文心, 马君霞, 熊伟丽. 带FIR滤波的非线性滑动平均动态软测量模型. 控制理论与应用, 2024, 41(4): 609–618

DOI: 10.7641/CTA.2023.11190

Nonlinear moving average dynamic soft sensor model with FIR filter

SUN Wen-xin¹, MA Jun-xia¹, XIONG Wei-li^{1, 2†}

(1. School of the Internet of Things Engineering, Jiangnan University, Wuxi Jiangsu 214122, China;
2. China Key Laboratory of Advanced Process Control for Light Industry Ministry of Education,
Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: The nonlinear moving average (NMA) model has been widely applied in the field of dynamic soft sensing. However, due to the high model complexity, the averaging horizon of the NMA model is relatively short. As a result, its accuracy could be compromised if the industrial process embodies long time delay or strong measurement noise. To solve this problem, a novel soft sensor modeling strategy, with relatively strong anti-interference ability and long averaging horizon, is proposed by combining the NMA model with the FIR filter. Besides, parameters in the NMA model and FIR filter are optimized synchronously based on the Adam algorithm. Meanwhile, the layer whitening strategy is designed to avoid the parameter coupling phenomenon between the two structures mentioned. Finally, numerical simulations and sulfur recovery process modeling experiments have been conducted to verify the prediction accuracy of the proposed model and the rationality of model design.

Key words: dynamic soft sensor; NMA model; FIR filter; parameter decoupling

Citation: SUN Wenxin, MA Junxia, XIONG Weili. Nonlinear moving average dynamic soft sensor model with FIR filter. *Control Theory & Applications*, 2024, 41(4): 609–618

1 引言

质量变量的实时采样是化工过程闭环控制系统不可缺少的组成部分. 但由于测量技术、环境及成本等因素, 其中一些质量变量(称为主导变量)难以直接测量. 软测量是解决上述问题的可行技术, 使用可测量的工业变量(称为辅助变量)和预测模型实时估计主导变量. 目前, 对软测量的研究大多集中在静态建模上, 即假定建模对象满足静态假设^[1], 当操作变量变化时,

主导变量能立即收敛到与之对应的稳态. 然而, 工业过程往往具有很强的动态特性. 例如, 管道中的材料运输可能会导致大时间延迟响应; 低反应速率的化学反应可能会减慢主导变量的收敛速度. 这些因素将影响静态模型的准确性. 为了解决这一问题, 动态建模在软测量领域得到了广泛研究.

根据结构, 动态软测量模型可分为3类: 递归模型、非线性滑动平均(nonlinear moving average, NM-

收稿日期: 2021–12–06; 录用日期: 2023–03–30.

†通信作者. E-mail: greenpre@163.com; Tel.: +86 15861683378.

本文责任编辑: 阳春华.

国家自然科学基金项目(61773182), 国家重点研发计划子课题项目(2018YFC1603705–03)资助.

Supported by the National Natural Science Foundation of China (61773182) and the Subtopics of National Key Research and Development Program of China (2018YFC1603705–03).

A)模型及基于动态特征的预测模型. 其中, 带外输入自回归模型 (nonlinear autoregressive with exogenous inputs model, NARX)^[2]是具有代表性的递归软测量模型, 其当前预测结果与以往多时刻的预测相关, 以此往复不断递归. 该结构的优点在于能设计合适的预测补偿策略, 当某时刻主导变量测量得到时, 能根据高效的算法对当前估计进行更新^[3]. 近年来, 一些具有递归结构的神经网络也在动态软测量中成功应用, 如: Xie等人^[4]基于门控循环单元构造了一种双通道递归神经网络, 精确预测了酯化反应中的熔体粘度指数; 针对长短期记忆模型(long short-term memory, LSTM)不考虑变量间相关性的问题, Yuan等人^[5]提出了基于注意力机制的LSTM软测量模型, 能自适应地挖掘与质量变量相关的特征变量, 提高了递归预测的准确率.

递归模型的优点在于可以记忆一些关键历史信息, 因此, 可以基于相对简单的模型结构对系统动态进行建模. 然而, 一些干扰信息(比如测量噪声、预测误差)也可能在递归结构中进行累积^[3]. NMA模型和基于动态特征的软测量模型摆脱了递归结构限制, 利用有限时长内的辅助变量信息进行预测, 适合于长期软测量任务^[6]. 其中, NMA模型是滑动平均模型的非线性扩展, 最早由Fortuna等人^[7]将其引入软测量领域. 在该模型基础上延伸出了许多改进的软测量模型, 如: 考虑到多率采样问题, Ando等人^[8]利用深度置信网络(deep belief network, DBN)构造了新的深层NMA模型, 该模型采用半监督训练方法进行优化, 充分利用了数据集集中的无标签样本信息, 并得到了比浅层NMA模型更高的预测精度; Chen等人^[9]在NMA模型基础上设计了一种自适应局部域划分方法, 能够将历史数据集自适应划分为多个局部域, 并确保基准窗口内的样本属于同一过程状态. 由于同时考虑了多种辅助变量在多时刻的采样信息, NMA模型的输入维度一般很大, 这使得模型的输入延迟项不应设置太大, 否则, 将由于模型参数过多而导致过拟合. 相比之下, 基于动态特征提取的模型同属于有限时间动态模型, 但通过精简的模型设计避开了这一问题. 此类模型首先从时间序列中提取出较为精炼的动态特征, 然后, 基于这些特征构建预测模型, 当特征提取方式设计得当时, 此类模型将具有更强的抗干扰能力. Chiplunkar和Huang^[10]建立了一种有监督慢特征分析模型, 同时, 从时序关联性以及预测目标关联性角度出发提取过程的动态特征, 相比普通慢特征分析, 该方法更有助于提升预测模型精度; 马勇等人^[11]提出了一种基于有限脉冲响应(finite impulse response, FIR)的动态软测量模型, 该模型利用FIR滤波来提取系统的动态特性, 通过将FIR模型与普通非线性模型串联来实现非线性动态关系的拟合. 在此基础上, Shang等人^[11]利用贝叶斯

框架实现了FIR模型参数的自动整定, 该方法能从数据中辨识出系统的时间滞后、惯性环节时间常数等动态系统参数.

综上, NMA模型和基于动态特征的预测模型具有较为鲜明的特点: 基于动态特征的模型具有较为精简的模型结构, 能够以此设计出时序边界较宽的动态软测量模型, 但是多数仅能描述线性的动态关系; NMA模型则与之相反, 能够描述非线性动态关系, 但受限于模型复杂度, 仅适用于无滞后、低噪声且收敛较快的过程. 本文结合两类模型的优点, 提出一种非线性、强抗干扰的软测量建模策略. 本文主要贡献有以下3方面: 1) 提出一种新的动态模型, 采用FIR滤波来描述工业过程中较简单的线性动态环节, 并滤除数据中的采样噪声, 并采用NMA模型来描述过程中相对复杂的非线性动态环节; 2) 提出一种白化层结构来连接FIR滤波器与NMA模型, 能够有效避免二者间的参数耦合问题, 提高所提模型的训练效率以及预测精度; 3) 给出了该模型结构下的梯度求解方式以及参数优化方法, 能同步优化NMA模型参数以及FIR参数.

2 符号定义与相关理论

2.1 序列的描述及其运算

采用小写粗体字母(如 \mathbf{a})来表示序列, 利用下标来表示序列中元素或片段(如 a_i 或 $\mathbf{a}_{i:j} = [a_i \ a_{i+1} \ \dots \ a_j]^T$). 定义 \mathbf{a}' 为序列 \mathbf{a} 的反向序列, 满足 $a'_t = a_{-t}$. 此外, 定义序列的有限片段均为列向量. 进一步, 定义卷积运算“*”如下:

$$[\mathbf{a} * \mathbf{b}]_t \triangleq \sum_{\tau \in \mathbb{Z}} a_{t-\tau} b_{\tau}, \quad (1)$$

在定义有限序列时会明确说明其有效范围, 范围外的元素都将视为0. 此外, 假定 \mathbf{a} 的有效范围为 $[t_0, t_1]$, 那么定义 \mathbf{a}' 的有效范围为 $[-t_1, -t_0]$; 假定 $\mathbf{c} = \mathbf{a} * \mathbf{b}$, 其中 \mathbf{b} 的有效区间为 $[\tau_0, \tau_1]$, 那么, 定义 \mathbf{c} 的有效范围为 $[t_0 + \tau_1, t_1 + \tau_0]$, 如图1所示.

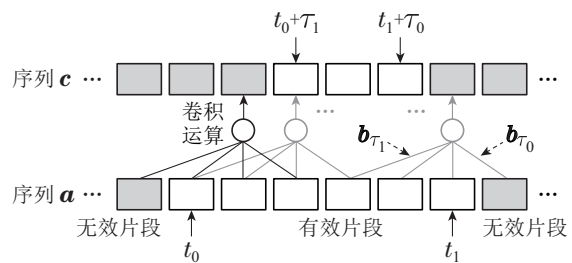


图1 卷积运算后有效范围的变化

Fig. 1 Change of effective range after convolution

利用自相关序列来描述平稳随机序列的时序特性. 假定 \mathbf{a} 是一个平稳随机序列, 在其有效范围内定义自相关序列为 \mathbf{r}^a , 其中 $\mathbf{r}_{\tau}^a \triangleq \text{Cov}(a_t, a_{t-\tau})$.

2.2 FIR滤波介绍

FIR滤波器是一种线性单输入单输出估计模型, 通过带噪声的测量信号来预测动态系统中的关键变量, 如图2所示. 其中: \mathbf{d} 表示线性动态环节的脉冲响应序列; \mathbf{y} 表示主导变量时间序列; \mathbf{s} 表示辅助变量序列, 三者间存在动态关系 $\mathbf{y} = \mathbf{s} * \mathbf{d}$. 此外, $\boldsymbol{\varepsilon}$ 表示测量噪声序列; $\mathbf{u} = \mathbf{s} + \boldsymbol{\varepsilon}$ 表示辅助变量量测序列.

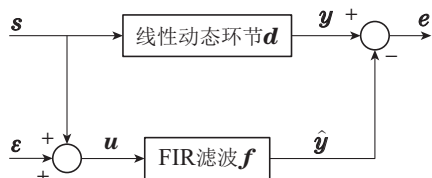


图2 FIR滤波问题示意图

Fig. 2 Schematic diagram of the FIR filtering problem

FIR滤波的目的是通过卷积操作 $\hat{\mathbf{y}} = \mathbf{u} * \mathbf{f}$ 得到相对精确的主导变量估计, 其中 \mathbf{f} 是FIR序列, 是一段有限参数序列. 设 \mathbf{f} 的有效范围为 $[0, l]$, 则卷积可简写为

$$\hat{y}_t = [\mathbf{u} * \mathbf{f}]_t = \mathbf{u}_{t-l:t}^T \mathbf{f}'_{-l:0}, \quad (2)$$

由于预测 \hat{y}_t 仅考虑有限时长内的信号量测, 简化了滤波参数 \mathbf{f} 的设计与分析. 比如, 当过程机理、噪声特性已知时, 以最小化预测方差 $E(e_t^2)$ 为指标设计滤波参数, 一般通过解Wiener-Hopf方程^[12]来获取最优参数. 其中 $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ 表示预测误差序列. 当机理未知, 而只给定一组带标签数据 $\{\mathbf{u}_t, \mathbf{y}_t\}_{0 \leq t \leq n}$ 时, 以最小化误差平方和 $\|\mathbf{e}_{1:n}\|_2^2$ 目标设计滤波器参数, 典型方法有最小二乘配置法.

2.3 NMA模型

NMA模型可视为FIR滤波的非线性扩展, 该模型同样用于解决图2所示的质量变量估计问题, 但无需假定 \mathbf{d} 为线性动态环节. 相比于FIR滤波, 该模型具有以下两个特点: 首先, 以多个辅助变量的时间片段作为输入, 即 $[\mathbf{u}_{t-l:t}^{(0)} \ \mathbf{u}_{t-l:t}^{(1)} \ \dots \ \mathbf{u}_{t-l:t}^{(m)}]$. 其中 $\mathbf{u}^{(i)}$ 表示第 i 个辅助变量的量测序列, 辅助变量数量为 $m + 1$. 此外, 该模型采用非线性回归模型(如: 神经网络、SVM(support vector machine))作为预测模型来拟合动态非线性关系. 记非线性预测模型的参数为 θ , 模型输入输出关系为 $g(\cdot; \theta)$, 则模型预测可表示为 $\hat{y}_t = g[\mathbf{u}_{t-l:t}^{(0)} \ \mathbf{u}_{t-l:t}^{(1)} \ \dots \ \mathbf{u}_{t-l:t}^{(m)}; \theta]$.

然而, 输入片段长度 l 的增大会导致NMA模型复杂度急剧上升. 以单隐层神经网络(neural network, NN)为例, 假定NN中包含 q 个隐含单元, 那么, 模型参数的数量将多达 $O(l, m, q) \triangleq mql + mq + ql + 3q + 1$ 个.

3 基于FIR滤波的NMA建模策略

3.1 问题描述以及模型概述

为了使模型考虑更长时段辅助变量信息而同时避免过拟合问题, 本文设计特殊模型结构来降低传统NMA模型的复杂度.

考虑到许多工业过程可以看作为线性与非线性动态环节的串联组合, 如图3所示. 本文考虑将FIR滤波与NMA模型进行互补来实现精确建模. 首先, 利用FIR滤波来滤除数据中的量测噪声, 并辨识出一些简单的线性动态环节(如时滞或惯性环节). 随后, 采用NMA模型来处理建模对象中复杂的非线性动态关系, 模型结构和其中参数如图4所示. 为加以区分, 记FIR滤波和NMA模型的输入片段长度分别为 l 和 k . 图4中卷积结果 $\mathbf{h}^{(i)}$ 和信息矩阵 $X^{(i)} \in \mathbb{R}^{n \times (k+1)}$ 是模型在处理辅助变量 $\mathbf{u}^{(i)}$ 时得到的中间变量, t 时刻的信息向量为 $X_{t,:}^{(i)} = [x_t^{(i,0)} \ x_t^{(i,1)} \ \dots \ x_t^{(i,k)}]$.

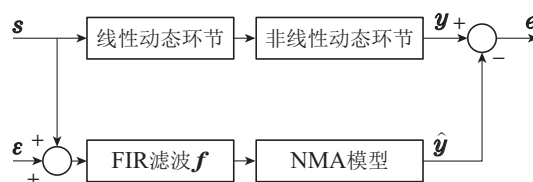


图3 建模问题描述

Fig. 3 Description diagram of the modeling problem

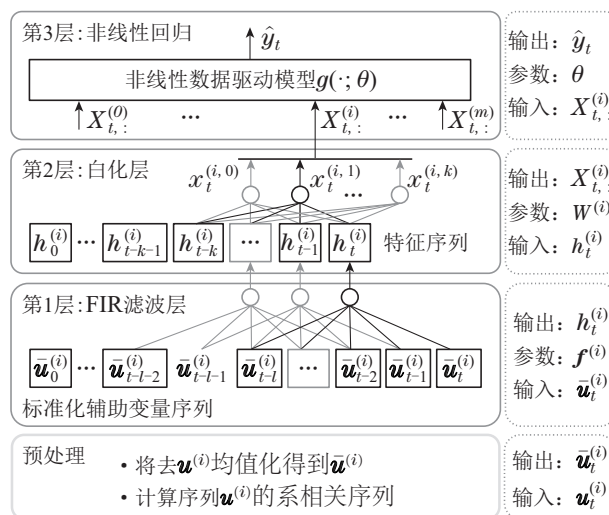


图4 模型结构示意图

Fig. 4 Model structure diagram

模型分为3层: FIR滤波层、白化层和非线性回归模型. 其中, 非线性回归模型构与白化层构成了一个输入信息平衡的NMA模型. 在该模型中, 白化层的设计主要用于解决模型优化时参数间的耦合问题^[13]. 此外, 该模型为每个辅助变量 $\mathbf{u}^{(i)}$ 都设置了各自的FIR滤波序列 $\mathbf{f}^{(i)}$ 和白化矩阵 $W^{(i)}$, 下文以辅助变量 $\mathbf{u}^{(i)}$ 为例介绍辅助变量的滤波及白化过程. 为了书写简便,

在不加说明情况下 $\mathbf{u}, \mathbf{h}, X$ 及 $\mathbf{x}^{(j)}$ 将分别表示 $\mathbf{u}^{(i)}, \mathbf{h}^{(i)}, X^{(i)}$ 和 $\mathbf{x}^{(i,j)}$, W 和 \mathbf{f} 同理。

3.2 模型预测过程

1) 预处理. 首先, 对辅助变量进行标准化, 使其均值为零, 如式(3)所示:

$$\bar{u}_t = u_t - \frac{1}{n+1} \sum_{\tau=0}^n u_\tau, \quad (3)$$

此外, 近似估计出该辅助变量的自相关序列. 在本文方法中, 只需估计出范围 $[-l-k, l+k]$ 以内 \hat{r}^u 的取值即可, 如式(4)所示:

$$\hat{r}_\tau^u = \frac{1}{n-|\tau|} \sum_{\tau=0}^{n-|\tau|} \bar{u}_t \bar{u}_{t+|\tau|}, \quad |\tau| \leq l+k. \quad (4)$$

2) 第1层(FIR滤波层). 该层利用 FIR 滤波实现噪声滤除, 以及简单动态环节的建模, 如式(5)所示:

$$\mathbf{h} = \bar{\mathbf{u}} * \mathbf{f}, \quad (5)$$

此外, 计算出序列 \mathbf{h} 的自相关序列, 如式(6)所示:

$$\hat{r}^h = \hat{r}^u * \mathbf{f} * \mathbf{f}', \quad (6)$$

由于 $\bar{\mathbf{u}}, \mathbf{f}$ 和 \hat{r}^u 都是有限时间序列, 因此, 滤波结果 \mathbf{h} 的有效范围为 $[l, n]$, 而 \hat{r}^h 的有效范围为 $[-k, k]$.

3) 第2层(变量白化). 白化的目的是从时序 \mathbf{h} 中分离出 $k+1$ 个线性无关的序列 $\mathbf{x}^{(j)}$. 分两步来实现这一目的: 首先, 计算出白化矩阵为 $W \in \mathbb{R}^{(k+1) \times (k+1)}$. 然后, 基于矩阵 W , 利用卷积运算计算出 $\mathbf{x}^{(j)}$. 式(7)显示了白化矩阵的计算过程, 即

$$\begin{cases} C_{a,b} = \hat{r}_{|a-b|}^h, & a, b = 0, 1, \dots, k, \\ W = (C + \sigma I)^{-\frac{1}{2}}, \end{cases} \quad (7a) \quad (7b)$$

其中: C 为 Toeplitz 矩阵^[14], 若假定 $\hat{r}_{0:k}^h = r_{0:k}^h$, 那么 $C = D[\mathbf{h}_{t-k:t}]$; σ 是为防止数值计算病态而设置的常数, 通常设置为很小的数, 如 $\sigma = 10^{-8}$. 接下来, 定义 $k+1$ 个有效范围为 $[0, k]$ 的有限时间序列 $\mathbf{w}^{(j)}$, 如式(8)所示:

$$\mathbf{w}_{0:k}^{(j)} = W_{:,j}, \quad j = 0, 1, \dots, k, \quad (8)$$

于是, 通过式(9)便能实现白化操作, 即

$$\mathbf{x}^{(j)} = \mathbf{h} * \mathbf{w}^{(j)}, \quad X = [\mathbf{x}^{(0)} \quad \mathbf{x}^{(1)} \quad \dots \quad \mathbf{x}^{(k)}], \quad (9)$$

在模型训练过程中, 变量 W 会随着迭代而不断更新, 如式(8)–(9)所示. 当模型训练完成后, W 将视为常量固定不变. 此外, 前文提到“白化层与非线性回归模型构成了一个输入信息平衡的 NMA 模型”. 这是因为当 $W = I$ 时, 恰好 $X_{t,:} = \mathbf{h}_{t-k:t}$ 成立, 与 NMA 模型的时序扩展一致. 而通过式(7)来决定 W 则能进一步平衡 X_t 中的信息, 因为, 当 σ 可忽略且数据量 n 足够大时, $\text{Cov}(X_{t,:})$ 将依概率趋于单位阵 I .

4) 第3层(非线性回归). 与 NMA 模型的逻辑相似, 采用单隐层神经网络模型来描述建模对象的非线性

动态特性, $\hat{\mathbf{y}}_t = g[X_{t,:}^{(0)} \quad X_{t,:}^{(1)} \quad \dots \quad X_{t,:}^{(k)}; \theta]$. 许多文献已介绍了 NN 的原理及逼近特性, 限于篇幅不再给出相关介绍. 此外, 由于序列 \mathbf{h} 及 $\mathbf{w}^{(j)}$ 的有效范围分别为 $[l, n]$ 和 $[0, k]$, 因此, 在预测序列中只有片段 $\hat{\mathbf{y}}_{l+k:n}$ 是有效预测.

以上是模型的总体介绍, 下面对白化层合理性进行解释. 假定辅助变量个数 $m = 1$, 在不设置白化层时预测算法为 $\hat{\mathbf{y}}_t = g([\bar{\mathbf{u}} * \mathbf{f}]_t \quad \dots \quad [\bar{\mathbf{u}} * \mathbf{f}]_{t-k}; \theta)$. 而这种形式存在很强的参数耦合现象, 因为滤波器参数与 NN 中的连接权重存在直接相乘关系, 会提升模型训练的困难^[13]. 而设置了白化层后, 模型算法则变为 $\hat{\mathbf{y}}_t = g([h_{t-k} \quad h_{t-k+1} \quad \dots \quad h_t] W; \theta)$. 此时, 滤波器参数将不再影响第3层输入的一阶矩、二阶矩统计量^[15], 第3层模型也无需考虑输入的分布情况, 由此, 缓解了层与层间的耦合现象.

综上, 相比于普通 NMA 模型, 所提模型具有两个特点: 1) 参数数量更加少. 在对主导变量 y_t 进行预测时, 所提模型考虑了时间范围 $[t-l-k, t]$ 内的辅助变量信息, 所包含参数数量为 $O(k, m, q) + ml$. 而在同样条件下, 普通 NMA 模型需要 $O(l+k, m, q)$ 个参数; 2) 所提模型更适合解决图3所示的建模问题, 这一点将在第4节和第5节中进行验证.

3.3 参数梯度求解

本文以最小化预测误差 $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ 为目标, 对各层参数进行整定. 目标函数如式(10)所示:

$$J = \|\mathbf{e}_{l+k:n}\|_2^2 + \alpha \sum_{j=0}^k \|D \mathbf{f}_{0:l}^{(j)}\|_2^2, \quad (10)$$

其中: $D \in \mathbb{R}^{(l+1) \times (l+1)}$ 是一个对角元素为 -1 的约当阵; 目标函数 J 可视为误差平方和与平滑正则项^[16] 的加权相加, 其中正则项使得 FIR 序列在模型训练时尽可能平滑; α 是正则项系数, 用于平衡模型的拟合能力与泛化性能. 采用 Adam 算法^[17] 来优化各层参数, 本小节重点介绍白化层与 FIR 滤波层的梯度求解. 图5展示了模型中各变量、常量及参数间的关系, 也展示了模型在优化时前向、后向的运算过程. 以下利用链式法则对梯度进行求解.

1) 第3层梯度求解. 第3层是一个标准 NN 模型, 可以利用反向传播(back propagation, BP)^[18] 算法得到其中参数的梯度 $\frac{\partial J}{\partial \theta}$, 以及信息矩阵的梯度 $\frac{\partial J}{\partial X}$. 需要说明, 在序列 $\mathbf{x}^{(j)}$ 中, 只有片段 $\mathbf{x}_{l+k:n}^{(j)}$ 有关, 因此, 只有当 $l+k \leq t \leq n$ 时, $\frac{\partial J}{\partial x_t^{(j)}} = \frac{\partial J}{\partial X_{t,j}}$, 否则, $\frac{\partial J}{\partial x_t^{(j)}} = 0$.

2) 第2层梯度求解. 该层中将利用反向求导方法, 根据 $\frac{\partial J}{\partial X}$ 分别求出梯度 $\frac{\partial J}{\partial \mathbf{h}}$ 和 $\frac{\partial J}{\partial \mathbf{r}^h}$. 首先, 根据式(9)可以得到 $\frac{\partial J}{\partial \mathbf{h}}$ 和 $\frac{\partial J}{\partial W}$, 如式(11)所示. 其中 $j = 0, 1, \dots, k$.

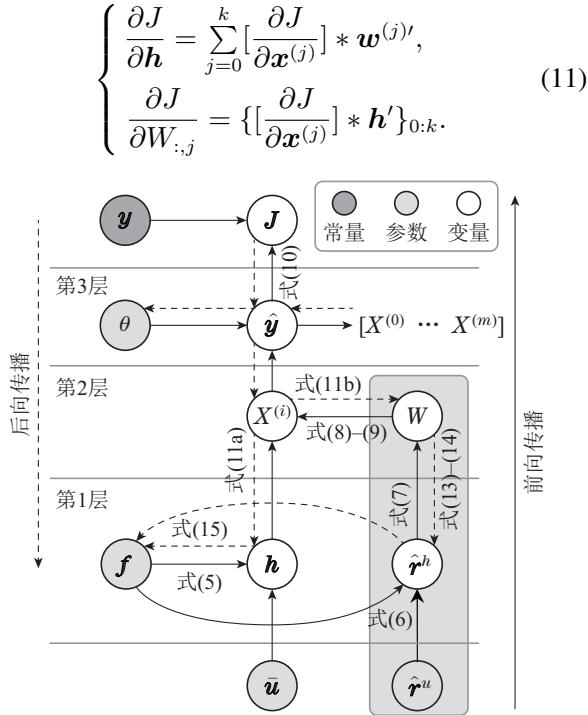


图5 前向后向计算示意图

Fig. 5 Overview of the forward and back propagation

进一步, 根据 $\frac{\partial J}{\partial W}$ 求解梯度 $\frac{\partial J}{\partial \hat{r}^h}$. 首先, 通过特征分解将 $C + \sigma I$ 分解成式(12)形式. 其中 λ_j 和 $K_{:,j}$ 分别为 $C + \sigma I$ 的特征值与特征向量, 即

$$C = V \text{diag}\{\lambda_{0:k}\} V^T, \quad (12)$$

由此, 可得到式(13), 其中“ \circ ”表示矩阵元素相乘, 即

$$\begin{cases} \frac{\partial J}{\partial C} + \frac{\partial J}{\partial C^T} = -V [(R + R^T) \circ K] V^T, \\ K_{a,b} \triangleq \frac{1}{\sqrt{\lambda_a \lambda_b} + \sqrt{\lambda_b \lambda_a}}; R \triangleq V^T \frac{\partial J}{\partial W} V. \end{cases} \quad (13)$$

进一步, 再根据式(7a)可以最终得到式(14), 其中 $\tau = 0, 1, \dots, k$.

$$\frac{\partial J}{\partial \hat{r}_\tau^h} = \sum_{|a-b|=\tau} \frac{\partial J}{\partial C_{a,b}}, \quad (14)$$

需要说明, 在序列 r^h 中, 仅有片段 $r_{0:k}^h$ 与 J 有关, 因此, 当 $\tau > k$ 或 $\tau < 0$ 时, $\frac{\partial J}{\partial \hat{r}_\tau^h} = 0$.

3) 第1层梯度求解. 根据 $\frac{\partial J}{\partial h}$ 和 $\frac{\partial J}{\partial r^h}$ 求解FIR参数的梯度. 根据式(5)–(6)可最终求得 $\frac{\partial J}{\partial f}$, 如式(15)所示:

$$\begin{aligned} \frac{\partial J}{\partial f_{0:l}} &= \{\hat{r}^h * f * [\frac{\partial J}{\partial \hat{r}^h} + (\frac{\partial J}{\partial \hat{r}^h})']\}_{0:l} + \\ &[\frac{\partial J}{\partial h} * \bar{u}']_{0:l} + 2\alpha DD^T f_{0:l}. \end{aligned} \quad (15)$$

最终, 经过以上3层梯度运算后便可求得所有模型参数的梯度, 再结合Adam算法^[17]既能完成模型参数的同步优化. 需说明, 模型的白化层是针对优化过程

中参数的耦合问题而设计的解耦方案, 主要用于提高模型的训练效率, 当模型训练完成后, 白化层参数 $W^{(i)}$ 将视作为常量参与运算. 因此, 图5中灰色框内的运算只有在模型优化过程中才会执行.

4 数值仿真实验

本节通过数值仿真验证所提方法能有效处理第3.1节所提问题. 为了模拟实际过程中的时间滞后与观测噪声问题, 构造如图6所示的建模问题.

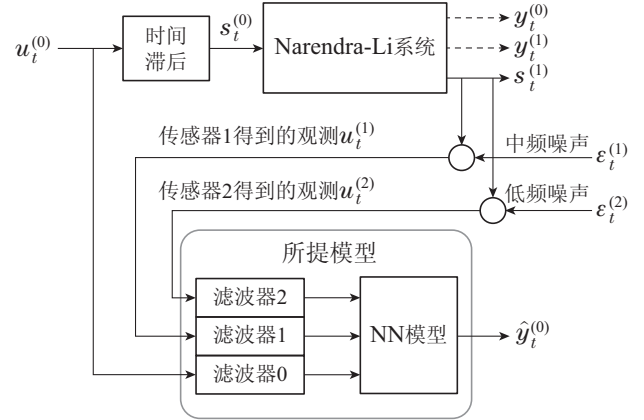


图6 数值仿真实验设计

Fig. 6 Diagram of the numerical experiment

如图6, 建模对象可视作为时滞环节与Narendra-Li系统^[19]的串联. 系统中包含两个辅助变量: $s^{(0)}$ 和 $s^{(1)}$, 其中: $s^{(0)}$ 只受操作变量 $u^{(0)}$ 影响, 两者间存在 d 时刻时间滞后, 即 $s_t^{(0)} = u_{t-d}^{(0)}$; 变量 $s^{(1)}$ 是系统的观测变量, 与系统状态变量 $y^{(0)}$, $y^{(1)}$ 有关. 上述变量的动态关系如式(16)所示:

$$\begin{cases} y_{t+1}^{(0)} = (1.8 + \frac{y_t^{(0)}}{1 + y_t^{(0)2}}) \sin y_t^{(1)}, \\ y_{t+1}^{(1)} = y_t^{(1)} \cos y_t^{(1)} + y_t^{(0)} e^{-2(y_t^{(0)2} + y_t^{(1)2})} + \\ \frac{2s_t^{(0)3}}{2 + 2s_t^{(0)2} + \cos(y_t^{(0)} + y_t^{(1)})}, \\ s_t^{(1)} = \frac{2y_t^{(0)}}{2 + \sin y_t^{(1)}} + \frac{2y_t^{(1)}}{2 + \sin y_t^{(0)}}, \end{cases} \quad (16)$$

此外, 假定采用两种传感器对可观测变量 $s^{(1)}$ 进行测量, 其观测结果 $u^{(1)}$ 和 $u^{(2)}$ 分别受到两种观测噪声 $\epsilon^{(1)}$ 和 $\epsilon^{(2)}$ 的干扰, 噪声的统计特性如式(17)所示:

$$\begin{cases} [\epsilon_t^{(1)}, \epsilon_t^{(2)}] \sim \mathcal{N}(0, 0.5I), \\ r_\tau^{\epsilon^{(1)}} = 0.5 \exp(-0.999|\tau|), \\ r_\tau^{\epsilon^{(2)}} = 0.5 \exp(-0.95|\tau|) \cos \tau. \end{cases} \quad (17)$$

该实验中, 设采样周期为1s, 令滞后时间 $d = 8s$, 以状态量 $y^{(0)}$ 作为待预测主导变量, 并令 $u^{(0)}$, $u^{(1)}$ 和 $u^{(2)}$ 为辅助变量, 其余变量(包括 d 在内)在训练及测试过程中都不可知. 令 $u_t^{(0)} \sim \mathcal{N}(0, 1)$ 生成一组长度为

7000的训练数据集,再令 $u_t^{(0)} = \sin(0.2\pi t) + \sin(0.08\pi t)$ 生成一组长度为300的测试数据序列,下面对实验结果进行描述.

4.1 建模有效性验证

通过数值仿真,本小节主要论证所提模型的滤波层能提取出有效精炼的特征信息,基于这些信息能建立出更精确的预测模型.首先,基于验证集(从训练集中分离得到)确定所提模型的超参数: $l = 10, k = 4$, NN隐层单元数为50, NN激活函数为“tanh”、迭代次数为2000.然后,根据训练集对所提模型进行优化,算法基于“Numpy 1.19.4”实现.训练完成后, FIR参数如图7所示.

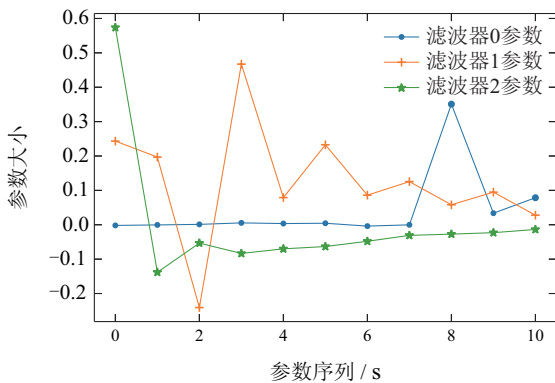


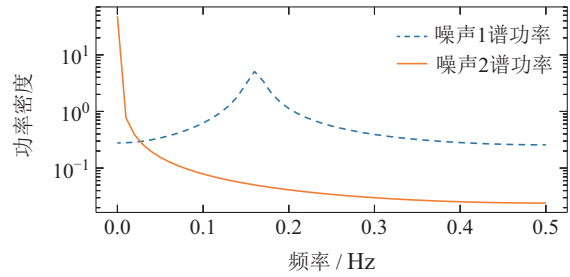
图7 FIR参数展示图

Fig. 7 Diagram of FIR parameters

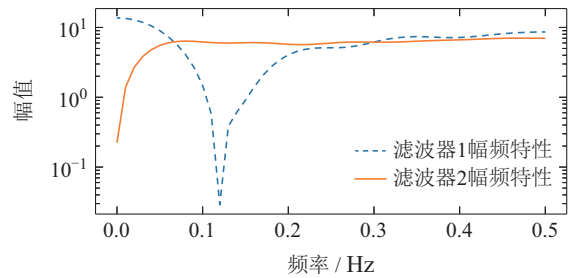
显然, $f_s^{(0)}$ 远大于序列 $f^{(0)}$ 中的其它参数,说明模型正确辨识了系统中的滞后环节.此外,滤波器1,2的特性难以从时域上进行分析,因此,对其进行傅里叶变换并从频域的角度进行分析,如图8所示.图8(a)展示了测量噪声的谱功率特性,可以看出,噪声1的功率则主要集中于 $[0, 0.05]$ Hz范围内,属于低频噪声,而噪声2的功率主要集中于 $[0.1, 0.3]$ Hz范围内.经过模型训练后, FIR滤波器能有效抑制这些噪声.如图8(a)所示,滤波器1被训练为一个高通滤波器,能够有效抑制 $[0, 0.05]$ Hz频段的噪声并能较好保留其他频段的信息,滤波器2则被训练为一个带阻滤波器,主要抑制 $[0.1, 0.3]$ Hz频段内的信息.

接下来,对模型预测精度进行验证.在两种建模策略下建立动态软测量模型:1)直接基于NMA结构建立预测模型;2)基于FIR滤波器建立NMA模型(FIR-NMA),该策略下首先利用图7所示滤波器对数据进行滤波,然后建立NMA模型.在两种建模策略下建立3种数据驱动模型从而显示卷积滤波的有效性,分别为深度置信网络(DBN)、极限梯度提升树(XGBoost)及NN模型.采用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)、最大绝对误差(max absolute error, ME)来评价各模型在

测试集中的预测性能,实验结果如表1所示.以下介绍了各模型实现细节.



(a) 噪声功率谱



(b) 滤波器幅频特性

图8 滤波器、噪声的频域特性

Fig. 8 Frequency domain characteristics of filters and noise

表1 各模型预测结果

Table 1 Predicting result of each model

建模策略	模型名称	超参数	预测误差($\times 10$)
NMA结构	NN	$k = 10$	2.08/1.64/8.01
	DBN	$k = 11$	1.63/1.29/4.96
	XGBoost	$k = 7$	2.43/1.93/7.13
FIR-NMA结构	NN		1.45/1.13/4.99
	DBN	$l = 10$ $k = 4$	1.48/1.14/5.74
	XGBoost		1.88/1.45/6.77

在基于NMA策略训练模型时,各模型的最优阶次都不尽相同,因此令预测模型的输入时序边界为 $k = 5, 6, \dots, 14$,并分别建立预测模型,并利用验证样本集确定最优 k 值.DBN包含两个隐含层,两者均包含10个隐含单元,采用开源代码¹和“Tensorflow 2.4.1”实现;XGBoost模型使用“XGBoost 1.3.0”实现,最大深度设为5、集成数量设为2000,学习率设为0.1.NN模型的超参数配置与第4.1节描述一致.此外,为了保证NMA结构下各模型的训练效果,预先对数据进行了白化操作.

在表1中, FIR-NAM结构下建立的NN预测模型为本文方法,预测指标中所包含的内容为 $RMSE \times 10 / MSE \times 10 / ME \times 10$.如表1所示,本文所提模型在各指标上均取得了最好的预测效果.此外,在FIR滤波的作用下,各模型也取得了更好的预测性能,这一结果验

¹<https://github.com/fuzimaoxinan/Tensorflow-Deep-Neural-Networks>

证了FIR滤波的有效性,同时,也验证了图3模型设计的合理性.

4.2 白化层有效性验证

为了验证白化操作的有效性,记录了迭代优化中所提模型的误差下降过程.基于相同训练、测试样本,分别在普通标准化操作^[20]及无任何操作条件下训练所提模型,并记录模型的误差下降过程作为对比.误差收敛过程如图9所示,其中:图9(a)描述了模型针对训练样本的误差收敛过程;图9(b)则描述了针对测试样本的误差收敛过程.

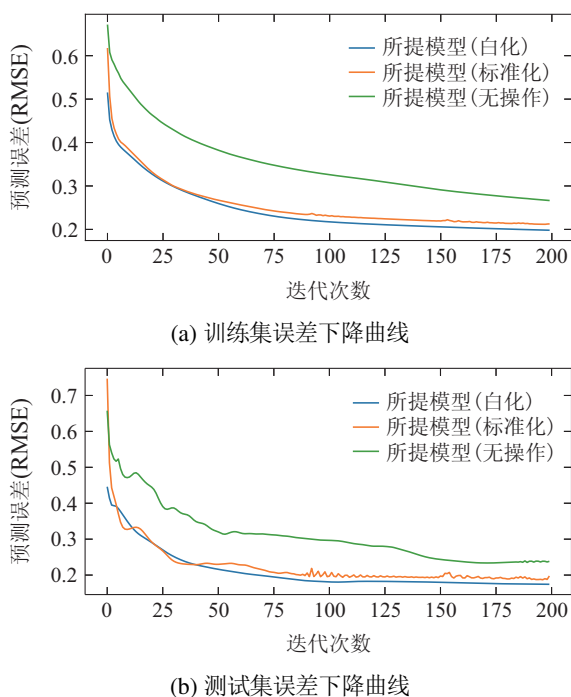


Fig. 9 Training loss curves

可以看出,经过标准化操作后优化的收敛速度得到了较大提升.而通过白化操作,模型的训练不仅具有较快的收敛速度,且收敛后模型也具有更好的预测性能.这是因为白化操作规范了NMA模型的输入分布,使得FIR滤波层得到了更好的优化.为了论证这一观点,对FIR滤波层的训练结果进行了观察.

首先,将带测试集中噪声集辅助变序列输入FIR滤波层,经过白化或标准化后统计其主元方差.然后将原辅助变量序列送入FIR滤波层,经过白化或标准化后统计输出信号对上述主元的贡献.若贡献较大则说明滤波器较好清除了噪声信号并相对完整地保留了有用信息,反之亦然.实验结果如图10所示.

可以看出,在不经任何操作情况下,卷积层的输出信息大多集中于一个主元,且主元的最大方差甚至达到了246.7,这一分布特点容易造成梯度爆炸现象,使得FIR滤波层优化不充分.针对这一问题,层标准化操作可以减缓该问题,从图10(b)可以看出,在层归一化

操作下,最大主元方差降至5.3,并有效提高了滤波效果.然而,归一化操作仅考虑各变量自身的方差,而不考虑变量间的相关性,因此,大多数信息依然集中于一个主元.白化操作可以进一步去除变量间的线性相关性,从图10(c)可以看出,在该操作下,最大主元方差降至2.35,辅助变量信息也被更加均匀地分配给了各个主元,使训练过程能更精细地调整卷积滤波层参数.

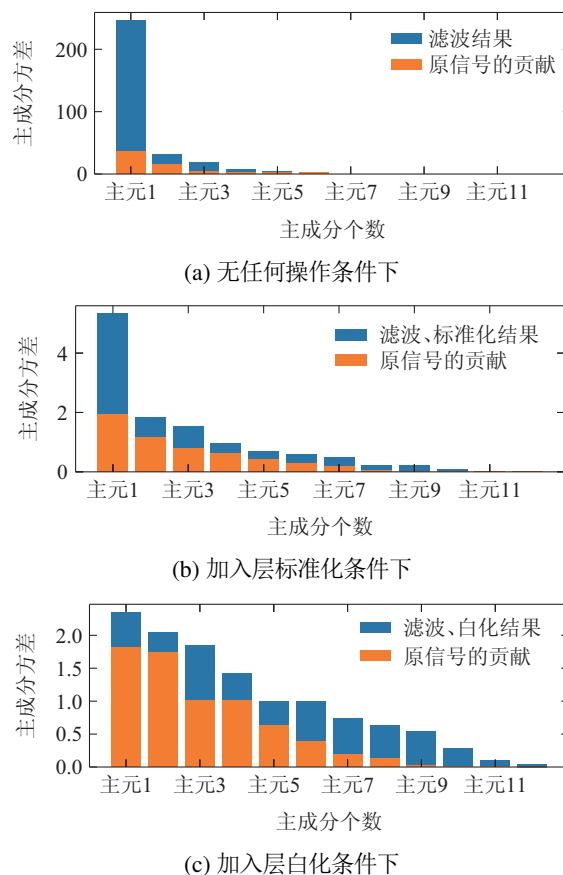


图10 不同条件下FIR的滤波效果

Fig. 10 Filtering effects of FIR under different conditions

5 硫回收建模实验

硫回收(sulfur recovery unit, SRU)装置负责从酸性气体中提取硫化物,从而减轻工业废气对空气质量的影响,其工艺流程的简化图如图11所示.

如图,SRU装置主要吸收两种气体:第1种气体富含 H_2S ,称为MEA气体,来自气体洗涤装置;第2种称为SWS气体,富含 H_2S 和 NH_3 ,来自酸性水汽提(sour water stripping, SWS)装置.两种气体通入加热器B106与B103升温,并于焚烧炉F101中与空气反应.其中涉及两种反应:部分 H_2S 与氧气燃烧,生成 SO_2 与水; NH_3 与氧气完全燃烧,生成氮气与水.随后,反应产物依次通入水冷凝器(E101, E102和E103),以及催化反应器(R101和R102),这些单元通过反应吸收 H_2S 和 SO_2 并生成水和纯硫.从冷凝器E103出来的尾气,经D104尾气分液罐后进入尾气处理单元,回收硫

化物循环处理,以提高硫回收率并确保排放气体达到环保要求.更详细过程介绍见文献^[21].

可以看出,如果F101中通入空气太少,尾气中的H₂S就会过多,反之,SO₂则会过多.为了使反应尽可

能充分,混合气体中H₂S和SO₂的摩尔浓度比例应该控制为2:1,为此,需要对尾气中H₂S与SO₂浓度进行实时监测.然而,酸性气体通常会影检测仪器寿命,当硬件传感器维修时需要用软测量来完成预测任务.

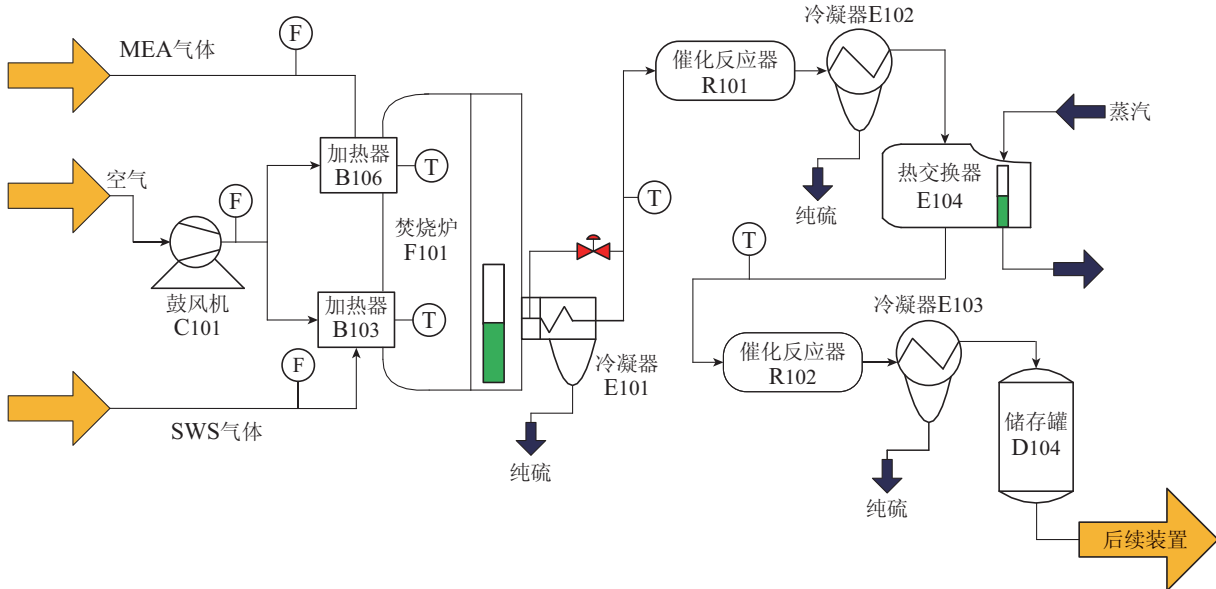


图 11 硫回收过程工艺流程图

Fig. 11 The simplified scheme of the SRU

SRU数据一共包含10081个数据,其采样周期为1 min一个样本.本文取前7000个样本作为训练样本,其后3081个样本作为测试样本.每个样本都包含5个辅助变量及2个主导变量,其描述见表2.在对SRU进行建模前,一般需要根据过程知识对数据进行滤波处理^[21],而本文模型则利用原始数据同时对FIR滤波器以及预测模型进行了优化.基于SRU建模实验对模型的精确性进行了验证.

表 2 变量介绍

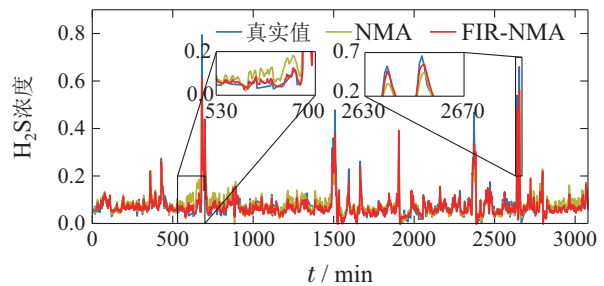
Table 2 Variable introduction

变量	变量名称	变量	变量名称
$u^{(0)}$	MEA_GAS气体流	$u^{(4)}$	SWS区域空气流
$u^{(1)}$	AIR_MEA空气流	$y^{(0)}$	H ₂ S浓度
$u^{(2)}$	AIR_MEA2二次空气流	$y^{(1)}$	SO ₂ 浓度
$u^{(3)}$	SWS区域气体流	-	-

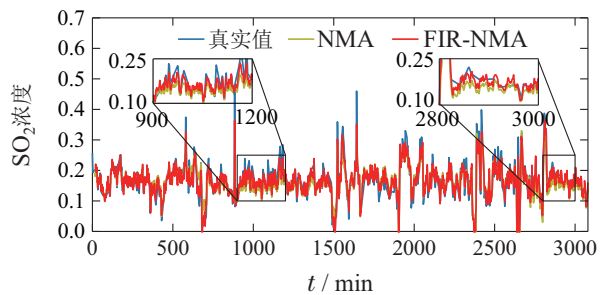
首先,以NN为例分别构造本文所提模型和普通NMA模型.经过模型训练后,两者在测试集中的预测效果如图12所示.模型的超参数设置如下:令NN的隐层单元数为20;在建立NMA策略构建模型时,令模型输入时序边界 $k = 15$,在建模前对数据进行了白化处理;在构造本文所提模型时,令 $l = 10, k = 5$,正则项系数 $\alpha = 1$.

如图12(a)所示, H₂S浓度在[2630, 2670] min内出

现了两次峰值,在这两次峰值上FIR-NMA模型预测精度均高于NMA模型.此外,当NMA策略下模型预测逐渐偏离真实值时FIR-NMA模型则仍能保持精度,如图12(b)所示.由此可见本文方法有效提高了传统NMA模型的预测精度.



(a) H₂S浓度预测结果



(b) SO₂浓度预测结果

图 12 软测量模型的测试结果

Fig. 12 Prediction result of testing data

随后,为了验证FIR滤波器在特征提取方面的有效性,采用第4.1节的实验方法,在多种模型结构框架下构建不同预测模型.在建模策略方面,考虑了常用的特征提取、降维方法来降低传统NMA模型的高复杂度,考虑了主元分析(principle component analysis, PC

A)和独立成分分析(independent component analysis, ICA).此外,比较了普通基于FIR滤波的动态模型结构^[20],由于该结构可看作为FIR滤波与静态模型的串联组合,将其记作为FIR-S.在预测模型方面,考虑了第4.1节所介绍的3种模型.结果如表3所示.

表3 各建模策略下的建模预测结果

Table 3 The modeling prediction results under each modeling strategy

建模策略	各预测模型对H ₂ S浓度预测结果 ^a			各预测模型对SO ₂ 浓度预测结果 ^a		
	NN	DBN	XGBoost	NN	DBN	XGBoost
NMA	2.53/1.89/2.89	2.74/1.61/4.78	3.28/1.91/5.40	2.75/2.13/2.15	2.40/1.82/2.14	3.90/3.01/2.01
NMA-PCA	2.44/1.56/3.76	2.64/1.62/4.85	3.29/1.97/5.14	2.38/1.82/1.71	2.46/1.92/1.38	3.95/3.07/1.97
NMA-ICA	2.74/1.66/4.28	2.59/1.62/3.73	3.19/1.72/5.42	2.40/1.83/1.31	2.63/1.99/1.79	2.86/2.06/1.97
FIR-S	2.39/1.66/3.20	2.59/1.58/3.05	2.80/1.74/2.48	2.18/1.71/1.28	2.28/1.81/1.10	2.64/2.04/1.55
FIR-NMA	2.10/1.45/2.38	2.23/1.43/2.48	2.57/1.68/1.99	2.11/1.65/1.24	2.19/1.70/1.14	2.64/2.05/1.56

^a表中结果为: RMSE×100/MSE×100/ME×10.

表3中,PCA,ICA能从辅助变量分布的角度提取最具代表性的特征,在多数情况下能提升NMA模型的预测精度.而相比之下,FIR-S策略以有监督方式实现滤波以及特征提取,能实现对原始数据的自适应滤波,因此,能取得更好的预测效果.本文方法在FIR-S基础上进行改进,使其具有更好的动态非线性描述能力.如表3所示,无论在预测H₂S浓度方面还是预测SO₂方面,本文方法在多个指标上均获得了较好的测试结果.这一结果同时验证了FIR滤波的有效性以及本文模型设计的合理性.

6 结论

本文工作主要集中于两点,首先,通过将FIR滤波器与NMA模型互补,提出了一种新的软测量模型结构;然后,设计了合适的层级模型结构来减轻参数耦合现象,并根据链式求导法则推导出了所提模型的优化方式.根据实验结果可以得到以下3个结论:

1) 所提方法能够较好辨识出系统中的滞后环节,并且有效滤除辅助变量中的观测噪声,能够解决带长时滞、强观测噪声的建模问题;

2) 训练完成后,本文模型的FIR滤波能输出有效的建模特征,并且基于这些特征,能构建出相比于NMA更精确的预测模型;

3) 采用白化层来连接FIR滤波器和NMA模型能有效提高模型的训练效率以及预测精度,验证了本文模型设计的合理性.

最后,本文研究工作说明较为复杂的动态模型可以与成熟的线性滤波方法进行互补,在结构设计得当时能进一步提高软测量模型的预测性能.

参考文献:

- [1] MA Yong, HUANG Dexian, JIN Yihui. Discussion about dynamic soft-sensing modeling. *Journal of Chemical Industry & Engineering*, 2005, 56(8): 1516 – 1519.
(马勇, 黄德先, 金以慧. 动态软测量建模方法初探. 化工学报, 2005, 56(8): 1516 – 1519.)
- [2] LIN T, HORNE B G, TINO P, et al. Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks*, 1996, 7(6): 1329 – 1338.
- [3] XIONG Weili, SUN Wenxin, MA Junxia. Autocorrelation constrained NARX dynamic soft sensing model. *Control and Decision*, 2020, 35(4): 51 – 57.
(熊伟丽, 孙文心, 马君霞. 带自相关约束的NARX动态软测量模型. 控制与决策, 2020, 35(4): 51 – 57.)
- [4] XIE R, HAO K, HUANG B, et al. Data-driven modeling based on two-stream λ gated recurrent unit network with soft sensor application. *IEEE Transactions on Industrial Electronics*, 2020, 67(8): 7034 – 7043.
- [5] YUAN X, LI L, YURI A W S, et al. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Transactions on Industrial Electronics*, 2021, 68(5): 4404 – 4414.
- [6] FORTUNA L, GRAZIANI S, RIZZO A, et al. *Soft Sensor for Monitoring and Control of Industrial Processes, chapter5: Choice of the Model Structure*. London: Springer-Verlay, 2007.
- [7] FORTUNA L, RIZZO A, SINATRA M, et al. Soft analyzer for a sulfur recovery unit. *Control Engineering Practice*, 2003, 11(2003): 1491 – 1500.
- [8] ANDO B, GRAZIANI S, XIBILIA M G. Low-order nonlinear finite-impulse response soft sensors for ionic electroactive actuators based on deep learning. *IEEE Transactions on Instrumentation and Measurement*, 2019, 68(5): 1637 – 1646.
- [9] CHEN X, MAO Z, JIA R, et al. Ensemble regularized local finite impulse response models and soft sensor application in nonlinear dynamic industrial processes. *Applied Soft Computing Journal*, 2019, 85(2019): 105806.
- [10] CHIPLUNKAR R, HUANG B. Siamese neural network-based supervised slow feature extraction for soft sensor application. *IEEE Transactions on Industrial Electronics*, 2021, 68(9): 8953 – 8962.

- [11] SHANG C, GAO X, YANG F, et al. Novel bayesian framework for dynamic soft sensor based on support vector machine with finite impulse response. *IEEE Transactions on Control Systems Technology*, 2014, 22(4): 1550 – 1557.
- [12] SIMON H. *Adaptive Filter Theory, Fifth Edition, Chapter 2.4: Wiener-Hopf Equations*. New Jersey: Pearson Education, 2013.
- [13] GOODFELLOW I, BENGIO Y, COURVILLE A. *Deep Learning, Chapter 8.7: Optimization Strategies and Meta-Algorithms*. Massachusetts: MIT Press, 2016.
- [14] ROBERT M. Toeplitz and circulant matrices: A review. *Foundations & Trends in Communications & Information Theory*, 2006, 2(3): 155 – 239.
- [15] DESJARDINS G, SIMONYAN K, PASCANU R. Natural neural networks. *Advances in Neural Information Processing Systems 28*. New York, USA: Curran Associates Inc., 2015.
- [16] SHANG C, HUANG X, JOHAN A K S, et al. Enhancing dynamic soft sensors based on DPLS: A temporal smoothness regularization approach. *Journal of Process Control*, 2015, 28(2015): 17 – 26.
- [17] KINGMA D P, BA J. Adam: A method for stochastic optimization. *ArXiv Preprint*, 2014: ArXiv:1412.6980.
- [18] BAYDIN A G, PEARLMUTTER B A, RADUL A A, et al. Automatic differentiation in machine learning: A survey. *Journal of Machine Learning Research*, 2018, 18(2018): 1 – 43.
- [19] NARENDRA K S. *Neural Networks in Control Systems, Chapter 11: Mathematical Perspectives on Neural Networks*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1996.
- [20] XIONG W, SUN W. Dynamic soft sensor modeling for time sequence data based on convolution filters. *IEEE Transaction on Control Systems Technology*, 2021, DOI: 10.1109/TCST.2020.3045845.
- [21] BELLA A D, FORTUNA L, GRAZIANI S, et al. Soft sensor design for a sulfur recovery unit using genetic algorithms. *IEEE International Symposium on Intelligent Signal Processing*. Alcalá de Henares, Spain: IEEE, 2007: 1 – 6.

作者简介:

孙文心 博士研究生, 目前研究方向为复杂工业过程软测量建模,

E-mail: 13373647858@163.com;

马君霞 副教授, 硕士生导师, 目前研究方向包括系统辨识、参数估计和过程控制, E-mail: junxia.20@163.com;

熊伟丽 教授, 博士生导师, 目前研究方向为复杂工业过程的建模和优化、智能软测量技术以及应用和故障检测, E-mail: greenpre@163.com.