

中图分类号: TK3; TP183 文献标志码: 文章编号:

# 基于 LSTM 和参数正则化的 NO<sub>x</sub> 排放 自学习多步预测方法

莫正阳, 郭文康, 李益国

(东南大学大型发电装备安全运行与智能测控国家工程研究中心, 南京, 210096)

## Self-Learning Multi-Step Prediction Method for NO<sub>x</sub> Emission based on LSTM and Parameter Regularization

MO Zhengyang, GUO Wenkang, LI Yiguo

(National Engineering Research Center of Power Generation Control and Safety, School of Energy and Environment,  
Southeast University, Nanjing 210096, China)

**ABSTRACT:** The establishment of an accurate NO<sub>x</sub> prediction model is the foundation for the optimal control of boiler combustion system and the reduction of pollutant emissions. However, the prediction performance of the offline NO<sub>x</sub> emission dynamic model gradually deteriorates with time due to the variation of coal type and equipment characteristics, which makes it difficult to satisfy the requirements of combustion optimization. In order to address this problem, this paper proposes an online self-learning dynamic modeling method for NO<sub>x</sub> emission, which adopts long short-term memory network with encoding and decoding structure (LSTM Encoder-Decoder) for the dynamic model, and completes the model updating by using the most recent data and the parameter regularization method (MAS) when the prediction error exceeds the threshold. The NO<sub>x</sub> emission prediction results for a 650 MW coal-fired boiler show that the proposed method has stable and reliable multi-step prediction accuracy and fewer model updates than its own ablation methods and adaptive least squares support vector machine (ALSSVM). The mean absolute error (MAE) and mean absolute percentage error (MAPE) for multi-step prediction (280 s) for 38000 test samples (17.6 days) are 5.49 mg/m<sup>3</sup> and 1.82%, respectively, which can well overcome the influence of time-varying characteristics of the boiler combustion system and satisfy the requirements of combustion optimization control on real-time and prediction accuracy.

**KEY WORDS:** dynamic model; online learning; NO<sub>x</sub> emission; long short-term memory network; parameter regularization

**摘要:** 建立精确的 NO<sub>x</sub> 预测模型是实现锅炉燃烧优化控制, 减少污染物排放的基础。然而由于煤种和设备特性变化导致离线建立的 NO<sub>x</sub> 排放动态模型的预测性能随时间逐渐降低, 难以满足燃烧优化控制的要求。针对该问题, 本文提出了一种 NO<sub>x</sub> 排放在线自学习动态建模方法, 其动态模型采用编码解码结构的长短期记忆网络(LSTM Encoder-Decoder), 并在预测误差超限时利用最近的历史序列数据和参数正则化方法完成模型更新。针对某 650 MW 燃煤锅炉的 NO<sub>x</sub> 排放预测结果表明, 与自身的消融方法和自适应最小二乘支持向量机(ALSSVM)相比, 所提方法具有稳定可靠的多步预测精度和更少的模型更新次数, 对 38000 个测试样本(17.6 天)的多步预测(280 秒)的平均绝对误差(MAE)和平均绝对百分比误差(MAPE)分别为 5.49 mg/m<sup>3</sup> 和 1.82%, 能很好地克服锅炉燃烧系统时变特性的影响, 并满足燃烧优化控制对模型实时性和预测精度的要求。

**关键词:** 动态模型; 在线学习; NO<sub>x</sub> 排放; 长短期记忆网络; 参数正则化

## 0 引言

燃煤电厂产生的 NO<sub>x</sub> 是主要的大气污染源之一, 对人类健康和空气质量造成了严重影响。为进一步降低锅炉燃烧过程中的 NO<sub>x</sub> 排放, 以数据驱动为主流的锅炉燃烧过程建模, 并结合启发式搜索算法对燃烧相关操作量进行优化的燃烧优化系统(combustion optimization system, COS)被广泛研究<sup>[1-4]</sup>。

建立准确可靠的锅炉燃烧系统(boiler

combustion system, BCS)动态模型是实现燃烧优化的基础<sup>[5]</sup>。为应对 BCS 的大迟延和时变特性,实现闭环优化控制,该模型需要具备以下特征:1)良好的多步预测性能;2)在线自学习能力;3)较低的计算量。

与传统的机理建模方法和计算流体力学仿真建模方法相比,机器学习建模,尤其是神经网络建模方法,具有较低的计算成本、较高的预测精度,因此展示出强大的应用潜力。目前,大部分基于机器学习方法的 BCS 的 NO<sub>x</sub> 排放建模都是针对稳态燃烧阶段<sup>[2,3,6,7]</sup>。然而,随着电网中可再生能源比例不断提高,承担调峰任务的燃煤锅炉需频繁调整负荷,因此建立 NO<sub>x</sub> 排放动态模型,并利用多步预测指标评价其捕捉系统动态特性的能力,对同时实现锅炉稳态和动态过渡过程的燃烧优化至关重要。Smrekar 等<sup>[8]</sup>通过对比多种机器学习建模方法对 NO<sub>x</sub> 排放的多步预测性能,表明机器学习模型可成功捕获 BCS 的动态特性。Yang 等<sup>[9]</sup>和 Tan 等<sup>[10]</sup>利用长短期记忆网络(long short-term memory, LSTM)实现对燃煤锅炉 NO<sub>x</sub> 排放的动态建模,并通过实际数据证实了该方法在预测精度方面优于支持向量机(support vector machine, SVM)。Tuttle 等<sup>[5]</sup>采用 10 种基于数据驱动的动态算法对 NO<sub>x</sub> 排放进行多步预测,结果显示门控循环单元(gate recurrent unit, GRU)展示出最佳的多步预测精度和稳定性。这些研究表明,由于循环神经网络更擅长捕捉时序依赖关系和处理高维数据,在 BCS 动态建模方面具有明显的优势。

尽管对 BCS 的 NO<sub>x</sub> 排放预测已经取得了较高的精度,但现有的研究大多忽略了系统由于炉膛结渣、设备磨损、煤质变化等因素造成的时变特性,导致实际数据分布随时间逐步偏离训练集,从而降低了模型的预测精度<sup>[6,11]</sup>。因此,为了确保 COS 能长期在线闭环运行而无需人工干预,需要采用低计算量和内存开销的在线学习算法,使模型能够从不断产生的数据流中持续学习<sup>[12]</sup>。目前已有一些学者提出了针对 BCS 稳态模型的在线学习算法,例如, Lv 等<sup>[13]</sup>提出增量式的自适应最小二乘支持向量机模型(adaptive least squares support vector machine, ALSSVM),实现锅炉 NO<sub>x</sub> 排放的实时单步预测。Tuttle 等<sup>[6]</sup>建立了在线前馈神经网络模型,并通过每 8 小时对运行数据进行评估,以决定是否再训练网络,从而应对系统特性变化。

然而,由于忽略了动态变化过程,这些方法的多步预测精度不高。

BCS 模型的在线学习算法需同时考虑对新知识的快速学习和对有效旧知识的保留<sup>[13]</sup>,以实现精准预测系统的局部特性和保持全局特性的泛化能力。这种对模型的“可塑性(获取新知识)-稳定性(保存旧知识)”的权衡类似于连续学习<sup>[14]</sup>,因此,可以对一些低资源消耗的连续学习算法进行适当改进,以适应 BCS 动态模型的在线更新需求。例如,连续学习中的弹性权重整合(elastic weight consolidation, EWC)<sup>[15]</sup>、无遗忘学习(learning without forgetting, LwF)<sup>[16]</sup>和记忆感知突触(memory aware synapses, MAS)<sup>[17]</sup>等基于正则化的方法,通过估算模型参数的重要性,并在更新时对重要参数的改变进行惩罚,以防止与历史任务相关的重要知识被覆盖;或者基于经验回放的方式<sup>[18,19]</sup>,将新样本与历史序列中的代表样本混合,共同更新模型,以维持对旧知识的长期记忆力。

经验回放通常需要存储大量的历史样本,导致再训练模型时间复杂度增加,因此本文采用 MAS<sup>[17]</sup>方法,并将其整合到 LSTM Encoder-Decoder 在线更新过程。相较其它正则化方法<sup>[15,16]</sup>,MAS 同时具有以下优点:1)固定的内存开销,只需保存网络每个参数的重要性权重,避免了内存消耗随时间增加;2)无监督学习, MAS 计算重要性权重不依赖于样本标签,扩大了其适用范围;3)低计算开销,只需通过一次反向传播计算梯度来估计网络各参数的重要性;4)优异的在线连续学习性能<sup>[17,20]</sup>。

本文主要贡献包括:1)采用 LSTM Encoder-Decoder 建立了 NO<sub>x</sub> 排放动态模型,并利用多步预测指标评估模型的精度;2)提出了适用于 BCS 的非平稳数据流场景的在线学习方法,并采用 MAS 正则化方法,以较少的模型更新次数实现了稳定持续且高精度的 NO<sub>x</sub> 多步预测,为实现闭环燃烧优化控制打下了基础。

## 1 NO<sub>x</sub> 排放在线自学习建模方法

### 1.1 基于 LSTM 的动态建模方法

本文采用 LSTM Encoder-Decoder<sup>[21]</sup>架构完成端到端的 NO<sub>x</sub> 排放多步预测建模。图 1(a)展示了由两个 LSTM 组成的沿时间展开的 Encoder-Decoder 架构,其中 Encoder 将已知历史信息序列

编码为一个包含上下文信息的固定张量, Decoder 则结合上下文和额外的信息生成目标变量的多步预测结果。模型  $t$  时刻输入  $\tilde{x}_t$  包括编码器部分  $x_{t-e:t-1}$  和解码器部分  $x_{t:t+d-1}$ , 其中,  $t-e:t-1$  表示连续  $e$  个采样时刻  $[t-e, t-e+1, \dots, t-1]$ ,  $e$  表示编码序列长度;  $t:t+d-1$  同理,  $d$  表示解码长度。对于 NOx 排放模型, Encoder 输入  $[x_i] = [u_i, y_i]$ ,  $i = t-e:t-1$ , 其中,  $u_i$  表示锅炉状态, 分为运行人员无法直接控制的机组负荷等状态变量  $u_i^s$ , 以及可以直接控制的二次风门开度、燃烬风门开度和给煤量偏置等操作变量  $u_i^c$ ;  $y_i$  为目标变量, 即 SCR 入口 NOx 浓度。Decoder 训练阶段输入为  $[x_i] = [u_i, y_i]$ ,  $i = t:t+d-1$ , 因输入中包含目标变量, 在预测阶段将采用迭代方式完成多步预测<sup>[22]</sup>。需注意的是, 由于 Decoder 输入中包含的锅炉未来操作量  $u_{t+1:t+d-1}^c$  可由 COS 在线求解优化问题得到, 所以这些变量在 NOx 多步预测过程中采用实际观测值。另外, Decoder 输入中包含的无法直接控制的状态量  $u_{t+1:t+d-1}^s$  将采用最近观测值  $u_t^s$ 。

相较于常规循环神经网络, LSTM 能缓解梯度爆炸问题, 同时捕获变量间长期的时间依赖关系。记当前时刻为  $t$ , LSTM 的单元结构图见图 1(b), 对应的计算公式如下:

$$i_t = \text{sigmoid}(W_i[h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \text{sigmoid}(W_f[h_{t-1}, x_t] + b_f) \quad (2)$$

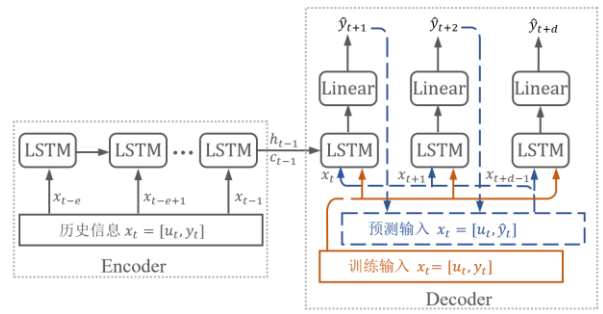
$$o_t = \text{sigmoid}(W_o[h_{t-1}, x_t] + b_o) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (4)$$

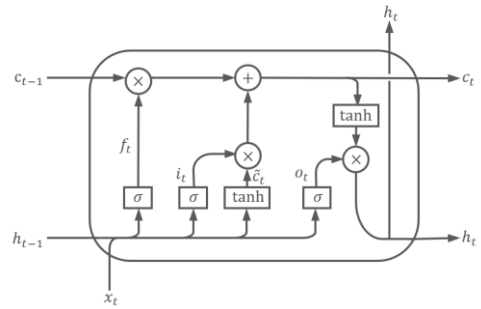
$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (5)$$

$$h_t = o_t \tanh(c_t) \quad (6)$$

式中,  $x_t$  为当前输入;  $c_{t-1}$  和  $h_{t-1}$  是上一时刻 LSTM 输出的记忆状态和隐藏状态;  $W$  和  $b$  为网络的权重和偏置;  $i_t$ ,  $f_t$ ,  $o_t$  分别为输入门、遗忘门和输出门, 因经过 sigmoid 激活函数, 其取值范围均限制在  $[0, 1]$ , 决定着对历史信息 and 当前信息的倾向权重;  $\tilde{c}_t$ 、 $c_t$  为当前候选记忆和输出记忆;  $h_t$  为输出隐藏状态, 在 Decoder 中经过全连接层被转化为预测目标  $\hat{y}_{t+1}$ 。



(a) 沿时间展开的 Encoder-Decoder



(b) LSTM 单元结构

图 1 LSTM Encoder-Decoder 网络架构

Fig.1 LSTM Encoder-Decoder network architecture

## 1.2 基于 MAS 的在线自学习方法

为应对 NOx 排放的时变特性, 解决 BCS 非平稳数据流场景下的连续学习问题, 本文在 LSTM 动态建模基础上, 提出一种基于 MAS 的在线自学习方法, 流程如图 2 所示, 以实现持续稳定和高精度的 NOx 多步预测。

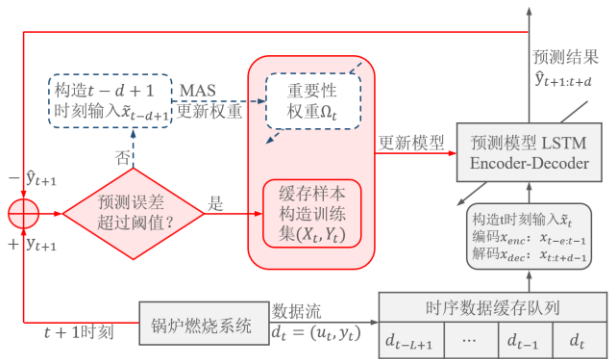


图 2 NOx 排放模型的在线学习方法

Fig.2 An online learning approach for NOx emission modeling

其主要步骤包括:

- 1) 在  $t$  时刻, 采集数据  $d_t$  并保存在长度为  $L$  的数据缓存队列中;
- 2) 构造 LSTM Encoder-Decoder 架构  $t$  时刻的输入  $\tilde{x}_t$  (包括编码部分和解码部分), 并进行 NOx 排

放浓度多步预测；

3) 在  $t+1$  时刻, 将  $\text{NO}_x$  浓度的单步预测结果  $\hat{y}_{t+1}$  与当前观测输入  $y_{t+1}$  对比, 并根据预测误差大小决定是否对模型进行再训练或者更新模型参数的重要性权重。

(a) 如果单步预测误差超过允许阈值, 那么利用缓存队列中的所有数据构造训练集和当前模型参数的重要性权重  $\Omega_t$ , 对模型进行再训练, 以实时跟踪  $\text{NO}_x$  排放特性;

(b) 否则采用 MAS 算法, 并利用缓存队列构造  $t-d+1$  时刻的模型输入  $\tilde{x}_{t-d+1}$  来更新网络参数的重要性权重。

在线学习方法的详细步骤见算法 1。

算法 1:  $\text{NO}_x$  排放模型在线学习方法的详细步骤

```

1: 采用离线训练集  $(X, Y)$ , 预训练模型  $f$ 
   输入: 预测误差允许阈值  $th$ , 衰减因子  $\alpha$ , 系数  $\lambda$ ,
2: 更新再训练次数  $epochs$ , 缓存序列  $buffer$  长度  $L$ 
3: 初始化:  $\{\Omega_i\} = 0$ 
4:  $buffer$  接收时序数据  $d_t = (u_t, y_t)$ 
5:  $\tilde{x}_t = buffer.ci(t)$  # 利用  $buffer$  数据构造  $t$  时刻模型输入
6:  $\hat{y}_{t+1+d} = f(\tilde{x}_t)$  # 迭代多步预测
7:  $e_t = |y_t - \hat{y}_t|$  # 上一时刻单步预测误差
8: if  $e_t > th$ : # 检测模型是否失配
9:    $(X_t, Y_t) = buffer.dataloader()$  # 构造训练集
10:  for  $epoch$  in  $epochs$ :
11:      $loss = L(f(X_t; \theta), Y_t) + \lambda \sum_i \Omega_i (\theta_i - \theta_i^*)^2$ 
12:      $backprob(loss)$  # 计算梯度
13:      $update(\theta)$  # 更新模型参数
14:  end for
15: else: # 预测准确时更新  $\Omega_i$ 
16:    $\tilde{x}_{t-d+1} = buffer.ci(t-d+1)$  #  $t-d+1$  时刻模型输入
17:    $g_i = backprob(\|f(\tilde{x}_{t-d+1})\|_2^2)$  # 计算梯度
18:    $\Omega_i = (1 - \alpha)\|g_i\| + \alpha\Omega_i$  # 更新重要性权重
19: end if
    
```

在非平稳数据流场景下, 通过最近缓存的样本对模型再训练会导致模型仅关注于系统的某一局部过程, 从而直接导致模型失去全局泛化能力。为此, 本文在上述在线学习方法中采用 MAS<sup>[17]</sup> 算法对模型进行再训练, 以保持网络重要参数的稳定性, 其主要原理描述如下。

根据离线训练集  $(X, Y)$  以式(7)作为损失函数获得预训练模型  $f$ 。

$$\min_{\theta} L(f(X; \theta), Y) \quad (7)$$

式中,  $\theta$  为模型参数;  $L$  为损失函数。

对于模型输入  $\tilde{x}_t$ , MAS 使用模型对其参数微小变化的敏感性的线性近似来估计  $f$  各参数的重要性权重, 如式(8)和(9)。

$$f(\tilde{x}_t; \theta + \delta) - f(\tilde{x}_t; \theta) \approx \sum_i g_i(\tilde{x}_t) \delta_i \quad (8)$$

$$\Omega_{t,i} = (1 - \alpha)\|g_i(\tilde{x}_t)\| + \alpha\Omega_{t-1,i} \quad (9)$$

式中,  $\delta_i$  表示模型参数  $\theta_i$  的微小扰动;  $g_i(\tilde{x}_t) = \partial \|f(\tilde{x}_t; \theta)\|_2^2 / \partial \theta_i$ ,  $f(\tilde{x}_t; \theta) \in \mathbb{R}^d$  为多步预测结果。  $\Omega_{t,i}$  表示  $t$  时刻模型参数  $\theta_i$  的重要性权重, 初值取 0, 由当前输入  $\tilde{x}_t$  估计的重要性权重  $\|g_i(\tilde{x}_t)\|$  和上一时刻重要性权重  $\Omega_{t-1,i}$  加权得到;  $\alpha$  为衰减因子, 通过合理设置  $\alpha$ , 可使旧知识逐渐被遗忘。

采用 MAS 方法后, 模型在线更新时的损失函数将增加一个对模型参数改变进行惩罚的正则化项:

$$L(\theta) = L(f(X_t; \theta), Y_t) + \lambda \sum_i \Omega_i (\theta_i - \theta_i^*)^2 \quad (10)$$

式中,  $(X_t, Y_t)$  为当前缓存队列中的数据所构造的训练集;  $\theta^*$  为模型更新前的参数;  $\lambda$  为正则化系数;  $\Omega_i$  为当前模型参数  $\theta_i$  的重要性权重, 其决定模型参数适应新任务的可调水平。

## 2 $\text{NO}_x$ 在线多步预测

### 2.1 锅炉对象与数据说明

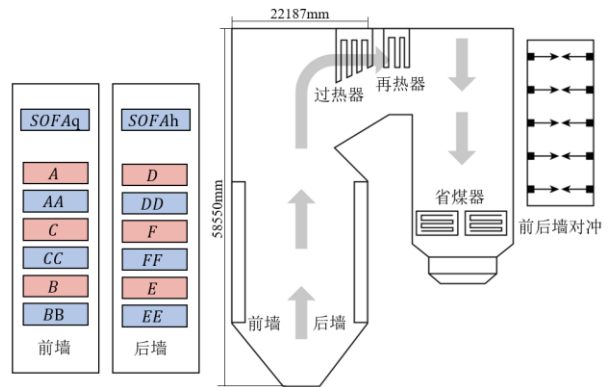


图 3 锅炉燃烧系统及风门布置示意图

Fig.3 Schematic diagram of boiler combustion system and damper arrangement

以某 650 MW 超临界燃煤直流锅炉为研究对象。该锅炉采用低  $\text{NO}_x$  轴向旋流燃烧器, 前后墙对冲燃烧方式, 配备六台中速磨煤机。BCS 及风门布置如图 3 所示, 燃烧器(A、B、C、D、E、F)和二次风喷口(AA、BB、CC、DD、EE、FF)每层各 5 只, 采用交替方式分三层前、后墙对冲布置。在

表 1 建模变量及变化范围

Tab.1 Modeling variables and range of variation

变量	说明	变化范围
状态变量 $u^s$	$LOAD$	机组负荷 [270,630] MW
$A、B、C、D、E、F$	6 台给煤机给煤量百分比	[0,100] %
操作变量 $u^c$	$AA、BB、CC、DD、EE、FF$	3 层前、后墙二次风门开度 [0,100] %
$SOFA$	前、后墙燃烬风门开度均值	[30,100] %
$O_2$	省煤器出口氧量设定	[1.2,8] %
目标变量 $y$	$NOx$	SCR 入口 NOx 浓度 [150,650] mg/m <sup>3</sup>

最上层燃烧器上方，前后墙各布置 1 层燃烬风口 ( $SOFA_q$ 、 $SOFA_h$ )。

建模数据来自电厂的分散式控制系统(DCS)，采样周期为 40 s，共计 46000 组。模型的输入考虑为 BCS 的主要操作变量、机组负荷和历史 NOx 排放浓度，共计 16 维，输出为未来 NOx 排放浓度，详细的变量信息见表 1。

## 2.2 建模过程

### 2.2.1 数据划分与训练集构建

数据集集中的前 8000 组用于模型离线预训练，并按 3:1 的比例划分为训练集和验证集，其中，训练集用于建立模型，验证集用于超参数寻优，后 38000 组(17.6 天)用于模型在线测试。图 4 展示了离线样本的负荷变化情况，可见样本基本覆盖了机组正常运行的负荷范围，且变化频繁。因此，建模样本能充分反映 BCS 的动态特性。

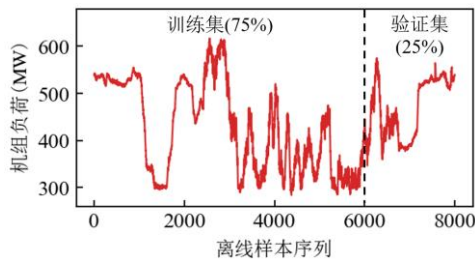


图 4 离线预训练样本负荷变化曲线

Fig.4 Load variation curve for offline pretrained samples

接着采用合适方式把时间序列数据转化为 LSTM Encoder-Decoder 的训练集。图 5 描述了构建训练集样本的过程，其中， $x_{enc} \in \mathbb{R}^{e \times 16}$ ，和  $x_{dec} \in \mathbb{R}^{d \times 16}$  分别为 Encoder 和 Decoder 输入； $y_{dec} \in \mathbb{R}^{d \times 1}$  为 Decoder 输出对应的真实标签； $stride$  为两样本间采样间隔，本文设置为 1，以获得充足的训练样本。由于 SCR 入口的 NOx 浓度对各操作量响应的过渡时间均在 10 min 内，因此将编码长度  $e$  设置为 20，将解码长度  $d$  设置为 15。所以  $t$  时

刻 LSTM Encoder-Decoder 训练集输入及对应的标签可表示为： $x_{enc} = [x_i] = [u_i, y_i]$ ， $i = t - 20 : t - 1$ ， $x_{dec} = [x_i] = [u_i, y_i]$ ， $i = t : t + 14$ ， $y_{dec} = y_{t+1:t+15}$ 。

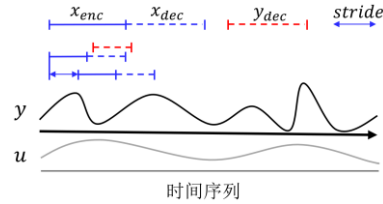


图 5 LSTM Encoder-Decoder 训练集样本构建示意图

Fig.5 Schematic of training set sample construction for LSTM Encoder-Decoder

### 2.2.2 评价指标

采用平均绝对误差(mean absolute error, MAE)和平均绝对百分比误差(mean absolute percentage error, MAPE)评估模型的预测精度，它们的定义如下：

$$MAE_h = \frac{1}{N} \sum_{i=1}^N |y_{k+h}^i - \hat{y}_{k+h}^i| \quad (11)$$

$$MAPE_h = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_{k+h}^i - \hat{y}_{k+h}^i}{y_{k+h}^i} \right| * 100 \quad (12)$$

式中， $h$  为前向预测步数， $N$  为评估时域内的总样本数。

### 2.2.3 参数设置与超参数寻优

采用离线样本和网格搜索技术优化 LSTM Encoder-Decoder 架构的超参数，超参数的搜索空间见表 2。表中 hidden size 确定 LSTM 循环层的隐藏单元数量；layers 确定 LSTM 循环层层数；weight decay 和 dropout 为权重衰退和 Dropout 正则化方法的超参数。其他常见设置如下：激活函数为 ReLU，batch size 为 128，最大 epoch 为 150，优化算法为 Adam，学习率为 0.001，损失函数为 L1Loss(即 MAE)。表 3 展示了超参数搜索结果，模型在每个网格点重复训练 5 次，综合考虑验证集预



测结果的准确性、稳定性和模型自身的泛化性能，选择[64,2,0.4,1e-3]为优化后模型参数。

在线更新时，LSTM Encoder-Decoder 的误差允许阈值  $th$  设置为  $1.9+r \times 1.5 \text{ mg/m}^3$ ，并令  $r=1$ ，其中，1.9 为离线验证集单步预测的 MAE 均值，1.5 为标准差。同时设置衰减因子  $\alpha = 0.999$ ，正则化系数  $\lambda = 1$ ，再训练次数  $epochs = 1$ ，缓存长度  $L = 1000$ 。

表 2 LSTM Encoder-Decoder 的超参数搜索范围

Tab.2 Hyperparametric search range for LSTM Encoder-Decoder

超参数	搜索范围
hidden size	16, 32, <b>64</b>
layers	<b>2</b> , 3, 4
dropout	0.1, 0.2, <b>0.4</b> , 0.6
weight decay	1e-1, 1e-2, <b>1e-3</b> , 1e-4

注:粗体表示最优参数或指标，下同

表 3 超参数搜索结果

Tab.3 Search results for hyperparameters

模型参数	训练集 MAE(mg/m <sup>3</sup> )	验证集 MAE(mg/m <sup>3</sup> )
[32,2,0.1,0.01]	1.716±0.352	2.124±0.438
[32,2,0.1,0.001]	1.808±0.378	2.216±0.582
<b>[64,2,0.4,0.001]</b>	<b>2.041±0.181</b>	<b>2.264±0.213</b>
[32,3,0.1,0.0001]	1.733±0.128	2.290±0.187
[64,2,0.1,0.0001]	1.964±0.536	2.300±0.936

注：结果按验证集 MAE 由小到大排列；指标表示为 mean±std。

## 2.3 预测结果

### 2.3.1 对比方法

为评估所提自学习多步预测方法的有效性，对下列方法进行对比分析：

1) Offline<sup>[4,10]</sup>：使用 LSTM Encoder-Decoder

建立 BCS 动态模型，该模型仅经过离线预训练，不进行在线学习。

2) Online：将 Offline 模型用于在线学习，使用缓存区中的数据对模型再训练，但不采用 MAS 方法，即将正则化系数  $\lambda$  设置为 0。

3) Online+MAS(本文)：除再训练时采用 MAS 方法外，其余设置与 Online 一致。

4) ALSSVM<sup>[13]</sup>：使用增量式的自适应最小二乘支持向量机算法，对 NO<sub>x</sub> 排放浓度在线预测。

ALSSVM 模型采用径向基核函数(radial basis function, RBF)，样本容量设置为 1000，惩罚系数  $c$  和核参数  $k$  同样采用离线样本和网格搜索技术优化得到，搜索范围均为[1e-3,1e-2,...,1e2,1e3]，优化结果为  $c=1000$ ， $k=0.01$ 。在线更新时，误差允许阈值  $th$  设置为  $4.2+r \times 3.5 \text{ mg/m}^3$ ，并令  $r=1$ ，同样，4.2 为离线验证集单步预测的 MAE 均值，3.5 为标准差。

$t$  时刻训练样本输入表示为[LOAD( $t$ ),A( $t-5$ ),B( $t-3$ ),C( $t-4$ ),D( $t-4$ ),E( $t-4$ ),F( $t-4$ ),AA( $t-4$ ),BB( $t-3$ ),CC( $t-3$ ),DD( $t-1$ ),EE( $t-4$ ),FF( $t-5$ ),SOFA( $t-3$ ),O2( $t$ ),NO<sub>x</sub>( $t$ )，输出为[NO<sub>x</sub>( $t+1$ )]，其中，各操作量与输出目标之间的最佳迟延阶数由最大互信息系数<sup>[23]</sup>(maximum information coefficient, MIC)确定，具体数值见图 6。为评估 ALSSVM 捕捉系统动态特性的性能，采用递推方式进行前向 15 步预测。与 LSTM Encoder-Decoder 一致，递推时输入中的操作变量  $u^c$  使用真实值(实际部署中由 COS 优化求解得到)，状态变量保持为最近观测值，目标变量  $y$  采用前一时刻的预测结果。

Time(s)	Delay	A	B	C	D	E	F	AA	BB	CC	DD	EE	FF	SOFA	O2
40	1	0.1901	0.2883	0.4818	0.1429	0.5041	0.4933	0.2843	0.3080	0.4484	0.3006	0.3063	0.3891	0.3271	<b>0.6473</b>
80	2	0.1948	0.2920	0.4945	0.1461	0.5159	0.5043	0.2872	0.3101	0.4485	<b>0.3059</b>	0.3134	0.3953	0.3268	0.6199
120	3	0.1964	0.2958	0.5091	0.1526	0.5337	0.5238	0.2873	0.3112	0.4485	0.3053	0.3174	0.3967	0.3265	0.6021
160	4	0.1988	<b>0.3038</b>	0.5280	0.1569	0.5568	0.5469	0.2889	<b>0.3145</b>	<b>0.4535</b>	0.3053	0.3240	0.3999	<b>0.3280</b>	0.5929
200	5	0.2015	0.3037	<b>0.5390</b>	<b>0.1591</b>	<b>0.5712</b>	<b>0.5608</b>	<b>0.2899</b>	0.3129	0.4525	0.3029	<b>0.3254</b>	0.4000	0.3247	0.5975
240	6	<b>0.2019</b>	0.3022	0.5353	0.1542	0.5692	0.5567	0.2889	0.3134	0.4502	0.3009	0.3215	<b>0.4026</b>	0.3229	0.6061
280	7	0.2005	0.2961	0.5227	0.1499	0.5535	0.5439	0.2887	0.3099	0.4515	0.3014	0.3191	0.3986	0.3215	0.6162
320	8	0.1977	0.2920	0.5070	0.1463	0.5414	0.5292	0.2874	0.3075	0.4492	0.3000	0.3161	0.3949	0.3210	0.6263
360	9	0.1955	0.2870	0.4989	0.1446	0.5308	0.5182	0.2846	0.3052	0.4511	0.2989	0.3128	0.3923	0.3231	0.6361
400	10	0.1934	0.2825	0.4901	0.1442	0.5190	0.5053	0.2854	0.3084	0.4490	0.3007	0.3135	0.3913	0.3251	0.6384
440	11	0.1919	0.2811	0.4833	0.1430	0.5115	0.4973	0.2835	0.3059	0.4477	0.2984	0.3116	0.3924	0.3261	0.6397
480	12	0.1908	0.2808	0.4818	0.1435	0.5085	0.4952	0.2811	0.3072	0.4451	0.3019	0.3109	0.3916	0.3241	0.6377
520	13	0.1906	0.2845	0.4850	0.1454	0.5099	0.4993	0.2833	0.3085	0.4460	0.3035	0.3095	0.3906	0.3201	0.6327
560	14	0.1909	0.2899	0.4919	0.1455	0.5172	0.5071	0.2830	0.3092	0.4464	0.3029	0.3098	0.3898	0.3208	0.6281
600	15	0.1892	<b>0.2933</b>	0.4915	0.1450	0.5186	0.5080	0.2834	0.3083	0.4435	0.3018	0.3096	0.3888	0.3181	0.6260

图 6 不同迟延阶数下的操作量与 NO<sub>x</sub> 浓度之间的 MIC

Fig.6 Maximum Information Coefficient between operating variables and NO<sub>x</sub> concentration at different delay orders

表4 不同预测步数和区间下 NO<sub>x</sub> 排放预测精度Tab.4 Accuracy of NO<sub>x</sub> emission prediction at different prediction steps and intervals

对比方法	离线验证集 6000-8000				在线测试集 8000-46000				
	H=1	H=1:7	H=8:15	H=1:15	H=1	H=1:7	H=7	H=8:15	H=1:15
MAP(mg/m <sup>3</sup> )									
ALSSVM	5.38±4.22	16.45±15.98	38.74±31.79	29.19±28.14	5.46±4.70	16.67±17.37	29.72±26.04	39.90±35.87	29.93±31.23
Offline	1.93±2.38	6.39±8.36	27.61±22.84	18.33±20.59	2.31±2.29	7.54±9.08	18.23±16.28	<b>30.04±28.65</b>	20.44±24.66
Online	1.93±2.38	6.39±8.36	27.61±22.84	18.33±20.59	1.22±1.46	5.62±7.70	15.86±14.97	31.77±30.24	20.20±26.16
Online+MAS	1.93±2.38	6.39±8.36	27.61±22.84	18.33±20.59	<b>1.22±1.46</b>	<b>5.49±7.52</b>	<b>15.38±14.6</b>	30.56±29.34	<b>19.52±25.31</b>
MAPE(%)									
ALSSVM	1.53±1.04	4.46±3.98	10.81±7.49	8.17±6.86	1.79±1.53	5.48±5.7	9.79±8.59	13.16±11.93	9.87±10.36
Offline	0.54±0.53	1.80±2.10	7.97±6.12	5.27±5.60	0.82±0.93	2.58±3.15	6.15±5.48	10.08±9.03	6.82±7.90
Online	0.54±0.53	1.80±2.10	7.97±6.12	5.27±5.60	0.41±0.48	1.86±2.50	5.24±4.88	10.49±9.96	6.69±8.61
Online+MAS	0.54±0.53	1.80±2.10	7.97±6.12	5.27±5.60	<b>0.41±0.48</b>	<b>1.82±2.39</b>	<b>5.07±4.88</b>	<b>10.07±9.40</b>	<b>6.43±8.16</b>

注: H=i:j 代表前向 i 步至前向 j 步预测范围内的平均统计结果; 指标表示为 mean±std。

### 2.3.2 结果讨论

在线数据流场景下, 四种方法对离线验证集和在线测试集, 在不同预测步数和区间下的 MAP 和 MAPE 的均值和标准差见表 4。通过分析结果可得出以下结论:

1) Offline 在测试集上的各预测步数和区间下的预测误差均显著大于离线验证集, 表明离线的静态模型并不适用于非平稳条件下的 NO<sub>x</sub> 预测。

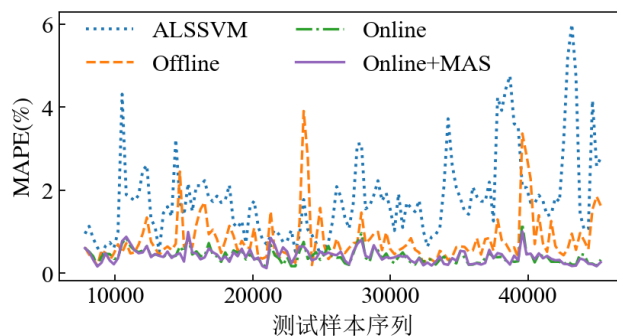
2) 与 Offline 相比, Online+MAS 在测试集上展现出更高的预测精度, 尤其是预测范围较小时, 如 H=1:7, 其预测误差指标 MAP 和 MAPE 分别为 5.49 mg/m<sup>3</sup> 和 1.82 %, 明显低于 Offline 的 7.54 mg/m<sup>3</sup> 和 2.58 %。

3) ALSSVM 能够适应非平稳环境, 在验证集和测试集上展示出相似的预测性能, 但受限于 LSSVM 本身不擅长处理高维大数据, 其在各预测步数和区间下的预测误差均明显大于 Online 和 Online+MAS, 甚至大于未在线学习的 Offline。

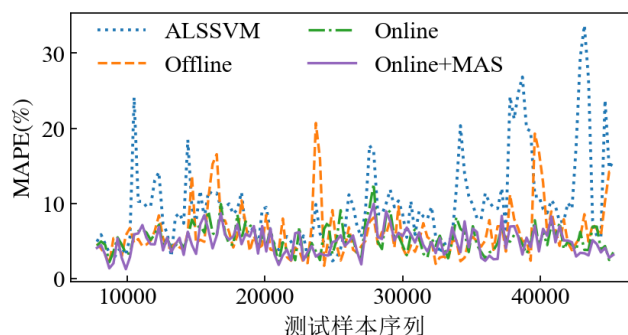
4) Online+MAS 在预测范围 H=1:15 内的 MAP 和 MAPE 分别为 19.52 mg/m<sup>3</sup> 和 6.43 %, 在对比的四种方法中多步预测精度最高。

图直观展示了四种方法对非平稳环境的适应性。图(a)为前向 1 步预测结果的 MAPE 随测试集序列的变化曲线, 由图可知: 1) Online 和 Online+MAS 在整个测试时域内保持了较高且稳定的预测精度, 表明了模型在线更新的有效性。2) Offline 由于受系统时变特性的影响, 预测精度波动较大。3) ALSSVM 的预测精度波动最为剧烈, 说明该方法对非平稳环境的适应性能较差。图(b)

展示了前向 7 步预测误差的变化情况, 结论与之前基本一致, 其中 Online+MAS 对非平稳数据流的适应性最好, 多步预测精度也最高。



(a)前向 1 步预测结果的 MAPE



(b)前向 7 步预测结果的 MAPE

图 7 四种对比方法的 MAPE 随测试样本序列变化情况

Fig.7 Variation of MAPE with test sample sequence for the four compared methods

图直观展示了不同方法对未来 15 步 NO<sub>x</sub> 排放趋势变化的捕捉能力。由图可知, 基于 LSTM Encoder-Decoder 方法的预测结果多数情况下与观测趋势基本一致, 展示出良好的趋势捕捉能力,

这得益于 LSTM 模型中隐状态的时序传递。而 Online 和 Online+MAS 的趋势捕捉更强, 精度更高, 这归功于其采取的在线学习机制, 使得模型能及时调整并适应 NO<sub>x</sub> 排放特性的变化。

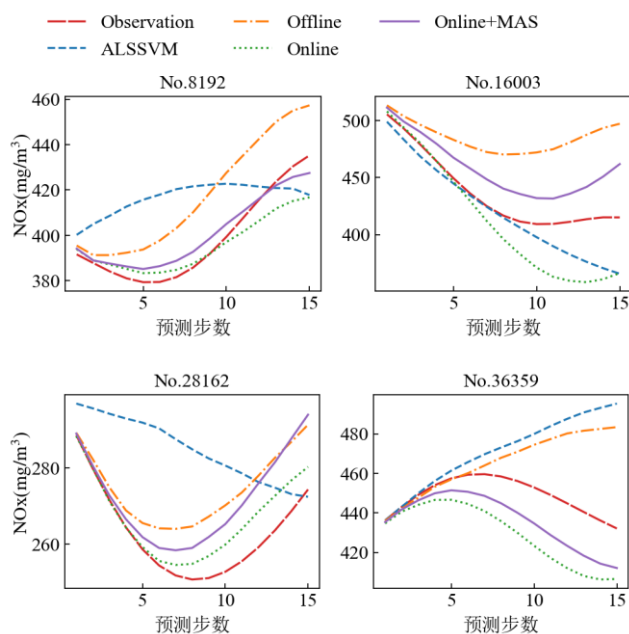


图8 不同样本点的前向15步预测曲线  
Fig.8 Forward 15-step prediction curves for different sample points

在运行环境 CPU AMD Ryzen7 4800H 2.9GHz, GPU GeForce RTX3070 8G 下, 模型更新过程的时间消耗和总更新次数见表5。因采用正则化方法保留了模型对典型工况的预测性能, Online+MAS 有效减少了模型更新次数且实现了更高精度的多步预测。模型单次平均更新时间不超过 0.1 秒, 完全满足闭环燃烧优化控制对模型更新的实时性要求。

表5 对比方法在测试集上性能的综合比较

Tab.5 Comprehensive comparison of performance of comparison methods on test set

对比方法	MAE(mg/m <sup>3</sup> ) H=1:15	MAPE(%) H=1:15	更新次数	平均更新时间(s/次)
ALSSVM	29.93	9.87	9654	0.0718
Offline	20.44	6.82	/	/
Online	20.20	6.69	1813	<b>0.0683</b>
Online+MAS	<b>19.52</b>	<b>6.43</b>	<b>1678</b>	0.0836

### 3 结论

面对非平稳数据流, 传统的锅炉燃烧系统 NO<sub>x</sub> 排放模型不能很好地适应系统的时变特性, 实现对 NO<sub>x</sub> 排放的准确预测。为解决该问题, 本

文提出了一种基于 LSTM Encoder-Decoder 和 MAS 的在线自学习动态建模方法。针对某 650 MW 燃煤锅炉实际运行数据开展的仿真结果表明, 所提方法能持续保持很高的多步预测精度, 尤其是在前向 1 步至 7 步预测范围内, 其 MAP 和 MAPE 仅为 5.49 mg/m<sup>3</sup> 和 1.82 %, 远低于 ALSSVM 和常规的 LSTM Encoder-Decoder 模型。同时, 所提方法通过限制网络重要参数的变化, 保持了网络预测性能的稳定, 降低了模型在线更新次数, 并进一步提高了预测精度。该方法可为实现闭环燃烧优化控制提供稳定且精确的动态模型, 并可应用于其他复杂和时变工业过程的动态建模和优化控制。

### 参考文献:

- [1] 徐文韬, 黄亚继, 曹歌瀚, et al. 基于 BP-改进 NSGA-II 锅炉燃烧多目标优化[J]. 东南大学学报(自然科学版), 2022, 52(05): 943–952.  
Xu Wentao, Huang Yaji, Cao Gehan, et al. Multi-objective combustion optimization for boiler based on BP-improved NSGA-II[J]. Journal of Southeast University (Natural Science Edition), 2022, 52(05): 943–952(in Chinese).
- [2] Wang C, Liu Y, Zheng S, et al. Optimizing combustion of coal fired boilers for reducing NO<sub>x</sub> emission using Gaussian Process[J]. Energy, 2018, 153: 149–158.
- [3] Shi Y, Zhong W, Chen X, et al. Combustion optimization of ultra supercritical boiler based on artificial intelligence[J]. Energy, 2019, 170: 804–817.
- [4] Blackburn L D, Tuttle J F, Andersson K, et al. Development of novel dynamic machine learning-based optimization of a coal-fired power plant[J]. Computers & Chemical Engineering, 2022, 163: 107848.
- [5] Tuttle J F, Blackburn L D, Andersson K, et al. A systematic comparison of machine learning methods for modeling of dynamic processes applied to combustion emission rate modeling[J]. Applied Energy, 2021, 292: 116886.
- [6] Tuttle J F, Vesel R, Alagarsamy S, et al. Sustainable NO<sub>x</sub> emission reduction at a coal-fired power station through the use of online neural network modeling and particle swarm optimization[J]. Control Engineering Practice, 2019, 93: 104167.
- [7] 唐振浩, 朱得宇, 李扬. 基于数据驱动的燃煤锅炉 NO<sub>x</sub> 排放浓度动态修正预测模型[J]. 中国电机工程学报, 2022, 42(14): 5182–5194.  
Tang Zhenhao, Zhu Deyu, Li Yang. Data Driven based Dynamic Correction Prediction Model for NO<sub>x</sub> Emission of Coal Fired Boiler[J]. Proceedings of the CSEE, 2022, 42(14): 5182–5194(in Chinese).



- [8] Smrekar J, Potočník P, Senegačnik A. Multi-step-ahead prediction of NO<sub>x</sub> emissions for a coal-based boiler[J]. Applied Energy, 2013, 106: 89–99.
- [9] Yang G, Wang Y, Li X. Prediction of the NO<sub>x</sub> emissions from thermal power plant using long-short term memory neural network[J]. Energy, 2020, 192: 116597.
- [10] Tan P, He B, Zhang C, et al. Dynamic modeling of NO<sub>x</sub> emission in a 660 MW coal-fired boiler with long short-term memory[J]. Energy, 2019, 176: 429–436.
- [11] Gu Y, Zhao W, Wu Z. Online adaptive least squares support vector machine and its application in utility boiler combustion optimization systems[J]. Journal of Process Control, 2011, 21(7): 1040–1048.
- [12] Ditzler G, Roveri M, Alippi C, et al. Learning in Nonstationary Environments: A Survey[J]. IEEE Computational Intelligence Magazine, 2015, 10(4): 12–25.
- [13] Lv Y, Yang T, Liu J. An adaptive least squares support vector machine model with a novel update for NO<sub>x</sub> emission prediction[J]. Chemometrics and Intelligent Laboratory Systems, 2015, 145: 103–113.
- [14] Cossu A, Carta A, Lomonaco V, et al. Continual Learning for Recurrent Neural Networks: an Empirical Evaluation[J]. Neural Networks, 2021, 143: 607–627.
- [15] Kirkpatrick J, Pascanu R, Rabinowitz N, et al. Overcoming catastrophic forgetting in neural networks[J]. Proceedings of the National Academy of Sciences, 2017, 114(13): 3521–3526.
- [16] Li Z, Hoiem D. Learning without Forgetting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 2935–2947.
- [17] Aljundi R, Babiloni F, Elhoseiny M, et al. Memory Aware Synapses: Learning What (not) to Forget[C]. Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 144–161.
- [18] Hayes T L, Cahill N D, Kanan C. Memory Efficient Experience Replay for Streaming Learning[C]. 2019 International Conference on Robotics and Automation (ICRA). 2019: 9769–9776.
- [19] Rolnick D, Ahuja A, Schwarz J, et al. Experience replay for continual learning[G]. Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2019: 350–360.
- [20] Aljundi R, Kelchtermans K, Tuytelaars T. Task-Free Continual Learning[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 11246–11255.
- [21] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]. Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. Cambridge, MA, USA: MIT Press, 2014: 3104–3112.
- [22] Noman F, Alkaws G, Alkahtani A A, et al. Multistep short-term wind speed prediction using nonlinear auto-regressive neural network with exogenous variable selection[J]. Alexandria Engineering Journal, 2021, 60(1): 1221–1229.
- [23] Albanese D, Riccadonna S, Donati C, et al. A practical tool for maximal information coefficient analysis[J]. GigaScience, 2018, 7(4): 1–8.



莫正阳

作者简介：莫正阳(1999)，男，硕士研究生，主要从事火电厂节能减排技术及锅炉燃烧优化控制的研究工作。