

自回归神经网络的预测值反馈再训练策略及应用

莫正阳 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)

摘要: 带外生变量的非线性自回归神经网络(NARX-NN)是一种可通过迭代实现动态参数多步预测的技术,但在训练阶段其训练样本输入均采用实测值,而在多步预测阶段,模型输入则采用存在误差的预测值。种差异会导致预测误差随着迭代过程迅速累积。为缓解上述问题,本文提出一种预测值反馈再训练策略(FR),其基本思路如下:首先采用常规训练策略对 NARX-NN 网络进行训练,然后用单步预测值替换实测值重新构造训练样本,并指导网络再次训练,以提高网络在进行多步预测时对预测误差的鲁棒性;最后采用所提 FR 训练策略训练 NARX-DNN、LSTM Encoder-Decoder 和 DeepAR 三种典型的 NARX-NN,应用于电站锅炉 NO_x 浓度的预测或综合能源系统用户电负荷的预测。与常规训练策略和计划采样策略(SS)的对比结果表明,FR 训练策略能显著提高各模型多步点预测精度。

关键词: 深度学习; NARX-NN; 多步预测; 误差积累; 训练策略

中图分类号: TK3; TP183 **文献标志码:** **文章编号:**

Feedback Retraining Strategy for Predicted Values of Autoregressive Neural Networks and Application

Mo Zhengyang Li Yiguo

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

ABSTRACT: Nonlinear autoregressive neural network with exogenous variables (NARX-NN) is a technique that can achieve multi-step prediction of dynamic variables through iteration, but in the training phase, its training sample inputs are used with measured values, while in the multi-step prediction phase, the model inputs are used with predicted values with errors. This discrepancy leads to a rapid accumulation of prediction errors with the iterative process. To alleviate the above problems, this paper proposes a feedback retraining strategy (FR) for the predicted values, whose basic idea is as follows: firstly, the conventional training strategy is used to train the NARX-NN network, then the measured values are replaced with single-step predicted values to reconstruct the training samples and guide the network to be trained again to improve the robustness of the network against prediction errors when making multi-step predictions; finally, the proposed FR training strategy is used to Three typical NARX-NNs, NARX-DNN, LSTM Encoder-Decoder and DeepAR, are trained and applied to the prediction of NO_x concentration in power plant boilers or the prediction of customer electrical load in integrated energy systems. The comparison results with the conventional training strategy and the scheduled sampling strategy (SS) show that the FR training strategy can significantly improve the multi-step point prediction accuracy of each model.

Key Words: deep learning; NARX-NN; multi-step prediction; error accumulation; training strategy

时间序列数据多步预测能给出目标变量未来一段时间内的估计值,较单步预测更具应用价值。例如,日前逐时风光出力 and 负荷预测应用于电力市场经济调度^[1,2],洪水流量^[3]和台风风速^[4]多步预测以提前防治自然灾害,锅炉主蒸汽温度^[5]和 NO_x 排放多步预测^[6]以控制其平稳、经济运行。

典型的多步预测方法可分为迭代多步(IMS)和直接多步(DMS)。IMS 将模型单步预测结果作为下一时刻输入完成迭代多步输出,因此 IMS 仅需建立一个单步预测模型就能实现对未来任意步预测。目前,对复杂、高维、机理未知的动态系统完成 IMS 建模常采用深度学习技术,包括基础网络和混合架构。例如 Hu^[3]通过在时间维度上扩展 DNN 网络的输入,实现锅炉未来 5 分钟内的主汽温度趋势预测。Yuan^[7]提出一种能准确捕

捉目标变量动态特性的有监督 LSTM,并在两个工业过程软测量应用中实现多步预测。Lim^[8]考虑到时序建模输入数据的异质性,将输入分为历史已知信息、未来协变量信息和静态常量,并使用有针对性的网络组件对每种类型输入提取特征,用 Self-Attention 对各种特性进行融合,使该架构在广泛的数据集上均表现出显著的性能提升。Li^[2]构建了 CNN-GRU 混合架构,用 CNN 提取光伏发电相关数据的特征,用 GRU 捕捉特征间的时间关联性,结合数据增强后的聚类算法,该架构对光伏发电功率表现出较高预测精度。

多步迭代预测神经网络模型可视为一个带有协变量输入(协变量集合可以为空)的非线性自回归神经网络(NARX-NN),该网络的输入中包含目标变量的历史值,进行多步预测时需将预测值反馈作为模型的输入。

Noman^[9]的实验结果表明,相较于仅使用协变量预测,将预测结果反馈到输入有助于实现更精确的多步预测。显然,对于动态系统,将目标变量历史值嵌入到输入空间能更完善地表征系统所处的状态,从而提高预测精度。尽管NARX-NN已应用于各个领域,但常规训练策略,如NARX-DNN采用单步预测误差梯度下降法、RNN采用teacher forcing,其训练样本输入均采用实测值,这与多步预测阶段,模型被迫采用存在误差的预测值作为输入,存在很大差异。这一差异使得预测误差会随着迭代过程逐渐累积^[10]。这是常规训练策略存在的一个基本问题。

针对RNN的上述问题,Bengio^[11]提出计划采样(Scheduled Sampling, SS)策略,其基本思想是在训练阶段使用预测值逐步替换下一时刻输入中的实测值。在多步预测任务^[12,13]中,SS有效提高了预测精度,但其预测值和实测值混合教学的方式会显著降低模型的收敛速度^[14],且对非循环结构的NARX-DNN,SS策略将不适用。

为此,本文以提高多步预测精度为目标,提出预测值反馈再训练策略(FR)以提高网络对预测误差的鲁棒性能。该策略首先采用常规训练策略完成网络训练,然后利用训练样本执行单步预测,再用单步预测值替换实测值重新构造训练样本,最后进行网络再训练。通过采用FR策略训练NARX-DNN、LSTM Encoder-Decoder和DeepAR三种典型的NARX-NN,应用于对锅炉NO_x浓度或综合能源系统用户电负荷的多步预测中,验证了该策略的有效性。

1 NARX-NN 预测值反馈再训练策略

1.1 NARX 模型

典型的包含协变量输入的非线性动力系统能被表示为:

$$\begin{aligned} y(k) = & f(y(k-1), \dots, y(k-n_y); \\ & u(k-\tau), \dots, u(k-n_u-\tau)) \end{aligned} \quad (1)$$

式中, y 为目标变量, u 为协变量, k 代表某一采样周期 T 下的离散时刻, $n_y \geq 1$, $n_u \geq 1$ 和 $\tau \geq 1$ 分别为系统目标变量的阶数、协变量阶数和协变量纯时延步数, f 为非线性映射函数,表征系统的动态特性。利用NARX结构,多步预测可通过式(2)迭代实现。

$$\begin{aligned} \hat{y}(k) = & f(\hat{y}(k-1), \dots, \hat{y}(k-n_y); \\ & u(k-\tau), \dots, u(k-n_u-\tau)) \end{aligned} \quad (2)$$

目前,因神经网络(NN)具有拟合任意非线性函数的能力而被广泛用于建立映射 f , 并使用式(1)完成训练过程,式(2)完成预测过程。这种训练和预测阶段的差异被称为暴露偏差^[17](exposure bias)。由于常规训练策略在训练过程中并未考虑到预测误差,网络可能对预测误差较为敏感而导致其在式(2)中迅速放大。

1.2 NARX-NN 模型

接下来首先介绍三种典型的NARX-NN多步预测模型,其中NARX-DNN和LSTM Encoder-Decoder用于点预测,而DeepAR用于概率预测。

1.2.1 NARX-DNN

NARX-DDN^[15]网络基本架构见图1,其实质为一单步预测DNN。在常规训练阶段,网络输入包括目标变量和协变量,网络输出值与实测值之间的差异将作为损失对网络权值进行梯度更新。在预测阶段需将输出预测值反馈到输入层,完成前向计算,实现多步预测。

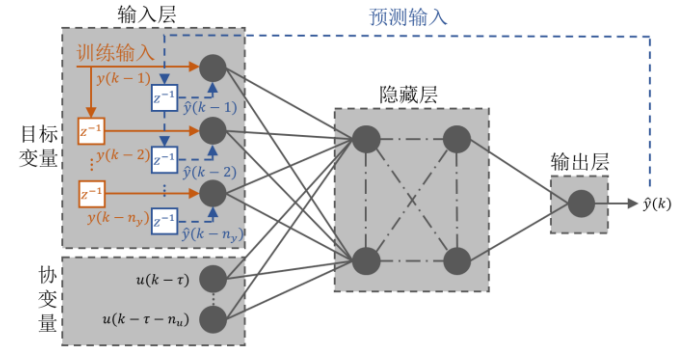


图1 NARX-DNN 网络框图

DNN网络架构不具备捕获时序依赖关系的能力,因此为提高预测精度,网络输入中需要包括各变量的多阶历史值。常见确定输入和输出变量阶数的策略有相空间重构^[16]和剪枝策略^[17]。

1.2.2 LSTM Encoder-Decoder

基于长短期记忆网络(LSTM)的编码器-解码器^[18](Encoder-Decoder)能直接完成端到端的多步预测学习。如图2(a)所示,该架构由两个LSTM组成,分别对应Encoder和Decoder,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序列。相较于常规循环神经网络(RNN),LSTM能缓解梯度爆炸问题,同时捕获变量间长期的依赖关系^[19]。记当前时刻为 k , LSTM的单元结构图见图2(b),对应的计算公式如下:

$$i_k = \text{sigmoid}(W_i[h_{k-1}, x_k] + b_i) \quad (3)$$

$$f_k = \text{sigmoid}(W_f[h_{k-1}, x_k] + b_f) \quad (4)$$

$$o_k = \text{sigmoid}(W_o[h_{k-1}, x_k] + b_o) \quad (5)$$

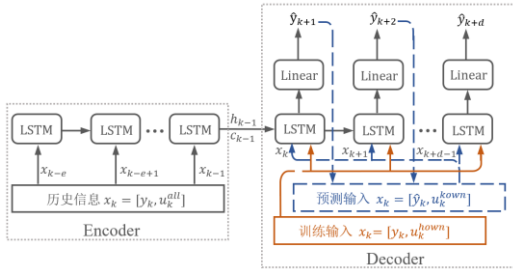
$$\tilde{c}_k = \tanh(W_{\tilde{c}}[h_{k-1}, x_k] + b_{\tilde{c}}) \quad (6)$$

$$c_k = f_k c_{k-1} + i_k \tilde{c}_k \quad (7)$$

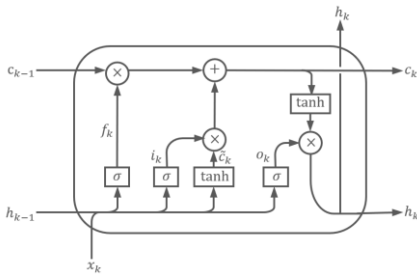
$$h_k = o_k \tanh(c_k) \quad (8)$$

式中, x_k 表示当前输入,是由当前时刻协变量和观测到的目标变量组成的张量,即 $x_k = [y_k, u_k]$; c_{k-1} 和 h_{k-1} 是上一时刻LSTM输出的记忆状态和隐藏状态; i_k , f_k , o_k 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],决定着对历史信息和当前信息的倾向权重; \tilde{c}_k 、 c_k 为当前候选记忆和输出记忆; h_k 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}_{k+1} 。

在 RNN 的 teacher forcing 常规训练策略中, 沿时间维展开的 Decoder 架构中各 LSTM 单元输入为目标实测值和未来协变量信息, 即 $x_k = [y_k, u_k^{known}]$, 如图 2(a)所示, 图中 e 为编码长度, d 为解码长度。而在预测阶段则采用目标预测值参与构成单元输入 $x_k = [\hat{y}_k, u_k^{known}]$ 。



(a) 沿时间展开的 Encoder-Decoder



(b) LSTM 单元结构

图 2 LSTM Encoder-Decoder 网络架构

1.2.3 DeepAR

DeepAR 是 Salinas 等^[20]提出的一种自回归多步概率密度预测模型, 其采用 LSTM 单元构成 Encoder-Decoder 架构。与图 2(a)中点预测结果不同, DeepAR 通过极大对数似然函数(9), 直接对随机变量 y 在条件 h 下的概率分布参数 θ 进行估计。本文假设 y 满足高斯分布, 其在参数 $\theta = \{\mu, \sigma^2\}$ (μ 为均值, σ 为标准差) 下的似然函数见式(10)。

$$L = \sum_{i=1}^N \sum_{t=k}^d \log p(y_t^i | \theta(h_t^i)) \quad (9)$$

$$p(y_t | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \mu)^2}{2\sigma^2}\right) \quad (10)$$

式中, N 为批量样本个数, d 为编码器长度。

DeepAR 的 Encoder 与图 2(a) 相同, 其 Decoder 结构见图 3。

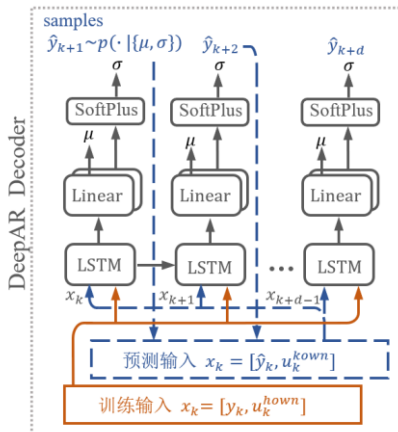


图 3 DeepAR Decoder 架构

此结构中全连接层输出 k 时刻似然参数 $\{\mu, \sigma\}$, 并采用 softplus 激活函数将 σ 转换到 $(0, +\infty)$, 计算公式见 (11), (12)。常规训练阶段, 用(9)式作为损失函数完成对模型参数更新。预测阶段, 根据似然分布 $p(\cdot | \mu, \sigma)$ 进行抽样获得 \hat{y}_k , 并参与构成下一步预测的模型输入。

$$u(h_k) = w_\mu^T h_k + b_\mu \quad (11)$$

$$\sigma(h_k) = \log(1 + \exp(w_\sigma^T h_k + b_\sigma)) \quad (12)$$

式中, h_k 为 k 时刻 LSTM 单元隐藏状态; w 和 b 为全连接网络参数。

1.3 预测值反馈再训练策略

针对 NARX-NN 模型的“曝光偏差”问题, 本文以提高多步预测精度为目标, 提出 FR 策略, 其基本思想为: 利用存在误差的单步预测值重新构造训练样本, 并重新训练模型以增强模型对预测误差的鲁棒性能。其流程如图 4 所示。

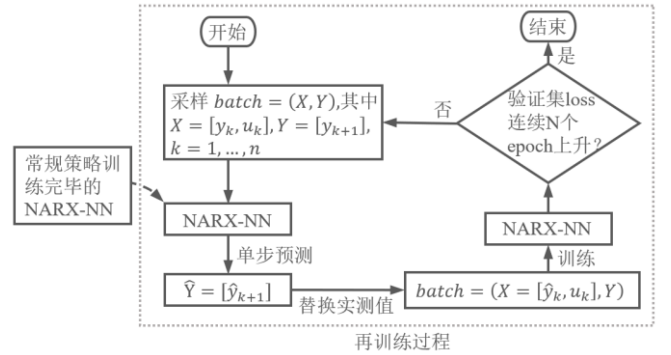


图 4 FR 方法流程图

图中, 首先采用常规训练策略完成对 NARX-NN 的初始训练; 接着对每一批训练样本 $batch = (X, Y)$, 使用 NARX-NN 进行单步预测, 获得预测结果 \hat{Y} ; 然后, 将模型输入 $X = [y_k, u_k]$ 中实测目标值 y_k 替换为对应时刻的单步预测值 \hat{y}_k , 获得新的训练样本 $batch = ([\hat{y}_k, u_k], Y)$; 最后采用 $batch$ 再次训练 NARX-NN。

FR 方法通过不断重复训练-单步预测-样本替换-再训练操作, 直到满足停止规则。需要指出的是, 此处再训练过程中采用单步预测是防止用过大的多步预测误差替换实测目标值后, 造成 $batch$ 偏离原来的数据分布, 破坏模型所表征的统计关系。

本文将常规训练策略和计划采样(SS)策略作为 FR 对比方法应用于多步预测任务中。SS 以一定概率决定训练阶段输入目标变量选择真实值还是上一时刻预测结果。本文选择线性衰减函数^[11]来决定选择真实值的概率, 函数表达式(13)。

$$p_k = \max(0, 1 - ci) \quad (13)$$

式中, p_k 为选择真实值概率, 最小为 0; c 为超参数, 依赖于模型训练期望的收敛速度; i 为模型更新次数。随着 i 增加, p_k 将由 1 以 c 为速率减小到 0。

2. 应用案例

2.1 锅炉 NOx 浓度预测

选择 NARX-DNN 和 LSTM Encoder-Decoder(下文简称为 DNN 和 LSTM)建立锅炉燃烧系统出口 NO_x 浓度的非线性动力学模型,以预测 NO_x 浓度变化趋势。NO_x 浓度数据集为某电厂前后墙对冲锅炉连续运行 7 天的实时燃烧数据,采样周期 40s,共 15000 组,变化范围[200.7, 647.6]mg/m³。

本文数据集均按 7:1.5:1.5 划分训练集、验证集和测试集,NO_x 浓度数据集按时间顺序依次划分,结果见图 5。验证集将用于超参数寻优和早停策略,测试集用于模型最终评价。在训练过程中,为防止过拟合降低模型泛化性能,所有模型均采用早停策略,验证集损失连续 10 个 epoch 不下降时停止训练,并选择损失最小的模型作为最佳模型。

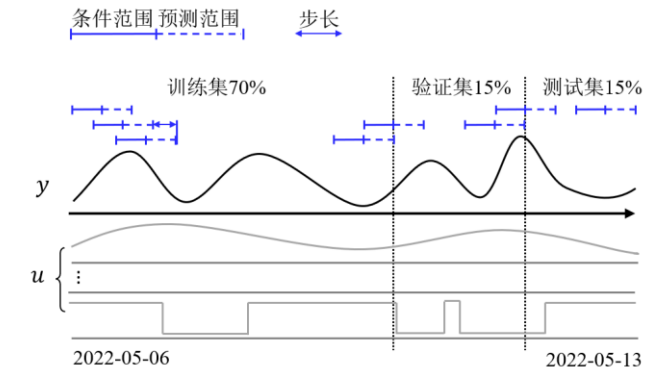


图 5 NO_x 浓度数据集数据划分方式及样本集构建描述

对于 Encoder-Decoder 架构,数据样本的构建方式见图 5 中蓝线,其中蓝色实线代表条件范围,即编码器输入步数,蓝色虚线代表预测范围,即解码器输入步数。在构建验证集和测试集样本时,保证预测范围在对应划分范围内即可。本文采用滑动窗口切分数据,并设置滑动步长为 1 来构建样本集。

2.1.1 输入变量和迟延步数选择

燃烧系统出口 NO_x 浓度受到给煤量(Fuel)、二次风门开度(Sec)、燃烬风门开度(Sofa)和省煤器出口氧量(O₂)的影响,图 6 展示了该系统的输入和输出变量。考虑将基于 NO_x 的多步预测用于实现多步预测控制的应用场景,建模时将控制变量视为已知的协变量信息,状态变量视为常量。

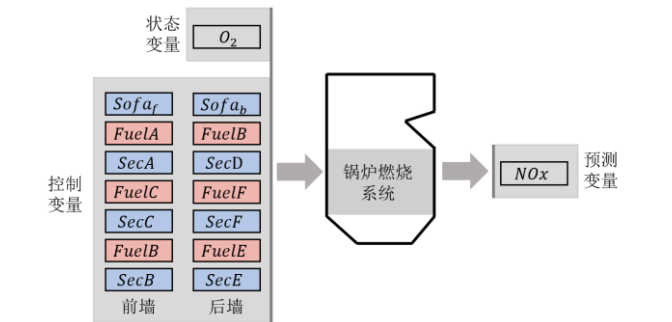


图 6 锅炉燃烧系统 NO_x 浓度预测的输入输出变量

由于为防止火焰偏斜,中层、底层二次风和燃烬风的前后墙风门开度一般相同,故分别采用平均值合

并为 SecCF、SecBE 和 Sofa。因 NO_x 采样值存在较大噪声,因此通过设计巴特沃斯 IIR 滤波器对分辨率为 1s 的初始 NO_x 采样值进行滤波,之后再间隔采样降低为 40s 采样周期,滤波结果见图 7。

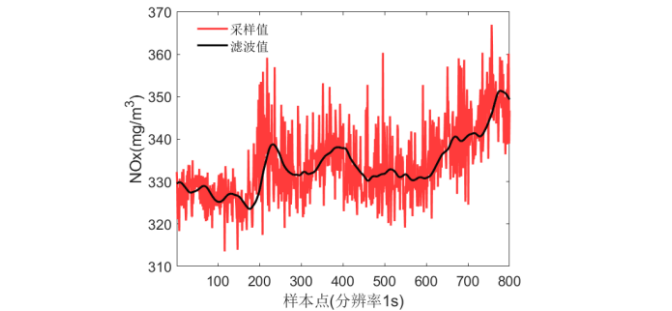


图 7 分辨率为 1s 的 NO_x 采样曲线和滤波结果

考虑到锅炉对象的大时滞特性,采用互信息(MI)算法^[21]确定各输入到输出的最佳迟延步数,如图 8 所示。

Time(s)	Delay	SecA	SecD	SecCF	SecBE	Sofa	O2	FuelA	FuelD	FuelC	FuelF	FuelB	FuelE
40	1	0.0077	0.0089	0.0074	0.0068	0.0129	0.8467	0.0518	0.0564	0.0383	0.0396	0.0996	0.0983
80	2	0.0077	0.0093	0.0079	0.0066	0.0134	0.7004	0.0529	0.0599	0.0395	0.0412	0.1040	0.1014
120	3	0.0078	0.0097	0.0076	0.0069	0.0134	0.6006	0.0530	0.0634	0.0427	0.0438	0.1132	0.1094
160	4	0.0076	0.0093	0.0071	0.0070	0.0133	0.5657	0.0555	0.0656	0.0468	0.0477	0.1259	0.1204
200	5	0.0074	0.0089	0.0071	0.0069	0.0127	0.5667	0.0585	0.0658	0.0494	0.0504	0.1357	0.1293
240	6	0.0074	0.0089	0.0074	0.0067	0.0129	0.5753	0.0587	0.0631	0.0487	0.0500	0.1356	0.1290
280	7	0.0072	0.0086	0.0073	0.0069	0.0130	0.5363	0.0576	0.0609	0.0458	0.0471	0.1276	0.1216
320	8	0.0072	0.0080	0.0071	0.0068	0.0136	0.5130	0.0562	0.0579	0.0430	0.0439	0.1181	0.1131
360	9	0.0073	0.0080	0.0071	0.0068	0.0143	0.5195	0.0552	0.0553	0.0407	0.0419	0.1104	0.1066
400	10	0.0073	0.0079	0.0068	0.0065	0.0159	0.5471	0.0531	0.0526	0.0396	0.0403	0.1058	0.1025
440	11	0.0071	0.0083	0.0065	0.0065	0.0156	0.5570	0.0518	0.0531	0.0393	0.0397	0.1044	0.1020
480	12	0.0076	0.0083	0.0063	0.0067	0.0149	0.5694	0.0517	0.0520	0.0393	0.0394	0.1040	0.1018
520	13	0.0077	0.0081	0.0061	0.0070	0.0144	0.5841	0.0516	0.0515	0.0385	0.0392	0.1027	0.1007
560	14	0.0074	0.0081	0.0059	0.0070	0.0134	0.5971	0.0525	0.0517	0.0386	0.0392	0.1033	0.1007
600	15	0.0072	0.0086	0.0059	0.0069	0.0134	0.5946	0.0519	0.0516	0.0389	0.0389	0.1040	0.1012

图 8 各输入与 NO_x 浓度时间序列在时间延迟坐标下的互信息

图中,Delay 为输入迟延阶数,例如 Delay=5,表示 $u(k-5)$ 序列与 $y(k)$ 之间 MI; 粗体数值为列最大 MI,对应迟延阶数将用于重构 NARX-DNN 输入。

2.1.2 超参数优化

网格搜索技术被用来搜寻各数据集下不同模型的最优超参数,所搜索的超参数空间见表 1。表中 Hidden size 用来决定 DNN 网络初始隐藏层和 LSTM 循环层隐藏单元数量; Layers 决定 DNN 隐藏层和 LSTM 循环层层数。所有模型的激活函数均采用 ReLU, batch size 设置为 128,最大 epoch 为 200,优化算法为 Adam,学习率设置为 0.001。

表 1 模型超参数搜索范围	
超参数	搜索范围
Hidden size	16, 32, 64
Layers	1, 2, 3, 4
Dropout	0.1, 0.2, 0.3

经过网格搜索后的网络结构见表 2。考虑到锅炉燃烧系统出口 NO_x 浓度对各控制量响应的过渡过程时间均在 10min 内,故将 NO_x 的条件范围e和预测范围d均设置为 15。

表 2 NO _x 浓度数据集的网络结构					
模型	Hidden size	Layers	Dropout	e	d
DNN	32	3	0.1	/	/
LSTM	32	2	0.1	15	15

2.1.3 评价指标

用平均绝对误差（MAE）和平均绝对百分比误差（MAPE）来评价模型的点预测精度。对第*h*步预测，它们的定义分别如下：

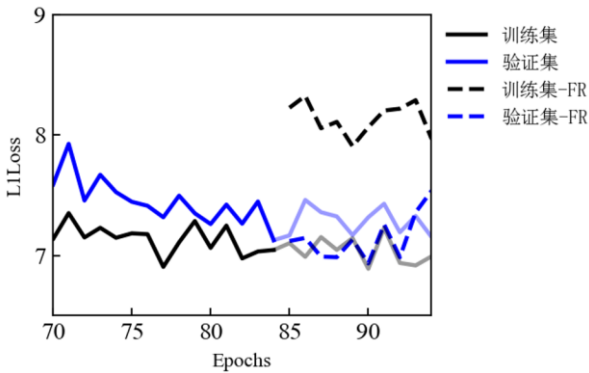
MAE_h = \frac{1}{N} \sum_{i=1}^N |y_{k+h}^i - \hat{y}_{k+h}^i| \tag{14}

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 \tag{15}

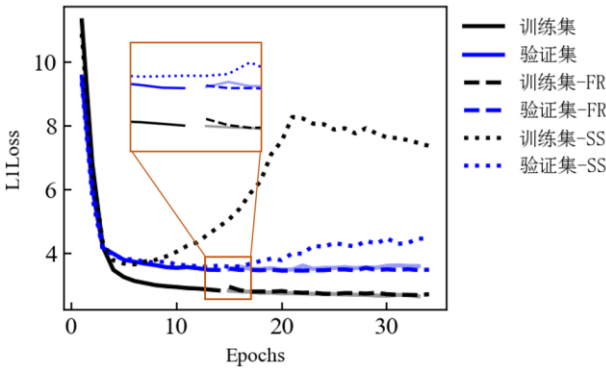
式中，*N*为测试集总样本数。

2.1.4 NOx 浓度多步预测结果

在不同训练策略下，DNN 和 LSTM 在训练集和验证集上的损失变化见图 9，损失函数为 L1 损失。图例中无任何后缀代表常规训练策略；FR 后缀代表预测值反馈再训练策略；SS 后缀代表计划采样策略（下同）。由图可知，DNN 模型在常规训练策略下经过 84 个 epoch 后收敛，然后通过 FP 再训练使训练集损失增加，验证集损失略微下降。LSTM 模型在常规训练策略下经过 14 个 epoch 后收敛，通过再训练后也有类似效果。这表明 FR 策略能在常规训练基础上进一步提升模型的泛化性能。需指出的是，LSTM 在 SS 训练策略下的收敛情况并不理想，其训练集损失随着选择真值概率的降低而先下降后上升，验证集损失经过 12 个 epoch 后收敛，之后出现过拟合。



(a)DNN 不同训练策略下收敛过程



(b)LSTM 不同训练策略下收敛过程

图 9 训练集和验证集收敛过程

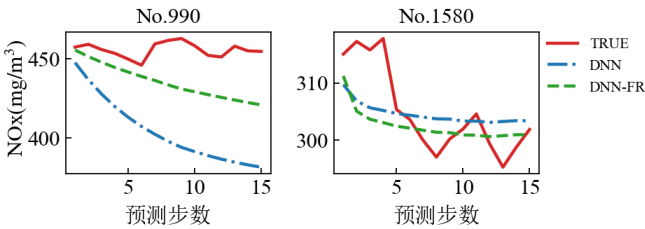
DNN 和 LSTM 在不同训练策略下对 NOx 浓度的多步预测性能见表 3，表中展示了不同预测步数下各模型

对测试集预测结果的 MAE 和 MAPE 指标及均值。图 10(a)展示了 4 组不同测试样本在 DNN 下采用常规训练策略和 FR 策略的前向 15 步预测曲线。图 10(b)展示了与(a)中相同的 4 组样本点在 LSTM 下采用常规训练策略、SS 策略和 FR 策略的前向 15 步预测曲线。

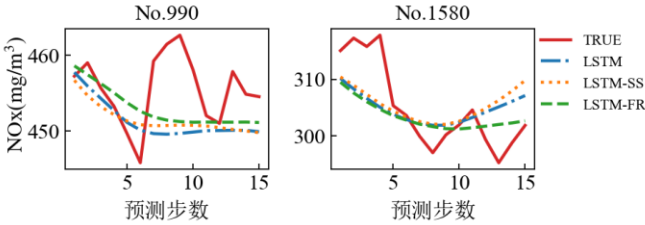
表 3 DNN 和 LSTM 在不同训练策略下

对 NOx 浓度的多步预测精度

模型	测试集				1-15 均值
	预测步数 (h)				
	1	5	10	15	
MAE(mg/m ³)					
DNN	11.56	35.83	42.77	44.81	36.96
DNN-FR	8.27	27.37	37.21	42.58	30.15
LSTM	4.30	12.97	18.88	26.32	16.28
LSTM-SS	4.48	13.61	19.28	23.75	16.23
LSTM-FR	4.17	12.34	17.59	22.15	14.89
MAPE(%)					
DNN	2.87	8.56	10.30	10.84	8.89
DNN-FR	2.05	6.77	9.15	10.32	7.32
LSTM	1.16	3.54	5.16	7.16	4.44
LSTM-SS	1.21	3.74	5.32	6.54	4.47
LSTM-FR	1.12	3.32	4.74	5.97	4.01



(a) DNN 在不同训练策略下预测结果



(b) LSTM 在不同训练策略下预测结果

图 10 NOx 浓度测试集多步预测结果

从上述结果可得出如下结论：

- 1) FR 策略能有效提升 DNN 和 LSTM 对 NOx 浓度的多步预测精度，而 SS 策略并未展示出提升作用。LSTM-FR 在预测范围内的平均 MAE 为 14.89 mg/m³、MAPE 为 4.01%，较常规训练策略分别降低 8.5%、9.7%。同样，DNN-FR 的 MAE 和 MAPE 平均降低 18.4%、17.6%。
- 2) 相比于 DNN，LSTM 在各预测步数下均具有较高预测精度。在预测范围内，LSTM 平均 MAE、MAPE 分别为 16.28 mg/m3、4.44%，较 DNN 降低 56.0%、50.1%。这得益于 LSTM 通过传递隐状态和保留记忆状态，捕获锅炉燃烧系统动态特性。

2.2 用户电负荷预测

选择 DeepAR 模型对综合能源系统中用户电负荷进行多步概率预测。用户电负荷数据来自美国亚利桑那

州立大学的综合能源系统(IES)。选取该 IES 2019 年逐时电负荷数据，并从气象中心下载对应气象数据，结合日期序列构成电负荷数据集，共 8760 组，电负荷变化范围[3.3, 32.5]MW。气象数据包含气温、湿度、气压，日期序列包含时、周、月。

考虑到用户电负荷有较强的月度趋势，将数据集按月份依次划分训练集、验证集和测试集，如图 11 所示，其余设置均同 NOx。

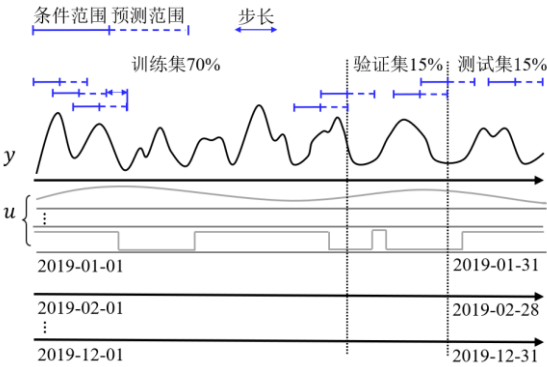


图 11 电负荷数据集数据划分方式及样本集构建描述

2.2.1 Deep AR 模型结构优化

网格搜索空间见表 1，并增加条件范围超参数 e ，搜索范围设置为[24,48,72]。采用DeepAR对电负荷进行前向 24 步概率密度预测，优化得到的模型结构参数见表 4，其余设置同 NOx。

表 4 DeepAR 的优化结构					
模型	Hidden size	Layers	Dropout	e	d
DeepAR	64	2	0.1	72	24

2.2.2 电负荷多步概率预测结果

DeepAR 训练过程损失变化见图 12，损失函数为高斯负对数似然损失。模型在常规训练策略下经 6 个 epoch 收敛，之后采用 FR 策略，训练集损失明显升高，验证集损失则轻微下降。

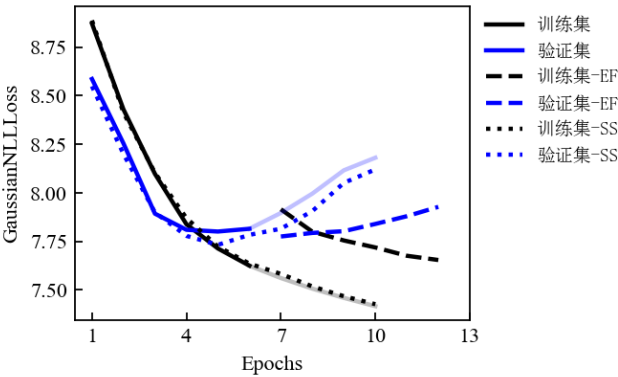


图 12 DeepAR 在不同训练策略下训练集和验证集收敛过程

表 5 展示了 DeepAR 在不同训练策略下对电负荷的多步预测性能。DeepAR 的点预测结果来自于预测分布的中位数，即 0.5 分位数。图 13 绘制了测试集中不同时刻的前向 24 步概率密度预测结果，包括预测范围内的实际电负荷、中位数、99%预测区间、90%预测区间和

50%预测区间。

表 5 DeepAR 在不同训练策略下对电负荷的预测精度

模型	测试集					1-24 均值
	预测步数（ h ）					
	1	6	12	18	24	
MAE(kW)						
DeepAR	423.8	744.4	849.5	886.9	971.6	811.8
DeepAR-SS	427.2	743.0	863.0	878.5	945.0	808.3
DeepAR-FR	430.0	740.3	830.6	853.4	900.8	792.2
MAPE(%)						
DeepAR	2.40	4.21	4.84	4.97	5.48	4.59
DeepAR-SS	2.45	4.20	4.89	4.93	5.35	4.56
DeepAR-FR	2.43	4.18	4.73	4.79	5.10	4.48

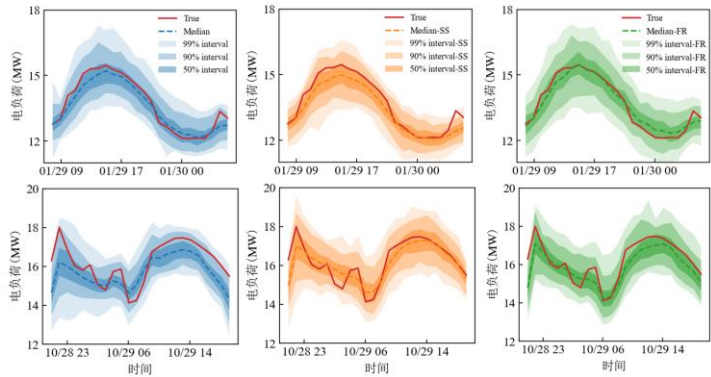


图 13 DeepAR 在不同训练策略下对电负荷测试集不同时刻的多步预测结果

根据表 5 和图 13 中内容，可得出如下结论：FR 策略能进一步提升 DeepAR 对电负荷的点预测精度，且随预测步数增加提升效果越明显。DeepAR-FR 第 12、16、24 步预测结果的 MAE 为 830.6kW、857.4kW 和 900.8kW，较采用常规训练策略的 DeepAR 分别降低 2.3%，3.8%，7.3%。SS 策略未表现出更优异的多步预测性能，这与文献^[20]中实验结论一致。

2.3 显著性检验

神经网络训练过程中因权重随机初始化、训练样本随机抽样等操作导致不同时刻训练的网络会收敛到不同的局部最优解，从而使评价指标在一定范围波动。本节将所有网络在不同随机种子下重复训练 15 次，并采用配对样本 T 检验分析不同训练策略下预测精度差异的显著性。

图 14-图 16 分别展示了 DNN、LSTM 和 DeepAR 预测精度指标的箱线图和 T 检验结果。箭头表示其两侧数据存在显著性差异，且箭头指向均值较低的数据。箭头上方 “*”、“**” 和 “***” 分别表示 T 检验结果的 $p < 0.05$ 、 $p < 0.01$ 和 $p < 0.001$ 。无箭头标注表明预测精度无显著差异。

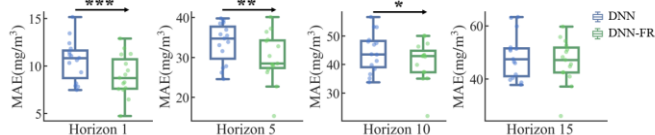


图 14 DNN 在常规训练策略和 FR 下不同预测步数预测精度 (MAE) 差异的显著性

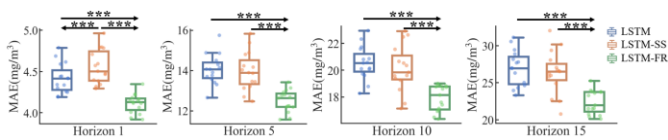


图 15 LSTM 在常规训练策略、SS 和 FR 下
不同预测步数预测精度 (MAE) 差异的显著性

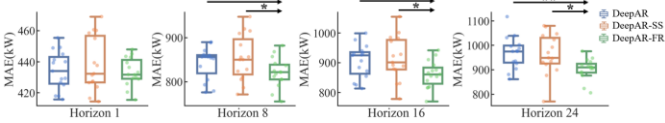


图 16 DeepAR 在常规训练策略、SS 和 FR 下
不同预测步数预测精度 (MAE) 的差异显著性

由图 14 可知, 在 NO_x 浓度预测任务中, FR 策略对 DNN 预测精度的提升程度的显著性随预测步数增大而减弱。这是因为 DNN 网络结构对时序数据预测能力有限。从图 15 中可以看出, FR 策略对擅长处理序列数据的 LSTM 在各预测步数下的 MAE 指标均存在显著性降低 (显著性水平均为 0.001)。同时图中显示 SS 策略并不能显著提升 NO_x 浓度的预测精度, 如图中 Horizon1 的情况甚至导致预测精度下降。由图 16 可知, 在电负荷预测任务中, FR 策略使 DeepAR 的较大预测步数的点预测指标 MAE 显著下降; 而在较小预测步数下, 常规策略、SS 和 FR 无显著性差异。

2.4 FR 训练策略评价

由 2.1-2.3 节中实验和分析可知, FR 策略可应用于任何训练好的 NARX-NN 模型, 对 NARX-DNN、LSTM Encoder-Decoder 和 DeepAR 的多步点预测精度均有显著性提升。通过将预测结果反馈到训练集输入并指导模型重新训练, FR 策略提高了各模型对预测误差的鲁棒性能。

在计算性能方面, FR 较常规训练策略仅多出一个涉及前向传播的单步预测过程, 因此对网络的每次梯度更新, 其计算时间复杂度与常规训练策略相当。在收敛性方面, 上述应用案例中 DNN、LSTM Encoder-Decoder 和 DeepAR 分别经 8、6、4 个 epoch 后收敛, 占总训练时间的约 10%、28% 和 39%。

3. 结论

针对 NARX-NN 常规训练策略, 在训练过程和预测过程中, 由于模型输入的差异导致预测误差易在预测过程中累积的问题, 提出了 FR 训练策略, 以降低网络对预测误差的敏感程度, 从而提高其多步预测精度。针对 NARX-DNN、LSTM Encoder-Decoder 和 DeepAR 三种典型 NARX-NN 模型分别采用 FR 策略进行训练, 并用于对电站锅炉 NO_x 浓度或综合能源系统电负荷进行多步预测。计算结果表明, 相较于常规训练策略和 SS 策略, FR 策略能显著提高各模型的多步点预测精度。

尽管 FR 策略对本文的案例均有显著提升, 但其在

更多数据集上的应用表现还需进一步探究。如果常规训练策略所训练的网络本身具有较大的单步预测误差, 那么所构建的新样本会偏离原始训练集的分布, 从而导致 FR 策略失效。因此接下来还需对“适当”的单步预测误差给予定量分析, 以明确应用 FR 策略的前提条件。

参考文献:

- [1] Ospina J, Newaz A, Faruque M O. Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model[J]. IET Renewable Power Generation, 2019.
- [2] Li Q, Zhang X, Ma T, et al. A Multi-step ahead photovoltaic power forecasting model based on TimeGAN, Soft DTW-based K-medoids clustering, and a CNN-GRU hybrid neural network[J]. Energy Reports, 2022, 8: 10346–10362.
- [3] Zou Y, Wang J, Lei P, et al. A novel multi-step ahead forecasting model for flood based on time residual LSTM[J]. Journal of Hydrology, 2023: 129521.
- [4] 李春祥, 李洲. 基于多种信号分解的台风风速多步预测[J]. 振动. 测试与诊断, 2019, 39(5): 1103-1110+1139.
- [5] Hu Q, Long J, Wang S, et al. A novel time-span input neural network for accurate municipal solid waste incineration boiler steam temperature prediction[J]. Journal of Zhejiang University-SCIENCE A, 2021, 22(10): 777–791.
- [6] Smrekar J, Potočník P, Senegačnik A. Multi-step-ahead prediction of NO_x emissions for a coal-based boiler[J]. Applied Energy, 2013, 106: 89–99.
- [7] Yuan X, Li L, Wang Y. Nonlinear Dynamic Soft Sensor Modeling With Supervised Long Short-Term Memory Network[J]. IEEE Transactions on Industrial Informatics, 2020, 16(5): 3168–3176.
- [8] Lim B, Arık S Ö, Loeff N, et al. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting[J]. International Journal of Forecasting, 2021, 37(4): 1748–1764.
- [9] 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.
- [10] Shi X, Yeung D-Y. Machine Learning for Spatiotemporal Sequence Forecasting: A Survey[J]. arXiv, 2018.
- [11] Bengio S, Vinyals O, Jaitly N, et al. Scheduled sampling for sequence prediction with recurrent Neural networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. Cambridge, MA, USA: MIT Press, 2015: 1171–1179.
- [12] Meng F, Pei H. Cyclic shear stress-strain prediction of saturated sand based on the unrolled seq2seq model and scheduled sampling[J]. Soil Dynamics and Earthquake Engineering, 2023, 165: 107665.
- [13] Li Y, Yu R, Shahabi C, et al. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting[J]. arXiv, 2018.

- [14] Mihaylova T, Martins A F T. Scheduled Sampling for Transformers[J]. arXiv, 2019.
- [15] Haris M, Noman Hasan M, Qin S. State of health prediction of supercapacitors using multi-trend learning of NARX neural network[J]. Materials Today Sustainability, 2022, 20: 100201.
- [16] 黄伟建, 李永涛, 黄远. 基于混合神经网络和注意力机制的混沌时间序列预测[J]. 物理学报, 2021, 70(1): 235–243.
- [17] Di Piazza A, Di Piazza M C, Vitale G. Solar and wind forecasting by NARX neural networks[J]. Renewable Energy and Environmental Sustainability, 2016, 1: 39.
- [18] Sutskever I, Vinyals O, Le Q V. Sequence to Sequence Learning with Neural Networks[J]. arXiv, 2014.
- [19] Chung J, Gulcehre C, Cho K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. arXiv, 2014.
- [20] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: Probabilistic forecasting with autoregressive recurrent networks[J]. International Journal of Forecasting, 2020, 36(3): 1181–1191.
- [21] 唐振浩, 朱得宇, 李扬. 基于数据驱动的燃煤锅炉 NO_x 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.
- [22] Seeger M W, Salinas D, Flunkert V. Bayesian Intermittent Demand Forecasting for Large Inventories[C]//Advances in Neural Information Processing Systems. Curran Associates, Inc., 2016, 29.