

# 基于 Transformer 模型的锅炉燃烧 NO<sub>x</sub> 排放多步预测

**摘要:** 火电机组锅炉燃烧过程会排放的大量氮氧化物(NO<sub>x</sub>), 对 NO<sub>x</sub> 浓度进行前向多步预测有助于优化燃烧过程。为实现精准的 NO<sub>x</sub> 预测, 本文采用 Transformer 模型架构建立了某 660MW 锅炉燃烧系统的 NO<sub>x</sub> 多步预测模型, 并优化了模型结构超参数, 包括编码器和解码器层数、特征升维后的维度和多头自注意力机制的头数等。结构优化后的 Transformer 模型在测试集上的前向 15 步预测的平均 MAE、MAPE 和 R<sup>2</sup> 分别为 10.999 mg/m<sup>3</sup>、7.715 %和 0.703。相比于目前广泛应用的 LSTM 以及 SVM 预测方法, Transformer 模型在 NO<sub>x</sub> 前向多步预测方面具有更高的多步预测精度。

**关键词:** NO<sub>x</sub> 排放、Transformer 模型、自注意力机制、多步预测

## Multi-step Prediction of NO<sub>x</sub> Emissions from Boiler Combustion based on Transformer Model

**Abstract:** Thermal power unit boilers emit a large amount of nitrogen oxides (NO<sub>x</sub>) during combustion, and forward multi-step prediction of NO<sub>x</sub> concentration helps to optimize the combustion process. In order to achieve accurate NO<sub>x</sub> prediction, this paper uses the Transformer model architecture to establish a NO<sub>x</sub> multi-step prediction model for a 660MW boiler combustion system, and optimizes the hyperparameters of the model structure, including the number of encoder and decoder layers, the dimension after feature upgrading, and the number of heads of the multi-head self-attention mechanism. The average MAE, MAPE and R<sup>2</sup> predicted by the Transformer model in the forward 15 steps on the test set were 10.999 mg/m<sup>3</sup>, 7.715 % and 0.703, respectively. Compared with the widely used LSTM and SVM prediction methods, the Transformer model has higher multi-step prediction accuracy in NO<sub>x</sub> forward multi-step prediction.

**Key Words:** NO<sub>x</sub> emission; Transformer model; Self-attention mechanisms; Multi-step forecasting

### 1 引言

氮氧化物(NO<sub>x</sub>)是大气中主要的气态污染物之一, 在中国已被严格限制排放<sup>[1]</sup>。火电机组锅炉燃烧过程会产生大量的 NO<sub>x</sub> 排放, 直接排向大气会对环境造成严重的影响, 因此准确预测 NO<sub>x</sub> 排放量, 并指导调整操作量来减少 NO<sub>x</sub> 排放就显得尤为重要。

由于燃煤锅炉炉膛内复杂的化学反应和换热过程, 基于能量、质量和动量守恒的第一性原理对炉膛出口 NO<sub>x</sub> 排放的动态预测仍然较为困难<sup>[2]</sup>。神经网络<sup>[3]</sup>、支持向量机<sup>[4]</sup>等机器学习则不需要求解复杂的守恒方程, 其通过数据驱动的方法从历史数据中提取操作变量和目标变量之间的相互关系, 构造输入输出模型, 从而达到预测输出的目的。

Cass 和 Radl 在 20 世纪 90 年代引入了人工神经网络来模拟 NO<sub>x</sub> 排放<sup>[5]</sup>, Ilamathi 等<sup>[6]</sup>利用人工神经网络建立了 210 MW 煤粉锅炉满负荷工况下的 NO<sub>x</sub> 排放模型, 发现所提出的方法可以在运行工况下生成可行的操作条件。李鹏辉等<sup>[7]</sup>利用 BP 神经网络对 NO<sub>x</sub> 排放进行建模, 该模型具有较强的泛化能力, 但预测精度较差。如今, 随着深度学习在数据驱动建模领域的快速发展, 使得训练深层次、非线性拟合能力更强的神经网络成为可能。循环神经网络(RNN)作为一种流行的深度学习方法, 它沿着序列形成一个有向图, 节点之间有连接, 从而可以跟踪时间序列问题的动态时间行为<sup>[8]</sup>。长短期记忆(LSTM)神经网络通过引入记忆元和门控结构, 解决了 RNN 模型存在的长期信息保存和短期输入缺失的问题<sup>[9]</sup>。LSTM 作为一种改进型递归神经网络已经应用于 NO<sub>x</sub> 预测问题, 并取得了一定的成果<sup>[10]</sup>。除人工神经网络外, 支持向量机(SVM)<sup>[11, 12]</sup>也被广泛用于建立煤粉电厂的 NO<sub>x</sub> 排放模型, 与人工神经网络相比, 支持向量机能够克服前者训练时间长且训练结果存在随机性等缺点, 但其很难捕捉样本之间的时序信息。

相较于其他机器学习算法, Transformer 模型出现时间较晚, 由谷歌团队在 2017 提出<sup>[13]</sup>。该模型采用编码器-解码器架构, 且完全基于自注意力机制, 不包含任何卷积层或循环网络层。值得注意的是, 自注意力机制同时具有高并行计算能力和最短最大路径长度这两个优势, 因此, 相较于其他深度神经网络, Transformer 模型具有较强的并行计算能力且能更轻松地学习序列中的远距离依赖关系。尽管 Transformer 最初是应用于文本数据上的序列到序列学习, 但现在已经推广到各种现代的深度学习中, 例如语言、视觉和强化学习领域。Liu 等人提出基于 Transformer 的胶囊网络模型, 通过提取社交媒体的语义特征预测股价走势<sup>[14]</sup>, 证明了 Transformer 应用于回归任务的可能性。

本文采用 Transformer 模型对 660MW 煤粉电站锅炉的炉膛出口 NO<sub>x</sub> 排放浓度进行了预测。利用某实际电站锅炉连续运行 7 天共 15000 组样本建立了基于 Transformer 模型结构的 NO<sub>x</sub> 多步预测模型, 并对模型结构的超参数进

行网格搜索寻优，在此基础上，与目前广泛应用于锅炉燃烧系统的 SVM 和 LSTM 多步预测方法进行了对比研究。

## 2 Transformer 模型

Transformer 模型基于编码器-解码器结构，该结构起初用于输入输出数量不匹配问题，后逐渐被应用于各个领域。

### 2.1 位置编码

LSTM 等递归神经网络对输入数据进行按序迭代计算，而 Transformer 模型中的自注意力机制因采用并行计算方式忽略了输入数据的时序关系。对时间序列问题，时序关系往往反映了因果机理，对模型输出具有重要影响。为融合数据自身的时序信息，Transformer 在输入特征中添加了位置编码<sup>[13]</sup>，位置编码的计算方法如下：

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)$$

其中， $pos$ 代表位置， $i$ 代表第 $i$ 个特征， $d_{model}$ 代表对输入特征进行升维转换后的维度， $i$ 的取值范围是 $[0, d_{model}/2]$ 。位置嵌入函数的周期从 $2\pi$ 到 $10000 * 2\pi$ 变化，而每一位在不同特征维度上都会得到不同周期的 $\sin$ 和 $\cos$ 函数的取值组合，从而产生唯一的位置信息，以便模型学习到样本之间的时序关系。

### 2.2 自注意力机制

注意力机制是指从大量信息中筛选出少量重要信息，忽略大多数不重要的信息，并聚焦到重要信息上。自注意力机制是注意力机制的变体，其减少了对外部信息的依赖，更擅长捕捉数据或特征的内部相关性。Transformer 的核心是自注意力机制，自注意力机制通过三组不同的线性变换分别将输入矩阵转换成查询矩阵 $Q$ ，键矩阵 $K$ 和值矩阵 $V$ ，其计算公式如下：

$$Q = Linear(X) = XW^Q$$

$$K = Linear(X) = XW^K$$

$$V = Linear(X) = XW^V$$

其中，输入矩阵 $X \in R^{n \times d}$ ， $n$ 为样本数， $d$ 为特征维度， $W^Q$ 、 $W^K$ 、 $W^V$ 为权重矩阵。

本研究采用缩放点积注意力计算，首先使用 $Q$ 和 $K$ 做点积，为防止内积过大，将点积后的结果除以 $\sqrt{d}$ 进行缩放操作，最后通过 softmax 函数计算将结果转化为概率矩阵，该矩阵反映了输入样本之间的相关程度，并与值矩阵 $V$ 做点积从而得到最终的输出矩阵，计算公式如下：

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

其中， $Q \in R^{n \times d_k}$ ， $K \in R^{n \times d_k}$ ， $V \in R^{n \times d_k}$ ， $d_k$ 为权重矩阵每一行的维度。计算过程如图 1 所示。

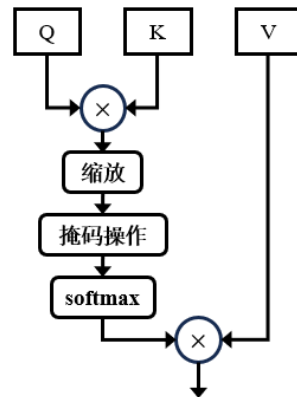


图 1 缩放点积注意力机制

### 2.3 多头自注意力机制

多头自注意力机制旨在基于相同的注意力机制学习到不同的样本信息并将不同的信息组合起来。具体计算过程如下：给定相同的查询、键和值时，通过多组不同的线性变换来变换 $Q$ 、 $K$ 、 $V$ ，并将其并行地送入到缩放点积注意力运算中，最后将多组注意力计算的输出拼接在一起以产生最终输出。这就是多头自注意力机制。

$$MutiHead(Q,K,V) = Concat(head_1, head_2, ..., head_h)W^0$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

其中， $W^0$ 是多头自注意力矩阵拼接后的线性变换矩阵， $W_i^Q$ 、 $W_i^K$ 、 $W_i^V$ 分别是 $Q$ 、 $K$ 、 $V$ 第 $i$ 个头的线性变换权重矩阵。多头自注意力机制计算过程如图 2 所示。

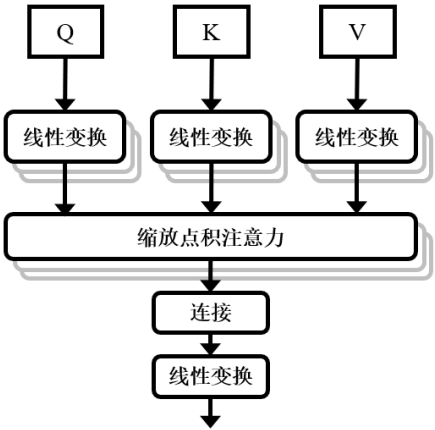


图 2 多头自注意力机制

### 2.4 Transformer 结构及原理

Transformer 的编码器由多个相同的层叠加而成，每个层包含两个子层，其中第一个子层是多头自注意力层，第二个子层是基于位置的前馈网络层，受残差网络(ResNet)<sup>[15]</sup>的启发，每个子层均采用了残差连接和层规范化，目的是避免模型在训练过程中出现梯度消失或梯度爆炸等问题。需要指出的是，在计算编码器自注意力时，查询、键和值都来自前一个编码器层的输出。解码器也是由多个相同的层叠加而成，除了编码器中描述的两个子层外，解码器还在两个子层之间插入了第三个子层，称为编码器-解码器注意力层，在该层中，查询来自前一个解码器层的输出，而键和值来自整个编码器的输出，而在解码器自注意力计算中，查询、键和值都来自于上一个解码器层的输出。为使 Transformer 模型能够达到预测锅炉燃烧系统 NOx 排放的目的，本研究对原始的 Transformer 模型架构进行了一定程度的改变，在整体编码器-解码器结构不变的情况下，由于编码器和解码器的样本输入为特征向量形式，因此该模型略去 embedding 步骤。其次，由于该模型用于回归任务，因此在模型的输出上并未使用 softmax 将输出值映射成概率形式，而是通过全连接层用线性函数来拟合预测值。本研究所采用的 Transformer 模型结构如图 3 所示。

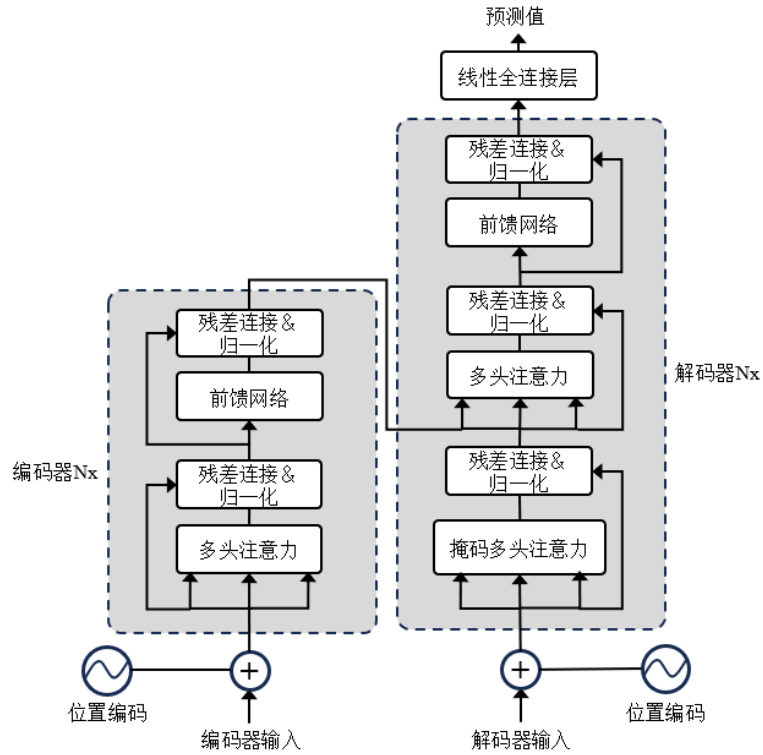


图 3 Transformer 结构图

本文针对锅炉 NO<sub>x</sub> 多步预测问题所设计的 Transformer 模型工作流程如下：将一定时间内的历史输入输出样本添加位置编码后输入到编码器中，输入信息通过多头自注意力机制来捕获历史样本上下文信息，将注意力机制学习到的结果与原始输入结合后再送入前馈网络中进一步学习，前馈网络输出的结果同样与前馈网络输入进行融合得到第一个编码器输出，将上一个编码器输出作为下一个编码器的输入，依次类推，经过多个编码器后得到最终的编码器输出；将未来预测时域内的样本输入变量集合添加位置编码后输入到解码器中，输入信息经过掩码多头注意力计算和残差连接归一化得到输出结果，其中掩码操作可以避免在预测当前时刻 NO<sub>x</sub> 值时用到未来时刻的输入变量。将该输出结果与整个编码器输出分别作为查询、键值对送入包含残差连接及归一化的多头注意力层和前馈网络层中，经过多个解码器后将输出送入线性全连接层，最后得到目标变量的预测结果。

### 3 NO<sub>x</sub> 多步预测建模

#### 3.1 数据准备与数据预处理

本文的研究样本取自某电厂 660WM 超临界燃煤直流锅炉。该锅炉配备六台中速磨煤机，采用直吹式正压冷一次风制粉系统和前后墙对冲燃烧方式，前后墙各三层煤粉燃烧器（上层 A、D，中层 C、F，下层 B、E），每层 5 只，共 30 只旋流煤粉燃烧器。二次风喷口（AA、BB、CC、DD、EE、FF）与燃烧器分布方式相同。在最上层燃烧器上方，前后墙各布置 1 层燃尽风口。

从该电站锅炉的监控信息系统中获取 15000 个连续样本，采样间隔为 40 秒，时间跨度约为 7 天。采用该样本数据对基于 Transformer 的 NO<sub>x</sub> 排放模型进行训练和测试。如图 4 所示，所选样本涵盖了锅炉机组稳定运行以及负荷变化频繁的范围。

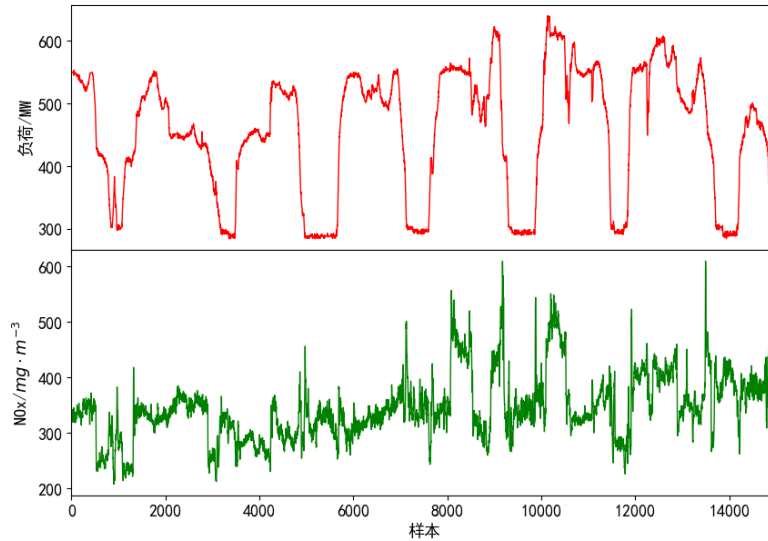


图 4 采样样本负荷及 NOx 排放浓度变化

样本数据分为两部分，将前 75%的数据设置为训练集，后 25%的数据设置为测试集。此划分保证了训练集和测试集都包含不同的运行条件，避免了不同样本分布对模型精度和泛化性的影响。

为提高数据质量，使数据更适合用于训练和测试模型，本文对原始数据进行了数据预处理。首先，为简化模型输入且避免重复特性，对存在多个测点的变量求取均值作为代表，如二次风、燃尽风门开度和氧量。其次，对数值波动较为剧烈的变量进行滤波处理，如目标变量 NOx。由于不同变量间的数值和单位差异会对模型性能产生影响，因此对所有变量依据其变化范围规范化到[0,1]之间。影响燃煤锅炉 NOx 排放的因素有很多，结合以上预处理过程，在本研究中，模型的输入考虑燃烧系统的 15 个主要输入变量，变量信息见表 1。

表 1 建模变量

	变量	说明
输入变量 $u$	LOAD	机组负荷
	A、B、C、D、E、F	6 台给煤机给煤量百分比
	AA、BB、CC、DD、EE、FF	3 层前、后墙二次风门开度
	SOFA	前、后墙燃烬风门开度均值
	O <sub>2</sub>	省煤器出口氧量设定
目标变量 $y$	NOx	燃烧系统出口，脱硝入口 NOx 浓度

### 3.2 训练集构建及模型参数设置

针对 Transformer 模型，考虑到 NOx 浓度对各操作量的响应过渡时间均在 10min 内，因此将编码器的输入样本维度 enc\_in 设置为 15，特征维度 enc\_feature 为 16。为进行间隔为 40s 的前向 10min 多步预测，解码器输入样本维度 dec\_in 同样设置为 15，由于在实际多步预测过程中设置锅炉负荷和氧量保持不变，且忽略 NOx 自回归项可能会更好地捕捉 NOx 的动态性能，因此解码器输入特征维度 dec\_feature 设置为 13。因此在训练过程中，编码器输入表示为 $[u_{t-15:t-1}, y_{t-15:t-1}]$ ，解码器输入表示为 $[u'_{t:t+14}]$ ，其中 $u'$ 表示去除LOAD和O<sub>2</sub>后的输入变量集合，真实标签为 $[y_{t+1:t+15}]$ 。

考虑模型应用于 NOx 排放多步预测任务，模型中的常见训练参数见表 2。

表 2 Transformer 模型训练参数配置

参数名称	值
batch_size	128
epoch	150
激活函数	ReLU
优化算法	Adam
学习率	0.002
损失函数	MSELoss

对于模型超参数，采用网格搜索技术进行寻优，编码器和解码器层数的寻优区间为[3,5,7]，升维转化后维度的

寻优区间设置为[32,64,128], 多头自注意力机制头数的寻优区间为[4,6,8], 通过对超参数间的不同组合进行寻优, 寻优结果如下: 编码器层数和解码器层数 $num\_layers = 3$ , 样本特征升维后的特征维度为 $d_{model} = 64$ , 多头自注意力机制头数为 $num\_heads = 8$ 。

### 3.3 评价指标

采用平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE) 和拟合优度 ( $R^2$ ) 来评估各模型的预测精度, 其中拟合优度是模型拟合样本数据的程度, 反应了模型的整体预测效果, 取值范围为[0,1], 其值越接近 1, 代表模型拟合效果越好。它们的定义如下:

$$MAE_h = \frac{1}{N} \sum_{i=1}^N |y_{t+h}^i - \hat{y}_{t+h}^i|$$

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

$$R^2_h = 1 - \frac{\sum_{i=1}^N (y_{t+h}^i - \hat{y}_{t+h}^i)^2}{\sum_{i=1}^N (y_{t+h}^i - \bar{y}_{t+h})^2}$$

式中,  $h$ 为前向预测步数,  $N$ 为评估时域内的总样本数。

## 4 结果与讨论

### 4.1 Transformer 模型预测结果

经过训练集训练后, Transformer 模型在测试集上的 NOx 排放多步预测值和实际值的对比如图 5 所示。图中的红色虚线表示理想情况下与测量值一致的预测结果, 没有误差。蓝点代表 Transformer 模型预测值。该图展示了在测试集上 Transformer 模型对 NOx 的第 1、5、10、15 步的前向预测结果, 可见所有蓝点都紧密地分布在红色虚线附近, 这表明 Transformer 模型具有较好的 NOx 排放多步预测能力。

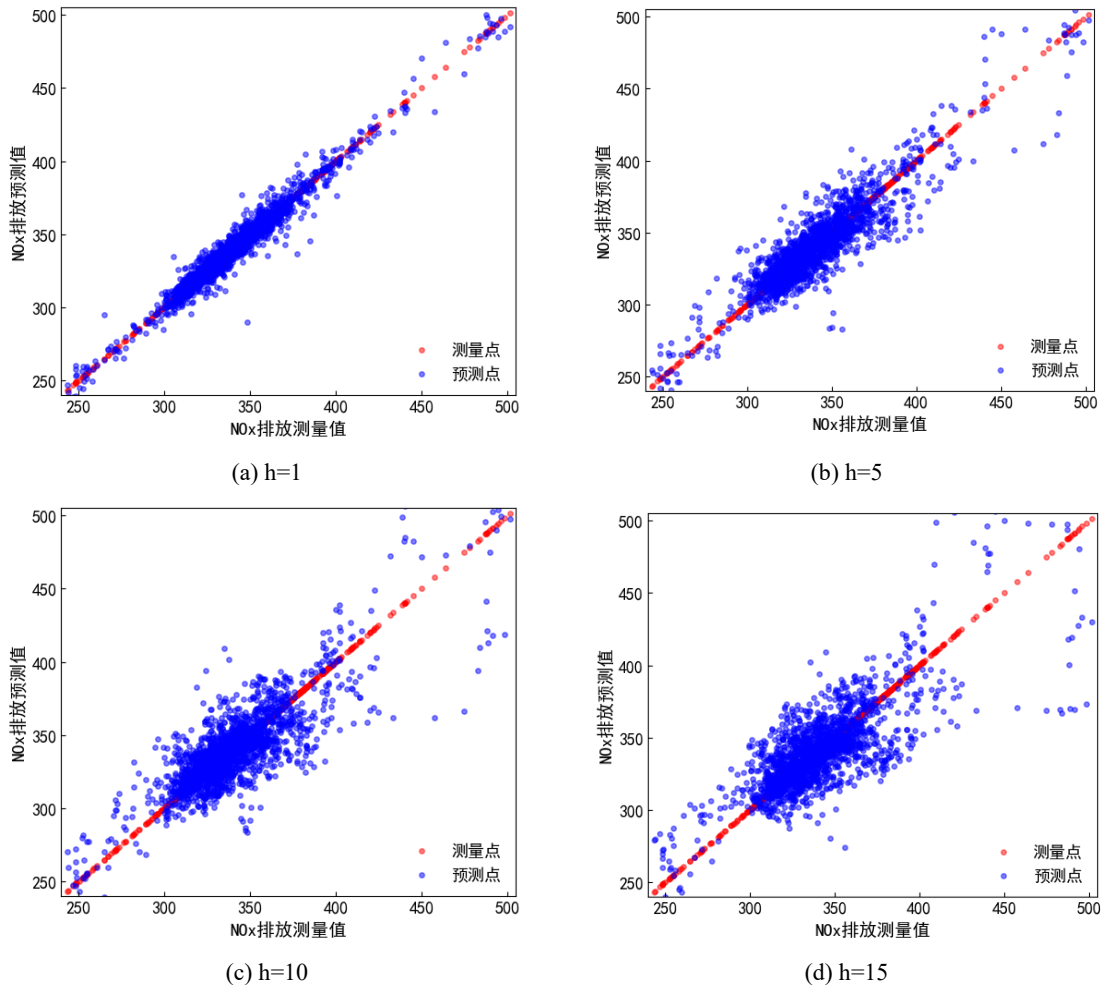


图 5 Transformer 模型分别在第 1、5、10、15 步的预测值与真实值对比



4.2 与 LSTM 以及 SVM 模型的预测效果对比

为了证明 Transformer 模型在火电厂 NOx 排放预测中的性能，本文将目前广泛应用的 LSTM 模型以及 SVM 模型与 Transformer 模型进行了对比研究。

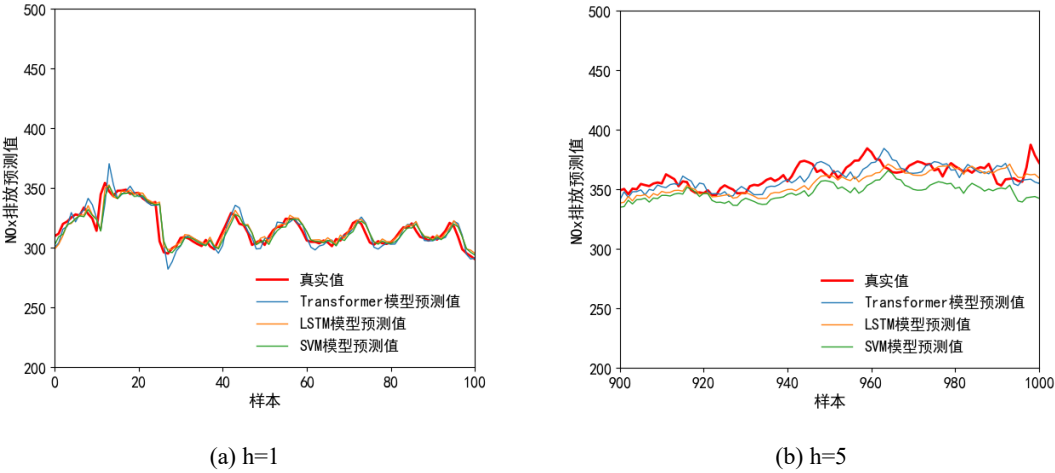
LSTM 模型可以通过存储单元中的门结构自动确定延迟时间顺序<sup>[16]</sup>。因此，输入输出数据集可表示为  $[u_{t-k:t}; y_{t+1}]$ 。采用网格搜索技术优化 LSTM 模型的超参数:隐藏层数 $hide\_layers = 2$ ,隐藏单元数 $hide\_size = 32$ 。对于其他常见设置，激活函数为 ReLU，batch size 为 16，迭代周期为 200 时，训练集和测试集的误差基本保持不变，故 epoch 设置为 200，采取的优化算法为 Adam，学习率为 0.001，损失函数为 MSELoss。

对于 SVM 模型，各输入变量与目标变量之间的最佳迟延系数根据最大互信息系数<sup>[17]</sup>确定。因此 $t$ 时刻训练样本输入表示为  $X = [LOAD(t), O_2(t), A(t - 6), B(t - 4), C(t - 5), D(t - 5), E(t - 5), F(t - 5), AA(t - 5), BB(t - 4), CC(t - 4), DD(t - 2), EE(t - 5), FF(t - 6), SOFA(t - 4), NO_x(t)]$ ，输出  $Y = [NO_x(t + 1)]$ 。SVM 模型采用径向基函数，样本容量为 1000，通过训练集样本等距抽样取得。惩罚系数 $C$ 和核参数 $g$ 采用网格搜索技术得到最优值，搜索范围均为 $[1e-3, 1e2, \cdots, 1e2, 1e3]$ ，优化结果为 $C = 1000$ ， $g = 0.01$ 。

采用参数优化后的 LSTM 模型以及 SVM 模型实现对 NOx 排放的前向 15 步预测。Transformer、LSTM 以及 SVM 三种预测模型的预测结果如表 3 所示，该表展示了不同预测步数和预测区间下三种模型的 MAE 和 MAPE 以及  $R^2$ 。

表 3 Transformer、LSTM、SVM 模型前向多步预测结果对比										
模型	评价指标	前向预测步数 (h)					前向预测范围 (1-h)			
		1	2	5	10	15	1-5	6-10	11-15	1-15
SVM	MAE (mg/m³)	4.033	6.814	13.083	20.040	25.791	8.866	17.426	23.548	16.613
LSTM		3.836	6.661	12.613	17.538	20.671	8.631	15.914	19.447	14.664
Transformer		4.391	3.606	9.273	13.311	15.626	6.185	12.168	14.642	10.999
SVM	MAPE (%)	2.871	4.783	9.006	13.334	17.034	6.164	11.737	15.542	11.148
LSTM		2.732	4.706	8.872	12.217	14.451	6.089	11.113	13.566	10.256
Transformer		3.202	2.623	6.681	9.265	10.749	4.477	8.576	10.093	7.715
SVM	R²	0.964	0.898	0.635	0.173	-0.302	0.809	0.358	-0.117	0.350
LSTM		0.965	0.896	0.641	0.308	0.078	0.808	0.427	0.166	0.467
Transformer		0.957	0.970	0.819	0.617	0.446	0.908	0.685	0.516	0.703

由表 3 可知，LSTM 模型在测试集上的单步预测精度最高，MAE 达到 3.836 mg/m³，相应的 MAPE 和  $R^2$  分别为 2.732 %和 0.965。SVM 模型的多步预测效果最差，这可能是由于该模型未考虑输入变量间的时序特性。相较其他两个模型，Transformer 模型在单步预测精度上没有优势，但该模型在 NOx 前向多步预测方面的整体精度上远高于其他两个模型，前 5 步平均 MAE 为 6.185 mg/m³，相应的 MAPE 和  $R^2$  分别为 4.477 %和 0.908，预测时域内的平均 MAE 达到 10.999 mg/m³，相应的 MAPE 为 7.715 %， $R^2$  为 0.703，接近于 1，可见 Transformer 模型对时间序列前向多步的拟合效果更好，对燃煤锅炉 NOx 排放具有更强的多步预测能力。



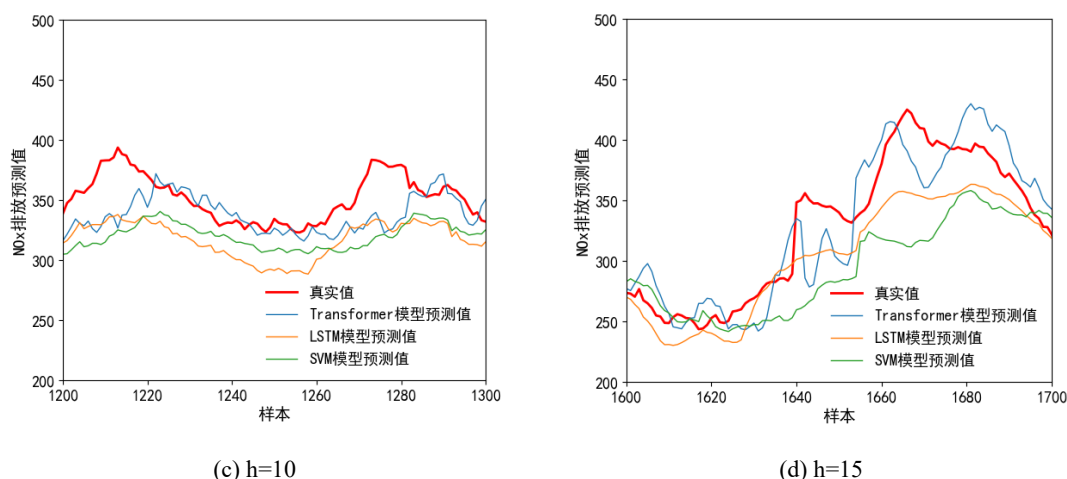


图 6 三种模型在预测步数为 1、5、10、15 时的预测曲线对比

图 6 直观地展示了各个模型的多步预测效果，三种模型的单步预测曲线与真实值曲线几乎重合，说明三种对比模型的单步预测精度均较高，Transformer 模型的单步预测精度略低一点，但其多步预测精度明显高于 LSTM 和 SVM 模型，该模型更好地拟合了当前预测步数的真实值曲线。

## 5 结论

为实现精准的锅炉燃烧系统 NO<sub>x</sub> 排放多步预测，本文提出一种基于 Transformer 结构的 NO<sub>x</sub> 预测模型，并利用 660MW 燃煤电厂锅炉的实际运行数据来构建输入输出集并预测 NO<sub>x</sub> 排放，对模型的编码器和解码器层数、样本特征升维后的特征维度、多头自注意力机制头数和学习率等超参数进行了优化，通过对模型进行训练，该模型在测试集上的多步预测性能表现良好，前 5 步平均 MAE 为 6.185 mg/m<sup>3</sup>，相应的 MAPE 为 4.477%，R<sup>2</sup> 为 0.908，未来 15 步预测的平均 MAE 为 10.999 mg/m<sup>3</sup>，相应 MAPE 为 7.715%，R<sup>2</sup> 为 0.703。进一步比较了目前广泛使用的 SVM 及 LSTM 序列模型预测方法，结果表明 Transformer 对时间序列信息的利用能力远强于 SVM 及 LSTM 模型，多步预测精度远高于其他两个对比模型。

综上所述，本文所研究的 Transformer 模型预测方法在 NO<sub>x</sub> 排放多步预测问题上具有较高的精度和可靠性，随着 DCS 系统的广泛部署和实时数据采集能力的增强，该方法具有良好的应用前景。

## 参考文献

- [1] 胡秀莲. 中国电力生产及环境问题 [J]. 中国能源, 2005, (11): 15-21.
- [2] 卓建坤, 焦伟红, 宋少鹏, et al. 锅炉燃烧优化中 NO<sub>x</sub> 预测模型研究进展 [J]. 燃烧科学与技术, 2016, 22(06): 531-40.
- [3] CHU J-Z, SHIEH S-S, JANG S-S, et al. Constrained optimization of combustion in a simulated coal-fired boiler using artificial neural network model and information analysis☆ [J]. Fuel, 2003, 82(6): 693-703.
- [4] 李楠, 卢钢, 李新利, et al. 基于集成深度玻尔兹曼机和最小二乘支持向量回归的燃烧过程 NO<sub>x</sub> 预测算法 [J]. 动力工程学报, 2016, 36(08): 615-20.
- [5] CASS R, RADL B. Adaptive process optimization using functional-link networks and evolutionary optimization [J]. Control Engineering Practice, 1996, 4(11): 1579-84.
- [6] BALAMURUGAN I, SELLADURAI V, KULENDRAN B, et al. ANN-GA approach for predictive modeling and optimization of NO<sub>x</sub> emission in a tangentially fired boiler [J]. Clean Technologies and Environmental Policy, 2012, 15: 125-31.
- [7] 李鹏辉, 刘冉, 余廷芳. 基于支持向量机和 BP 神经网络的燃煤锅炉 NO<sub>x</sub> 排放预测 [J]. 热能动力工程, 2016, 31(10): 104-8+29-30.
- [8] KOMBRINK S, MIKOLOV T, KARAFIÁT M, et al. Recurrent Neural Network Based Language Modeling in Meeting Recognition [M]. 2011.
- [9] HOCHREITER S, SCHMIDHUBER J. Long Short-term Memory [J]. Neural computation, 1997, 9: 1735-80.
- [10] 杨国田, 张涛, 王英男, et al. 基于长短期记忆神经网络的火电厂 NO<sub>x</sub> 排放预测模型 [J]. 热力发电, 2018, 47(10): 12-7.
- [11] LV Y, LIU J, YANG T, et al. A novel least squares support vector machine ensemble model for NO<sub>x</sub> emission prediction of a coal-fired boiler [J]. Energy, 2013, 55: 319-29.



- [12] 牛培峰, 麻红波, 李国强, et al. 基于支持向量机和果蝇优化算法的循环流化床锅炉 NO<sub>x</sub> 排放特性研究 [J]. 动力工程学报, 2013, 33(04): 267-71.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [J]. Adv Neur In, 2017, 30.
- [14] LIU J, LIN H, LIU X, et al. Transformer-Based Capsule Network For Stock Movement Prediction; proceedings of the Proceedings of the First Workshop on Financial Technology and Natural Language Processing, F, 2019 [C].
- [15] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition; proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, F, 2016 [C].
- [16] TAN P, HE B, ZHANG C, et al. Dynamic modeling of NO X emission in a 660 MW coal-fired boiler with long short-term memory [J]. Energy, 2019, 176: 429-36.
- [17] ALBANESE D, RICCADONNA S, DONATI C, et al. A practical tool for Maximal Information Coefficient analysis [J]. GigaScience, 2018, 7.