

基于Wavenet的燃煤锅炉NO_x排放浓度预测模型

段俏玉

长沙理工大学, 数学与统计学院, 湖南 长沙

收稿日期: 2023年11月13日; 录用日期: 2023年12月7日; 发布日期: 2023年12月15日

摘要

基于数据驱动的方法对国内西北地区某火电厂660 MW燃煤机组的SCR脱硝入口NO_x浓度建立预测模型。首先对火电厂DCS系统采集的数据进行数据预处理, 包括开机判定、LOF算法异常值剔除, 然后采用随机森林进行特征选择, 最后分别采用Wavenet (波网)、LSTM (长短期记忆网络)、SVR (支持向量回归) 三种方法分别对SCR脱硝入口NO_x排放浓度进行预测。结果表明: 无论是预测时间还是各项评价指标, Wavenet都表现更为优秀, 其预测结果的平均百分比误差(Mean Absolute Percentage Error, MAPE)仅为1.31%, r²决定系数达到了0.97, 充分显示Wavenet有比较好的预测能力。

关键词

燃煤锅炉, 氮氧化物排放, 随机森林算法, Wavenet算法

NO_x Emission Concentration Prediction Model for Coal-Fired Boilers Based on Wavenet

Qiaoyu Duan

School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha Hunan

Received: Nov. 13th, 2023; accepted: Dec. 7th, 2023; published: Dec. 15th, 2023

Abstract

Based on the data-driven method, a prediction model was established for the NO_x concentration at the SCR denitrification inlet of a 660MW coal-fired unit of a thermal power plant in northwest China. Firstly, the data collected by the DCS system of the thermal power plant is preprocessed, in-

cluding the start-up judgment and LOF algorithm to screen out the data for outliers. Then, the random forest algorithm is used for feature selection. Finally, the NO_x concentration at the SCR denitrification inlet was predicted by Wavenet model, LSTM (Long Short-Term Memory Network) model and SVR (Support Vector Regression) model. The results show that Wavenet performs better in terms of prediction time and evaluation indicators. The mean absolute percentage error (MAPE) is only 1.31%, and the r^2 coefficient of determination reaches 0.97. This fully shows that the proposed Wavenet model has relatively good predictive ability.

Keywords

Coal-Fired Boiler, NO_x Emissions, Random Forest Algorithm, Wavenet Algorithm

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

全球环境污染日益严重，世界各地越来越注重环境保护， NO_x 作为环境主要污染物之一，国家对于其在火电厂燃煤锅炉的排放标准也越来越严格。而锅炉脱硝主要是依靠 SCR 脱硝系统，其需要根据入口 NO_x 的浓度加入适量的氨水进行处理，因此精确预测 SCR 脱硝系统入口 NO_x 的浓度对于脱硝效率而言非常重要。

预测建模通常分为机理建模和数据驱动建模，与 SCR 脱硝系统入口 NO_x 浓度相关的指标非常庞杂，因此难以建立准确的机理模型。所以国内外近几年都比较倾向于数据驱动建模。SCR 脱硝系统入口 NO_x 浓度预测作为一个典型的时间序列预测问题，对其的数据驱动建模主要分为两类不同的方法，一是机器学习，目前比较主流的有 SVR [1] [2]、随机森林[3]、XGBoost [4]，机器学习的方法优点是可解释性强、相比于传统的统计学方法做时间预测对非线性相关性处理的更好，但是缺点是单一的机器学习方法比较依赖于特征工程的好坏，计算效率较慢且往往准确性不够高。第二类是神经网络方法，对于时间序列预测问题，前些年许多学者大量使用 LSTM [5] [6] [7]或者 GRU [8] [9] (门控循环单元)这种 RNN (循环神经网络)，相比于机器学习方法，并不完全依赖于特征工程的好坏，准确度也提高不少，模型预测效果往往不错，但是其特有的循环网络结构，使得计算量较大，导致计算效率低下。因此在因果卷积[10]的出现之后，CNN (卷积神经网络)类的方法——Wavenet 在时间序列预测方面的应用[11]也得以实现。因卷积网络特有的卷积运算，以及空洞卷积的使用，使得计算速度得到大幅度的提升，且相比于单一的 LSTM 方法，预测精度更高。

为了提高 SCR 脱硝系统入口 NO_x 浓度的预测效果，基于提出的 Wavenet 条件预测[12]，提出了一种随机森林 - Wavenet 预测 SCR 脱硝系统入口 NO_x 浓度的模型。首先通过 LOF 算法[13]对数据进行异常点剔除，然后用随机森林算法[14]对机理分析后的变量进行特征选择，最后利用 Wavenet 模型对 SCR 脱硝系统入口 NO_x 浓度进行预测。采用西北地区某火电厂的 660 MW 燃煤机组 2022 年的数据进行实验，并与 LSTM、SVR 这两种预测模型做了对比，实验结果表明 Wavenet 模型具有比较良好的预测效果，且计算效率也更高。

2. 数据预处理及特征选择

整个燃煤锅炉机组长时间运行过程中难免会出现设备故障、节假日休息等特殊情况。如果将所有

数据不经处理全部用于神经网络的训练与预测，可能造成模型泛化性较差，预测效果也不会太好，因此需要对原始采集的数据进行异常点剔除及降维。

2.1. NO_x 排放的机理分析

对 NO_x 的生成机理进行分析，可用于初步降维。锅炉燃烧过程中会生成 NO_x，其成分主要为 NO 和 NO₂，但是通常 NO 占了其中的大部分。按照其来源可分为三种氮氧化物：燃料性型 NO_x、热力型 NO_x、快速型 NO_x，影响因素有：① 燃料中含氮量多少，② 过量空气系数；③ 炉内温度。

通过机理分析，我们从 DCS 系统采集的 157 个变量变成了包含机组负荷、总风量、二次风门反馈等 70 余个变量，虽然已经减少了不少变量，但是实验设备有限，这么多变量仍然需要进一步降维才能投入模型中训练。

2.2. 异常值处理

先对数据进行开机判定，通过开机判定的两个条件：1) 机组负荷 > 8；2) 汽机侧转速 > 2900，将未开机时的数据剔除。然后用 LOF (局部离群因子) 算法[13]进行异常值筛选。因为数据是燃煤机组一整年的数据，机组在长期使用过程中难免会出现某些异常情况，因此需要采取合适的方法进行异常点筛选。LOF 算法是 Breuning [13] 等人在 2000 年提出的基于密度的识别异常点的方法，其通过定义了一个局部可达密度给出以下公式用于计算每个点的局部孤立程度从而进行异常值筛选：

$$\text{LOF}_{\text{MinPts}}(p) = \frac{\sum_{o \in N_{\text{MinPts}}(p)} \frac{\text{lrd}_{\text{MinPts}}(o)}{\text{lrd}_{\text{MinPts}}(p)}}{|N_{\text{MinPts}}(p)|}$$

其中 $N_k(p)$ 表示 p 的第 k 邻域，即所有与 p 点之间距离都小于 p 的第 k 距离的所有点的集合，MinPts 表示指定的 k ， $\text{lrd}_{\text{MinPts}}(p)$ 表示点 p 的第 MinPts 局部可达密度。在实际应用中，仅需要给定参数 k 和取一个异常点比例的参数 c ，然后根据异常点比例确定取异常值的阈值 m ，当 p 的局部离群因子 $\text{LOF}_k(p) \geq m$ 时， p 被确定为异常点剔除。本文 k 取 20， c 取 0.1，距离取欧氏距离。将异常值所在的一整行删除，最后余下 421,764 组数据，用于接下来的实验。

2.3. 基于随机森林的特征选择

高维数据容易引起“梯度爆炸”、模型泛化性差等问题，通过机理分析虽然已经减少了一半的变量，但是仍然有 70 个变量，这 70 个变量里面存在不少无关变量以及冗余的变量，保留这些变量对于模型的训练以及测试百害而无一利，因此选择集成式特征选择方法——随机森林对数据接着降维。

随机森林是 Breiman [14] 在 2001 年提出的一种 bagging 类集成式方法，基于其特征选择具体步骤如下：

步骤一、通过对原数据集进行有放回的自助采样法获得 n 个训练集用于训练出 n 个决策树组成随机森林；

步骤二、对 \forall 决策树 $i \in 1, 2, \dots, n$ ，计算袋外数据(Out of Bag, OOB)误差，记为 errOOB1_i ；

步骤三、对 $\forall j \in 1, 2, 3, \dots, m$ ， m 为特征个数，随机对袋外数据中所以样本的特征 X_j 加入噪声干扰，再次计算袋外数据误差，记为 $\text{errOOB2}_i^{(j)}$ ；

步骤四、计算特征 X_j 的重要性为 $p_j = \left(\sum_{i=1, 2, \dots, n} (\text{errOOB2}_i^{(j)} - \text{errOOB1}_i) \right) / n$ ，将其按照 $p_j / \left(\sum_{j=1, 2, \dots, m} p_j \right)$

计算结果进行降序排序；

步骤五、按照一定比例删除部分排序靠后的特征；

步骤六、将删除后的新数据重复上述所有步骤，直到剩余目标个数的特征。

本文对随机森林的参数选择为：弱学习器即决策树为 10,000 个，选择回归随机森林，最后得到 24 个筛选后的变量，见表 1。

Table 1. Feature variables after random forest feature selection

表 1. 随机森林特征选择后的特征变量

采集点名称	单位	采集点名称	单位
3 号高加下端差	℃	前左上层中心风门反馈	%
D 磨煤机电流	A	空预器入口烟气含氧量	%
排烟氧量	%	滑压函数	Pa
前左 GG 燃尽风门反馈	%	E 冷风阀门开度	%
B 热风阀门开度	%	前左 B 二次风门反馈	%
C 磨煤机电流	A	前左 D 二次风门反馈	%
E 磨煤机电流	A	给水温度	℃
排烟温度	℃	6 号低加上端差	℃
C 冷风阀门开度	%	后左 GG 燃尽风门反馈	%
前左 FF 燃尽风门反馈	%	6 号低加下端差	℃
一次风机动叶反馈	%	机组负荷	MW
锅炉总风量	t/h	锅炉主控 P (主蒸汽压力)	MPa

3. 基于 Wavenet 的 NO_x 排放浓度预测模型

3.1. 基本原理

参考的[11]的 $NARX(p)$ (p 阶非线性自回归外生)结构确定的 NO_x 排放浓度预测框架：

$$\hat{y}_i = f(y_{t-1}, \dots, y_{t-p}, X_{t-1}, \dots, X_{t-p}), \quad (1)$$

其中 \hat{y}_i 为目标序列的预测值， $f(y_{t-1}, \dots, y_{t-p}, X_{t-1}, \dots, X_{t-p})$ 是目标时间序列 y_{t-1}, \dots, y_{t-p} 和外生时间序列 (即特征变量时间序列) X_{t-1}, \dots, X_{t-p} 的非线性函数。

使用 MSE 类型的损失函数，目标函数为：

$$E(\omega) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \frac{\gamma}{2} \sum_{l=0}^L \sum_{h=1}^{M_{l+1}} (w_h^l)^2, \quad (2)$$

其中 \hat{y}_i 为目标序列的预测值， w_h^l 为权重矩阵，右边第二项为 L2 正则化，用于防止模型过拟合。

$f(y_{t-1}, \dots, y_{t-p}, X_{t-1}, \dots, X_{t-p})$ 是一系列的卷积和非线性激活函数的复合。基础卷积神经网络的结构见图 1：

见图 1，卷积神经网络相比于全连接神经网络结构是稀疏的，卷积神经网络的权重在每一层上共享，这种结构极大的提高了计算效率。卷积神经网络的权重矩阵也称为滤波器(别称卷积核)，每一层的卷积神经网络是权重矩阵与神经元做卷积。

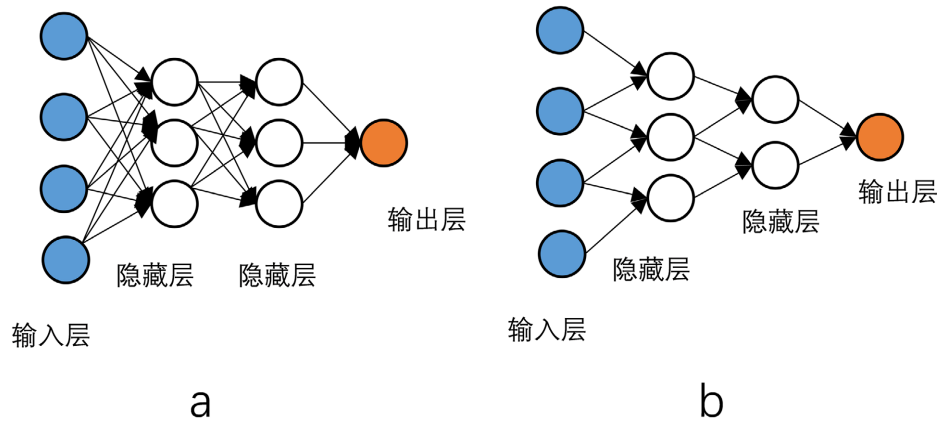


Figure 1. (a) is a fully connected neural network, and (b) is a convolutional neural network
 图 1. (a)为全连接神经网络, (b)为卷积神经网络

假设输入为 $x = (x_0, x_1, \dots, x_{n-1})$, 通过第一层的滤波器 $w_h^1, h = 1, 2, \dots, M_1$, M_1 为第一层卷积核的个数, 则第一个的卷积输出为:

$$a^1(i, h) = (w_h^1 * x)(i) = \sum_{j=-\infty}^{\infty} w_h^1(j) x(i - j), \quad (3)$$

其中 $w_h^1 \in R^{1 \times k \times 1}$, $a^1 \in R^{1 \times (n-k-1) \times M_1}$, k 为卷积核的尺寸, 第一层输出为: $f^1 = \sigma(a^1)$, σ 为激活函数. $l = 2, \dots, L$, $f^{l-1} \in R^{1 \times N_{l-1} \times M_{l-1}}$, N_{l-1} 与卷积核的尺寸有关, 而 M_{l-1} 是与 $l-1$ 层的卷积核的个数, l 层的卷积输出为:

$$a^l(i, h) = (w_h^l * f^{l-1})(i) = \sum_{j=-\infty}^{\infty} \sum_m^{M_{l-1}} w_h^l(j, m) f^{l-1}(i - j, m), \quad (4)$$

而对于 Wavenet 模型最主要的部分是其中的因果卷积和空洞卷积, 其最初是 Oord 等人在 2016 年 [10] 提出的, 图 2 表示了一个四层的因果卷积:

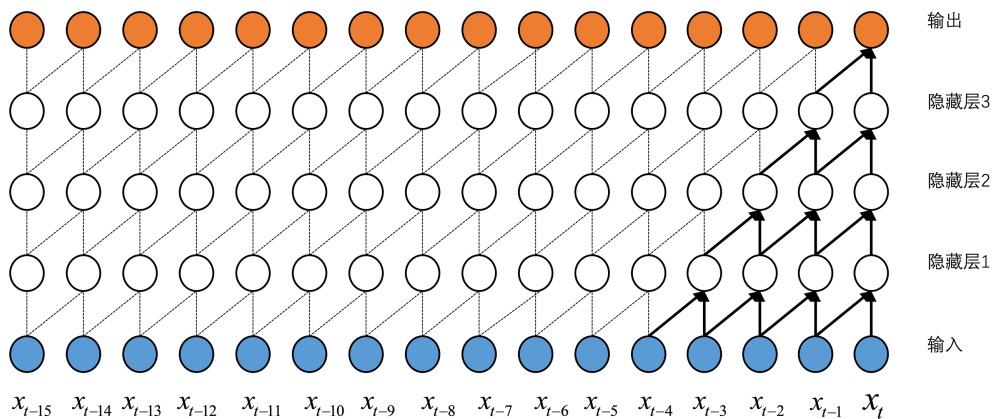


Figure 2. Causal convolution with four layers
 图 2. 四层的因果卷积

如图 2, 因果卷积保证了模型在时间 t 处的预测 $p(x_{t+1} | x_1, \dots, x_t)$ 不会依赖于任何未来时间步 x_{t+1}, \dots, x_T 的信息, 保证了时间序列问题提取信息时的合理性, 所以由因果卷积的模型可以处理时间序列问题。但是因为因果卷积的模型, 但是因果卷积的感受野(感受野 = 层数 + 过滤器长度 - 1)较小,

能提取的信息有限，需要许多层堆叠或者大的滤光片来保证提取到足够长的信息，但是这种方法可能会导致计算量成倍增加。因此[10]提出了空洞因果卷积，见图3：

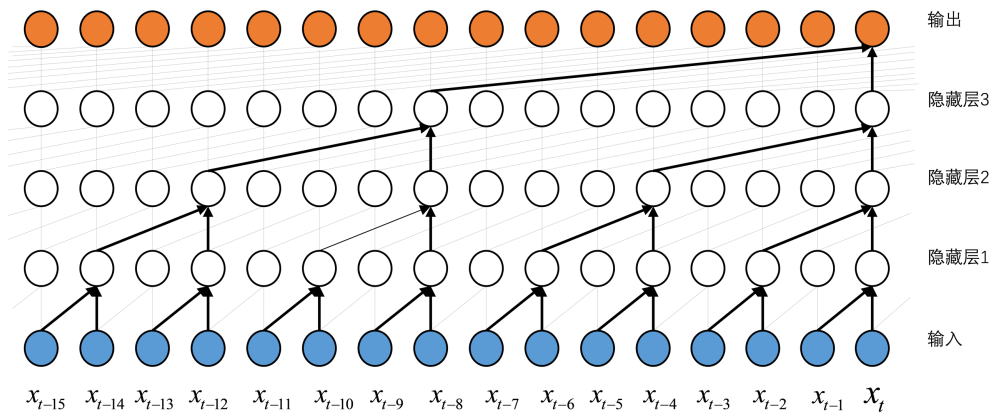


Figure 3. Four-layer void causal convolutional neural network
图3. 四层的空洞因果卷积神经网络

如图3中，空洞因果卷积仅仅用少量的层数即可获得非常大的感受野，同时保留了输入信息的完整性，极大的提高了计算效率。这得益于空洞卷积提出的扩张率，如上图所示以指数增加的扩张率会让感受野随着层数呈指数增长，空洞因果卷积的感受野计算方式为： $r = 2^{L-1}k$ ，其中 L 为网络层数， k 为过滤片的尺寸。第 l 层空洞因果卷积的输出为：

$$(w_h^l *_{d} f^{l-1})(i) = \sum_{j=-\infty}^{\infty} \sum_{m=1}^{M_{l-1}} w_h^l(j, m) f^{l-1}(i - d \cdot j, m), \tag{5}$$

其中 $d = 2^{l-1}$ 为扩张率。

3.2. 模型建立

由公式(1)可知每个 y_t 的预测值 \hat{y}_t 都是由非线性函数 $f(y_{t-1}, \dots, y_{t-p}, X_{t-1}, \dots, X_{t-p})$ 确定的，而这个非线性函数即我们的Wavenet模型，其主要结构见图4：

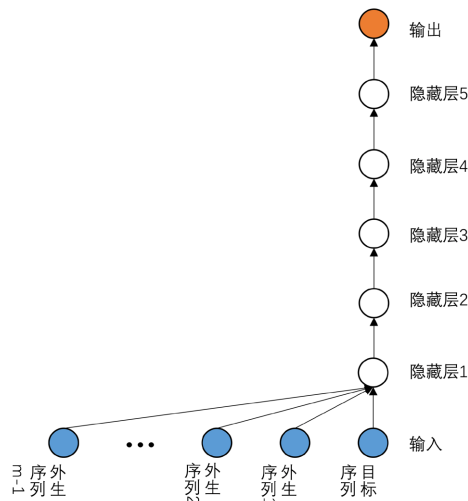


Figure 4. Wavenet neural network model stacked with 6-layer
图4. 堆叠了6层的Wavenet神经网络模型

以上预测模型考虑了外生时间序列的影响，但仅仅在第一层将外生时间序列和目标时间序列分开处理，从第二层到最后一层都合并处理。

Wavenet 模型的具体结构见图 5、图 6、图 7：

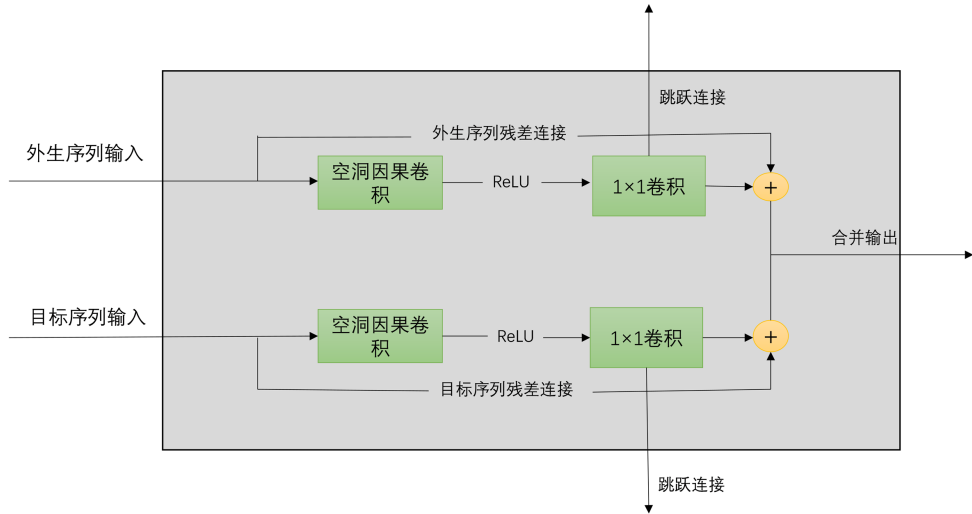


Figure 5. Structure of the first layer of Wavenet
图 5. Wavenet 第一层的结构

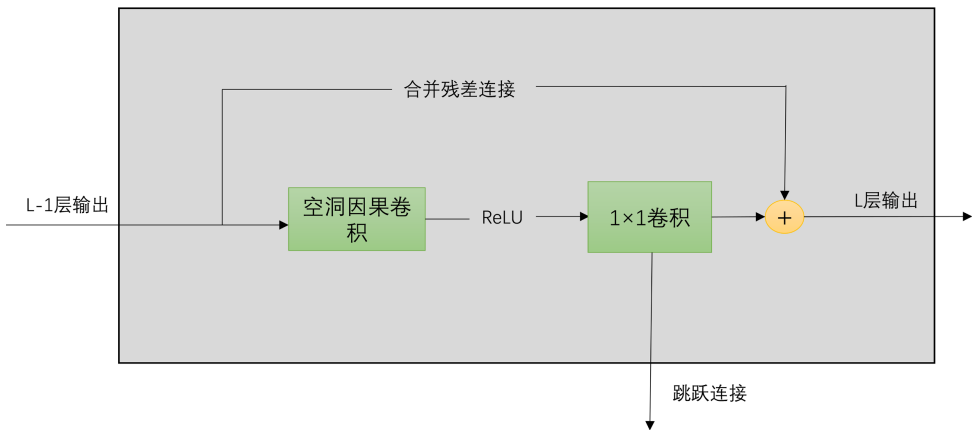


Figure 6. Structure of layers 2 to 5
图 6. 第 2 层到第 5 层的结构

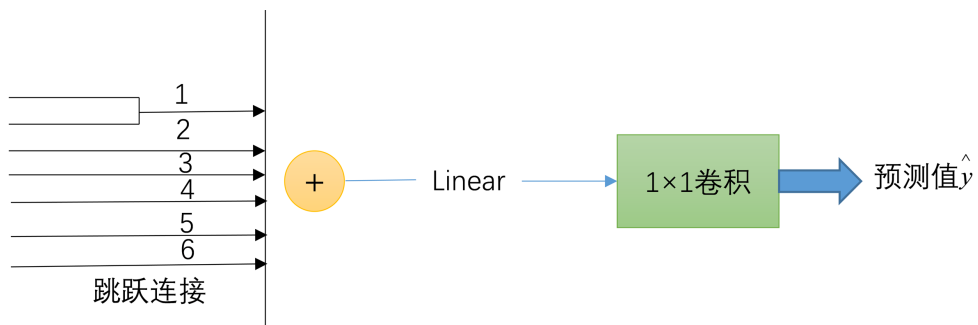


Figure 7. Structure of the output layer
图 7. 输出层的结构

如图 5、图 6，每一层的残差连接之后的输出作为下一层的输入，每一层一维卷积后的跳跃连接到输出层。前五层都使用 ReLU 作为扩张因果卷积的激活函数。最后一层见图 7，选择 Linear 作为激活函数。其中一维卷积网络用于残差连接和跳跃连接时的维度一致。因为每一层的结果都进行了跳跃连接，在最后的输出层可将每一层结果进行逐元素相加后再通过激活函数和一个 1×1 的卷积进行输出最后的预测结果。每一层的跳跃连接可以加快收敛速度，而残差连接保证了可以堆叠更多层而不发生梯度爆炸问题。

4. 实验过程和结果

本文采用 Wavenet 对 NO_x 排放建立了预测模型，取时间长度为 $p = 64$ ，Wavenet 本质上是属于 CNN (卷积神经网络) 类的方法，为了显示模型性能的好坏，选取了 RNN (循环神经网络) 类的方法 LSTM (长短期记忆网络) 与机器学习类的方法 SVR (支持向量回归)。因为 $p = 64$ ，滤光片大小为 2，Wavenet 按照空洞卷积的规则一共堆叠了 6 层，优化器选择 Adam，初始学习率为 0.005，正则化参数为 0.05；LSTM 选取 64 个隐藏节点，只有一层时间步长也取 64，优化器取 Adam，初始学习率为 0.005；SVR 的核函数取 RBF 函数， ε (目标误差) 取 0.01， C (惩罚系数) 取 4.56， γ (核函数的系数) 取 0.7。

数据处理

本文按照 7:2:1 的比例划分训练集、验证集、测试集。为了在训练时更快收敛，需要对数据的每一列先进行归一化处理：

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

按照图 8 的方法将归一化后的特征变量数据和目标变量数据分别处理成 $l \times p \times (m-1)$ 和 $l \times p \times 1$ 的时间序列结构：

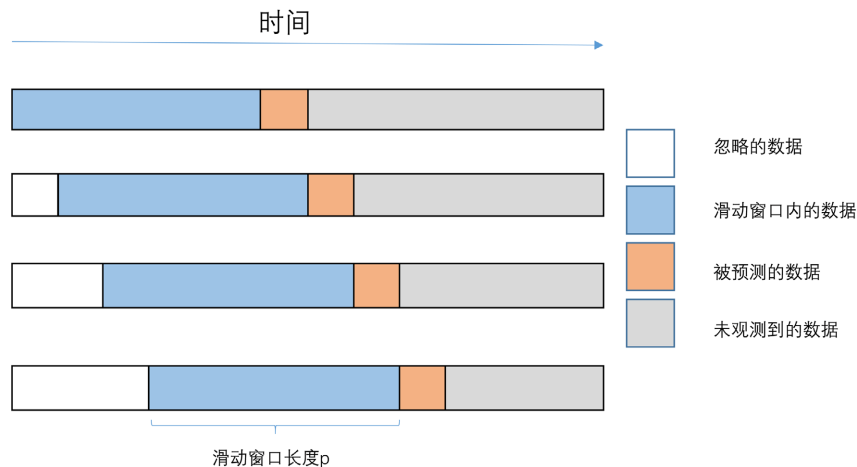


Figure 8. Time series processing
图 8. 时间序列处理

其中 l 为序列长度， p 为滑动窗口长度， m 为特征变量维度，将处理好的时间序列按照 128 的批次放入 Wavenet 模型中进行训练，设定为最多 1000 次迭代：

如图 9，模型收敛很快，训练集和验证集大概在 20 次左右就已经收敛了，而且每次迭代时间仅为 19 s，计算效率较高。

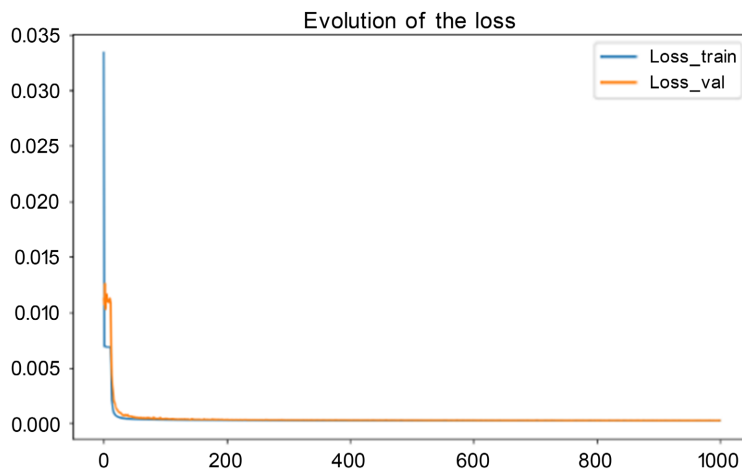
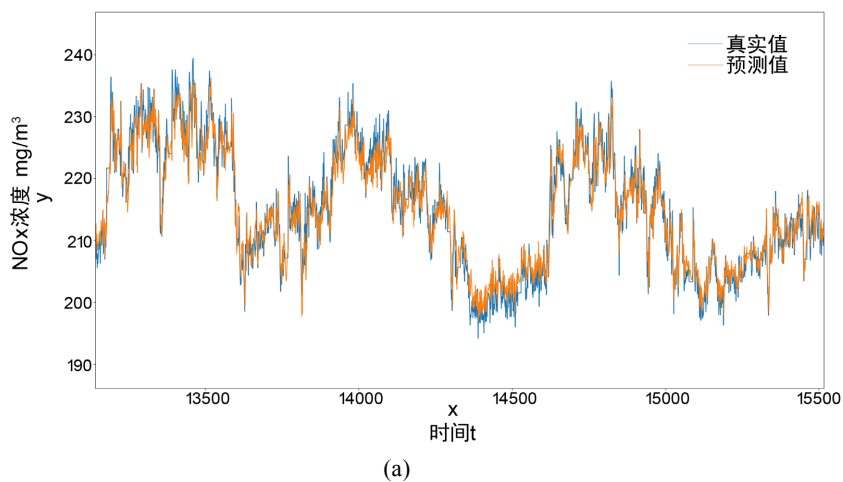
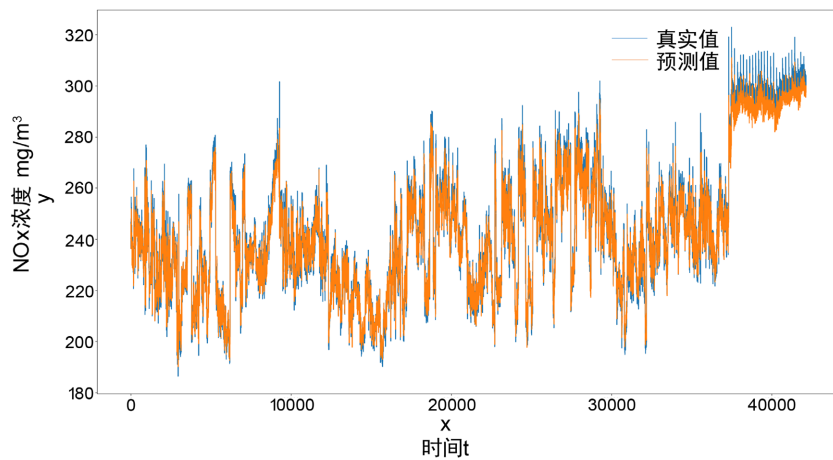


Figure 9. Epoch decline plot of training set loss and validation set loss
图 9. 训练集损失与验证集损失关于 Epoch 下降图

基于 Wavenet 模型在测试集上的预测结果见图 10:



(a)



(b)

Figure 10. (a) is the fitting diagram of the test set, and (b) is the fitting diagram of one of the segments
图 10. (a)为测试集拟合图, (b)为截取其中某一段拟合图

由上图 10 可知, Wavenet 模型拟合效果比较好, 同时其具有高效的计算效率, 但是只是根据拟合图我们没办法得到预测的量化结果, 因此对 Wavenet 模型以及对比模型的 NO_x 浓度预测值与真实值分别计算了四种评价指标——RMSE (均方根误差)、MAE (平均绝对误差)、MAPE (平均绝对百分比误差)、 R^2 决定系数, 实验结果见表 2:

Table 2. Comparison of evaluation indicators

表 2. 各项评价指标对比

模型	RMSE	MAE	MAPE(%)	R^2
Wavenet	4.27	3.20	1.28	0.97
LSTM	8.4414	6.1605	2.3766	0.90
SVR	13.4296	9.6845	3.6754	0.75

时间对比见表 3:

Table 3. Comparison of the iteration times of each round

表 3. 每轮迭代时间对比

模型	时间
Wavenet	19 s
LSTM	46 s

5. 结论

通过实验及对比实验, 得到: 1) Wavenet 模型与 LSTM 模型、SVR 模型相比较, 其计算效率更高, 计算成本更低, 预测效果最好。均方根误差(RMSE)仅为 4.27, 满足燃煤锅炉排放实际生产需要。且 Wavenet 采用的空洞因果卷积既能保证解决时间序列问题的合理性, 又能极大的提高计算效率, 所以从各方面来讲, 本文的预测模型有着不错的效果。2) 特征工程往往能够使得计算量大幅降低, 但由于实验条件有限, 未做是否进行特征选择的对比实验, 未能得到除了关于特征工程对预测精度的提升是否有明显效果。

参考文献

- [1] 钱志. 基于改进型 SVR 的电网短期负荷预测[J]. 中国电力, 2016, 49(8): 54-58.
- [2] 袁秀芳, 张征, 郑伯川, 焦伟超. 基于 SVR 的多变量电力消费预测[J]. 西华师范大学学报(自然科学版), 2015, 36(3): 289-294.
- [3] 刘旭. 火电厂大气污染物监测及空气质量预测系统研究[D]: [硕士学位论文]. 淮南: 安徽理工大学, 2022.
- [4] 邢红涛, 郭江龙, 张颖, 等. 基于变量选择和 XGBoost 组合模型的 NO_x 排放预测[J]. 自动化与仪器仪表, 2021(7): 21-25.
- [5] 杨国田, 王英男, 李新利, 等. 基于互信息变量选择与 LSTM 的电站锅炉 NO_x 排放动态预测[J]. 华北电力大学学报(自然科学版), 2020, 47(3): 66-74.
- [6] 杨青. 基于 LSTM 的燃煤锅炉 NO_x 排放预测研究[J]. 机电信息, 2019(5): 21-22.
- [7] 邢红涛, 郭江龙, 刘书安, 等. 基于 CNN-LSTM 混合神经网络模型的 NO_x 排放预测[J]. 电子测量技术, 2022, 45(2): 98-103.
- [8] 王文广, 赵文杰. 基于 GRU 神经网络的燃煤电站 NO_x 排放预测模型[J]. 华北电力大学学报(自然科学版), 2020, 47(1): 96-103.

- [9] 王珑迪, 何超. 基于 PCA-GRU 的轻型汽油车 NO_x 排放预测[J]. 时代汽车, 2023(10): 10-12.
- [10] Oord, A.V.D., Dieleman, S., Zen, H., *et al.* (2016) WaveNet: A Generative Model for Raw Audio. <https://arxiv.org/abs/1609.03499>
- [11] Börjesson, L. and Singull, M. (2020) Forecasting Financial Time Series through Causal and Dilated Convolutional Neural Networks. *Entropy*, **22**, Article 1094. <https://doi.org/10.3390/e22101094>
- [12] Borovykh, A., Bohte, S. and Oosterlee, C.W. (2017) Conditional Time Series Forecasting with Convolutional Neural Networks. <https://arxiv.org/abs/1703.04691>
- [13] Breuning, M.M., Kriegel, H.P., Ng, R.T., *et al.* (2000) LOF: Identifying Density-Based Local Outliers. *Proceeding of the 2000 ACM SIGMO International Conference on Management of Data*, **29**, 93-104. <https://doi.org/10.1145/335191.335388>
- [14] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>