

王昱婕, 郭小燕. 基于 CEEMDAN 分解的 TCN-GRU 空气质量预测 [J]. 智能计算机与应用, 2025, 15(11): 45-52. DOI: 10.20169/j. issn. 2095-2163. 251107

基于 CEEMDAN 分解的 TCN-GRU 空气质量预测

王昱婕¹, 郭小燕²

(1 甘肃农业大学 理学院, 兰州 730070; 2 甘肃农业大学 信息科学技术学院, 兰州 730070)

摘要: 针对空气质量数据序列噪声强和稳定性弱的问题, 本文提出了一种融合完全自适应噪声集合经验模态分解与时间卷积网络-门控循环单元的组合预测模型(CEEMDAN-TCN-GRU), 旨在提高空气质量小时级别预测的准确性。首先, 对原始 AQI 序列进行 CEEMDAN 分解处理, 得到多个固有模态分量, 以降低数据的复杂性; 其次, 计算各分量的排列熵值, 并按照数值大小进行分量重构; 最后, 将重构序列以及 PM_{2.5}、PM₁₀、SO₂、NO₂、O₃ 和 CO 六种污染指标输入至组合预测模型 TCN-GRU, 利用 TCN 以提取时间特征和局部短期信息, GRU 捕获长期依赖信息得到最终预测值, 提升模型预测精度。实验结果表明: 相较于单一预测模型, 本文提出的组合预测模型的预测效果更佳, 经过模态分解处理的模型预测效果优于未经分解的预测模型, 在均方根误差上相较于 TCN-GRU 和 EEMD-TCN-GRU 分别下降了 5.81 和 2.94, 拟合精度分别增加了 10.3% 和 6.1%。

关键词: 空气质量小时级别预测; 自适应噪声集合经验模态分解; 排列熵; 时间卷积网络; 门控循环单元

中图分类号: X813 文献标志码: A 文章编号: 2095-2163(2025)11-0045-08

Research on TCN-GRU air quality prediction model based on CEEMDAN decomposition

WANG Yujie¹, GUO Xiaoyan²

(1 College of Science, Gansu Agricultural University, Lanzhou 730070, China;
2 College of information science and technology, Gansu Agricultural University, Lanzhou 730070, China)

Abstract: In response to the problems of strong noise and weak stability in air quality data sequences, this paper proposes a combined prediction model (CEEMDAN-TCN-GRU) that integrates fully adaptive noise set empirical mode decomposition and time convolutional network gated recurrent unit, aiming to improve the accuracy of hourly level prediction of air quality. Firstly, perform CEEMDAN decomposition on the original AQI sequence to obtain multiple intrinsic modal components, in order to reduce the complexity of the data; Secondly, calculate the permutation entropy of each component and reconstruct the components according to their numerical values; Finally, the reconstructed sequence and six pollution indicators including PM_{2.5}, PM₁₀, SO₂, NO₂, O₃, and CO are input into the combined prediction model TCN-GRU. TCN is used to extract temporal features and local short-term information, while GRU captures long-term dependency information to obtain the final prediction value, improving the model's prediction accuracy. The experimental results show that compared with a single prediction model, the combined prediction model proposed in this paper has better prediction performance. The model processed by modal decomposition has better prediction performance than the undecomposed prediction model. Compared with TCN-GRU and EEMD-TCN-GRU, the root mean square error decreased by 5.81 and 2.94, respectively, and the fitting accuracy increased by 10.3% and 6.1%, respectively.

Key words: hourly air quality level forecast; adaptive noise set empirical mode decomposition; permutation entropy; time convolutional network; gate recurrent unit

0 引言

城市化进程不断加速, 工业、交通、建筑等活动对空气质量产生了严重影响, 空气质量下降不仅影响人的健康, 还会对环境和经济造成负担。利用统计学原

理建立气象模型是目前常用的空气质量预测方法, 王君悦等^[1]利用大气化学模式 WRF-Chem (Weather Research and Forecasting - Chem) 对长三角地区的 PM_{2.5} 和 O₃ 污染浓度进行模拟; Chadha A 等^[2]用自回归模型对造成空气污染的主要污染物的个体浓度

作者简介: 王昱婕(2000—), 女, 硕士, 主要研究方向: 大数据处理与分析。

通信作者: 郭小燕(1976—), 女, 硕士, 教授, 主要研究方向: 机器视觉, 大数据处理与分析。Email: guoxy@gsau.edu.cn。

收稿日期: 2024-02-26

哈尔滨工业大学主办 ◆ 学术研究与应用

进行估算;沈劲等^[3]采用气象元素分群及多元线性回归方法对顺德日平均污染物浓度及最长 8 小时污染物浓度的趋势变化进行了模拟;田文婷等^[4]建立 ARIMA (Autoregressive Integrated Moving Average Model) 乘法季节模型对未来 12 个月空气质量进行预测并拟合未来变化趋势。这些统计模型在空气质量预测方面取得了较好的效果,但研究均建立在空气质量数据是平稳的且输入值与输出值存在着某种线性关联的基础上,实际情况是空气质量原始数据属于时序数据,时间依赖性强,此时使用统计学模型无法更好地提取时间特征,预测效果会受到影响。

随着人工智能技术的飞速发展,深度学习技术应用于空气质量指数的预测^[5]。Du 等^[6]考虑到统计学方法捕捉时序数据关系能力较弱,采用基于时序特征的长短时记忆神经网络(LSTM)进行空气质量预报,以便求解多个输入时间变量的问题,但其在时间特征提取方面仍有改进空间;黄春桃等^[7]用门控循环单元(GRU)模型对 PM_{2.5}、PM₁₀ 日均浓度值进行预测,相较于应用较多的传统机器学习模型和 LSTM 模型,GRU 表现出更优的预测结果,其结构更简单,提取时间信息能力更强,但仅适用于小规模数据的训练和预测,对于数据量大的情况拟合效果会下降;彭豪杰等^[8]提出了 PM2.5 浓度预测混合模型,运用深度学习手段从训练数据中挖掘出深层的隐性特征,有效应对了输入与输出之间非线性关系的挑战,但预测的准确性依旧不尽如意,分析认为预测误差的主要原因在于空气质量指数(AQI)会受到污染物浓度变化、季节更替及其他因素的影响,造成 AQI 的数据序列波动性强并且不具备稳定性。

针对上述问题,有学者提出模态分解方法,模态分解主要分为两大类即变分模态分解和经验模态分解^[9-10]。变分模态分解(VMD)寻找具体模态与其各自对应的核心频率,通过这些模态的集合合作精确地复原出原始输入信号,并确保每个模态经去调制至基带后呈现出连续光滑的特性;经验模态分解(EMD)数字处理办法把信号内的震荡看成是局部的,将一个信号分解为多个显著信号的本征模态成分之和,但是该方法存在一定模态混叠的问题。为了解决模态混叠和噪声残留的问题,赵星宇等^[11]运用经验模态分解法(CEEMDAN)对序列进行预处理,并运用时间序列预测相关的组合模型对数据进行预测,取得了较好的预测结果。本文运用完善的自适应噪声集合的经验模态分解法(CEEMDAN)来处理信号,通过向目标信号注入配对的正负高斯白噪声,有效降低了经验模

态分解(EMD)在分解过程中可能出现的模态重叠问题。在每次分解过程中逐渐减小噪声水平来提高结果的稳定性,以更好地提取序列信息,去除空气质量数据包含的噪声。

综合上述研究,发现经过模态分解后的组合预测模型能够有效提高模型的预测精度。本文将空气质量信息中的 AQI 指数,利用完全自适应噪声集合经验模态分解(CEEMDAN)为 10 个相对平稳的子序列,以解决数据中的噪声对预测结果的影响;通过计算子序列的排列熵对子序列按照复杂度大小进行聚合重构,形成新序列,由于空气质量数据属于时间序列数据,将时间卷积网络(TCN)和门控循环单元(GRU)进行融合,建立一种组合空气质量预测模型(CEEMDAN-TCN-GRU),模型的输入为模态分解重组的序列和 PM_{2.5}、PM₁₀、SO₂、NO₂、O₃ 以及 CO 污染指标,模型的输出为小时级别的空气质量预测结果。为了验证模型的有效性,将本文提出的预测模型与 TCN、GRU 以及 TCN-GRU、EEMD-TCN-GRU 等模型进行对比实验,实验结果表明 CEEMDAN-TCN-GRU 模型的各项评价指标均优于其他对比模型,预测效果最佳。

1 数据与方法

1.1 研究区域与数据来源

本研究区域聚焦于武汉市城区,如图 1 所示。武汉常住人口超千万,市内江河纵横、湖港交织、地形复杂,同时被京津冀、长三角、珠三角和四川盆地等污染严重区域所围绕的工业化城市,其空气质量表现为结构性污染,秋冬季重污染持续存在^[12-13]。数据选取武汉市 2022 年 1 月~2022 年 12 月共 8 761 条小时级别的空气质量数据,包含时间、空气质量指数(AQI)、PM_{2.5} 浓度、PM₁₀ 浓度、SO₂ 浓度、NO₂ 浓度、O₃ 浓度和 CO 浓度,数据集来自 <https://quotsoft.net/air/> 网站。



图 1 研究区域

Fig. 1 Survey region

对 2022 年 365 天中的逐小时数据取均值,发现 PM_{10} 浓度、 SO_2 浓度和 CO 浓度整体波动较小; NO_2 浓度早上 8:00 开始呈现下降趋势,并在 14:00~16:00 达到最小; O_3 浓度受温度和紫外线强度影响,随着日

照时间增加, O_3 浓度在 15:00~17:00 达到峰值; $\text{PM}_{2.5}$ 浓度在午夜处于较高位置,于 16:00~18:00 出现最低值; AQI 数值在 12:00 处于最大值,最小值出现在 19:00,各个指标逐小时统计分析,见表 1。

表 1 各指标逐小时统计分析
Table 1 Statistical describe of hourly data

时间	AQI	$\text{PM}_{2.5}/(\mu\text{g} \cdot \text{m}^{-3})$	$\text{PM}_{10}/(\mu\text{g} \cdot \text{m}^{-3})$	$\text{SO}_2/(\mu\text{g} \cdot \text{m}^{-3})$	$\text{NO}_2/(\mu\text{g} \cdot \text{m}^{-3})$	$\text{O}_3/(\mu\text{g} \cdot \text{m}^{-3})$	$\text{CO}/(\mu\text{g} \cdot \text{m}^{-3})$
0:00	58.88	37.05	61.21	8.12	41.94	51.80	0.87
1:00	58.13	36.84	59.85	8.19	40.20	50.24	0.87
2:00	57.55	36.50	58.60	8.21	38.81	48.67	0.87
3:00	57.00	36.29	57.52	8.23	37.76	46.89	0.86
4:00	56.72	36.23	56.76	8.28	36.69	45.34	0.85
5:00	56.11	36.03	56.20	8.30	35.92	43.43	0.85
6:00	56.13	36.04	56.10	8.34	36.09	41.12	0.86
7:00	56.52	36.33	56.64	8.49	37.59	38.65	0.89
8:00	57.35	36.66	57.99	8.93	38.30	40.39	0.92
9:00	58.54	37.17	59.84	9.53	35.06	49.64	0.94
10:00	59.60	37.26	61.58	9.98	32.21	61.37	0.94
11:00	60.13	37.18	61.99	9.94	29.21	74.99	0.92
12:00	60.65	36.45	60.32	9.72	26.24	87.69	0.89
13:00	61.35	35.39	57.32	9.28	23.62	97.12	0.85
14:00	60.63	34.08	54.19	8.93	22.34	102.62	0.81
15:00	59.38	32.59	51.15	8.64	22.28	105.56	0.79
16:00	58.52	31.67	49.32	8.44	22.96	106.08	0.77
17:00	57.51	31.33	48.91	8.35	24.96	103.41	.77
18:00	56.11	31.70	49.60	8.04	29.57	95.39	0.78
19:00	55.23	33.12	51.37	7.95	36.47	82.52	0.82
20:00	55.97	34.96	55.10	7.93	41.70	69.98	0.85
21:00	58.32	36.34	59.25	7.97	43.97	61.94	0.86
22:00	59.60	36.95	61.56	8.02	44.62	56.63	0.87
23:00	59.57	37.05	61.89	8.03	43.78	53.57	0.87

1.2 研究方法

1.2.1 模态分解

将 t 时刻原始复杂的非线性、非平稳空气质量数据信号 $\text{AQI}(t)$ 通过自适应噪声完备集合经验模态分解(CEEMDAN)分解成多个简单的振动模式,每个振动模式代表了信号中特定频率范围内的能量。

1) 将均值为 0 的高斯白噪声 $\omega_i(t)$ 加入原始信号 $\text{AQI}(t)$ 中,得到预处理序列 $\text{AQI}_i(t)$:

$$\text{AQI}_i(t) = \text{AQI}(t) + \beta_0 \omega_i(t) \quad (1)$$

其中, β_0 为控制初始信号与附加信号之间噪声比的系数。

2) 利用 EMD 方法对预处理序列进行 N 次分解,获得初始成分 $\text{IMF}_1^i(t)$,之后利用公式(2)对初始成分求均值以获得 CEEMDAN 的首个 $\text{IMF}_1(t)$,分解一次之后,剩余成分依照下式展开:

$$\overline{\text{IMF}}_1(t) = \frac{1}{N} \sum_{i=1}^N \text{IMF}_1^i(t) = \frac{1}{N} \text{EMD}_1[\text{AQI}_i(t)] \quad (2)$$

$$R_1(t) = \text{AQI}(t) - \overline{\text{IMF}}_1(t) \quad (3)$$

3) 重复上述步骤,将分解后得到的第 k 阶段的残余分量中分别添加正、负添加两对高斯白噪声,获得一个新的信号,并以此为载体进行 EMD 分解,直至信号 $R_k(t)$ 无法再进行 EMD 分解,公式如下:

$$\overline{\text{IMF}_k}(t) = \frac{1}{N} \sum_{i=1}^N \text{EMD}_i\{R_{k-1}(t) + \beta_{k-1} \text{EMD}_{k-1}[\omega_i(t)]\} \quad (4)$$

$$R_k(t) = R_{k-1}(t) - \overline{\text{IMF}_k}(t) \quad (5)$$

其中, EMD_{k-1} 是指将数据序列经过 EMD 方法分解所得的第 $k-1$ 阶 IMF 成分; β_{k-1} 表示 CEEMDAN 算法在处理第 $k-1$ 次迭代的剩余部分时引入的噪声幅度; $R_k(t)$ 是经历了 k 次分解后得到的残留部分。

1.2.2 时间卷积网络

Bai 等^[14] 将卷积神经网络 (Convolutional Neural Network, CNN) 的原理融入到时间序列数据的分析之中, 提出一种时间卷积网络 (Temporal Convolutional Networks, TCN), 这个网络主要构建于含有扩张因果卷积的若干残差单元上, 不仅保证了时间序列的连

贯性, 也有效地避免了递归神经网络在处理数据时可能遇到的梯度爆炸或消失的难题。

1) 因果卷积

由于时序数据的特性要求模型只能基于过去的观测值进行预测, 而不能使用未来的信息来预测现在, 因此在处理时序数据时, 采用因果卷积限制卷积操作只作用于过去的输入数据而不能访问未来的输入数据, 保证数据在时间上的顺序性。

2) 空洞卷积

普通卷积中每个输入位置的像素仅与卷积核中心位置的像素相乘, 而在空洞卷积中, 滤波器跳过一些位置以增大感受野, 其中包含的膨胀率决定滤波器每次跳过的像素数量, 可以控制感受野的大小。

空洞因果卷积如图 2 所示。

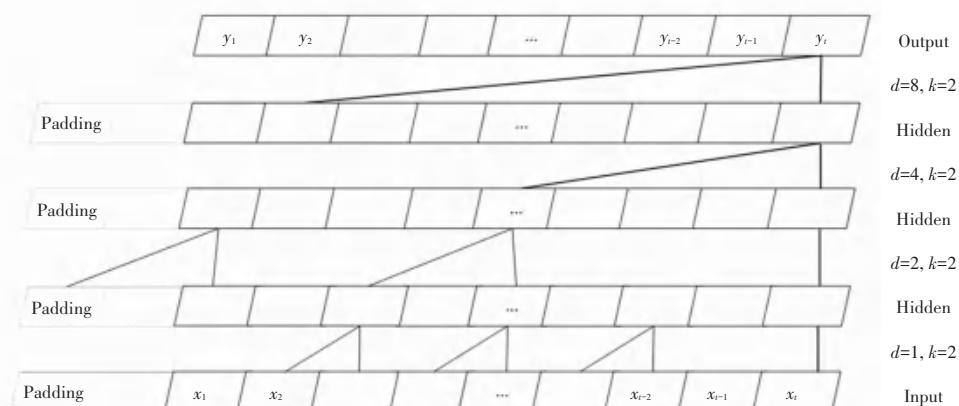


图 2 空洞因果卷积

Fig. 2 Dilated causal convolution

3) 残差连接

残差连接通过在卷积层之间添加连接, 将前一层的输出直接添加到后一层的输入中, 使信息在网络中更快地传播, 防止梯度逐渐消失的问题, 有助于

加速模型收敛, 提高模型性能, 帮助网络更好地处理长期依赖关系。残差模块包括具有膨胀效果的因果卷积层、权重标准化层, 并选择 ReLU 作为非线性激励函数, Dropout 层以减少过拟合, 如图 3 所示。

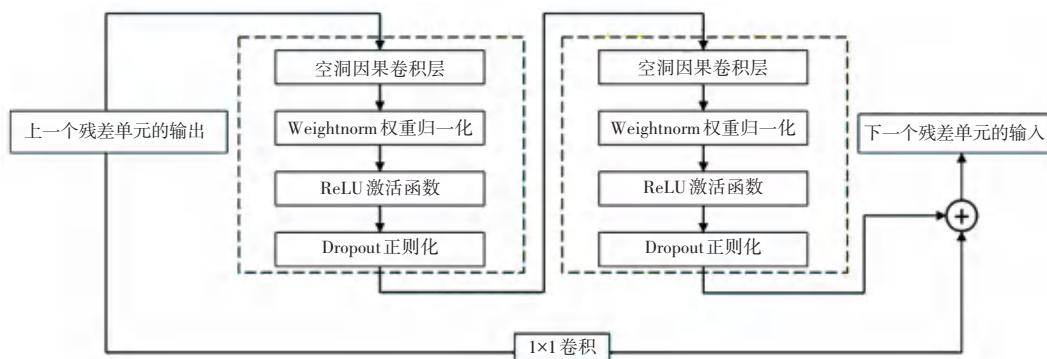


图 3 TCN 残差模块

Fig. 3 TCN residual module

1.2.3 门控循环单元(GRU)

门控循环单元(GRU)是一类特殊的循环神经网络(RNN)结构, 用于解决传统循环神经网络中的梯度消失和长时相关性获取困难等问题^[15]。GRU 结构如图 4 所示, 融入了重置门与更新门, 这两个机制通常受前一时间步的隐藏态和现阶段输入数据的影响。

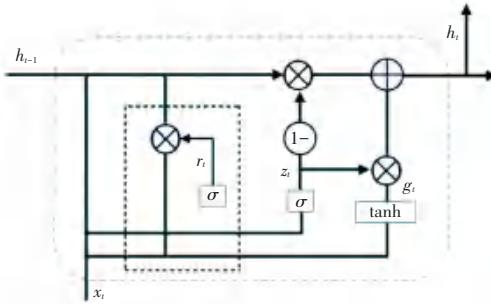


图 4 GRU 结构图

Fig. 4 GRU structure diagram

具体算式如下:

重置门决定了如何将新的输入信息与之前的记忆结合, r_t 映射重置门的输出:

$$r_t = \sigma(\mathbf{W}_r x_t + \mathbf{W}_r h_{t-1}) \quad (6)$$

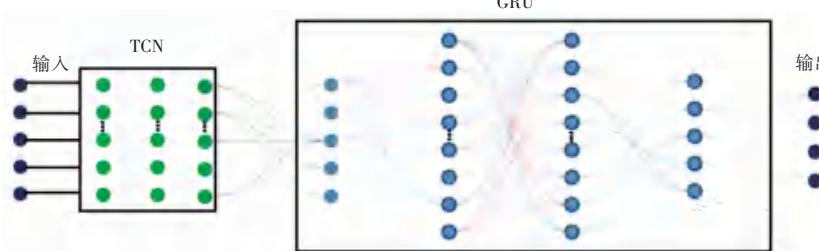


图 5 TCN-GRU 模型结构

Fig. 5 Structure of the TCN-GRU model

1.4 预测流程

CEEMDAN-TCN-GRU 组合预测模型的模型结构如图 6 所示。

详细预测流程:

1) 通过对原始空气质量 AQI 序列进行 CEEMDAN 分解, 获得 n 个不同的相对平稳 IMF 分量和 1 个残余信号;

2) 选择排列熵算法计算分解后每个序列的复杂度信息, 并根据复杂度大小对子序列进行类似重组, 得到 k 个分量($k \leq n$);

3) 将重组后的新序列输入到 TCN-GRU 模型进行训练和预测, 再将残差序列以及 $\text{PM}_{2.5}$ 、 PM_{10} 、CO 浓度等污染物指标送入 TCN-GRU 分别进行预测;

4) 在获得每个网格的预测结果后, 对其进行序列重构, 输入至模型中以获得最终的空气质量预测

其中, x_t 为当前输入数据; h_{t-1} 为前一时间点的潜在状态; \mathbf{W}_r 为重置门中的权重矩阵。

z_t 对应更新门的输出结果, 更新门控制前一时刻的状态信息对当前时刻状态的影响如下式:

$$z_t = \sigma(\mathbf{W}_z x_t + \mathbf{W}_z h_{t-1}) \quad (7)$$

其中, \mathbf{W}_z 为更新门中的权重矩阵。

g_t 代表当前时间点的候选隐藏状态, 候选隐藏状态是通过当前输入和重置门的输出计算得出, 如下式:

$$g_t = \tanh[\mathbf{W}x_t + \mathbf{W}(r_t \times h_{t-1})] \quad (8)$$

其中, \mathbf{W} 用于生成候选潜在状态的权重矩阵, $\tanh(\cdot)$ 为激活函数。

h_t 为最终输出, 由更新门和候选隐藏状态共同决定:

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times g_t \quad (9)$$

1.3 模型结构

利用 TCN 在时间上并行处理的能力, 提取原始信号中的特征, 再将结果送入 GRU 模型, 利用其时序特征, 进行武汉市空气质量预测任务, 组合预测模型结构如图 5 所示。

GRU

结果。

1.5 模型评价指标

为了能够更合理地评估模型的预测性能, 本文选择 RMSE、MAE 和 R^2 三种指标对预测模型展开评价, 具体计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (a_i - \hat{a}_i)^2}, \in [0, +\infty] \quad (10)$$

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |a_i - \hat{a}_i|, \in [0, +\infty] \quad (11)$$

$$R^2 = 1 - \frac{\frac{1}{m} \sum_{i=1}^m (a_i - \hat{a}_i)^2}{\frac{1}{m} \sum_{i=1}^m (a_i - \bar{a}_i)^2}, \in [0, 1] \quad (12)$$

其中, \hat{a}_i 为预测值; a_i 为真实值; m 为样本数据个数。

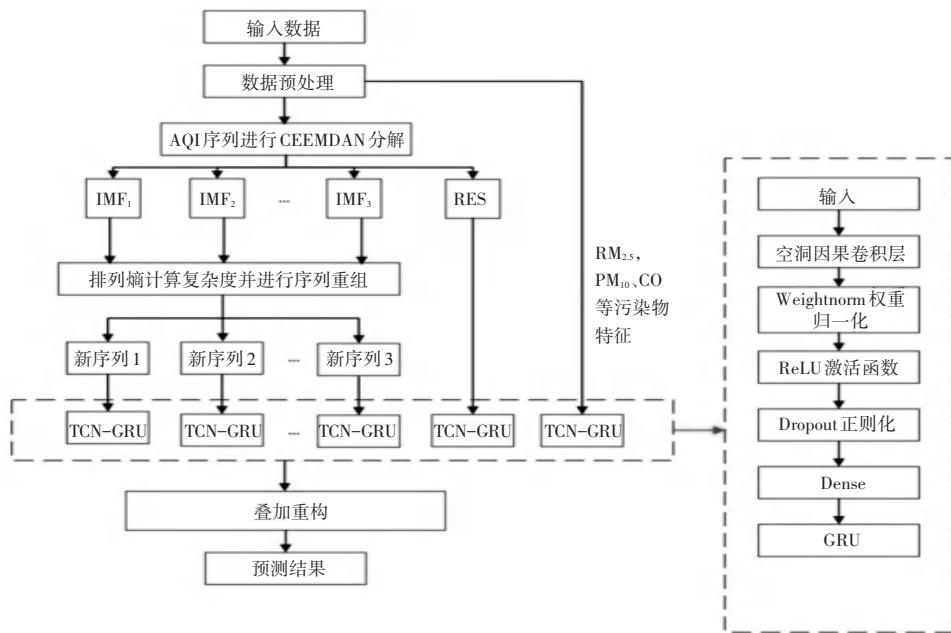


图 6 CEEMDAN-TCN-GRU 模型结构

Fig. 6 CEEMDAN-TCN-GRU model structure

2 结果与分析

2.1 CEEMDAN 分解结果

使用 CEEMDAN 方法对 AQI 进行分解, 分解结

果如图 7 所示, 横轴表示 AQI 样本数, 纵轴表示 IMF 成分的频率, 用信号表示原始 AQI 数据, 通过分解得到了振荡频率从高到低的 10 组 IMF 分量, Residue 为残余信号。

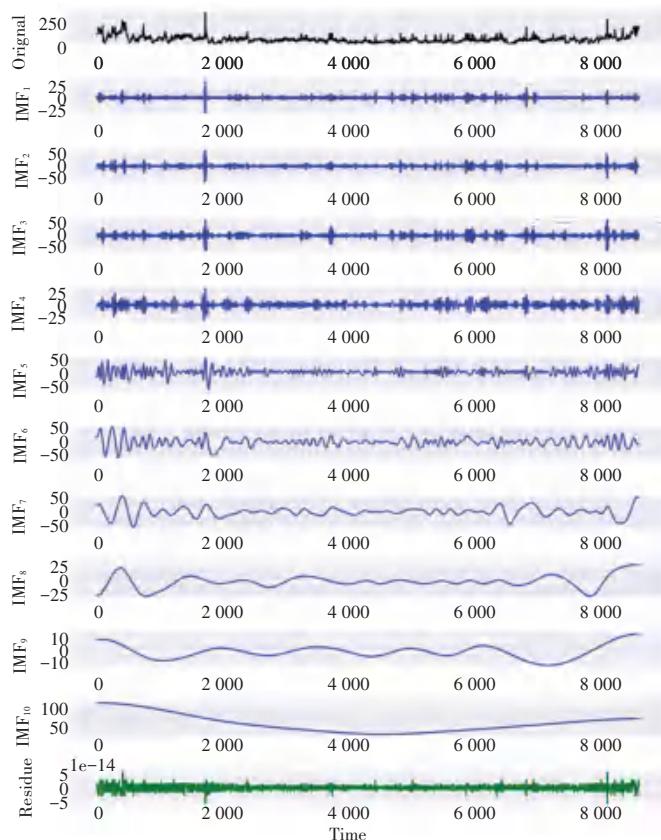


图 7 CEEMDAN 分解结果

Fig. 7 CEEMDAN decomposition results

根据图 7 的分析, 相邻两个 IMF 分量的变化趋势非常接近, 因此可以将这些具有类似复杂性的部分进行整合计算。选择排列熵用于研究和分析序列的复杂度, 通过计算数据序列的不同排列模式数量来量化数据的复杂性, 评估分解效果。排列熵数值

越大说明分解后的数据序列越复杂随机。分解所得 IMF 分量的排列熵数值见表 2, 可见各 IMF 部分的排列熵(PE)值逐一递减, 表明该序列的随机性与复杂度亦随之逐步下降。

表 2 各分量排列熵值

Table 2 Permutation entropy values for each component

分量	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
PE 值	0.986	0.799	0.648	0.553	0.479	0.436	0.410	0.395	0.390	0.387

2.2 模型预测结果分析

本文选择数据集前 60% 作为训练集, 20% 作为验证集, 20% 作为测试集。将本文提出目标预测模型 CEEMDAN-TCN-GRU 与 TCN、GRU、EEMD-TCN、EEMD-GRU、CEEMDAN-TCN、CEEMDAN-GRU、TCN-GRU、EEMD-TCN-GRU 8 种不同的模型进行对比实验。此外, 考虑到不同模型的超参数设置会对最终预测结果产生影响, 本文选取了不同的参数进行结果对比, 确定 epochs 为 200, batch_size 大小为 64, 最终模型的设置如下:

(1) TCN 模型的扩张因子确定为 [1, 2, 4], 卷积核大小设置选择 5×5, 卷积核个数为 128;

(2) GRU 采用两层模型, 其中神经元节点数为

64, 学习率为 0.001, 优化器选择 Adam, 激活函数采用 Sigmoid 函数。

将所选数据输入至预测模型中得到 AQI 序列的预测结果, 预测结果和真实结果的情况如图 8 所示, 可见 EEMD 分解后的 TCN 和 GRU 模型的极端值预测效果有所提升, 但整体拟合效果不佳, 预测偏差较大; 经过 CEEMDAN 分解后的 TCN 和 GRU 模型的真实结果与预测结果重合度明显提升, 证明 CEEMDAN 分解更加有效; 使用组合模型进行预测, 其真实值与预测值的偏差进一步缩小; 目标模型 CEEMDAN-TCN-GRU 的预测结果最接近于真实值, 且对于各端点处的真实值预测更加精确。

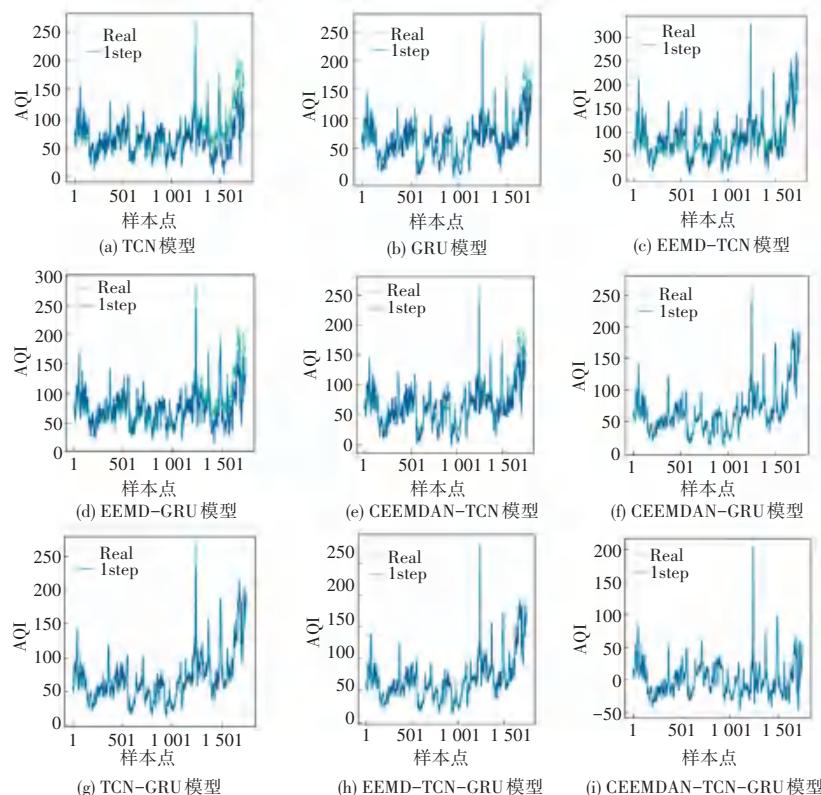


图 8 预测模型的真实值和预测值

Fig. 8 Real and predicted values of prediction model

相同数据样本在单一预测对比模型中的评价指标见表 3, 可见分解后的预测模型性能明显优于未分解的模型, 经 EEMD 方法处理后的 TCN 和 GRU 在 RMSE、MAE 上分别降低了 2.010 1、2.626 3 和 0.949 5、0.113 5, R^2 增加了 0.078 9 和 0.112 6。为了进一步验证本文使用到的 CEEMDAN 分解方法的优势, 对单一模型进行了 CEEMDAN 处理, 结果表明该分解方法的评价指标数值同样优于未经分解处理的预测模型, 且相较于 EEMD-TCN 和 EEMD-GRU 两个模型, CEEMDAN-TCN 和 CEEMDAN-GRU 结果更佳, 其 RMSE 分别减少至 14.691 9 和 10.129 1, MAE 分别减少为 11.364 7 和 7.933 2, 而 R^2 分别提升至 0.823 5 和 0.832 9, 证明 CEEMDAN 分解在提升单一模型预测精度方面是有效的。

表 3 单一预测模型的评价指标

Table 3 Evaluation indicators for a single prediction model

模型	RMSE	MAE	R^2
TCN	22.614 5	17.510 4	0.587 3
GRU	20.364 6	14.071 1	0.655 5
EEMD-TCN	20.604 4	16.560 9	0.666 2
EEMD-GRU	17.738 3	13.957 6	0.768 1
CEEMDAN-TCN	14.691 9	11.364 7	0.823 5
CEEMDAN-GRU	10.129 1	7.933 2	0.832 9

再次进行组合模型预测, 结果见表 4, 可见与单一预测模型相比, 组合预测模型 TCN-GRU 在 RMSE、MAE 和 R^2 3 个评价指标上具有更佳性能, 说明组合模型在实际预测过程中具备重有意义; 而相较于未进行分解处理的组合预测模型, 评价指标结果表明在经过 EEMD 和 CEEMDAN 分解后, 模型的预测性能仍有一定幅度的提升, 且本文选择的 CEEMDAN 方法可使预测模型精准度达到最佳, 相比于 EEMD-TCN-GRU 预测模型, 本文提出的 CEEMDAN-TCN-GRU 空气质量组合预测模型的各项评价指标均优于其他比较模型, 其 RMSE、MAE 分别由 8.926 9 降低至 6.630 1、6.583 3 降低至 4.471 6, 同时 R^2 由 0.910 4 提高为 0.964 1。

表 4 组合预测模型的评价指标

Table 4 Evaluation indicators for combination prediction models

模型	RMSE	MAE	R^2
TCN-GRU	12.445 4	7.774 9	0.873 9
EEMD-TCN-GRU	9.573 2	6.800 1	0.908 7
CEEMDAN-TCN-GRU	6.630 1	4.471 6	0.964 1

3 结束语

针对 AQI 序列存在复杂性和非平稳性的问题, 本文提出了一种将自适应噪声完备集成经验模态分

解(CEEMDAN)与 TCN-GRU 组合模型相结合的空气质量预测模型。利用 CEEMDAN 分解 AQI 时序数据, 可有效减少原始数据的不平稳与复杂度, 从而改善后续数据的质量; 组合模型中, 选择 TCN 利用卷积神经网络的并行结构进行时序建模, 可以捕捉序列中的短期依赖和局部模式信息; 结合 GRU 更有利于处理时间序列数据中的长序列关系和时序演变, 从而提高建模质量, 使模型具有更好的泛化性和适应性。

本文所提出的组合时序模型 CEEMDAN-TCN-GRU 与 TCN、GRU、CEEMDAN-TCN、CEEMDAN-GRU、TCN-GRU 等模型进行对比实验, 结果表明 CEEMDAN-TCN-GRU 模型具有更佳地 RMSE、MAE 和 R^2 , 分别为 6.630 1、4.471 6 和 0.964 1, 取得了较好的预测结果。

参考文献

- [1] 王君悦, 刘朝顺. 基于 WRF-Chem 的长三角地区 $\text{PM}_{2.5}$ 和 O_3 污染协同控制研究 [J]. 环境科学学报, 2022, 42(7): 32–42.
- [2] CHADHA A, GUPTA S. An empirical study of machine learning techniques for air quality prediction [J]. Journal of Critical Reviews, 2020, 7: 1468–1475.
- [3] 沈劲, 钟流举, 何芳芳, 等. 基于聚类与多元回归的空气质量预报模型开发 [J]. 环境科学与技术, 2015, 38(2): 63–66.
- [4] 田文婷, 宇世航. 缺失数据下空气质量预测的 ARIMA 乘法季节模型 [J]. 齐齐哈尔大学学报(自然科学版), 2022, 38(4): 83–89.
- [5] GU K, ZHOU Y, SUN H, et al. Prediction of air quality in Shenzhen based on neural network algorithm [J]. Neural Computing and Applications, 2020, 32: 1879–1892.
- [6] DU Zhehua, LIN Xin. Air quality prediction based on neural network model of long short-term memory [J]. IOP Conference Series Earth and Environmental Science, 2020, 508: 12–13.
- [7] 黄春桃, 范东平, 卢集富, 等. 基于深度学习模型的广州市大气 $\text{PM}_{2.5}$ 和 $\text{PM}10$ 浓度预测 [J]. 环境工程, 2021, 39(12): 135–140.
- [8] 彭豪杰, 周杨, 胡校飞, 等. 基于深度学习与随机森林的 $\text{PM}_{2.5}$ 浓度预测模型 [J]. 遥感学报, 2023, 27(2): 430–440.
- [9] DRAGOMIRETSKIY K, ZOSKO D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531–544.
- [10] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society of London. Series A, 1998, 454: 903–995. DOI: 10.1098/rspa.1998.0193
- [11] 赵星宇, 吴泉军, 朱威. 基于 CEEMDAN 和 TCN-LSTM 模型的短期电力负荷预测 [J]. 科学技术与工程, 2023, 23(4): 1557–1564.
- [12] 王瑶, 唐晓, 陈科艺, 等. 新冠疫情期间武汉空气质量变化的多维观测分析 [J]. 气候与环境研究, 2022, 27(6): 756–768.
- [13] LI H, WANG J, LI R, et al. Novel analysis-forecast system based on multi-objective optimization for air quality index [J]. Journal of Cleaner Production, 2019, 208: 1365–1383.
- [14] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [J]. arXiv, 1803.01271, 2018.
- [15] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint arXiv, 1406.1078, 2014.