

集成 AEC 和时空特征的工业园区 PM_{2.5} 浓度预测

董红召^{1*}, 廖世凯¹, 杨强², 应方³ (1.浙江工业大学, 智能交通系统研究所, 浙江 杭州 310014; 2.杭州环研科技有限公司, 浙江 杭州 311122; 3.浙江省杭州生态环境监测中心, 浙江 杭州 310004)

摘要: 为实现工业园区企业污染排放精细化管理, 捕捉工业园区内企业污染排放与污染物浓度之间的响应关系, 提出一种集成大气环境容量(AEC)和时空特征的工业园区 PM_{2.5} 浓度预测模型. 通过有限体积法获得工业园区日均大气自净能力指数(ASI), 结合工业园区日排放数据作为 AEC 特征; 同时利用小波分析和 Pearson 相关系数法提取时空特征, 包括目标监测站 PM_{2.5} 浓度的时间变化特征和其与周围监测点 PM_{2.5} 的空间相关特征. 通过 CNN 获取训练数据中 PM_{2.5} 的关联特征, 并利用 BILSTM 充分反映时间序列训练数据中隐含的关键历史长短期依赖关系, 确保快速准确的预测性能. 以 2018~2020 年濮阳市工业园区大气污染物观测数据、气象数据及排放数据进行实验验证. 结果表明: 本文提出的 CNN-BILSTM 预测模型相较于传统 LSTM 模型预测精度提升 10%; AEC 特征和时空特征有利于提高模型精度和稳定性, 集成 AEC 和时空特征的 CNN-BILSTM 预测模型在 PM_{2.5} 污染天数预测准确率最高, 达 93%; 分季节预测结果表明, 秋冬季的预测精度最高.

关键词: PM_{2.5} 预测; CNN-BILSTM; 大气环境容量; 工业园区

中图分类号: X513 文献标识码: A 文章编号: 1000-6923(2022)10-4537-10

PM_{2.5} concentration prediction in industrial parks integrating AEC and spatio-temporal characteristics. DONG Hong-zhao^{1*}, LIAO Shi-kai¹, YANG Qiang², YING Fang³ (1. Institute of Intelligent Transportation Systems, Zhejiang University of Technology, Hangzhou 310014, China; 2. Hangzhou Huanyan Technology Co., Ltd., Hangzhou 311122, China; 3. Zhejiang Hangzhou Ecological and Environmental Monitoring Center, Hangzhou 310004, China). *China Environmental Science*, 2022, 42(10): 4537~4546

Abstract: With the purpose of realizing the fine management and control of enterprise pollution emissions and capturing response relationship between enterprise pollution emissions and pollutant concentrations in industrial parks, a PM_{2.5} concentration prediction model was proposed to integrate atmospheric environmental capacity (AEC) and spatio-temporal characteristics. Firstly, the daily average atmospheric self-purification capacity index (ASI) was obtained using finite volume method and combined with the daily emission data acted as AEC characteristics. At the same time, temporal characteristics at the target monitoring station and spatial characteristics including its surrounding monitoring points of PM_{2.5} concentration were showed by wavelet analysis and Pearson correlation coefficient method. Then, in order to guarantee the fast and accurate prediction performance, the correlation characteristics of PM_{2.5} in training data were obtained by CNN, and BILSTM was used to fully reflect the key historical long short-term dependencies implied in time series training data. The air pollutant observation data, meteorological data and emission data of Puyang Industrial Park from 2018 to 2020 were applied to experimental verification. The results show that the CNN-BILSTM prediction model proposed in this paper improves the prediction accuracy by 10% compared with the traditional LSTM model. AEC and spatio-temporal features are well situated to improve accuracy and stability of model. The CNN-BILSTM prediction model integrating AEC and spatio-temporal features has the highest prediction accuracy during PM_{2.5} pollution period, which is up to 93%. Moreover, the seasonal prediction results performed the highest prediction accuracy in autumn and winter.

Key words: PM_{2.5} prediction; CNN-BILSTM; atmospheric environment capacity; industrial park

研究表明,工业生产和废气排放与区域 PM_{2.5} 浓度高度正相关^[1].从大气环境容量(AEC)的角度看,大气污染程度取决于污染物排放量与大气环境容量间的平衡,在无源的情景下,大气中污染物浓度衰减的快慢取决于大气自净能力的大小^[2-4].通过大气自净能力可得到大气污染物排放量与污染物浓度间的影响规律,继而在大气自净能力不变时,由污染物浓度反演得到理想排放限值,实现对工业园区排放量的精准管控.

目前 PM_{2.5} 浓度预测模型主要包括机理模型和统计学模型^[5].机理模型通过物理化学原理对大气污染物的理化反应进行模拟分析,需要有各类气象数据和污染源排放清单.主要包括通用多尺度空气质量模型^[6](CAMQ 模式)、气象-化学耦合模型^[7](WRF-Chem 模式)和嵌套空气质量预报模式系统^[8]

收稿日期: 2022-03-20

基金项目: 浙江省科技计划项目(LGF20F030001)

* 责任作者, 教授, its@zjut.edu.cn

(NAQPMS 模式)等,在缺少各类气象数据和污染源排放清单时,机理模型难以精准预测 $PM_{2.5}$ 浓度,统计学模型则利用机器学习、深度学习和统计学分析其历史监测数据发现变化规律并给出合理预测.模型预测精度高,在缺少探空气象数据和污染源排放清单时也能保证较好的预测效果.目前统计模型主要有:支持向量回归模型^[9-11]、随机森林^[12-14]、BP 神经网络^[15]以及 LSTM^[16-19]、BILSTM^[20-21]神经网络等单机器学习模型.为进一步提高预测精度,多机器学习模型应运而生.Liu 等^[22]构建了 DBN、LSTM 和多层神经网络(MLP)的组合模型.康俊锋等^[23]建立 XGBoost、LSTM 组合模型进行 $PM_{2.5}$ 浓度短期预测.梁泽等^[24]耦合遗传算法和 RBF 神经网络构建 $PM_{2.5}$ 预测模型,具有依赖变量少、预测精度高和运算效率高等特征. AEC 源于机理模型,其变化反映出污染物在大气中的物理化学反应,但现有研究未实现机理模型和非机理模型的有机结合,未考虑 AEC 变化对污染物浓度的影响,且往往笼统的将所有站点看作独立个体,未考虑大气污染物浓度的时空特征,即站点的空间相关性和污染物浓度变化的周期性^[25-26].

本研究提出一种集成 AEC 和时空特征的工业园区 $PM_{2.5}$ 预测模型,将工业园区气象数据、空气污染物数据、 $PM_{2.5}$ 历史浓度数据及排放数据结合,在大气自净能力-排放源- $PM_{2.5}$ 浓度的大气环境容量理念下,构建 CNN-BILSTM 网络来捕捉工业园 $PM_{2.5}$ 浓度的变化规律,对未来 24h 平均 $PM_{2.5}$ 浓度进行预测,为工业园区大气重污染预警和企业排放控制提供理论支持.

1 数据与方法

1.1 研究区与数据源

濮阳市工业园区坐落于河南省濮阳市,濮阳市为石油化工型城市,污染物排放量较大,重污染天气频发,化工、新能源新材料和装备制造为其三大主导产业.濮阳市工业园区二、三类工业企业数量较多,对城区环境质量影响较大^[27].工业园区企业粉尘、废气无组织排放的监管督办力度自 2018 年起逐步加大,违者停业整顿,但污染排放量基数庞大,空气质量状况不容乐观^[28-29].根据濮阳市工业园区 2018 年大气源清单,各类源 $PM_{2.5}$ 主要前体物排

量如表 1 所示.

表 1 工业园区 2018 年 $PM_{2.5}$ 主要前体物排放量统计(t)
Table 1 Emission statistics of main precursor of $PM_{2.5}$ in industrial parks in 2018 (t)

| 排放量 | TSP | SO ₂ | NO _x |
|-----|------|-----------------|-----------------|
| 工业源 | 5655 | 225 | 488 |
| 移动源 | 9 | 3 | 102 |
| 农业源 | 0 | 0 | 0 |
| 扬尘源 | 635 | 0 | 0 |
| 其它 | 1 | 0 | 0 |

由表 1 可知,工业园区污染排放主要来自工业源, $PM_{2.5}$ 主要前体物,如烟尘、二氧化碳、氮氧化物的排放量均为最高,平均占比达到 90%.工业源排放主要来自于各化工或其它制造企业的生产活动,相比于移动源、扬尘源等,可控性更强,将其作为输入变量构建模型更利于实现工业园区排放管控.

使用的大气污染物浓度数据为 2018 年 1 月 1 日~2020 年 12 月 31 日濮阳市工业园区和周围共计 12 个空气自动监测站点(图 1)逐小时监测数据,包括 $PM_{2.5}$ 、 PM_{10} 、CO、SO₂ 及 O₃ 质量浓度.其中气象数据来源于濮阳市地面气象站,主要包括气温、相对湿度、气压、降水、风速风向及露点温度等.排放数据来源于河南省重点污染源自动监控系统,监测排口主要为园区内大型石油化工、火力发电企业等重点污染源的窑炉排放口、烟囱排放口和脱硫出口等,主要排放物包括二氧化硫、氮氧化物及烟尘等.

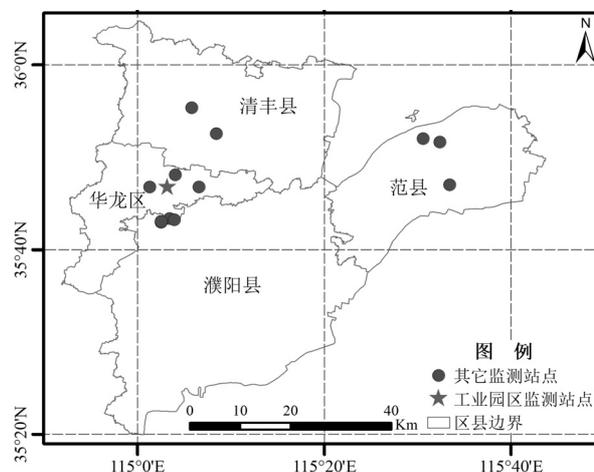


图 1 空气自动监测站点位置分布

Fig.1 Position of air automatic monitoring points

1.2 研究方法

1.2.1 大气自净能力指数及其计算方法 大气自净指数用来表征大气自身运动中对大气中污染物的扩散、稀释和清除能力.根据平均浓度预报方程和箱模型,采用平均体积法可推导出单位时间、单位面积上大气平流扩散和降水所能清除的最大污染物总量大气自净能力指数 ASI^[3],即

$$ASI = Q/S = (\frac{\sqrt{\pi}}{2} V_E + W_l R \sqrt{S}) \cdot C_s / \sqrt{S} \quad (1)$$

式中:Q 为大气污染物在一定空气体积内的排放量,kg;S 为底面积,km²;V_E 为大气通风量,m²/s;W_l 为雨洗常数,无量纲;R 为单位时间内的降水量,mm/h; C_s 为污染物达标浓度,mg/m³.

大气自净能力指数越大,代表大气对污染物的清除能力越强,反之,代表其越弱.

1.2.2 CNN-BILSTM 预测模型 气象、大气污染物浓度和污染源强之间的变化呈高度复杂性、非线性,采用传统模拟方法难以快速准确的达到模拟效果.因此采用融合历史气象、排放和空气质量数据的深度学习模型来拟合三者变化规律.大气污染物浓度的影响因子变化复杂且数量繁多,包括气象因子、排放因子和其他大气污染物,故利用卷积神经网络

(CNN)进行关键特征提取,减少冗余信息;污染物浓度变化具有强自相关性和周期性,污染物浓度变化不仅受过去的信息特征影响,而且与未来的特征变化亦息息相关.而双向长短期记忆网络(BILSTM)兼顾双向时间序列中所有值的关键历史长期和短期依赖关系.因此采用 CNN-BILSTM 组合模型拟合工业园区内 PM_{2.5} 浓度变化规律.

如图 2 所示,数据建模包括时间序列建模和归一化处理,时间序列建模将原始数据划分为不同时间步样例,与 PM_{2.5} 浓度相关的气象、排放和其他大气污染物浓度数据作为特征集,数据归一化采用 Min-Max 标准化方法. CNN 包括卷积层、池化层和全连接层;BILSTM 包括前向 LSTM 层和后向 LSTM 层.模型输入为气象、排放和大气污染物浓度时序数据,输出为未来 1hPM_{2.5} 浓度.预测建模包括 CNN 卷积层将时序数据映射为 n×s 矩阵,其中,n 代表时序数据样例数,s 代表数据维度,经过滤波器处理后进入池化层,提取局部特征,减少冗余信息,将提取后的多组特征向量输入 BILSTM 层、Dropout 层和全连接层中,得到未来 PM_{2.5} 浓度输出结果.此外,为了增强 CNN-BILSTM 网络模型的泛化性和鲁棒性,选择 Adam 作为优化算法.

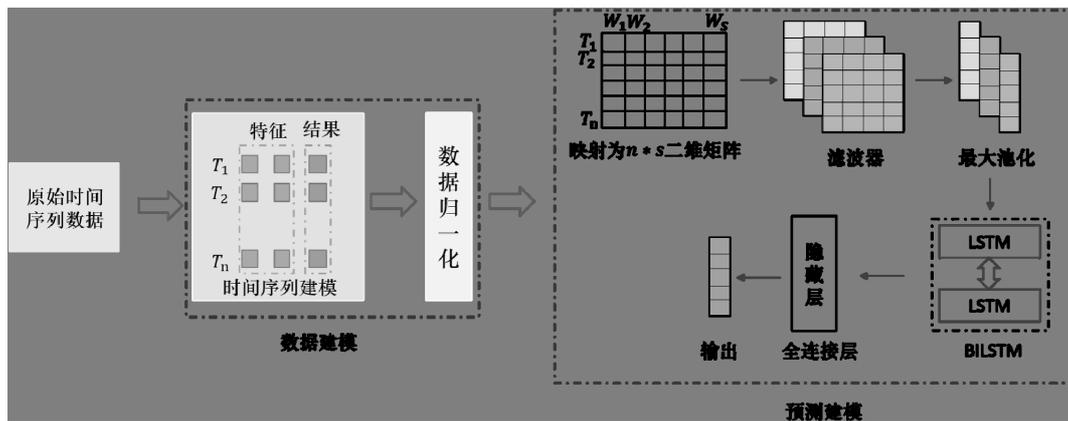


图 2 CNN-BILSTM 框架
Fig.2 The framework of CNN-BILSTM

1.2.3 组合预测模型流程 预测模型构建流程如图 3 所示,包括预处理、模型构建和模型预测.

(1)预处理:预处理包括数据预处理、特征获取和构建训练集.数据预处理主要包括对气象数据、大气污染物数据和排放数据进行异常值处理、缺失值填补和归一化处理,异常值处理方法为 3σ 准则,缺失

值填补采用均值填充和 MICE 算法,归一化处理采用 Min-Max 标准化方法;特征获取包括通过有限体积法得到大气自净能力指数,结合排放数据组成大气自净能力-排放变化特征,通过小波分析得到 PM_{2.5} 浓度变化的主周期,作为时间特征,相关性分析得到其浓度变化相似的站点数据和辅助气象及大

气污染物,作为空间特征和辅助特征;最后利用获取的特征构建训练集.

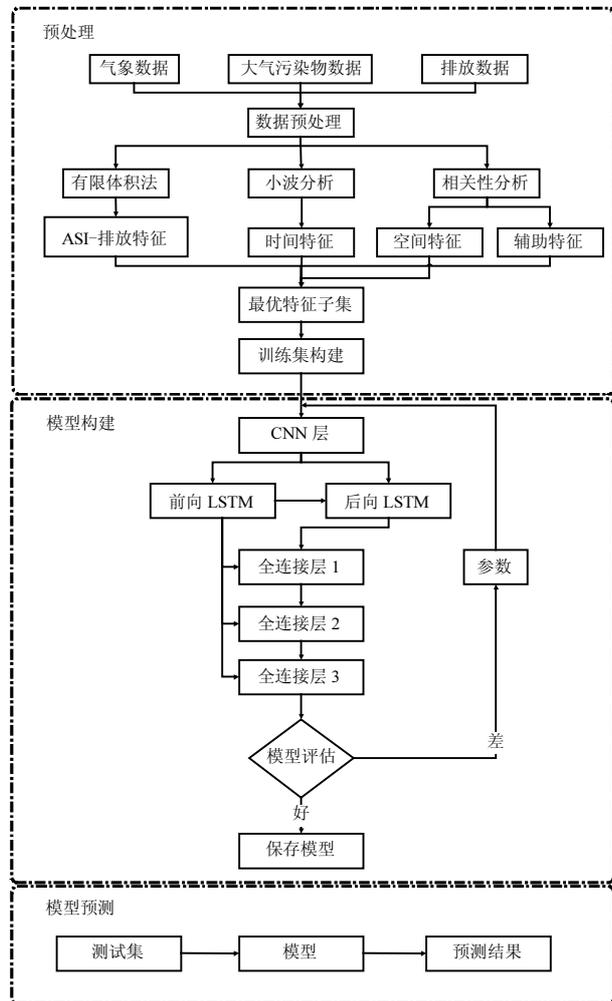


图3 CNN-BiLSTM 模型预测流程

Fig.3 CNN-BiLSTM model prediction process

(2)模型构建:秋冬季重污染天气频发,为了保证模型污染预警能力,测试集选择在秋冬季,数据集按照训练集、验证集和测试集为 6:2:2 的比例划分,划分方法为留出法,算法优化器为 Adam,以实测值和预测值的相关系数作为标准对结果进行评估,最终获取模型的最佳参数,保存最优模型.

(3)模型预测:将测试集数据输入模型中,对预测值和实测值进行数学分析,统计其对应污染等级情况并分析预测结果.

1.2.4 评价指标 组合预测模型精度通过均方根误差(RMSE)、平均绝对误差(MAE)、相关系数(R^2)来对预测结果进行评价.

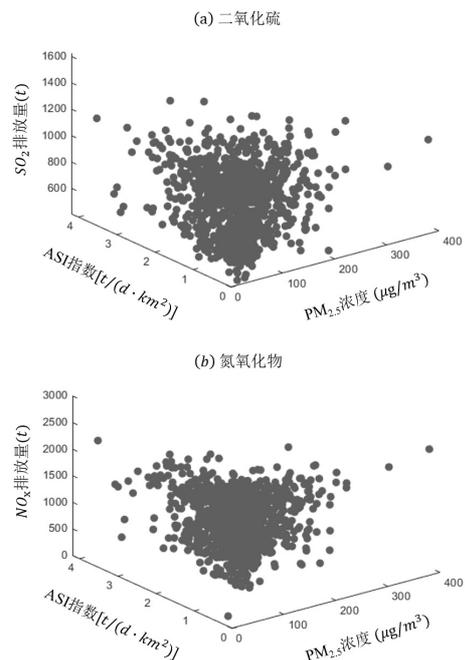
2 结果与分析

2.1 特征分析

2.1.1 AEC 特征 如表 2 所示, ASI 与 $PM_{2.5}$ 浓度呈负相关,当大气自净能力增强时,大气扩散清除能力增强, $PM_{2.5}$ 浓度呈下降趋势; 3 种主要污染排放与 $PM_{2.5}$ 呈显著正相关, 表明当园区内排放量增加会导致 $PM_{2.5}$ 浓度上升.由图 4 可知,污染物排放量与 $PM_{2.5}$ 浓度之间呈明显的正比关系,当污染物排放量增大时, $PM_{2.5}$ 浓度呈上升趋势;当排放量在一定范围内时,随着污染物浓度增大,ASI 呈减小趋势;即当 $PM_{2.5}$ 浓度呈上升趋势时,排放量呈增大趋势,而 ASI 逐渐减小.此外,大气自净能力与排放都呈现明显的季节变化,濮阳市工业园区属于暖温带半湿润季风型大陆性气候,不同季节气候差异明显,园区企业不同季节生产计划不同,污染排放量也不尽相同.

表 2 $PM_{2.5}$ 日均浓度与 ASI 和排放量相关性统计
Table 2 Correlation statistics of $PM_{2.5}$ daily average concentration with ASI and emissions

| 参数 | 二氧化硫 | 氮氧化物 | 烟尘 | ASI |
|-----------|--------|--------|--------|--------|
| R | 0.48 | 0.49 | 0.47 | -0.31 |
| 显著性水平 (P) | < 0.01 | < 0.01 | < 0.01 | < 0.01 |



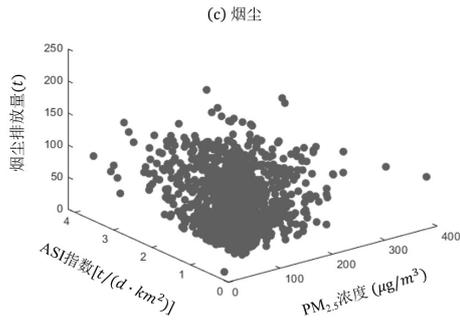


图4 PM_{2.5}、ASI 和排放量的日均数据分布

Fig.4 Daily data distribution of PM_{2.5}, ASI and emissions

季节划分:3~5 月为春季,6~8 月为夏季,9~11 月为秋季,12 月和次年的 1、2 月为冬季^[30].如图 5 所示,随着春夏秋冬时间线的推进,PM_{2.5} 浓度逐渐上升,污染物排放量逐渐增大.将 ASI、大气污染物排放和季节作为 AEC 特征输入模型.

2.1.2 时间特征 如图 6 所示,2018~2020 年濮阳市工业园区 PM_{2.5} 日均浓度时间序列呈现明显的周期性特征.对 PM_{2.5} 时间序列进行小波分析,绘制小波方差图(图 7),小波方差图中波峰对应的尺度为 PM_{2.5} 的主周期,由图 7 可知,在 70d 以内,PM_{2.5} 时间

序列存在 2 个振荡周期,分别为 10~20d 和 30~50d,第 1 主周期为 37d,第 2 主周期为 14d.结合图 8 可知,PM_{2.5} 时间序列具有较强的自相关性,但是在滞后天数大于 52d 时,相关系数低于 0.2,在滞后天数为 14d 时相关性最高,因此选择日前 14d 的 PM_{2.5} 浓度作为时间特征输入模型.

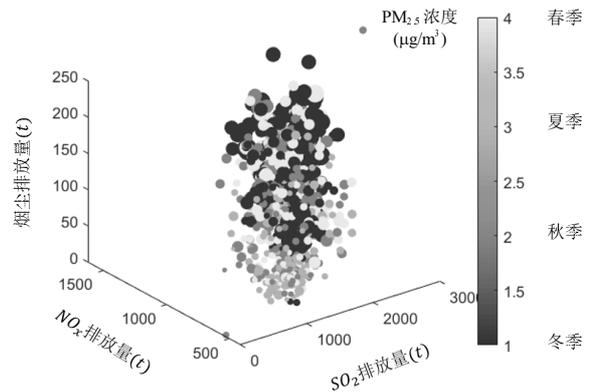


图5 PM_{2.5} 和大气污染物排放的季节变化

Fig.5 Seasonal variations in PM_{2.5} and air pollutant emissions
气泡大小代表PM_{2.5}浓度,气泡越大,PM_{2.5}浓度越高;颜色变化代表一年四季,颜色越深,越趋近于冬季

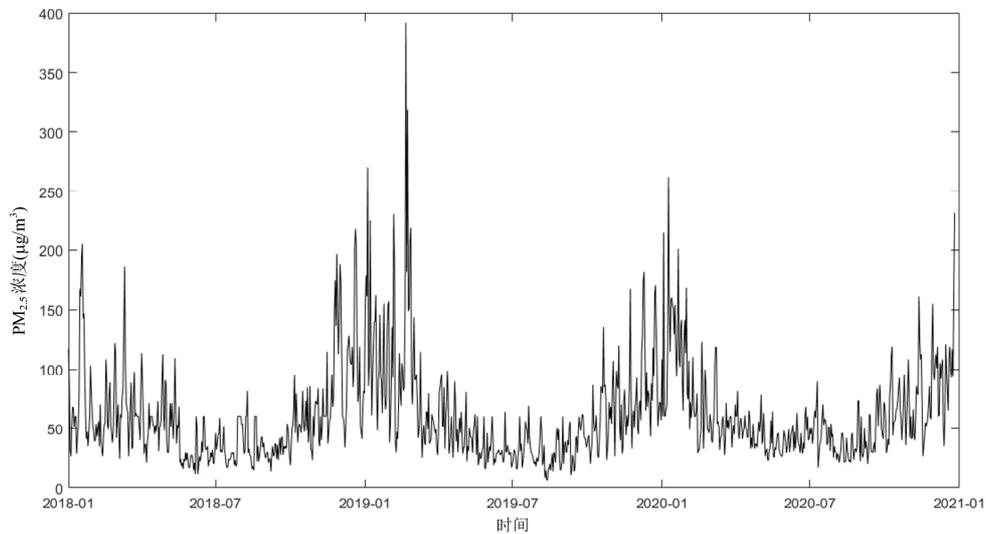


图6 2018~2020 年工业园区 PM_{2.5} 日均浓度时间序列

Fig.6 Time series of PM_{2.5} daily average concentration in industrial parks from 2018 to 2020

2.1.3 空间特征 空间特征分析主要是工业园区内监测站点与周围监测站点 PM_{2.5} 浓度序列的相关性分析.由于工业园区紧邻的华龙区内有其小工业园,区域污染可能有其他贡献源,故华龙区 3 个站点数据均用于分析;工业园区北面、南面及东面的清丰县、范县及濮阳县均有 3 个空气质量监测站点,

由于监测站点位置集中及与工业园区站点较远,因此取各县区域内 3 个站点的平均值用于分析.如表 3 所示,华龙区的 3 个站点 PM_{2.5} 浓度变化与工业园区 PM_{2.5} 浓度变化有极强的相关性,经过实验分析,相关性过高会产生冗余,导致精度降低^[24].故只将清丰县、范县和濮阳县 3 个县区的站点平均值作为

空间特征输入。

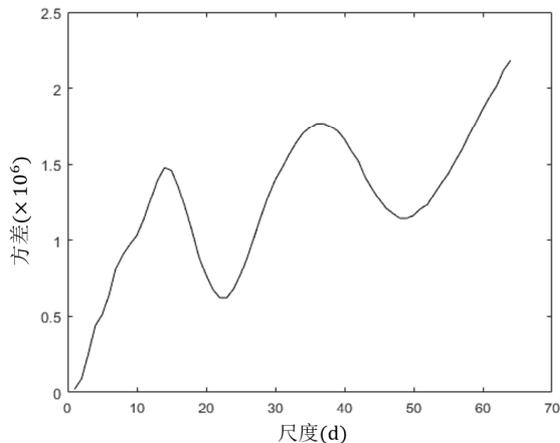


图7 $PM_{2.5}$ 时间序列小波方差

Fig.7 Wavelet variance diagram of $PM_{2.5}$ time series

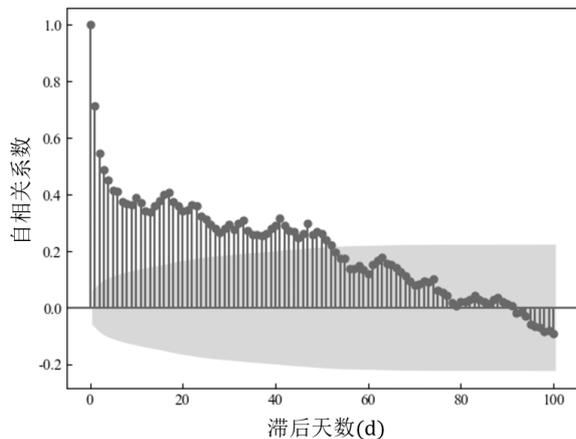


图8 $PM_{2.5}$ 时间序列时间滞后自相关系数

Fig.8 Lagging self-correlation coefficient of $PM_{2.5}$ time series

表3 工业园区站点与相邻站点 $PM_{2.5}$ 浓度相关性统计

Table 3 Correlation statistics of $PM_{2.5}$ concentration between industrial park sites and adjacent sites

| 参数 | 华龙区 | | 清丰县 | 范县 | 濮阳县 |
|----------|---------|---------|---------|---------|---------|
| | 油田运输公司 | 油田物探公司 | 环保局 | | |
| 相关系数 | 0.88 | 0.88 | 0.87 | 0.52 | 0.50 |
| 显著性水平(P) | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 |

2.1.4 辅助特征 $PM_{2.5}$ 浓度变化的辅助因素包括气象因素和其他大气污染物浓度.如表4所示, $PM_{2.5}$ 与降水、相对湿度和风向相关性不高,濮阳市位于中纬地带,常年受东南季风环流的控制和影响,属暖温带半湿润季风型大陆性气候.年平均日照时数2300~2500h,是河南省日照高值区.由于气温变化大,降水时空分布不均,气象灾害频繁,干旱占突出地位^[31];年平均风速2.7m/s,常年主导风向是南风、东南风,因此降水、相对湿度和风向的变化与 $PM_{2.5}$ 相关性不高. $PM_{2.5}$ 与 PM_{10} 、 NO_2 和 CO 相关性很强,与 SO_2 的相关性较低,这与前人的研究结果相同^[32-33]. $PM_{2.5}$ 与 O_3 呈负相关,这是因为高浓度的 $PM_{2.5}$ 和 PM_{10} 会导致气溶胶光学厚度增大而降低光化学速率,从而减少 O_3 生成^[26].

综上所述,将气温、气压、风速和露点温度4个气象因素及 PM_{10} 、 NO_2 、 CO 和 O_3 4个污染物变量作为辅助特征输入。

表4 $PM_{2.5}$ 浓度与气象因素和辅助大气污染物相关性统计

Table 4 Correlation statistics of $PM_{2.5}$ concentration with meteorological factors and air pollutants

| 项目 | 气温 | 降水 | 气压 | 风向 | 风速 | 相对湿度 | 露点温度 | PM_{10} | SO_2 | NO_2 | CO | O_3 |
|----------|--------|--------|--------|--------|--------|--------|--------|-----------|--------|--------|--------|--------|
| 相关系数 | -0.59 | -0.15 | 0.49 | 0.05 | -0.24 | 0.17 | -0.44 | 0.87 | 0.17 | 0.68 | 0.57 | -0.44 |
| 显著性水平(P) | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 | < 0.01 |

2.2 预测结果分析与对比

2.2.1 $PM_{2.5}$ 预测结果 为了实现大气污染预警且保证工业园区管控策略有效实施,选择未来1d $PM_{2.5}$ 的日均浓度作为预测变量.考虑到计算时长及BILSTM的双向搜索,设置当前模型输入步长为1,BILSTM中前向和后向LSTM隐层状态的维度 $l=l'=32,64,128$,测试训练后的编码器-解码器结构,结果如表5所示.当BILSTM隐层状态维度最大为128时,预测精度最高.随着维度增加,结构参数和训练难度均在增加,不过,当维度从32升高到64时,精

度略有下降,因此,简单提升网络规模不能够显著增加预测准确度。

CNN模型和LSTM模型在短序列预测中表现优异^[34-37],为检验集成AEC和时空特征的CNN-BILSTM预测模型能否提高 $PM_{2.5}$ 浓度预测精度,分别建立CNN传统模型、LSTM传统模型、CNN-BILSTM传统模型、含时空特征的CNN-BILSTM预测模型、集成AEC特征的CNN-BILSTM预测模型、集成AEC和时空特征的CNN-BILSTM预测模型进行对比验证.如图9所示,

其中 CNN 模型、LSTM 模型和集成 AEC 和时空特征的 CNN-BILSTM 预测模型作纵向对比,用于验证组合模型优越性;CNN-BILSTM 传统模型、仅含时空特征或 AEC 特征的 CNN-BILSTM 预测模型及集成 AEC 和时空特征的 CNN-BILSTM 预测模型作横向对比,验证 AEC 和时空特征对模型的精度增益。

表 5 不同隐层状态维度的预测结果评价

Table 5 Prediction results evaluation of different hidden layer state dimensions

| 隐层状态维度 | RMSE | MAE | R ² |
|--------|--------|--------|----------------|
| 32 | 13.451 | 10.710 | 0.833 |
| 64 | 13.568 | 10.797 | 0.830 |
| 128 | 11.625 | 9.982 | 0.871 |

由图 9 可见,CNN 和 LSTM 模型在浓度预测精度上逊色于 CNN-BILSTM 模型;在加入时空特征后,预测值的波动减小,加入 AEC 和时空特征后,浓度预测精度在高值区和低值区表现均优异,集成 AEC 特征

的 CNN-BILSTM 模型表现较好,集成 AEC 特征而未集成时空特征的 CNN-BILSTM 模型发生预测值大于实测值的情况较多,而集成时空特征未集成 AEC 特征的 CNN-BILSTM 模型捕捉浓度高值的能力较差,集成 AEC 和时空特征的 CNN-BILSTM 模型既能在实测浓度范围内预测,又可以捕捉污染突变,综合表现最好.同时,实验表明增加了工业源排放数据的深度学习模型预测 PM_{2.5} 浓度的精度大幅提升,显然,只有深度学习的数据集完备且准确的情况下,预测精度才能明显提升,这也侧面印证了研究的园区内工业源是造成 PM_{2.5} 污染峰值的主要来源。

如图 10 所示,集成 AEC 和时空特征的 CNN-BILSTM 模型拟合效果最好,各项评价指标最优.从拟合线和辅助线的偏离情况来看,集成 AEC 和时空特征的 CNN-BILSTM 模型在高值区的拟合效果最好,其 RMSE 值为 12.912,相比传统的 LSTM 模型,精度提高 10%.证明集成 AEC 和时空特征的 CNN-BILSTM 模型有更高的预测精度。

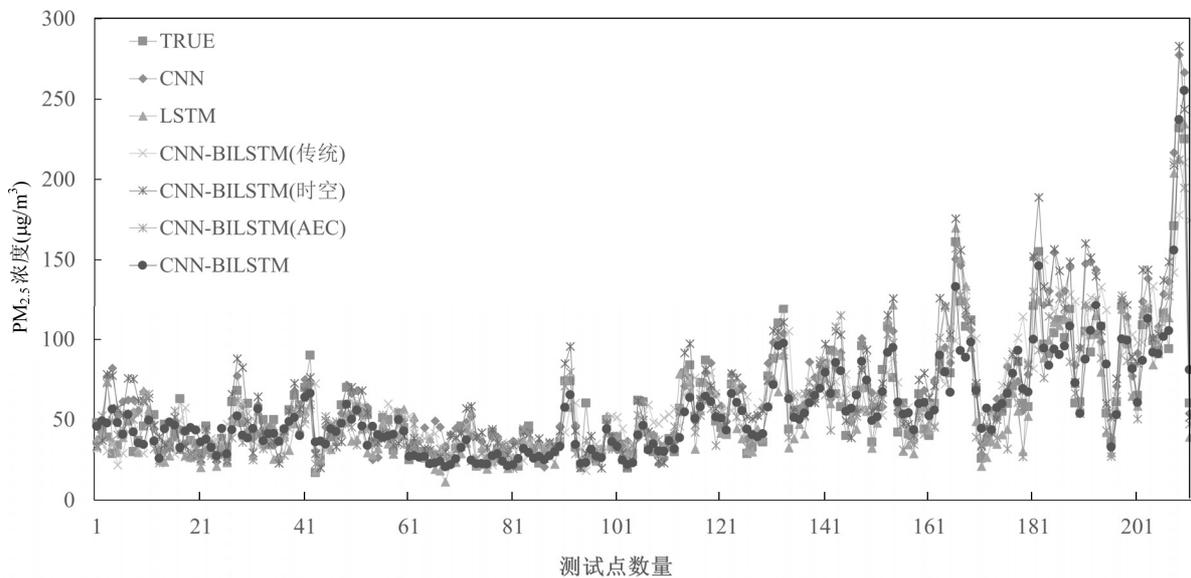


图 9 预测和实测结果对比

Fig.9 Comparison of predicted and true values

2.2.2 PM_{2.5} 污染等级预测结果 根据国家环境保护标准空气质量指数技术规范^[38],分别对 PM_{2.5} 浓度的真实值和预测值进行污染程度划分,统计污染天数和中度污染(4 级)以上的预测准确率,为了量化估计模型的不确定性,利用置信区间来获得模型的污染等级分类的真实误差,置信半径的计算公式如下

所示:

$$R_{\text{int}} = z\sqrt{v_{\text{acc}}(1-v_{\text{acc}})/n} \quad (5)$$

式中, R_{int} 是置信空间半径; v_{acc} 是分类准确率; n 是样本大小; z 是高斯分布临界值,当显著性水平为 95% 时取值 $z=1.96$ 。

如表 6 所示,CNN-BILSTM 模型对 PM_{2.5} 的浓

度高值捕捉能力明显提升,在污染天数和重度污染及严重污染的预测准确率均为最高,分别为 93%和 83%.根据其置信区间可知其真实误差为±3.4%和 ±0.3%.其中集成 AEC 和时空特征后的 CNN-

BILSTM 预测模型比传统模型在预测污染天数上的准确率平均提高 13.5%;集成 AEC 和时空特征的 CNN-BILSTM 模型比传统模型在重度污染和严重污染以上的预测准确率平均提升 50%.

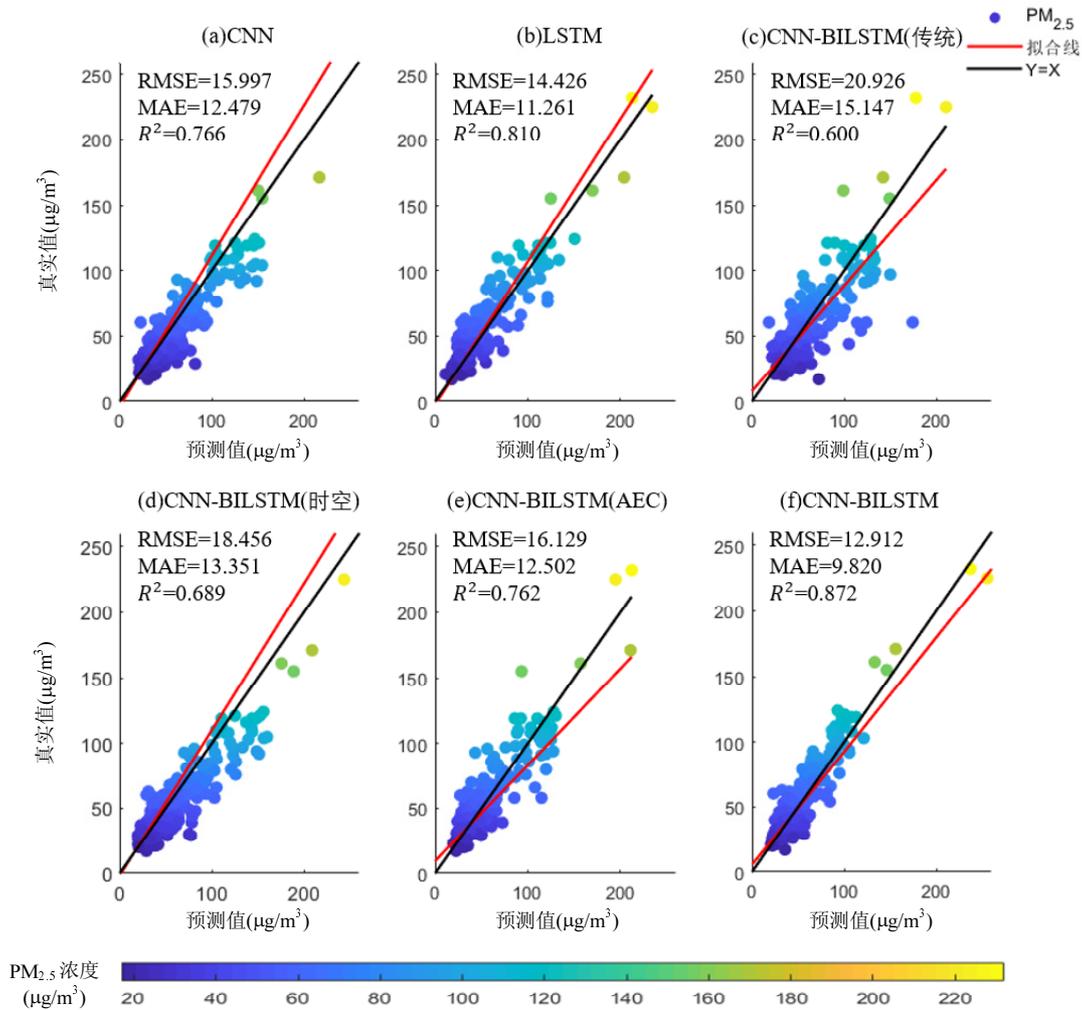


图 10 不同模型预测性能对比

Fig.10 Comparison of prediction performance of different models

表 6 不同模型污染预测准确率对比

Table 6 Comparison of pollution prediction accuracy of different models

| 项目 | 模型 | | |
|----------------|-------|-------|------------|
| | CNN | LSTM | CNN-BILSTM |
| 污染天数(d) | 33 | 40 | 43 |
| 准确率(%) | 72 | 87 | 93 |
| 置信区间半径 | 0.060 | 0.045 | 0.034 |
| 重度污染和严重污染天数(d) | 4 | 4 | 10 |
| 准确率(%) | 33 | 33 | 83 |
| 置信区间半径 | 0.063 | 0.063 | 0.003 |

2.2.3 不同季节预测 由图 11 可知,模型在秋季和冬季的预测结果较好,其中冬季的预测精度最高, RMSE、MAE 和 R² 各指标值分别为 13.846、11.482 和 0.889;而在春夏季的预测结果较差,其中夏季的预测结果最差,各指标值为 12.569、10.033、0.253. 结合濮阳市气候特点即工业园区排放特点,造成夏季预测结果较差原因是由于夏季雨量充沛,大气扩散清除能力好;而在秋冬季预测精度高是由于大气扩散条件差,园区排放量大,PM_{2.5} 浓度与影响因子的相关性更好.

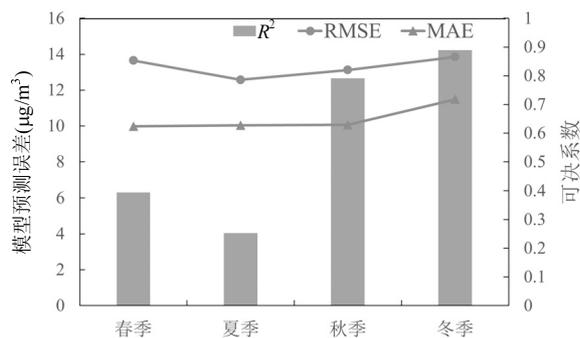


图 11 组合模型不同季节预测结果
Fig.11 Combination model forecast results in different seasons

3 结论

3.1 工业园区日均 PM_{2.5} 浓度存在显著的周期性变化且时间变化主周期为 14d;空间上 PM_{2.5} 日均浓度变化存在差异性,相邻站点的 PM_{2.5} 日均浓度变化具有高度相似性。

3.2 CNN-BILSTM 组合模型具有更高的预测精度和稳定性.与 CNN 和 LSTM 模型相比,在污染天数上的预测准确率平均提高 13.5%,在重度污染和严重污染天数的预测准确率平均提升 50%.可用于工业园区重污染事件预警预报。

3.3 AEC 特征和时空特征不仅考虑了气象、工业排放和 PM_{2.5} 浓度间的理化反应,又兼顾 PM_{2.5} 浓度变化的周期性、自相关性和空间相关性,因此集成 AEC 和时空特征的工业园区 PM_{2.5} 预测模型相对于未集成 AEC 和时空特征的模型具有更高的预测精度,且由于模型建立了工业排放和 PM_{2.5} 浓度的映射关系,既提高了模型对 PM_{2.5} 峰值的捕捉能力,亦有助于实现工业园企业污染排放精准管控。

3.4 由于季节特征差异,模型在不同季节的预测能力有所差异,秋冬季预测结果较好,而在春夏季预测能力较差。

参考文献:

- [1] Lu D, Xu J, Yang D, et al. Spatio-temporal variation and influence factors of PM_{2.5} concentrations in China from 1998 to 2014 [J]. Atmospheric Pollution Research, 2017,8(6):1151-1159.
- [2] 王 郁,徐大海,孙俊英.新疆地区大气环境容量系数的气候特征及其在空气质量变化中的作用 [J]. 环境科学学报, 2021,41(12):5073-5082. Wang Y, Xu D H, Sun J Y. Climatic characteristics of atmospheric environmental capacity coefficient in Xinjiang and its role in air quality changes [J]. Acta Scientiae Circumstantiae, 2021,41(12):5073-5082.
- [3] 朱 蓉,张存杰,梅 梅.大气自净能力指数的气候特征与应用研究 [J]. 中国环境科学, 2018,38(10):3601-3610. Zhu R, Zhang C J, Mei M. Research on climate characteristics and application of atmospheric self-purification ability index [J]. China Environmental Science, 2018,38(10):3601-3610.
- [4] 徐大海.大气环境容量系数 A 值频率曲线拟合及其应用 [J]. 中国环境科学, 2016,36(10):2913-2922. Xu D H. Frequency curve fitting and application of atmospheric environment capacity coefficient A value [J]. China Environmental Science, 2016,36(10):2913-2922.
- [5] 孙宝磊,孙 磊,张朝能,等.基于 BP 神经网络的大气污染物浓度预测 [J]. 环境科学学报, 2017,37(5):1864-1871. Sun B L, Sun H, Zhang C N, et al. Prediction of atmospheric pollutant concentration based on BP neural network [J]. Acta Scientiae Circumstantiae, 2017,37(5):1864-1871.
- [6] Dennis R L, Byun D W, Novak J H. The next generation of integrated air quality modeling: EPA's models-3 [J]. Atmospheric Environment, 1996,30(12):1925-1938.
- [7] 周广强,谢 英,吴剑斌,等.基于 WRF-Chem 模式的华东区域 PM_{2.5} 预报及偏差原因 [J]. 中国环境科学, 2016,36(8):2251-2259. Zhou G Q, Xie Y, Wu J B, et al. WRF-Chem based PM_{2.5} forecast and bias analysis over the East China Region [J]. China Environmental Science, 2016,36(8):2251-2259.
- [8] Qingxin W, Qiaolin Z, Jinhua T, et al. Estimating PM_{2.5} concentrations based on MODIS AOD and NAQPMS data over Beijing Tianjin Hebei. [J]. Sensors, 2019,19(5):1207.
- [9] Sun W, Sun J. Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm [J]. Journal of Environmental Management, 2017,188:144-152.
- [10] 宋国君,国潇丹,杨 啸,等.沈阳市 PM_{2.5} 浓度 ARIMA-SVM 组合预测研究 [J]. 中国环境科学, 2018,38(11):4031-4039. Song G J, Guo X D, Yang X, et al. ARIMA-SVM combination prediction of PM_{2.5} concentration in Shenyang [J]. China Environmental Science, 2018,38(11):4031-4039.
- [11] 李建新,刘小生,刘 静,等.基于 MRMR-HK-SVM 模型的 PM_{2.5} 浓度预测 [J]. 中国环境科学, 2019,39(6):2304-2310. Li J X, Liu X S, Liu J, et al. Prediction of PM_{2.5} concentration based on MRMR-HK-SVM model [J]. China Environmental Science, 2019, 39(6):2304-2310.
- [12] 侯俊雄,李 琦,朱亚杰,等.基于随机森林的 PM_{2.5} 实时预报系统 [J]. 测绘科学, 2017,42(1):1-6. Hou J X, Li Q, Zhu Y J, et al. Real-time forecasting system of PM_{2.5} concentration based on spark framework and random forest model [J]. Science of Surveying and Mapping, 2017,42(1):1-6.
- [13] 任才溶,谢 刚.基于随机森林和气象参数的 PM_{2.5} 浓度等级预测 [J]. 计算机工程与应用, 2019,55(2):213-220. Ren C R, Xie G. Prediction of PM_{2.5} concentration level based on random forest and meteorological parameters [J]. Computer Engineering and Applications, 2019,55(2):213-220.
- [14] 夏晓圣,陈善善,王佳佳,等.基于随机森林模型的中国 PM_{2.5} 浓度影响因素分析 [J]. 环境科学, 2020,41(5):2057-2065. Xia X S, Chen J J, Wang J J, et al. PM_{2.5} concentration influencing factors in China based on the random forest model [J]. Environmental

- Science, 2020,41(5):2057-2065.
- [15] 王敏, 邹滨, 郭宇, 等. 基于 BP 神经网络的城市 PM_{2.5} 浓度空间预测 [J]. 环境污染与防治, 2013,35(9):63-66.
Wang M, Zhou B, Guo Y, et al. BP artificial neural network-based analysis of spatial variability of urban PM_{2.5} concentration [J]. Environmental Pollution and Control, 2013,35(9):63-66.
- [16] Dai H, Huang G, Zeng H, et al. PM_{2.5} concentration prediction based on spatiotemporal feature selection using XGBoost-MSCNN-GA-LSTM [J]. Sustainability, 2021,13(21):12071.
- [17] Li S, Xie G, Ren J, et al. Urban PM_{2.5} concentration prediction via attention-based CNN-LSTM [J]. Applied Sciences, 2020,10(6):1953.
- [18] 邓丽, 郭群勇, 杨水荣. 融合 SSAE 深度特征学习和 LSTM 网络的 PM_{2.5} 小时浓度预测 [J]. 环境科学学报, 2020,40(9):3422-3434.
Deng L, Wu Q Y, Yang S R. Prediction of PM_{2.5} hourly concentration using SSAE deep feature learning and LSTM network [J]. Acta Scientiae Circumstantiae, 2020,40(9):3422-3434.
- [19] Liao Q, Zhu M, Wu L, et al. Deep learning for air quality forecasts: a review [J]. Current Pollution Reports, 2020,6(4):399-409.
- [20] 武艳, 杜景林, 全力, 等. 一种基于 PCA-IPSO 的 BiLSTM 神经网络预测 PM [J]. 现代电子技术, 2022,45(1):119-124.
Wu Y, Du J L, Quan L, et al. A BiLSTM neural network based on PCA-IPSO to predict PM [J]. Modern Electronic Technology, 2022,45(1):119-124.
- [21] Zhang M, Wu D, Xue R. Hourly prediction of PM_{2.5} concentration in Beijing based on Bi-LSTM neural network [J]. Multimedia Tools and Applications, 2021,80(16):24455-24468.
- [22] Liu H, Dong S. A novel hybrid ensemble model for hourly PM_{2.5} forecasting using multiple neural networks: a case study in China [J]. Air Quality, Atmosphere & Health, 2020,13(12):1411-1420.
- [23] 康俊锋, 谭建林, 方雷, 等. XGBoost-LSTM 变权组合模型支持下短期 PM_{2.5} 浓度预测—以上海为例 [J]. 中国环境科学, 2021,41(9):4016-4025.
Kang J F, Tan J L, Fang L, et al. Short-term PM_{2.5} concentration prediction under the support of XGBoost-LSTM variable weight combination model: Taking Shanghai as an example [J]. China Environmental Science, 2021,41(9):4016-4025.
- [24] 梁泽, 王玥瑶, 岳远素, 等. 耦合遗传算法与 RBF 神经网络的 PM_{2.5} 浓度预测模型 [J]. 中国环境科学, 2020,40(2):523-529.
Liang Z, Wang Y Y, Yue Y X, et al. PM_{2.5} concentration prediction model coupled with genetic algorithm and RBF neural network [J]. China Environmental Science, 2020,40(2):523-529.
- [25] 董红召, 王乐恒, 唐伟, 等. 集成时空特征的 PCA-PSO-SVM 臭氧 (O₃) 预测方法研究 [J]. 中国环境科学, 2021,41(2):596-605.
Dong H Z, Wang L H, Tang W, et al. Research on PCA-PSO-SVM ozone (O₃) prediction method integrating spatiotemporal features [J]. China Environmental Science, 2021,41(2):596-605.
- [26] 王平, 张红, 秦作栋, 等. 基于 wavelet-SVM 的 PM₁₀ 浓度时序数据预测 [J]. 环境科学, 2017,38(8):3153-3161.
Wang P, Zhong H, Qin Z D, et al. Prediction of PM₁₀ concentration time series data based on wavelet-SVM [J]. Environmental Science, 2017,38(8):3153-3161.
- [27] 杨玉璟, 李彬, 李源昊. 濮阳市西部工业园区对城区大气环境影响分析 [J]. 濮阳职业技术学院学报, 2015,28(4):155-157.
Yang Y J, Li B, Li Y H. Analysis of the impact of the western industrial park in Puyang on the atmospheric environment in the urban area [J]. Journal of Puyang Vocational and Technical College, 2015,28(4):155-157.
- [28] 濮阳新闻网. 市污染防治攻坚战关于工业园区管委会对工 [EB/OL]. <http://www.xiaoyukun.com/a/puyangxinwen/20191007/4019.html>2019-09-28/2019-10-07.
Puyang News Network. The Municipal Environmental Pollution Prevention and Control Office on the Industrial Park Management Committee's work [EB/OL]. <http://www.xiaoyukun.com/a/puyangxinwen/20191007/4019.html>2019-09-28/2019-10-07.
- [29] 中华人民共和国生态环境部. 濮阳市工业园区督办问题清单 [EB/OL]. <https://www.mee.gov.cn/xxgk/2018/xxgk/xxgk12/201811/W020181106387569066991.pdf>.
Ministry of Ecology and Environment of the People's Republic of China. List of Issues for Supervision of Puyang Industrial Park [EB/OL]. <https://www.mee.gov.cn/xxgk/2018/xxgk/xxgk12/201811/W020181106387569066991.pdf>.
- [30] 曾昭亮, 郭建平, 马大喜. 基于江西地区多卫星数据的气溶胶立体分布研究 [J]. 大气与环境光学学报, 2016,11(5):75-84.
Zeng Z L, Guo J P, Ma D X. Research of aerosol three-dimensional distribution based on multi-satellite data over Jiangxi [J]. Journal of Atmospheric and Environmental Optics, 2016,11(5):75-84.
- [31] 中国天气网. 濮阳 [EB/OL]. www.weather.com.cn/html/2010-05-10.
China Weather Network. Puyang [EB/OL]. www.weather.com.cn/html/2010-05-10.
- [32] 侯俊雄, 李琦, 朱亚杰, 等. 基于随机森林的 PM_{2.5} 实时预报系统 [J]. 测绘科学, 2017,42(1):1-6.
Hou J X, Li Q, Zhu Y J, et al. Real-time forecasting system of PM_{2.5} concentration based on spark framework and random forest model [J]. Science of Surveying and Mapping, 2017,42(1):1-6.
- [33] 赵文芳, 林润生, 唐伟, 等. 基于深度学习的 PM_{2.5} 短期预测模型 [J]. 南京师大学报: 自然科学版, 2019,42(3):32-41.
Zhao W F, Lin R S, Tang W, et al. PM_{2.5} short-term prediction model based on deep learning [J]. Journal of Nanjing Normal University (Natural Science Edition), 2019,42(3):32-41.
- [34] Yang B, Liu R, Zio E. Remaining useful life prediction based on a double-convolutional neural network architecture [J]. IEEE Transactions on Industrial Electronics, 2019,66(12):9521-9530.
- [35] Liu R, Yang B, Hauptmann A G. Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network [J]. IEEE Transactions on industrial informatics, 2019,16(1):87-96.
- [36] Pa A, Hk B, Bkp A. Prediction, and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India [J]. Chaos, Solitons & Fractals, 2020,139:110017.
- [37] Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM [J]. Chaos Solitons & Fractals, 2020,140:110212.
- [38] HJ 633-2012 环境空气质量指数(AQI)技术规定(试行) [S].
HJ 633-2012 Technical specification of ambient air quality index (AQI) (Trial) [S].
- 作者简介:** 董红召(1969-),男,河南滑县人,教授,博士,主要研究方向为智能环保、智能交通系统.发表论文 100 余篇.