

DOI:10.13232/j.cnki.jnju.2023.02.015

融合空间和统计特征的 CNN-GRU 臭氧浓度预测模型研究

杨雨佳¹, 肖庆来², 陈 健^{3,4}, 曾松伟^{5*}

(1. 浙江农林大学数学与计算机科学学院, 杭州, 311300; 2. 松阳县自然资源和规划局, 丽水, 323400;
3. 浙江农林大学省部共建亚热带森林培育国家重点实验室, 杭州, 311300;
4. 浙江农林大学林业与生物技术学院, 杭州, 311300; 5. 浙江农林大学光机电工程学院, 杭州, 311300)

摘 要:近年来, 空气污染问题备受关注, 近地面臭氧逐渐成为我国部分城市的首要污染物, 因此对臭氧浓度的精准预测尤为重要. 为了进一步提高臭氧浓度预测的精度, 提出一种融合空间特征和统计特征的卷积神经网络和门控循环单元 (Convolutional Neural Networks and Gate Recurrent Unit, CNN-GRU) 臭氧浓度组合预测模型. 首先, 通过对时空因素以及其他大气污染物与臭氧浓度进行相关性分析, 利用基于统计域的方法和克里金插值法对臭氧浓度时序数据进行预处理来提取臭氧浓度数据的时空特征, 采用并联合 CNN 和 GRU 结构的组合预测模型得到最终的臭氧浓度预测结果. 实验结果表明, CNN-GRU 组合预测模型预测未来一小时的臭氧浓度可决系数、均方根误差和均方误差的值分别为 0.9598, 11.9508 和 8.2753, 未来两小时的臭氧浓度可决系数、均方根误差和均方误差的值分别为 0.8985, 18.5373 和 13.0045, 优于独立的 CNN、长短期记忆 (Long Short Term Memory, LSTM) 网络、GRU、卷积-长短期记忆网络 (Convolutional LSTM Network, ConvLSTM)、CNN-LSTM 和 CNN-GRU 预测模型, 这是由于 CNN-GRU 组合预测模型融合了空间和统计特征, 可以多角度提取特征并采用并联合的网络结构, 所以预测精度较高, 且具备较好的鲁棒性.

关键词:臭氧浓度预测, 卷积神经网络, 门控循环单元, 空间特征, 统计域

中图分类号: X51

文献标志码: A

Research on CNN-GRU ozone prediction considering spatial features and statistical features

Yang Yujia¹, Xiao Qinglai², Chen Jian^{3,4}, Zeng Songwei^{5*}

(1. College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou, 311300, China;
2. Songyang County Natural Resources and Planning Bureau, Lishui, 323400, China;
3. State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, 311300, China;
4. College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, 311300, China;
5. College of Optical Mechanical and Electrical Engineering, Zhejiang A&F University, Hangzhou, 311300, China)

Abstract: In recent years, air pollution has attracted much attention and near ground ozone gradually becomes the primary pollutant in some cities in China. Therefore, accurate prediction of ozone concentration is particularly important. In order to improve the prediction accuracy of ozone concentration, In this paper, we propose a combined prediction model of ozone concentration based on CNN-GRU (Convolutional Neural Network and Gate Recurrent Unit) by fusing spatial and statistical features. Firstly, by analyzing the correlation between temporal and spatial factors as well as other atmospheric pollutants and ozone concentration, the statistical domain based method and Kriging interpolation method are used to preprocess the ozone concentration time series data to extract the spatial-temporal characteristics of the ozone concentration data. The combined

基金项目: 国家自然科学基金(41471442)

收稿日期: 2022-11-14

* 通讯联系人, E-mail: zsw@zafu.edu.cn

prediction model of parallel hybrid CNN (Convolutional Neural Networks) and GRU (Gate Recurrent Unit) structure is used to obtain the final ozone concentration prediction results. The results show that the combined prediction model of ozone concentration based on CNN-GRU has the highest accuracy and the best prediction effect, the values of R^2 , RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) of ozone concentration in the next hour are 0.9598, 11.9508 and 8.2753, in the next two hours are 0.8985, 18.5373 and 13.0045, outperform the independent CNN, LSTM (Long Short-Term Memory), GRU, ConvLSTM (Convolutional LSTM Network), CNN-LSTM and CNN-GRU prediction models. Because the prediction model combines spatial and statistical features, multi angle feature extraction and parallel hybrid network structure are used to obtain high prediction accuracy and good robustness.

Key words: ozone concentration prediction, Convolutional Neural Network, Gated Recurrent Unit, spatial feature, statistical domain

近年来,我国大气污染问题备受关注,其中,近地面的臭氧是大气污染物的一种,也是光化学烟雾的组成之一,是氮氧化物(NO_x)^[1]和挥发性有机物(VOCs)等污染物在大气中发生光化学反应生成的二次污染物^[2].高浓度臭氧不仅影响人体健康,导致多种疾病发生^[3],还阻碍农作物的生产,影响昆虫的发育和行为^[4],对生态环境产生严重的危害^[5].目前,大气环境监测站为臭氧相关研究提供了数据支持,通过监测数据可以了解当前的空气质量,但无法直接预测未来空气质量的发展趋势.对未来臭氧浓度进行预测,可为受臭氧影响的户外工作者及政府决策提供科学支持,并对生态环境起保护作用.因此,对臭氧浓度的精准预测具有重要的现实意义.

当前对臭氧浓度预测的研究主要分两类.一类是基于传统方法的预测模型,例如,吴莹和王玉祥^[6]利用嵌套空气质量预报模式系统和通用多尺度空气质量模型对臭氧进行预报,两种模型对臭氧预报的相关系数分别为0.67和0.79.此方法能对臭氧浓度进行大致的预测,但需要大量人力物力,且预测精度不高.另一类是基于机器学习的预测模型,和传统的预测模型相比,它使用臭氧浓度和相关气象因子的时间序列数据作为模型输入,众多学者搭建了随机森林^[7-9]、误差反向传播算法(Error Back Propagation, EBP)^[10-11]、决策树^[12]、分布式梯度增强库(Extreme Gradient Boosting, XGBoost)^[13-15]等数值关系模型来预测大气污染物的浓度,此类模型构造简单,便于操作.贾鹏程^[16]通过分析气象因子、前体物以及颗粒物对近地面臭氧浓度的影响,利用序列至序列

(Sequence to Sequence, seq2seq)模型对未来一小时的臭氧浓度进行预测,预测均方根误差为 $12.40 \mu\text{g}\cdot\text{m}^{-3}$,平均绝对误差为 $9.27 \mu\text{g}\cdot\text{m}^{-3}$.王馨陆等^[7]基于成都市的气象数据,使用随机森林等机器学习方法,建立了成都市夏季臭氧污染潜势模型,对臭氧的短期污染潜势有良好的预报效果,测试集的相关系数在0.73~0.76.蔡旺华^[14]运用XGBoost模型预测空气中的臭氧浓度,充分考虑臭氧浓度的时序变化规律,将气象条件和生成臭氧的重要前体物纳入模型考虑后,取得了90%的预报等级准确率.目前,臭氧浓度的预测方法着重研究臭氧的时间自相关性和臭氧与其他环境因子之间的关系,使用的机器学习模型强调提高输入数据的质量和有效性,或优化单个模型的参数.但现有的臭氧浓度组合预测模型大都简单地将两个或多个预测方法进行结合,所以对未来臭氧浓度的预测精度仍有较大的提升空间.

针对上述问题,本文提出一种融合空间特征和统计特征的CNN-GRU(Convolutional Neural Networks and Gate Recurrent Unit)臭氧浓度组合预测模型.该混合模型由统计模型和机器学习模型组合而成,采用并联杂交方法作为混合结构来整合个体预测模型,提高了模型的预测准确率.

1 数据与方法

1.1 数据及预处理 杭州市($29^{\circ}11'\text{N}\sim 30^{\circ}34'\text{N}$, $118^{\circ}20'\text{E}\sim 120^{\circ}37'\text{E}$)位于中国东部沿海的长江三角洲地区,属于亚热带季风气候,其大气污染以秋冬季 $\text{PM}_{2.5}$ 和夏季的臭氧超标为主,空气质量问题备受关注.本研究选取杭州市九个环境空气质量

国控站点 2017 年 5 月 1 日至 2019 年 6 月 29 日的同时段逐小时的气象监测数据,其中,环境污染因子主要为 O_3 , CO , SO_2 , NO_2 , $PM_{2.5}$ 和 PM_{10} . 研究区域如图 1 所示. 将 2017 年 5 月 1 日至 2018 年 4 月 30 日的 8735 条数据作为训练样本集,2018 年 5 月 1 日至 2019 年 6 月 29 日的 10175 条数据作为测试样本集. 表 1 为部分样本数据的示例.

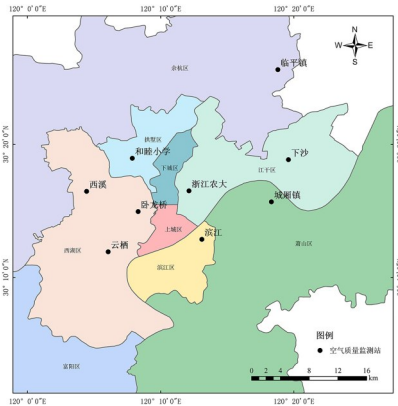


图 1 研究区域

Fig. 1 Research area

1.2 融合空间和统计特征的 CNN-GRU 模型网络结构 本文提出一种融合空间特征和统计特征的基于 CNN 和 GRU 的臭氧浓度并行组合预测的模型,图 2 为 CNN-GRU 模型的网络结构图. 为了研究邻近区域对预测站点臭氧浓度的影响,使

用克里金插值法作为数据预处理方法,并利用二维 CNN 提取空间信息,基于 CNN 和 GRU 构造网络结构分支,将两个分支和统计特征的输出以并联的方式作为全连接神经网络的输入(图 2a).

1.3 卷积神经网络(CNN)分支 CNN 在传统意义上是为图像数据设计的,它在面对复杂的网络结构时能进行高效的特征提取,被广泛运用在计算机视觉网络等领域^[17-18]. 为了使 CNN 适应时间序列数据集,增加模型精度,进一步提高网络的泛化能力,本文采用多层因果卷积来确保仅使用历史信息对臭氧浓度进行预测^[19-21]. 对于隐藏层 l 的中间特征,滤波器的表达式如下:

$$h_t^{l+1} = A((W * h)(l, t)) \quad (1)$$

$$(W * h)(l, t) = \sum_{\tau=0}^k W(l, \tau) h_{t-\tau}^l \quad (2)$$

其中, h_t^l 表示在 t 时刻 l 层的中间状态, $*$ 表示卷积运算符, W 表示 l 层的过滤器权重, $A(\cdot)$ 是例如 sigmoid 的激活函数,代表任何特定于架构的非线性处理.

在一维 CNN 的情况下,式(2)与数字信号处理中的有限长单位冲激响应滤波器存在强相似性,这也导致 CNN 在处理时间序列问题时会产生两个关键性影响. 首先,在处理时间序列问题时 CNN 保持空间不变性,使每个时间步使用的滤波器权重集相同. 此外, CNN 在处理时间序列数据

表 1 部分样本数据的示例

Table 1 Some of the data samples

日期	城市	监测点名	$PM_{2.5} (\mu g \cdot m^{-3})$	$PM_{10} (\mu g \cdot m^{-3})$	$SO_2 (\mu g \cdot m^{-3})$	$NO_2 (\mu g \cdot m^{-3})$	$O_3 (\mu g \cdot m^{-3})$	$CO (\mu g \cdot m^{-3})$
2018/3/3 10:00	杭州	浙江农大	18	25	6	37	24	1.3
2018/3/3 11:00	杭州	浙江农大	17	22	7	34	36	1.2
2018/3/3 12:00	杭州	浙江农大	15	20	7	31	56	1.2
2018/3/3 13:00	杭州	浙江农大	16	20	7	30	78	1.2
2018/3/3 14:00	杭州	浙江农大	15	19	7	30	95	1.1
2018/3/3 15:00	杭州	浙江农大	17	18	7	33	107	1.1
2018/3/3 16:00	杭州	浙江农大	20	20	7	31	123	1.1
2018/3/3 17:00	杭州	浙江农大	25	NULL	8	44	112	1.0
2018/3/3 18:00	杭州	浙江农大	33	33	10	72	75	1.1
2018/3/3 19:00	杭州	浙江农大	47	51	20	68	78	1.2
2018/3/3 20:00	杭州	浙江农大	58	64	20	81	59	1.3
2018/3/3 21:00	杭州	浙江农大	66	74	12	59	94	1.1
2018/3/3 22:00	杭州	浙江农大	67	70	12	54	84	1.1
2018/3/3 23:00	杭州	浙江农大	69	73	11	60	55	1.1

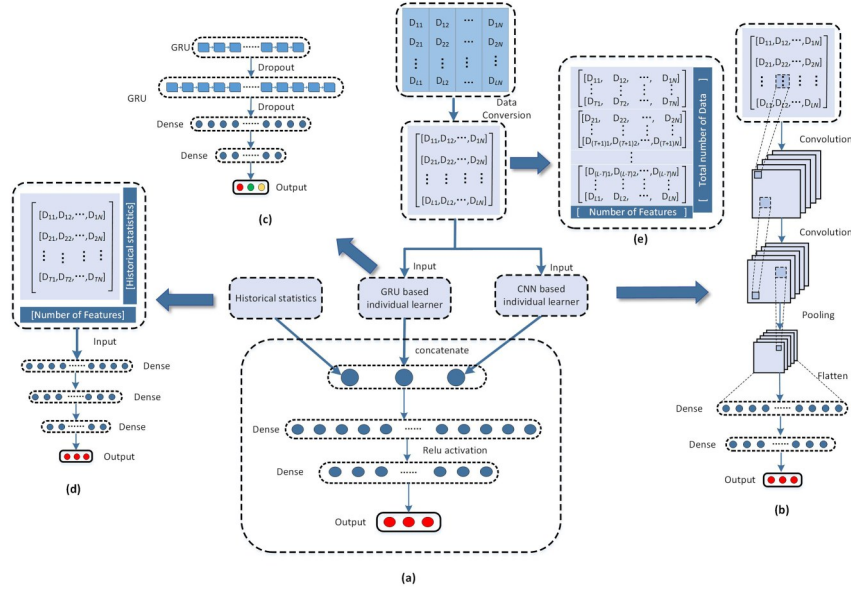


图2 融合空间和统计特征的CNN-GRU模型网络结构

Fig. 2 Network structure diagram of CNN-GRU model integrating spatial and statistical features

时,只能通过其定义的回溯窗口或接受域中的输入进行预测,因此需要对感受野大小 k 进行精准调节,以确保模型对历史信息的充分利用。

为了保证预测效果的稳定性并降低网络模型的复杂度,在CNN的网络架构上,本文将两层重复的卷积层连接在输入层之后,用较少的网络层尽可能多地提取特征,避免经过池化层后丢失重要的特征。经过连续的卷积和池化后,模型将获得的特征平化为一维,选用 3×3 的卷积核,池化层大小为 2×2 ,神经元数量分别设置为32个和2个,最后将数据输入Dense层(图2b)。使用线性整流函数作为激活函数,其具有的稀疏激活特性可以避免过拟合现象,其表达式如下:

$$f(x) = \max(0, x) \quad (3)$$

1.4 门控循环单元(GRU)分支 2014年Cho et al^[22]提出的GRU是基于循环神经网络(Recurrent Neural Network, RNN)的一种门控机制,在保留过去信息影响的同时,遗忘一些非必要的信息。和长短期记忆(Long Short-Term Memory, LSTM)网络相比,GRU解决了处理长时间序列数据时梯度消失的问题,同时除去了输出门,使参数数量更少,进一步提高了训练效率。因此,本文使用的模型采用双层GRU结构,采用一对一和多对一两种递归单元展开形式,使模型在每一个时间步对输

入的时间序列数据产生新的记忆状态,并能在最后的一个时间步对记忆进行更新,其计算式如下:

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (4)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (5)$$

$$\hat{h}_t = \tanh(W[\hat{h}_t, x_t]) \quad (6)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \hat{h}_t \quad (7)$$

其中, x_t 表示神经元的输入, W 表示候选隐藏状态权重矩阵, W_z 表示更新门权重矩阵, W_r 表示重置门权重矩阵, σ 表示sigmoid函数, h_t 表示隐藏层输出。式(4)和式(5)分别为更新门和重置门的计算式,GRU将LSTM中的遗忘门和输入门整合成一个更新门,结构上更优化。更新门决定是否抛弃前一时刻的隐藏状态 h_{t-1} 的无用信息。GRU的模型内部结构如图3所示。由图可见,通过当前

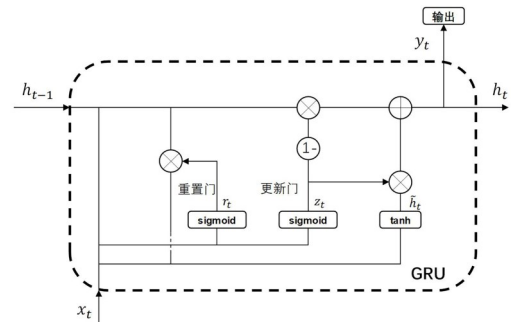


图3 GRU模型的结构图

Fig. 3 The structure of GRU model

时刻的输入 x_t 和前一时刻的隐藏状态 h_{t-1} , GRU 就能计算此时隐藏节点的输出 y_t 和传送给下一节点的隐藏状态 h_t .

1.5 基于统计域的时序特征分支 时间序列存在自相关性,其某一时刻的值会受到另一时刻的值的影响.本文使用基于统计域的时间特征作为混合模型的另一个分支,统计预测时间点前一小时臭氧浓度值、前三小时平均臭氧浓度值、24 h 前同一时间点与下一小时的差值和 24 h 前下一小时与下两小时的差值,将特征提取后作为分支与 CNN 分支和 GRU 分支并行,将得到的输出输入全连接神经网络.

1.6 评估指标 为了更直观地评价模型的预测精度,使用平均绝对误差(MAE)、均方根误差(RMSE)、可决系数(R^2)对臭氧浓度的预测结果进行评价.表达式如下:

$$MAE = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (\bar{y} - y_i)^2} \quad (10)$$

其中, \hat{y}_i 为预测值, y_i 为真实值, \bar{y} 为真实值的平均值, m 为样本数量.

2 实验结果与分析

2.1 臭氧与气象因子相关性分析 图 4 为臭氧浓度与气象因子的相关性热力图,气象因子包括湿度、累计日照、气温、风速、CO、NO₂、SO₂、PM_{2.5} 和 PM₁₀. 由图可知,气温与臭氧浓度呈正相关,其相关系数比累计日照、平均风速高,这是因为温度高导致光化学反应增强,使臭氧浓度升高^[23]. 湿度与臭氧浓度呈现负相关,高湿度的环境容易抑制臭氧的产生^[24]. 空气颗粒物之间存在物理化学层面的相互转化,同时在运输过程中也会相互产生影响,它们之间存在较复杂的关系^[25]. 还可以看出,臭氧与其他的大气污染物呈现明显的负相关关系,其中与 NO₂ 的相关系数最高,为 0.593. NO₂ 和 CO 属于臭氧的前体物,在发生光化学反应后生成臭氧,因此其与臭氧浓度呈负相关关系,相关系数分别为 0.593 和 0.309. 当空气中 PM_{2.5}

和 PM₁₀ 含量较高时,会使气溶胶的光学厚度加大,能减缓臭氧的形成,因此其与臭氧浓度同样呈负相关关系^[26],相关系数分别为 0.177 和 0.22. 综上所述,这些气象因子均对臭氧浓度有一定影响,因此将这些气象因子作为辅助变量加入模型,作为特征输入再进行臭氧浓度预测,可以提升臭氧浓度预测的精度.

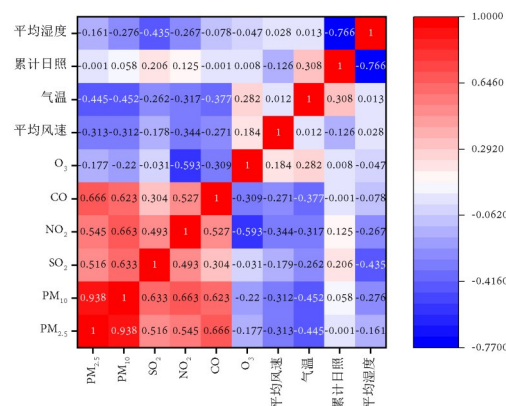


图 4 臭氧浓度与其余气象因子相关性分析

Fig. 4 Correlation analysis between ozone concentration and other meteorological factors

2.2 时间特征分析 分析时间序列数据的自相关性能更好地掌握目标因子的变化规律.图 5 展示了 2017—2019 年浙江农大(浙江大学华家池校区)站的臭氧浓度变化,其呈现明显的季节性特征,是夏季高、冬季低,有很强的时间相关性.

各个站点臭氧浓度的日平均变化如图 6 所示.由图可知,臭氧浓度的日平均变化呈典型的单峰型分布^[27],夜间浓度较低,自上午起臭氧的

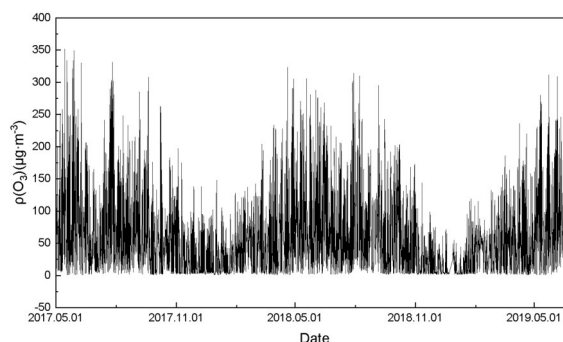


图 5 2017—2019 年浙江农大站点臭氧浓度的变化

Fig. 5 Concentration of ozone at Zhejiang Agricultural University station from 2017 to 2019

光化学反应增强,臭氧浓度逐渐增加,一般在下午达到峰值,随后持续下降.

2.3 空间特征分析 由于国控监测站点分布不均,监测站点的稀疏性以及存在的监测空白地区无法满足对臭氧浓度的日常监测需求.为了探究臭氧浓度的空间特性,对杭州市九个环境空气质量国控站点进行臭氧浓度预测研究.由图6可知,九个临近的站点基于时间序列的臭氧浓度日趋势基本保持一致,表现出较高的空间相关性.

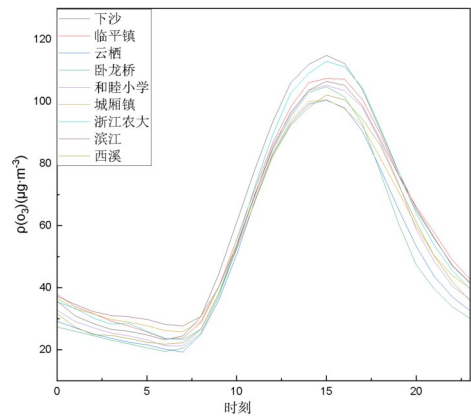


图6 各站点臭氧浓度的日平均变化
Fig. 6 Daily average of ozone concentration at each station

由于污染物分散于不同站点,分析相邻站点的数据可以对臭氧浓度的空间相关性进行补充.计算各站点间臭氧浓度的皮尔逊相关系数,其热力图如图7所示.由图可见,各站点之间的臭氧浓度相关性各不相同.相关系数在0.86~0.96,说

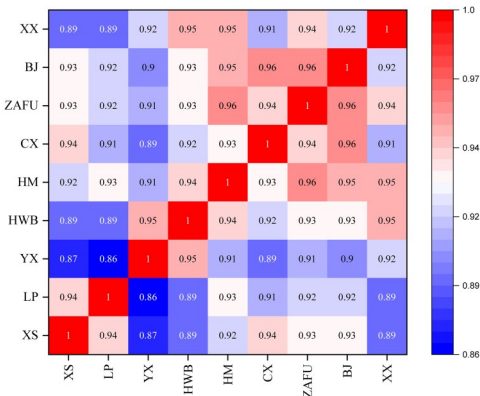


图7 站点与站点之间的臭氧浓度相关性分析
Fig. 7 Correlation analysis of ozone concentration between stations

明一定程度上各站点的臭氧浓度会彼此影响.因此,在模型中加入空间特征可以提高预测精度.

2.4 空间特征对预测结果的影响 通过空间特征分析可知,在一定程度上,每个站点所在的区域会对周围站点有不同程度的影响,因此可以提取预测点周围站点的空间信息以提高臭氧浓度预测的准确率.浙江农大站位于杭州市九个站点分布的偏中心位置,所以预测模型选用浙江农大站的臭氧浓度数据为代表.使用不基于空间特征的CNN, LSTM, GRU 和基于空间特征的 CNN-LSTM, CNN-GRU 模型对未来一小时和两小时的臭氧浓度进行预测,并进行比较分析,结果如表2所示.由表可知,在预测未来一小时的臭氧浓度时基于空间特征的CNN-GRU模型效果最佳, R^2 可达0.9468, $RMSE$ 为13.4237, MAE 为9.6685, 和不基于空间特征的模型(NSpa-LSTM)相比, R^2 , $RMSE$ 和 MAE 分别提升1.53%, 10.21% 和 6.74%. 证明基于空间特征的CNN-GRU模型优于基于空间特征的CNN-LSTM, 远优于不基于空间特征的单一模型.实验对未来两小时的臭氧浓度的预测效果更显著, R^2 , $RMSE$ 和 MAE 分别提升1.89%, 4.60% 和 7.65%, 充分证明增加空间特征对提升臭氧浓度预测精度有良好的效果.

2.5 统计特征对预测结果的影响 通过时间特征分析可知,时间序列存在强自相关性,某一时刻的值会影响另一时刻的值.如表3所示,统计前一

表2 空间特征对臭氧浓度预测结果的影响
Table 2 Influence of spatial characteristics on the prediction of ozone concentration

对比实验 模型	预测未来 时间 (h)	指标		
		R^2	$RMSE$	MAE
NSpa-CNN	1	0.9380	14.2636	10.4514
	2	0.8698	21.9167	14.9628
NSpa-LSTM	1	0.9325	14.9497	10.3668
	2	0.8633	21.5153	14.7754
NSpa-GRU	1	0.9359	14.7534	10.6444
	2	0.8626	21.2264	14.4919
Spa-CNN-LSTM	1	0.9422	13.9816	10.0346
	2	0.8709	20.9081	13.7600
Spa-CNN-GRU	1	0.9468	13.4237	9.6685
	2	0.8789	20.2510	13.3830

小时臭氧浓度预测未来一小时臭氧浓度的模型的 R^2 达到 0.9209, $RMSE$ 和 MAE 分别为 16.3637 和 10.5260, 统计前三小时臭氧浓度平均值预测未来一小时臭氧浓度模型的 R^2 也达到 0.7826. 因此, 考虑将该基于统计域的时间特征作为混合模型的另一个分支. 基于时间特征和统计特征的 CNN-GRU 模型预测未来一小时臭氧浓度模型的 R^2 提高到 0.9598, $RMSE$ 和 MAE 分别为 11.9508 和 8.2753, 预测未来两小时臭氧浓度模型的 R^2 , $RMSE$ 和 MAE 比传统的统计模型分别提升 56.45%, 51.19% 和 48.68%, 证明将统计特征作为模型分支的方法有更好的预测精度.

表 3 统计特征对臭氧浓度预测结果的影响

Table 3 Influence of statistical characteristics on the prediction of ozone concentration

对比实验 模型	预测未来 时间 (h)	指标		
		R^2	$RMSE$	MAE
Statical_last1	1	0.9209	16.3637	10.5260
	2	0.7540	28.8671	18.8714
Statical_last3	1	0.7826	27.1368	17.8993
	2	0.5743	37.9750	25.3378
Spa-sta-	1	0.9598	11.9508	8.2753
CNN-GRU	2	0.8985	18.5373	13.0045

2.6 网络模型结构对预测结果的影响 为了探究不同的网络模型结构对预测精度的影响, 将 CNN-GRU 并行模型、CNN-LSTM 并行模型和 ConvLSTM 模型进行对比, 实验结果如表 4 所示. 由表可知, 预测结果的精度从高到低分别为 CNN-GRU 并行模型、CNN-LSTM 并行模型、ConvLSTM 串行模型. CNN-GRU 并行模型预测未来一小时臭氧浓度的 R^2 , $RMSE$ 和 MAE 比串行模型分别提升 2.01%, 13.98% 和 17.07%, 预测未来两小时臭氧浓度的 R^2 , $RMSE$ 和 MAE 比 ConvLSTM 串行模型分别提升 5.68%, 15.17% 和 23.95%. 证明通过结合 CNN 的高维特征提取能力和 GRU 的时间序列存储能力可以提高模型在处理复杂数据过程中的拟合能力, 本文提出的预测模型通过分支结构的并行, 和传统的串行结构相比, 可以多角度抽取特征来达到更好的预测效果.

表 4 网络模型结构对臭氧浓度预测结果的影响

Table 4 Influence of network model structure on the prediction of ozone concentration

对比实验模型	预测未来 时间 (h)	指标		
		R^2	$RMSE$	MAE
Spa-	1	0.9281	15.6056	11.6580
ConvLSTM	2	0.8317	23.8724	17.5970
Spa-CNN-LSTM (parallel)	1	0.9422	13.9816	10.0346
	2	0.8709	20.9081	13.7600
Spa-CNN-GRU (parallel)	1	0.9468	13.4237	9.6685
	2	0.8789	20.2510	13.3830

2.7 不同模型对臭氧浓度预测结果对比 为了验证基于空间特征和统计特征的 CNN-GRU 组合预测模型的预测精度, 选择统计前一小时的臭氧浓度、统计前三小时平均臭氧浓度、无空间特征的 CNN、LSTM、GRU、基于空间特征的 Conv-LSTM、基于空间特征的 CNN-LSTM 和基于空间特征的 CNN-GRU 组合预测模型进行对比实验. 不同模型未来一小时、两小时的预测值与实际值的对比如图 8~10 所示.

由图 8 可知, 基于空间特征和统计特征的 CNN-GRU 组合预测模型的精度比其他模型更高, 预测未来两小时臭氧浓度效果更明显, R^2 提升了 8.0%.

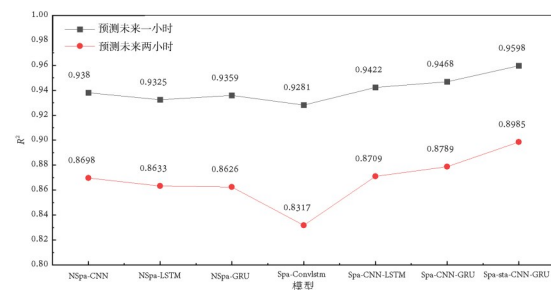


图 8 不同模型的预测准确率

Fig. 8 Prediction accuracy of different models

由图 9 和图 10 可知, 和统计模型相比, 基于机器学习预测模型的拟合效果更好, 准确率更高, 使用单一预测模型的拟合效果明显劣于组合模型. 在组合模型中, ConvLSTM 模型采用串联的网络结构, 预测精度不如其他机器学习预测模型. 调整模型的网络结构, 将串联的网络结构改为并联, 改进后并联的 CNN-GRU 组合预测模型的拟合

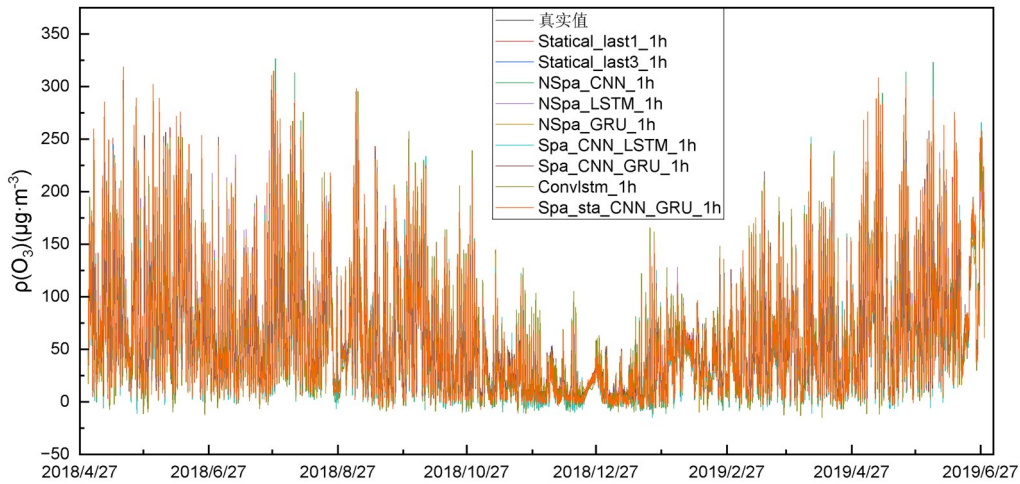


图9 各模型对未来一小时臭氧浓度预测和实测结果的对比

Fig. 9 Predicted and measured results of ozone concentration in the next hour by each mode

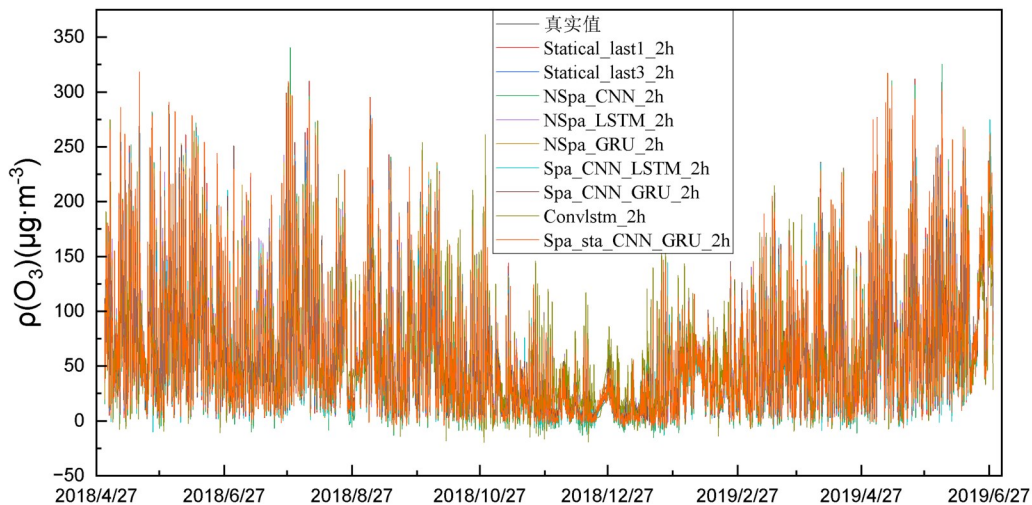


图10 各模型对未来两小时臭氧浓度预测和实测结果的对比

Fig. 10 Predicted and measured results of ozone concentration in the next two hours by each model

效果更好.而在增加空间特征和统计特征后,拟合效果进一步提升,表明基于空间特征和统计特征的CNN-GRU组合预测模型具有更高的预测精度.

2.8 不同季节预测结果分析 研究位于典型亚热带季风气候的区域的不同季节对臭氧浓度预测的影响,针对不同季节进行预测分析,分别为春季(3,4,5月)、夏季(6,7,8月)、秋季(9,10,11月)和冬季(12,1,2月).

实验结果如表5所示.由表可知,本研究的基于空间特征和统计特征的CNN-GRU臭氧浓度组合预测模型在春季和冬季的预测效果较好,其

表5 组合预测模型对臭氧浓度的季节预测结果

Table 5 Seasonal prediction of ozone concentration by the combined prediction model

季节	预测未来 时间 (h)	指标		
		R^2	RMSE	MAE
春季	1	0.9458	13.8409	9.9245
	2	0.8630	22.0168	15.6293
夏季	1	0.9445	15.0720	10.7980
	2	0.8629	23.7048	16.9528
秋季	1	0.8890	7.9106	5.6158
	2	0.7794	11.1586	7.6534
冬季	1	0.9520	9.8347	6.6402
	2	0.8877	15.0586	10.0924

中冬季的臭氧浓度预测精度最高, R^2 , $RMSE$ 和 MAE 分别为 0.9520, 9.8347 和 6.6402, 优于其他季节. 夏季和秋季的预测结果较差, 其中夏季的预测结果最差, R^2 , $RMSE$ 和 MAE 分别为 0.9445, 15.0720 和 10.7980. 夏季预测结果较差的原因是夏季光化学反应的增强有利于夏季的臭氧生成, 而夏季大气对流旺盛, 形成了较好的大气污染物扩散的气象条件^[28].

2.9 适用性分析 为了检验基于空间特征和统计特征的 CNN-GRU 臭氧浓度组合预测模型的适用性和泛化性, 对站点未来 3~7 天的臭氧浓度日均值进行预测, 并以 R^2 , $RMSE$ 和 MAE 三个评价指标对预测结果进行评价. 将模型的输入设置为预测时间前 30 天的臭氧日均值, 统计预测时间前一天的臭氧浓度日均值、前三天平均臭氧浓度日均值及前七天平均臭氧浓度日均值输入模型. 实验结果如表 6 所示. 由表可知, 该模型对未来 3~7 天的臭氧浓度仍有较高的预测精度, 证明本文模型具有较好的泛化性, 加入时间特征使模型能学习时序变化, 加入空间特征使模型能结合区域性的影响, 加入统计特征使模型的精度得到了进一步提升. 通过设置不同的输入参数来训练模型, 使模型能达成短期甚至中长期的臭氧浓度预测. 实验结果证明, 本模型可应用于未来更长时间的臭氧浓度预测.

表 6 组合预测模型对未来 3~7 天臭氧浓度的预测结果
Table 6 Prediction of ozone concentration in the next 3~7 days by the combined prediction model

预测未来 时间 (d)	指标		
	R^2	$RMSE$	MAE
3	0.9071	17.6538	13.4826
4	0.8981	16.5204	15.0985
5	0.8650	20.7062	15.3523
6	0.8559	22.8132	17.0422
7	0.8267	24.2813	19.4327

3 结论

本文研究针对臭氧浓度的短期预测. 通过臭氧浓度的时间和空间特征分析, 发现杭州市的臭氧浓度有显著的周期性变化和自相关性, 同时, 气

象因子对臭氧浓度变化也存在一定影响. 在空间上, 相邻站点的臭氧浓度有区域相关性. 本文的组合预测模型充分利用站点臭氧浓度的空间特征, 提高了预测精度. 基于统计域的时序特征可以消除臭氧浓度度量时间序列数据的相关性, 增加统计特征作为网络结构分支进行模型预测, 进一步提高了本文模型的预测精度.

本文基于空间特征和统计特征的 CNN-GRU 臭氧浓度组合预测模型从多角度提取特征, 并用并行的网络结构将特征进行融合. 通过对比实验, 对臭氧浓度的预测结果明显优于其他对比模型, 预测未来一小时臭氧浓度的 R^2 , $RMSE$ 和 MAE 分别为 0.9598, 11.9508 和 8.2753, 预测未来两小时的臭氧浓度 R^2 , $RMSE$ 和 MAE 分别为 0.8985, 18.5373 和 13.0045. 模型对未来更长时效的预测也具有较高的预测精度.

参考文献

- [1] 叶绿萌, 樊少芬, 常鸣, 等. 珠三角地区秋季臭氧生成敏感性时空变化模拟研究. 南京大学学报(自然科学), 2016, 52(6): 977—988. (Ye L M, Fan S F, Chang M, et al. Spatial and temporal distribution of ozone sensitive district in Pearl River Delta region during autumn. Nanjing University (Natural Science), 2016, 52(6): 977—988.)
- [2] 董红召, 王乐恒, 唐伟, 等. 融合时空特征的 PCA-PSO-SVM 臭氧(O_3)预测方法研究. 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 prediction considering spatial-temporal features. China Environmental Science, 2021, 41(2): 596—605.)
- [3] 郭禹慧. 中国臭氧污染的时空格局及健康影响评估研究. 硕士学位论文. 西安: 西北大学, 2021. (Guo Y H. Study on the spatial and temporal patterns of ozone pollution and health impact assessment in China. Master Dissertation. Xi'an: Northwest University, 2021.)
- [4] 毕洁, 文明明, 喻莉君, 等. 臭氧在储粮害虫防治中的应用研究进展. 河南工业大学学报(自然科学版), 2022, 43(1): 131—138. (Bi J, Wen M M, Yu L J, et al. Research progress of ozone application in the control of stored grain pests. Journal of Henan

- University of Technology (Natural Science Edition), 2022, 43(1): 131—138.)
- [5] 冯兆忠,袁相洋,李品,等.地表臭氧浓度升高对陆地生态系统影响的研究进展.植物生态学报,2020,44(5):526—542. (Feng Z Z, Yuan X Y, Li P, et al. Progress in the effects of elevated ground-level ozone on terrestrial ecosystems. Chinese Journal of Plant Ecology, 2020, 44(5): 526—542.)
- [6] 吴莹,王玉祥. NAQPMS 和 CMAQ 模式在臭氧预报应用中的效果检验.四川环境,2019,38(1):81—84. (Wu Y, Wang Y X. The effects of NAQPMS model and CMAQ model in ozone forecasting applications. Sichuan Environment, 2019, 38(1): 81—84.)
- [7] 王馨陆,黄冉,张雯娴,等.基于机器学习方法的臭氧和PM_{2.5}污染潜势预报模型——以成都市为例.北京大学学报(自然科学版),2021,57(5):938—950. (Wang X L, Huang R, Zhang W X, et al. Forecasting ozone and PM_{2.5} pollution potentials using machine learning algorithms: A case study in Chengdu. Acta Scientiarum Naturalium Universitatis Pekinensis, 2021, 57(5): 938—950.)
- [8] 马润美,张亚一,班婕,等.基于随机森林模型的京津冀地区近地面臭氧站点浓度预测.环境与健康杂志,2019,36(11):954—957. (Ma R M, Zhang Y Y, Ban J, et al. Prediction of ground ambient ozone site concentration in Beijing-Tianjin-Hebei region based on random forest model. Journal of Environment and Health, 2019, 36(11): 954—957.)
- [9] 黄厘博.重庆市主城区空气质量分析及PM_{2.5}浓度预测.硕士学位论文.重庆:西南大学,2020. (Huang L B. Analysis of air quality and prediction of PM_{2.5} concentration in the main district of Chongqing. Master Dissertation. Chongqing: Southwest University, 2020.)
- [10] 张文方.北京城区臭氧浓度变化规律及预测方法研究.硕士学位论文.北京:北京林业大学,2010. (Zhang W F. Study on the variation and prediction method of ozone concentration in Beijing Urban Area. Master Dissertation. Beijing: Beijing Forestry University, 2010.)
- [11] 刘宇轩.基于后向传播神经网络的典型道路PM_{2.5}和臭氧预测模型研究.硕士学位论文.杭州:浙江大学,2020. (Liu Y X. Study on prediction of roadside PM_{2.5} and ozone by using back propagation neural network. Master Dissertation. Hangzhou: Zhejiang University, 2020.)
- [12] 丁慷,陈报章,王瑾,等.基于决策树的统计预报模型在臭氧浓度时空分布预测中的应用研究.环境科学学报,2018,38(8):3229—3242. (Ding S, Chen B Z, Wang J, et al. An applied research of decision-tree based statistical model in forecasting the spatial-temporal distribution of O₃. Acta Scientiae Circumstantiae, 2018, 38(8): 3229—3242.)
- [13] 郑旭曼.基于集成学习的O₃浓度逐小时预测模型研究.硕士学位论文.上海:华东师范大学,2018. (Zheng X M. Hourly prediction model of ozone concentrations based on ensemble learning. Master Dissertation. Shanghai, China: East China Normal University, 2018.)
- [14] 蔡旺华.运用机器学习方法预测空气中臭氧浓度.中国环境管理,2018,10(2):78—84. (Cai W H. Using machine learning method for predicting the concentration of ozone in the air. Chinese Journal of Environmental Management, 2018, 10(2): 78—84.)
- [15] 虞益军,曾国辉,黄勃,等.改进的Prophet融合误差预测模型应用于大气二氧化硫时序预测.南京大学学报(自然科学),2022,58(3):440—447. (Yu Y J, Zeng G H, Huang B, et al. Improved prophet fusion error prediction model applied to atmospheric sulfur dioxide time series prediction. Journal of Nanjing University (Natural Science), 2022, 58(3): 440—447.)
- [16] 贾鹏程.基于深度学习的长三角地区臭氧临近预报技术研究.硕士学位论文.南京:南京信息工程大学,2021. (Jia P C. Study on ozone near forecast technology in Yangtze River Delta Based on deep learning. Master Dissertation. Nanjing: Nanjing University of Information Science and Technology, 2021.)
- [17] Selby B, Kockelman K M. Spatial prediction of traffic levels in unmeasured locations: Applications of universal kriging and geographically weighted regression. Journal of Transport Geography, 2013 (29): 24—32.
- [18] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84—90.
- [19] Van Den Oord A, Dieleman S, Zen H G, et al.

- WaveNet: A generative model for raw audio. 2016, arXiv:1609.03499.
- [20] Bai S J, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. 2018, arXiv: 1803.01271.
- [21] Borovykh A, Bohte S, Oosterlee C W. Conditional time series forecasting with convolutional neural networks. 2017, arXiv:1703.04691.
- [22] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014, arXiv:1406.1078.
- [23] 曹不凡. 影响中国东部夏季臭氧污染的大气环流及其预估. 硕士学位论文. 南京:南京信息工程大学, 2021. (Cao B F. Atmospheric circulation affecting ozone pollution in eastern China in summer and its prediction. Master Dissertation. Nanjing: Nanjing University of Information Science and Technology, 2021.)
- [24] 魏海茹, 薛鹏, 薛铭, 等. 运城市静稳天气状况下环境空气主要污染物与气象因子的相关关系研究//中国环境科学学会 2022 年科学技术年会论文集(一). 中国环境科学学会, 2022: 540—545.
- [25] 侯俊雄, 李琦, 朱亚杰, 等. 基于随机森林的 $\text{PM}_{2.5}$ 实时预报系统. 测绘科学, 2017, 42(1): 1—6. (Hou J X, Li Q, Zhu Y J, et al. Real-time forecasting system of $\text{PM}_{2.5}$ concentration based on spark framework and random forest model. Science of Surveying and Mapping, 2017, 42(1): 1—6.)
- [26] 吴锴, 康平, 于雷, 等. 2015—2016 年中国城市臭氧浓度时空变化规律研究. 环境科学学报, 2018, 38(6): 2179—2190. (Wu K, Kang P, Yu L, et al. Pollution status and spatio-temporal variations of ozone in China during 2015—2016. Acta Scientiae Circumstantiae, 2018, 38(6): 2179—2190.)
- [27] 毛毛, 彭珍. 2018 年中国中东部地区臭氧浓度分布的时空特征. 南京大学学报(自然科学), 2021, 57(3): 460—465. (Mao M, Peng Z. Spatial and temporal variations of ozone concentrations in the mid-east part of China in 2018. Journal of Nanjing University (Natural Science), 2021, 57(3): 460—465.)
- [28] 黄智磊, 王振楠, 张玉娜, 等. 温县大气污染特征及气象因素分析//中国环境科学学会 2022 年科学技术年会论文集(一). 中国环境科学学会, 2022: 523—534.

(责任编辑 杨可盛)