

黎嘉明. 空气质量预测的深度学习模型研究与实践[J]. 智能计算机与应用, 2024, 14(7): 182-189. DOI: 10.20169/j.issn.2095-2163.240729

# 空气质量预测的深度学习模型研究与实践

黎嘉明

(广东省生态环境监测中心, 广州 510000)

**摘要:** 及时和准确的空气质量预测数据对于环境管理至关重要, 尤其是在空气重污染期间, 预测数据可以为政府生态环境管理部门应对污染状况、精准地调配社会资源的决策提供数据支撑。本文提出的基于人工智能的深度学习模型 AirNet6, 可以兼顾准确性和实时性, 实现臭氧、二氧化硫、一氧化碳等因子的 7 天甚至更长时间的空气质量预测。与传统的化学模型演算不同, 本模型使用时空图卷积网络 (STGCN), 捕获历史监测数据、天气预测数据、社会活动等数据的规律, 在 2 min 内完成一百多个点位未来 168 h 数据的预测。实验表明, AirNet6 模型在速度、节能和准确度上, 比传统的化学模型及时间序列 AI 模型均有明显进步。

**关键词:** 空气质量预测; 人工智能; 深度学习模型; 时空图卷积网络

中图分类号: X824; TP183

文献标志码: A

文章编号: 2095-2163(2024)07-0182-08

## Research and practice of deep learning model for air quality prediction

LI Jiaming

(Guangdong Province Ecological Environment Monitoring Center, Guangzhou 510000, China)

**Abstract:** Timely and accurate air quality prediction data is very important for environmental management, especially, during the period of heavy air pollution. The prediction data can provide data support for the decision-making of the government's ecological environment management departments to cope with the pollution situation and accurately allocate social resources. The artificial intelligence based deep learning model AirNet6 proposed in this article can give consideration to both accuracy and real-time performance to achieve 7-day or longer air quality prediction for ozone, sulfur dioxide, carbon monoxide and other factors. Unlike traditional chemical model calculations, this model base on Spatio-Temporal Graph Convolutional Networks (STGCN), which captures the laws of historical monitoring data, weather prediction data, social activities and other data, and completes the prediction of more than one hundred points for the next 168 hours in two minutes. Experiments show that the AirNet6 model has made significant progress in speed, energy efficiency, and accuracy compared to traditional chemical models and time series AI models.

**Key words:** air quality prediction; artificial intelligence; depth learning model; STGCN

## 0 引言

传统的基于化学模型的空气质量预测方法, 一般是基于污染源排放数据、天气预测数据通过复杂的模拟运算得到结果<sup>[1]</sup>。庞大的计算量需要大规模服务器集群甚至超级计算机等计算资源, 并耗时数个小时才能生成结果, 因此通常只能做到每天预测一次<sup>[2]</sup>。而空气质量预测的数据对环境管理很重要, 是政府做出决策、调配资源、采取应对措施的基本依据, 在重污染天气时, 管理机构和专家需要实时应对天气和污染状况作出决策, 需要更加及时反映当前时刻的预测数据, 化学模型的计算速度很难满足需求<sup>[3]</sup>。

而随着人工智能技术的发展, 基于机器学习的时间序列预测模型可以根据历史的数据, 推理出未来一段时间的预测数据。比如在气象预报中, 可以使用 LSTM (Long Short-Term Memory) 模型, 通过前一段时间序列数据, 推测出后续的气温数据变化<sup>[4]</sup>。经改进的 LSTM 模型, 融合栈式稀疏自编码器 (Stack Sparse Auto-Encoder, SSAE), 在预测京津冀地区 PM2.5 数据时, 也取得很好的效果<sup>[5]</sup>。但时间序列的模型只是单纯地捕捉测点的数据随时间变化的规律, 并没有考虑测点之间的关联因素, 也没有关联其他环境的因素, 因此准确度仍有不足。

时空图卷积网络 (Spatio-Temporal Graph Convolutional Networks, STGCN) 最早是用于交通网

络的流量预测,后来也延伸至其他领域<sup>[6]</sup>。如 Stańczyk 等<sup>[7]</sup>使用 STGCN 对丹麦和荷兰两个地区的多个观测点位的风速数据进行预测,取得了比传统卷积网络更高的准确度;而 Zhang 等<sup>[8]</sup>使用的 STGCN 模型预测河流水质中溶解氧、水温等多个因子的数据。本文提出的新的深度学习模型 AirNet6,不只是利用了测点的历史空气质量数据以及其空间关联,还加入了对应的天气预测数据、社会活动数据等影响空气质量的因子,能快速预测出 168 h 内的空气质量数据。

## 1 模型设计

### 1.1 模型输入

传统的化学模型需要空气质量数据、气象数据和污染排放数据作为运算的基础,例如 CMAQ (Community Multiscale Air Quality) 模型的输入数据就包括各监测因子的实测数据、气象预报数据和污染源排放清单<sup>[9]</sup>。类似的,AirNet6 模型的输入数据也包括历史的监测数据及天气预测数据和排放数据,但与 CMAQ 输入的二维平面数据不同,AirNet6 输入的是多位的点数据;另外,排放数据在 AirNet6 中用社会活动数据统计数据进行替代。模型数据流图如图 1 所示。模型输入的数据是一个 3 维张量 (tensor),其中第一维是时间,粒度为一小时,例如以一年的数据作为输入,时间维度的长度为  $365 \times 24 = 8760$ ;第二维是点位,例如广东的国控空气监测的点位有 125 个,点位维度的长度就为 125;第三维是通道,对应不同的因子,第一个通道是监测数据,例如 ozone 的监测数据,第二个通道开始的 7 个通道是天气预测的数据,分别是气压、光照、降水、温度、湿度、风速 u、风速 v,最后一个通道是社会活动数据。

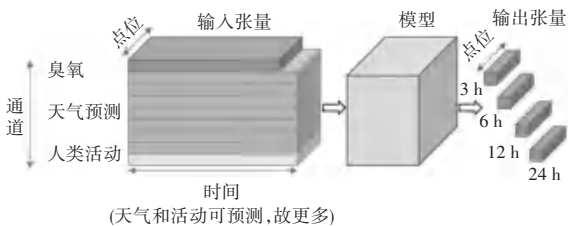


图 1 模型数据流图

Fig. 1 Model data flow diagram

天气预测数据与社会活动数据是可预测的,比历史监测数据多一些,例如:当需要预测 7 天后空气质量数据时,模型可以使用已经获取到的 7 天后的天气预测数据以及社会活动数据,而不是只使用当前的历史数据,这样预测的结果更为准确。

天气预测有多种参量数据,其中地面风场、大气

层结及天气形势等主要气象参量与环境空气质量的相关关系较强<sup>[10]</sup>。因此本模型选用气压、光照、降水、温度、湿度、u 风速、v 风速作为模型输入数据。另外,空气质量数据是基于监测点位的数据,而天气预测数据是平面网络数据,例如本模型使用的 NOAA (National Oceanic and Atmospheric Administration) 的 GFS (Global Forecast System) 数据,是沿经纬度每 0.25 度一个数据。为了将这两种数据融入到一个模型中,需要把其空间域统一。因此,本文使用监测点位作为统一的空间域,数据量可以大为减少,每个监测点位的天气预测数据,根据其位置通过插值得出。

排放清单数据反映了每种污染物排放的时空分布,是空气质量预报重要的输入数据<sup>[11]</sup>。经研究测试,相关的社会活动数据如污染源排放、经济活动、交通流量等微观数据很难准确、完整地收集,且这些数据的时间与空间域跟监测数据相差很远,难以像气象数据那样整合。而在对空气质量产生影响的社会经济因素中,社会活动强度是较为明显的,可以用反映社会活动强度的数据作为排放清单的替代<sup>[12]</sup>。例如,电网负荷指标就可以直接反映社会活动强度变化,而对用电负荷特性影响很大的因素是重大社会活动和节假日<sup>[13]</sup>。因此,本模型使用一个宏观的统计值作为活动强度的可预测的指标数据。由于社会活动的频繁程度主要与“工作日(w)”及“休息日(r)”交替相关,通过对“前一日”、“当日”、“后一日”不同情况的枚举,得出 www、wwr、wrw、wrr、rww、rwr、rrw、rrr 这 8 种排列,通过对历史数据中这 8 种情况的监测数据统计,就可以得出每个点位在这 8 种情况下 24 小时中的平均值,作为代表其社会活动频繁程度的经验值,如图 2 所示。

### 1.2 网络结构

模型的网络结构如图 3 所示,模型的网络结构由三层时空卷积块和一层连接层组成。三层时空卷积层中,前面两层的时空卷积层只有一个时空卷积块,后面第三层的时空卷积层与连接层对接,有 4 个时空卷积块分别对第四层连接层的 4 个连接块。时空卷积块是模型的核心,由时间层和空间层组成。时间层用于捕捉数据组的时序变化特征,与 LSTM 的作用类似,但由于其基于卷积网络而非 LSTM 的循环网络,可以更好的使用并发计算机制,运算的速度更快<sup>[14]</sup>;Kipf<sup>[15]</sup>提出空间层,建立多个点位间关联图谱,实现图卷积网络,点位数据之间的空间关联特征。

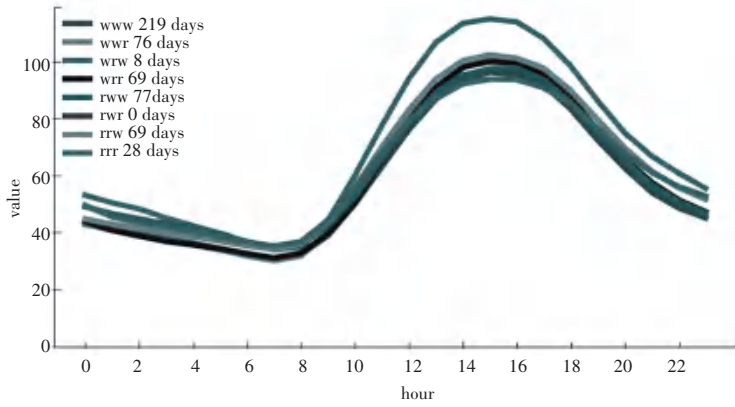


图2 在8类工作日情况下的监测数据均值

Fig. 2 Mean monitoring data under 8 kinds of working day

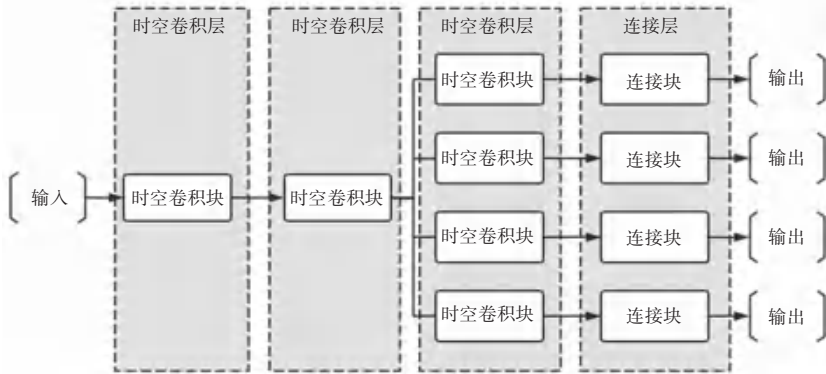


图3 模型的网络结构图

Fig. 3 Network structure diagram of the model

### 1.3 模型输出

最初的模型输出的数据是一个1维张量,第一维是点位,数据代表这些点位在 $n$ 个小时后的时间点的空气质量,这就需要为每一个时间点训练一个模型,如果需要预测7天的空气质量数据,就要训练 $7 \times 24 = 168$ 个模型,而6个监测因子,就要训练 $168 \times 6 = 1\ 008$ 个模型。

使用时间移位的方法,每个因子只需要训练8个模型。每个模型只预测 $n = [3, 6, 12, 24, 48, 72, 120, 168]$ 8个时间点的空气数据,只需要 $8 \times 6 = 48$ 个模型。而这些时间点之间的数据,使用最靠近的

后面的时间点模型计算得出,但输入数据需要相应前移获取。例如,时间点10没有对应的模型,就是用后面最靠近的12的模型,但输入数据需要前移2小时来取。

模型数据输入输出映射图如图4所示。假设时间点12的模型需要输入24个小时的数据,当前时刻为 $t$ , $t$ 以及 $t$ 以前的数据都已就绪;而预测 $t + 7 \sim t + 12$ 的数据,都使用时间点12的模型计算,其中预测 $t + 12$ 时,输入 $t - 23 \sim t$ 的数据;预测 $t + 10$ 时,输入 $t - 25 \sim t - 2$ 的数据。

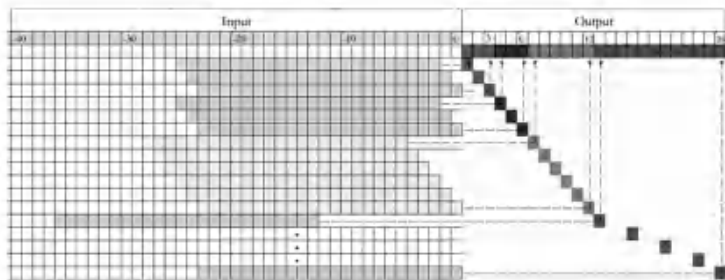


图4 模型数据输入输出映射图

Fig. 4 Model data input output mapping diagram

同理,时间点  $n$  的模型可覆盖的预测时间,见表1。

表1 时间点覆盖表  
Table 1 Time point coverage

时间点 $n$ 模型	覆盖预测时间/h
3	1~3
6	4~6
12	7~12
24	13~24
48	25~48
72	49~72
120	73~120
168	121~168

经进一步的研究迭代,将模型结构改为多输出,每个因子只需要训练3个模型,其中一个用于同时预测  $n = [3, 6, 12, 24]$  4个时间点后的空气数据,一个同时预测  $n = [48, 72]$  两个时间点,还有一个同时预测  $n = [120, 168]$  两个时间点,这样就只需要  $3 \times 6 = 18$  个模型。经测试对比,将时间点作4-2-2的分割,可以减少模型和训练模型的时间,而且准确率几乎没有下降。但若作4-4分割,后面4个时间点(48, 72, 120, 168)的准确率有些下降,如果是用1个模型直接输出8个时间点,准确率显著下降,因此4-2-2的分割是相对更好的选择。

最后,迭代后模型的输出为一个2维张量,第一维是时间点,长度为2或者4,代表2或者4个时间点;第二维度是点位。

## 2 实验

### 2.1 数据集

本文使用广东省125个国控点的空气质量小时数据作为训练和验证模型的数据,包括臭氧、二氧化氮、一氧化碳、二氧化硫、PM10、PM2.5六个监测因子,点位的经纬度坐标用于与天气预测数据关联插值。

天气预测的数据,本文使用NOMADS(NOAA Operational Model Archive and Distribution System)的GFS0.25数据,空间范围是覆盖广东省的经纬度矩形("north": 26, "west": 109, "south": 20, "east": 118)。近8天的GFS数据可以通过NOAA官方的NOMADS渠道获取,但更久远的历史数据需通过大气科研组织UCAR(University Corporation for Atmospheric Research)取得。这两个平台均支持WCS(Web Coverage Service)和OPeNDAP(Open-

source Project for a Network Data Access Protocol)等通用数据接口,OPeNDAP更适合长期使用需求且性能更好<sup>[16]</sup>。因此为了适配常规化运行的场景,本模型的数据源使用OPeNDAP接口。

社会活动数据使用“工作日-休息日”宏观经验值,加上空气质量历史统计数据组成。

数据集的范围为2021年4月1日~2022年10月1日,其中2021年4月1日~2022年7月1日的数据用于训练,2022年7月1日~2022年10月1日的数据用于测试。即5个季度的数据用于训练,1个季度的数据用于测试。

### 2.2 数据预处理

算法和数据是人工智能模型的核心,高质量的数据对提高预测的准确率至关重要,因此在模型输入数据之前,要对数据做清洗、补足等预处理操作<sup>[17]</sup>。由于国控点的空气质量监测数据由自动监测设备产生,通常存在缺数据的现象<sup>[18]</sup>;而模型不能输入空数据,所以需要先把缺失的数据补足。补足的方法是先用空间补足,即使用同一时间周边有数据的点位的数据,通过效果相对较好的克里金插值方法得出缺数据点位的数值<sup>[19]</sup>。如果空间补足后仍有缺数据(这种情况一般为网络或者中心服务故障,导致同一时间大部分点位都没有数据),此时使用时间补足,用这个点位前一天的数据补足缺失的数据。

天气预测数据缺失不常见,而且因为天气预测数据冗余非常多,只需要把前一天和后一天相应时间的数据赋值过来即可。

社会活动的宏观指标数据没有缺失,不需要补足。

所有的输入数据在进入模型之前,必须规格化。规格化就是把所有数值按比例压缩到 $[0, 1]$ 的范围,以利于神经网络计算。规格化的上下限值一般设置为样本数据(供训练的历史数据)的最大最小值,但为了训练出的模型能够适应样本以外的数据,本文将上下限值扩展一倍。设 $S_{\max}$ 、 $S_{\min}$ 、 $S_{\text{avg}}$ 分别是所有样本数据的最大值、最小值和平均值,规格化的上限和下限如公式(1)和公式(2):

$$\text{Normalized}_{\max} = 2 \times S_{\max} - S_{\text{avg}} \quad (1)$$

$$\text{Normalized}_{\min} = 2 \times S_{\min} - S_{\text{avg}} \quad (2)$$

### 2.3 实验设置

本文使用3台PC进行训练及验证,每天PC配备一个nvidia 3080或3090GPU以提供算力,在训练时功耗约为300W,每个模型的训练时间大概需要

2 h。模型程序运行在 ubuntu20.04 和 python3.9 环境中, AI 框架使用 pytorch1.12.0, CUDA 版本为 11.7。为了加快计算速度,减少 GPU 内存使用,模型全程使用 float32 类型进行运算。

## 2.4 实验结果

根据《环境空气质量数值预报技术规范(HJ 1130-2020)》中的“6.2 单项污染物浓度预报统计评估”,计算 4 个模型在预测 2022 年 7 月 1 日~2022

年 10 月 1 日各因子预报结果与实况的偏差、误差和相关性。

标准化平均偏差结果比较见表 2, AirNet6 模型在 6 个因子的预测结果的平均偏差绝对值中,  $O_3$ 、 $NO_2$ 、CO、PM10 等 4 个因子是最小的,而 PM2.5 和  $SO_2$  分别排在 2、3 位。另外,由于化学模型偏差较大,对其数据做初步的修偏处理,得到“化学模型(修偏)”数据,用于后续的结果对比。

表 2 标准化平均偏差结果比较

因子	模型	Day1	Day2	Day3	Day4	Day5	Day6	Day7	平均
$O_3$	AirNet6	-6	-11	-9	-13	-13	-13	-13	-11
	Prophet	-18	-18	-17	-17	-18	-18	-19	-18
	LSTM	-22	-24	-23	-23	-23	-23	-24	-23
	化学模型	50	57	65	71	74	73	66	65
	化学模型(修偏)	-9	-5	0	4	5	5	1	0
$NO_2$	AirNet6	0	1	3	0	-1	2	2	1
	Prophet	-14	-14	-15	-15	-16	-16	-16	-15
	LSTM	22	21	21	21	20	20	20	21
	化学模型	49	55	55	53	55	50	48	52
	化学模型(修偏)	-2	2	2	1	2	-1	-3	0
CO	AirNet6	-6	-8	-6	-4	-4	-2	-2	-4
	Prophet	-8	-8	-8	-8	-8	-8	-8	-8
	LSTM	35	37	36	36	36	35	35	36
	化学模型	-47	-47	-47	-46	-45	-45	-47	-46
	化学模型(修偏)	-1	-2	-1	1	3	1	-1	0
$SO_2$	AirNet6	-9	-13	-12	-13	-13	-12	-13	-12
	Prophet	-9	-10	-10	-10	-10	-10	-10	-10
	LSTM	9	10	9	9	9	9	9	9
	化学模型	7	10	14	17	18	13	9	13
	化学模型(修偏)	-5	-2	2	4	5	0	-2	0
PM10	AirNet6	-4	-5	-8	-7	-7	-9	-9	-7
	Prophet	-23	-23	-23	-23	-24	-24	-24	-23
	LSTM	-13	-13	-13	-13	-13	-12	-13	-13
	化学模型	-43	-43	-42	-40	-41	-42	-46	-42
	化学模型(修偏)	-2	-1	-1	4	2	-1	-7	-1
PM2.5	AirNet6	-10	-23	-13	-18	-17	-17	-17	-16
	Prophet	-29	-29	-29	-29	-30	-30	-30	-30
	LSTM	-13	-13	-13	-13	-14	-15	-16	-14
	化学模型	-29	-29	-29	-24	-25	-27	-33	-28
	化学模型(修偏)	-1	-2	-1	6	3	1	-7	0

均方根误差结果比较见表 3, AirNet6 模型在 6 个因子的预测结果的均方根误差中, 6 个因子均是

表 3 均方根误差结果比较  
Table 3 Comparison of root mean square error results

因子	模型	Day1	Day2	Day3	Day4	Day5	Day6	Day7	平均
O <sub>3</sub>	AirNet6	33.55	36.37	35.32	38.65	38.78	38.96	39.01	37.23
	Prophet	42.80	44.57	45.59	46.50	47.22	47.74	48.19	46.09
	LSTM	51.79	53.06	52.97	52.71	52.91	52.97	53.04	52.78
	化学模型	55.19	56.14	60.57	63.84	65.01	64.36	63.73	61.26
	化学模型(修偏)	43.78	42.79	43.63	43.91	43.35	43.45	45.31	43.75
NO <sub>2</sub>	AirNet6	9.45	9.76	10.02	9.94	9.95	10.22	10.21	9.94
	Prophet	10.66	10.79	10.86	10.92	10.98	11.06	11.13	10.91
	LSTM	18.90	14.93	14.94	14.92	14.91	14.91	14.93	15.49
	化学模型	24.48	25.10	25.89	25.44	25.64	24.62	24.79	25.14
	化学模型(修偏)	16.44	16.73	17.28	17.05	17.09	16.55	16.79	16.85
CO	AirNet6	0.37	0.40	0.40	0.38	0.37	0.37	0.37	0.38
	Prophet	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37
	LSTM	0.49	0.50	0.49	0.49	0.49	0.49	0.49	0.49
	化学模型	0.47	0.48	0.48	0.48	0.47	0.45	0.47	0.47
	化学模型(修偏)	0.50	0.50	0.51	0.51	0.51	0.48	0.49	0.50
SO <sub>2</sub>	AirNet6	14.71	14.98	14.92	14.99	14.90	14.71	14.69	14.84
	Prophet	14.71	14.95	14.91	14.91	14.82	14.71	14.70	14.82
	LSTM	15.37	15.31	15.26	15.25	15.15	15.03	15.00	15.20
	化学模型	16.18	16.00	16.11	16.41	16.19	15.92	15.84	16.09
	化学模型(修偏)	15.80	15.60	15.68	15.97	15.73	15.49	15.43	15.67
PM10	AirNet6	15.32	16.72	16.70	18.38	18.43	17.58	17.60	17.25
	Prophet	19.87	20.70	21.17	21.58	21.88	22.15	22.46	21.40
	LSTM	21.96	22.26	22.41	22.44	22.60	22.73	22.75	22.45
	化学模型	22.87	22.80	23.72	23.60	23.06	23.39	24.32	23.39
	化学模型(修偏)	25.85	25.70	27.15	27.80	26.54	26.52	26.17	26.53
PM2.5	AirNet6	10.07	11.41	10.31	12.00	12.04	11.66	11.66	11.31
	Prophet	12.82	13.48	13.80	14.09	14.31	14.51	14.75	13.97
	LSTM	14.23	14.32	14.30	14.39	14.43	14.40	14.32	14.34
	化学模型	14.01	13.99	14.86	15.00	14.29	14.47	14.59	14.46
	化学模型(修偏)	16.44	16.31	17.32	18.01	16.87	17.00	16.43	16.91

多种模型 1~7 天预测结果的相关系数见表 4, AirNet6 模型在 6 个因子的预测结果的相关系数中, O<sub>3</sub>、NO<sub>2</sub>、PM10、PM2.5 等 4 个因子是最小的, 而 CO 和 SO<sub>2</sub> 排在 2 位。

根据实验的结果可知, AirNet6 在预测 6 个监测因子中, O<sub>3</sub>、NO<sub>2</sub>、PM10、PM2.5 的准确率都明显优

于其他模型, 而 CO 和 SO<sub>2</sub> 这两个因子的准确率与第一名 Prophet 很接近且明显优于其他模型。

### 3 常规化运行

模型训练完后通常可使用一个季度到一年, 因为模型只是捕捉了训练数据(本文使用的是 2021 年

4月1日~2022年7月1日的数据)的规律,当内、外环境因素有变化后,模型的预测准确率会下降,届时需要重新使用训练数据重新训练模型,以获得更匹配的预测结果。

表4 多种模型1~7天预测结果的相关系数

Table 4 Correlation coefficient results of 1 to 7 days prediction with different models

因子	模型	Day1	Day2	Day3	Day4	Day5	Day6	Day7	平均
O <sub>3</sub>	AirNet6	0.79	0.76	0.77	0.73	0.73	0.72	0.72	0.74
	Prophet	0.66	0.61	0.59	0.57	0.55	0.54	0.53	0.58
	LSTM	0.42	0.40	0.40	0.41	0.41	0.41	0.41	0.41
	化学模型	0.44	0.45	0.43	0.42	0.45	0.48	0.45	0.45
	化学模型(修偏)	0.44	0.45	0.43	0.42	0.45	0.48	0.45	0.45
NO <sub>2</sub>	AirNet6	0.70	0.67	0.67	0.67	0.67	0.65	0.65	0.67
	Prophet	0.61	0.60	0.59	0.59	0.59	0.58	0.57	0.59
	LSTM	0.21	0.28	0.27	0.27	0.27	0.27	0.27	0.26
	化学模型	0.34	0.32	0.30	0.30	0.30	0.32	0.32	0.32
	化学模型(修偏)	0.34	0.32	0.30	0.30	0.30	0.32	0.32	0.32
CO	AirNet6	0.26	0.15	0.15	0.13	0.13	0.13	0.13	0.16
	Prophet	0.29	0.27	0.26	0.25	0.24	0.23	0.22	0.25
	LSTM	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	化学模型	0.09	0.07	0.05	0.05	0.06	0.08	0.08	0.07
	化学模型(修偏)	0.09	0.07	0.05	0.05	0.06	0.08	0.08	0.07
SO <sub>2</sub>	AirNet6	0.18	0.16	0.16	0.13	0.13	0.16	0.16	0.16
	Prophet	0.18	0.17	0.17	0.16	0.16	0.16	0.16	0.17
	LSTM	0.04	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	化学模型	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	化学模型(修偏)	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
PM10	AirNet6	0.70	0.63	0.63	0.53	0.53	0.58	0.58	0.59
	Prophet	0.50	0.43	0.39	0.35	0.33	0.31	0.29	0.37
	LSTM	0.16	0.13	0.10	0.09	0.07	0.05	0.04	0.09
	化学模型	0.21	0.17	0.08	0.10	0.12	0.15	0.16	0.14
	化学模型(修偏)	0.21	0.17	0.08	0.10	0.12	0.15	0.16	0.14
PM2.5	AirNet6	0.68	0.61	0.66	0.51	0.50	0.54	0.54	0.58
	Prophet	0.49	0.40	0.36	0.32	0.29	0.28	0.25	0.34
	LSTM	0.08	0.06	0.06	0.04	0.03	0.04	0.05	0.05
	化学模型	0.20	0.18	0.08	0.11	0.13	0.16	0.19	0.15
	化学模型(修偏)	0.20	0.18	0.08	0.11	0.13	0.16	0.19	0.15

模型在常规化运行时,不需要训练,只需要使用训练好的模型进行推演,因此消耗的资源非常小,一台普通配置的PC(CPU AMD Ryzen 7 1700, GPU Nvidia 970),在2 min内就可完成6个因子168 h预测,可以在生产中实现密度为每小时一次的预测,有助于重污染天气的应对。

#### 4 资源消耗比较

人工智能模型与传统的化学模型的架构和运行机制相差很大,因此在研发和实际运行时所需的资源也各不相同,人工智能模型需要更多的运算资源用于前期训练,但训练完成后在常规运行阶段只需要极少的资源用于推理运算,而化学模型在运行阶

段需要大量服务器计算资源,所以总体来说人工智能模型会比传统的化学模型节省大量资源<sup>[20]</sup>。不同模型的资源消耗见表 5。

表 5 资源消耗比较

Table 5 Comparison of Resource consumption

模型	训练耗时/d	训练硬件需求	常规运行耗时/min	常规运行硬件需求	源数据需求
AirNet6	2	3080 GPU	2	普通 PC	历史数据、天气预测
Prophet			360	普通 PC	历史数据
LSTM	1	3 080 GPU	1	普通 PC	历史数据
化学模型			360	服务器集群	历史数据、天气预测、源清单

## 5 结束语

本文提出了一种新的深度学习模型 AirNet6,可以捕捉一片区域的空气质量随时间变化及随天气和社会活动的规律,快速而准确的预测未来的空气质量数据。实验表明,在真实的广东省近年的国控数据集上,新的模型优于其他模型,这对于实施更加精细的环境管理是十分有用的。

未来,会继续优化迭代模型,提高模型的准确性和适用范围,如河流水质监测、噪声监测等。

## 参考文献

[1] 薛文博,王金南,杨金田,等. 国内外空气质量模型研究进展[J]. 环境与可持续发展,2023,38(3):14-20.

[2] 王扬锋,马雁军,权维俊,等. 东北区域空气质量数值预报模式应用研究进展[J]. 气象与环境学报,2020,36(6):130-136.

[3] 王子钊. 冬季突发性 PM<sub>2.5</sub> 重污染天气预警技术研究[J]. 环境科学与管理,2022,47(10):134-138.

[4] 花凡,李莉,蔡鑫楠,等. 长短期记忆网络在气温预测中的应用[J]. 智能计算机与应用,2022,12(11):92-102.

[5] 邓丽,邬群勇,杨水荣. 融合 SSAE 深度特征学习和 LSTM 网络的 PM<sub>2.5</sub> 小时浓度预测[J]. 环境科学学报,2020,40(9):3422-3434.

[6] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[J]. arXiv preprint arXiv:1709.04875, 2017.

[7] STAŃCZYK T, MEHRKANOON S. Deep graph convolutional networks for wind speed prediction[J]. arXiv preprint arXiv:2101.10041, 2021.

[8] ZHANG Y F, PETER J, THORBLIR N, et al. Multi-task temporal convolutional network for predicting water quality sensor data [C]// Proceedings of International Conference on Neural Information. Cham: Springer, 2019. DOI: 10.1007/978-3-030-36808-1\_14

[9] CHEN X, ZHANG Y, WANG K, et al. Evaluation of the offline-coupled GFSv15-FV3-CMAQv5.0.2 in support of the next-generation National Air Quality Forecast Capability over the contiguous United States [J]. Geoscientific Model Development, 2021, 14(6): 3969-3993.

[10] 张裕芬,朱坦,冯银厂,等. 地面风场、大气层结及天气形势等主要污染气象参量的特征及其与环境空气质量的相关关系[J]. 城市环境与城市生态,2006,19(4):33-36.

[11] 程兴宏,徐祥德,安兴琴,等. 2013 年 1 月华北地区重霾污染过程 SO<sub>2</sub> 和 NO<sub>x</sub> 的 CMAQ 源同化模拟研究[J]. 环境科学学报,2016,36(2):638-648.

[12] 张晓平,林美含. 中国城市空气污染区域差异及社会经济影响因素分析——基于两种空气质量指数的比较研究[J]. 中国科学院大学学报,2020,37(1):39-50.

[13] 龚爱国,李燕波,邓晨. 张家界地区节假日负荷特性分析[J]. 低碳世界,2016(26):60-61.

[14] GEHRING J, AULI M, GRANGIER D, et al. Convolutional sequence to sequence learning [C]//Proceedings of the 34<sup>th</sup> International Conference on Machine Learning. IEEE, 2017: 1243-1252.

[15] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks [J]. arXiv preprint arXiv:1609.02907, 2016.

[16] BAART F, DE BOER G, DE HAAS W, et al. A comparison between WCS and OPeNDAP for making model results and data products available through the internet [J]. Transactions in GIS, 2012, 16(2): 249-265.

[17] 李细生,陈媛,罗慧妮,等. 湖南空气质量预报中的数据预处理和特征工程[J]. 中国环境监测,2023,39(4):185-195.

[18] 王莉华,安欣欣,景宽,等. 大气网格化监测运行维护管理现状与展望[J]. 中国环境监测,2021,3(2):16-22.

[19] 靳国栋,刘衍聪,牛文杰. 距离加权反比插值法和克里金插值法的比较[J]. 长春工业大学学报,2003,3:53-57.

[20] 王茜,黄蕊珠,肖宇. 长三角区域空气质量数值预报系统及其在重大活动保障中的应用[J]. 中国环境监测,2020,36(2):15-21.