

面向深度学习的空气质量预测研究进展^①



赵小明^{1,2}, 顾珂铭¹, 张石清²

¹(浙江理工大学 信息学院, 杭州 310018)

²(台州学院 智能信息处理研究所, 台州 317000)

通信作者: 赵小明, E-mail: tzxyzxm@163.com

摘要: 空气污染是影响公共卫生的重要因素, 空气质量预测是空气污染预警的关键, 是近年来环境学、统计学、计算机科学等领域中的热点研究课题. 本文综述了空气质量预测方法的研究现状与进展, 尤其对近年来新发展起来的深度学习方法在空气质量预测方面的应用进行了系统分析与总结. 首先, 介绍了空气质量预测方法的演变历程和空气污染数据集. 然后, 阐述了传统空气质量预测方法. 随后, 从时间信息、时空信息、注意力机制等角度出发, 重点分析和比较了现有面向深度学习的空气质量预测方法的进展. 最后, 对空气质量预测方法的未来发展趋势进行了总结与展望.

关键词: 空气污染; 空气质量预测; 深度学习; 注意力机制

引用格式: 赵小明, 顾珂铭, 张石清. 面向深度学习的空气质量预测研究进展. 计算机系统应用, 2022, 31(11): 49-59. <http://www.c-s-a.org.cn/1003-3254/8847.html>

Research Progress of Air Quality Prediction Based on Deep Learning

ZHAO Xiao-Ming^{1,2}, GU Ke-Ming¹, ZHANG Shi-Qing²

¹(School of Informatics Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

²(Institute of Intelligent Information Processing, Taizhou University, Taizhou 317000, China)

Abstract: Air pollution is an important factor affecting public health, and air quality prediction is the key to air pollution early warning and a hot research topic in the fields of environmental science, statistics, and computer science in recent years. This study reviews the research status and progress of air quality prediction methods, with a special focus on the systematical analysis and summarization of the applications of the newly-emerged deep learning methods in recent years in air quality prediction. Specifically, the evolution process of air quality prediction methods and air pollution datasets are outlined. After the traditional air quality prediction methods are described, the progress of existing deep learning-based air quality prediction methods is analyzed and compared in detail from the perspectives of temporal information, temporal-spatial information, and attention mechanisms. Finally, the development trend of air quality prediction methods is summarized and predicted.

Key words: air pollution; air quality prediction; deep learning; attention mechanism

空气质量对人们的生活具有至关重要的影响, 随着社会的科技化与商业化, 空气质量难以避免地产生了一些不可预料的发展趋势, 这引起了人们与相关部门的关注. 越来越多的国家开始建设大量的空气质量监测

站. 根据 2017 年 10 月 27 日世界卫生组织国际癌症研究机构 (international agency for research on cancer, IARC) 公布的致癌物清单, 空气污染物属于致癌物的一类, 因此预测空气质量的走势已经成为现今科学研究的热点

① 基金项目: 国家自然科学基金 (61976149); 浙江省自然科学基金 (LZ20F020002)

收稿时间: 2022-03-07; 修改时间: 2022-04-12; 采用时间: 2022-05-16; csa 在线出版时间: 2022-07-29

问题。

2021年,习近平总书记在主持召开中央财经委员会第九次会议强调,要把“碳达峰、碳中和”纳入生态文明建设整体布局.减少碳排放、改善空气质量,应对气候变化成为普遍共识.大气污染防治背景下,中国气象局印发《“十四五”公共气象服务发展规划》,强调提升气象服务数字化、智能化水平的重要意义,同时提出加强气象服务核心技术研发,深化信息技术融合应用的要求^[1].

空气质量预测是气象服务中的重要一环,在防污攻坚战中对空气质量持续进行了“健康管理”,对研判污染物排放、空气质量变化趋势具有重要意义.提升空气质量精准预测,有利于早日实现清洁空气及碳中和的目标,因此亟需发挥系统模型在空气质量预测技术中的支撑作用.

如何准确地进行空气质量预测十分具有挑战性.空气质量预测来源广泛,可以来源于地面监测站,气象卫星,监控飞机等,这些数据是空间与时间上高度不均匀的污染物采样,很难将它们汇总去对污染物变化进行预测.因为地面空气质量监测站由于接近污染物排放地点,是研究空气质量变化的重要数据来源,但仍存在许多问题.首先,尽管现在部分地区不仅设立了许多区域级空气质量监测站,还在街道上安装了相互连接的微型空气质量监测站^[2],但仍有许多地区没有建立完善的空气质量监测系统,导致数据样本的稀少与稀疏.其次,空气污染物的产生具有大量的化学过程,并具有相当的时间与空间性,难以通过固定的公式对空气质量进行预测.

针对上述问题,近年来,常用的空气质量预测方法可分为两大类:传统预测方法和机器学习算法.其中,传统预测方法主要包括数值模型预测方法^[3]和统计方法^[4].数值模型方法基于大气动力学与化学过程,通过气象方法使用方程组构建数学模型模拟水平或垂直方向的污染数据.统计方法较少关注于空气污染物的物理化学属性,而是通过前一个步长的气象数据和空气质量数据直接结合,通过使用基于统计的模型来预测空气质量.这些方法可以有效地结合多领域知识,但它们都有一定的局限性,如在处理大量的数据时缺乏灵活的多尺度框架,并且大量的运算可能会导致计算误差的出现,因此研究人员开始使用非线性机器学习方法,如支持向量回归(support vector regression, SVR)^[5],

随机森林回归(random forest regression, RFR)^[6]等.虽然非线性机器学习方法在空气质量预测中取得了不错的性能,但大部分的机器学习方法很难捕捉空气质量的时间序列模式,并且不能做到在空气污染的长期影响中学习规律,因此影响了一定的预测精度.

近年来,深度学习^[7]逐渐走进了人们的视野,深度学习可以学习样本的内在规律,可以从中提取出高层次的特征信息.代表性的深度学习模型主要包括深度信念网络(deep belief network, DBN)^[8]、卷积神经网络(convolutional neural network, CNN)^[9]和循环神经网络(recurrent neural network, RNN)^[10]模型等.目前,深度学习方法在计算机视觉^[11]、语音识别^[12]、数据挖掘^[13]等领域取得了很大成功.近年来,深度学习方法也开始应用于空气污染预测^[14],大大提高了空气质量预测的效率与精准度.从2010年到2021年,空气质量预测算法经历了从传统方法到深度学习方法两个不同阶段.同时,数据集已经从单空气站点过渡到多空气站点和城市群.图1展示了空气质量预测方法及数据集的演变历程.本文从国内外空气污染数据集入手,先介绍传统空气质量预测方法,然后对基于深度学习的空气质量预测研究进行系统总结与归纳,包括代表性的深度学习技术回顾、基于时空要素和注意力机制的空气质量预测方法等方面,详述国内外研究进展状况,指出面临的挑战及未来的发展方向.

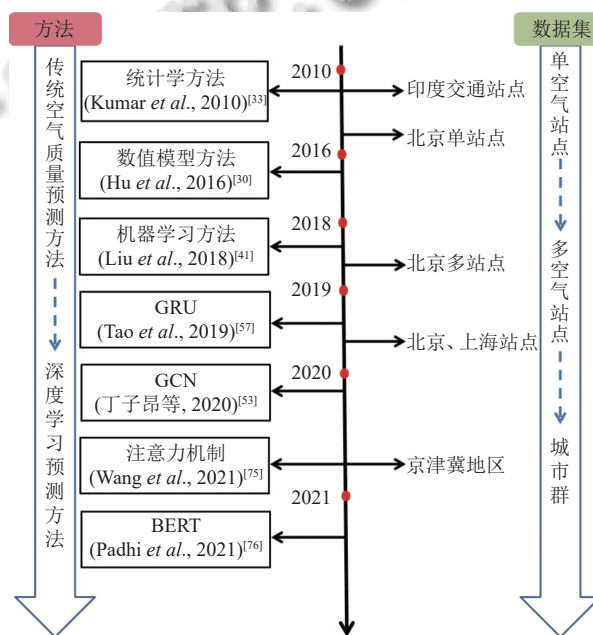


图1 空气质量预测方法与数据集的演变历程

1 空气质量数据集

开展空气质量预测研究,一个完善综合的数据库是必不可少的.近年来,研究者已经相继建立了一些空气质量数据库,其中代表性的数据库有10个,包括城市单站点,多站点,以及城市群空气质量数据集,如表1

表1 代表性的空气质量数据集

研究地点	时间	来源	简要介绍
北京单站点 ^[15]	2010/01/01–2014/12/31	美国驻北京大使馆	包含北京单站点小时级PM _{2.5} 浓度数据及温度、气压等气象数据
北京多站点 ^[16]	2015/01/01–2016/12/31	中国国家环境监测中心	包含北京12个站点小时级PM _{2.5} 、PM ₁₀ 等空气质量数据
北京多站点 ^[17]	2017/01/01–2018/05/01	北京空气质量监测站	包含北京35个站点小时级PM _{2.5} 、PM ₁₀ 等空气质量数据
上海多站点 ^[18]	2015/03/01–2017/02/28	上海环境监测中心	包含上海2个站点小时级PM _{2.5} 、PM ₁₀ 等空气质量数据
淮海经济区 ^[19]	2017/03/01–2019/02/28	中国国家生态环境部	包含淮海经济区10个城市小时级PM _{2.5} 、PM ₁₀ 等空气质量数据
兰州多站点 ^[20]	2017/11/01–2019/01/01	兰州空气微监测站	包含兰州359个站点小时级PM _{2.5} 、PM ₁₀ 等空气质量数据
石家庄单站点 ^[21]	2019/01/01–2019/12/31	石家庄空气监测站	包含石家庄单站点小时级PM _{2.5} 、PM ₁₀ 等空气质量数据
美国各州 ^[22]	2000/01/01–2016/12/31	美国环境保护署	包含美国各州PM _{2.5} 、PM ₁₀ 等空气质量数据
德里多站点 ^[23]	2018/12/01–2019/11/30	印度中央控制污染委员会	包含印度首都德里PM _{2.5} 、PM ₁₀ 等空气质量数据及温度、气压等气象数据

2 基于传统方法的空气质量预测

早期空气质量预测方法主要包括数值模型方法和统计方法以及以SVR为代表的传统机器学习方法.数值模型方法接近传统的天气预报方式,以大气物理,大气化学为基础,将时间、空间维度划分为网格通过气象数据与大气污染数据通过计算机模拟,常见的数值模型方法有多尺度空气质量模型 (community multiscale air quality modeling system, CMAQ)^[24]、扩展综合空气质量模型 (comprehensive air quality model with extensions, CAMx)^[25]、嵌套网格空气质量预报模式系统 (nested air quality prediction modeling system, NAQPMS)^[26]等.统计方法往往基于线性依赖、平稳性等假设,通过曲线拟合和参数估计来建立数学模型.最常见的统计模型有自回归滑动平均模型 (autoregressive moving average model, ARMA)^[27]、自回归整合滑动平均模型 (autoregressive integrated moving average model, ARIMA)^[28]、多元线性回归模型 (multivariable linear regression model, MLR)^[29]、时间序列回归模型等.经典的机器学习算法通过设置一定的参数,消除了上述假设引起的误差,是目前使用较为广泛的空气质量预测方法之一.常用的机器学习算法有SVR、RFR等.

Hu等^[30]基于天气研究和预报模型 (weather research and forecasting, WRF)^[31]以及CMAQ模型对中国空气质量进行了一年(2013)的模拟.该研究表明CMAQ模型可以成功重现大多数城市的PM_{2.5}以及

所示.现有的数据集大多数来源于区域的空气质量监测站,一般包含细颗粒物 (PM_{2.5})、一氧化碳 (CO)、二氧化硫 (SO₂)、二氧化氮 (NO₂)、臭氧 (O₃)等大气污染物及大气气温、大气气压、相对湿度、降水量等气象数据,这些数据一般以小时级存储.

O₃浓度,而且可以重现中国的空气污染并确定改进的方向.谢磊等^[32]将多个空气质量数值预测模型 (CMAQ、CAMx、WRF-CHEM)的预测数据作为实验样本,基于最优定权组合法构建了对SO₂的预测模型并与多元线性回归法和动态权重更新法进行比较,实验结果表明最优定权组合法可以有效提高SO₂预测模型的预测精度.

Kumar等^[33]使用ARMA、ARIMA等模型预测印度德里一个交通站点的空气污染数据.该方法通过不同信息准则的组合如贝叶斯信息准则 (Bayesian information criterion, BIC)^[34]等来确定模型的自回归参数 (p)及滑动平均参数 (q),最终取得了令人满意的预测结果.Cheng等^[35]提出一种基于滑动平均 (moving average, MA)^[36]和有序加权平均 (order weight average, OWA)^[37]的MA-OWA模型.该模型通过OWA算子的加权聚合及MA的简化计算高效准确地计算台湾城市空气站点的臭氧浓度,可以获得比MA、ARMA等模型更好的预测结果.

Lei等^[38]提出了采用分类回归树 (classification and regression trees, CART)^[39]和多元回归 (multiple regression, MR)^[40]来建立多污染物统计模型,在对澳门空气质量的预测中获得了较高的置信水平.Liu等^[41]比较了多线性回归 (MLR)、随机森林回归 (RFR)、支持向量回归 (SVR)等机器学习算法预测空气质量指数 (air quality index, AQI)的性能.实验结果表明,

MLR 性能较为一般, 而由于随机森林变量的交互作用, RFR 在预测北京区域 AQI 具有更好的泛化能力和更高的准确性。

Nieto 等^[42]采用 SVR 来建立了西班牙城市非线性动态空气质量预测模型。该模型能非常准确地捕捉到研究地点内主要污染物之间的依赖机制, 呈现出对高度非线性模型卓越的建模能力。Kumar 等^[43]通过采用 ARIMA、主成分回归模型 (principal component regression, PCR)^[44] 以及两者的综合模型分别对印度德里地区每个季节的每日 AQI 进行预测。结果表明, ARIMA 和 PCR 的组合可以正确处理空气变量的自相关性, 而基于主成分分析 (principal component analysis, PCA)^[45] 的气象参数的引入可以大大增加模型的正确率和稳定性。

综上所述, 数值模型对历史气象数据需求较少, 但想要达到较高的准确性需要准确且完整的空气污染源数据和巨大的计算能力。这种较高的成本导致该方法的普遍性较低, 常用于硬件设备比较齐全的科研机构等场所。统计学方法具有计算成本较低, 实现较为简单的优点, 但常用的统计学模型以线性内核为主, 很难去拟合非线性问题, 而单纯的基于统计方法的空气质量预测近些年较为少见。传统的机器学习方法具有较高的准确度而且所需数据集较小, 但它们扩展性较差, 难以捕捉空气污染数据中不稳定的规律模式。这些传统预测方法虽然有其优点, 但或多或少存在一定的缺陷, 与现今大数据时代空气质量预测的要求存在一定的脱节, 因而探索一种新型的、适宜复杂特征工程的空气质量预测模型是近年来空气质量预测的发展趋势。

3 基于深度学习的空气质量预测

深度学习作为机器学习的一个子领域, 是一种功能强大且更适合大数据的机器学习方法。深度学习模型通过模仿人脑的机制对图像、声音、文本等数据进行处理, 并展现出巨大潜力。同样, 空气质量预测作为领域内的研究热点, 深度学习也已经取得不俗的成果。目前有关空气质量预测的深度模型主要有两种: 以 RNN、LSTM (long short term memory)^[46] 等时间深度网络为主要组成部分对时间信息建模的空气质量预测和通过 DNN、CNN 等网络加入空间信息的考虑时空信息的空气质量预测。除此之外, 以注意力机制为首的新兴深度学习方法也在逐渐应用于此领域, 具体表述

如下。

3.1 考虑时间信息的空气质量预测

考虑时间信息的空气质量预测往往把空气污染数据看成一维时间序列或一维时间信号, 通过一定的预处理方式如离散傅里叶变换 (discrete Fourier transform, DFT)^[47]、小波变换 (wavelet transform, WT)^[48] 等将数据分解或平稳化, 最后将处理好的数据通过深度学习模型提取时间特征并输出预测结果。最常用于该过程的深度模型主要包括 RNN、LSTM、GRU (gated recurrent neural network)^[49] 及 1D CNN 等。

Zhang 等^[50]提出一种结合变分模态分解 (variational mode decomposition, VMD)^[51] 与双向长短期记忆网络 (BiLSTM)^[52] 的一种混合深度学习模型 VMDBiLSTM, 用于实现基于时间数据对北京站点的 PM_{2.5} 预测。所研究的时间序列通过 VMD 方法将原始的 PM_{2.5} 数据自适应地分解为几个子信号分量, 降低了原有数据的波动性与非线性从而实现了数据的平稳化。该模型采用 BiLSTM 分别对各分量进行预测, 显著提高了预测精度。实验结果表明, BiLSTM 可以有效地从时间序列数据中提取特征, 对于非线性且非平稳的 PM_{2.5} 数据, 考虑使用 VMD 等信号处理方法, 可以大大提升预测性能。

丁子昂等^[53]提出一种基于互补集合经验模态分解方法-皮尔逊相关分析 (complementary ensemble empirical mode decomposition, CEEMD)^[54] 及 LSTM 的一种 PM_{2.5} 预测方法, 使用 CEEMD 方法对时序数据分解并筛选, 最后输入到深度网络进行训练并输出预测结果, 该模型在真实数据集中获得了 90% 以上的预测精度, 结果表明 CEEMD 方法可以挖掘历史数据的隐藏特征, 增强数据的时序性, 可以使 LSTM 网络获得最佳训练效果, 有效地进行 PM_{2.5} 预测。

Jin 等^[55]研究提出一种结合多嵌套长短期记忆网络 (multiple nested long short term memory networks, MTMC-NLSTM) 的新型深度学习框架, 采用小波变换将原始数据分解为去噪低频分量和高频分量, 分别提取出数据的总体模式特征与短期特征, 将处理后的数据输入到 NLSTM^[56] 中输出北京 12 个站点的 AQI 值预测结果, 结果表明该框架在多元空气质量时间预测中具有显著的优势, 且由于训练时间较短, 几乎可以实时跟进实际的 AQI 值。

Tao 等^[57]结合 1D CNN 与双向门控循环单元网络 (BiGRU) 提出一种基于深度学习的 PM_{2.5} 预测方法

(CBGRU), 该模型结合卷积神经网络和循环神经网络的时间预测能力, 通过 CNN 对数据下采样减少数据的复杂性, 并使用 BiGRU 挖掘时间信息特征输出预测结果. 孙蒙等^[58] 基于 DBN 和差分进化算法对北京地区气象与 AQI 数据建立预测模型 (DE-DBN), 使用基于 BP 神经网络和统计的网络模型的预测结果作为对比. 结果表明深度网络模型比起传统的统计模型, 机器学习方法具有更好地预测水平, 寻优算法可以有效优化深度学习模型. Retta 等^[59] 比较了 LSTM、GRU、ConvLSTM^[60] 等网络对 PM_{2.5} 时间序列的序列学习和预测能力, 实验结果发现时间卷积网络 (temporal convolutional networks, TCN)^[61] 在单变量时间序列处理中也具有和 LSTM、GRU 相似的性能.

综上所述, 空气质量预测是一个多变量时间序列预测问题. 考虑时间信息的空气质量预测以捕获数据中的时间模式为目的, 通过深度网络挖掘空气污染数据与气象数据等不同数据源提供的信息特征, 建立多变量时间序列与预测污染物之间的非线性关系, 是最为简单、最为常见的深度学习预测方法, 可以取得较好的预测性能. 但空气污染数据除了具有显著的时间特性, 其大部分来源于地理空间相邻的空气质量监测站, 单纯考虑时间信息的空气质量预测没有考虑监测站之间的空间依赖性. 例如, 某一区域 PM_{2.5} 浓度可能会受到其他区域的影响. 空间信息的丢失会导致预测精度、稳定性的降低. 所以, 将空间特征处理并嵌入到深度网络中是更为先进、有效的做法.

3.2 考虑时空信息的空气质量预测

由于空气污染数据的时空特性, 在深度模型中加入空间信息往往可以提高预测的精准度^[62], 因而现在大部分的空气质量预测深度学习模型都会同时对数据的空间和时间依赖性进行建模. 对空间模型建模的深度网络主要有 DNN、CNN 等, 图神经网络 (graph neural networks, GNN)^[63] 以及引入卷积操作的 GNN 图卷积神经网络 (graph convolutional network, GCN)^[64] 也是近年较为流行的空间信息建模的深度学习模型. 现有的考虑时空信息的模型主要有 CNN-LSTM、CNN-GRU、GCN-LSTM 等.

Wu 等^[65] 提出一种名为多尺度空间时间网络 (multi-scale spatial temporal network, MSSTN) 的新的深度卷积神经网络模型, 该网络分为 3 个子网, 扩张卷积网络来提取时间尺度的特征, 图卷积网络用来提取空间特

征, 融合网络用于特征融合输出预测结果, 通过 3 个网络的分工可以更好地发现多尺度时空模式特征及区域间的相互作用. 实验结果表明对空间关系以及站间关系的建模可以大大提升预测的精准度.

Han 等^[66] 提出了多对抗时空递归图神经网络 (multi-adversarial spatiotemporal recurrent graph neural networks, MasterGNN) 用于联合空气质量和天气预测. 他们提出一种模拟空气质量与监测站之间时空自相关关系的递归图神经网络, 通过对抗图学习框架来对抗时空建模引入的噪声. 为了更好地训练网络, 他们将多对抗学习作为一个多任务学习问题, 自适应的训练. 实验结果表明, 图神经网络与对抗学习的结合可以有效提升预测精度, 该网络在空气质量与天气预测的任务上优于其他先进的方法.

Huang 等^[67] 提出一种名为空间注意力嵌入式循环神经网络 (spatio-attention embedded recurrent neural network, SpAttRNN) 的 AQI 预测模型, 对于每个空气监测站分别输入时间特征和空间特征, 通过综合利用监测站之间的动态时空相关性进行预测. 该方法引入归一化邻接矩阵, 矩阵包含目标站与其他站点的地理距离及几何关系, 通过注意力矩阵表示观测站之间的重要相关性, 完成对空间相关性的建模. LSTM 用来对来自目标站点的原始时间特征进行编码. 注意力矩阵与 LSTM 连接输出预测结果, 完成预测任务. Yeo 等^[68] 使用 5 层 CNN 架构和 3 层 GRU 模型来生成首尔地区 25 个站点的 PM_{2.5} 实时预测, 通过多层模型提高深度学习模型的计算效率. 同时, 本研究还提出一种地理模型, 可以确定提高目标站点预测准确性所需相邻站点的最佳数量.

Ge 等^[69] 提出一种多尺度时空图卷积网络 (multi-scale spatiotemporal graph convolution network, MST-GCN), 该网络由一个多尺度模块以及几个时空模块组成, 最后用一个融合模块融合多个组. 为了对空间以及长时间依赖性进行建模, 其中的时空模块包含一个图卷积层和一个时间卷积层. Lin 等^[70] 通过基于地理上下文的扩散卷积递归神经网络 (geo-context based diffusion convolutional recurrent neural network, GC-DCRNN) 对短期 PM_{2.5} 进行预测, 该网络通过地理环境相似性的图形来描述空间关系, 其中相似性根据每个位置对空气质量影响的重要性计算, 该实验在两个真实空气质量数据集中证明了有效性.

综上所述,同时考虑检测站之间历史数据的时间性和空间性比常规考虑时间信息的预测表现出更好的性能.考虑时空信息可以学习多变量空气污染数据的局部变化趋势及时空依赖特征.现今大部分基于深度学习的空气质量预测研究都以挖掘数据的时空性为主.虽然考虑时空信息的空气质量预测是目前较为成熟的预测方法,但大部分的时空深度模型仍有一定的局限性,例如 CNN-LSTM、GCN-LSTM 等停留在传统深度神经网络的结合,避免不了传统时间深度网络难以捕捉长时间依赖模式的弊端.此外,空间特性和时间依赖性之间同样存在复杂的相互作用,不同的污染物也可能具有不同的扩散速度.因此,探索适应性更强的深度学习模型框架来应对复杂的时空相互作用是基于深度学习空气质量预测的发展趋势.

3.3 基于注意力机制的空气质量预测

近年来,RNN、LSTM 等循环结构网络模型已经在序列预测、自然语言处理等应用中获得了不俗的效果,但由于 RNN 固有的顺序框架使得训练样本无法并行化,所以在长时间序列的批量处理中往往会受到内存的限制,注意力机制 (attention model)^[71] 源于人类视觉注意力,注意力模型可以应用于任何类型的输入而不管其形状如何,可以解决 RNN 等网络常常出现的计算资源分配不均问题,因而逐渐成为深度学习领域最热门的模型之一. Transformer^[72] 是基于自注意力机制 (self-attention) 的全新神经网络架构,比起常规的 RNN 等网络,Transformer 拥有更高的预测质量却仅需要较少的算力,作为一个纯注意力机制模型,结合了多头注意力机制 (multi-head attention) 与残差模块,并加入了位置编码 (position embedding, PE) 来表示序列中元素的相对或绝对位置关系,完全依赖于自注意力机制对全局关系建模,可以更好地学习远程依赖性,更重要的是,可以并行计算从而充分利用 GPU 资源. Transformer 在自然语言处理和计算机视觉领域已经大放异彩,在处理时间序列数据时,同样可以取得不错的成果^[73]. Informer^[74] 是基于 Transformer,专为长时间序列预测设计的模型,有效解决了 Transformer 在内存使用量,时间复杂度的一定局限性,打破了编解码器结构的固有限制,模型具有较强的解决长距离依赖问题的能力,可以很好地解决空气质量预测在时间跨度较长时无法获得满意的结果的问题.

Wang 等^[75] 将北京与天津各空气预测站点之间的

空间邻接关系、时间模式相似性以及功能相似性编码到图中,提出了一种注意时间图卷积网络 (attentive temporal graph convolutional network, ATGCN),该网络通过并行编码块将图卷积操作合并于 GRU 中,利用注意力单元最终在解码模块输出所有站点的多步预测.结果表明,图卷积网络可以聚合站点间的特征,达到更好的预测效果. Padhi 等^[76] 基于 Transformer 提出一种层次表 BERT (bidirectional encoder representations from Transformers)^[77] 来学习表格时间序列,通过学习到的编码层来预测空气污染浓度,并在真实数据集中证明了有效性.

Wang 等^[78] 提出一种加入注意力机制的混合序列到序列空气质量预测模型,用于预测区域地面臭氧浓度.该模型通过序列到序列注意力模型 (attention-based Seq2Seq)^[79] 提取输入的原始数据以及辅助数据的时间依赖,同时使用深度网络提取网络中的空间信息,最后通过 DNN 将时间与空间模型结合输出预测结果.实验说明考虑预测模型的时空相关性对于区域臭氧的预测具有显著的优势.

综上所述,注意力机制可以灵活地聚合站点间的历史信息,并随时间自适应的调整站点之间的依赖关系权重,挖掘时空信息的深层特征.最为重要的是,注意力机制对长期依赖的捕捉不受序列长度的限制,在中、长期空气质量预测中取得了令人满意的结果.注意力机制模型简单却拥有较高的辨别能力,可以从多尺度、宏观的角度把握空气数据的时空关系.以注意力机制为核心,从空间和时间角度处理并分析数据源是现今该领域较为成熟的深度学习模型框架.表 2 总结了 3 种基于深度学习的空气质量预测方法.

近年来,随着硬件设备的发展,计算能力迅速提升,同时,大量城市相继建立起空气质量监测站以对空气污染物实时监控,扩充了空气污染研究的数据集.由于深度学习可以有效提取大规模数据的内在特征,越来越多的人将深度学习作为空气质量预测的主要方法.但是,深度学习方法并不能完全取代传统空气质量预测.首先,在空气污染排放源数据来源全面、真实且设备完善的情况下,数值模型方法可以达到惊人的准确率.其次,在预测多频次短时空气污染数据时,统计方法比起深度学习方法更有效率,如对多地区空气质量数据的实时监控.最后,传统的预测方法拥有较高的可解释性,而深度学习大部分为数据驱动的黑盒模型,

难以保证质量和调试. 随着深度学习的发展, 上述的不足也逐渐被弥补. 但现今阶段, 深度学习方法可以通过与传统预测方法结合, 取长补短, 发挥深度学习优势的同时用传统方法的优势去弥补深度学习模型的不足. 例如, 结合统计方法与深度学习方法的混合模型以及传统机器学习与深度学习方法的集成模型^[80].

表2 基于深度学习的空气质量预测方法的总结与比较

方法类别	时间	作者	研究地点	模型名称
考虑时间信息	2018	孙蒙等 ^[58]	北京	DE-DBN
	2019	Tao等 ^[57]	北京	CBGRU
	2021	Zhang等 ^[50]	北京	VMDBiLSTM
	2021	Jin等 ^[55]	北京	MTMC-NLSTM
	2021	Retta等 ^[59]	北京	TCN
考虑时空信息	2018	Lin等 ^[70]	北京、洛杉矶	GC-DCRNN
	2019	Wu等 ^[65]	中国华北地区	MSSIN
	2021	Han等 ^[66]	北京、上海	MasterGNN
	2021	Huang ^[67]	北京	SpAttRNN
	2021	Yeo等 ^[68]	首尔	CNN-GRU
	2021	Ge等 ^[69]	北京	MST-GCN
考虑注意力机制	2020	Wang等 ^[78]	北京	Attention-based Seq2Seq
	2021	Wang等 ^[75]	北京、天津	ATGCN
	2021	Padhi等 ^[76]	北京	BERT

4 挑战与机遇

近年来, 使用深度学习进行空气质量预测的技术已经越来越成熟, 但仍存在着大量的挑战阻碍了研究的进展, 笔者总结了空气质量预测面临的挑战以及如何应对将挑战转换为未来的发展方向, 内容主要如下.

(1) 数据集的片面性与稀疏性. 数据集的综合性与全面性一直是深度学习研究的一个重要影响因素, 虽然各地区逐渐建立起大量的空气质量监测站, 但如今尚未有一个包含空气污染数据与气象数据、交通情况、海拔高度等全面的大型数据集, 而且由于空气质量预测的研究往往需要足够的历史数据集, 而因为时间的久远, 数据中往往有大量的缺失值, 数据较为稀疏, 必须在数据预处理阶段对数据进行填充, 难免会影响到预测结果, 因此需要探索一种能够克服或减轻新建站点数据短缺问题的预测模型. 迁移学习 (transfer learning)^[81,82] 的主要思想是将相关领域的标注数据或知识结构迁移来改进目标任务的学习效果, 在深度学习中往往表现为冻结预训练模型的部分网络层, 只训练自己的全连接层. 迁移学习可以很好地解决空气质

量预测训练成本较高, 数据集较小的问题. Ma 等^[83] 提出了一个基于迁移学习的堆叠双向长短期记忆网络 (transfer learning based stacked bidirectional long short term memory, TLS-BLSTM), 通过将在已有站点对 BLSTM 层进行预训练并冻结, 然后使用新站点的数据对剩下的网络层进行微调, 经过在安徽所有新的空气质量监测站的测试表明, 迁移学习可以大大提升新站点的预测性能.

(2) 训练成本问题. 现有的经典深度学习网络结构往往存在训练参数过多, 开销较大的问题, 而空气质量预测由于其时效性, 快速而准确地预测十分重要, 如何在不影响网络性能的同时, 尽可能地减少训练成本, 提高训练效率, 是目前亟待解决的关键问题. 动态神经网络 (dynamic neural networks)^[84] 可以根据输入数据动态自适应地调整自身的结构与参数, 可以有效地减少计算成本, 增加模型的通用性和可解释性. 时间自适应动态网络 (temporal-wise dynamic networks)^[85] 被设计用来对不同时间位置的数据进行自适应计算, 可以从多方面减少冗余计算, 空气污染数据作为典型的时序数据可以与动态神经网络有效结合从而减少计算量, 增加计算效率.

(3) 如何挖掘空气质量监测站之间的深层关系. 现有的基于深度学习的空气质量预测模型通常旨在优化特定预测任务的目标函数, 这些方法往往忽略了空气质量监测站之间潜在的非线性空间相关性, 如何对相邻区域具有高相似性的监测站的特征信息进行信息共享是解决问题的关键. 多任务学习 (multi-task learning, MTL)^[86] 是一种将多个相关任务放在一起学习的一种机器学习方法, 显著提高了学习的泛化效果. 优点是可以通过其相互作用帮助逃离单任务学习可能陷入的局部最小值, 且多个相关的辅助任务可以提高主任务的预测性能^[87]. 在空气污染预测中, 数据往往来源于具有很多相关性的密集监测站点, 通过多任务学习可以实现任务间的信息共享以及特定信息的融合, 可以更好地准确地预测空气污染物之间复杂的相互作用.

(4) 增加不同特征输入. 空气污染数据往往受到多方面的影响, 例如, 高风速会降低 PM_{2.5} 的浓度, 高湿度通常会加重空气污染, 高大气压通常会导致良好的空气质量, 上述天气数据特征对空气质量预测具有重要意义, 同样, 地理地形信息, 交通信息也会影响空气质量预测的结果, 未来可以探索结合卫星、遥感图像或

加入地形、POI 信息等数据对空气质量预测的影响。

5 总结

空气质量预测是一个多学科交叉的研究课题,其中基于深度学习的空气质量预测已经成为近年机器学习领域的研究热点. 本文从介绍近年来国内外空气污染数据集入手,详细介绍了传统空气质量预测方法并重点阐述了深度学习在空气质量预测上的应用. 最后,本文指出了该领域未来的发展趋势,整理并归纳了空气质量预测未来的挑战与机遇.

参考文献

- 1 中国气象局. “十四五”公共气象服务发展规划. 北京: 中国气象局, 2021.
- 2 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. 中国环境监测, 2021, 37(2): 16–22. [doi: 10.19316/j.issn.1002-6002.2021.02.03]
- 3 An JL, Huang MY, Wang ZF, *et al.* Numerical regional air quality forecast tests over the Mainland of China. *Water, Air, and Soil Pollution*, 2001, 130(1–4): 1781–1786.
- 4 Shahraiyni HT, Sodoudi S. Statistical modeling approaches for PM₁₀ prediction in urban areas; A review of 21st-century studies. *Atmosphere*, 2016, 7(2): 15. [doi: 10.3390/atmos7020015]
- 5 Zhu SL, Lian XY, Liu HX, *et al.* Daily air quality index forecasting with hybrid models: A case in China. *Environmental Pollution*, 2017, 231: 1232–1244. [doi: 10.1016/j.envpol.2017.08.069]
- 6 Jaiswal JK, Samikannu R. Application of random forest algorithm on feature subset selection and classification and regression. *Proceedings of the 2017 World Congress on Computing and Communication Technologies (WCCCT)*. Tiruchirappalli: IEEE, 2017. 65–68.
- 7 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436–444. [doi: 10.1038/nature14539]
- 8 Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, 18(7): 1527–1554. [doi: 10.1162/neco.2006.18.7.1527]
- 9 LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]
- 10 Elman JL. Finding structure in time. *Cognitive Science*, 1990, 14(2): 179–211. [doi: 10.1207/s15516709cog1402_1]
- 11 Forsyth D, Ponce J. *Computer Vision: A Modern Approach*. 2nd ed., New York: Pearson, 2011.
- 12 Yu D, Deng L. *Automatic Speech Recognition*. London: Springer, 2015.
- 13 Nguyen G, Dlugolinsky S, Bobák M, *et al.* Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey. *Artificial Intelligence Review*, 2019, 52(1): 77–124. [doi: 10.1007/s10462-018-09679-z]
- 14 Li X, Peng L, Hu Y, *et al.* Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, 2016, 23(22): 22408–22417. [doi: 10.1007/s11356-016-7812-9]
- 15 Du SD, Li TR, Yang Y, *et al.* Deep air quality forecasting using hybrid deep learning framework. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(6): 2412–2424. [doi: 10.1109/TKDE.2019.2954510]
- 16 Yan R, Liao JQ, Yang J, *et al.* Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, 2021, 169: 114513. [doi: 10.1016/j.eswa.2020.114513]
- 17 Wang JS, Song GJ. A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing*, 2018, 314: 198–206. [doi: 10.1016/j.neucom.2018.06.049]
- 18 Wang DS, Wang HW, Li C, *et al.* Roadside air quality forecasting in Shanghai with a novel sequence-to-sequence model. *International Journal of Environmental Research and Public Health*, 2020, 17(24): 9471. [doi: 10.3390/ijerph17249471]
- 19 Zhang KF, Thé J, Xie GY, *et al.* Multi-step ahead forecasting of regional air quality using spatial-temporal deep neural networks: A case study of Huaihai Economic Zone. *Journal of Cleaner Production*, 2020, 277: 123231. [doi: 10.1016/j.jclepro.2020.123231]
- 20 Zhang Q, Wu S, Wang XW, *et al.* A PM_{2.5} concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations. *Journal of Cleaner Production*, 2020, 275: 122722. [doi: 10.1016/j.jclepro.2020.122722]
- 21 Wang JY, Li JZ, Wang XX, *et al.* Air quality prediction using CT-LSTM. *Neural Computing and Applications*, 2021, 33(10): 4779–4792. [doi: 10.1007/s00521-020-05535-w]
- 22 Dairi A, Harrou F, Khadraoui S, *et al.* Integrated multiple directed attention-based deep learning for improved air pollution forecasting. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 3520815.
- 23 Abirami S, Chitra P. Regional air quality forecasting using spatiotemporal deep learning. *Journal of Cleaner Production*,

- 2021, 283: 125341. [doi: [10.1016/j.jclepro.2020.125341](https://doi.org/10.1016/j.jclepro.2020.125341)]
- 24 Byun D, Schere KL. Review of the governing equations, computational algorithms, and other components of the Models-3 community multiscale air quality (CMAQ) modeling system. *Applied Mechanics Reviews*, 2006, 59(2): 51–77. [doi: [10.1115/1.2128636](https://doi.org/10.1115/1.2128636)]
- 25 Tesche TW, Morris R, Tonnesen G, *et al.* CMAQ/CAMx annual 2002 performance evaluation over the eastern US. *Atmospheric Environment*, 2006, 40(26): 4906–4919. [doi: [10.1016/j.atmosenv.2005.08.046](https://doi.org/10.1016/j.atmosenv.2005.08.046)]
- 26 Wang ZF, Li J, Wang Z, *et al.* Modeling study of regional severe hazes over mid-eastern China in January 2013 and its implications on pollution prevention and control. *Science China Earth Sciences*, 2014, 57(1): 3–13. [doi: [10.1007/s11430-013-4793-0](https://doi.org/10.1007/s11430-013-4793-0)]
- 27 Akaike H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 1973, 60(2): 255–265. [doi: [10.1093/biomet/60.2.255](https://doi.org/10.1093/biomet/60.2.255)]
- 28 Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 2003, 50: 159–175. [doi: [10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)]
- 29 Genc DD, Yesilyurt C, Tuncel G. Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. *Environmental Monitoring and Assessment*, 2010, 166(1–4): 11–27. [doi: [10.1007/s10661-009-0981-y](https://doi.org/10.1007/s10661-009-0981-y)]
- 30 Hu JL, Chen JJ, Ying Q, *et al.* One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system. *Atmospheric Chemistry and Physics*, 2016, 16(16): 10333–10350. [doi: [10.5194/acp-16-10333-2016](https://doi.org/10.5194/acp-16-10333-2016)]
- 31 Skamarock WC, Klemp JB. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, 2008, 227(7): 3465–3485. [doi: [10.1016/j.jcp.2007.01.037](https://doi.org/10.1016/j.jcp.2007.01.037)]
- 32 谢磊, 铁治欣, 宋飞扬, 等. 基于最优定权组合法的大气污染物 SO₂ 预测. *计算机系统应用*, 2019, 28(3): 80–87. [doi: [10.15888/j.cnki.csa.006814](https://doi.org/10.15888/j.cnki.csa.006814)]
- 33 Kumar U, Jain VK. ARIMA forecasting of ambient air pollutants (O₃, NO, NO₂ and CO). *Stochastic Environmental Research and Risk Assessment*, 2010, 24(5): 751–760. [doi: [10.1007/s00477-009-0361-8](https://doi.org/10.1007/s00477-009-0361-8)]
- 34 Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 1978, 6(2): 461–464.
- 35 Cheng CH, Huang SF. Forecasting of ozone concentration using frequency MA-OWA model. *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics*. San Antonio: IEEE, 2009. 3461–3466.
- 36 Chiarella C, He XZ, Hommes C. A dynamic analysis of moving average rules. *Journal of Economic Dynamics and Control*, 2006, 30(9–10): 1729–1753. [doi: [10.1016/j.jedc.2005.08.014](https://doi.org/10.1016/j.jedc.2005.08.014)]
- 37 Yager RR. Applications and extensions of OWA aggregations. *International Journal of Man-Machine Studies*, 1992, 37(1): 103–122. [doi: [10.1016/0020-7373\(92\)90093-Z](https://doi.org/10.1016/0020-7373(92)90093-Z)]
- 38 Lei MT, Monjardino J, Mendes L, *et al.* Macao air quality forecast using statistical methods. *Air Quality, Atmosphere & Health*, 2019, 12(9): 1049–1057.
- 39 Loh WY. Classification and regression trees. *WIREs Data Mining Knowledge Discovery*, 2011, 1(1): 14–23. [doi: [10.1002/widm.8](https://doi.org/10.1002/widm.8)]
- 40 Kelley K, Bolin JHH. Multiple regression. *Handbook of Quantitative Methods for Educational Research*. In: Teo T, ed. Rotterdam: Springer, 2013. 71–101.
- 41 Liu B, Shi C, Li J, *et al.* Comparison of different machine learning methods to forecast air quality index. *Proceedings of the International Conference on Frontier Computing*. 2018. 235–245.
- 42 Nieto PG, Combarro EF, del Coz Díaz JJ, *et al.* A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study. *Applied Mathematics and Computation*, 2013, 219(17): 8923–8937. [doi: [10.1016/j.amc.2013.03.018](https://doi.org/10.1016/j.amc.2013.03.018)]
- 43 Kumar A, Goyal P. Forecasting of daily air quality index in Delhi. *Science of the Total Environment*, 2011, 409(24): 5517–5523. [doi: [10.1016/j.scitotenv.2011.08.069](https://doi.org/10.1016/j.scitotenv.2011.08.069)]
- 44 Al-Alawi SM, Abdul-Wahab SA, Bakheit CS. Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. *Environmental Modelling & Software*, 2008, 23(4): 396–403.
- 45 Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2016, 374(2065): 20150202. [doi: [10.1098/rsta.2015.0202](https://doi.org/10.1098/rsta.2015.0202)]
- 46 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- 47 Weinstein S, Ebert P. Data transmission by frequency-division multiplexing using the discrete Fourier transform. *IEEE Transactions on Communication Technology*, 1971, 19(5): 628–634. [doi: [10.1109/TCOM.1971.1090705](https://doi.org/10.1109/TCOM.1971.1090705)]
- 48 Zhang DS. Wavelet transform. *Fundamentals of Image Data*

- Mining. Cham: Springer, 2019. 35–44.
- 49 Cho K, van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha: ACL, 2014. 1724–1734.
- 50 Zhang ZD, Zeng YK, Yan K, *et al.* A hybrid deep learning technology for PM_{2.5} air quality forecasting. Environmental Science and Pollution Research, 2021, 28(29): 39409–39422. [doi: [10.1007/s11356-021-12657-8](https://doi.org/10.1007/s11356-021-12657-8)]
- 51 Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Transactions on Signal Processing, 2014, 62(3): 531–544. [doi: [10.1109/TSP.2013.2288675](https://doi.org/10.1109/TSP.2013.2288675)]
- 52 Zhang B, Zhang HW, Zhao GM, *et al.* Constructing a PM_{2.5} concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. Environmental Modelling & Software, 2020, 124: 104600.
- 53 丁子昂, 乐曹伟, 吴玲玲, 等. 基于 CEEMD-Pearson 和深度 LSTM 混合模型的 PM_{2.5} 浓度预测方法. 计算机科学, 2020, 47(S1): 444–449.
- 54 Yeh JR, Shieh JS, Huang NE. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. Advances in Adaptive Data Analysis, 2010, 2(2): 135–156. [doi: [10.1142/S1793536910000422](https://doi.org/10.1142/S1793536910000422)]
- 55 Jin N, Zeng YK, Yan K, *et al.* Multivariate air quality forecasting with nested long short term memory neural network. IEEE Transactions on Industrial Informatics, 2021, 17(12): 8514–8522. [doi: [10.1109/TII.2021.3065425](https://doi.org/10.1109/TII.2021.3065425)]
- 56 Moniz JRA, Krueger D. Nested LSTMs. Proceedings of the 9th Asian Conference on Machine Learning. Seoul: PMLR, 2017. 530–544.
- 57 Tao Q, Liu F, Li Y, *et al.* Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU. IEEE Access, 2019, 7: 76690–76698. [doi: [10.1109/ACCESS.2019.2921578](https://doi.org/10.1109/ACCESS.2019.2921578)]
- 58 孙蒙. 基于优化深度置信网络的空气质量指数预测模型研究 [硕士学位论文]. 北京: 中国地质大学(北京), 2018.
- 59 Retta S, Yarramsetti P, Kethavath S. Comprehensive analysis of deep learning approaches for PM_{2.5} forecasting. Proceedings of International Conference on Computational Intelligence and Data Engineering. Singapore: Springer, 2021. 311–322. [doi: [10.1007/978-981-15-8767-2_27](https://doi.org/10.1007/978-981-15-8767-2_27)]
- 60 Shi XJ, Chen ZR, Wang H, *et al.* Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: ACM, 2015. 802–810.
- 61 Lea C, Flynn MD, Vidal R, *et al.* Temporal convolutional networks for action segmentation and detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1003–1012.
- 62 Liao Q, Zhu MM, Wu L, *et al.* Deep learning for air quality forecasts: A review. Current Pollution Reports, 2020, 6(4): 399–409. [doi: [10.1007/s40726-020-00159-z](https://doi.org/10.1007/s40726-020-00159-z)]
- 63 Xu K, Hu WH, Leskovec J, *et al.* How powerful are graph neural networks? Proceedings of the 7th International Conference on Learning Representations. New Orleans: ICLR, 2018.
- 64 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. Proceedings of the 5th International Conference on Learning Representations. Toulon: ICLR, 2017.
- 65 Wu ZY, Wang Y, Zhang L. MSSTN: Multi-scale spatial temporal network for air pollution prediction. Proceedings of the 2019 IEEE International Conference on Big Data (Big Data). Los Angeles: IEEE, 2019. 1547–1556. [doi: [10.1109/BigData47090.2019.9005574](https://doi.org/10.1109/BigData47090.2019.9005574)]
- 66 Han JD, Liu H, Zhu HS, *et al.* Joint air quality and weather prediction based on multi-adversarial spatiotemporal networks. Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2021. 4081–4089.
- 67 Huang Y, Ying JJC, Tseng VS. Spatio-attention embedded recurrent neural network for air quality prediction. Knowledge-based Systems, 2021, 233: 107416. [doi: [10.1016/j.knsys.2021.107416](https://doi.org/10.1016/j.knsys.2021.107416)]
- 68 Yeo I, Choi Y, Lops Y, *et al.* Efficient PM_{2.5} forecasting using geographical correlation based on integrated deep learning algorithms. Neural Computing and Applications, 2021, 33(22): 15073–15089. [doi: [10.1007/s00521-021-06082-8](https://doi.org/10.1007/s00521-021-06082-8)]
- 69 Ge L, Wu KY, Zeng Y, *et al.* Multi-scale spatiotemporal graph convolution network for air quality prediction. Applied Intelligence, 2021, 51(6): 3491–3505. [doi: [10.1007/s10489-020-02054-y](https://doi.org/10.1007/s10489-020-02054-y)]
- 70 Lin YJ, Mago N, Gao Y, *et al.* Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Seattle: ACM, 2018. 359–368.
- 71 Niu ZY, Zhong GQ, Yu H. A review on the attention mechanism of deep learning. Neurocomputing, 2021, 452: 48–62. [doi: [10.1016/j.neucom.2021.03.091](https://doi.org/10.1016/j.neucom.2021.03.091)]

- 72 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 6000–6010.
- 73 Bilgin O, Mąka P, Vergutz T, *et al.* TENT: Tensorized encoder transformer for temperature forecasting. arXiv: 2106.14742, 2021.
- 74 Zhou HY, Zhang SH, Peng JQ, *et al.* Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2021. 11106–11115.
- 75 Wang CY, Zhu YM, Zang TZ, *et al.* Modeling inter-station relationships with attentive temporal graph convolutional network for air quality prediction. Proceedings of the 14th ACM International Conference on Web Search and Data Mining. ACM, 2021. 616–634.
- 76 Padhi I, Schiff Y, Melnyk I, *et al.* Tabular transformers for modeling multivariate time series. Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto: IEEE, 2021. 3565–3569.
- 77 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2018. 4171–4186.
- 78 Wang HW, Li XB, Wang DS, *et al.* Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. Journal of Cleaner Production, 2020, 253: 119841. [doi: [10.1016/j.jclepro.2019.119841](https://doi.org/10.1016/j.jclepro.2019.119841)]
- 79 Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: ACM, 2014. 3104–3112.
- 80 Luo ZP, Huang JQ, Hu K, *et al.* AccuAir: Winning solution to air quality prediction for KDD Cup 2018. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 1842–1850. [doi: [10.1145/3292500.3330787](https://doi.org/10.1145/3292500.3330787)]
- 81 Yosinski J, Clune J, Bengio Y, *et al.* How transferable are features in deep neural networks? Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: ACM, 2014. 3320–3328.
- 82 Oquab M, Bottou L, Laptev I, *et al.* Learning and transferring mid-level image representations using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 1717–1724.
- 83 Ma J, Li Z, Cheng JCP, *et al.* Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. Science of the Total Environment, 2020, 705: 135771. [doi: [10.1016/j.scitotenv.2019.135771](https://doi.org/10.1016/j.scitotenv.2019.135771)]
- 84 Xu ZH, Li S, Zhou XF, *et al.* Dynamic neural networks for motion-force control of redundant manipulators: An optimization perspective. IEEE Transactions on Industrial Electronics, 2021, 68(2): 1525–1536. [doi: [10.1109/TIE.2020.2970635](https://doi.org/10.1109/TIE.2020.2970635)]
- 85 Han YZ, Huang G, Song SJ, *et al.* Dynamic neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 7436–7456. [doi: [10.1109/TPAMI.2021.3117837](https://doi.org/10.1109/TPAMI.2021.3117837)]
- 86 Zhang Y, Yang Q. A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 2021. [doi: [10.1109/TKDE.2021.3070203](https://doi.org/10.1109/TKDE.2021.3070203)]
- 87 Xu XH, Yoneda M. Multitask air-quality prediction based on LSTM-autoencoder model. IEEE Transactions on Cybernetics, 2021, 51(5): 2577–2586. [doi: [10.1109/TCYB.2019.2945999](https://doi.org/10.1109/TCYB.2019.2945999)]

(校对责编: 牛欣悦)