



DOI:10.13364/j.issn.1672-6510.20230047

基于多尺度时空优化的空气质量预测方法

董梅, 张贤坤, 黄文杰, 秦锋斌, 宋琛
(天津科技大学人工智能学院, 天津 300457)

摘要: 本文提出一种基于多尺度时空优化的空气质量预测方法 (multi-scale spatial-temporal network for air quality prediction, MSSTN-AQP), 结合空气质量系统中存在的长短期时间依赖关系和动态空间依赖性, 提高长期空气质量预测的准确性。首先, 通过构建多尺度时空特征提取模块, 从多源异构数据中提取时空特征。其次, 构建动态空间特征提取模块。通过将图卷积网络与注意力机制进行有效结合, 捕捉空气质量网络中的全局空间特征, 用于对多种空间依赖关系的联合建模。最后, 构建时间特征提取模块, 对 Transformer 模型进行改进与优化。自适应时间 Transformer 模块主要用于模拟跨多个时间步长的双向时间依赖关系。此外, 将上述时空特征提取模块进行有效集成化, 构建端到端的空气质量预测模型。为了验证模型的有效性, 在两个真实数据集中进行实验验证。实验结果表明, MSSTN-AQP 在预测精度上更具优势, 尤其是在长期的空气质量预测任务中优势更加明显。

关键词: 空气质量预测; 多尺度时空特征提取; 图卷积网络; 自适应时间 Transformer

中图分类号: TP391 文献标志码: A 文章编号: 1672-6510(2024)02-0071-10

Multi-Scale Spatial-Temporal Optimization Method for Air Quality Prediction

DONG Mei, ZHANG Xiankun, HUANG Wenjie, QIN Fengbin, SONG Chen
(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: In this article we propose an air quality prediction method based on Multi-Scale Spatial-Temporal Network for air quality prediction (MSSTN-AQP) to improve the accuracy of long-term air quality prediction by combining the long- and short-term time dependence and dynamic spatial dependence in the air quality system. First, the spatial-temporal features are extracted from multi-source heterogeneous data by constructing a multi-scale spatial-temporal feature extraction module. Second, the dynamic spatial feature extraction module is constructed. With an effective combination of the graph convolutional network with the attention mechanism, the global spatial features in the air quality network are captured and used for the joint modeling of multiple spatial dependencies. Finally, the temporal feature extraction module is constructed, which is to improve and optimize the Transformer model. The adaptive time Transformer module is mainly used to simulate bidirectional time dependencies across multiple time steps. Moreover, the above temporal feature extraction module is effectively integrated to construct an end-to-end air quality prediction model. To verify the effectiveness of MSSTN-AQP, extensive experiments were conducted on two real data sets. The experimental results showed that MSSTN-AQP was more advantageous in prediction accuracy, especially in long-term air quality prediction tasks.

Key words: air quality prediction; multi-scale spatial-temporal feature extraction; graph convolutional network; adaptive time Transformer

引文格式:

董梅, 张贤坤, 黄文杰, 等. 基于多尺度时空优化的空气质量预测方法[J]. 天津科技大学学报, 2024, 39(2): 71-80.

收稿日期: 2023-03-07; 修回日期: 2023-06-07

基金项目: 天津市自然科学基金项目(19JCYBJC15300); 天津市研究生科研创新项目(2021YJSS04)

作者简介: 董梅(1997—), 女, 甘肃临洮人, 硕士研究生; 通信作者: 宋琛, 讲师, songchen@tust.edu.cn

DONG M, ZHANG X K, HUANG W J, et al. Multi-scale spatial-temporal optimization method for air quality prediction [J]. Journal of Tianjin university of science & technology, 2024, 39(2): 71-80.

空气污染严重影响着人们的日常生活,对人类的身体造成损伤,甚至会引起心血管疾病、呼吸道疾病以及诱发肺部癌变等^[1-4]。近些年,空气质量监测站的不断建立并每小时公布空气质量数据,人们可以实时查询当前空气质量指数。除了实时公布当前的空气质量指数之外,作为空气污染防治工作的重要一环,空气质量预测问题也受到越来越多研究者的关注。通过空气质量监测站收集到的数据可以对未来的空气质量进行预测,以便提前准备相应的预防措施^[5-6]。

空气质量系统的复杂性、多样性和动态性等特征使挖掘空气中污染物浓度的变化规律以及精准地预测未来的空气质量更具难度^[7-8]。空气污染由细颗粒物(PM_{2.5})、可吸入颗粒物(PM₁₀)以及部分气体(NO₂、CO、O₃和SO₂)的污染共同构成,对人类健康有持续性的危害。造成空气污染的影响因素众多,主要包括气象因素、地理位置和兴趣点(POI)信息等,而且不同类型污染物排放的频率和周期也不同。与此同时,空气质量系统中存在复杂的时空相互作用^[9]。因此,解决空气质量预测问题必须要考虑不同时间和不同位置污染扩散的复杂时空相互作用。为此,将空气质量监测站建模成时空网络是一种比较贴合实际的研究思路。在该时空网络图中,节点可由空气质量监测站表示,边的权重由不同站点之间连接性和欧氏距离所决定。

深度学习技术在计算机视觉、智能推荐、自然语言处理等领域被广泛应用。在空气质量预测问题研究中也有很多学者引入其先进的技术。目前的研究工作重心主要集中在利用深度学习技术分析其存在空间属性关系或时间依赖关系。Liang等^[10]考虑监测站之间空间和时间的关系,通过计算没有拓扑结构的监测站之间的重要性学习空间依赖性。Qi等^[11]通过同时考虑空气质量指数(AQI)预测的空间属性和时间依赖性,使用多个子模型包括一个图卷积神经网络和循环神经网络编码几个特征向量的空间属性和时间依赖性。这些模型生成特征向量后,直接将空间特征向量和时间特征向量进行拼接,利用全连通层将拼接后的向量转化为预测结果。但是,编码器将这些空间和时间相关性特征分别压缩成几个特征向量时,往往会忽略一些关键的相关性,因此很难解决复杂的时空相互作用问题。这主要原因在于空间相关特征

不能直接干预与时间相关的预测,反之亦然。另外,还有一些工作将卷积网络与基于卷积的序列学习模型或递归神经网络(recurrent neural network, RNN)相结合,共同捕捉空间和时间上的依赖关系^[12-14]。在空气质量预测问题中,尤其是在长期预测方面,仍然存在以下两个问题:

(1)忽略动态空间依赖性。在空气污染物的传播过程中产生的空间依赖关系会发生动态变化。因此,有效捕捉这些动态的空间依赖关系十分必要。

(2)忽略长期时间依赖性。在现有的空气质量预测方法中,长期的时间依赖性很容易被忽略。空气质量与时间往往存在相关性,存在周期性空气污染状况。因此,挖掘其时间维度上的长期关联模式很有必要。

为了解决上述问题,本文提出了一种基于多尺度时空优化的空气质量预测方法(multi-scale spatial-temporal network for air quality prediction, MSSTN-AQP),旨在解决空气质量预测任务存在的时空相关性问题。首先,提出一个基于多尺度时空优化的空气质量预测模型,动态建模长期的时空依赖关系。其次,将图神经网络与注意力机制有效结合,构建时空优化图神经网络,有效提取空气质量网络中的全局空间特征,并对空间依赖关系进行联合建模。结合空气质量预测中多变量时间序列预测特征,提出自适应时间Transformer模块,与RNN及其变体相比,改进之后的Transformer可以轻松扩展到长序列,并且能够并行处理长程依赖关系。

为了验证模型的有效性,本文使用两个真实的数据集,从不同的角度评估MSSTN-AQP模型和其他多种模型的预测性能,验证该模型的预测效果以及泛化能力。

1 相关工作

空气质量预测模型主要分为数值物理模型和数据驱动模型^[8]。传统的数值物理方法是一种确定性方法,具有良好的可解释性,如高斯羽流模型^[15]、街道峡谷模型^[16]和计算流体动力学模型^[17]等。由于该类模型要求污染源数据的完整性,因此模型的泛化能力较差。为了获得更准确的预测结果,研究者们提出了数据驱动的空气品质预测方法,包括传统的线性统计

模型、机器学习技术和深度学习方法^[11]。统计预测方法利用时间序列空气质量历史数据,采用统计建模策略对未来空气质量进行预测^[18-20]。统计预测方法可分为两种模型(线性模型和非线性模型),对空气质量的预测范围更广。线性统计预测模型基于真实观测数据的假设线性进行预测,包括自回归移动平均(auto regression moving average, ARMA)^[18]和自回归综合移动平均(auto regressive integrated moving average, ARIMA)^[19]。现实世界中的观测数据大多是非线性的,传统的非线性统计预测方法为机器学习(machine learning, ML)模型。多种常用的机器学习算法,包括多元线性回归(multiple linear regression, MLR)^[20]、人工神经网络(artificial neural network, ANN)^[21-22]、支持向量回归(support vector regression, SVR)^[23-24]、随机森林(random forest, RF)^[13]以及ML模型的集成学习^[25]已被用于PM_{2.5}质量浓度的预测。机器学习方法具有独特的数学逻辑,输入和输出数据之间的相关性相对确定。由于网络结构较浅,所以依赖时间序列空气质量历史数据进行建模的能力有限。近年来,深度学习在一些领域中取得了长足的进步。深度学习模型通过叠加多个神经网络,拟合从输入到输出的非线性转换,实现特征自动学习。GeoMAN模型^[10]利用注意力机制对同一传感器上不同度量值之间的相关性和不同传感器上相同度量值之间的相关性进行建模。这些模型通常使用RNN及其变体在时间维度上模拟复杂依赖关系,使用卷积神经网络(convolutional neural network, CNN)或图神经网络(graph neural network, GNN)在空间维度上模拟复杂依赖关系^[26-30]。Ong等^[31]选择K近邻城市,使用动态预训练的深度循环神经网络预测PM_{2.5}质量浓度。这些模型可以最大限度地从数据中提取有用的信息,有效处理和学习数据之间的时间依赖性。然而,这些模型大多只考虑相邻站点在地理空间区域的时空依赖性,忽略了远距离位点之间可能由其他复杂影响因素引起的强相关性。

空气质量预测问题是典型的多变量时间序列预测问题,在以往的工作中大多以RNN及其变体为核心模型进行研究,但是也存在一些问题,如其在训练中存在梯度爆炸或梯度消失问题,因此RNN对时间依赖的建模是有限的^[32-33]。为了克服这些缺点,门控递归单元(gated recurrent unit, GRU)^[34]和长短期记忆(long short term memory, LSTM)^[35]被用来进行长距离依赖关系建模。这些序列模型仍然存在耗时的

训练过程和对长序列建模的有限可扩展性的问题。基于卷积的序列学习模型被作为一种替代方案,但需要多个隐藏层覆盖有限的感受野下的大背景。研究者采用了扩张卷积的WaveNet扩大感受野,从而减少了隐藏层的数量^[12]。由于模型隐藏层的数量与输入序列的长度呈线性增长,所以模型的可扩展性在长输入序列中受到限制。此外,序列中各部分之间路径长度的增加,使捕捉长距离依赖关系的效率受到更深层的影响^[12,31-35]。Transformer^[36]可以通过高度可并行的自注意力机制实现高效的序列学习,长距离时间变化的依赖性可以通过一个单层从不同长度的输入序列中自适应捕获。

2 MSSTN-AQP模型

将空气质量预测任务表述为一个时空预测问题。为了解决这一问题,描述了本文模型的整体架构,包括多尺度特征提取层、基于编码器-解码器的时空网络特征提取模块以及时空特征融合层。

2.1 问题分析与定义

空气质量预测是空气污染预警和智能控制的关键,其目的是预测目标监测站随时间变换的未来空气污染物浓度,是一个典型的时空序列预测问题。

根据 T 时刻前的历史数据,预测未来 $T+1$ 时刻的空气污染物浓度 $X_{i,T+1}$,预测未来 $T+n$ 时刻的空气污染物浓度 $X_{i,T+n}$,其中未来空气污染物浓度是基于与空气质量相关的历史时刻序列数据集建模后进行预测。在本研究中,将空气质量预测问题中的相关特征划分为时间特征和空间特征。时间特征包括空气污染物(如PM_{2.5}、PM₁₀、SO₂、CO、O₃等)和气象信息(如温度、压力和湿度等),用 X_T^{tempo} ($X_T^{\text{tempo}} \in \mathbf{R}^F$)表示 T 时刻监测站的时间特征信息,其中 F 表示时间特征的数量,包含空气污染物和气象数据信息;空间特征包括监测站的地理空间信息,其不随时间发生变化。空间特征可以从POI数据中提取(例如目标区域周围特定位置类型的数量),记为 X_T^{spatial} ($X_T^{\text{spatial}} \in \mathbf{R}^S$,其中 S 为空间特征的数量)。利用空间地图信息构建各监测站的站间关系。

因此,每个时间戳的最终输入特征矩阵为

$$X_T \in \mathbf{R}^{N+M} \quad (1)$$

$$M = F + S \quad (2)$$

其中: N 为研究区域内监测站的数量, M 为输入特征的维度。

为了表示不同监测站之间的空间关系,定义包含 N 个节点(研究区域内的所有监测站)的无向图 G , $G=(V,E,A)$, 其中 $V(v_i \in V)$ 是站点的集合, $E(<v_i, v_j > \in E)$ 是站点之间潜在交互的边的集合, $A(A \in R^{N \times N})$ 表示一个空间权重矩阵, 其中每个元素 $A_{i,j}$ 表示 v_i 和 v_j 之间的定量空间相关性, 边的权重表示站点之间的关系强度。在本文提出的模型中采用欧氏距离表示关系, 这意味着监测值越低, 其相关性越强。在图构建中, 其站点之间的边关系根据两个监测站之间的地理位置的欧氏距离计算得到, 距离阈值设置为 300 km, 即若两个监测站之间的距离大于 300 km, 则表示这两个监测站之间不存在边关系。另外, 站点之间的边属性由风向以及风速决定。

2.2 MSSTN-AQP 模型结构

MSSTN-AQP 模型由编码器-解码器构成的端到

端模型, 其核心组件包括多尺度特征提取层、时空优化图卷积网络模块、自适应时间 Transformer 模块以及时空特征融合模块。MSSTN-AQP 模型可以对目标区域污染物以及周围邻近区域内的空气质量进行动态转化, 并对目标站点空气质量预测产生的影响进行分析研究。在不用人为处理数据提取特征的情况下, MSSTN-AQP 模型通过反复训练可以从大量有监督数据集中自动学习空气质量数据中的动态时空依赖特征, 进而对目标站点进行细粒度空气质量预测。模型中堆叠的时空特征编码块和时空特征解码块可以进一步对堆叠形成的深层时空特征在动态依赖的背景下联合提取时空特征。预测层通过利用池化操作和多层感知机聚合这些时空特征, 最终的输出结果为空气中目标污染物的浓度预测值。该模型结构如图 1 所示。

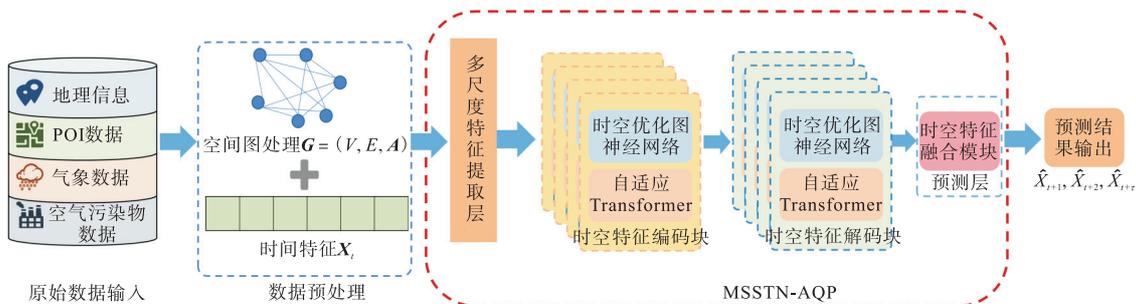


图 1 MSSTN-AQP 模型总体框架

Fig. 1 Overall framework of MSSTN-AQP model

2.3 时空特征编码块

一个监测站未来的空气质量状况由其邻近监测站的空气质量状况、观测的时间步长以及天气状况等决定。在本小节中, 通过开发一个时空特征编码块和解码块集成空间和时间特征提取器, 以联合建模空气质量网络中的空间依赖性和时间依赖性, 实现准确预测。

由于并非所有监测数据都对预测目标位置的空气质量有相同的贡献, 因而同时对空间和时间特征进行建模, 本文提出一种新的时空优化图神经网络和自适应时间 Transformer 作为时空依赖编码器的核心组件。

2.3.1 时空优化图神经网络

时空优化图神经网络由动态图卷积层和注意力机制融合而成。将构建好的多尺度时空图作为图神经网络的输入, 在图神经网络层中嵌入注意力机制, 使其更有效地提取空间特征信息。通过固定的图卷积层和一个动态图卷积层, 分别探索空间依赖的固定

和有向的动态组件。利用注意力机制融合学习到的全局空间特征。时空优化图神经网络可以被视为用于动态图构建和特征学习的通用消息传递 GNN。

图卷积是将经典的基于网格的卷积推广到图域。在学习到的权重和预定义的图的基础上, 通过聚合相邻节点的信息获取节点特征。采用基于切比雪夫多项式近似的图卷积学习结构感知的节点特征, 从网络拓扑中捕获平稳的空间依赖。

定义 g 的度矩阵为 D , 其对角元素为

$$D_{ii} = \sum_j A_{ij} \tag{3}$$

归一化拉普拉斯矩阵为

$$L = I_n - D^{-1/2} A D^{-1/2} \tag{4}$$

可缩放的拉普拉斯矩阵 $\tilde{L} = 2L / \lambda_{\max} - I_n$ 为切比雪夫多项式, 其中 λ_{\max} 为 L 的最大特征值。给定嵌入特征 \hat{X}^s , 通过切比雪夫多项式逼近的图卷积得到结构感知节点特征 \hat{X}^g ($\hat{X}^g \in R^{N \times d_g}$), 每个时间步长的阶数为 $k = 1, \dots, K$ 。

$$\hat{X}_{:,j}^g = \sum_{i=1}^{d_g} \sum_{k=1}^K \theta_{ij,k} T_k \hat{L} \hat{X}_{:,j}^s \quad \forall j=0, \dots, d_g \quad (5)$$

其中: $\hat{X}_{:,j}^g$ 为 \hat{X}^g 的第 j 个通道, $\theta_{ij,k}$ 为可学习参数。由于 g 是基于监测站之间的物理连通性和距离构建的, 因此可以通过固定的图卷积层显式地探索由网络拓扑决定的固定空间依赖关系。

基于图卷积网络的模型只能建模静态空间依赖性。为了捕获随时间演化的隐藏空间依赖关系, 将注意力机制的原始结构进行了改进, 旨在实现高维潜在子空间的训练和建模。通过学习线性映射, 将每个节点的输入特征投影到高维潜在子空间。投影特征采用自注意力机制, 根据变化的图信号有效建模节点之间动态的空间依赖关系, 然而预定义的网络拓扑结构并不能充分反映空气质量网络内部的动态空间依赖关系。通过并行化处理, 同时学习多个线性映射, 对受各种潜在子空间中各种因素影响的动态有向空间依赖关系进行建模。

将每个时间步的嵌入特征 \hat{X}^s 映射到高维潜在子空间, 在这里的映射过程主要通过前馈神经网络实现。当注意力机制考虑一种空间依赖模式时, 基于 \hat{X}^s 可以同时为每个节点训练子空间。此外, 为减少计算时的计算成本和存储成本, 在该模块中使用点积的方式, 利用 Softmax 对空间依赖关系进行归一化, d_A^s 防止了 Softmax 函数导致的饱和性。因此, 节点特征 M^s ($M^s \in \mathbb{R}^{N \times d_g}$) 用 S^s 进行更新。

$$M^s = S^s V_s \quad (6)$$

注意机制可以通过引入子空间学习多种模式的空间依赖关系, 从而可以从不同的潜在子空间中建模不同的隐藏空间依赖关系。

此外, 在每个节点上应用具有非线性激活的共享三层前馈神经网络, 进一步提高学习到的节点特征 M^s 的预测能力。为了探索特征通道之间的相互作用, 使用 U^s ($U^s \in \mathbb{R}^{N \times d_g}$) 更新 M^s 。

$$U^s = \text{ReLU}(\text{ReLU}(M^s W_0^s) W_1^s) W_2^s \quad (7)$$

其中: M^s 为可扩展训练的残差连接, $M^s = \hat{X}^s + M^s$, M^s 采用 $Y^s = U^s + M^s$ 结合门控机制进行特征融合。为了提高模型的空间依赖能力, 该模型中堆叠了多个动态图卷积层。

2.3.2 自适应时间 Transformer

为了有效捕获长程时间依赖, MSSTN-AQP 模型中引入 Transformer, 并提出自适应时间 Transformer。模型详细结构如图 2 所示。与 RNN 及其变体相比, 自适应 Transformer 可以轻松扩展到长序列, 并且能够并行处理长程依赖关系。输入特征为

$$X^T = X^s + Y^s \in \mathbb{R}^{M \times N \times d_g} \quad (8)$$

将输入特征与时间嵌入 D^T 进行连接可得到

$$X^T = G_t([X^T, D^T]) \in \mathbb{R}^{M \times N \times d_g} \quad (9)$$

其中 D^T 是一个 1×1 卷积层, 在每个时间步为每个节点产生 d_g 维向量。考虑到 g 中任意一个节点的二维空间特征张量 $X^T \in \mathbb{R}^{M \times d_g}$, 因此对每个节点进行并行化处理可以有效对时序依赖关系进行建模。

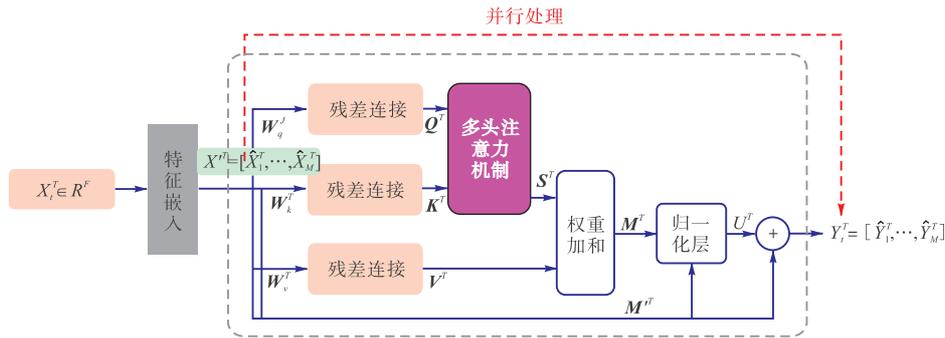


图 2 自适应时间 Transformer
Fig. 2 Adaptive time Transformer

采用自注意力机制对时序依赖关系进行建模, 其目的是有效进行长程的时间特征提取。自适应 Transformer 的输入是一个时间序列

$$X^T \in \mathbb{R}^{M \times d_g} \quad (10)$$

其具有长度为 M 和 d_g 通道的滑动窗口, 自适应 Transformer 是由查询子空间 $Q^T \in \mathbb{R}^{N \times d_A^T}$ 、关键子空间

$K^T \in \mathbb{R}^{N \times d_A^T}$ 和值子空间 $V^T \in \mathbb{R}^{N \times d_A^T}$, 动态计算时序依赖关系。

$$\begin{aligned} Q^T &= \hat{X}^T W_q^T \\ K^T &= \hat{X}^T W_k^T \\ V^T &= \hat{X}^T W_v^T \end{aligned} \quad (11)$$

其中 $W_q^T \in \mathbb{R}^{d_g \times d_A^T}$ 、 $W_k^T \in \mathbb{R}^{d_g \times d_A^T}$ 和 $W_v^T \in \mathbb{R}^{d_g \times d_A^T}$ 分别为可学习的线性映射。

与以往研究不同的是,基于 RNN 的模型仅限于考虑基于之前时间步骤的时序依赖,而自适应时间 Transformer 模型将值 V^T 与时间特征 M^T 的权重 S^T 聚合。

$$M^T = S^T + V^T \quad (12)$$

在自适应时间 Transformer 的每一层中都有效捕获了长程双向时序依赖关系,这是因为每个时间步都关注滑动窗口内的剩余时间步,可以很容易地通过增加 M 扩展到长序列,而不会牺牲太多的计算效率。相比之下,基于 RNN 的模型将面临梯度消失或梯度爆炸的问题,而基于卷积的模型必须明确指定随着 M 增长的卷积层数。

2.4 模型优化

空气质量预测任务的目的是预报下一个时间段监测站的空气质量状况。本文使用均方误差(MSE)损失(用符号 L_{MSE} 表示)计算最小化预测值之间的误差,损失函数定义为

$$L_{MSE} = \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{N} \sum_{i=1}^N (\hat{X}_i^t - X_i^t)^2 \right) \quad (13)$$

基于 MSSTN-AQP 模型的输出结果得到预测结果,在预测层中使用池化操作和一个多层感知机进行特征融合和预测。

3 实验

为了验证本文方法在空气质量预测任务中的有效性,本次研究中采用两个真实的空气质量数据集进行空气质量预测实验。

3.1 数据集及实验环境

本研究中使用的数据来源于中国空气质量数据监测平台和国家气候数据中心。本次实验共收集并处理了两个真实数据集。

数据集 1(京津冀):共收集 2017 年 1 月 1 日至 2019 年 6 月 30 日期间京津冀地区 GMT+8 时区气象监测站和大气污染物浓度监测站的数据,每 3 小时记录一次。气象数据属性包括温度、露点、海平面气压、风向、风速、天空条件总覆盖度(云量)和 $PM_{2.5}$ 质量浓度。该数据集包括来自 21 个监测站的 152 565 个历史数据。

数据集 2(北京):收集 2014 年 5 月 1 日至 2015 年 4 月 30 日期间中国北京的 36 个监测站每小时的

空气污染物浓度数据,包括 $PM_{2.5}$ 、 PM_{10} 、 SO_2 、 NO_2 、 CO 和 O_3 。数据集中包括来自 36 个监测站的 341 532 个历史数据。

数据预处理:从监测站收集的数据可能由于设备维护或其他原因中断而丢失。首先,对原始数据进行预处理。本研究中通过沿时间维度的线性插值填充空气质量数据中的缺失值。选取站点所在的网格或区域,并获取每个站点的历史气象数据,按照时间顺序以 7:1:2 的比例分割数据集,以产生不重叠的训练集、验证集和测试集数据。本研究中将 $PM_{2.5}$ 作为目标污染物。两个数据集的统计数据见表 1。

表 1 数据集统计数据

Tab. 1 Statistics of the datasets

数据描述	数据集 1	数据集 2
空气质量监测站/个	21	36
历史观测数据	152 565	314 532
空气质量影响因素/个	7	12
时间跨度	2017-01-01— 2019-06-30	2014-05-01— 2015-04-30
数据采集时间间隔/h	3	1
输入长度	4	12
训练集	106 806	220 176
验证集	15 246	31 464
测试集	30 513	62 892

实验环境配置:本次实验在单个 NVIDIA TI-TANX 上运行。通过在 PyCharm 开发环境中安装配置 PyTorch 框架并使用 Python 编程语言完成相关实验。所有的项目在带有 CPU(Intel Xeon Platinum 8163)、32 GB RAM 和 GPU(NVIDIA Tesla V100)的服务器上完成。

3.2 评价指标

采用空气质量预测问题中被广泛使用的 3 个评估指标平均绝对误差(mean absolute error, MAE, 用符号 E_{MAE} 表示)、均方根误差(root mean square error, RMSE, 用符号 E_{RMSE} 表示)和平均绝对百分比误差(mean absolute percentage error, MAPE, 用符号 E_{MAPE} 表示)衡量预测模型的性能,将 MSSTN-AQP 与其他方法进行比较。

假设 $x = x_1, \dots, x_n$ 表示当前时刻下监测站监测到的真实值, $\hat{x} = \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ 表示模型预测值,则模型性能评估指标定义为

$$E_{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (14)$$

$$E_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (15)$$

$$E_{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{x}_i - x_i}{x_i} \right| \quad (16)$$

3.3 基线模型

本研究选用以下 7 个模型与 MSSTN-AQP 模型进行比较。

HA: 历史平均水平。该方法使用历史观测值的平均值作为预测值。

ARMA^[19]: 时间序列数据预测是一种传统的线性统计方法。由于 ARMA 是有限的多步空气质量预测策略,所以本文中仅将其用于单步空气质量预测。

SVR^[22]: 向量自回归模型。该模型是用于回归任务的传统数值模型。

DCRNN^[12]: 扩散卷积递归神经网络。编码器和解码器都包含 2 个循环层。

GC-LSTM^[11]: 将 LSTM 和图卷积网络(GCN)相结合,分别对时间依赖性和空间依赖性进行建模。

PM_{2.5}-GNN^[29]: 利用图神经网络引入领域知识,并于 RNN 集成化建模,捕获复杂的空气质量传播过

程中的细粒度和长期依赖关系。

HighAir^[26]: 一种基于层次图神经网络的空气质量预测方法。采用编码器-解码器结构,并考虑了复杂的空气质量影响因素,从层次的角度构建城市级图和站点级图。

3.4 实验结果与分析

为了验证模型的有效性和可行性,对 MSSTN-AQP 模型与现有模型的整体性能进行比较,通过 RMSE、MAPE 和 MAE 评估指标进行详细比较,结果见表 2。由表 2 可知:MSSTN-AQP 模型与其他经典模型相比,在预测性能上提升非常明显;在不同数据集中,无论是短时间预测还是长时间预测,该模型都表现出良好的预测优势。在北京 36 个站点的数据集中,MSSTN-AQP 模型在 24 h 预测中,RMSE、MAPE 和 MAE 指标分别比现有最好的模型均有不同程度提升,表明在空气质量预测任务中有效引入 Transformer 模型的优越性。

表 2 不同模型的总体性能比较

Tab. 2 Comparison of overall performance of different models

数据集	模型	1 h			6 h			12 h			24 h		
		RMSE	MAPE	MAE									
北京	HA	75.87	211.59	55.32	75.87	211.59	55.32	75.87	211.59	55.32	75.87	211.59	55.32
	ARMA	24.90	23.04	11.12	36.62	37.06	17.84	44.42	50.01	23.09	50.28	62.59	27.39
	SVR	24.31	26.53	11.77	34.90	41.89	18.39	42.35	56.55	23.69	48.14	70.80	28.21
	DCRNN	26.83	30.39	14.49	33.05	37.77	18.10	38.11	44.98	21.18	42.37	52.18	23.87
	GC-LSTM	25.67	31.30	13.81	32.34	39.28	17.53	38.04	47.86	20.85	42.95	56.62	23.87
	GCRNN	25.56	27.45	13.60	31.95	34.09	17.13	37.87	40.62	20.19	41.32	47.02	22.89
	PM _{2.5} -GNN	26.26	27.37	16.04	34.11	33.37	21.23	37.06	40.42	22.02	41.28	47.67	22.63
	HighAir	25.74	25.83	13.38	32.11	33.13	16.97	37.50	40.60	20.18	41.95	47.90	23.05
	MSSTN-AQP	24.37	21.60	11.37	29.23	25.72	13.19	33.50	35.79	16.45	39.67	34.61	19.83
京津冀	HA	54.23	154.00	38.40	54.23	154.00	38.40	54.23	154.00	38.40	54.23	154.00	38.40
	ARMA	25.27	38.40	14.42	34.58	60.21	20.75	39.58	73.66	24.50	42.80	82.65	26.90
	SVR	22.84	43.11	14.17	29.84	64.01	19.34	33.52	75.91	22.06	35.83	83.97	23.78
	DCRNN	16.21	39.16	12.93	18.98	46.22	14.75	20.12	53.66	16.50	24.13	54.32	19.43
	GC-LSTM	16.88	40.20	14.29	18.18	48.72	16.12	20.33	55.40	18.24	22.56	60.95	23.86
	GCRNN	15.86	36.05	12.32	17.92	50.42	19.95	24.76	56.57	21.93	26.89	61.29	23.42
	PM _{2.5} -GNN	15.33	36.22	10.77	17.48	43.32	12.76	20.38	54.91	14.55	22.23	60.21	16.99
	HighAir	15.14	36.43	10.45	17.26	41.51	12.84	18.63	45.66	14.79	21.47	58.86	16.44
	MSSTN-AQP	14.89	34.21	8.43	16.51	39.79	9.52	17.58	42.57	10.53	18.72	44.93	12.36

3.5 消融研究

为了评估本文模型对空间相关性建模和时间依赖性建模的有效性,分别将两种变体模型与 MSSTN-AQP 模型的预测结果进行了实验比较。

MSSTN-AQP/noT: 与 MSSTN-AQP 结构相同,少了时间特征提取组件,即仅能够学习空间特征。

MSSTN-AQP/noS: 与 MSSTN-AQP 结构相同,

少了空间特征提取组件,即仅能够学习时间特征。

本实验选择北京市的数据集进行实验验证,因为相较于京津冀地区的城市样本数据集,该数据集具有更大的数据规模以及复杂的时空依赖关系,更具研究挑战性。为了有效评估,实验过程中对核心组件进行了统一的变量控制,展示了 MSSTN-AQP 模型核心组件的性能,基于文章篇幅,本文仅展示 MAE 的结

果对比。图 3 实验结果说明 MSSTN-AQP 模型在很大程度上优于单纯的时间建模和空间建模, 预测性能均表现为最佳, 尤其在长时间的预测中, 预测性能的提升更加明显。因此, 本文模型的时空优化特征提取模块可以有效建模动态空间依赖关系和长程时间依赖关系, 联合时空特征提取可以提高空气质量预测的精确性。

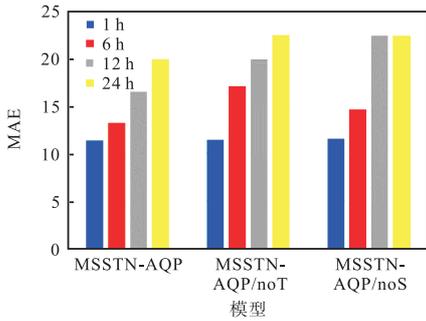


图 3 消融实验结果对比

Fig. 3 Comparison of ablation experiment results

3.6 模型参数设置

MSSTN-AQP 模型的参数设置包括时空特征编码块以及解码块的数量、特征通道的设置以及隐藏层的数量和多头注意力的数量, 通过调参实验, 保留最佳参数组合, 使模型的预测性能达到最优。

在 MSSTN-AQP 模型中通过编码器-解码器的方式, 级联多个时空特征提取块, 构建时空依赖模型, 使模型预测中的 MAE 下降, 但当时空块堆叠数量大于 2 时, MAE 不再降低(图 4)。

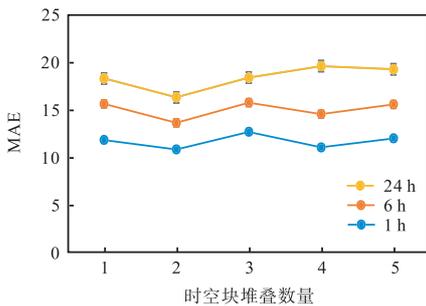


图 4 模型参数优化

Fig. 4 Model parameter optimization

在文中图神经网络以及 Transformer 的参数调优过程中, 实验结果对比发现在特征通道数为 64×64 、隐藏层设置为 2、注意力头数量为 4 时会得到最优的评估指标值, MSSTN-AQP 模型的预测性能达到最优。由于文章篇幅有限, 具体的参数调优过程这里不再详细展开论述。

4 结 语

本文提出一种基于多尺度时空优化的空气质量预测方法, 利用时空优化图神经网络和注意力机制的有效结合建模动态的空气质量系统中的时空依赖关系, 并通过改进的 Transformer 建模空气质量预测任务中的时间依赖关系, 然后将时空特征信息进行有效提取和整合, 进而进行下一步目标污染物的预测。该方法既考虑了目标站点历史数据中的时间特征信息, 又通过动态提取其潜在的时空依赖特征信息。在接下来的工作中, 将进一步优化模型的框架结构以及参数设置, 使其尽可能用于更多时序预测任务中。

参考文献:

- [1] 吴一鸣, 周怡静, 田贺忠, 等. 我国城市生活垃圾处理处置全过程大气排放研究进展[J]. 环境科学研究, 2018, 31(6): 991-999.
- [2] 中国气象局. “十四五”公共气象服务发展规划[M]. 北京: 气象出版社, 2021.
- [3] 张莹. 我国典型城市空气污染特征及其健康影响和预报研究[D]. 兰州: 兰州大学, 2016.
- [4] 刘杰. 北京大气污染物时空变化规律及评价预测模型研究[D]. 北京: 北京科技大学, 2015.
- [5] CHEN L, XU J, WU B, et al. Group-aware graph neural network for nationwide city air quality forecasting [EB/OL]. [2023-01-20]. <https://doi.org/10.48550/arXiv.2108.12238>.
- [6] SILLMAN S. The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments [J]. Atmospheric environment, 1999, 33(12): 1821-1845.
- [7] ZHENG Y, LIU F, HSIEH H P. U-Air: when urban air quality inference meets big data[C]//ACM. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2013: 1436-1444.
- [8] PAN L, SUN B, WANG W. City air quality forecasting and impact factors analysis based on grey model[J]. Procedia engineering, 2011, 12: 74-79.
- [9] DU S, LI T, GONG X, et al. Traffic flow forecasting based on hybrid deep learning framework[C]//IEEE. 2017 12th International Conference on Intelligent Sys-

- tems and Knowledge Engineering (ISKE). New York: IEEE, 2017: 1–6.
- [10] LIANG Y, KE S, ZHANG J, et al. GeoMAN: multi-level attention networks for geo-sensory time series prediction [C]//IJCAI. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: IJCAI, 2018: 3428–3434.
- [11] QI Y, LI Q, KARIMIAN H, et al. A hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory[J]. Science of the total environment, 2019, 664: 1–10.
- [12] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 922–929.
- [13] CHENG W, SHEN Y, ZHU Y, et al. A neural attention model for urban air quality inference: learning the weights of monitoring stations[EB/OL]. [2023–01–20]. <https://aaai.org/ocs/AAAI/AAAI18/paper/viewPaper/16607>.
- [14] LIN Y, MAGO N, GAO Y, et al. Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning[C]//ACM. Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2018: 359–368.
- [15] RAKOWSKA A, WONG K C, TOWNSEND T, et al. Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon[J]. Atmos environ, 2014, 98: 260–270.
- [16] SCAAR H, TEODOROV T, ZIEGLER T, et al. Computational fluid dynamics(CFD) analysis of air flow uniformity in a fixed-bed dryer for medicinal plants[J]. IST International symposium on CFD applications in agriculture, 2012, 1008: 119–126.
- [17] KINDAP T, UNAL A, CHEN S H, et al. Long-range aerosol transport from Europe to Istanbul, Turkey[J]. Atmospheric environment, 2006, 40(19): 3536–3547.
- [18] CEKIM H O. Forecasting PM₁₀ concentrations using time series models: a case of the most polluted cities in Turkey[J]. Environmental science and pollution research, 2020, 27: 25612–25624.
- [19] DONNELLY A, MISSTEAR B, BRODERICK B. Real time air quality forecasting using integrated parametric and non-parametric regression techniques[J]. Atmospheric environment, 2015, 103: 53–65.
- [20] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results[J]. Journal of transportation engineering, 2003, 129(6): 664–672.
- [21] CHU J, DONG Y, HAN X, et al. Short-term prediction of urban PM_{2.5} based on a hybrid modified variational mode decomposition and support vector regression model[J]. Environmental science and pollution research, 2021, 28: 56–72.
- [22] GARIAZZO C, CARLINO G, SILIBELLO C, et al. A multi-city air pollution population exposure study: combined use of chemical-transport and random-forest models with dynamic population data[J]. Science of the total environment, 2020, 724: 138102.
- [23] XIAO Q, CHANG H H, GENG G, et al. An ensemble machine-learning model to predict historical PM_{2.5} concentrations in China from satellite data[J]. Environmental science & technology, 2018, 52: 13260–13269.
- [24] GUO Y, TANG Q, GONG D Y, et al. Estimating ground-level PM_{2.5} concentrations in Beijing using a satellite-based geographically and temporally weighted regression model[J]. Remote sensing of environment, 2017, 198: 140–149.
- [25] OPREA M, MIHALACHE S F, POPESCU M. Applying artificial neural networks to short-term PM_{2.5} forecasting modeling[C]//IFIP International Conference on Artificial Intelligence Applications and Innovations. Berlin: Springer, 2016: 204–211.
- [26] XU J, CHEN L, LÜ M, et al. Highair: a hierarchical graph neural network-based air quality forecasting method[EB/OL]. [2023–01–20]. <https://doi.org/10.48550/arXiv.2101.04264>.
- [27] CHEN L, DING Y, LYU D, et al. Deep multi-task learning based urban air quality index modelling[J]. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 2019, 3(1): 1–17.
- [28] GE L, WU K, ZENG Y, et al. Multi-scale spatiotemporal

- graph convolution network for air quality prediction[J]. Applied intelligence, 2021, 51: 3491–3505.
- [29] WANG S, LI Y, ZHANG J, et al. PM_{2.5}-GNN: a domain knowledge enhanced graph neural network for PM_{2.5} forecasting[C]//ACM. Proceedings of the 28th International Conference on Advances in Geographic Information Systems. New York: ACM, 2020: 163–166.
- [30] ZHENG Y, YI X, LI M, et al. Forecasting fine-grained air quality based on big data[C]//ACM. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 2267–2276.
- [31] ONG B T, SUGIURA K, ZETTSU K. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM_{2.5}[J]. Neural computing and applications, 2016, 27(6): 1553–1566.
- [32] MA X, TAO Z, WANG Y, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. Transportation research part C: emerging technologies, 2015, 54: 187–197.
- [33] WU Y, TAN H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework[EB/OL]. [2023-01-20]. <https://doi.org/10.48550/arXiv.1612.01022>.
- [34] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. [2023-01-20]. <https://doi.org/10.48550/arXiv.1412.3555>.
- [35] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735–1780.
- [36] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. [2023-01-20]. <https://doi.org/10.48550/arXiv.1706.03762>.

责任编辑: 郎婧

(上接第 33 页)

- 应用研究[D]. 杭州: 浙江工商大学, 2020.
- [41] 王雅妮. 三明治型 KGM/SA/KGM 复合涂膜对三文鱼

鱼片的保鲜性能研究[D]. 锦州: 渤海大学, 2021.

责任编辑: 郎婧