

基于自适应层级图卷积的多站点空气质量预测模型^①



张石清^{1,2}, 胡 炜¹, 赵小明²

¹(浙江理工大学 计算机科学与技术学院, 杭州 310018)

²(台州学院 智能信息处理研究所, 台州 318000)

通信作者: 张石清, E-mail: tzczsq@163.com

摘 要: 时空预测任务在污染治理、交通、能源、气象等领域应用广泛。PM_{2.5} 浓度预测作为典型的时空预测任务, 需要对空气质量数据中的时空依赖关系进行分析和利用。现有时空图神经网络 (ST-GNNs) 研究所使用的邻接矩阵使用启发式规则预定义, 无法准确表示站点之间的真实关系。本文提出了一种自适应分层图卷积神经网络 (AHGCNN) 用于 PM_{2.5} 预测。首先, 引入了一种分层映射图卷积架构, 在不同层级上使用不同的自学习邻接矩阵, 以有效挖掘不同站点之间独特的时空依赖。其次, 以基于注意力的聚合机制连接上下层邻接矩阵, 加速收敛过程。最后, 将隐藏的空间状态与门控循环单元相结合, 形成一个统一的预测架构, 同时捕捉多层次的空间依赖关系和时间依赖关系, 提供最终的预测结果。实验中, 我们与 7 种主流预测模型进行对比, 结果表明该模型可以有效获取空气监测站点之间的时空依赖, 提高预测精确度。

关键词: 空气质量; PM_{2.5}; 深度学习; 图卷积; 时空依赖

引用格式: 张石清, 胡炜, 赵小明. 基于自适应层级图卷积的多站点空气质量预测模型. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9484.html>

Multi-site Air Quality Forecasting Model Using Adaptive Hierarchical Graph Convolution

ZHANG Shi-Qing^{1,2}, HU Wei¹, ZHAO Xiao-Ming²

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

²(Institute of Intelligent Information Processing, Taizhou University, Taizhou 318000, China)

Abstract: Spatiotemporal forecasting finds extensive applications in domains such as pollution management, transportation, energy, and meteorology. Predicting PM_{2.5} concentration, as a quintessential spatiotemporal forecasting task, necessitates the analysis and utilization of spatiotemporal dependencies within air quality data. Existing studies on spatiotemporal graph neural networks (ST-GNNs) either employ predefined heuristic rules or trainable parameters for adjacency matrices, posing challenges in accurately representing authentic inter-station relationships. This study introduces the adaptive hierarchical graph convolutional neural network (AHGCNN) to address these issues concerning PM_{2.5} prediction. Firstly, a hierarchical mapping graph convolutional architecture is introduced, employing distinct self-learning adjacency matrices at different hierarchical levels, efficiently uncovering unique spatiotemporal dependencies among various monitoring stations. Secondly, an attention-based aggregation mechanism is employed to connect adjacency matrices across different hierarchical levels, expediting the convergence process. Finally, the hidden spatial states are fused with gated recurrent unit (GRU), forming a unified predictive framework capable of concurrently capturing multi-level spatial and temporal dependencies, ultimately delivering the prediction results. In the experiments,

① 基金项目: 国家自然科学基金 (62276180); 浙江省自然科学基金 (LZ20F020002)

收稿时间: 2023-11-06; 修改时间: 2023-12-04; 采用时间: 2023-12-20; csa 在线出版时间: 2024-03-22

the proposed model is comparatively analyzed with seven mainstream models. The results indicate that the model can effectively capture the spatiotemporal dependencies between air monitoring stations, improving predictive accuracy.

Key words: air quality; $\text{PM}_{2.5}$; deep learning; graph convolution; spatiotemporal dependence

在工业化和城市化的背景下,空气污染已成为人们关注的焦点.其中, $\text{PM}_{2.5}$ (指空气动力学直径小于或等于 2.5 μm 的大气颗粒物)是一种特别危险的因素,可以直接进入人体并对健康造成危害^[1].许多研究表明,高水平的 $\text{PM}_{2.5}$ 暴露与呼吸系统和心血管疾病密切相关^[2,3].此外, $\text{PM}_{2.5}$ 还对出生率和婴儿健康产生不良影响^[4,5].然而,控制雾霾需要更多实时的空气质量监测数据.为了有效预防和控制空气污染,并进行城市管理规划,准确高效地预测近期 $\text{PM}_{2.5}$ 浓度是必要的.

由于 $\text{PM}_{2.5}$ 浓度预测的重要性,近年来在该领域取得了显著进展^[6].主要的发展可以分为 3 个不同的阶段:(1)早期的研究将数据挖掘方法与统计分析相结合^[7,8].Zhang 等^[9]研究了美国东部天气研究与预报(WRF)/社区多尺度空气质量(CMAQ)系统的性能.然而,由于这些方法研究区域、研究数据的限制,很难去拟合非线性特征,从而导致预测准确性降低^[10].(2)近年来,深度学习的迅速发展,为从输入数据中学习时空依赖关系提供了新的解决方案.研究人员将整个城市视为图像并将其划分网格,利用卷积神经网络(convolutional neural network, CNN)来提取序列内的空间相关性^[11,12].此外,郭宇辰等^[13]利用将循环神经网络(recurrent neural network, RNN)与 CNN 组合来捕捉序列内的时间依赖关系,取得了不错的效果.然而,基于 CNN 的方法对空间上邻居的高效聚合使其对远距离依赖关系不敏感,并且仅适用于欧几里德空间结构^[14].Lin 等^[15]基于门控递归单元(gated recurrent unit, GRU)及其变体,设计了 MLEGRU(multiple linear regression based GRU)对多站点数据进行预测.但基于循环神经网络或其变体忽视了数据在空间上的依赖,容易导致预测精度受影响.(3)图卷积网络(graph convolutional network, GCN)^[16]是 CNN 在图结构上的推广,适用于处理非欧几里德空间数据.由于空气质量监测站的拓扑结构,GCN 已在 $\text{PM}_{2.5}$ 浓度预测领域取得了成功且广泛的应用^[17,18].有研究者将 GCN 用于提取 $\text{PM}_{2.5}$ 浓度预测中的空间依赖关系^[19].此外,还尝试了将 GCN 与 Bi-LSTM 结合

以捕捉时空依赖关系^[20].Wang 等^[21]将多图注意力图卷积与 GRU 结合,将不同图之间相关站点的时空特征聚合,有效提高了预测精度.

尽管 GCN 在 $\text{PM}_{2.5}$ 预测中的有效性已得到验证,但仍然存在 4 个重要问题尚未得到充分讨论:(1)邻接矩阵决定了图卷积网络中的聚合方法,但通常是固定的,并通过基于空间距离或图连接性的启发式方法生成.这些方法无法捕捉真实的时空依赖.(2)现有方法忽视了空气质量预测中的层次依赖性.空气质量受全球、区域和局部尺度上多种因素的影响.这些因素以层次方式相互作用,意味着一个层次上的预测或输入依赖于更高层次的预测.例如,周边环境状况可能会受到意外事件(如工厂污染泄漏)的影响,但这种干扰未必会对整个城市的空气质量产生影响.(3)当前的图卷积方法主要遵循图信号处理的观点,倾向于平滑节点的输入信号.在这种情况下,只使用一个邻接矩阵的可叠加图卷积层难以有效获得多层次的数据高级表示.(4)不同层次表示对最终空气质量的贡献不应是静态的,而应在时间上动态变化.例如,特定的天气异常(如雾霾天气)可能会增加低层次特征的影响.这些问题可以用来改进当前的空气质量预测研究.

为了解决问题,本文提出了一种新的深度学习框架,自适应分层图卷积神经网络(adaptive hierarchical graph convolutional neural network, AHGCNN)用于 $\text{PM}_{2.5}$ 预测.具体而言,为了解决主流图卷积网络在捕获多层次空间依赖方面的低效率和不准确性,引入了自适应分层图卷积(hierarchical mapping graph convolution, HMGC).通过自学习邻接矩阵,该网络可以在每个层次上动态提取空间依赖.此外,对层次拓扑结构相关性进行建模,提供了一种基于注意力的层次聚合机制,以较小的计算成本实现了图卷积结构.然后将 GRU 中的线性单元替换成 HMGC,构成自适应分层图卷积门控循环单元(adaptive hierarchical graph convolution gated recurrent unit, AHGC-GRU),多层堆叠搭建网络主体结构,输出预测结果.

1 理论基础

1.1 PM_{2.5} 浓度预测问题定义

通常情况下, PM_{2.5} 浓度预测被视为一个时空序列预测问题. 给定 N 个监测站的历史空气数据记录, 任务是预测每个监测站未来的 PM_{2.5} 浓度情况. 基于之前的研究, 本文将 N 个监测站及其两两连接定义为一个加权有向图 $G = (V, E, A)$, 其中 V 表示节点集合, $|V| = E$, E 表示边的集合, $A \in \mathbb{R}^{N \times N}$ 表示加权邻接矩阵, 描述了任意节点对之间的关系紧密程度. 在任意时间 t , G 的历史数据被编码为图信号 $X_t \in \mathbb{R}^{N \times C}$, 其中 C 表示每个节点属性的维度. PM_{2.5} 预测问题的核心在于学习一个函数 f , 该函数能够利用 S 个历史图信号 $x = [X_{(t-S+1)}, \dots,$

$X_t] \in \mathbb{R}^{S \times N \times C}$ 以及监测站网络的邻接矩阵 $A \in \mathbb{R}^{N \times N}$, 来预测 T 个未来的信号 $\hat{y} = [\hat{X}_{(t+1)}, \dots, \hat{X}_{(t+T)}] \in \mathbb{R}^{(T \times N \times C)}$. 数学上可以表述为:

$$\hat{y} = f(x, A) \quad (1)$$

其中, $\hat{y}$ 应尽可能接近 y .

2 模型设计

本节描述 AHGCNN 模型的基本组件. 如图 1 所示, 包含邻接矩阵生成、自适应动态图的构造、多层信息融合、AHGC-GRU 多层堆叠形成的编码器和解码器. 下面详细介绍 AHGCNN 模型各个模块的具体实现.

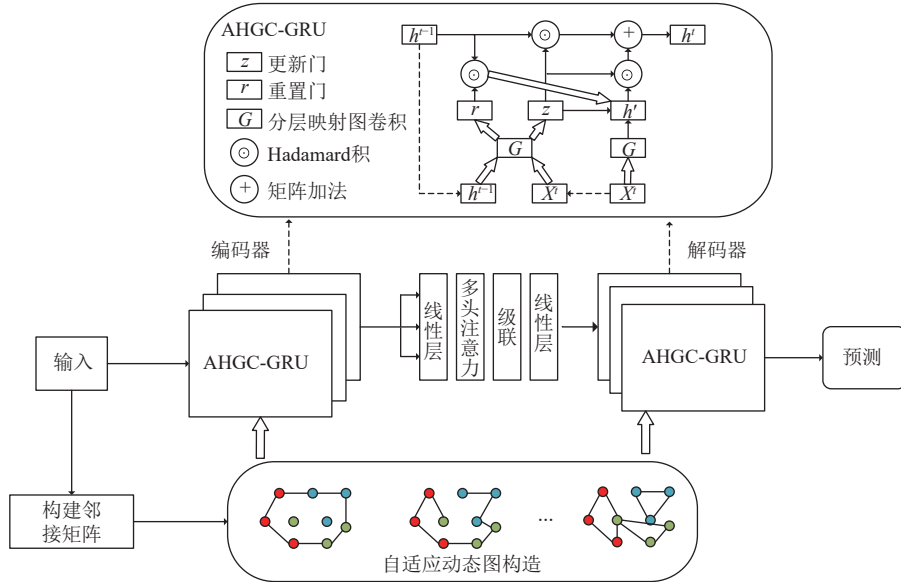


图 1 AHGCNN 模型架构

2.1 构建邻接矩阵

邻接矩阵决定了在图卷积中节点及其邻居是如何被聚合的. 现有的研究依赖先验知识来事先构建图结构, 本文的方法是完全基于数据驱动的, 因此能够捕捉数据之间的时序依赖性.

对于给定的图信号集合 $X_{t_a:t_a+\tau-1} \in \mathbb{R}^{\tau \times N \times d}$, 将其转换为一个维度为 $(\tau \times d) \times N$ 的二维矩阵 X^l . 为了有效地捕获不同站点之间的内部相似性, 并过滤掉站点之间的冗余信息. 本文使用奇异值分解 (singular value decomposition, SVD) 将 X^l 分解为两部分. 即:

$$SVD(X^l) \rightarrow (X^s, X^t) \quad (2)$$

其中, X^s 和 X^t 分别表示按时间维度和站点维度划分的矩阵.

矩阵 X^l 包含大量的冗余信息. 通过使用 SVD 进行降维, 可以过滤掉这些冗余信息, 从而揭示出隐藏的模式. 矩阵 $X^s \in \mathbb{R}^{N \times \gamma}$ 包含了每个站点特征的高级表示, 其中 γ 表示特征的维度. 本文采用基于高斯核的方法来计算矩阵 X^s 中第 x 行和第 y 列之间的相似性, 并将其用作生成的邻接矩阵中的边权重. 公式定义如下:

$$A_{xy} = \text{Similarity}(X_x^s, X_y^s) = \exp\left(-\frac{\|X_x^s - X_y^s\|^2}{2\beta^2}\right) \quad (3)$$

其中, A_{xy} 指的是邻接矩阵 A 的第 x 行第 y 列的元素, β 是标准偏差.

2.2 分层映射图卷积

为了高效准确地捕捉多层空间依赖关系, 本文引入了一种新的图卷积网络: 分层映射图卷积网络 (HMGC).

HMGC 的特点在于每个层级都包含独特的邻接矩阵,从而更灵活地建模底层图结构.这种结构可以进行递归定义:

$$Z^{(m+1)} = Z^{(m)} \star_G g_\theta^{(m)} = A^{(m)} Z^{(m)} \theta^{(m)} \quad (4)$$

$Z^{(m)}$ 代表 $m+1$ 层的输入,而 $Z^{(m+1)}$ 不仅是第 $m+1$ 层的输出,还是第 $m+2$ 层的输入.参数 $g_\theta^{(m)}$ 和 $\theta^{(m)}$ 表示第 m 层的图卷积滤波器参数, $\star_G$ 表示图卷积操作.

节点之间的层级关系由 $A^{(m)}$ 建模, $A^{(m)}$ 的值取决于所处的层级.利用分层映射函数 φ 构建第 $m+1$ 层的上层邻接矩阵:

$$A^{(m+1)} = \varphi^{(m)}(A^{(m)}) \quad (5)$$

对由式 (3) 生成的初始图结构进行标准化处理,将其用于邻接矩阵的初始化,并采用随机梯度下降进行优化. HMGC 初始层的定义如下:

$$A^{(0)} = D^{-1} A_{xy} \quad (6)$$

$$Z^{(1)} = A^{(0)} Z^{(0)} \theta^{(0)} \quad (7)$$

其中, D 代表节点度的对角矩阵. $Z^{(0)} = X$, 在每个时间步中,特征矩阵 $X \in R^{N \times d}$ 作为 HMGC 初始层的输入.

2.3 自适应动态图构建

在现实中,图结构并不总是可用的,节点之间的关联性经常随着时间变化.此外, $PM_{2.5}$ 浓度与天气条件密切相关.在类似的天气条件下,空气质量往往比不同的天气条件下更加均匀.因此,本文遵循模型训练过程来构建不同层之间的动态邻接矩阵.天气数据的时间变化性使得动态邻接矩阵具有潜在的时间相关性.

首先,节点的数量导致更新 $N \times N$ 的邻接矩阵和映射函数 φ 在计算方面的代价很高,从而在训练过程中导致过度参数化.为了解决这个问题,本文采用 SVD 将 $A^{(0)} \in R^{N \times N}$ 分解为两个小矩阵:

$$SVD(A^{(0)}) \rightarrow (E_1^{(0)}, E_2^{(0)}) \quad (8)$$

$E_1^{(0)} \in R^{N \times L}$ 表示首层的源节点嵌入,而 $E_2^{(0)} \in R^{N \times L}$ 则代表目标节点嵌入. L 表示维度.值得注意的是,可训练的参数数量从 $N \times N$ 减少到了 $2 \times N \times L$ ($N \gg L$).

函数 φ 通过全连接层进行实现,以模拟层间相关性.全连接层的参数在 E_1 和 E_2 之间共享.

$$\begin{cases} E_1^{(m)} = E_1^{(m-1)} W^{(m-1)} + b^{(m-1)} \\ E_2^{(m)} = E_2^{(m-1)} W^{(m-1)} + b^{(m-1)} \end{cases} \quad (9)$$

在这种情况下, $E_1^{(m)}$ 和 $E_2^{(m)}$ 分别代表第 m 层中的源

节点嵌入和目标节点嵌入.符号 $W^{(m-1)}$ 和 $b^{(m-1)}$ 分别表示权重矩阵和偏置向量.为了简化模型,不同层次的 E_1 和 E_2 的特征维度被设定为 L ,从而将每个映射的参数数量减少到 $L \times (L+1)$.

基于式 (4) 和式 (9),对于分层映射图卷积 (HMGC),其数学表达式可以重新定义为:

$$Z^{(m+1)} = (E_1^{(m)} E_2^{(m)T}) Z^{(m)} \theta^{(m)} \quad (10)$$

2.4 基于注意力机制的分层信息融合

为了从所有图卷积层中汇集信息,而不仅仅从一个固定的层中提取信息,本文通过注意力机制实现多层聚合策略,以选择性地选取对当前预测任务准确性产生更大影响的信息.

通过 HMGC 获得的图信号的多层输出表示为 $Z = \{Z^{(1)}, Z^{(2)}, \dots, Z^{(M)}\}$, $Z \in R^{(M \times N \times \tau)}$, 其中 M 表示图卷积层的总数, τ 表示特征维度.本文使用多头注意力来调节各层间的权重比以此获得最终的特征向量 h :

$$\begin{cases} Q_i = W_i^Q F(\widehat{Z}) \\ K_i = W_i^K F(\widehat{Z}) \\ V_i = W_i^V F(\widehat{Z}) \\ head_i = attention((K_i, V_i), Q_i) \\ h = Concat(head_1, \dots, head_i) W_0 \end{cases} \quad (11)$$

其中, F 表示线性转换, $\widehat{Z}$ 代表 Z 的扁平化输出, W_i^Q, W_i^K, W_i^V 分别表示映射 Q, K, V 使用的 3 个线性层的权重, W_0 表示注意力模块最后一个线性层的权重.

2.5 时间依赖建模

门控循环单元 (GRU) 是循环神经网络 (RNN) 的一种变体,相对简单而又强大,它解决了 RNN 中梯度消失和梯度爆炸的问题.根据之前研究^[22],本文将 GRU 中的线性变换替换为 HMGC,以获得更好的性能.因此,提出了自适应分层图卷积门控循环单元 (AHGC-GRU).

$$\begin{cases} r^{(t)} = \sigma(\Theta_r \star_G [h^{(t)}, H^{(t-1)}] + b_r) \\ u^{(t)} = \sigma(\Theta_u \star_G [h^{(t)}, H^{(t-1)}] + b_u) \\ c^{(t)} = \tanh(\Theta_c \star_G [h^{(t)}, (r^{(t)} \odot H^{(t-1)})] + b_c) \\ H^{(t)} = u^{(t)} \odot H^{(t-1)} + (1 - u^{(t)}) \odot c^{(t)} \end{cases} \quad (12)$$

$h^{(t)}$ 和 $H^{(t)}$ 分别代表时间步 t 处的注意力分数和门控循环单元 (GRU) 的输出,对两者进行 Hadamard 乘积 $\odot$ 运算. σ 表示 Sigmoid 激活函数, $\tanh$ 表示 $\tanh$ 激活函数,重置门 $r^{(t)}$ 帮助网络遗忘不相关信息.同时,更新门 $u^{(t)}$ 控制着给定时间步的 GRU 输出.符号 $\star_G$ 表示

图卷积操作, 而 $\Theta_r, \Theta_u, \Theta_c$ 则表示相应滤波器的参数.

在多步预测模型中, 编码器和解码器均由 AHGC-GRU 构成, AHGCNN 模型的伪代码如算法 1.

算法 1. AHGCNN 模型框架

输入: 图 $G=(V,E,A)$; 图卷积深度 m .

输出: 表征向量 Z .

```

1.  $A^{(0)} \leftarrow G(V,E,A)$ 
2. for  $m=1, \dots, m$  do
3.    $Z^{(m)}, E_1^{(m)}, E_2^{(m)} \leftarrow \text{AHGCGRU}(Z^{(m-1)}, G, A^{(m)})$ 
4.    $A^{(m+1)} \leftarrow \text{ReLU}(E_1^{(m)} \cdot E_2^{(m)T})$ 
5.    $Z \leftarrow \{Z^{(1)}, \dots, Z^{(m)}\}$ 
6.    $h \leftarrow \text{mulatt}(Z)$ 
7. for  $m=1, \dots, m$  do
8.    $Z^{(m)}, E_1^{(m)}, E_2^{(m)} \leftarrow \text{AHGCGRU}(h, G, A^{(m)})$ 
9.    $A^{(m+1)} \leftarrow \text{ReLU}(E_1^{(m)} \cdot E_2^{(m)T})$ 
10.   $Z \leftarrow \{Z^{(1)}, \dots, Z^{(m)}\}$ 

```

3 实验设计与验证

3.1 数据集描述

为了评估模型捕捉时空依赖性的能力, 本文选择了位于中国北京市及周边地区的 39 个气象监测站作为研究对象. 通过考虑从这些监测站收集到的空气质量、气象数据的时空特性, 可以获得精确的预测结果. 这些预测结果可以显著地惠及易受影响的人群, 并有助于制定适当的空气污染控制政策.

依据之前的工作方式^[23], 本文使用了两种类型的数据: 空气污染物浓度数据和气象数据. 空气污染物浓度数据来源于北京市环境保护监测中心网站 (<http://www.bjmemc.com.cn/>), 其数据类型包括 $\text{PM}_{2.5}$ 、 PM_{10} 、 SO_2 、 NO_2 、 O_3 、 CO 和 AQI . 气象数据来自欧洲中期气象预报中心 (ECMWF) (<https://cds.climate.copernicus.eu/>), 其中包含地面气压、气温、露点温度、风向和风速等参数.

从 39 个站点收集了为期 3 年的数据 (2018 年 1 月 1 日–2020 年 12 月 31 日), 数据每小时采集一次. 鉴于数据的不完整性 (例如缺失值), 本文采用线性插值技术填补短期的缺失值. 最终经过处理的数据根据 7:1:2 的比例被划分为训练集、验证集和测试集.

3.2 评价指标

本文采用两种评估指标来评估所提出模型在预测结果方面的性能, 包括均方根误差 ($RMSE$) 和平均绝对误差 (MAE). 其定义如下:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

其中, 变量 y 和 $\hat{y}$ 分别表示实际和预测的 $\text{PM}_{2.5}$ 值, 而 n 表示测试数据的总数量. 较小的 MAE 和 $RMSE$ 值意味着更好的最终预测结果. 值得注意的是, 在均方根误差 ($RMSE$) 的情况下, 每个误差都会被平方, 这意味着各个误差会呈二次增长, 并对最终 $RMSE$ 值产生不同的影响. 因此, 通常会同时使用 MAE 和 $RMSE$ 来综合评估所采用方法的性能.

3.3 基线模型

为了测试 AHGCNN 模型的预测效果, 本文与常见的深度学习网络模型进行了比较, 包括 LSTM^[24]、GRU^[25]、RNN^[26]、Informer^[27]、FFTransformer^[28]; 同时为了说明图结构对于捕捉空气数据时空依赖性的重要性, 与扩散递归神经网络 (diffusion convolutional recurrent neural network, DCRNN)^[29] 和 Graph WaveNet^[30] 等模型做实验对比.

3.4 实验参数

本次实验模型使用 PyTorch 1.13.1 实现, 并在单块 GPU (NVIDIA RTX 3090, 24 GB 显存) 上进行训练, 操作系统为 Ubuntu 22.04. 当预测模型各项参数设置如表 1 所示时, 模型的预测性能达到最佳.

表 1 最优参数组合

参数	数值
历史数据长度	24
预测长度	24
自适应矩阵维度	13
学习率	0.0005
批次大小	64
优化策略	Adam
图卷积层数	3
注意力头数	3

模型使用 MSE 作为损失函数, 损失值的训练迭代情况如图 2 所示. 从图 2 中可以观察到, 在前 10 轮迭代中, 训练集和测试集的损失值迅速下降. 随着迭代次数的增加, 训练集的损失值持续缓慢下降. 在经过 15 轮迭代后, 模型基本完成拟合, 损失值趋于平稳, 训练精度达到 0.1 左右. 随后的 20 轮迭代中, 测试集的损失值基本稳定, 保持在 0.4 左右.

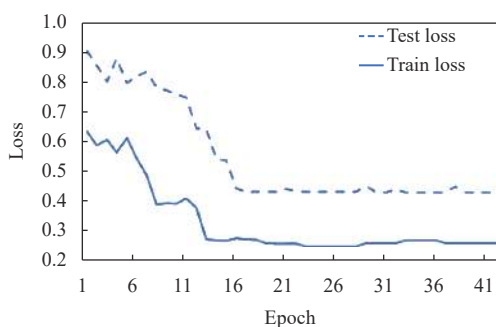


图2 模型 Loss 变化曲线

3.5 预测结果对比分析

如表2所示, RNN、LSTM和GRU模型表现最差。这并不令人意外, 因为它们只能访问节点表示, 无法获取邻居信息。说明在 $PM_{2.5}$ 预测中捕获邻居信息的必要性。

AHGCNN模型在性能上优于现有的预测模型。在短期预测方面, AHGCNN的改进效果更为显著。与DCRNN相比, 在预测的1 h、3 h、6 h、12 h和24 h

时间点, AHGCNN的MAE分别降低了5.56%、6.12%、5.91%、13.5%和3.18%, RMSE分别降低了9.08%、5.38%、3.96%、8.49%和0.47%。与Graph WaveNet相比, AHGCNN在预测的1 h、3 h、6 h、12 h和24 h时间点, 分别将MAE降低了1.9%、0.91%、1.59%、0.87%和0.52%, RMSE分别降低了1.13%、4.31%、4.09%、3.87%和3.35%。这表明动态邻接矩阵能够有效地捕获随时间变化的时空依赖关系, 即使底层的图结构是未知的。与长序列预测模型Informer相比, 在预测的1 h、3 h、6 h、12 h和24 h时间点, AHGCNN的MAE分别降低了10.95%、12.11%、9.76%、7.97%和3.73%, RMSE分别降低了17.65%、9.46%、6.26%、3.25%和0.47%。与FFTransformer相比, AHGCNN在预测的1 h、3 h、6 h、12 h和24 h时间点, 分别将MAE降低了11.51%、9.77%、13.24%、8.07%和3.16%, RMSE分别降低了4.9%、15.70%、10.43%、5.86%和2.39%。这个结果证实了AHGCNN在利用时空依赖知识方面的优势, 并展示了它可靠的预测能力。

表2 我们的模型与基线预测性能比较

模型	1 h		3 h		6 h		12 h		24 h		Average	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	7.39	12.34	12.69	20.77	17.32	28.14	23.21	34.44	26.25	37.63	20.97	33.13
GRU	7.25	12.27	12.57	20.49	17.27	27.56	22.49	34.27	26.11	37.48	20.67	32.81
RNN	7.49	12.51	13.36	21.17	17.86	28.69	23.63	35.42	27.68	38.62	25.68	37.62
Graph WaveNet	5.80	9.72	9.60	15.99	13.24	21.22	17.24	26.56	20.97	30.74	18.37	27.77
DCRNN	6.45	10.57	10.57	16.17	14.06	21.19	19.88	27.90	21.96	29.85	17.45	28.18
Informer	6.39	11.67	11.67	16.90	14.44	21.71	18.57	26.39	21.67	29.85	16.47	25.27
FFTransformer	6.43	10.11	10.11	18.15	15.02	22.72	18.59	27.12	21.54	30.44	16.78	25.27
AHGCNN	5.69	9.61	9.61	15.30	13.03	20.35	17.09	25.53	20.86	29.71	16.42	24.52

图3展示了在多个监测站点中, AHGCNN、RNN、LSTM、GRU、DCRNN、Graph WaveNet、Informer和FFTransformer等多个模型之间的预测值与实际值之间的差异。测试期间, RNN、LSTM和GRU模型在预测值与实际值之间存在显著差异, 尤其是在空气质量数据发生明显波动的时期。尽管DCRNN、Graph WaveNet、Informer和FFTransformer模型在稳定期间的性能有所提高, 但在突然变化时仍存在变化和与实际值的大偏差。相反, AHGCNN模型表现出卓越的预测能力, 在各个时期表现出色。这些发现说明其他模型无法充分利用监测站之间的非线性空间特性, 导致在数据剧烈变化时模型适应性不足。然而, 通过多层信息的聚合, AHGCNN模型可以更有效地提取信息, 捕捉站点之间的异质空间和时间特征, 从而在空气质量数

据突然变化时避免了其他模型中出现的显著预测误差。

我们对模型学习到的自适应邻接矩阵进行了分析和比较, 如图4所示。与由数据生成的原始邻接矩阵相比, 学习到的邻接矩阵展现出更加丰富的站点之间的相关性, 为之前的邻接矩阵提供了补充。这表明单个矩阵无法完全捕捉到随着时间变化而节点之间联系的变化。

4 结语

本文提出了一种基于自适应分层图卷积神经网络(AHGCNN)的 $PM_{2.5}$ 浓度预测方法。该模型能够在不需要预先确定图结构的情况下, 捕捉数据之间的隐式空间依赖关系。具体而言, 为了捕捉多层次的空间依赖关系, 引入HMGC的图卷积结构。HMGC中的邻接矩

阵是自学习的,并且每一层都是独特的,有助于处理随时间变化的潜在时空依赖关系.此外,本文采用分层映射机制,将上层图结构与下层图结构连接起来.这使得模型能够融合不同数据层次的信息,同时减少模型参数的数量.随后,单元网络将这些组件综合起来形成最

终的预测.在真实世界数据集上进行的实验验证了AHGCNN在 $PM_{2.5}$ 浓度预测任务中取得了良好的性能.在未来的工作中,将探索将更多外部知识纳入,以构建动态邻接矩阵,并引入更有效的注意机制来处理多层次信息.

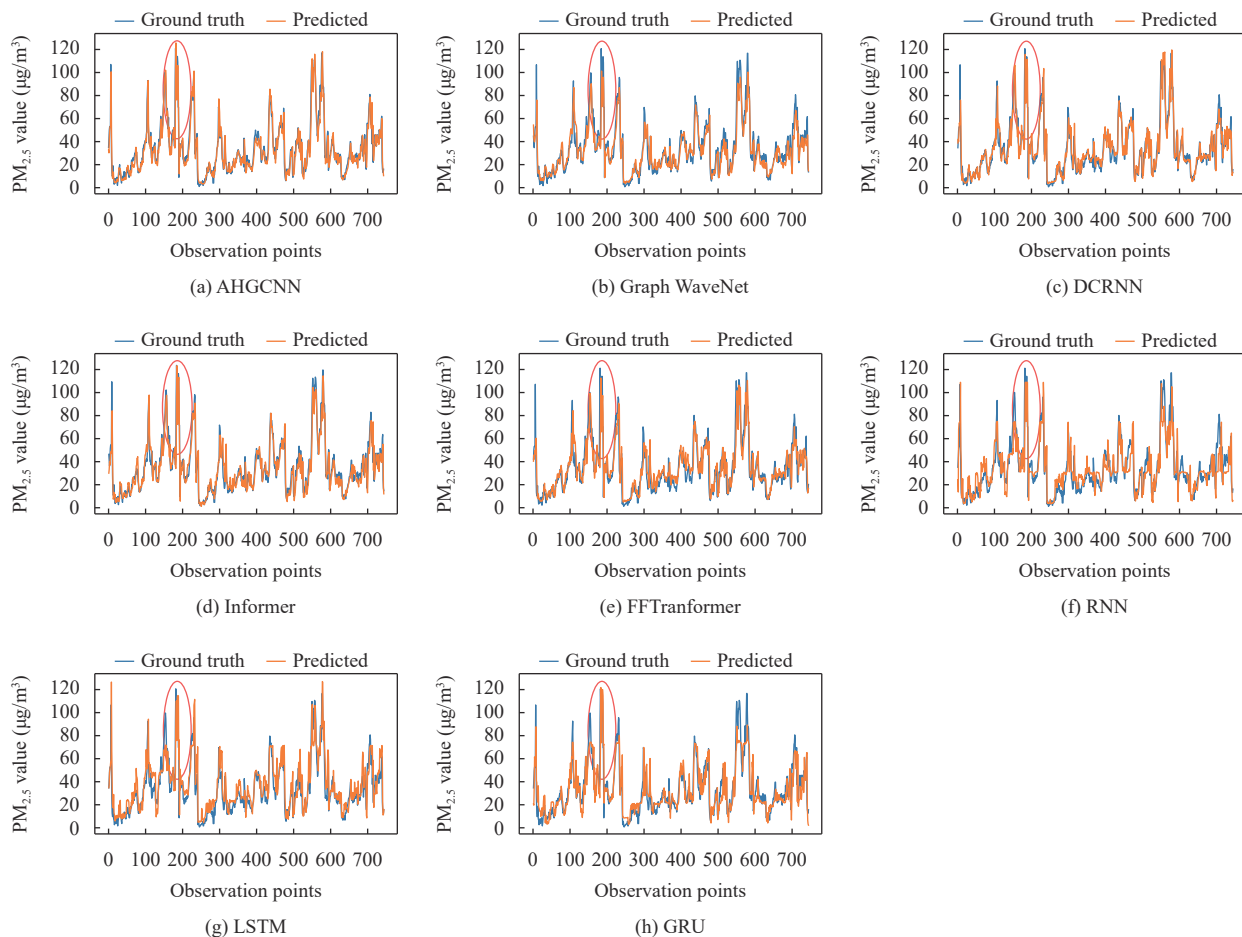
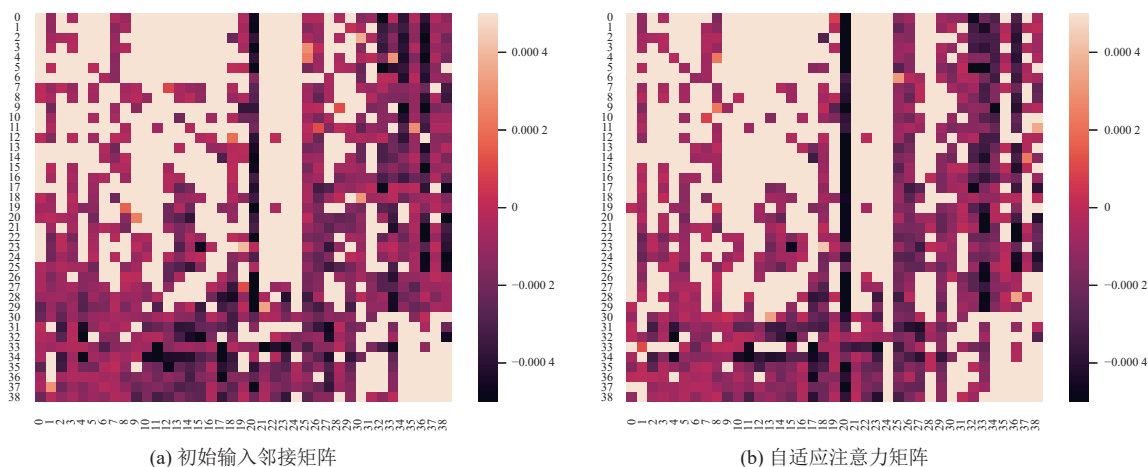


图3 站点1上我们的模型与基线模型一个月内预测值与真实值对比(折线图)



(a) 初始输入邻接矩阵

(b) 自适应注意力矩阵

图4 数据集中39个节点的初始输入邻接矩阵和学习到的自适应注意力矩阵比较

参考文献

- 1 孟聪申, 刘静怡, 刘悦, 等. 2018 年中国五城市大气细颗粒物暴露所致人群超额死亡风险评估. 卫生研究, 2021, 50(4): 593–599. [doi: [10.19813/j.cnki.weishengyanjiu.2021.04.010](https://doi.org/10.19813/j.cnki.weishengyanjiu.2021.04.010)]
- 2 Leclercq B, Kluza J, Antherieu S, *et al.* Air pollution-derived PM_{2.5} impairs mitochondrial function in healthy and chronic obstructive pulmonary diseased human bronchial epithelial cells. *Environmental Pollution*, 2018, 243: 1434–1449. [doi: [10.1016/j.envpol.2018.09.062](https://doi.org/10.1016/j.envpol.2018.09.062)]
- 3 Wright N, Newell K, Chan KH, *et al.* Long-term ambient air pollution exposure and cardio-respiratory disease in China: Findings from a prospective cohort study. *Environmental Health*, 2023, 22(1): 30. [doi: [10.1186/s12940-023-00978-9](https://doi.org/10.1186/s12940-023-00978-9)]
- 4 Panunzi S, Marchetti P, Stafoggia M, *et al.* Residential exposure to air pollution and adverse respiratory and allergic outcomes in children and adolescents living in a chipboard industrial area of Northern Italy. *Science of the Total Environment*, 2023, 864: 161070. [doi: [10.1016/j.scitotenv.2022.161070](https://doi.org/10.1016/j.scitotenv.2022.161070)]
- 5 Yang WH, Johnson MB, Liao HS, *et al.* Combined effect of preconceptional and prenatal exposure to air pollution and temperature on childhood pneumonia: A case-control study. *Environmental Research*, 2023, 216: 114806. [doi: [10.1016/j.envres.2022.114806](https://doi.org/10.1016/j.envres.2022.114806)]
- 6 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. [doi: [10.15888/j.cnki.csa.008847](https://doi.org/10.15888/j.cnki.csa.008847)]
- 7 Hasnain A, Sheng YH, Hashmi MZ, *et al.* Time series analysis and forecasting of air pollutants based on prophet forecasting model in Jiangsu province, China. *Frontiers in Environmental Science*, 2022, 10: 945628. [doi: [10.3389/fenvs.2022.945628](https://doi.org/10.3389/fenvs.2022.945628)]
- 8 杨锦伟, 孙宝磊. 基于灰色马尔科夫模型的平顶山市空气污染物浓度预测. 数学的实践与认识, 2014, 44(2): 64–70. [doi: [10.3969/j.issn.1000-0984.2014.02.008](https://doi.org/10.3969/j.issn.1000-0984.2014.02.008)]
- 9 Zhang HL, Chen G, Hu JL, *et al.* Evaluation of a seven-year air quality simulation using the weather research and forecasting (WRF)/community multiscale air quality (CMAQ) models in the eastern United States. *Science of the Total Environment*, 2014, 473–474: 275–285.
- 10 Gao X, Li WD. A graph-based LSTM model for PM_{2.5} forecasting. *Atmospheric Pollution Research*, 2021, 12(9): 101150. [doi: [10.1016/j.apr.2021.101150](https://doi.org/10.1016/j.apr.2021.101150)]
- 11 Hou JW, Wang YJ, Hou B, *et al.* Spatial simulation and prediction of air temperature based on CNN-LSTM. *Applied Artificial Intelligence*, 2023, 37(1): 2166235. [doi: [10.1080/08839514.2023.2166235](https://doi.org/10.1080/08839514.2023.2166235)]
- 12 张波, 陆云杰, 秦东明, 等. 一种卷积自编码深度学习的空气污染多站点联合预测模型. 电子学报, 2022, 50(6): 1410–1427.
- 13 郭宇辰, 加鹤萍, 余涛, 等. 基于 CNN-LSTM 组合模型的碳价预测方法. 科技管理研究, 2023, 43(11): 200–206. [doi: [10.3969/j.issn.1000-7695.2023.11.025](https://doi.org/10.3969/j.issn.1000-7695.2023.11.025)]
- 14 Ye JC, Sun LL, Du BW, *et al.* Coupled layer-wise graph convolution for transportation demand prediction. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. AAAI, 2021. 4617–4625.
- 15 Lin CY, Chang YS, Abimannan S. Ensemble multifeatured deep learning models for air quality forecasting. *Atmospheric Pollution Research*, 2021, 12(5): 101045. [doi: [10.1016/j.apr.2021.03.008](https://doi.org/10.1016/j.apr.2021.03.008)]
- 16 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *Proceedings of the 5th International Conference on Learning Representations*. Toulon: ICLR, 2017.
- 17 马俊文, 严京海, 孙瑞雯, 等. 基于 LSTM-GCN 的 PM_{2.5} 浓度预测模型. 中国环境监测, 2022, 38(5): 153–160. [doi: [10.19316/j.issn.1002-6002.2022.05.17](https://doi.org/10.19316/j.issn.1002-6002.2022.05.17)]
- 18 祁柏林, 郭昆鹏, 杨彬, 等. 基于 GCN-LSTM 的空气质量预测. 计算机系统应用, 2021, 30(3): 208–213. [doi: [10.15888/j.cnki.csa.007815](https://doi.org/10.15888/j.cnki.csa.007815)]
- 19 Qi YL, 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. *Science of the Total Environment*, 2019, 664: 1–10. [doi: [10.1016/j.scitotenv.2019.01.333](https://doi.org/10.1016/j.scitotenv.2019.01.333)]
- 20 Kunar P, Nithish R, Sudha. Enhancing PM_{2.5} predictions using combination of graph convolutional network and Bi-LSTM. *Proceedings of the 9th International Conference on Advanced Computing and Communication Systems*. Coimbatore: IEEE, 2023. 1218–1221.
- 21 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.
- 22 Ouyang XC, Yang Y, Zhang YL, *et al.* Spatial-temporal dynamic graph convolution neural network for air quality prediction. *Proceedings of the 2021 International Joint Conference on Neural Networks*. Shenzhen: IEEE, 2021. 1–8.

- 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 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)]
- 25 Chung J, Gulcehre C, Cho K, *et al.* Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555, 2014.
- 26 Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization. arXiv:1409.2329, 2015.
- 27 Zhou HY, Zhang SH, Peng JQ, *et al.* Informer: Beyond efficient Transformer for long sequence time-series forecasting. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. AAAI, 2021. 11106–11115.
- 28 Bentsen LØ, Warakagoda ND, Stenbro R, *et al.* Spatio-temporal wind speed forecasting using graph networks and novel Transformer architectures. *Applied Energy*, 2023, 333: 120565. [doi: [10.1016/j.apenergy.2022.120565](https://doi.org/10.1016/j.apenergy.2022.120565)]
- 29 Li YG, Yu R, Shahabi C, *et al.* Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *Proceedings of the 6th International Conference on Learning Representations*. Vancouver: OpenReview.net, 2018.
- 30 Wu ZH, Pan SR, Long GD, *et al.* Graph WaveNet for deep spatial-temporal graph modeling. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao: IJCAI.org, 2019. 1907–1913.

(校对责编: 孙君艳)