

项目编号: S202410559009

暨南大学大学生创新创业训练计划项目

结题验收申请表

项目级别 省级

项目名称 基于 GAT-TSDFormer 的 PM2.5 时空预测

——基于轻量化与解释性视角

项目类型 创新训练

项目负责人 刘威腾

学院专业 经济学院经济统计学

联系电话 15217271378

电子邮件 liuwt4408@126.com

指导教师 谢贤芬, 王国长

指导教师学院 经济学院

填表日期 2025 年 04 月 08 日

暨南大学本科学院教务处

2025 年 3 月

填表说明

一、《暨南大学大学生创新创业训练计划项目结题验收申请表》由项目负责人根据项目实施情况实事求是填写，表达明确严谨。

二、填写本表内容统一使用宋体，五号字，单倍行距。表格各栏目可根据内容进行调整，要确保表格的样式没有被改变；填写完成后使用 A4 纸双面打印，于左侧装订成册。

三、封面项目级别选填国家级、省级、校级三者之一；项目类型选填创新训练、创业训练、创业实践三者之一。

四、项目负责人填写完本表后，必须请指导教师审核，并填写意见和签名。

五、填报时请参考《暨南大学大学生创新创业训练计划项目管理办法》（暨教〔2025〕15号）及其他相关文件要求。

六、依托大创项目发表的论文需在首页或致谢部分严格按规定进行标注，否则不予认定结题成果。标注要求如下：

（一）中文文章：国家级/广东省/暨南大学大学生创新创业训练计划支持项目（项目编号：）

（二）英文文章：

1. 国家级项目：National Innovation and Entrepreneurship Training Program For Undergraduate(NO.)

2. 省级项目：Innovation and Entrepreneurship Training Program For Undergraduate of Guangdong Province(NO.)

3. 校级项目：Innovation and Entrepreneurship Training Program For Undergraduate of Jinan University(NO.)

项目名称		基于 GAT-TSDFormer 的 PM2.5 时空预测——基于轻量化与解释性视角					
项目经费总额		10000		实际起止年月		2024 年 5 月至 2025 年 5 月	
负责人	姓 名	刘威腾	学 号	2021103160		生源情况	✓内招 □外招
	学院专业	经济学院经济通继续				联系电话	15217271378
	项目中承担的工作	负责统筹队伍工作，GAT 模块编写				E-mail	liuwt4408@126.com
项目组成员	姓 名	学 号	学院及专业	生源情况		分 工	
	王国杰	2022102173	信息科学技术学院软件工程	✓内招 □外招		TSDFormer 模块编写	
	招铸峰	2022102815	经济学院经济统计	✓内招 □外招		Baseline 模型代码	
	杨添欣	2021102002	经济学院金融学	✓内招 □外招		收集文献，查找数据	
	王萌	2021104222	经济学院金融学	✓内招 □外招		查找数据，财务报销	
指导教师(第一导师)	姓 名	谢贤芬	学 院	经济学院		职务/职称	高级实验师
	联系电话	15915747152		E-mail		txiexianfen2009@jnu.edu.cn	
<div>一、项目成果简介（重点介绍特色及创新点，200 字左右）</div> <p>本项目针对 PM2.5 预测领域现有研究中特征工程深化、预测模型轻量化、时空发展传播机制可解释性工作上的空白，提出了使用 GAT-TSDFormer 模型作为 PM2.5 预测模型，具体来讲，首先深化了特征工程：将影响 PM2.5 的相关因素划分为垂直扩散和水平传播因素；接着建立了 GAT-TSDFormer 模型：基于 GAT 模型进行空间依赖性建模、基于 TSDFormer 模型进行时间依赖性建模；然后进行消融实验、敏感性实验、基线模型对比实验，对 GAT-TSDFormer 性能进行全面的分析和对比，最终模型 RMSE、MAE、MAPE、CSI、POD、FAR 分别为 37.40、22.25、0.32、0.88、0.97、0.09，各项实验模型表现优异；接着深入揭示了 PM2.5 的时空发展传播规律：通过模型中的 GAT 模块中的注意力机制组件挖掘 PM2.5 空间传播规律，通过 TSDFormer 中的序列成分分解以及各推理算子组件来分析 PM2.5 的时间发展规律；</p>							

二、预期指标及完成情况（预期指标应按照《暨南大学大学生创新创业训练计划项目申报书》内容填写）

1.收集数据集，深化特征工程

本研究使用了 KnowAir 数据集，该数据集涵盖了 2015 年至 2018 年的中国严重污染区域，提供了详细的气象和 PM2.5 数据。数据集的创建过程结合了现场监测和 ERA5 气候再分析数据，确保了数据的完整性和准确性。数据集主要用于空气质量监测和 PM2.5 浓度的时空预测，旨在帮助制定有效的污染控制策略和早期预警系统。

本项目的主要目标是准确预测每小时 PM2.5 的浓度，从前面文献综述可知，存在大量的影响 PM2.5 发展传播的因素，为了提升我们后续模型的预测精度，我们充分挖掘已有的因素信息，深化特征工程，将关于大气环境和污染物的传播信息纳入到输入变量之中。

在本项目中，我们将影响 PM2.5 发展传播的因素分为两类:垂直扩散因素和水平传播因素，其中垂直扩散因素是指 PM2.5 通过复杂的动态或热力学过程在垂直方向上的扩散过程，主要为某城市当地的气象特征状态,在后续模型中我们将它们作为节点属性。通过查阅相关文献,我们整理了 PM2.5 垂直扩散因素如下表所示:

表 1 PM2.5 垂直扩散因素表

相关因素	影响机制
行星边界层高度（PML）	影响 PM2.5 垂直稀释
K 指数	湿空气分层不稳定(K 指数)影响 PM2.5 聚集
950ha 风速	950ha 处风速直接影响 PM2.5 垂直扩散
2m 温度	通过化学效应和冷锋通风影响 PM2.5 浓度
相对湿度	空气水分会促进硝酸铵(PM2.5 主要成分)的形成
降水	通过拖曳效应产生下沉气流，抑制 PM2.5 垂直扩散
大气压强	大气压强较高表面压强意味着稳定大气条件，易导致 PM2.5 累计形成

水平扩散因素是指 PM2.5 通过借助风的动能在水平方向上，从一个城市传播到另外一个城市的传播过程，主要为城市与城市之间的相对位置、风向风速等气象特征状态,在后续模型中我们将它们作为边缘属性。通过查阅相关文献，我们整理了 PM2.5 水平传播因素如下表所示(假设此时研究的是从城市 i 向城市 j 的传播过程):

表 2 PM2.5 水平扩散因素表

相关因素	影响机制
城市 i 地面风速	地面风速越大，越容易传播到其它城市
城市 i,j 直线距离	距离越短，越容易传播到其它城市
城市 i 风向与城市 i 至城市 j 方向差值的绝对值	其差值的绝对值越大，借助风的动能传播到 j 城市可能性越大
平流系数 S	公式(1)(2)

$$S_{j \rightarrow i} = ReLU\left(\frac{|v|}{d} \cos(\alpha)\right)$$

$$\alpha = |\gamma - \beta|$$

首先考虑到 PM2.5 浓度受到多种因素影响，且不同因素对 PM2.5 扩散传播影响机制存在

差异我们依据对垂直扩散因素和水平传播因素的分类，定义有向图的节点属性和边缘属性并根据城市距离和是否有高山阻挡来定义邻接矩阵。

首先定义节点属性 P^i :我们将城市 i 的节点属性 P^i 定义为 PM2.5 垂直扩散因素与城市的经纬度,具体如下表所示:

表 3 节点属性表

变量名称	单位
行星边界层高度（PML）	m
K 指数	K
经线方向风速	m/s
纬线方向风速	m/s
2m 处的气温	K
相对湿度	%
总降水	M
大气压	Pa

接着定义边缘属性 $Q^{j \rightarrow i}$:我们将城市 j 到城市 i 的边缘属性 $Q^{j \rightarrow i}$ 定义为 PM2.5 水平传播因素，其方向与此时气象风向相同(假设此时方向为从城市 i 到城市 j)，具体如下表所示:

表 4 边属性表

变量名	单位
源节点风速	km/h
源节点与宿节点距离	km
源节点风向	度 (°)
源节点与宿节点方向	度 (°)
平流系数	%

最后定义邻接矩阵 A：我们根据以下两个因素来构建邻接矩阵：首先，大多数气溶胶污染物分布在地面以上的一定范围内，因此距离是衡量节点相关性的一个重要因素；其次两个城市沿线的山脉会阻碍污染物的 PM2.5 的水平传播，若山体过高，则会直接阻碍其水平传播，故邻接矩阵 A 定义如下：

$$A_{ij} = H(d_{\theta} - d_{ij}) \cdot H(m_{\theta} - m_{ij})$$

$$d_{ij} = \|\rho_i - \rho_j\|$$

$$m_{ij} = \sup_{\lambda \in (0,1)} \{h(\lambda \rho_i + (1 - \lambda) \rho_j) - \max\{h(\rho_i), h(\rho_j)\}\}$$

在本文中，我们距离阈值设置为 $d_{\theta} = 300 \text{ km}$ ，为高度阈值设置 $m_{\theta} = 1200 \text{ m}$ 。这意味着：只有当这两个城市之间的距离小 300km，并且它们之间的山脉高度小于 1200m 时，PM2.5 才能从一个城市传输到另一个城市。

我们抽取京津冀地区作为试验对象，从月份、城市、年度三个维度，对 PM2.5 的时空分布规律进行分析。见图 1，在京津冀地区中保定、唐山、北京的中位数较高以及箱体的宽度和高度也更高，说明该地的 PM2.5 浓度较高。此三座城市位于京津冀地区的核心且 PM2.5 浓度较高，可能作为核心 PM2.5 来源对周边城市造成影响。见图 2，2 月、4 月、7 月 PM2.5 的偏离异常值

较低，3月、8月、11月、12月的PM2.5的偏离异常值较高，这在月份上具有一定的时空分布规律。由图3可知，不同年份之间，PM2.5浓度在不同城市的分布具有较大的差别。

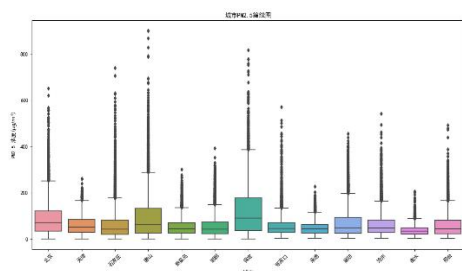


图1 各城市PM2.5箱线图

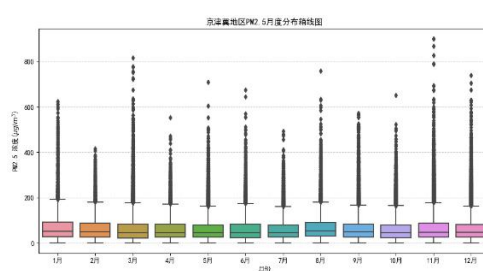


图2 各月份PM2.5箱线图

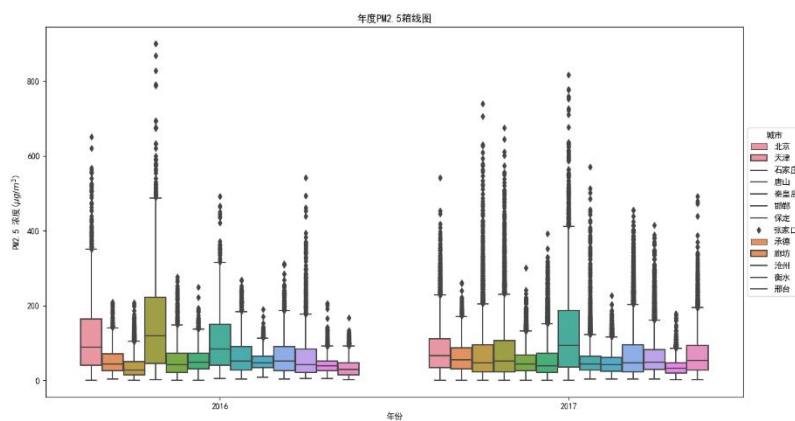


图3 年度PM2.5箱线图

2.构建预测模型，编写模型代码

项目所使用的 GAT-TSDFormer 模型结构如下:

(1) 空间依赖性建模

在空间依赖性的建模中，我们结合知识增强型图神经网络（GNN）和注意力机制，通过利用邻近信息和更新节点的表示来捕获 PM2.5 的空间传播，从而让模型学习 PM2.5 在空间上即不同城市之间的传播效应。具体结构如下图所示：

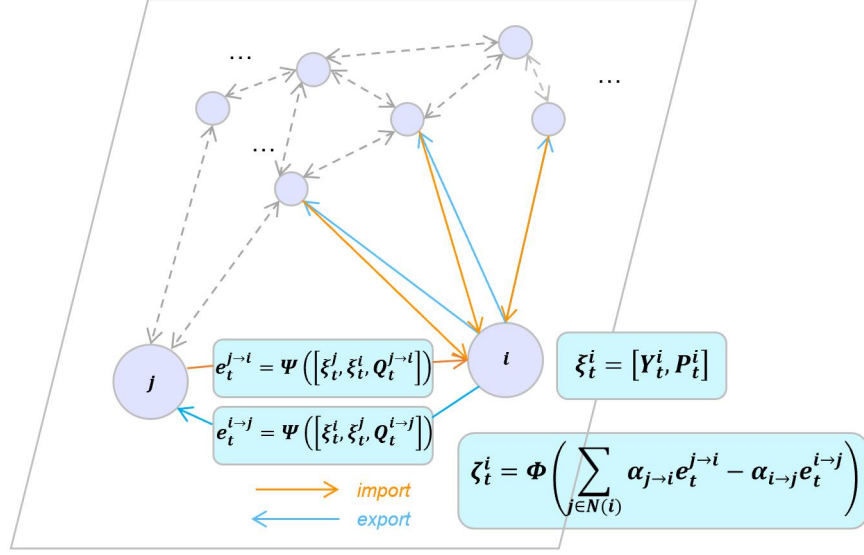


图 4 空间依赖性建模

空间依赖性建模的具体迭代过程的公式如下所示：

$$\xi_t^i = [Y_t^i, P_t^i]$$

$$e_t^{j \rightarrow i} = \Psi([\xi_t^j, \xi_t^i, Q_t^{j \rightarrow i}])$$

$$\zeta_t^i = \Phi\left(\sum_{j \in N(i)} (\alpha_{j \rightarrow i} e_t^{j \rightarrow i} - \alpha_{i \rightarrow j} e_t^{i \rightarrow j})\right)$$

其中， ξ_t^i 为节点表示； $e_t^{j \rightarrow i}$ 为边缘表示，与边缘属性 $Q_t^{j \rightarrow i}$ 同方向； ζ_t^i 为节点的空间相关性，通过汇总所有邻居的影响来计算，具体而言，给定一个节点 i ，来自其邻居节点集合 $\{j: j \in N(i)\}$ 的传输等于权重参数与边缘表示 $\alpha_{j \rightarrow i} e_t^{j \rightarrow i}, \alpha_{i \rightarrow j} e_t^{i \rightarrow j}$ 之间的差异，如图 4 中的橙色线和蓝色线所示。

经过多次递归操作后，图上的每个节点都可获取其他节点的信息实现空间依赖性建模，最终将 ξ_t^i, ζ_t^i 进行连接操作后得到城市 i 在空间依赖性模块的输出：

$$H_t^i = [\xi_t^i, \zeta_t^i]$$

$$H_{t-T_{in}:t}^i = \{H_{t-T_{in}}^i, H_{t-T_{in}+1}^i, \dots, H_t^i\}$$

(2) 时间依赖性建模

在时间依赖性建模方面，我们通过推理组件和外推组件来有效地对输入的时间序列有效分解为趋势成分、季节成分和短期波动成分并对各个部分进行有效外推。图 5 顶部分支（红虚线上部）说明了用于导出各序列成分的推理过程，而底部分支（红虚线下部）则描述了外推过程，用于预测各成分的潜在演变。最终获得的各成分的分量和残差被组合成一个宽向量，最后输入到多项式回归层以捕获它们复杂的相互关系。

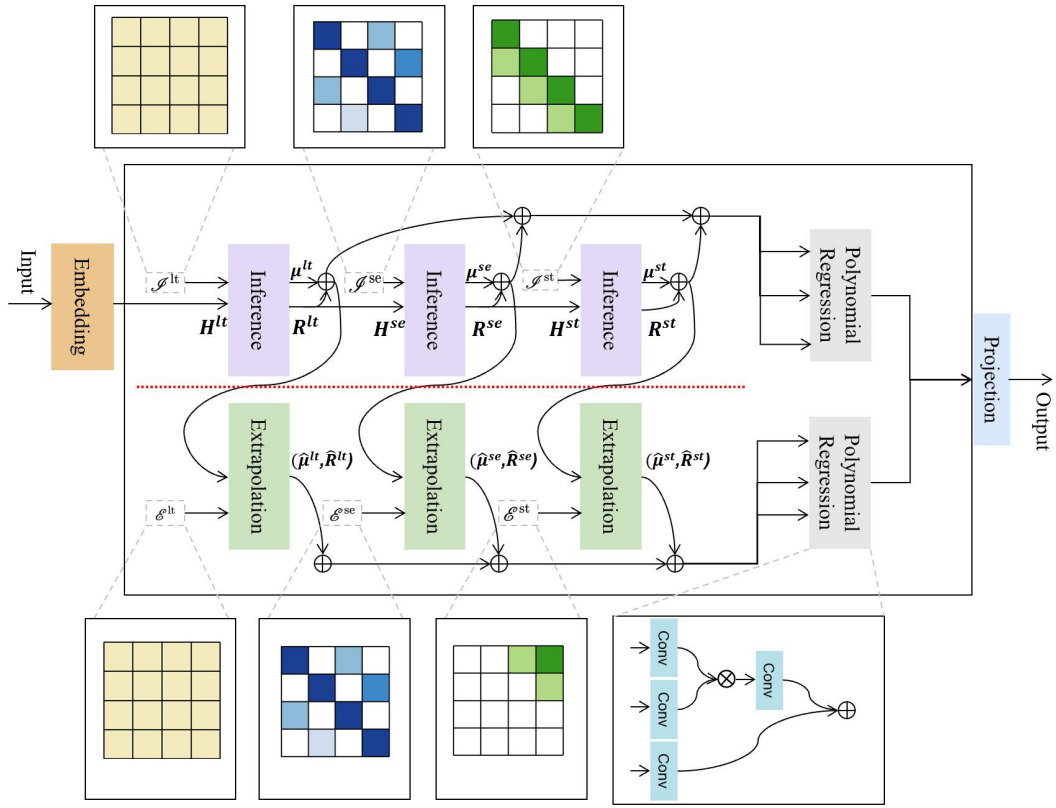


图 5 时间依赖性建模

在推理组件方面，我们通过推理算子 $\mathcal{J}^* \in \mathbf{R}^{T_{in} \times T_{in}}$ 沿着时间维度分别应用于每个序列，用于推断相应的时间成分，在整个系列中共享：

$$\mu_i^* = \mathcal{J}^* H_i^*$$

$$\sigma_i^{*2} = \mathcal{J}^* H_i^* - \mu_i^{*2} + \epsilon$$

其中 $*$ $\in \{lt, se, st\}$ 表示趋势、季节、短期变动； $H_i^* \in \mathbf{R}^{T_{in} \times d}$ 表示第 i 个序列的输入特征，具

体为空间依赖性模型的输出或上个推理模块的残差输出； $\mu_i^*, \sigma_i^{*2} \in R^{N \times T_{in} \times d}$ 为推理第 i 个序列对应成分的均值和标准差；

接着，通过从数据中减去平移效应并将差值除以缩放效应将其余成分从数据中分离出来，得到对应的推理组件的残差 $R_i^* \in R^{N \times T_{in} \times d}$ ：

$$R_i^* = \frac{H_i^* - \mu_i^*}{\sigma_i^*}$$

在外推组件方面，我们通过外推算子 \mathcal{E}^* 对 μ_i^* 与 R_i^* 进行线性映射进行外推：

$$\hat{\mu}_i^* = \mathcal{E}^* \mu_i^*$$

$$\hat{R}_i^* = \mathcal{E}^* R_i^*$$

其中 $\hat{\mu}_i^* \in R^{N \times T_{in} \times d}$ 为外推第 i 个序列对应成分的均值和标准差； \hat{R}_i^* 为对应的外推组件的残差

在时间序列的趋势、季节和短期波动成分的推理和外推中，推理组件和外推组件的主要差别在于推理算子 \mathcal{J}^* 和外推算子 \mathcal{E}^* 的不同，其具体定义如下。

①趋势成分推理和外推算子

趋势成分旨在表征时间序列数据的趋势模式，通过汇总多个季节收集的样本，消除仅影响少数时间步长的短期影响，并获得偏差较小的长期成分的估计。长期分量的推理和外推算子公式定义如下：

$$\mathcal{J}_{t_1, t_2}^{lt} = \frac{1}{T_{in}}$$

$$\mathcal{E}_{t_1, t_2}^{lt} = \frac{1}{T_{in}}$$

②季节成分推理和外推算子

在推理算子方面，在季节持续时间一致的假设下，定义 τ_{in} 为输入序列所包含的季节数， c 为单个季节的长度，则有 $T_{in} = \tau_{in} * c$ 。与趋势成分推理和算子相反，季节性成分采用扩大窗口策略，该窗口由 τ_{in} 大小和 c 膨胀因子定义，捕获跨越多个季节的数据点以反映数据的周期性特征。

此外，我们利用注意力机制，根据输入序列中的顺序为不同季节分配不同的权重，从而可

以准确地衡量它们与季节成分的相关性。权重定义为参数矩阵 $W^{se} \in R^{\tau_{in} \times \tau_{in}}$ ，接着通过进行归一化得到 $A^{se} \in R^{\tau_{in} \times \tau_{in}}$ 。其中 A^{se} 代表季节相关性，用于评估季节之间存在的内在相互联系，季节成分推理算子定义如下：

$$\mathcal{I}_{t_1, t_2}^{se}(A) = \begin{cases} A_{i,j}^{se} & t_1 - t_2 \equiv 0 \pmod{c} \\ 0 & \text{其它} \end{cases}$$

其中：

$$i = \left\lfloor \frac{t_1}{c} \right\rfloor, j = \left\lfloor \frac{t_2}{c} \right\rfloor$$

$$A_{i,j}^{se} = \frac{\exp(W_{i,j}^{se})}{\sum_{k=0}^{\tau_{in}-1} \exp(W_{i,k}^{se})}$$

在外推算子方面，若推理算子所得的季节性成分稳定，则可直接重复进行直接外推，但这一情况并不多见，因此为适应不断变化的季节性成分，我们构建了动态的外推算子。首先我们定义输出序列的最小季节数表示为 τ_{out} ，定义 $\hat{W}^{se} \in R^{\tau_{out} \times \tau_{in}}$ 为参数矩阵来捕获输入和输出序列中每对季节之间的季节相关性，与推理算子类似，我们对 \hat{W}^{se} 进行归一化，得到 $\hat{A}^{se} \in R^{\tau_{out} \times \tau_{in}}$ ，故季节成分外推算子定义如下：

$$\mathcal{E}_{t_1, t_2}^{se}(A) = \begin{cases} \hat{A}_{i,j} & t_1 - t_2 \equiv 0 \pmod{c} \\ 0 & \text{其它} \end{cases}$$

其中：

$$i = \left\lfloor \frac{t_1}{c} \right\rfloor, j = \left\lfloor \frac{t_2}{c} \right\rfloor$$

$$\hat{A}_{i,j}^{se} = \frac{\exp(\hat{W}_{i,j}^{se})}{\sum_{k=0}^{\tau_{in}-1} \exp(\hat{W}_{i,k}^{se})}$$

③短期波动成分推理和外推算子

短期成分识别季节性和长期成分未解释的不规则现象和短暂现象，与季节成分推理外推相比，它需要一个较小的窗口大小 δ ，以包含具有直接相关性的最近观察结果。这些观察结果受当前短期因素的影响，根据相关滞后表现出不同程度的相关性。为捕获不同滞后阶数的相关性，我们定义参数向量 $w^{st} \in R^{\delta}$ 来构建注意力机制来利用了注意力加权的观察结果，最终短期波动成分外推算子定义如下：

$$\mathcal{I}_{t_1, t_2}^{st}(A) = \begin{cases} A_{t_1-t_2}^{st} & (t_1 - t_2 \geq 0) \wedge (t_1 - t_2 < \delta) \\ 0 & \text{其它} \end{cases}$$

$$A_i^{st} = \frac{\exp(w_i^{st})}{\sum_{i=0}^{\delta-1} \exp(w_i^{st})}$$

与季节成分的外推算子类似,我们首先定义 $\hat{w}_{st} \in R^{\delta \times \delta}$ 为参数矩阵来捕获窗口大小中各个时刻的相关性,并对 \hat{w}_{st} 进行归一化,故短期波动成分外推算子定义如下:

$$\mathcal{E}_{t_1, t_2}^{st}(A) = \begin{cases} \hat{A}_{t_1, T_{in}-1-t_2}^{st} & (t_1 < \delta) \wedge (t_2 > T_{in} - \delta - 1) \\ 0 & \text{其它} \end{cases}$$

$$\hat{A}_{i,j}^{st} = \frac{\exp(\hat{w}_{i,j}^{st})}{\sum_{k=0}^{\delta-1} \exp(\hat{w}_{i,k}^{st})}$$

我们使用 pytorch 库编写模型对应代码, 部分代码截图如下:

```

1 import torch
2 from torch import nn
3 import torch.nn.functional as F
4 from model.PM25_GNN import GraphGNN
5 import math
6
7 epsilon = 0.0001 # 数值稳定性常数
8
9 class TimeSeriesModule(nn.Module):
10     """使用SSCNN方式的时间序列建模模块"""
11     def __init__(self, input_dim, hidden_dim, num_nodes, hist_len, pred_len, cycle_len, short_period_len):
12         super(TimeSeriesModule, self).__init__()
13
14         # 这个cycle_len 是周期的长度, 影响后面的季节性建模的周期划分, hist_len和pred_len的是cycle_len的整数倍
15
16         # input_dim: 输入特征维度, 为26, self.gnn_out+self.in_dim=26
17         self.input_dim = input_dim
18         self.hidden_dim = hidden_dim # 隐藏层维度, 为64
19         self.num_nodes = num_nodes # 节点数量, 为13
20         self.hist_len = hist_len # 历史序列长度,
21         self.pred_len = pred_len # 预测长度
22         self.cycle_len = cycle_len # 周期长度, 用于季节性的分割
23         self.short_period_len = short_period_len # 短期依赖长度, 一天的数据点数
24
25
26         # 输入投影层
27         self.input_projection = nn.Conv2d(
28             in_channels=input_dim,
29             out_channels=hidden_dim,
30             kernel_size=1

```

图 6 GAT-TSDFormer 模型部分代码截图

3.进行模型相关实验

在完成 GNN-SCNN 代码撰写之后，我们进行了消融实验、敏感性实验和基线模型对比实验来全面了解模型的性能以及各个组件的有效性。以下是我们项目文件的目录结果：

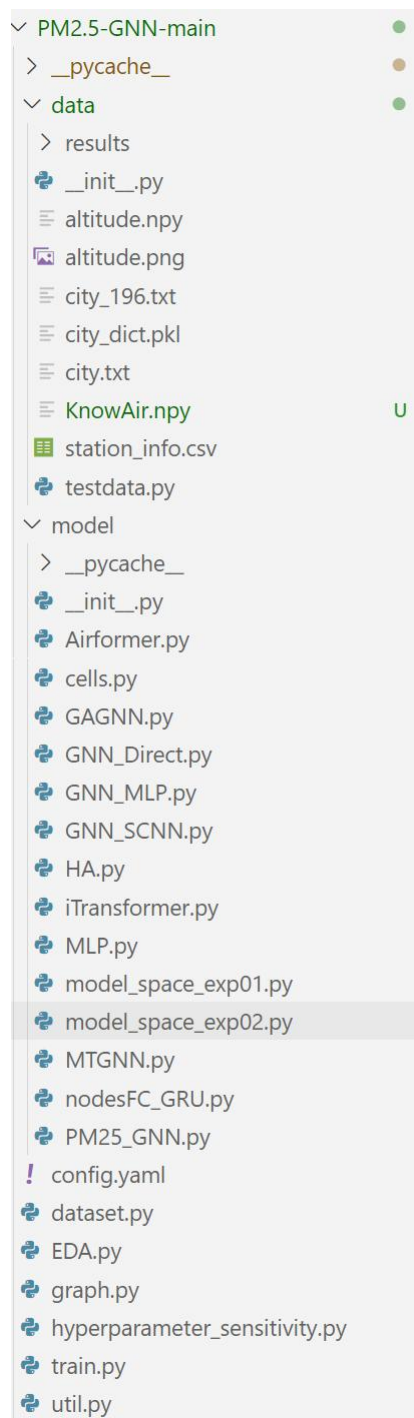


图 7 项目目录文件结构

(1) 消融实验

消融实验是指通过系统地移除模型的某一部分来评估该部分对模型整体性能的影响，从而

来验证模型某组件的合理性。本文以 GAT-TSDFomer 模型作为完整模型，分别以单独移除空间依赖性建组件、移除时间依赖性组件以及两者皆移除作为移除组件后模型，通过对比来了解被移除组件的作用来验证我们的观点，并根据结果对模型做进一步的调整和实验。消融实验的结果如下：

表 5 消融实验结果

Metric	移除空间依赖性组件	移除时间依赖性组件	两者皆移除后结果	完整模型
Train_Loss	0.3501 ± 0.0041	0.2921 ± 0.0092	0.3708 ± 0.0043	0.2784 ± 0.0102
Validate_Loss	0.3613 ± 0.0065	0.3111 ± 0.0028	0.3779 ± 0.0073	0.3044 ± 0.0033
Test_Loss	0.4003 ± 0.0070	0.3901 ± 0.0080	0.4227 ± 0.0076	0.4040 ± 0.0084
RMSE	38.0010 ± 0.5500	36.5010 ± 0.3500	39.3467 ± 0.5916	37.4006 ± 0.3669
MAE	22.5020 ± 0.4000	21.8012 ± 0.2200	23.6006 ± 0.4271	22.2471 ± 0.2347
MAPE	0.3712 ± 0.0180	0.3112 ± 0.0070	0.3842 ± 0.0208	0.3185 ± 0.0081
CSI	0.8801 ± 0.0025	0.8721 ± 0.0018	0.8659 ± 0.0038	0.8840 ± 0.0024
POD	0.9710 ± 0.0015	0.9701 ± 0.0008	0.9686 ± 0.0027	0.9721 ± 0.0043
FAR	0.0950 ± 0.0040	0.1250 ± 0.0030	0.1091 ± 0.0053	0.0930 ± 0.0048
内存消耗	3.5	4.7	2.0020	10.7476
训练时间	91.0478	80.0345	112.8595	96.1645

从实验结果可以看出，在移除了空间组件或时间组件之后，模型在各项性能指标上皆有降低，说明项目所设计的模型各组件是有效的。

(2) 敏感性实验

项目分别就优化器与学习率进行了敏感性测试，结果如下：

①优化器的敏感性测试，分别测试三组优化器，AdamW,GCD 和 RMSprop,测试结果如下，其中 SGD 训练损失下降速度最慢，最终仍然处于较高水平，验证损失保持在较高水平，表明模型未能很好学习到模式，相比之下，AdamW 和 RMSprop 表现优于 SGD，训练损失和验证损失都快速下降，并在较低水平稳定。

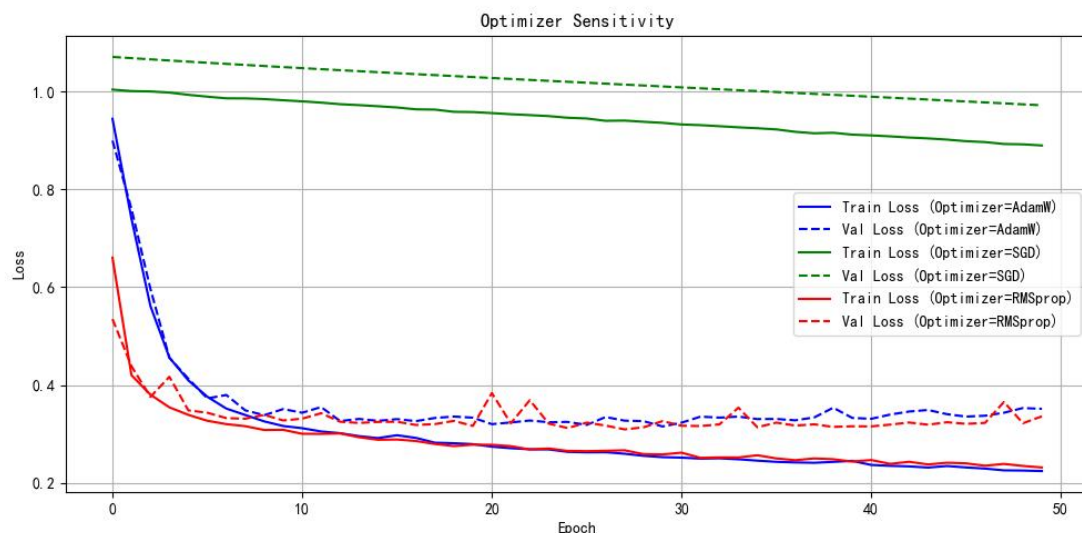


图 8 优化器的敏感性测试结果图

②学习率的敏感性测试，测试学习率为 0.001, 0.0005 和 0.0001 三组学习率下模型的损失函数的变化，学习率为 0.001 的时候，训练损失和验证损失都快速下降，并在较低水平波动，学习率为 0.0005 时，训练损失下降速度比 0.001 略慢，但最终能达到相近的损失水平，验证损失较低且稳定，泛化能力较好，学习率为 0.0001 时，训练损失下降速度最慢，并且未能达到较低损失水平，验证损失也较高，表明模型泛化能力不佳。

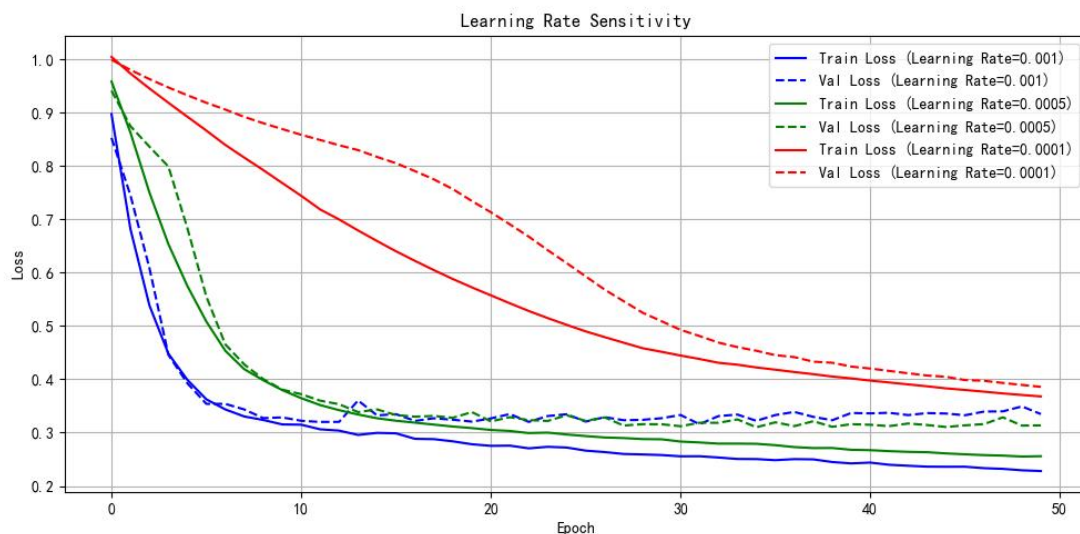


图 9 学习率的敏感性测试结果图

通过上述敏感性测试，我们发现：学习率选择在 0.001 能够快速收敛并保持泛化能力，在优化器选择上，AdamW 和 RMSprop 优于 SGD，在本任务中能够更快降低损失并稳定泛化，所以模型采用 AdamW+学习率 0.001 作为最佳组合。

(3) 基线模型对比试验

为更加准确地衡量模型的预测精度，我们将 GAT-STDFormer 模型与以下三类基准模型进行比较：

①历史平均值（HA）：通过相应时期的历史读数的平均值来预测空气质量。它不需要复杂的算法和大量的计算资源，只依赖于历史平均值。但由于没有考虑其他因素的影响，对于复杂

的时空关系和变化模式预测效果有限。

②多层感知机 (MLP)：它通过多个全连接层组成，来构建无时序性的模型来预测空气质量。MLP 的基本结构包括输入层、隐藏层和输出层。输入层接收原始数据或特征向量作为输入。它在处理非线性关系时具有较强的表达能力，但对输入数据的尺度和分布比较敏感，需要进行适当的预处理。

③STGNN 的变体模型：STGNN 及其变体是目前图论在 PM2.5 时空预测的重要模型，我们选择 MTGNN (Wu et al 2020) 和 GAGNN (Chen et al. 2021) 作为基线模型。这些模型利用图神经网络的能力来捕捉空气质量数据中的时空依赖关系，通过建模节点之间的关系和时间上的演化，实现对未来空气质量的预测，其优点在于能够处理复杂的时空数据，具有较高的预测准确性，但缺点在于计算复杂度高，以及对稀疏矩阵敏感。

④Transformer 的变体模型：目前基于 Transformer 变体模型在时空预测上获得了很大成功，我们选取 iTransformer (Q. Liu. et al.2022) 和 Airformer (D. Liang et al. 2023) 作为基线模型。这些模型利用 Transformer 的自注意力机制和多头注意力机制来捕捉时空序列数据中的依赖关系和模式，并进行预测。相比于传统的循环神经网络，Transformer 模型具有并行计算的优势，并且能够处理长距离依赖关系，但其缺点在于模型复杂度高，需要较长的训练时间以及计算资源。

最终实验结果如下：

表 6 基线模型对比试验表

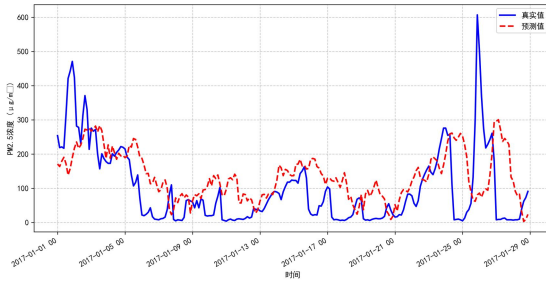
Metric	HA	MLP	MTGNN	GAGNN	iTransfor mer	Airforme r	GAT-TS DFormer
Train_Loss	0.4521±	0.3708±	0.2668±	0.2665±	0.1933±	0.2622±	0.2784±
	0.0119	0.0043	0.0058	0.0094	0.0119	0.0149	0.0102
Validate_Loss	0.4371±	0.3779±	0.3107±	0.2899±	0.3130±	0.2858±	0.3044±
	0.0049	0.0073	0.0070	0.0034	0.0064	0.0050	0.0033
Test_Loss	0.5717±	0.4227±	0.4233±	0.3849±	0.3971±	0.3821±	0.4040±
	0.0062	0.0076	0.0098	0.0065	0.0051	0.0051	0.0084
RMSE	46.7061	39.3467	38.3333	36.7328	37.0274	36.5493	37.4006
	±0.5004	±0.5916	±0.5417	±0.7144	±0.2300	±0.3288	±0.3669
MAE	28.5324	23.6006	22.8187	21.8300	22.0629	21.7190	22.2471
	±0.3397	±0.4271	±0.3897	±0.5178	±0.1752	±0.2441	±0.2347
MAPE	0.6192±	0.3842±	0.3307±	0.3254±	0.3022±	0.3211±	0.3185±
	0.0228	0.0208	0.0156	0.0195	0.0095	0.0149	0.0081
CSI	0.8196±	0.8659±	0.8801±	0.8845±	0.8867±	0.8851±	0.8840±
	0.0049	0.0038	0.0053	0.0054	0.0024	0.0032	0.0024
POD	0.9728±	0.9686±	0.9685±	0.9775±	0.9695±	0.9771±	0.9721±
	0.0035	0.0027	0.0039	0.0045	0.0031	0.0028	0.0043

FAR	0.1612±	0.1091±	0.0939±	0.0971±	0.0879±	0.0961±	0.0930±
	0.0071	0.0053	0.0060	0.0091	0.0036	0.0053	0.0048
内存消耗	111.9551	2.0020	10.8296	38.2180	10.3276	38.1240	10.7476
训练时间	336.9035	112.8595	88.4041	186.4534	191.5296	308.6492	96.1645

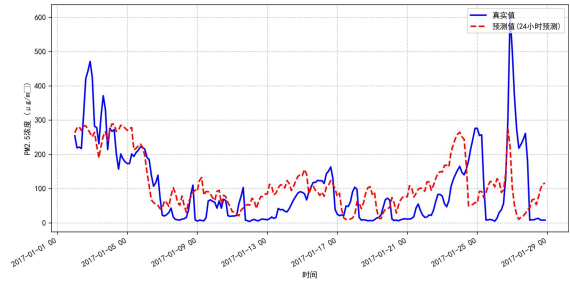
首先，我们展示了八个评估指标的结果，通过仔细分析揭示了所提出模型 GAT-TSDFormer 的特点。总体性能方面，如表 6 所示，HA、MLP 表现最差，这是由于它们只能访问节点表示。它们的较差表明 PM2.5 预测中使用邻域信息的必要性。

其他五个模型都能捕获时空依赖，我们比较这五个模型以确定哪种方法更有效地利用领域知识。尽管 iTransformer 训练损失最低，而测试损失相对较高，表明这是过拟合现象。相比之下，GAT-TSDFormer 在几乎所有预测信念指标上与其他模型相近，同时内存消耗和训练时间相较于其他模型大幅度降低，表明其在利用领域知识和预测能力上的优势且实现了模型轻量化。

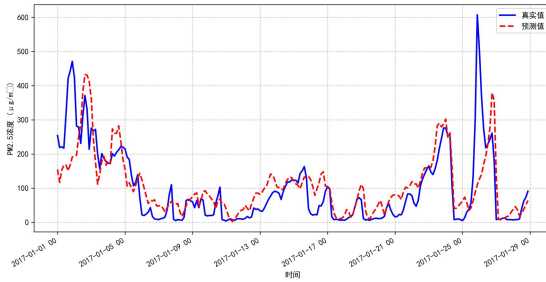
在更细粒度的分析中，我们选择了某些区域和城市（如北京）进行案例研究，通过 RMSE 和皮尔逊相关系数等量化指标验证了模型的有效性。结果表明，GAT-TSDFormer 在不同子数据集上的预测能力一致优异，尤其是在极端情况下表现出色，这得益于其对领域知识的充分利用和对长期依赖的建模能力。



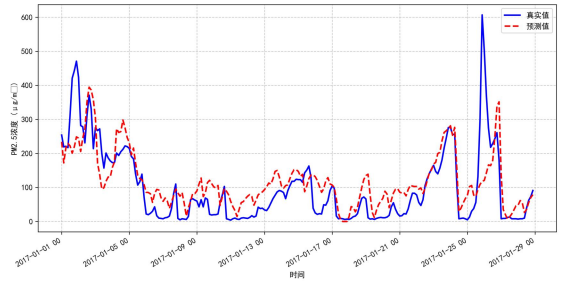
(a)HA



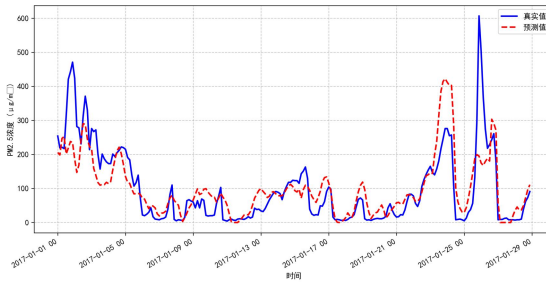
(b)MLP



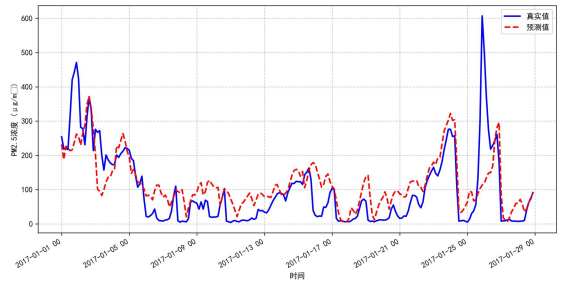
(c)MTGNN



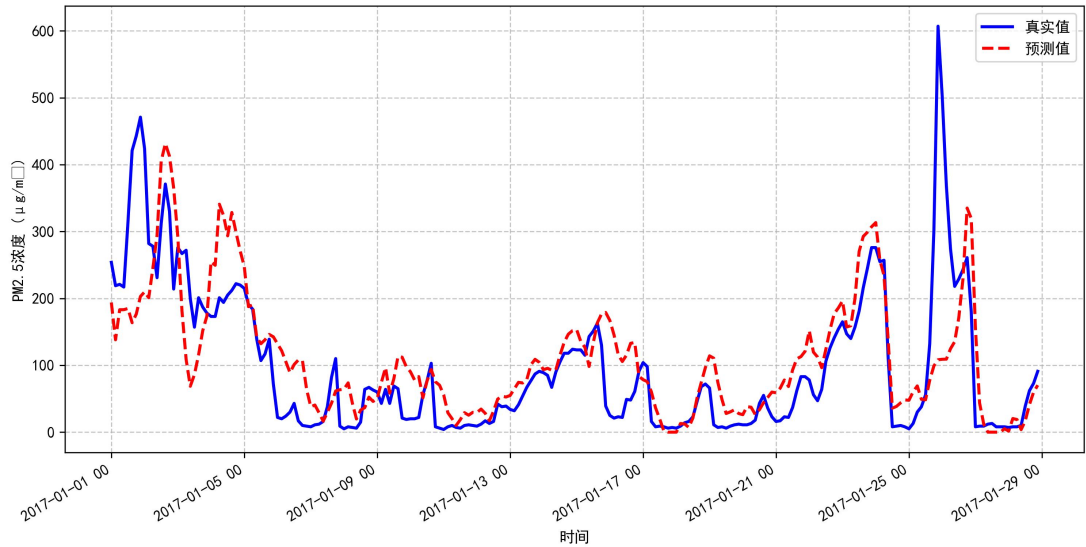
(d)GAGNN



(e)iTransformer



(f)Airformer



(e) GAT-TSDFormer

图 10 各模型北京 PM2.5 预测效果图

4.模型解释性分析

(1) 空间扩散规律

在空间传播规律中，我们将 GAT-TSDFormer 空间依赖性建模中城市相关性变量进行提取，以北京为例，结果如下：

表 7 空间扩散规律结果

城市名	与北京的空间相关性
北京	0.5001
天津	0.4822
石家庄	0.5167
唐山	0.5332
秦皇岛	0.4946
邯郸	0.5209
保定	0.522
张家口	0.5092
承德	0.432
廊坊	0.476
沧州	0.5177
衡水	0.5411
邢台	0.5421

从表 7 可以看出，北京周边城市与北京的空间相关性差异不大，其中衡水和邢台与北京的 PM2.5 浓度预测相关性最高，承德则相关性最低。

(2) 时间传播规律

在时间传播规律中，我们分别从长期性、季节性、短期成分上进行分析，结果如下：

长期依赖性建模，捕捉的是趋势性规律，在时间步 t 时，输入综合特征向量 x_t ，该向量由 13 个原始特征（PBL 高度、K 指数、相对湿度等）以及通过建模得到的空间特征组合而成，这些特征反映了该时刻各个节点的状态信息，并且融合了空间上的关联信息。需要求解出长期均值

μ_{lt} , 长期二阶矩 $E[x_{lt}^2]$, 长期方差 σ_{lt}^2 , 长期标准化残差 r_{lt} 以及长期预测均值 $\hat{\mu}_{lt}$ 和残差 \hat{r}_{lt} 。

$$\mu_{lt} = \frac{1}{T} \sum_{t=1}^T x_t$$

$$E[x_{lt}^2] = \frac{1}{T} \sum_{t=1}^T x_t^2$$

$$\sigma_{lt}^2 = E[x_{lt}^2] - \mu_{lt}^2 + \epsilon$$

$$r_{lt} = \frac{x_t - \mu_{lt}}{\sigma_{lt}}$$

$$\hat{\mu}_{lt} = \mu_{lt}(T), \hat{r}_{lt} = r_{lt}(T)$$

季节性时间建模中, 我们利用注意力机制, 根据输入序列中的顺序为不同季节分配不同的权重, 从而可以准确衡量它们与季节成分的相关性。 α_{se}^i 为季节性注意力权重, 由 I_{se}^i 通过 $softmax$ 函数处理得到, 满足 $\sum_{i=1}^N \alpha_{se}^i = 1$, 表示第 i 哥季节的数据在计算当前季节均值时所占的比重。

$$\alpha_{se}^i = \frac{\exp(I_{se}^i)}{\sum_{j=1}^N \exp(I_{se}^j)}$$

I_{se}^i , 是可学习的季节性注意力参数, 衡量不同季节的数据对当前季节均值计算的重要性。模型通过学习这些参数, 能够自动调整不同季节数据在计算季节性均值时的权重。 $hist_len$ 是历史信息的步长, $cycle_len$ 是一个周期的步长, 在 I_{se}^i 权重中, 我们设置 $hist_len$ 为 $cycle_len$ 的4倍, 也就是将原始的数据分为四个季节周期, 得到 I_{se}^i 的季节性注意力权重矩阵, 用于计算季节性均值和二阶矩, 该矩阵的形状是 $(hist_len // cycle, hist_len // cycle_len, 1, 1)$, 会在 $softmax$ 函数在第一个维度进行归一化, 举例一个归一化之后的权重矩阵:

$$\begin{vmatrix} 0.2138 & 0.2363 & 0.2612 & 0.2887 \\ 0.2363 & 0.2612 & 0.2887 & 0.2138 \\ 0.2612 & 0.2887 & 0.2138 & 0.2363 \\ 0.2887 & 0.2138 & 0.2363 & 0.2612 \end{vmatrix}$$

第一行 $[0.2138, 0.2363, 0.2612, 0.2887]$, 表示当计算第一个季节性周期的特征时, 第一、二、三和四个季节性周期的权重分别为 0.2138, 0.2363, 0.2612, 0.2887, 表明在计算第一个季节性周期的特征时, 第四个季节性周期的影响最大, 第一个季节性周期自身的影响相对最小。

β_{se}^i 为季节性预测注意力权重，由 E_{se}^i 通过 $softmax$ 函数处理得来，用于预测未来季节的均值和残差。反映了不同历史季节数据对未来季节预测的重要程度。

$$\beta_{se}^i = \frac{\exp(E_{se}^i)}{\sum_{j=1}^N \exp(E_{se}^j)}$$

E_{se}^i 考虑了预测长度和历史长度之间的关系，用于从历史数据中提取信息来预测未来。其中 E_{se}^i 表示在预测第 i 个未来季节性周期时，第 j 个历史季节性周期的权重。也就是说，在计算未来第 i 个季节性周期的均值和残差时，第 j 个历史季节性周期的信息所占的比重。下面为其归一化权重矩阵：

$$\begin{vmatrix} 0.1027 & 0.2991 & 0.3973 & 0.2009 \\ 0.0667 & 0.1889 & 0.3111 & 0.4333 \end{vmatrix}$$

由上面的权重矩阵，我们可知，在预测第一个未来周期的时候，主要由第3个历史周期（权重0.3973）和第2个历史周期（权重0.2991）决定，第1个和第4个历史周期的影响较小。在预测第2个未来周期的时候，主要由第4个历史周期和第3个历史周期决定。

短期成分识别季节性和长期成分未解释的不规则现象和短暂现象，和季节成分推理相比，需要一个更小的窗口大小，包含具有直接相关性的最近观察结果，这些观察结果受当前短期因素的影响，根据相关滞后表现出不同程度的相关性。

需要计算短期均值 μ_{st} ，短期二阶矩 $E[x_{st}^2]$ ，短期方差 σ_{st}^2 ，短期标准化残差 r_{st} ，短期预测均值 $\hat{\mu}_{st}$ 和残差 \hat{r}_{st} 。

$$\mu_{st} = \sum_{i=1}^M \alpha_{st}^i x_{t-M+i}$$

$$E[x_{st}^2] = \sum_{i=1}^M \alpha_{st}^i (x_{t-M+i})^2$$

$$\sigma_{st}^2 = E[x_{st}^2] - \mu_{st}^2 + \epsilon$$

$$r_{st} = \frac{x_t - \mu_{st}}{\sigma_{st}}$$

$$\hat{\mu}_{st} = \sum_{i=1}^M \beta_{st}^i \mu_{st}(t-M+i)$$

$$\hat{r}_{st} = \sum_{i=1}^M \beta_{st}^i r_{st}(t - M + i)$$

其中 M 是短期窗口的长度，较短的窗口长度可以更敏锐捕捉近期数据的变化，但是较长的窗口则能综合更多的信息。 I_{st}^i 是可学习的短期注意力参数，它的形状是 $(1,1,1,shot_period_len)$ ，用于衡量短期窗口内不同时间点的数据对当前的短期均值计算的重要性， α_{st}^i 短期注意力权重，由 I_{st}^i 通过 $softmax$ 函数处理得到。

$$\alpha_{st}^i = \frac{\exp(I_{st}^i)}{\sum_{j=1}^M \exp(I_{st}^j)}$$

β_{st}^i 短期预测注意力权重， E_{st}^i 是用于短期预测的可学习注意力参数 E_{st}^i 的形状是 $(pre_len, 1,1,short_period_len)$ ，在预测未来短期的均值和残差的时候，帮助模型根据近期数据的变化趋势进行预测。

$$\beta_{st}^i = \frac{\exp(E_{st}^i)}{\sum_{j=1}^M \exp(E_{st}^j)}$$

三、本项目取得的成果及数量（如参加竞赛、发表论文、申请专利情况等，须提供相关证明材料；请参考《暨南大学大学生创新创业训练计划项目结题成果验收标准》填写）

项目级别： ☐ 国家级 ☒ 省级 ☐ 校级

项目类型： ☒ 创新训练 ☐ 创业训练 ☐ 创业实践

1. ☐ 发表学术论文（ ）份； 2. ☐ 获受理专利（ ）份； 3. ☐ 获著作权（ ）份；

4. ☐ 学术报告（1）份； 5. ☐ 省部级及以上竞赛获得三等奖或以上奖项（ ）份；

6. ☐ 项目发展历程（ ）份； 7. ☐ 商业计划书（ ）份；

8. ☐ 项目注册成立公司的工商营业执照（ ）份；

9. ☐ 校级及以上竞赛获得三等奖或以上奖项（ ）份；

10. ☐ 入驻校内或校外创业孵化基地 3 个月以上（ ）件，

11. ☐ 其他（ ）件，名称： _____

附件清单：学术报告《基于 GAT-TSDFomer 的 PM2.5 时空预测——基于轻量化与解释性视角》

四、项目工作总结（对研究工作的评价、预期前景、下一步工作计划等）

项目组基于所提出的 GAT-TSDFomer 模型进行了各项实验，得出了最终的结论并撰写了学术报告，在申报阶段的预期结果基本完成，在部分领域甚至取得了超出设想的结果，对于研究问题获得更加深刻的认识，可以说本项目取得了较为圆满的成功。项目的顺利完成离不开项目组每一位成员的辛勤付出，每位成员认真完成自己的分工，遇到问题时积极讨论努力解决，克服了众多的困难，不断细化完善研究，使项目得到了较为深刻的学术成果。本次项目的过程为项目组成员提供了有益的学术训练，特别是在代码撰写、团队协作方面获取了宝贵的经验，为后续的学习研究打下了良好基础。

在本项目的所得成果上，可以将模型推广至各种时空预测场景问题之中，例如其他各项污染物的传播预测、人口流动预测等等问题，在未来中，我们将进一步调整模型，优化模型相关细节，已实现更高精度预测、更准确解释性的 PM2.5 时空预测。

五、经费使用情况（项目经费使用情况部分说明购置实验材料、试剂、药品、加工测试、资料、复印、调研、交通等已开支经费数额，可自行加行）

序号	用途	金额（元）
1		
2		
3		
4		
合 计		

我承诺在大学生创新创业训练计划项目实施过程中遵守学校有关规定，恪守学术规范，本项目的成果中除特别注明和引用外，均为本项目观点，不存在剽窃、抄袭他人学术成果，伪造、篡改实验数据的情况。如有违规行为，愿承担一切责任，接受学校的处理。

项目负责人签名：

年 月 日

六、指导教师意见

签名：

年 月 日

七、学院专家组验收意见

专家签名：

年 月 日

八、学院意见

验收结果：优秀☐ 合格☐ 不合格☐ 延期☐

签字盖章：

年 月 日

九、学校意见

年 月 日