

基于 Orion 模型的交通流量大数据建模与预测研究

摘要

国家统计局发布的《新中国 75 年经济社会发展成就系列报告》^①指出，2023 年末我国常住人口城镇化率达 66.16%，比 1949 年末提高 55.52%，年均提高 0.75%。随着城市化进程的加速和机动车保有量的增长，交通拥堵和事故频发等问题已成为全球城市面临的普遍难题。为减少交通拥堵情况，降低驾驶过程中产生的经济和时间成本，交通流量预测显得愈发重要。

本文借助 PeMS03、PeMS04、PeMS08、NYCTaxi(r)及 NYCBike2(r)的交通流量数据，构建交通预测深度学习模型（本文简称 Orion 模型）。为解决现有交通预测模型在长期预测时精度不足、性能不稳定、多特征识别能力有限；以及过度依赖通用注意力机制导致模型在交通预测领域针对性不足且计算复杂度高等问题。本文采用线性插值和多周期滑动检索方法处理交通流量数据，借助三周期（小时周期、日周期和周周期）输入，捕捉不同时间尺度下的流量模式，经动态图注意力模块、时序多头自注意力模块和前馈网络三个子模块组成的时空模块对三周期输入数据进行时空特征提取，在此基础上，借助多头自注意力融合模块融合三周期数据，由前馈网络和线性层进行最终调整后，完成模型的构建。选择 LSTM、STGCN、DCRNN、ASTGCN、Graph-WaveNet 以及 STID 模型作为基线模型，进行模型精度的评价，并在 Taxi BJ 数据集完成应用。

研究结果表明，Orion 模型在 MAE、MAPE 和 RMSE 三个指标上，较 LSTM、STGCN、DCRNN、ASTGCN、Graph-WaveNet 及 STID 模型展现出更高精度，尤其在长期预测中表现出更优越的精确度和稳定性。Taxi BJ 数据集的应用进一步揭示，北京交通预测周周期规律至关重要，预测时段前 10-30 分钟的数据对准确捕捉流量变化起关键作用，且不同节点间的强弱关系需被充分考

^① 国家统计局.沧桑巨变换新颜 城市发展启新篇——新中国 75 年经济社会发展成就系列报告之十九[2024-09-23]. https://www.stats.gov.cn/zt_18555/ztf/xzg75njshf/cj/202409/t20240923_1956630.html

虑以制定针对性方案。据此，相关部门可通过优化信号灯配时等措施，以改善北京市交通状况。

关键词：Orion 模型；交通流量预测；ST 模型；多头自注意力机制；图注意力网络

目 录

摘要.....	1
一、绪论.....	1
(一)问题的提出.....	1
(二)文献综述.....	2
1.传统统计与机器学习方法.....	2
2.深度学习与时空建模.....	2
3.注意力机制与高级混合模型.....	3
4.文献述评.....	4
(三)特色与创新点.....	4
1.设计基于 KNN 的先验和后验信息相结合的空间动态图注意力神经网络..	4
2.设计融合多头自注意力机制的多周期融合模块.....	4
3.设计可以充分利用多周期和多相关影响特征的交通流量预测模型.....	4
二、模型构建.....	5
(一)构建原理.....	5
(二)三周期数据输入.....	6
1.小时数据 (Xh)	6
2.日周期数据 (Xd)	6
3.周周期数据 (Xw)	7
(三)动态图注意力模块.....	7
4.注意力计算.....	8
(四)时序多头自注意力模块.....	10
1.缩放点积注意力机制.....	11
(五)多头自注意力融合模块.....	12
1.多头自注意力融合路径.....	13
2.线性投影路径.....	14
三、模型精度评价与应用.....	15

(一)数据来源及描述.....	15
(二)数据处理.....	16
1.基于 KNN 的加权邻接矩阵构建.....	17
(三)精度对比实验.....	17
1.任务概述.....	17
2.Orion 模型超参数解释、超参数调整实验结果及分析.....	17
3.基线模型超参数设置.....	19
4.精度对比实验结果及其分析.....	22
(四)模型应用.....	24
1.数据获取.....	24
2.实验描述.....	26
3.结果分析.....	27
四、结论与展望.....	30
(一)结论.....	30
1.精度对比实验.....	30
2.模型应用.....	30
(二)展望.....	30
1.应用领域扩展.....	30
2.短期预测效果欠佳.....	30
3.非结构化数据处理能力不足.....	30
4.预测逻辑改进.....	31
参考文献.....	32
附录.....	35
致谢.....	错误!未定义书签。

表格与插图清单

表 1	数据集统计信息.....	15
表 2	Orion 模型超参数解释	18
表 3	Orion 模型超参数调整实验结果（续 1）	19
表 4	LSTM 模型超参数解释与设置	20
表 5	STGCN 模型超参数解释与设置	20
表 7	ASTGCN 模型超参数解释与设置	21
表 8	Graph-WaveNet 模型超参数解释与设置	21
表 10	不同模型在五种数据集上的交通流量预测性能比较（续 1）	22
表 10	不同模型在五种数据集上的交通流量预测性能比较（续 2）	22
表 11	Taxi BJ 数据集统计信息	25
表 12	Taxi BJ 实验结果对比	26
表 13	流量预测值与真值对比.....	27
图 2	模型结构图.....	5
图 3	时间序列片段输入示例.....	6
图 4	动态图注意力结构图.....	7
图 5	注意力计算模块结构图.....	8
图 6	时序多头自注意力模块结构图.....	10
图 8	多头自注意力融合模块结构图.....	12
图 9	多头自注意力融合路径图.....	13
图 12	数据处理流程图.....	16
图 16	不同方法在 Taxi BJ 数据集上的性能随预测区间增大的比较	26
图 18	周周期动态图注意力模块节点间权重热力.....	28
图 19	周周期时序多头自注意力模块时间步间权重热力图.....	28

基于 Orion 模型的交通流量大数据建模与预测研究

一、绪论

(一)问题的提出

中新经纬 2025 年 2 月 17 日电^②，北京全网交通指数为 8.0，平均速度为 21.7 公里/时，达到严重拥堵。纵观世界，近年来，许多国家大力发展智能交通系统（ITS^③），以实现高效的交通管理。出行公路相对封闭，一旦发生拥堵，将严重影响其通行能力，若能够提前准确预测，交通管理部门就能据此更合理地引导车辆，提高公路网络的运行效率。

公路交通流量数据是在固定的时间点，由分布于连续空间中的固定位置记录的。显然，相邻位置和时间点的观测数据并非独立，而是彼此动态关联的。

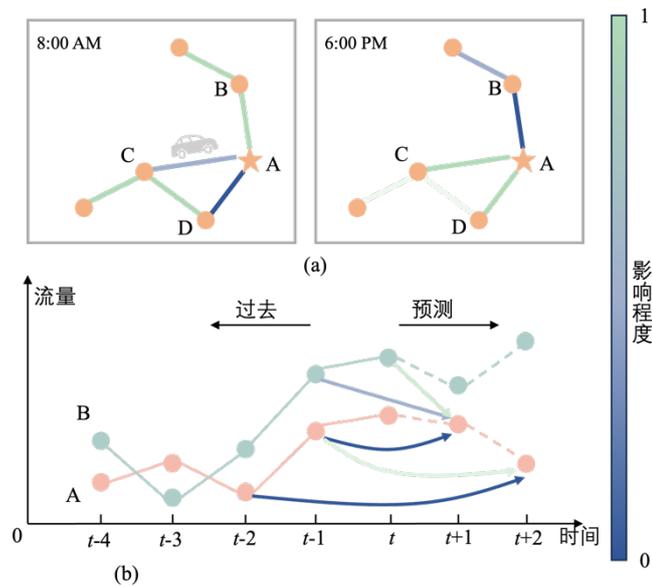


图 1 交通流量时空相关图

图 1 展示了交通流量（也可指车速、车道占用率等）的时空相关性。两点间粗线表示其相互影响强度，线的颜色越深，影响越大。在图 1(a)空间维度中，不

^② 中新经纬.出行注意！北京交通严重拥堵.[2025-02-17].<https://jw.jwview.com/jwview/content.jsp?id=616912>

^③ ITS(Intelligent Transportation Systems).智能交通系统.将先进技术整合到交通系统，提升安全、效率和可持续性的应用，如交通管理与自动驾驶。

同位置对点 A 的影响不同，甚至同一位置不同时间对点 A 的影响也不同。在图 1(b)时间维度中，不同位置的历史观测数据对点 A 未来不同时间的交通状态具有不同的影响。总之，公路网络上交通数据的相关性在空间和时间维度上均表现出强烈的动态性。如何探索非线性且复杂的时空数据，发现其内在的时空模式，并进行准确的交通流量预测，是一个极具挑战性的问题。

(二)文献综述

1.传统统计与机器学习方法

早期交通流量预测依赖简单且可解释性强的统计模型。Williams 等人(2003)提出的 ARIMA 模型通过差分和自回归捕捉时间依赖，但难以处理非线性和异质性^[1]。Zivot 等人(2006)提出的 VAR 模型进一步考虑多变量的相互关系，但其假设的线性关系限制了在复杂场景中的应用^[2]。这些方法在实际应用中预测精度往往达不到理想效果。

随着计算机技术进步，传统机器学习方法逐渐取代统计模型。COVER T 等人(1967)提出的 KNN 模型基于历史中的相似模式进行预测，但对噪声敏感^[3]。Jeong 等人(2013)提出的 SVM 利用核函数建模非线性关系，泛化能力较强，然而该方法高度依赖人工特征工程，导致复杂场景下的预测性能受限^[4]。Leshem 等人(2007)提出的随机森林通过结合多棵决策树建模非线性关系，展现出较强的适应性和稳定性^[5]。尽管如此，这些方法在捕捉交通流量的时空相关性方面仍显不足。

2.深度学习与时空建模

深度学习显著推进了交通流量预测，尤其是在时间序列建模方面。Hochreiter 等人(1997)提出的 LSTM 通过记忆单元捕捉长期依赖性^[6]。Cho K 等人(2014)提出的 GRU 简化了 LSTM 结构，降低了计算复杂度，同时保持了相似性能^[7]。Bai 等人(2019)验证了 GRU 在交通预测中的应用。这些模型擅长时间序列建模，但忽略了交通网络的空间结构^[8]。

为解决空间建模不足，研究者开发了时空深度学习方法。Zhang 等人(2018)设计的 ST-ResNet 模型基于残差卷积并结合 LSTM，展现了强大的空间特征提取

能力^[9]。Huang 等人（2018）提出的 CNN-LSTM 模型通过卷积提取空间特征，再利用 LSTM 建模时间依赖，显著提高了预测精度^[10]。然而，Kipf 等人（2017）的研究表明，这些方法通常假设为网格数据结构，难以适应非规则的交通网络^[11]。Bruna 等人（2013）提出的 GCN 通过建模非欧几里得结构应对这一问题^[12]。Defferrard 等人（2016）通过切比雪夫多项式优化了 GCN 计算效率^[13]。Yu 等人（2017）在此基础上提出的 STGCN 结合图卷集合时间卷积，首次实现了端到端学习^[14]。Cui 等人（2019）提出的图卷积循环神经网络进一步增强了网络规模预测能力^[15]。Li 等人（2017）提出的 DCRNN 通过扩散卷积建模动态空间依赖，但自回归机制限制了长期相关性捕捉^[16]。

3.注意力机制与高级混合模型

注意力机制的引入增强了时空模型的动态建模能力。Vaswani 等人（2017）提出的 Transformer 通过自注意力机制实现强大序列建模，启发了交通预测的应用^[17]。Velickovic 等人（2017）提出的 GAT 模型动态分配节点权重，为交通网络建模提供了新思路^[18]。Guo 等人（2019）提出的 ASTGCN 通过注意力机制自适应调整时空相关性，提升了预测性能^[19]。Liang 等人（2018）提出的多层次注意力网络通过多尺度建模提升了复杂场景的预测精度^[20]。Guo 等人（2021）提出的 ASTGNN 通过趋势感知多头自注意力与动态图卷积，增强了时空动态建模，但其高计算复杂度限制了实时应用^[21]。

近年来，研究者通过整合多种机制进一步提升了交通预测模型的性能。Wu 等人（2019）提出的 Graph-WaveNet 模型通过自适应邻接矩阵动态捕捉空间依赖，但其邻接矩阵生成方式缺乏对交通网络的深入挖掘^[22]。Song 等人（2020）优化的 GCN 框架提高了空间建模效率^[23]。Zhang 等人（2022）提出的 STID 模型通过并行多头注意力处理特征子空间，在长期预测表现优异，但计算成本高^[24]。Huang 等人（2020）、Li 和 Zhu（2021）以及 Zheng 等人（2020）通过融合多源数据（如天气、事件），进一步提升了模型精度^[25-27]。

4.文献述评

从统计模型到深度学习，交通流量预测技术取得显著进展。然而，现有方法仍面临挑战：模型难以同时精准建模短时波动与长期趋势，导致长序列预测误差累积，性能不稳定；过度依赖通用注意力框架，导致模型在交通预测领域针对性不足且计算复杂度高。未来研究应聚焦于开发高效、动态的建模方法，融合多源数据，同时优化短期精度与长期稳定性，为智能交通系统的发展提供支持。

(三)特色与创新点

1.设计基于 KNN 的先验和后验信息相结合的空间动态图注意力神经网络

提出动态图注意力模块，结合 KNN 生成的先验加权邻接矩阵与后验学习，动态调整节点间空间依赖。通过权重约束（ $\pm 50\%$ 浮动）与动态归一化降低计算复杂度，增强模型长期预测稳定性与泛化能力。

2.设计融合多头自注意力机制的多周期融合模块

设计多头自注意力融合模块，动态融合三周期特征，捕捉周期间复杂交互关系。通过多头自注意力机制自适应分配周期权重，显著提升多尺度时间模式建模精度，优化高峰期与异常场景预测效果。

3.设计可以充分利用多周期和多相关影响特征的交通流量预测模型

本模型系统化整合小时、日、周周期交通数据，通过多个创新模块，精准捕捉复杂时空动态。显著提升预测精度与鲁棒性。

二、模型构建

(一)构建原理

本文提出一种可应用于交通流量预测的 Orion 模型，针对交通时间序列与时空数据的特性进行优化，系统化处理复杂交通场景。Orion 模型整体架构如图 2 所示。

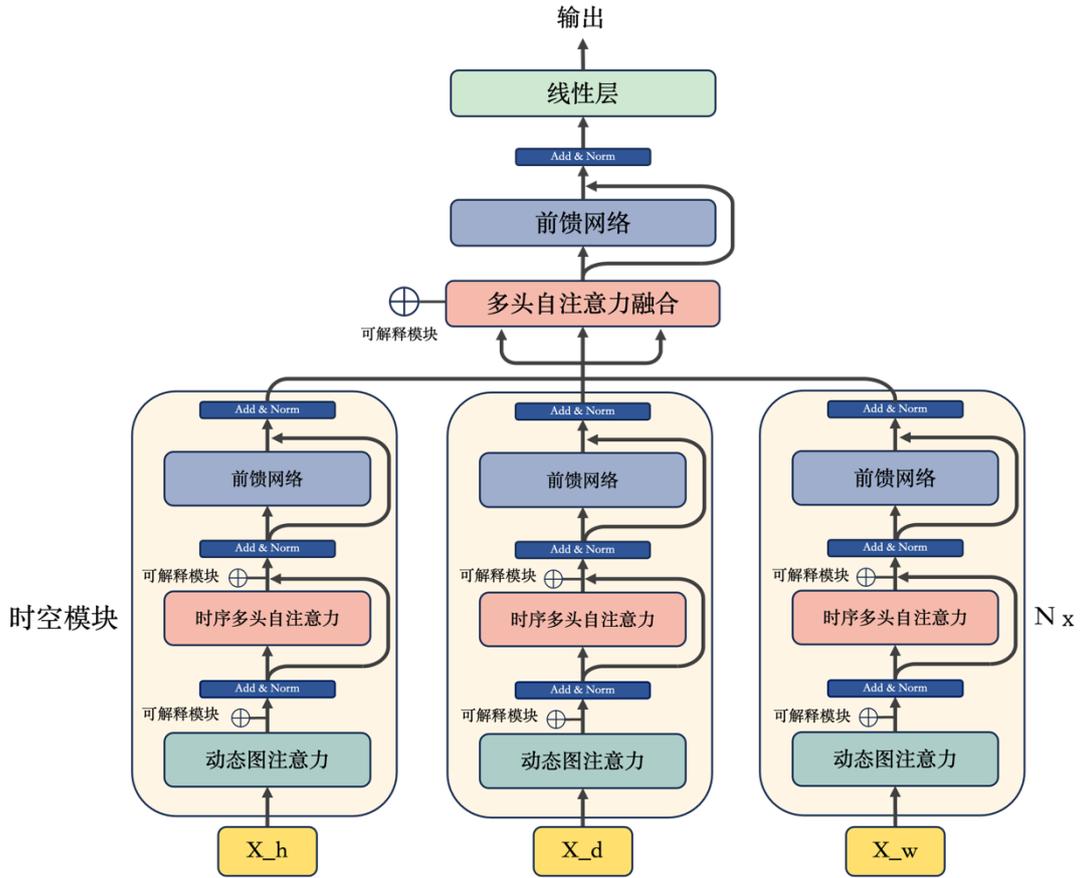


图 2 模型结构图

Orion 模型接收三周期数据输入（周周期特征 X_w 、日周期特征 X_d 和近期特征 X_h ），张量形状为 $\mathbb{R}^{B \times N \times F_{in} \times T}$ （其中 B 是批量大小， N 是节点（传感器）数量， F_{in} 是输入特征数量， T 是输入时间步长）。数据首先通过时空模块，该模块包含动态图注意力模块（捕捉空间依赖关系）、时序多头自注意力（建模长距离时间依赖）和前馈网络（增强非线性表达）。处理后的特征经多头自注意力融合模块动态融合，再通过前馈网络和线性层对局部的深度非线性特征进行捕捉。生成最终预测结果 $\mathbb{R}^{B \times N \times T_{pred}}$ （其中 T_{pred} 是预测时间步长）。模型各子模块采用残差连接、

层归一化（Add & Norm）[17]和丢弃机制（Dropout）[28]确保训练稳定性；每个注意力模块均集成可解释模块，通过可视化动态图注意力权重揭示空间、时间及周期交互的依赖模式。其中三周期数据输入、动态图注意力模块，时序多头自注意力模块，多头自注意力融合模块为本文在已有研究基础上改进得到，故做详细介绍。

(二)三周期数据输入

受 ASTGCN 模型启发，本文提出基于时间偏移的三周期输入设计，结合周周期、日周期和小时周期（近期数据）。相比仅依赖近期数据或简单滑动窗口的模型^[29]，该方法可捕捉更为复杂的时空动态。但与 ASTGCN 直接采样与预测目标完全一致时间段不同的，本文通过引入时间偏移避免对特定时间段的直接依赖，统一提取预测时刻前一小时的数据，显著降低过拟合风险。

对于输入数据 $X \in \mathbb{R}^{T \times N \times F_{in}}$ ，提取三种周期数据，形状均为 $X_h \in \mathbb{R}^{B \times N \times F_{in} \times T_h}$ ，如图 3 所示。

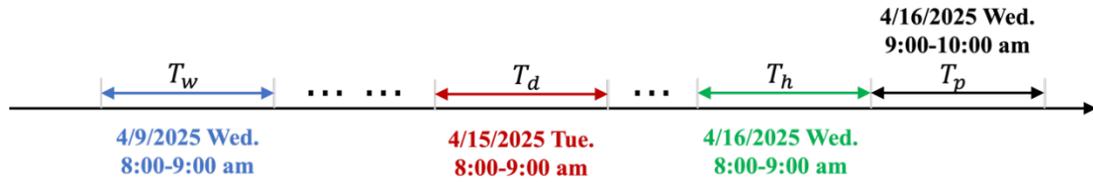


图 3 时间序列片段输入示例

1.小时数据 (X_h)

选取预测时刻前 T_h 个时间步，如图中绿色片段所示。旨在捕捉短期的流量突变，如突发事故或实时天气变化导致的流量骤增。

2.日周期数据 (X_d)

提取前一天同一时间段前 T_d 个时间步，如图中红色片段所示。旨在反映每日重复流量模式，如通勤高峰或午餐时间流量低谷。

3.周周期数据 (X_w)

提取前一周同星期同一时间段前 T_w 个时间步，如图中蓝色片段所示。旨在捕捉每周交通模式周期性规律，如工作日与周末的差异，这种周期性规律在长周期分析中尤为明显，前期筛选后输入模型更易被其识别。

(三)动态图注意力模块

本文提出了动态图注意力模块，其结构如图 4 所示。该模块将先验结构信息与后验学习数据相结合以优化计算效率与预测精度，借助基于 KNN 的加权邻接矩阵和权重约束机制筛选强相关节点，再经注意力机制动态调整权重。

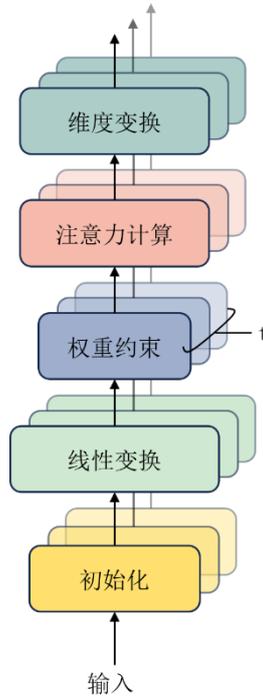


图 4 动态图注意力结构图

首先，该模块调用预先生成的基于 KNN 的加权邻接矩阵 $A_{ij} \in \mathbb{R}^{N \times N}$ （具体见 3-2-3）作为初始权重 W_{init} ；同时生成邻接掩码 $M \in \mathbb{R}^{N \times N}$ ，筛选每个节点的 K 个最近邻居，聚焦于强相关节点。同时，输入时间序列张量 $X \in \mathbb{R}^{B \times N \times F_{in} \times T}$ ，经过线性变换调整后与矩阵 $W \in \mathbb{R}^{F_{in} \times (H_s \cdot C)}$ 相乘得到张量：

$$H = WX_{perm} \in \mathbb{R}^{B \times T \times N \times (H_s \cdot C)} \quad \text{公式(1)}$$

并调整为 $H \in \mathbb{R}^{B \times T \times N \times H_s \times C}$ 。随后，采用权重约束机制对 W_{init} 的范围进行约束（本文将其约束在初始值的 $\pm 50\%$ ），并在每次前向传播后重新归一化。接着，借助注意力计算得到 $\alpha_{ij}^h \in \mathbb{R}^{B \times T \times N \times N}$ 。最终，借助维度调整通过加权聚合计算输出特征，并将多头结果拼接再调整，得到输出 $H_{out} \in \mathbb{R}^{B \times N \times (H_s \cdot C) \times T}$ 。其中注意力计算为本模块的核心部分，具体介绍如下。

4.注意力计算

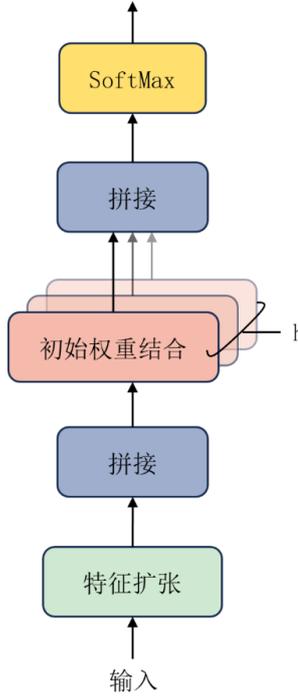


图 5 注意力计算模块结构图

如图 5 所示。首先，从特征张量 $H \in \mathbb{R}^{B \times T \times N \times H_s \times C}$ 中提取第 h 头的特征，得到 $H_h \in \mathbb{R}^{B \times T \times N \times C}$ 。接下来，在特征扩张步骤，将 H_h 分别扩展为两个张量通过广播操作进一步扩展为 $H_h^1 \in \mathbb{R}^{B \times T \times N \times N \times C}$ 和 $H_h^2 \in \mathbb{R}^{B \times T \times N \times N \times C}$ 。此操作的核心动机在于交通网络中的节点交互本质上是成对的，为计算节点 i 和 j 之间的注意力权重，需为每对节点构造对应的特征表示。随后，将 H_h^1 和 H_h^2 在第 4 维（特征维度）上拼接，生成张量：

$$e_h = \text{Concat}(H_h^1, H_h^2) \in \mathbb{R}^{B \times T \times N \times N \times 2C} \quad \text{公式(2)}$$

拼接操作通过组合节点 i 和 j 的特征，为注意力机制提供完整的联合特征表示。接着，在注意力分数计算步骤，利用可学习的注意力参数 $\mathbf{a}_h \in \mathbb{R}^{2C \times 1}$ 对拼接特征 \mathbf{e}_h 进行加权计算，得到注意力分数：

$$\mathbf{e}_{ij}^h = \text{LeakyReLU}(\mathbf{a}_h^T \mathbf{e}_h[:, :, i, j, :]) \in \mathbb{R}^{B \times T \times N \times N} \quad \text{公式(3)}$$

其中 $\mathbf{e}_h[:, :, i, j, :] \in \mathbb{R}^{2C}$ 是节点对 (i, j) 的拼接特征向量。LeakyReLU 激活函数通过在输入小于 0 时引入一个小的正斜率（通常为 0.01），使输出为 αx (α 为小正数)，而非直接输出 0，避免了其可能导致的“神经元死亡”问题。 \mathbf{a}_h 为可学习参数，模型自适应地学习哪些特征对节点交互影响更大，从而生成更有意义的注意力分数。进一步，为引入图结构的先验知识并结合后验信息，在初始权重结合步骤，本文将注意力分数 \mathbf{e}_{ij}^h 与邻接掩码 $\mathbf{M} \in \mathbb{R}^{N \times N}$ 和初始权重 $\mathbf{W}_{\text{init}} \in \mathbb{R}^{N \times N}$ 相结合，得到调整后的注意力分数为：

$$\mathbf{e}_{ij}^h = \mathbf{e}_{ij}^h \cdot \mathbf{M}[i, j] + \mathbf{W}_{\text{init}}[i, j] \in \mathbb{R}^{B \times T \times N \times N} \quad \text{公式(4)}$$

其中， $\mathbf{M}[i, j]$ 表示图的邻接关系（1 表示节点 i 和 j 相连，0 表示不相连）， $\mathbf{W}_{\text{init}}[i, j]$ 是预定义初始权重，用于引入基于历史数据的先验知识。交通网络作为图结构，不相邻的节点可能不直接交互（即 $\mathbf{M}[i, j] = 0$ ），因此通过邻接掩码 \mathbf{M} 可以屏蔽不合理的注意力分数，避免学习错误的交互模式；初始权重 \mathbf{W}_{init} 则在冷启动阶段利用先验知识，加速收敛并提升预测精度。

最终，对调整后的注意力分数 \mathbf{e}_{ij}^h 应用 SoftMax 函数，得到最终注意力权重 $\alpha_{ij}^h \in \mathbb{R}^{B \times T \times N \times N}$ 。SoftMax 归一化将注意力分数转化为概率分布，确保节点间交互权重之和为 1。

(四)时序多头自注意力模块

本文设计了时序多头自注意力模块，其结构如图 6 所示。相较于传统多头自注意力机制，本文针对交通流量预测数据优化了张量维度的处理方式，通过多次的维度变换确保了高效的并行计算。

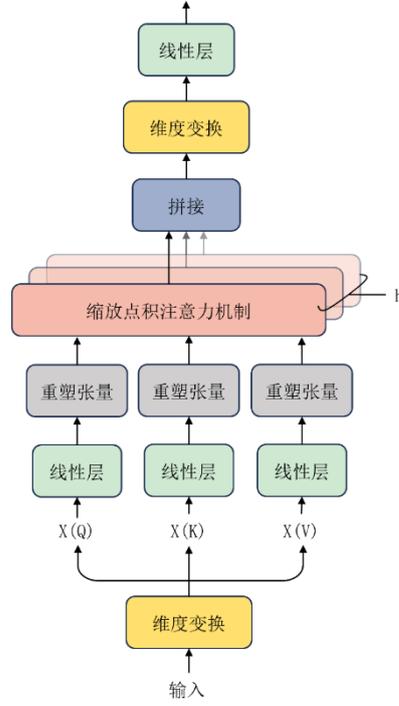


图 6 时序多头自注意力模块结构图

首先，输出张量经维度变换至 $X_{perm} \in \mathbb{R}^{B \times T \times N \times (H_s \cdot C)}$ ，并通过线性层生成查询、键和值张量：

$$Q = W_q X_{perm}, K = W_k X_{perm}, V = W_v X_{perm} \quad \text{公式(5)}$$

其中 $Q, K, V \in \mathbb{R}^{B \times T \times N \times (H_s \cdot C)}$ 。随后，通过重塑张量将 Q, K, V 分头为 $Q, K, V \in \mathbb{R}^{B \times N \times H_t \times T \times D_k}$ （其中 $D_k = (H_s \cdot C) / H_t$ ），以通过多头机制从不同子空间捕捉多样化依赖。接着，通过缩放点积注意力机制计算得到 $context \in \mathbb{R}^{B \times N \times H_t \times T \times D_k}$ ，并借助多头拼接和维度变换将结果调整为 $\in \mathbb{R}^{B \times T \times N \times (H_s \cdot C)}$ 以整合多头信息。最终，通过输出线性变换得到输出：

$$H_{out} = W_o context \in \mathbb{R}^{B \times N \times (H_s \cdot C) \times T} \quad \text{公式(6)}$$

其中 W_o 是可学习投影矩阵，稳定输出维度投影特征。模块内的缩放点积注意力机制为本模块核心部分，具体介绍如下。

1.缩放点积注意力机制

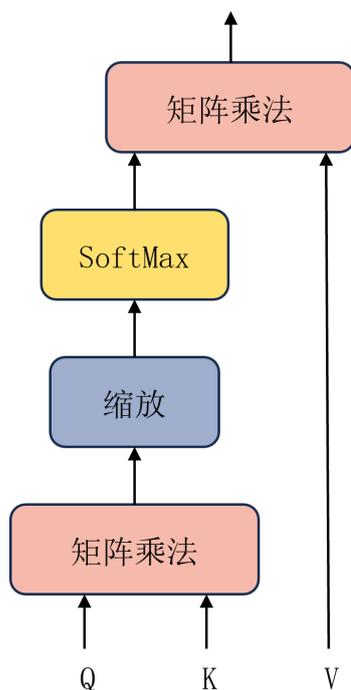


图 7 缩放点积注意力机制结构图

如图 7 所示，输入查询 Q 、键 K 和值 V 张量，形状为 $Q, K, V \in \mathbb{R}^{B \times N \times H_t \times T \times D_k}$ 。机制通过公式：

$$\text{context} = \text{softmax} \left(\frac{QK^T}{\sqrt{D_k}} \right) \cdot V \quad \text{公式(7)}$$

计算注意力权重并生成上下文表示。首先，点乘与缩放生成注意力分数 ($\mathbb{R}^{B \times N \times H_t \times T \times T}$)，缩放因子 $\sqrt{D_k}$ 缓解数值不稳定性。接着，SoftMax 归一化分数为概率分布。最终，加权融合值 V ，输出上下文表示 ($\mathbb{R}^{B \times N \times H_t \times T \times D_k}$)，自适应整合全局信息，避免信息丢失。

(五)多头自注意力融合模块

在完成多周期时空依赖关系的识别后，数据张量进入三周期信息融合阶段。本文设计了多头自注意力融合模块，通过多头自注意力机制实现多周期数据的动态融合，从而改进融合能力不足导致的信息丢失。其结构如图 8 所示。

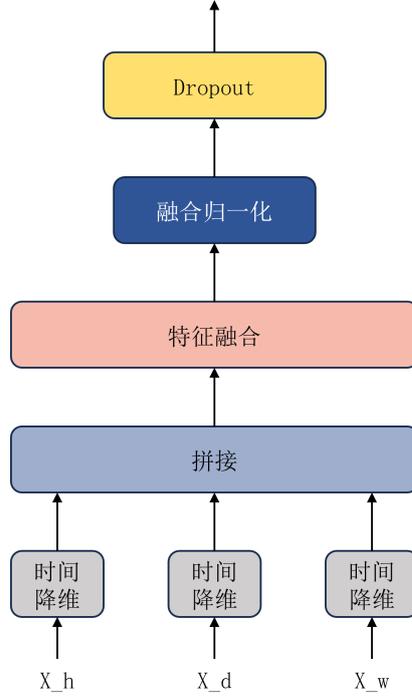


图 8 多头自注意力融合模块结构图

首先，三周期时空模块的输出 $X_h, X_d, X_w \in \mathbb{R}^{B \times N \times (H_s \cdot C) \times T}$ 通过时间降维模块，利用线性变换：

$$X_{\text{reduced}} = W X_{\text{reshaped}}, \quad W \in \mathbb{R}^{T \times 1} \quad \text{公式 (8)}$$

将时间维度压缩至 1，形状变为 $X_h, X_d, X_w \in \mathbb{R}^{B \times N \times (H_s \cdot C) \times 1}$ 。接着，数据张量沿时间维度拼接为 $X_{\text{fused}} \in \mathbb{R}^{B \times N \times (H_s \cdot C) \times 3}$ ，实现多周期数据的初步整合。随后， X_{fused} 进入特征融合部分，被分为多头自注意力路径和线性投影路径两条路径处理，分别输出 $X_{\text{fused_out}} \in \mathbb{R}^{B \times N \times C \times T_{\text{pred}}}$ 和 $X_{\text{proj}} \in \mathbb{R}^{B \times N \times C \times T_{\text{pred}}}$ ，两条路径输出借助融合归一化进行残差连接和层归一化，得到 X_{out} 。最终，Dropout 随机丢弃部分特征，输出 $X_{\text{out}} \in \mathbb{R}^{B \times N \times C \times T_{\text{pred}}}$ 。其中特征融合的多头自注意力路径和线性投影路径为本模块核心部分，具体讲解如下。

1.多头自注意力融合路径

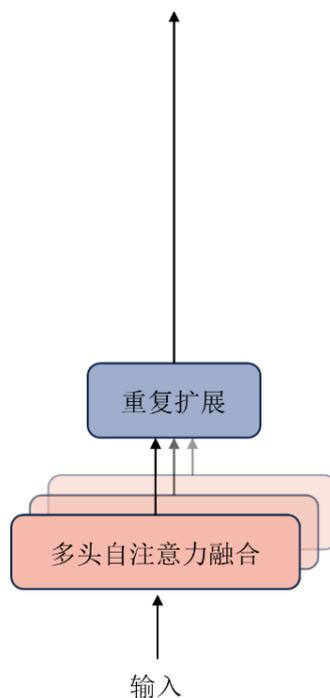


图9 多头自注意力融合路径图

如图9所示。具体而言，该路径接收融合特征 $X \in \mathbb{R}^{B \times N \times (H_s \cdot C) \times 3}$ ，在注意力融合步骤，进行多头自注意力计算（具体讲解见2-4），获取不同周期间交互权重。例如，在工作日早晨，模型可能赋予日周期特征（如常规早高峰模式）较高权重，而在节假日前夜提升周周期特征权重，输出增强特征 $X \in \mathbb{R}^{B \times N \times C \times 3}$ 。随后，通过扩展模块将特征扩展至预测长度 T_{pred} ，再经重复扩展获得最终输出：

$$X_{fused_out} = \text{Repeat}\left(X_{fused_out}, \frac{T_{pred}}{3}\right). \quad \text{公式(9)}$$

2.线性投影路径

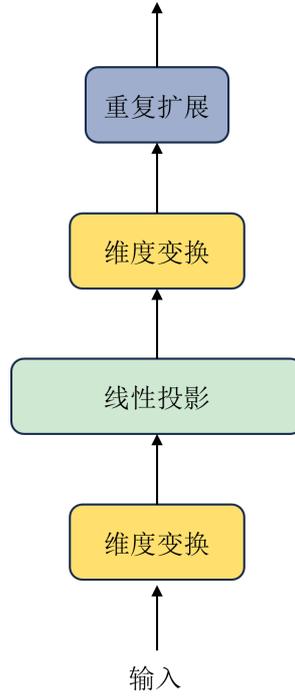


图 10 线性投影路径图

如图 10 所示。首先，通过维度调整重排输入至 $X \in \mathbb{R}^{B \times N \times 3 \times (H_s \cdot C)}$ 。接着，通过线性投影模块利用权重矩阵 $W_p \in \mathbb{R}^{(H_s \cdot C) \times C}$ 实现降维映射，例如将原始高维时空特征压缩为更加紧凑的表示形式，此方法在处理大规模路网时尤为重要，可以将数千个监测点采集的原始特征有效压缩，公式为：

$$X_{\text{proj}} = W_p X_{\text{perm}} \quad \text{公式(10)}$$

得到特征 $X \in \mathbb{R}^{B \times N \times 3 \times C}$ 。最终，经过第二次维度调整得到输出 $X \in \mathbb{R}^{B \times N \times C \times 12}$ 。

三、模型精度评价与应用

(一)数据来源及描述

本文使用的 PeMS03、PeMS04 和 PeMS08 来源于加州交通局的 PeMS^④数据集，分别对应加州不同区域的交通网络，包含流量、速度和占用率等特征。本文使用的 NYCTaxi(r)和 NYCBike2(r)来源于纽约市出租车数据集及纽约市自行车共享系统数据集。本文通过随机抽样生成子集以适配模型训练，标注为“(r)”。

各数据集均包含以下两类文件：空间连接数据文件，记录探测点间的连接关系，以三列形式（from、to、cost）表示（其中 from、to、cost 分别代表起点节点、终点节点和边距离），用于构建图神经网络的加权邻接矩阵；交通时间序列数据文件，以三维形状（T, N, F_{in}）存储，为模型训练和预测提供输入。具体统计信息见表 1。

表 1 数据集统计信息

数据集	步长	探测器数	特征数	特征名称	时间间隔	特征单位
PeMS03	26208	358	1	流量	5 分钟	辆/小时
PeMS04	16992	307	3	流量	5 分钟	辆/小时
				占用率		比例 (%)
				速度		英里/小时
PeMS08	17856	170	3	流量	5 分钟	辆/小时
				占用率		比例 (%)
				速度		英里/小时
NYCBike2(r)	11952	200	4	租车量	30 分钟	辆/小时
				还车量		辆/小时
				索引：半小时		无单位
				索引：星期		无单位
NYCTaxi(r)	11952	200	3	双向流量	30 分钟	辆/小时
				索引：半小时		无单位
				索引：星期		无单位

^④PeMS (Performance Measurement System) 数据集是由美国加州交通部门 (Caltrans) 开发并维护的交通流量监测数据集，广泛应用于交通流量预测、智能交通系统及相关研究领域。

为清楚的了解数据集特征值波动情况以提升模型训练效率和可信度，我们绘制了图 11。从图中可得出，PEMS03 和 PEMS08 的流量特征呈现显著周期性，PEMS04 的流量波动较无序，占用率和速度较平稳。NYCBike2(r)的租车量、还车量及索引特征均具周期性并伴随上升趋势。NYCTaxi(r)的流量特征表现为先降后升的高频周期性波动，时间索引特征较平稳。

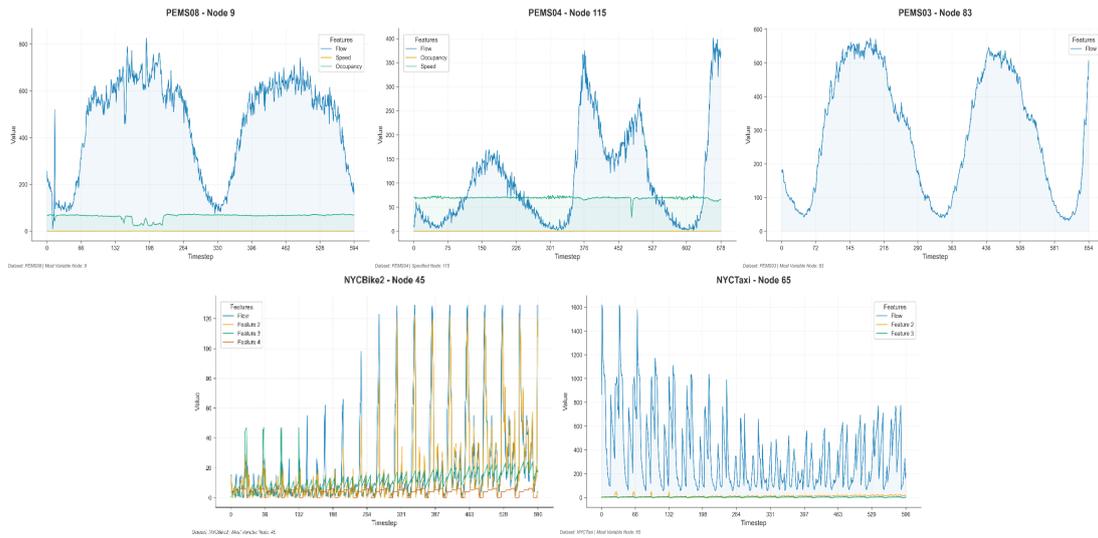


图 11 数据集随机节点特征值波动图

(二)数据处理

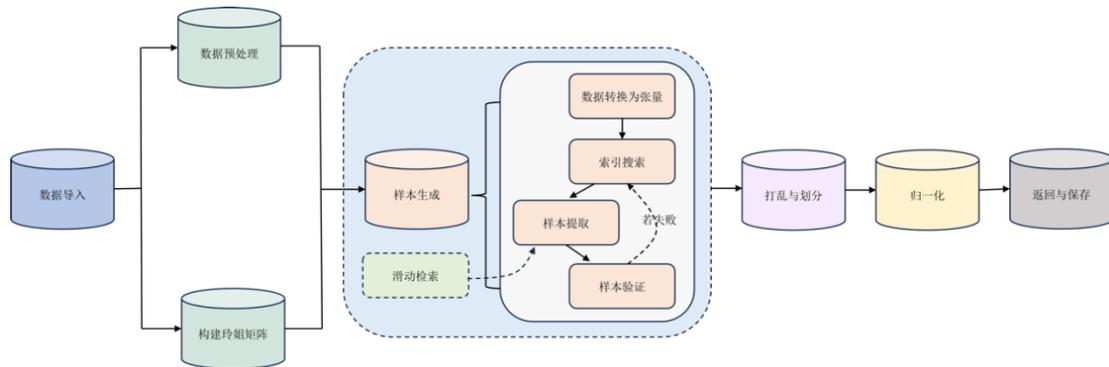


图 12 数据处理流程图

如图 12 所示。数据导入模块经交通时间序列数据文件导入时间序列张量 $X \in \mathbb{R}^{T \times N \times F}$ ，同时生成掩码张量 $M \in \mathbb{R}^{T \times N \times F}$ 以区分有效数据和无效数据，并在加权邻接矩阵构建和后续损失计算时传入，确保模型仅优化真实流量；经空间连接数据文件加载边列表，为后续构建加权邻接矩阵提供基础。数据预处理模块采用时间轴线性插值填补 X 中的零值与缺失值^[30]，并重塑张量为 $X' \in \mathbb{R}^{(N \times F) \times T}$ 。

邻接矩阵构建模块采用基于 KNN 方法构建加权邻接矩阵。样本生成模块通过多周期滑动检索策略生成输入张量 $X_w, X_d, X_h \in \mathbb{R}^{K \cdot T_{\text{input}} \times N \times F}$ （其中 K 表示为每个周期提取的子序列数量）和目标张量 $Y \in \mathbb{R}^{T_{\text{pred}} \times N \times 1}$ ，通过时间对齐和样本验证确保数据完整性。数据按 6:2:2 比例打乱并划分为训练、验证和测试集，所有数据及其特征值经标准化处理以消除量纲影响。其中邻接矩阵构建模块的基于 KNN 的加权邻接矩阵生成为本流程的核心模块，具体介绍如下。

1.基于 KNN 的加权邻接矩阵构建

本文提出了基于 KNN 策略的加权邻接矩阵构建方法，通过距离倒数动态分配权重，优先增强近邻节点的影响，并限制邻居数量以减少冗余连接，从而精确刻画局部空间依赖，降低计算复杂度。

首先，该方法对边列表中 cost 进行标准化以消除量纲影响。接着，对每个节点 i 及其邻居 j ，构建邻居列表，按 cost 升序排序并保留最近的 K 个邻居。对保留的邻居 j ，计算加权邻接矩阵的元素：

$$A_{ij} = \frac{1}{\text{cost}_{ij} + \epsilon}, \quad \epsilon = 10^{-6} \quad \text{公式(11)}$$

其中 ϵ 用于防止除零错误。对于不属于 i 的 K 个近邻的节点 j ，设置 $A_{ij} = 0$ ，随后对加权邻接矩阵进行归一化，确保后续模型稳定性。

(三)精度对比实验

1.任务概述

本文将所有数据集均按 6:2:2 的比例划分为训练集、验证集和测试集。设置输入序列长度为 12 ($T_{\text{input}} = 12$)，预测未来 12 个时间步 ($T_{\text{pred}} = 12$)。Orion 模型及基线模型在统一 GPU 环境训练，使用 Adam 优化器进行梯度更新，MAE 作为总体损失函数，训练 50 个 epoch。预测结果经反向归一化后逐时间步评估。

2.Orion 模型超参数解释、超参数调整实验结果及分析

本文以 PeMS04 数据集为基准，预先使用 10 个 epoch 进行了超参数调整实验，综合验证集的 MAE、RMSE、MAPE 及单 epoch 平均训练时间，筛选出了最

优超参数组合，从而提高全面训练的效率与效果。以下为模型中所有超参数解释（见表 2）及超参数调整实验结果（见表 3）。

表 2 Orion 模型超参数解释

超参数名称	超参数解释
KNN 邻居数 (KK)	控制 KNN 加权邻接矩阵每个节点的邻居数量，影响空间依赖建模
时空模块数量 (NB)	决定时空模块模块堆叠次数，影响时空特征捕捉深度
学习率 (LR)	控制参数更新步长，影响收敛速度与稳定性
空间多头注意力头数 (HS)	动态图注意力模块中多头注意力头数，影响空间特征子空间多样性
时间多头自注意力头数 (HT)	时序多头自注意力模块头数，影响时间依赖建模精度
融合多头自注意力头数 (HF)	多头自注意力融合头数，控制多周期特征融合灵活性
Dropout 率 (DP)	控制随机丢弃特征比例，防止过拟合
前馈网络隐藏层维度 (FH)	前馈网络隐藏层神经元数量，影响非线性变换能力

表 3 Orion 模型超参数调整实验结果

组号	KK	NB	LR	HS	HT	时间 (min:s/epoch)
1	3	2	0.0001	8	8	14:50
2	5	2	0.0001	8	8	15:42
3	3	3	0.0001	8	8	21:15
4	3	2	0.001	8	8	10:45
5	3	2	0.005	8	8	10:50
6	3	2	0.01	8	8	10:40
7	3	1	0.0001	8	8	5:42
8	3	2	0.0001	4	4	6:45
9	3	2	0.0001	4	8	8:20
10	3	2	0.0001	16	16	24:35:00
11	3	2	0.0001	8	8	8:15
12	3	2	0.0001	8	8	10:35

表 3 Orion 模型超参数调整实验结果 (续 1)

组号	HF	DP	FH	MAE	RMSE	MAPE	时间 (min:s/epoch)
1	4	0.2	768	23.765	28.487	21.689	14:50
2	4	0.2	512	24.312	29.156	22.345	15:42
3	4	0.2	512	23.765	28.432	21.678	21:15
4	4	0.2	512	26.123	31.245	24.567	10:45
5	4	0.2	512	28.456	33.789	26.123	10:50
6	4	0.2	512	NaN	NaN	NaN	10:40
7	4	0.2	512	27.234	32.567	25.456	5:42
8	4	0.2	512	25.678	30.987	23.892	6:45
9	4	0.2	512	24.987	29.876	22.978	8:20
10	8	0.3	1024	23.987	28.765	22.123	24:35:00
11	4	0.2	256	24.321	29.234	22.456	8:15
12	4	0.2	512	23.804	28.514	21.755	10:35

表 3 表明: KK=3 在高效捕捉局部空间依赖性方面表现最佳, 增至 5 会引入噪声并显著增加计算负担(+48.2%); NB=2 较好平衡精度与效率, 减至 1 会严重影响模型性能(MAE 等指标恶化 3.4-4.1), 增至 3 仅带来微小改善但计算成本翻倍; LR=0.0001 最适合 Orion 复杂结构, 较大学习率(尤其 0.01)导致梯度爆炸; HS=HT=8 在时空特征捕捉与计算效率间取得平衡, 减半会明显降低模型表现(MAE 等上升 1.9-2.5), 翻倍虽略有改善但训练时间激增(+132.2%); HF=4 足以实现多周期特征有效融合, 增大收益有限; FH=512 提供充分表达能力, 减半会限制模型性能, 增至 768 仅带来微小改善但计算成本增加 40%。综合评估, 组 1 参数配置在 MAE、RMSE、MAPE 和训练时间上达到最佳平衡, 适合大规模交通流量预测任务部署。

3.基线模型超参数设置

为确保实验公平性, 本文选用 LSTM、STGCN、DCRNN、ASTGCN、Graph-WaveNet 和 STID 六个具有代表性和可比性的模型作为基线模型。本文对基线模型进行了超参数调优, 确保其在与 Orion 模型参数规模相当的条件达到其性能上限, 实现公平且有说服力的对比评估。各基线模型及其主要超参数的解释与设置如下表所示。

表 4 LSTM 模型超参数解释与设置

超参数名称	超参数解释	设置
隐藏层大小 (HS)	LSTM 单元中隐藏状态的维度, 影响模型记忆容量	512
层数 (NL)	堆叠的 LSTM 层数, 增加模型深度和复杂度	2
丢弃率 (DR)	——	0.2
学习率 (LR)	——	0.0001

表 5 STGCN 模型超参数解释与设置

超参数名称	超参数解释	设置
空间卷积核大小 (KS)	空间卷积感受野大小, 决定节点获取邻居信息范围	3
时间卷积核大小 (KT)	时间卷积的感受野大小, 影响时序模式建模能力	3
时间间隔 (TI)	数据采样时间间隔, 影响模型对时间序列的理解	5
学习率 (LR)	——	0.0001

表 6 DCRNN 模型超参数解释与设置

超参数名称	超参数解释	设置
学习率衰减步数 (LDS)	控制学习率衰减, 隔步数调整学习率, 影响收敛稳定	2000
预测时间步长 (HOR)	模型预测的未来时间步数, 决定预测范围	12
最大扩散步数 (MDS)	扩散卷积的最大跳数, 决定图信息传播范围	2
RNN 层数 (NRL)	堆叠的循环神经网络层数, 增加模型深度	2
RNN 单元数 (RNU)	隐藏状态维度, 影响模型的容量	64
丢弃率 (DR)	——	0.2
学习率 (LR)	——	0.0001

表 7 ASTGCN 模型超参数解释与设置

超参数名称	超参数解释	设置
块数量 (NB)	时空注意力块的堆叠数量, 提取多层次特征	2
切比雪夫多项式阶数 (K)	控制图卷积感受野, 决定邻居信息范围	3
切比雪夫卷积滤波器数量 (NCF)	控制图卷积层特征维度, 影响特征提取能力	64
时间卷积滤波器数量 (NTF)	决定时间卷积层特征维度, 影响时序特征建模	64
学习率 (LR)	——	0.0001

表 8 Graph-WaveNet 模型超参数解释与设置

超参数名称	超参数解释	设置
隐藏单元数 (NHU)	隐藏层维度, 影响模型表达能力	32
权重衰减率 (WDR)	L2 正则化系数, 控制权重大小, 防止过拟合	0.0001
丢弃率 (DR)	——	0.2
学习率 (LR)	——	0.0001

表 9 STID 模型超参数解释与设置

超参数名称	超参数解释	设置
嵌入维度 (ED)	特征嵌入维度, 影响表示能力	32
层数 (NL)	堆叠的层数, 增加模型深度和复杂度	3
节点维度 (ND)	节点特征嵌入维度, 影响空间特征表示	32
时间维度-时间 ID (TDT)	时间 ID 编码的维度, 影响时间特征建模	32
时间维度-星期 (TDD)	星期几编码维度, 增强周期性建模	32
权重衰减率 (WD)	L2 正则化系数控制权重大小, 防止过拟合	0.0001
学习率衰减率 (GAM)	学习率调度器衰减系数, 影响优化稳定性	0.5
梯度裁剪阈值 (MN)	梯度裁剪最大范数值, 防止梯度爆炸	5
学习率 (LR)	——	0.001

4.精度对比实验结果及其分析

为全面评估 Orion 模型在交通流量预测任务中的性能，本文计算了 12 步预测值的平均评估指标如表 10 所示，并通过折线图展示各步预测误差趋势如图 13。

表 10 不同模型在五种数据集上的交通流量预测性能比较

Model	PeMS03			PeMS04		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
LSTM	29.63	29.18%	45.13	27.34	21.08%	42.09
STGCN	20.86	17.54%	33.32	36.84	62.96%	48.65
DCRNN	19.5	20.29%	31.01	27.99	20.58%	42.22
ASTGCN	20.7	19.14%	37.33	23.59	15.43%	37.09
Graph-WaveNet	17.63	18.93%	28.3	24.22	18.82%	37.37
STID	17.27	19.23%	27.5	26.91	21.75%	40.55
Orion	15.64	17.17%	23.21	20.31	18.79%	29.45

表 10 不同模型在五种数据集上的交通流量预测性能比较（续 1）

Model	PeMS08			NYCTaxi(r)		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
LSTM	21.74	14.36%	34.4	18.77	—	62.07
STGCN	25.89	19.41%	37.1	35.04	105.66%	98.29
DCRNN	23.13	16.69%	34.62	56.33	133.85%	138.66
ASTGCN	17.06	11.44%	26.7	39.98	70.79%	98.3
Graph-WaveNet	19.03	14.88%	29.78	49.22	76.42%	122.4
STID	21.99	15.35%	33.57	51.69	121.11%	124.43
Orion	15.18	11.40%	21.98	12.69	68.44%	34.07

表 10 不同模型在五种数据集上的交通流量预测性能比较（续 2）

Model	NYCBike2(r)		
	MAE	MAPE	RMSE
LSTM	1.68	—	6.2
STGCN	3.44	121.84%	8.47
DCRNN	8	78.15%	15.51
ASTGCN	1.47	16.67%	5.5
Graph-WaveNet	7.14	82.67%	14.07
STID	7.17	84.68%	13.96
Orion	2.88	42.86%	4.44

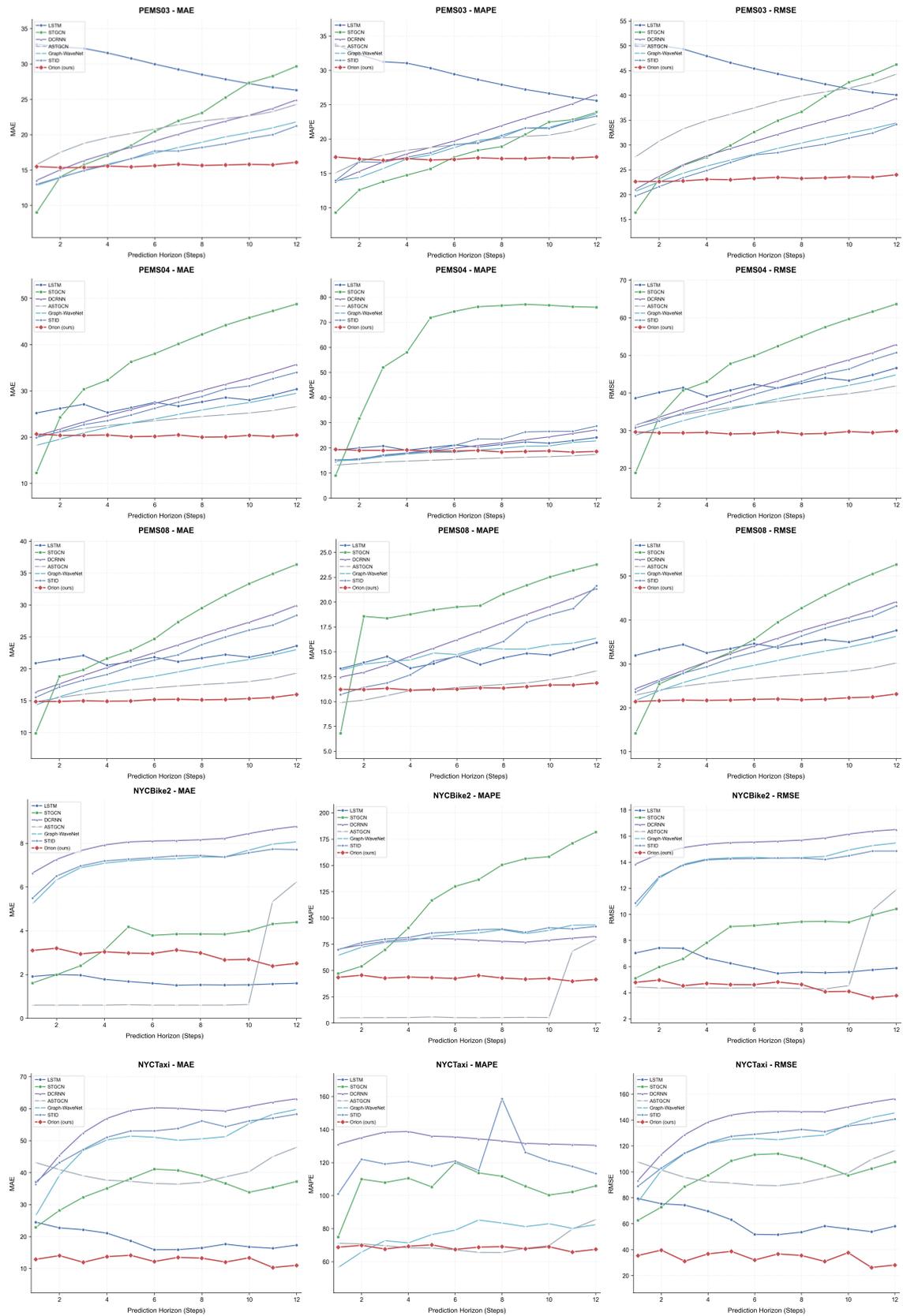


图 13 不同模型在五种数据集上交通流量预测性能随预测区间增大的比较

由表 13 及图 10 分析可得:时序模型 LSTM 在交通预测中表现受限,PeMS03 的 MAE 为 29.63, RMSE 为 45.13, NYCTaxi(r)的 RMSE 高达 62.07, 长期预测 (8-12 步) 误差累积明显, 因未有效捕捉交通数据的空间相关性, 如 PeMS03 中上游拥堵对下游流量的传播效应。STGCN 和 DCRNN 作为时空图卷积模型, 在 PeMS03 表现优于 LSTM(MAE 分别为 20.86 和 19.50, RMSE 为 33.32 和 31.01), 但在 NYCTaxi(r)城市场景中性能下降 (STGCN 的 MAPE 达 105.66%, DCRNN 的 RMSE 为 138.66), 或许与静态图建模难以适应动态事件 (如节庆活动) 驱动流量突增, 且 DCRNN 在长序列预测中误差累积明显有关。基于注意力机制的模型如 ASTGCN、Graph-WaveNet 和 STID 整体表现更优, ASTGCN 在 NYCBike2(r) MAE 为 1.47, MAPE 为 16.67%, 但在 NYCTaxi(r) MAPE 仍高达 70.79%; Graph-WaveNet 和 STID 在 PeMS03 的 MAE 分别为 17.63 和 17.27, RMSE 为 28.30 和 27.50, 但在 NYCTaxi(r)误差较高 (RMSE 分别为 122.40 和 124.43), 因动态依赖建模不足。Orion 模型表现最为突出, 在 PeMS03、PeMS04、PeMS08 和 NYCTaxi(r)的 MAE 和 RMSE 均最优, 且长期预测误差增长最缓, 稳定性最佳。其通过动态图注意力机制和多头自注意力融合模块, 有效捕捉复杂时空依赖和周期性规律, 如 PeMS08 中高峰期流量的动态扩散。虽在 NYCBike2(r) 略逊于 ASTGCN, 可能与数据周期性强而导致 Orion 稍显复杂有关, 但其在多场景的稳健性和对动态场景的适应性为智能交通系统提供了可靠支持。

(四)模型应用

1.数据获取

本研究采用的 Taxi BJ 数据集^⑤。由于原始数据量过大, 为提高训练和预测效率, 本文从中随机抽取部分数据, 记为 Taxi BJ(r)。选取以下四个特征进行建模与预测: 流入量、归一化温度、归一化风速及天气索引, 具体信息如表 11 及图 14 所示。流量特征基本表现出周期性, 气温、风速及天气索引基本平稳。

^⑤ TaxiBJ 数据集来源于 Liang Y, Ouyang K, Jing L, et al. Urbanfm: Inferring fine-grained urban flows[C]/Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 3132-3142.

表 11 Taxi BJ 数据集统计信息

数据集	步长	探测器数	特征数	特征名称	时间间隔	特征单位
Taxi_BJ	14163	128	4	流入量	30 分钟	辆/小时
				温度（归一化）		无单位（归一化值）
				风速（归一化）		无单位（归一化值）
				天气索引		无单位（索引）

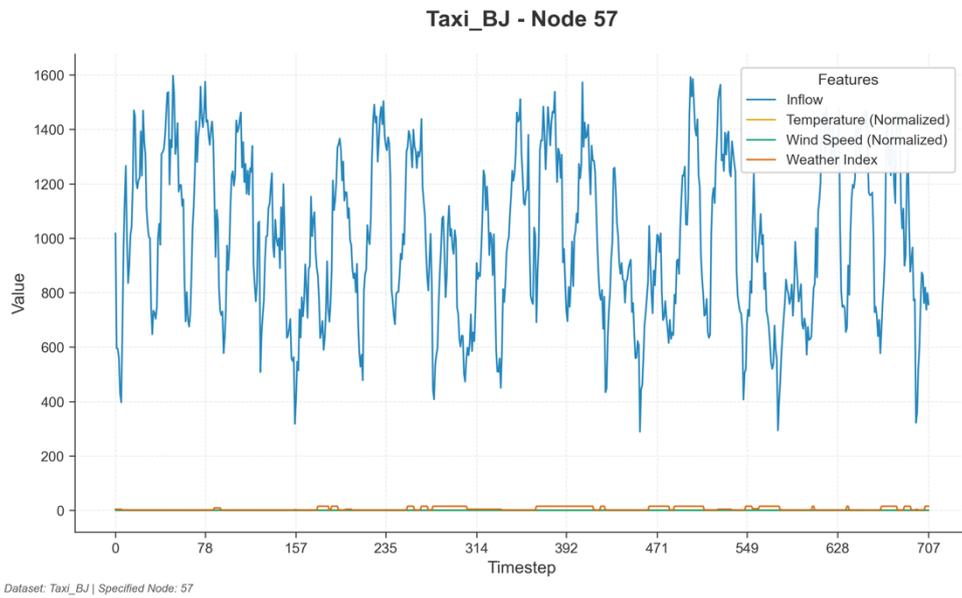


图 14 Taxi BJ 随机节点特征值波动图

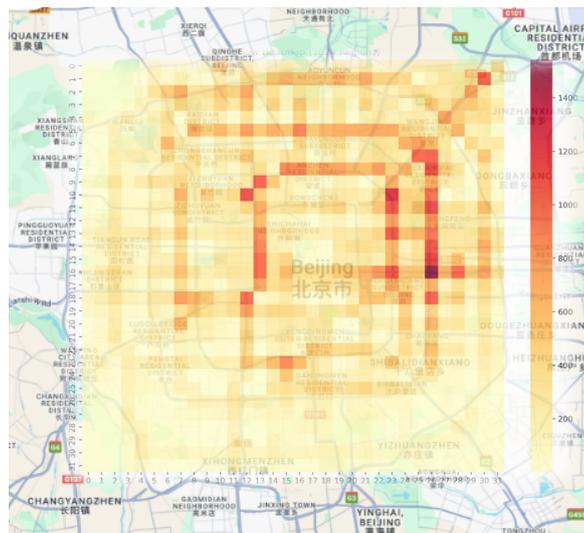


图 15 Taxi BJ 随机节点流量特征值热力图

图 15 热力图与北京主要交通干线基本吻合，验证 Orion 模型在预测精度与解释性方面具有双重优势。

2.实验描述

本文采用统一的超参数设置，对北京数据集进行训练与测试，如表 12 所示，Orion 以最低的 MAE (34.51)、RMSE (59.78) 和第二低的 MAPE (46.86%) 展现出较优的预测精度，优于其他基准模型。该结果证实了基于多头自注意力机制的时空图注意力神经网络在交通流量预测任务中的有效性。

表 12 Taxi BJ 实验结果对比

Model	MAE	MAPE	RMSE
LSTM	51.89	52.36%	90.88
STGCN	65.4	112.90%	93.72
DCRNN	50.72	60.13%	86.4
ASTGCN	39.31	43.43%	72.02
Graph-WaveNet	48.58	54.65%	82.39
STID	52.72	72.84%	86.37
Orion	34.51	46.86%	59.78

借助可视化分析如图 16 所示，Orion 模型的 MAE、MAPE 和 RMSE 指标均保持稳定，且数值显著低于其他模型。这表明 Orion 在长期预测中具备卓越的精确性和稳定性。

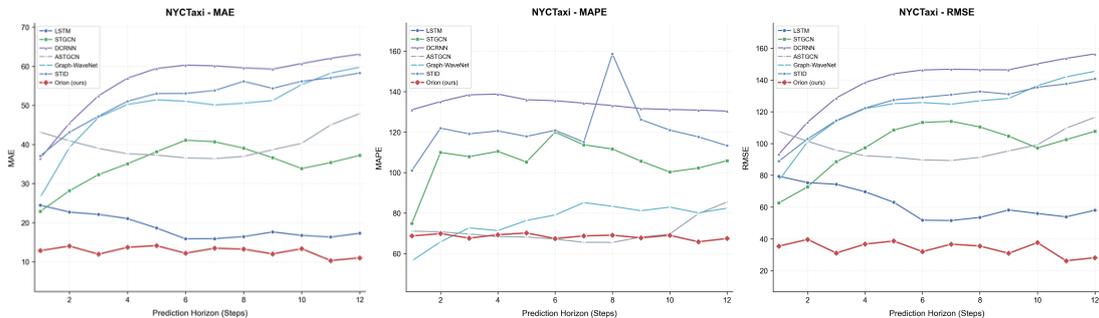


图 16 不同方法在 Taxi BJ 数据集上的性能随预测区间增大的比较

3.结果分析

为验证实验效果，本文随机选取节点 73 的预测结果与实际数据进行比对，如表 13 及图 17 所示。结果显示，Orion 模型的 MAE、MAPE 和 RMSE 指标均保持稳定，且数值显著低于其他模型。这表明 Orion 在长期预测中具备卓越的精确性和稳定性。

表 13 流量预测值与真值对比

数据类型	第 1 步	第 2 步	第 3 步	第 4 步	第 5 步	第 6 步	第 7 步	第 8 步
真实值	18	15	14	12	14	15	9	11
预测值	18.89	17.688	17.663	16.101	16.465	16.269	15.822	10.537

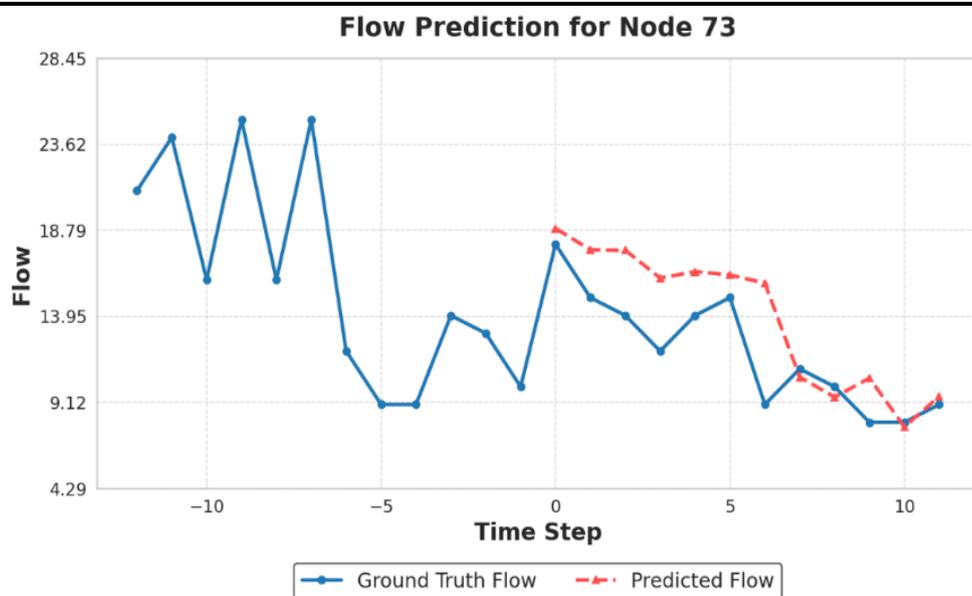


图 17 流量预测值与真值对比折线图

为进一步揭示 Orion 模型捕获的时块热力图输出进行了分析。在空间维度上，如图 18 所示，节点 56 与节点 90 相关性较高，权重约 0.016，而与节点 65 相关性较弱，权重约 0.008。这与实际网络中节点间距离（56 与 90 距离 12，56 与 65 距离 24）呈正相关，表明距离越近，相关性越强。

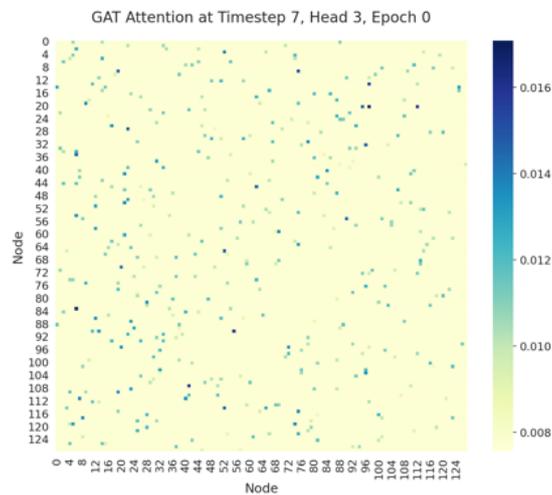


图 18 周周期动态图注意力模块节点间权重热力

在时间维度上，图 19 显示，图中第 6 列和第 10 列呈现出明显的高权重带，约 0.08345，表明该时间步为模型重点关注的交通流量变化关键点。此外，第 11 行与第 10 列交叉点的高权重反映了时间步 11 与 10 之间的强相关性，可能对应交通数据的连续性规律。热力图对角线附近的高权重分布进一步证明模型对相邻时间步的敏感性，同时通过远程依赖（如第 11 行与第 10 列的高权重区域）捕捉长期时序模式。

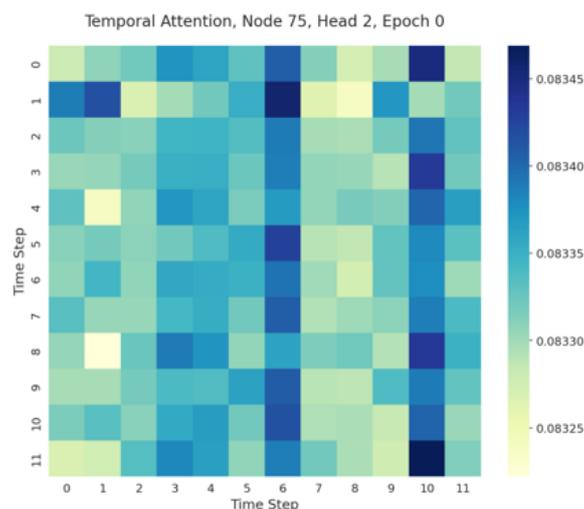


图 19 周周期时序多头自注意力模块时间步间权重热力图

图 20 表明，日和周周期的自相关权重较高（分别为 0.325 和 0.375），而小时周期权重较低。这反映了模型在融合时更依赖日周期和周周期的特征，可能是由于 Taxi BJ(r)数据集的周期性规律较强，而 Recent 周期的短期波动对长期预测贡献有限。

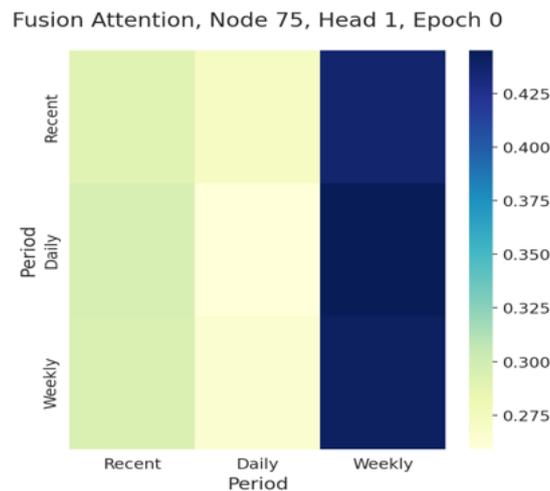


图 20 多头自注意力融合模块周期间权重热力图

基于上述可视化分析结果，本文提出以下政策性建议：图 20 表明，在交通流量预测与规划中，应重点关注周期性规律，特别是周周期影响（周周期权重高达 0.375，明显高于日和小时周期）。具体而言，城市交通管理部门可利用这一发现，制定针对性的交通疏导策略，如在每周固定时段调整红绿灯配时方案。例如，综合多张图表，结合图 20，对于预测本周四 9:00-10:00 任务，应重点关注周周期数据；结合图 19 中第 6 列（约 8:30）和第 10 列（约 8:50）的高权重带（权重约 0.08345），可重点利用上周四 8:30-8:50 期间的流量数据（即预测时段的前 10-30 分钟），预测并调整预测 9:00-10:00 的信号灯配时方案；再结合图 18 显示的节点间强弱关系，该政策特别应在第 56 节点与第 90 节点这类高相关性区域（权重约 0.016）的交叉路口或探测点实施。此外，由于模型捕捉到时间步 11 与时间步 10 之间的强相关性，建议根据上周同时段流量数据提前调整可变车道方向，实施动态车道管理措施，提高道路通行效率。同时，鉴于第 56 与第 65 节点间相关性较弱（权重约 0.008），可考虑在这类区域开通快速公交专线，减轻个体出行压力，从而构建更加高效、智能的城市交通网络系统。

四、结论与展望

(一)结论

1.精度对比实验

Orion 模型在交通流量预测中表现出色,显著优于五种基线模型,在 PeMS03、PeMS04、PeMS08、NYCTaxi(r)、NYCBike(r)数据集上三种指标均较低,且长期预测误差增长缓慢,稳定性强。

2.模型应用

Orion 以最低 MAE (34.51) 和 RMSE (59.78) 展现优异精度。可解释模块热力图分析揭示了周周期对北京交通预测的重要性,经多图分析,提出城市交通管理可利用周周期规律和节点间强相关性,优化信号灯配时和动态车道管理,构建高效智能交通网络的建议。

(二)展望

1.应用领域扩展

由于 Orion 模型为通用图结构时空预测框架,故其还可适用于其余时空依赖场景:如到达时间预估(ETA)中,图建模路网拓扑,时序预测交通动态,精准估算到达时间;再如能源消耗预测中,图分析电网节点关系,时序优化负荷模式,提升能源分配效率。

2.短期预测效果欠佳

Orion 模型在对短期交通流量预测的结果上,并未表现出如长期一样显著的优越性。未来可考虑借助简化模型结构,考虑用 GRU 替代 LSTM 或优化注意力机制,考虑用 Informer 替代 Transformer,从而使模型兼顾长短期高精度预测。

3.非结构化数据处理能力不足

Orion 模型对于非结构化数据(如社会突发事件)的识别和处理能力不足。若能结合大语言模型进行初步分析,可以从新闻报道、社交媒体和政策文件中提取影响交通流量的隐含因素。或可提高预测准确性和适应性。

4.预测逻辑改进

当前模型采用一次性预测全部时间序列的方法，这限制了模型对长序列预测的稳定性和准确性。通过引入一步预测策略，即利用前一刻的预测结果作为下一时刻的输入，或可更好地捕捉时序依赖关系，提升预测精度和稳定性。

参考文献

- [1] 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): 664-72.
- [2] ZIVOT E, WANG J. Vector autoregressive models for multivariate time series [J]. *Modeling financial time series with S-PLUS®*, 2006: 385-429.
- [3] COVER T, HART P. Nearest neighbor pattern classification [J]. *IEEE transactions on information theory*, 1967, 13(1): 21-7.
- [4] JEONG I-Y, TYAGI K, LEE K. MIREX 2013: An Efficient Paradigm for Audio Tag Classification Using Sparse Autoencoder and Multi-Kernel SVM [J]. 2013.
- [5] LESHEM G, RITOV Y. Traffic flow prediction using adaboost algorithm with random forests as a weak learner; proceedings of the Proceedings of world academy of science, engineering and technology, F, 2007 [C]. Citeseer.
- [6] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. *Neural computation*, 1997, 9(8): 1735-80.
- [7] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv preprint arXiv:14123555, 2014.
- [8] BAI L, YAO L, KANHERE S S, et al. Passenger demand forecasting with multi-task convolutional recurrent neural networks; proceedings of the Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part II 23, F, 2019 [C]. Springer.
- [9] ZHANG J, ZHENG Y, QI D, et al. Predicting citywide crowd flows using deep spatio-temporal residual networks [J]. *Artificial Intelligence*, 2018, 259: 147-66.
- [10] HUANG C-J, KUO P-H. A deep CNN-LSTM model for particulate matter (PM_{2.5}) forecasting in smart cities [J]. *Sensors*, 2018, 18(7): 2220.

- [11] BERG R V D, KIPF T N, WELLING M. Graph convolutional matrix completion [J]. arXiv preprint arXiv:170602263, 2017.
- [12] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs [J]. arXiv preprint arXiv:13126203, 2013.
- [13] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering [J]. Advances in neural information processing systems, 2016, 29.
- [14] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting [J]. arXiv preprint arXiv:170904875, 2017.
- [15] CUI Z, HENRICKSON K, KE R, et al. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(11): 4883-94.
- [16] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting [J]. arXiv preprint arXiv:170701926, 2017.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.
- [18] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [J]. arXiv preprint arXiv:171010903, 2017.
- [19] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting; proceedings of the Proceedings of the AAAI conference on artificial intelligence, F, 2019 [C].
- [20] LIANG Y, KE S, ZHANG J, et al. Geoman: Multi-level attention networks for geo-sensory time series prediction; proceedings of the IJCAI, F, 2018 [C].
- [21] GUO S, LIN Y, WAN H, et al. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(11): 5415-28.

- [22] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling [J]. arXiv preprint arXiv:190600121, 2019.
- [23] SONG C, LIN Y, GUO S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting; proceedings of the Proceedings of the AAAI conference on artificial intelligence, F, 2020 [C].
- [24] SHAO Z, ZHANG Z, WANG F, et al. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting; proceedings of the Proceedings of the 31st ACM international conference on information & knowledge management, F, 2022 [C].
- [25] HUANG R, HUANG C, LIU Y, et al. LSGCN: Long short-term traffic prediction with graph convolutional networks; proceedings of the IJCAI, F, 2020 [C].
- [26] LI M, ZHU Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting; proceedings of the Proceedings of the AAAI conference on artificial intelligence, F, 2021 [C].
- [27] ZHENG C, FAN X, WANG C, et al. Gman: A graph multi-attention network for traffic prediction; proceedings of the Proceedings of the AAAI conference on artificial intelligence, F, 2020 [C].
- [28] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. The journal of machine learning research, 2014, 15(1): 1929-58.
- [29] JIANG J, HAN C, ZHAO W X, et al. Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction; proceedings of the Proceedings of the AAAI conference on artificial intelligence, F, 2023 [C].
- [30] WANG Y, ZHANG J, ZHU H, et al. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics; proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, F, 2019 [C].

附录

本文全部项目文件均上传至百度网盘：

文件名字：数据及其他—TJJM20250303014424

链接：<https://pan.baidu.com/s/1GRQdZTXfJYumxgGNeFyv7w?pwd=xfp4>

提取码：xfp4