

LSFormer: 用于交通流预测的负载量感知空间 异质性变换器

李轩, 李艳红*, 徐昊翔, 黄健翔, 陈亮亮

(中南民族大学 计算机学院, 武汉 430074)

摘要 高精度的交通流预测可以有效缓解智能城市道路的拥堵压力. 然而, 交通流预测面临着如何有效揭示交通流数据中隐藏的时空依赖关系的挑战. 目前大多数方法都是基于图神经网络(GNN)或变压器模型. 前者只考虑短程空间信息, 无法捕捉长程空间依赖关系, 而后者虽然能够捕捉长程依赖关系, 但大多数研究都没有充分挖掘变压器架构的潜力. 为此, 提出了一种用于交通流预测的新型负载感知空间异质性变换器, 即LSFormer. 具体来说, 为空间自注意力模块设计了相对位置编码以优化空间位置信息感知问题, 使模型能更好地捕捉位置信息. 然后, 引入了负载感知模块, 以突出周边交通流对中心点的影响, 解决了现有方法对周边区域依赖关系建模不足的问题. 在5个真实世界公共交通数据集上的广泛实验结果表明: 文中所提模型可以达到先进的性能. 此外, 还将学习到的空间嵌入可视化, 使模型具有可解释性.

关键词 交通流预测; 时空特征; 变换器; 图神经网络

中图分类号 TP399.41 文献标志码 A 文章编号 1672-4321(2026)01-0086-11

doi:10.20056/j.cnki.ZNMDZK.20250830

LSFormer: workload-aware spatial heterogeneity transformer for traffic flow prediction

LI Xuan, LI Yanhong*, XU Haoxiang, HUANG Jianxiang, CHEN Liangliang

(College of Computer Science, South-Central Minzu University, Wuhan 430074, China)

Abstract Highly accurate traffic flow prediction can effectively alleviate the congestion pressure on smart city roads. However, traffic flow prediction faces the challenge of how to effectively reveal the hidden spatio-temporal dependencies in traffic flow data. Most of the current methods are based on graph neural network (GNN) or transformer models. The former only considers short-range spatial information and is unable to capture long-range spatial dependencies, while the latter is able to capture long-range dependencies, but most of the studies have not fully exploited the potential of the transformer architecture. To this end, a novel load-aware spatial heterogeneity transformer for traffic flow prediction, LSFormer, is proposed. Specifically, relative position coding is designed for the spatial self-attention module to optimize the spatial location information perception problem so that the model can better capture location information. Then, a load-aware module is introduced to highlight the influence of peripheral traffic flow on the centroid, which solves the problem of insufficient modeling of peripheral area dependencies by existing methods. Extensive experimental results on five real-world public transportation datasets show that the proposed model in this paper can achieve state-of-the-art performance. In addition, the learned spatial embeddings are visualized to make the model interpretable.

Keywords traffic flow prediction; spatio-temporal features; transformer; graph neural network

收稿日期 2024-12-19

* 通信作者 李艳红(1973-), 女, 教授, 博士, 研究方向: 时空数据处理, 网络大数据, DB与AI融合, E-mail: liyanhong@mail.scuec.edu.cn

基金项目 湖北省自然科学基金资助项目(2017CFB135); 中央高校基本科研业务费专项资金资助项目(CZY23019); 网络创新及应用型人才课程实践教学研究项目(2019年第一批)

智慧城市可以利用信息和通信技术(ICTs, Information and Communication Technologies)^[1]提高城市运行效率,改善居民生活质量.随着城市化进程的加快,现代城市面临着诸多挑战,其中交通管理是一个重要方面.智慧城市利用智能交通系统(ITS, Intelligent Transportation Systems)来解决交通管理问题,其中交通流预测是关键技术之一.交通流预测利用历史交通数据和实时信息来预测道路或交通节点的未来交通流.这项技术有助于优化交通规划和基础设施建设,从而改善交通拥堵状况.通过预测拥堵发生的时间和地点,可以采取交通引导和路线优化等措施来降低拥堵发生的可能性.随着技术的不断进步和数据的不断积累,交通流预测在智慧城市中的重要性将日益凸显.

最近的研究表明,使用各种深度学习技术进行交通流预测取得了重大进展,其中包括通过图神经网络捕捉空间依赖关系的STGCN^[2],在输入嵌入方面取得进展的STAEformer^[3],以及应用注意力机制的PDFormer^[4].尽管在改进交通流预测方面做出了巨大努力,但这些模型仍存在一些关键局限.例如,基于GNN(Graph Neural Network)模型在过度平滑问题上举步维艰,即在堆叠多层后,节点的表示趋于一致,导致难以捕获长距离的空间依赖关系.这种局限性显著限制了模型在复杂交通网络中的表现.

例如,STGCN仅通过一维卷积捕获空间和时间特征,对长程依赖的建模能力不足.虽然PDFormer在时间和空间方面都有更详细的构造,但其并行架构可能会导致时间和空间的脱节,从而有可能忽视时空信息的整体一致性.此外,在时间方面,STAEformer提供了丰富的嵌入信息.然而,在空间方面,仅仅依靠随机初始化的自适应嵌入很难取得理想的结果,并导致模型的实验结果出现明显的波动.

首先,交通系统中不同位置之间的空间依赖关系是密切相关的,而不是孤立的.例如,如图1(a)所示,传感器146和其他3个传感器(92、108、148)在空间位置上都直接相邻,但它们的趋势和数值显然有很大不同.空间上的相邻并不绝对影响其相邻节点,但考虑到传感器146所有相邻点的流量,并对它们进行运算后便获得了负载流量(图1(a)中红色曲线称为负载流量),可以看见在数值上负载流量更接近中心传感器的测量值.因此在建模时有必要考虑这一因素对预测任务的影响.然而,现有方法主要以孤立的方式模拟空间依赖关系,可能会忽略一些时空信息,限制了模型的预测能力.

其次,考虑到城市功能分区的影响,不同分区的传感器捕捉到的交通模式表现出明显的差异.如图1(b)所示,传感器146和85位于两个不同的城市

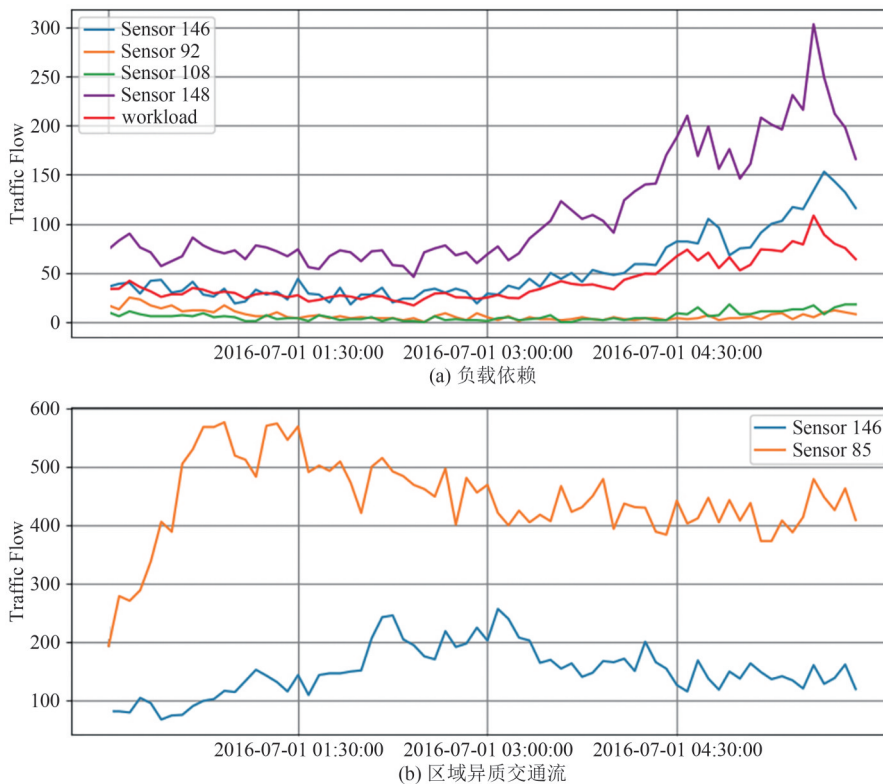


图1 交通流数据示例(传感器146、92、108和148空间上直接相邻)

Fig. 1 Examples of traffic flow data (Sensors 146, 92, 108 and 148 are directly adjacent in space)

功能分区,因此在同一时间观测到的交通流明显不同.由德国地理学家瓦尔特·克里斯塔勒于20世纪30年代提出的中心地理论以及空间分割理论可知,功能相似的区域往往在空间上相互远离,而相邻区域往往属于不同的功能区.这就意味着,功能相似区域内的地点之间的空间依赖关系通常是远距离的.然而,大多数现有方法主要关注局部特征(由于基于GNN的方法存在过度平滑问题),并没有充分考虑到功能区随地理位置变化而产生的变化,因此,这些方法无法捕捉长程依赖关系和特定区域的交通模式,这在一定程度上限制了模型的准确性.

为解决上述问题,本文提出了一种既能感知负载又能感知空间异构性的交通流预测模型,即LSFormer,主要贡献如下:

(1)介绍了一种利用时空自注意力机制的LSFormer模型,旨在精确预测交通流,去解决交通数据的动态性、区域差异性和可扩展性带来的复杂问题;

(2)设计了一个空间负载感知模块,将周边交通流负载信息整合到自注意力互动中,以帮助模型更好、更有效地建立空间依赖关系;

(3)在空间自注意力中引入了相对位置编码,使模型能够捕捉空间的相对位置关系;

(4)使用5个真实的公共数据集进行了广泛的交通流预测实验.

1 研究现状

1.1 用于交通预测的深度学习

早期,循环神经网络和卷积神经网络(CNN, Convolutional Neural Network)被广泛应用于交通流预测.前者能有效捕捉时间信息,后者能有效捕捉空间信息.

随后,随着图神经网络在处理图结构数据方面的出色表现^[5],也被广泛应用于交通流预测领域.近年来,注意力机制被广泛应用于挖掘交通流数据中隐藏的时空依赖关系.其他研究则侧重于探索数据

之间的潜在时间关系,以丰富模型的输入内容.还有人致力于解决传统变换器模型在位置编码方面的缺陷.文献[6]提供了对空间和时间因素如何随时间相互作用的深入的理解,从而建立了更准确的预测模型.文献[7]对多步骤依赖关系网络(MSDR, Multi-Step Dependency Relationship Network)的探索标志着时空预测领域的重大进展.MSDR强调了交通数据中多步依赖关系的重要性,提出了一种能够在不同预测视野中学习复杂时空关系的网络结构.

与上述研究相比,本文提出的LSFormer模型不仅通过时间嵌入层(Temporal Embedding Layer)探索数据中隐藏的时间信息,还通过使用空间嵌入层(Spatial Embedding Layer)发掘隐藏的空间信息.此外,还认识到了基于Transformer的模型的局限性,并为空间设计了有效的相对位置编码,以充分发挥该模型在交通流预测任务中的性能.

1.2 Transformer

Transformer是一种独特的网络架构,它完全基于自注意力机制,已在多变量时间序列预测中证明了其有效性.近来,基于Transformer的高效创新模型层出不穷.Crossformer突破了以往基于Transformer的模型的局限性,考虑到了跨维依赖性.Fedformer结合了频域的概念,在降低模型复杂度的同时提高了性能.另一个贡献是由Autoformer模型提供的,该模型基于深度分解架构和自相关机制,通过渐进分解和序列级连接,长期预测效率显著提高.

2 相关定义

本节将介绍后续章节中使用的一些术语,并将交通流预测问题形式化,以便更好地理解 and 解释采用的技术和方法.为了建立图模型并收集传感器位置和交通流数据,首先需要确定图的节点和边.在本文中,将根据收集到的传感器位置和交通流量数据构建一个图 G .每个传感器都是该图中的一个节点,其关系用 G 的结构来表示,具体方法如图2所示.

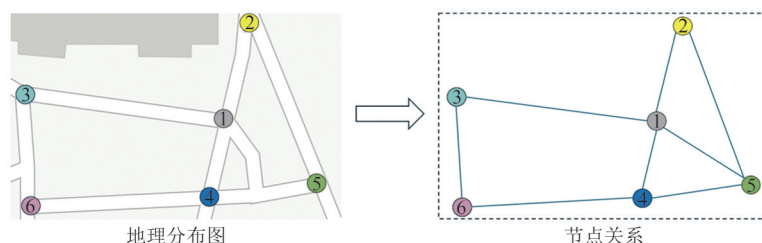


图2 道路图的构建

Fig. 2 Construction of road graph

2.1 术语

定义 1(道路网络) 道路网络被描述为一个图 $G = (V, \varepsilon, A)$, 其中 $V = \{v_1, \dots, v_N\}$ 表示一组包含 N 个节点 ($|V| = N$) 的集合, $\varepsilon \subseteq V \times V$ 表示边的集合, A 是网络 G 的邻接矩阵.

定义 2(交通流量) 让 $X_t \in R^{N \times C}$ 表示道路网络中节点 N 在时间 t 的交通流量, 其中 C 表示交通流量维度. 定义 $\chi = (X_t, X_{t+1}, \dots, X_{t+T}) \in R^{T \times N \times C}$ 为 T 个时间片段上所有节点的交通流量.

2.2 问题定义

根据到当前时间步长的历史交通流量 χ , 来构建一个预测函数 f , 目的是更准确地估计未来 T' 时间步长内所有区域的未来交通流量, 具体函数形式如下:

$$X_t, \dots, X_{t+T} \xrightarrow{f} X_{t+T+1}, \dots, X_{t+T+T'} \quad (1)$$

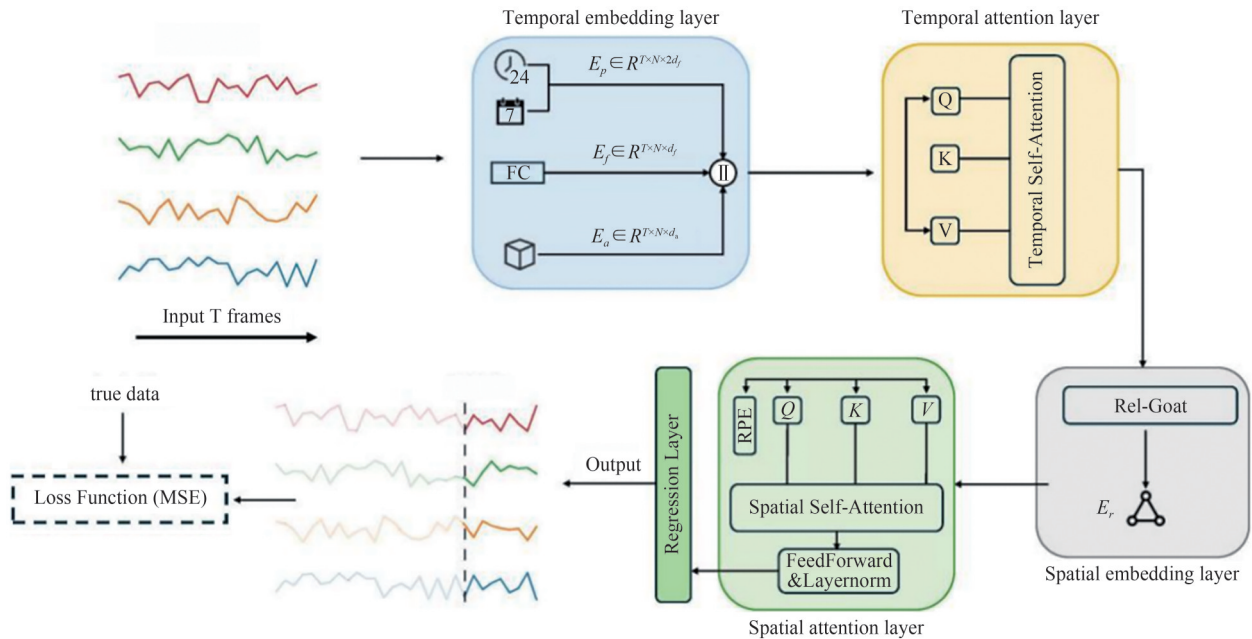


图 3 空间和负载感知转换器(LSFormer)的架构

Fig. 3 The architecture of spatial and workload aware transformer (LSFormer)

3.1 时间嵌入层

在时间嵌入层中, 原始交通流数据首先通过一个全连接层 $FC(\cdot)$ 进行维度扩展, 得到 $E_f \in R^{T \times N \times d_f}$:

$$E_f = FC(X_t; X_{t+T}). \quad (2)$$

然后, 为了使模型能够捕捉周期性信息, 引入了两个可学习的嵌入字典, 分别用 $T_\omega \in R^{N_\omega \times d_f}$ 和 $T_d \in R^{N_d \times d_f}$ 表示, 其中, T_ω 用于识别当前时刻的交通流数据是一周中的哪一天, 一般情况下, $N_\omega = 7$. 与 T_ω 类似, T_d 用于识别交通流数据是一天中的哪一时刻. 在本文使用的数据集中, $N_d = 288$. 然后, 参考 STAEformer, 得到周期嵌入信息 $E_p \in R^{T \times N \times 2d_f}$ 和时空自适应嵌入信息 $E_a = R^{T \times N \times d_a}$.

3 方法

图 3 展示了 LSFormer 的整体框架, 它由时间嵌入层、时间转换层、空间嵌入层、空间转换层和回归层组成. 为了准确捕捉交通流数据固有的周期性, LSFormer 在时间嵌入层中加入了 3 个时间嵌入. 嵌入时间信息的数据随后被送入时间转换层, 从而使 LSFormer 考虑到时间依赖性. 随后, 在空间嵌入层中实施负载感知模块, 以嵌入空间位置信息. 由此产生的输出被送入带有相对位置编码的空间转换层, 使模型能够捕捉空间依赖性. 通过上述模块, LSFormer 可以有效地将空间和时间信息结合起来, 从而实现高准确度的预测.

最后, 通过合并这 3 种嵌入, 得到了特征表示 $Z \in R^{T \times N \times d_h}$, 用于输入时间转换层, 其中 $d_h = 3d_f + d_a$, Z 的计算式如下:

$$Z = E_f || E_p || E_a. \quad (3)$$

3.2 空间嵌入层

在空间嵌入层中引入新模块: 负载感知模块, 用于生成嵌入向量 $E_g \in R^{T \times N \times d_g}$.

负载感知模块. 它考虑到了这样一个事实, 即在给定时刻 t , 空间中任何位置的交通流量都不是独立的, 都会受到周围所有位置交通流量的影响, 而且周围不同位置的交通流量也会对其产生不同的影响. 例如, 位于市中心的地点可能比位于人口稀少地区

的地点拥有更多的交通流量.同时,地点是在市中心还是在市中心边缘也会影响交通流量.因此,借鉴了图排序注意力网络(GOAT, Graph Ordering Attention Network)^[8]的思想,开发了一个名为 Rel-GOAT 层的

模块(如图 4 所示),以捕捉一个地点与其周围所有地点在交通流量方面的依赖关系.接下来介绍 Rel-GOAT 的机制,它由以下 3 部分组成:

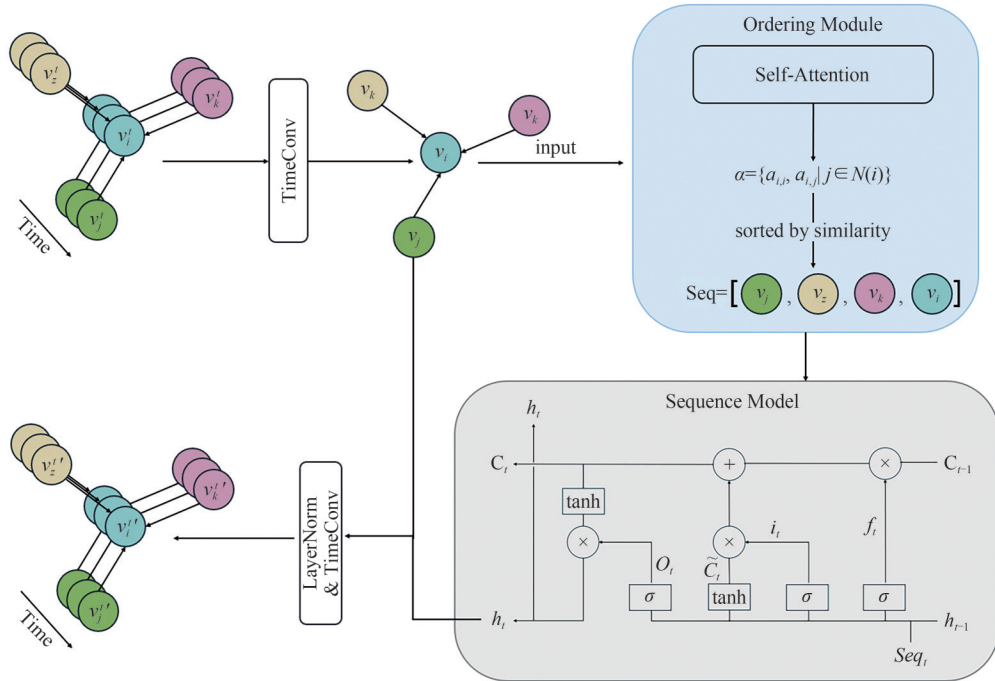


图 4 Rel-GOAT 层示意图

Fig. 4 Illustration of a Rel-GOAT layer

(1) GOAT 最初的设计侧重于对图结构数据建模,主要解决空间信息问题.然而,它并没有考虑到图中的信息依赖关系可能会随时间发生变化这一事实.因此,GOAT 在捕捉时空依赖性和对时变图结构的动态变化进行建模时存在局限性.为了使 Rel-GOAT 能够处理时空数据,首先对每个节点的所有时间步进行时间卷积运算,然后将结果作为下一步的输入.在 Rel-GOAT 的输出结果方面,使用了层归一化和残差连接技术,这使得 Rel-GOAT 在交通流预测任务中取得了优异的性能.

(2) 排序部分利用注意力机制将无序的邻居隐藏状态向量集(每个向量的维度为 d , 表示为 $\{v_1, \dots, v_Q\}$, 其中 Q 等于邻居数量加 1) 排列成顺序序列:

$$[v_{\pi_1}, \dots, v_{\pi_Q}] = \text{OrderingPart}(v_1, \dots, v_Q), \quad (4)$$

其中,排序由置换函数 $\pi(\cdot)$ 决定.

具体来说,与图注意力网络模型类似,对于图中的每个节点 $n_i \in V$, 首先应用一个共享的线性变换(参数为 $\beta \in R^{d \times d}$ 的权重矩阵), 然后,使用由 $\vec{\mu} \in R^{2d}$ 参数共享的自注意力机制来计算注意分数:

$$a_{ij} = \text{Re Lu}(\vec{\mu}^T [\beta v_i || \beta v_j]). \quad (5)$$

对于节点 i 的邻居 j , 有 $n_j \in N(n_i)$. 随后,对系数

进行降序排序,得到邻域节点的特定排列组合 π .

$$a_{i\pi(1)}, \dots, a_{i\pi(Q)} = \text{sort}(a_{i1}, \dots, a_{iQ}). \quad (6)$$

在确定了排列组合 π 之后,创建邻域隐藏状态的排序序列:

$$v_{\text{sorted}(i)} = \left[\frac{e^{a_{i\pi(1)}}}{\sum_{j=1}^Q e^{a_{i\pi(j)}}} \beta v_{\pi(1)}, \dots, \frac{e^{a_{i\pi(Q)}}}{\sum_{j=1}^Q e^{a_{i\pi(j)}}} \beta v_{\pi(Q)} \right]. \quad (7)$$

(3) 序列建模部分将排序部分生成的节点有序序列作为输入,它使用所有邻域共享的 RNN (Recurrent Neural Network) 处理这些序列,生成新的隐藏状态.在现有的 RNN 中,双向 LSTM 最为合适,它的遗忘门有助于消除冗余信息,输入门足以有效隔离独特信息,而记忆状态则有助于识别协同信息,其形式如下:

$$v_i^{\text{new}} = \text{LSTM}(v_{\text{sorted}(i)}) \in R^d. \quad (8)$$

具体来说,就是从时间自注意力层获得的结果 Z_i 输入到 Rel-GOAT 层,以获得空间负载嵌入 E_g , 该嵌入同时考虑了周围的交通流量和不同位置的不同影响:

$$E_i = \text{Rel-GOATlayer}(Z_i). \quad (9)$$

在得到上述空间嵌入后,将它们与 Z_i 进行维度

连接,得到最终的特征表示 Z_i' ,并将其输入空间转换层:

$$Z_i' = Z_i \| E_g. \quad (10)$$

3.3 自注意力层

对于通过时间和空间嵌入层处理后的特征表示,采用 vanilla transformer^[9]来捕捉时间和空间上的相互依存关系.给定隐藏表示 $Z \in R^{T \times N \times d}$,具有 T 个时间切片和 N 个空间节点,通过自注意力层获得查询、键和值矩阵如下:

$$Q = ZW_q, K = ZW_k, V = ZW_v, \quad (11)$$

其中 W_q 、 W_k 和 W_v 是可学习参数,那么计算注意力分数的方法如下:

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_h}} + rpe\right), \quad (12)$$

其中, A 反映了时间和空间关系, rpe 反映了不同节点之间的相对位置.然后,得到输出 Z_o :

$$Z_o = AV. \quad (13)$$

最后,将经过时间和空间自注意力层处理的表征 $Z_s \in R^{T \times N \times d}$,通过全连接层生成预测结果.线性回归层的表达式如下:

$$\hat{Y} = FC(Z_s), \quad (14)$$

其中 $\hat{Y} \in R^{T' \times N \times d}$ 是预测值, T' 是预测范围.

从正弦位置编码中汲取灵感,提出在时间和空间嵌入层中捕捉序列之间的绝对位置关系.然而,最近的研究^[10-11]强调,绝对位置编码方法不利于模型捕捉序列之间的相对位置关系.因此,设计了以下两种特定的相对位置编码,以增强模型对位置关系的感知.

空间相对位置编码.在时间序列中,前后关系明确存在于序列中,而在空间中,邻接关系不能直接从序列中得出.因此,使用邻接矩阵 $A \in R^{N \times N}$ 计算了跳数矩阵 $sh \in R^{N \times N}$,该矩阵计算了任意节点到达所有其他节点所需的节点数.为了提高模型的泛化性能,在这一部分引入一个截断阈值 τ_s ,具体来说,当 $r \in sh > \tau_s$ 时,将元素的值截断为 τ_s ,截断后,得到的矩阵为 rpe_s .跳数矩阵能够直观地反映节点之间的空间邻近性,而截断阈值 τ_s 则限制了远距离节点对注意力机制的干扰.通过引入相对位置编码,模型可以同时捕获局部和全局的空间依赖关系,从而提升预测精度并能够显式地量化每个节点对其他节点的相对重要性,直观反映节点间的空间关系,这种设计对后续嵌入可视化提供了良好的支持.

4 实验

4.1 数据集

实验将在 PEMS03、PEMS04、PEMS08、METR-LA 和 PEMS-BAY 五个实际公共交通数据集上验证 LSFormer 的性能.这 5 个数据集中的交通数据均以 5 min 为一个时间间隔,每小时共 12 个帧.每个数据集都包含大量的交通流量、速度和占用率数据,提供了丰富的交通动态信息.

为便于分析和处理,这些数据被统一为 5 min 的时间间隔,以确保一致性和可比性.每小时的交通数据分为 12 个时间段.这种精细的时间划分捕捉到了交通流量的细微变化,能够更准确地评估和预测交通状况.

这 5 个数据集的详细统计信息见表 1.时间跨度较长(如 METR-LA 和 PEMS-BAY)的数据集包含更多的周期性和趋势性变化,测试了模型对时序信息的捕捉能力.相对较短的数据集(PEMS04、PEMS08)则更适合验证模型的泛化性能.节点密集的数据集(PEMS04)对局部依赖建模提出了较高要求,而稀疏分布(PEMS03)则需要模型更关注全局的空间关系.节点线性分布(METR-LA)验证了相对位置编码在长程空间依赖捕获中的优势,而复杂分布(PEMS-BAY)检验了模型对异质性的适应能力.

表 1 数据集说明

Tab. 1 Data set description

数据集	节点数	时间戳	时间间隔/min	时间范围
PEMS03	358	26208	5	09/01/2018-11/30/2018
PEMS04	307	16992	5	01/01/2018-02/28/2018
PEMS08	170	17856	5	07/01/2016-08/31/2016
METR-LA	207	34272	5	03/01/2012-06/27/2012
PEMS-BAY	325	52116	5	01/01/2017-06/30/2017

4.2 基线模型

在本研究中,将所提出的方法与几种广泛使用的基线方法进行了比较,以确定其在所考虑的任务中的相对性能.

ST-WA^[6]提出了一种时空感知的交通时间序列预测框架.该框架通过生成位置特定和时间变化的模型参数,动态捕捉时空模式,并引入窗口注意力机制,将复杂度从二次降至线性,显著提升了预测效率和精度.

HI^[12]是一个典型的传统模型.

DCRNN^[13]是交通流预测领域的经典基线模型.该模型通过扩散过程模拟交通流随时间和空间变化

的动态特征,并利用扩散卷积运算识别空间相关性.

AGCRN^[14]使用自适应模块提高传统图卷积的性能,并将其纳入RNN以捕捉时空依赖性.

STGCN使用一维卷积来捕捉空间和时间依赖性,并且是基于图卷积的.

GTS^[15]认为通过探索变量之间的相关性和因果关系可以增强时间序列预测,并提出与GNN同时学习图中存在的信息.

STID^[16]深入而有效地分析了受周期性因素影响的当前时刻到预测时刻交通流的复杂过程,因此只需一个简单的全连接层就能实现非常理想的结果.

MTGNN^[17]是一种通用的多变量依赖图神经网络框架,它通过图学习模块学习多变量单向关系,并使用图卷积和时间序列卷积模块捕捉时空依赖特征.

STNorm^[18]对交通时间序列进行因式分解.

为了进一步证实模型的性能,有必要介绍一些为处理这项任务而设计的基于变压器的模型,其中,Informer、CrossFormer、FEDformer和Autoformer是分别由文献[19-22]提出的基于变换器架构的模型,这些模型在多变量时间序列预测任务中取得了优异的性能.然而,由于这些模型缺乏针对交通预测任务的具体设计,实用性受到了限制.因此,选择了一系列不同的替代模型作为本文的实验对比对象,如GMAN^[23],它是一种基于注意力的模型,采用编码器-解码器架构,以及PDFormer.

LSFormer在PEMS08上表现尤为突出,得益于负载感知模块在捕捉区域性交通模式方面的优势. PEMS03和PEMS04的性能差异反映了模型对稀疏(PEMS03)和密集(PEMS04)节点分布的适应能力,相对位置编码有效增强了全局和局部特征建模,实验结果见下表2.

表2 各数据集(PEMS03、PEMS04和PEMS08)上的性能对比

Tab. 2 Performance comparison on datasets (PEMS03, PEMS04 and PEMS08)

数据集	PEMS03			PEMS04			PEMS08		
	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE
HI	32.62	30.60	49.89	42.35	29.92	61.66	36.66	21.63	50.45
DCRNN	15.54	15.62	27.18	19.63	13.59	31.26	15.22	10.21	24.17
AGCRN	15.24	15.89	26.65	19.38	13.40	31.25	15.32	10.03	24.41
STGCN	15.83	16.13	27.51	19.57	13.44	31.38	16.08	10.60	25.39
GTS	15.41	15.39	26.15	20.96	14.66	32.95	16.49	10.54	26.08
MTGNN	14.85	14.55	25.23	19.17	13.37	31.70	15.18	10.20	24.24
STNorm	15.32	14.37	25.93	18.96	12.69	30.98	15.41	9.76	24.77
GMAN	16.87	18.23	27.92	19.14	13.19	31.60	15.31	10.13	24.92
PDFormer	14.94	15.82	25.39	18.36	12.00	30.03	13.58	9.05	23.41
STID	15.33	16.40	27.40	18.38	12.04	29.95	14.21	9.27	23.28
ST-WA	15.17	15.83	26.63	19.06	12.56	31.02	15.41	9.94	24.62
LSFormer	15.40	15.66	27.18	18.34	12.00	30.21	13.50	8.87	23.28

4.3 评估指标

为了科学评价不同模型的预测效果,本文选取了一组客观指标,包括均方根误差(RMSE, Root Mean Square Error)、平均绝对误差(MAE, Mean Absolute Error)和平均绝对百分比误差(MAPE, Mean Absolute Percentage Error).这些指标各有特点和应用场景,可以从不同角度综合评价模型的预测性能.其中, RMSE适用于对较大误差较为敏感的场景,因为它能放大较大的误差值,从而更准确地反映模型在极端情况下的表现. MAE适用于需要平衡每个误差的场景,它提供了误差的平均绝对值,可以直观地反映预测偏差.而MAPE则适用于需要测量相对误差的情况,尤其适用于比较模型在不同数据集上的表现,但当实际值接近于零时可能会不稳定.上述指

标用公式表示如下:

$$RMSE = \sqrt{\frac{1}{p} \sum_{i=1}^p (Y_{\text{real}(i)} - Y_{\text{Pred}(i)})^2}, \quad (15)$$

$$MAE = \frac{1}{p} \sum_{i=1}^p |Y_{\text{real}(i)} - Y_{\text{Pred}(i)}|, \quad (16)$$

$$MAPE = \frac{100\%}{p} \sum_{i=1}^p \left| \frac{Y_{\text{real}(i)} - Y_{\text{Pred}(i)}}{Y_{\text{real}(i)}} \right|. \quad (17)$$

4.4 实验设置

数据集预处理.将数据集划分为训练集、验证集和测试集对于评估模型的性能和确保其通用性至关重要.考虑到数据集之间的差异,本实验使用了两套不同的数据集划分比例来训练模型. PEMS-BAY和METR-LA数据集按7:1:2的比例分为训练集、验证集和测试集.同时,PEMS03、PEMS04和PEMS08

数据集按 6:2:2 的比例分割.此外,使用前一小时的数据来预测后一小时的交通流量.

模型设置.所有实验都由一台配备了英伟达™(NVIDIA®)GeForce 4090 GPU 的计算机支持,并使用 24 GB 内存提供充足的计算资源.实验中使用 PyTorch 1.10.0 和 Python 3.8.18 这两个工具包来构建和训练 LSFormer 模型.在模型训练过程中,根据验证集的表现选择最佳模型,以确保模型的准确性和泛化能力.具体来说,使用 AdamOptimizer 进行训练,批量大小为 16,最多训练 100 期,以保证模型的训练效果.

4.5 性能比较

表 2 和表 3 显示了本文提出的模型在 5 个数据

集上与基线模型的详细比较,最佳结果用蓝色表示,次佳结果用下划线表示.从表 2-3 可以得出以下结论:(1) LSFormer 在全部数据集上的所有指标几乎都优于所有基线模型;(2) 在基于 GNN 的模型中,MTGNN 的性能很有竞争力.这些基于 GNN 的模型在传递信息时只考虑了空间中节点的邻接性,相比之下,LSFormer 因为考虑了全局范围内区块之间的相似性和差异性而取得了更好的性能;(3) 在基于自注意力的模型方面,PDFormer 成为最佳基准.与 PDFormer 相比,LSFormer 引入了有效的空间嵌入,帮助自注意力模块更好地捕捉到交通模式,并取得了良好的性能.

表 3 各数据集(METR-LA, PEMS-BAY)上的性能对比

Tab. 3 Performance comparison on datasets (METR-LA, PEMS-BAY)

数据集	指标	HI	DCRNN	AGCRN	STGCN	GTS	MTGNN	STNorm	GMAN	PDFormer	STID	LSFormer	
METR-LA	Horizon 3 (15 min)	MAE	6.80	<u>2.67</u>	2.85	2.75	2.75	2.69	2.81	2.80	2.83	2.82	2.64
		RMSE	14.21	<u>5.16</u>	5.53	5.27	5.27	5.16	5.57	5.55	5.45	5.53	5.10
		MAPE/%	16.72	76.86	7.63	7.12	7.12	<u>6.89</u>	7.40	7.41	7.77	7.75	6.81
	Horizon 6 (30 min)	MAE	6.80	3.12	3.20	3.15	3.14	<u>3.05</u>	3.18	3.12	3.20	3.19	2.96
		RMSE	14.21	6.27	6.52	6.35	6.33	<u>6.13</u>	6.59	6.49	6.46	6.57	6.06
		MAPE/%	16.72	8.42	9.00	8.62	8.62	8.16	8.47	8.73	9.19	9.39	<u>8.17</u>
	Horizon 12 (60 min)	MAE	6.80	3.54	3.59	3.60	3.59	3.47	3.57	<u>3.44</u>	3.62	3.55	3.35
		RMSE	14.20	7.47	7.45	7.43	7.44	<u>7.21</u>	7.51	7.35	7.47	7.55	7.07
		MAPE/%	10.15	10.32	10.47	10.47	10.25	9.70	10.24	10.07	10.91	10.95	<u>9.80</u>
PEMS-BAY	Horizon 3 (15 min)	MAE	3.06	<u>1.31</u>	1.35	1.35	1.37	1.33	1.33	1.35	1.32	<u>1.31</u>	1.31
		RMSE	7.05	2.76	2.88	2.88	2.92	2.82	2.82	2.90	2.79	2.79	<u>2.77</u>
		MAPE/%	6.85	2.73	2.91	2.91	2.85	2.76	2.76	2.87	2.78	2.78	<u>2.76</u>
	Horizon 6 (30 min)	MAE	3.06	1.65	1.67	1.67	1.72	1.65	1.65	1.65	<u>1.64</u>	<u>1.64</u>	1.62
		RMSE	7.04	3.75	3.82	3.82	3.86	3.77	3.77	<u>3.82</u>	3.73	3.73	3.69
		MAPE/%	6.84	3.71	3.81	3.81	3.88	<u>3.66</u>	3.66	3.74	3.73	3.73	<u>3.67</u>
	Horizon 12 (60 min)	MAE	3.05	1.97	1.94	1.94	2.06	1.92	1.92	1.92	<u>1.91</u>	<u>1.91</u>	1.90
		RMSE	7.03	4.60	4.50	4.50	4.60	4.45	4.45	4.49	<u>4.42</u>	<u>4.42</u>	4.37
		MAPE/%	6.83	4.68	4.55	4.72	4.88	4.46	4.46	4.52	4.55	4.55	4.54

4.6 消融实验

为了验证 LSFormer 模型各部分对提高模型预测精度的帮助程度,本实验对模型的两个变体进行了消融研究,具体如下:

- ● w/o E_g : 该变体删除了工作量感知嵌入 E_g ;
- ● w/o rpe_s : 这一变体删除了空间相对位置编码 rpe_s .

图 5 显示了上述组件对本文模型性能的影响.考虑到 PEMS08 数据集在多个数据集中最具代表性,对其进行了消融研究.通过分析消融实验的结果,可以得出以下结论:

(1) 去除 E_g 会使模型在各指标上的性能下降最大,因此表明 E_g 对交通建模至关重要;

(2) 此外,当移除空间相对位置编码后,观察到性能大幅下降,表明位置编码增强了基于变压器结构的模型处理交通流相关数据的能力.

4.7 案例研究

在本节中,将对 LSFormer 学习到的负载感知嵌入 E_g 进行可视化处理,以提高其可解释性,并研究二者在提高交通流预测准确性方面的有效性.

E_g 的可视化.通过 GOAT 学习获得的负载感知嵌入的可视化结果显示,在数据集 PEMS08 上,大多数点呈现内外双环结构,中心有少数几个点(如图 6(a)所示).这种环形结构本身就具有连续性特征,这与相距不远的两个流量节点周围的负载不完全相同但差异不会太大的特性相一致.因此,最终的可视

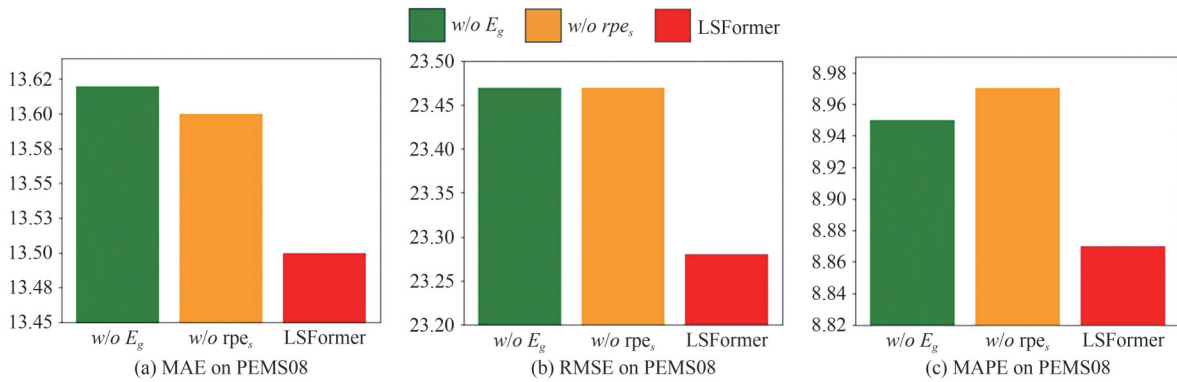


图5 PEMS08的消融研究

Fig. 5 Ablation study on PEMS08

化结果显示,大部分点分布在环形结构上,少数几个点孤立为中心,这表明大部分点受到周围交通流的影响,但也有少数地方受周围交通流的影响较小.然而,METR-LA数据集的可视化结果呈现出条状结构(如图6(b)所示).这种形状也反映了节点之间的连续性,但与PEMS08数据集相比存在显著差异.推

测这可能是由于两个数据集中传感器的空间分布不同造成的.METR-LA数据集中的传感器大多分布在相互交叉的直线上,类似于地铁线路的分布,而PEMS08数据集的空间布局相对更为复杂.这两项研究结果表明,所提出的 E_g 可以有效捕捉周围交通流对中心点交通流的影响.

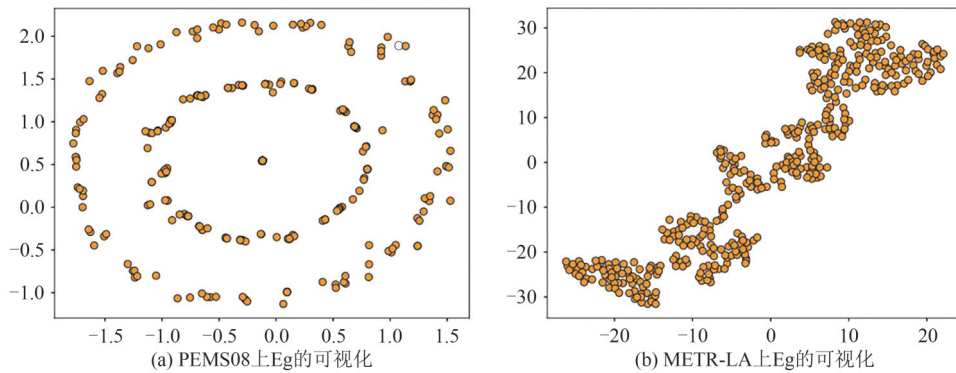


图6 PEMS08和METR-LA案例研究

Fig. 6 Case study on PEMS08 and METR-LA

注意力得分的可视化.图7(a)和图7(b)分别展示了验证集的空间注意力得分矩阵和时间注意力得分矩阵,并根据训练好的模型以热图的形式进行

了可视化.从图7(a)可以看出,每个节点除了给自己分配高注意力分数外,还对一定数量的其他节点分配了较高的注意力分数.这些高分节点并不局限

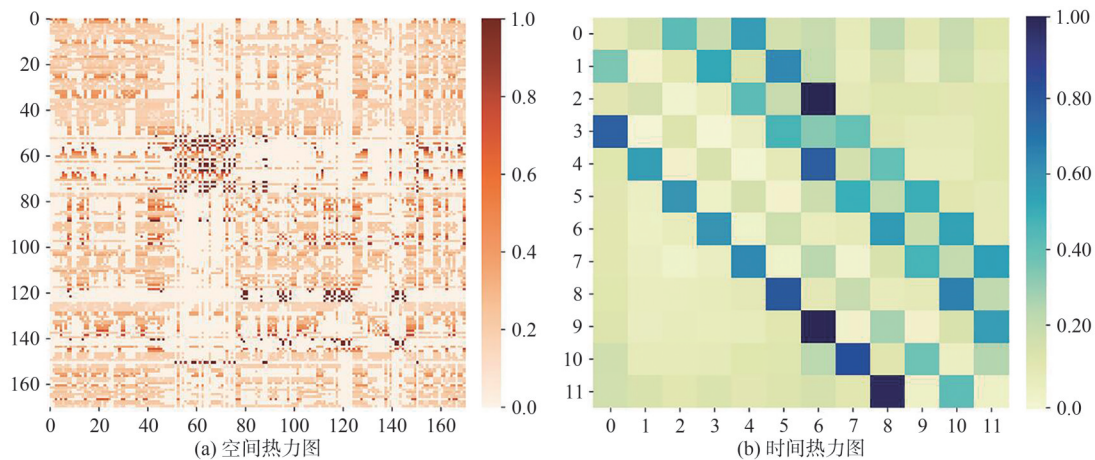


图7 PEMS08注意力得分可视化

Fig. 7 Visualization of attention scores on PEMS08

于附近的节点,也分布在较远的位置,这表明该模型能有效捕捉远距离空间依赖关系.相比之下,图7(b)揭示了一个不同的现象:在任何给定的时间步长 t 中,只有靠近时间步长 t 的交通流才会获得较高的注意力分数,而较远时间步长的注意力分数明显较低.这表明,在短期预测中,预测值与距其相邻的时间步长具有很强的依赖性.

5 结语

本文提出了用于交通流预测的LSFormer模型,具体来说,开发了一个负载感知模块,用于发现当前位置与其周围环境在交通流方面的依赖关系;进一步设计了空间位置编码,以提高模型的泛化能力.在5个真实世界数据集上进行了多次实验,证明了LSFormer在预测准确性方面优于其他模型,并将学习到的空间嵌入可视化,以帮助理解它.未来,将尝试把LSFormer应用于其他时空预测任务,例如光伏发电预测^[24-26].此外,还将探索比较学习方法在交通流预测中的潜力.

参 考 文 献

- [1] YIN C, XIONG Z, CHEN H, et al. A literature survey on smart cities[J]. *Science China Information Sciences*, 2015, 58(10): 1-18.
- [2] YU B, YIN H, ZHU Z, et al. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[C]//*Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*. Stockholm: ijcai.org, 2018:3634-3640.
- [3] LIU H, DONG Z, JIANG R. Spatio-temporal adaptive embedding makes vanilla transformer SOTA for traffic forecasting [C]//*ACM. Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*. Birmingham: ACM, 2023:4125-4129.
- [4] JIANG J, HAN C, ZHAO W X. PDFFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction [C]//*AAAI. Proceedings of the 2023 AAAI Conference on Artificial Intelligence*. Washington D C: AAAI Press, 2023: 1132-1139.
- [5] 姚静怡,李艳红,黄银峰,等.灵活的属性社区搜索方法[J].*中南民族大学学报(自然科学版)*,2024,43(3): 358-369.
- [6] CIRSTEA R G, YANG B, GUO C. Towards spatio-temporal aware traffic time series forecasting[C]//*IEEE. Proceedings of the 2022 IEEE International Conference on Data Engineering (ICDE)*. Kuala Lumpur: IEEE, 2022: 2900-2913.
- [7] LIU D, WANG J, SHANG S. MS DR: Multi-step dependency relation networks for spatial-temporal forecasting [C]//*ACM. Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining (KDD)*. Washington D C: ACM, 2022: 1042-1050.
- [8] CHATZIANASTASIS M, LUTZEYER J, DASOULAS G. Graph ordering attention networks[J]. *AAAI*, 2023, 37(6): 7006-7014.
- [9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//*31st Conference on Neural Information Processing Systems (NIPS)*. Long Beach: Curran Associates, 2017:1-11.
- [10] SHAW P, USZKOREIT J, VASWANI A. Self-attention with relative position representations [C]//*ACL. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. New Orleans: ACL, 2018: 464-468.
- [11] LIU Z, LIN Y, CAO Y. Swin transformer: Hierarchical vision transformer using shifted windows [C]//*IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal: IEEE, 2021: 9992-10002.
- [12] CUI Y, XIE J, ZHENG K. Historical inertia: A neglected but powerful baseline for long sequence time-series forecasting [C]//*ACM. Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*. Virtual Event: ACM, 2021: 2965-2969.
- [13] LI Y, YU R, SHAHABI C. Graph convolutional recurrent neural network: Data-driven traffic forecasting [C]//*9th International Conference on Learning Representations (ICLR)*. Vancouver: OpenReview.net, 2018:1-16.
- [14] BAI L, YAO L, LI C. Adaptive graph convolutional recurrent network for traffic forecasting [J]. *Advances in neural information processing systems*, 2020 (33): 17804-17815.
- [15] SHANG C, CHEN J, BI J. Discrete graph structure learning for forecasting multiple time series [C]//*6th International Conference on Learning Representations (ICLR)*. Virtual Event: OpenReview.net, 2021:1-14
- [16] SHAO Z, ZHANG Z, WANG F. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting [C]//*ACM. Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM)*. Atlanta: ACM, 2022: 3257-3260.
- [17] WU Z, PAN S, LONG G. Connecting the dots: Multivariate

- time series forecasting with graph neural networks [C]//ACM. Proceedings of the 2020 International Conference on Knowledge Discovery and Data Mining (KDD). Virtual Event: ACM, 2020: 753-763.
- [18] DENG J, CHEN X, JIANG R. ST-Norm: Spatial and temporal normalization for multi-variate time series forecasting [C]//ACM. Proceedings of the 2021 International Conference on Knowledge Discovery and Data Mining (KDD). Singapore: ACM, 2021: 2401-2410.
- [19] ZHOU H, ZHANG S, PENG J. Informer: Beyond efficient transformer for long sequence time-series forecasting [C]//AAAI. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Virtual Event: AAAI Press, 2021: 11106-11115.
- [20] ZHANG Y, YAN J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting [C]//ICLR. Proceedings of the International Conference on Learning Representations (ICLR). Kigali: OpenReview.net, 2023:1-21.
- [21] ZHOU T, MA Z, WEN Q. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting [C]//NeurIPS. Proceedings of Advances in Neural Information Processing Systems (NeurIPS). New Orleans: NeurIPS, 2022:27268-27286.
- [22] WU H, XU J, WANG J. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting [C]//NeurIPS. Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Virtual Event: NeurIPS, 2021:22419-22430.
- [23] ZHENG C, FAN X, WANG C. GMAN: A graph multi-attention network for traffic prediction [C]//AAAI. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). New York: AAAI Press, 2020: 1234-1241.
- [24] SOUHE F G Y, MBEY C F, KAKEU V J F, et al. Optimized forecasting of photovoltaic power generation using hybrid deep learning model based on GRU and SVM[J]. Electrical Engineering, 2024,106: 1-20.
- [25] ASGHAR R, FULGINEI F R, QUERCIO M, et al. Artificial neural networks for photovoltaic power forecasting: A review of five promising models [J]. IEEE Access, 2024(12):90461-90485.
- [26] HALJASMAA K I, BRAMM A M, MATRENIN P, et al. Weather condition clustering for improvement of photovoltaic power plant generation forecasting accuracy [J]. IEEE Access, 2024, 17(9):419.

(责编 曹东,校对 雷建云)