李博,崔高峰,冯泽涛.基于注意力机制的多图神经网络交通预测模型[J].智能计算机与应用,2025,15(3):56-63. DOI: 10.20169/j.issn.2095-2163.24082102

基于注意力机制的多图神经网络交通预测模型

李 博¹, 崔高峰², 冯泽涛¹

(1 山东工商学院 计算机科学与技术学院,山东 烟台 264005; 2 山东工商学院 统计学院,山东 烟台 264005)

摘 要:由于复杂的时空相关性和非线性的交通模式,实现精确的预测仍然是一个挑战。针对于此,本文提出了一种新颖的 多变量时间序列预测框架(MSTGCN+AL),尝试使用多图神经网络进行预测。引入2种新的图类型,一是通过自适应邻接矩 阵得到自适应邻接图,能更好地获取交通节点之间的位置关系;另一个是潜在图,通过使用全局变量拟合三角函数,可以更好 地从交通数据中提取周期性和上下文信息。为了对齐图节点及其时间戳,采用了一个基于注意力机制的多图融合模块,包括 多图空间嵌入、空间注意力和图注意力。为了验证本文方法的有效性,在 METR-LA 数据集上进行了广泛的实验。实验结果 表明,与基线方法相比,所提 MSTGCN+AL 模型在预测性能上表现更佳。

关键词:图神经网络;自适应邻接矩阵;注意力机制;多图融合;交通预测

中图分类号: TP181 文献标志码: A 文章编号: 2095-2163(2025)03-0056-08

Attention-based multi-graph neural network for traffic prediction

LI Bo¹, CUI Gaofeng², FENG Zetao¹

(1 College of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, Shandong, China; 2 College of Statistics, Shandong Technology and Business University, Yantai 264005, Shandong, China)

Abstract: Accurate traffic flow prediction is essential for intelligent transportation systems; however, due to intricate spatiotemporal correlations and nonlinear traffic patterns, attaining precise predictions remains challenging. To address these issues, this paper presents a novel framework for multivariate time series prediction (MSTGCN+AL), attempting to use multi–graph neural networks for prediction. The paper has introduced two new types of graphs. One is the adaptive adjacency graph obtained through an adaptive adjacency matrix, which can better capture the positional relationships between transportation nodes. The other is the latent graph, which utilizes global variables to fit trigonometric functions, enabling it to extract periodicity and contextual information from traffic data more effectively. To align graph nodes and their timestamps, the paper devises a multi–graph fusion module based on attention mechanisms, including multi–graph spatial embedding, spatial attention, and graph attention. To validate the effectiveness of the propsed approach, extensive experiments are conducted on the METR–LA dataset. The experimental results show the superior predictive performance of the proposed MSTGCN+AL model when compared to baseline methods.

Key words: graph neural networks; adaptive adjacency matrix; attention mechanism; multi-graph fusion; traffic prediction

0 引 言

近年来,围绕长期时间序列数据预测的研究日 益增多,时序预测已广泛应用于城市交通流量预 测^[1]、未来天气预报,以及地铁系统中的实际乘客 流量预测等,这些场景下的预测对人们的生活起着 至关重要的作用。如今,针对此类数据的预测方法 多种多样,传统的统计方法有灰色系统理论 (GM)^[2]、向量自回归(VAR)模型^[3]和自回归积分 滑动平均(ARIMA)^[4]模型等。然而,传统方法在适 应性方面存在一定的局限性,难以有效处理非平稳 数据,并且容易出现过拟合问题。这些方法未能充 分揭示变量之间的潜在关系,因此需要更先进的技 术来更好地探索变量之间的潜在关系,以解决传统 方法的局限性。

城市交通流量是一个涉及高维和大规模的数据 集^[5],并且特定位置的交通状态可能同时具有空间 依赖性(不仅受邻近区域影响)和时间依赖性(周期

作者简介:李 博(1980—),男,博士,副教授,硕士生导师,主要研究方向:机器学习;崔高峰(1998—),男,硕士研究生,主要研究方向:图机 器学习。

通信作者: 冯泽涛(1977—),男,讲师,主要研究方向:金融大数据。Email:1299515135@qq.com。 收稿日期: 2024-08-21

性),因此对交通流量的预测比其他时间序列预测 任务更具挑战性,为此本文将二者相结合,采用时空 图神经网络进行城市交通流量预测。

尽管时空图神经网络的性能表现优秀,但在处 理交通流量数据时仍面临以下挑战:

(1)现有的时空图神经网络严重依赖于预定义 的图结构。然而,传统的邻接图^[6]通常基于某些先 验知识或数据预处理步骤,而且初始化过程并不完 善。这些邻接图可能会受到缺失或噪声数据的影 响,并且可能无法从数据中完全捕获变量之间的隐 藏关系。Wu 等学者^[7]提出了一种自适应邻接矩 阵,用于发现隐藏的空间依赖性,但并未有效地利用 现有的先验知识来编码邻接矩阵,因此可能导致重 要信息的丢失。

(2)长期时空数据往往呈现出周期性模式。现 有的图模型仅表示静态的空间信息,无法捕获长期 的时空依赖性。Shao等学者^[8]通过创建直方图来 总结给定的时空训练数据,提出了一个启发式图,旨 在利用启发式知识和人类洞察力。然而,这项工作 并未考虑所利用的长期时间序列数据中潜在的周期 性,可能对预测结果产生显著影响。

应对上述挑战,本文开展了关于多图构建与融合的研究,并进行了改进和优化。为了解决邻接图问题,本文提出了一种结合图学习原理的自适应邻接图算法。该算法能够从输入数据中学习图的结构和特征,以完成诸如图分类、节点分类和社区检测等任务。同时,为了克服现有多图神经网络(MGNN)方法的局限性,如泛化能力低和上下文信息利用不足,开发了一种新的图模型,称为潜在式图(Latent graph)。

1 相关工作

1.1 长期时空预测

近年来,作为深度学习技术代表的图神经网络 (GNNs),在非欧几里得空间数据表示学习中取得了 显著突破,为预测各种复杂城市数据奠定了基础。 Wang 等学者^[9]回顾了截至 2020 年开发的时空数据 挖掘的深度学习方法,包括与预测学习相关的一些 STGNN 技术。Ye 等学者^[10]分析了几个实际问题, 总结了在城市交通系统中关于预测、检测和控制问 题的相关工作。Bui 等学者^[11]和 Jiang 等学者^[12]专 门介绍了最新的 STGNN 技术在交通预测任务中的 应用。Gao 等学者^[13]研究了生成对抗技术在时空 数据学习中的应用,包括与时空图数据结合的方法。 这些研究和调查提供了对基于 STGNN 在各个领域 中预测学习技术应用的深入理解。

1.2 多图神经网络

多图卷积网络(MGCNs)在处理时空数据预测 任务中也引起了广泛关注。Bruna 等学者^[14]首次将 卷积神经网络引入图结构。随后,Defferrard 等学 者^[15]通过使用快速局部卷积进一步推动了这一领 域的研究工作。然而,单个图观测通常存在偏差,使 用多图可以弥补和减轻这些偏差。为解决这一问 题,Chai 等学者^[16]提出了一种基于多图卷积网络的 自行车交通预测方法。类似地,Lv 等学者^[17]使用 多图来编码道路之间的非欧几里得空间和语义相关 性,用于交通流预测。

2 本文方法

2.1 问题阐述

在实际时间序列数据预测问题中,变量的历史 观测通常以连续的时间步长记录,例如预测下周的 销售额或者预测未来一年的股票价格^[18]。设 $y_t = R^N$ 表示一个 N 维多元变量的值在时间步 t 的值, $y_t[i] \in R$ 表示 i^{th} 变量在时间步 t 的值。给定过去 T 个时间步的观测值,预测下一个时间步的观测值。 可用公式表示为:

$$y_{t+1} = f(y_t, y_{t-1}, \cdots, y_{t-T+1})$$
(1)

其中, y_t 表示时间步 t 的观测值; f 表示预测函数; T 表示给定的历史窗口大小。

从基于图的角度来看,将多变量时间序列中的 变量看作图中的节点。节点之间的关系可以表示为 一个邻接矩阵。由于这个邻接矩阵大多数情况下是 由人的主观判断直接给出的,所以研究中需要通过 模型来学习节点之间关系的表示。

2.2 模型框架

模型整体结构如图 1 所示。图 1 中,包括多维 度构图、节点表征生成(MGSE)、基于注意力机制的 多图动态融合、时空网络。其中,时空网络部分采用 ST-GNN 中的 MSTGCN 结构,可以采用任何一种时 空预测基础模型,由于具体 ST-GNN 结构不是本文 重点,在此不做赘述。本文的核心是多维度构图和 基于注意力机制的多图动态融合。

2.2.1 图的构建

为了从多个角度刻画空间维度上各个节点之间 的关系,文中构造了5种类型的图。本节中,详细描 述了本文提出的2个新的图模型,即自适应邻接图 $G^{4} = \{V, E, W^{4}\}$ 和潜在式图 $G^{P} = \{V, E, W^{P}\}, 以及$ 距离图 $G^{D} = \{V, E, W^{D}\}$ 、等效图 $G^{E} = \{V, E, W^{E}\}$ 、 时间模式相似图 $G^{T} = \{V, E, W^{T}\}$,由此组成一个多 重图集 $G = \{G^{D}, G^{A}, G^{E}, G^{P}, G^{T}\}$ 。通过多种类型图 的构建,实现从多个维度提取节点之间的信息,方便 后续多角度的时空关系建模。在此基础上,给出研 究中拟将用到的概念定义如下。



图 1 MSTGCN+AL 系统概述 Fig. 1 Overview of the MSTGCN+AL system

(1)距离图 (*G^b*):根据节点之间的欧式距离构 图。推得公式为:

$$W_{ij}^{D} = \begin{cases} \exp(-\frac{d_{ij}^{2}}{\sigma_{D}^{2}}), \text{ for } i \neq j \text{ and } \exp(-\frac{d_{ij}^{2}}{\sigma_{D}^{2}}) \geq \varepsilon \\ \downarrow 0, & \text{otherwise} \end{cases}$$
(2)

其中, d_{ij} 表示 v_i 和 v_j 之间的欧氏距离。并用 2 个参数 ε 和 σ_n 来控制矩阵的稀疏性和分布。

(2) 自适应邻接图 (G^{4}): 首先将先验知识得到 的初始化邻接矩阵设为 W_{ij}^{0} , 然后迭代更新 $W_{ij}^{(t+1)} =$ $W_{ij}^{(t)} + \lambda \times (A_{ij} - W_{ij}^{(t)})(t = 0, 1, 2, \cdots)$ 。其中, A_{ij} 是 给定邻接矩阵中从节点 i 到节点 j 的值(通常是0或 1, 表示是否连接)。 迭代直到满足收敛条件 $\max_{i,j} \mid W_{ij}^{(t+1)} - W_{ij}^{(t)} \mid < \theta$,这里 θ 是收敛阈值,最后 根据一个权重阈值 ε 得到新的邻接矩阵:

$$\boldsymbol{W}_{ij}^{A} = \begin{cases} 1, & \text{for } i \neq j \text{ and } w_{ij}^{A} \ge \varepsilon \\ 0, & \text{otherwise} \end{cases}$$
(3)

其中, W_{ii}^{A} 表示权重矩阵, ε 表示权重阈值。

(3)等效图 (G^{E}):根据每个地点的等效功能构 图。首先取与第 i 个节点最相关的 K 个节点的功能 向量,每个节点 v_i 可以得到一个功能向量 $E_i = \{f_{i,1}, f_{i,2}, \dots, f_{i,K}\}$,然后计算第 i 个节点与其他节点的 Pearson 相关系数,根据该结果构图,功能越相似的 地点、邻接矩阵对应元素值越大。研究推得公式如 下:

$$W_{ij}^{E} = \frac{\frac{1}{2}}{\frac{1}{2}} \frac{\sum_{k=1}^{K} (f_{i,k} - \bar{E}_{i})(f_{j,k} - \bar{E}_{j})}{\sqrt{\sum_{i=1}^{k} (f_{i,k} - \bar{E}_{i})^{2}} \sqrt{\sum_{j=1}^{k} (f_{j,k} - \bar{E}_{j})^{2}}}, \text{ if } i \neq j,$$
otherwise
$$(4)$$

其中, $f_{i,p}$ 表示节点 v_i 的速度值。

(4) 潜在式图 (G^{ρ}):为提升对长周期性及时空关 系的捕捉能力,提出了一种新的图模型,称为潜在式 图。使用全局数据拟合一个直方图时序数据分布,用 函数 $f(x) = \cos(\alpha x + \beta) + \mu$ 来近似该分布。每个节 点都会获得一组相应的参数 (α, β, μ) ,利用欧几里 得距离 $d_{ij}^{H} = \sqrt{(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2 + (\mu_1 - \mu_2)^2}$ 计算 2 个节点之间分布距离。具体公式如下:

$$W_{ij}^{P} = \begin{cases} \exp(-\frac{\|d_{ij}^{H}\|^{2}}{\sigma_{H}^{2}}), & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases}$$
(5)

(5)时间模式相似图 (*G^T*):根据定义序列的长度 *p*,利用每个节点的时间序列数据,计算两两节点之间的 Pearson 相关系数,由此来定义时间模式相似图矩阵。具体公式如下:

$$W_{ij}^{T} = \begin{cases} \frac{1}{p} \sum_{p=1}^{p} (t_{i,p} - \bar{T}_{i})(t_{j,p} - \bar{T}_{j}) \\ \sqrt{\sum_{i=1}^{p} (t_{i,p} - \bar{T}_{i})^{2}} \sqrt{\sum_{j=1}^{k} (f_{j,p} - \bar{T}_{j})^{2}} \\ 0, & \text{otherwise} \end{cases}$$
(6)

其中, $t_{i,p}$ 表示节点 v_i 在时间步长 p 处的速度 值, $T_i = \{t_{i,1}, t_{i,2}, \dots, t_{i,p}\}$ 。

2.2.2 基于注意力的多图融合

近年来,多图融合^[19]的方法很多,如取平均和 或加权和合并等,但这些方法效果都有欠佳之处。 本文提出了一种多图融合方法,该方法的整个过程 参见图1。涉及到的伪代码描述如下。

- 算法1 基于注意力机制的多图融合
- 输入 权重矩阵 $W^{D}, W^{A}, W^{E}, W^{P}, W^{T}$
- 参数 批次数 Bt

输出 融合后的权重矩阵 W*

1. 将权重矩阵堆叠成张量 **T**⁽⁰⁾ ∈ ℝ^{|G|×N×N}

- 2. 训练 MSTGNN 模型时进行训练 T⁽⁰⁾
- 3. for i in 0: Bt 1 do
- 4. $\boldsymbol{T}^{(i+1)} \leftarrow \boldsymbol{T}^{(i)}$
- 5. i = i + 1

6.
$$W_{jk}^* = \sum_{i=1}^{|G|} T^{(i)}(i,j,k)$$

7. end for

8. 返回 W^{*}

研究中,构造了一个可训练的权重张量,作为基于注意力多图融合(MGFBOA)的输入;同时,在多 图空间嵌入(MGSE)中整合了空间和图形信息,作 为空间图形注意块(SG Att Block)的输入(见图2)。 图2中,SE、GE分别是 MGSE 中的节点嵌入和多图 嵌入,SA和 GA分别是 SG Att Block 中的空间注意 力和图注意力。此外,为了便于残余连接,所有层都 产生 D 维输出。接下来,将展开探讨论述如下。





Fig. 2 Multi-graph embedding and Spatial Graph attention module

(1)多图空间嵌入(MGSE)。由于构造了多个 图,每个节点在不同的图中有不同的角色,因此将道 路网络信息纳入预测模型至关重要。研究中,本文 采用了一种空间嵌入方法,将顶点编码为保留图结 构信息的向量,即利用 node2vec 方法^[20]来学习节点 表示 $E_{v_i}^s \in R^o$ 。但单独的空间嵌入仅提供静态表 示,无法表示不同图中节点之间的关系,进一步提出 了一种多图嵌入 $E_{c_i}^{MC} \in R^o$,其中 G^i 为任意图。因 此,本文采用了 2 个嵌入表示一个节点,将空间嵌入 和多图嵌入融合为多图空间嵌入 $E_{v_i,G^i} = E_{v_i}^s + E_{G^i}^{MC}$, 作为每个节点在每个图的嵌入。

(2)空间图形注意块(SG Att Block)。在得到 多种类型的图后,下一步就是如何实现每个节点以 及多个图上的信息融合。在上述的 5 个图中,任何 节点都会受到其他不同级别节点的影响,这种影响 是高度动态的。为了更好地刻画节点之间的关系, 本文建立了一种多图注意力块,其中包括空间注意、 图注意和一个门控融合(见图 2)。将第 l 块的输入 表示为 $H^{(l-1)}$,其中顶点 v_i 在图 G^i 的隐藏状态表示 为 $h_{v_i,G^i}^{(l-1)}$ 。第 l 块空间和时间注意机制的输出分别表 示为 $H_s^{(l)}$ 和 $H_c^{(l)}$ 。门控融合后,得到第 l 块的输出, 表示为 $H^{(l)}$ 。

(3)空间注意力机制(Spatial Attention)。通过 空间注意力机制来自适应地捕捉节点之间的相关 性。关键思想在于在每个图的内部,动态地为不同 节点分配不同的权重,如图 3 所示。对于不同的图 *Gⁱ* 计算所有顶点的加权和:

$$hs_{v_{i},G^{i}}^{(l)} = \sum_{v_{k} \in V} \alpha_{v_{i},v_{k}} \cdot h_{v_{k},G^{i}}^{(l-1)}$$
(7)

其中, α_{v_i,v_k} 表示顶点 v_k 对 v_i 重要性的注意力得分。



图 5 至问注意力 Fig. 3 Spatial Attention

对于上述注意力得分,考虑到一个图中的节点 不仅受到其他节点的影响,同时还要受到其他图中 节点的影响。对此,将隐藏状态与时空嵌入连接起 来,并采用缩放点积方法计算 v_k 与 v_i 之间的相关 性:

$$s_{v_i,v_k} = \frac{\left\langle h_{v_i,G^i}^{(l-1)} \parallel E_{v_i,G^i}, h_{v_k,G^i}^{(l-1)} \parallel E_{v_k,G^i} \right\rangle}{\sqrt{2D}} \qquad (8)$$

$$x_{v_{i},v_{k}} = \frac{\exp(s_{v_{i},v_{k}})}{\sum_{v_{i} \in V} \exp(s_{v_{i},v_{k}})}$$
(9)

其中, V表示所有顶点的集合。

为了稳定学习过程,将 N 个平行注意机制与不同的可学习投影联系起来,将空间注意机制扩展为 多头机制:

$$s_{v_{i},v_{k}}^{(n)} = \frac{\langle f_{s,1}^{(n)}(h_{v_{i},G^{i}}^{(l-1)} \parallel E_{v_{i},G^{i}}), f_{s,2}^{(n)}(h_{v_{k},G^{i}}^{(l-1)} \parallel E_{v_{k},G^{i}}) \rangle}{\sqrt{d}}$$
(10)

$$\alpha_{v_{i},v_{k}}^{(n)} = \frac{\exp(s_{v_{i},v_{k}}^{(n)})}{\sum_{v_{k} \in V} \exp(s_{v_{i},v_{k}}^{(n)})}$$
(11)

$$hs_{v_{i},G^{i}}^{(l)} = \| \sum_{n=1}^{N} \{ \sum_{v_{k} \in V}^{v_{k} \in V} \alpha_{v_{i},v_{k}}^{(n)} \cdot f_{s,3}^{(n)}(h_{v_{k},G^{i}}^{(l-1)}) \}$$
(12)

其中, $f_{s,1}^{(n)}$, $f_{s,2}^{(n)}$, $f_{s,3}^{(n)}$ 分别表示不同的激活函数 在第 n 个头部注意下的非线性投影。

(4)图注意力(Graph Attention)。通过图注意 力机制来获取同一个节点不同图中的自相关性,对 于每个节点,按照 Spatial Attention 类似的方法计算 不同图之间的注意力得分(如图 4 所示),以及缩放 点积方法计算对于节点 v_i、图 G^m和 Gⁿ之间的相关 性,并扩展为多头机制:

$$u_{G^{j},G^{k}}^{(n)} = \frac{\left\langle f_{G,1}^{(n)}(h_{v_{i},G^{j}}^{(l-1)} \| E_{v_{i},G^{j}}), f_{G,2}^{(n)}(h_{v_{i},G^{k}}^{(l-1)} \| E_{v_{i},G^{k}}) \right\rangle}{\sqrt{d}}$$
(13)

$$\beta_{C^{j},C^{k}}^{(n)} = \frac{\exp(u_{C^{j},C^{k}}^{(n)})}{\sum \exp(u_{C^{j},C^{k}}^{(n)})}$$
(14)

$$hg_{v_i,G^j}^{(l)} = \| \sum_{n=1}^{N} \{ \sum_{G^j, G^k \in G}^{G^j, G^k \in G} \beta_{G^j, G^k}^{(n)} \cdot f_{G,3}^{(n)}(h_{v_k, G^k}^{(l-1)}) \} (15)$$

(5) Gated Fusion。在得到空间注意和图注意 后,进一步提取同一图及不同图节点的相关性,本文 设计了一个门控机制来对二者融合。在第*l*块中, 空间注意 *H*^(*l*) 与图形注意 *H*^(*l*) 融合为:

$$H^{(l)} = z \odot H^l_S + (1 - z) \odot H^{(l)}_G$$
 (16)

$$\boldsymbol{z} = \boldsymbol{\sigma} \left(H_S^l \boldsymbol{W}_{z,1} + H_G^{(l)} \boldsymbol{W}_{z,2} + \boldsymbol{b}_z \right)$$
(17)

其中, $W_{z,1} \in \mathbb{R}^{D \times D}$, $W_{z,2} \in \mathbb{R}^{D \times D}$, $b_z \in \mathbb{R}^{D}$ 分 别表示可学习参数;"①"表示元素乘积; $\sigma(\cdot)$ 表示 Sigmoid 激活函数。

通过对空间注意和图形注意的融合,创建了空间图形注意块(SG Att Block)。



Fig. 4 Graph Attention

3 实验结果与分析

为了验证本文所提模型具有较好的性能,将改进后的模型 MSTGCN+AL 与其他多元时间序列模型 在 METR-LA 数据集上的结果进行比较,同时为证明创新部分的有效性进行消融实验和对比实验。

3.1 数据集

本研究中选用的是 METR-LA 数据集^[10](见表 1),该数据集由洛杉矶大都会交通管理局提供,其 中包括从环路检测器收集的交通数据。该数据集包 括 207 个环路检测器的测量数据。这些检测器安 装在高速公路上,用于监控车辆的平均速度。数据 通常以 5 min 为间隔进行收集,这就可以对交通模 式进行详细分析。

	表1	METR-LA	数据	集部分展示	
Table 1	Part	tial display o	of the	METR-LA	dataset

		1 0				
时间			检测器			
	773869	767541	767542	717447	717446	
2012/3/1 0:00:00	64.375	67.625	67.125	61.500	66.875	
2012/3/1 0:05:00	62.667	68.556	65.444	62.444	64.444	
2012/3/1 0:10:00	64.000	63.750	60.000	59.000	66.500	
2012/3/1 0:15:00	0	0	0	0	0	
2012/3/1 0:20:00	0	0	0	0	0	
2012/3/1 0:25:00	57.333	69.000	67.667	61.667	67.333	

首先对数据进行清洗预处理及数据集划分,将 70%的数据用于训练,10%用于验证,20%用于测 试。输入序列长度为12,目标序列包含接下来的12 个后续步骤。

3.2 实验设置

(1)平台。所有实验均在 Linux 系统上进行训练 和测试。(CPU: 鲲鹏 920 7265(128C)@3.0 GHz, GPU: NVIDIA ® Tesla ® A100 PCIe)。

(2)超参数。Adam 优化器对所提方法进行了 优化。学习率设为 1e-4。采用 L1 损失函数来衡量 所提模型的性能。批大小为 64,全局种子设为 0,重 复实验。所有的测试都经过 100 次的训练。

3.3 基线模型

为了对本文提出的模型进行分析和比较,本文 使用几种基线模型对数据集进行了模拟实验。基线 模型有 ARIMA-kal:^[21](FITEE 2017)、WaveNet^[22] (Google 2016)、STGCN^[23](IJCAI 2018)、GMAN^[24] (AAAI 2020)、TS2Vec^[25](AAAI 2021)、CTS - TS2Vec^[26] (IEEE CASE 2023), MSTGCN * [8] (ECAI 2022)

在基线的模型中,ARIMA-kal 是一个时间序列 预测模型,结合了自回归积分移动平均方法和卡尔 曼滤波器。WaveNet 同时采用因果卷积和扩张卷 积,这可能显著增加其计算成本。STGCN 大量依赖 预定义的图结构。GMAN 利用 node2vec 算法在执 行注意机制的同时保留节点的结构信息。TS2Vec 和 CTS-TS2Vec 都是用于时间序列的无监督预训练 方法。当图结构未预先定义时,这些方法可能无法 有效地对多变量时间序列数据进行建模。 MSTGCN*利用多图融合,但其多图仍基于自定义的 邻接矩阵,并未考虑周期性。

3.4 基线实验结果与分析

表 2 显示了不同基线方法在数据集 METR-LA 上的性能比较。本文采用 3 种广泛使用的指标来评估所提出的模型的性能,其中包括平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分比误差(MAPE)。

表 2 各种基准方法在不同时间步数预测下的实验结果

Table 2 Experimental results of various baseline methods conducted for predictions at different time steps

	Horizon 3		Horizon 6			Horizon 12			
Model	MAE	RMSE	MAPE /%	MAE	RMSE	MAPE /%	MAE	RMSE	MAPE /%
ARIMAkal	3.990	8.210	9.60	5.050	10.450	12.70	6.900	13.230	17.40
WaveNet	2.990	5.890	8.04	3.590	7.280	10.25	4.450	8.930	13.62
STGCN	3.330	5.770	8.92	3.740	6.650	10.39	4.590	9.400	12.70
GMAN	4.040	8.530	10.20	4.590	9.850	11.69	5.330	11.210	13.60
TS2Vec	6.000	8.920	13.15	6.340	9.650	14.04	6.930	10.440	15.95
CTS- TS2Vec	3.310	6.630	7.80	3.800	7.510	9.44	4.770	9.020	12.23
MSTGCN *	2.973	5.685	7.92	3.513	7.017	9.93	4.194	8.466	12.27
MSTGCN+AL	2.950	5.629	7.90	3.479	6.832	9.80	4.116	8.231	12.20

通过分析表 2 可以看出,在 METR-LA 数据集的 30 min、45 min 和 60 min 预测中, MAE 都取得了最佳结果,并且在 60 min 的预测中, MSTGCN+AL 的所有评价指标均达到了最佳。在 30 min 预测时, MAE 和 RMSE 取得了最佳结果。对于 MSTGCN+AL,随着时间长度的增加,模型的预测表现仍然最佳。

需要注意的是, MSTGCN*和 GMAN 都具有类 似的多图融合模块,并广泛采用注意力机制,但本文 所提的模型在所有指标上的表现都优于前两者,显 示出本文所提模型架构的有效性。因此,可以说明 本文的 MSTGCN+AL 模型表现出优越的预测能力, 特别是在较长时间预测方面, 如提前1h的预测。 这种特性对实际应用非常有利,为交通管理部门提 供充足的时间来实施基于预测洞察的优化策略,从 而增强交通管理能力。

3.5 消融实验

为验证所提模型关键组件的有效性,设计了几种 MSTGCN*变体,在 METR-LA 数据集上进行了消融研究。具体阐释如下。

(1) MSTGCN+A:含有自适应邻接图的 MSTGCN*, 用创新的自适应邻接图代替基础的邻接定义来得到 邻接图。

(2) MSTGCN+L: 含有潜在式图的 MSTGCN*, 用创新的潜在试图代替原先线性拟合直方图来表示 时空训练数据的概述。其中,每个 bin 表示一个预 定义的时间范围,条形高度测量落入每个 bin 的数 据记录的数量。

为了评估本文所提模型中每个组件的实际性能,按照上述实验设置来运行消融实验,每个模型也运行100 epoch,实验结果见表3。

表 3 消融实验 Table 3 Results of ablation experiments

Methods	MSTGCN*	MSTGCN+A	MSTGCN+L
MAE	3. 477 4	3. 467 3	3. 462 5
RMSE	6.8714	6. 863 3	6. 839 2
MAPE	0.097 5	0.097 3	0.0976

METR-LA 数据在模型 MSTGCN*上的消融实

验结果如图 5 所示。从图 5 可以看出,通过消融实 验,证明了 2 个图的改进方法是有效的。2 个图的 性能如下:

(1)自适应邻接图的性能:自适应邻接图的引入极大地改善了结果,因在其中结合了自适应学习和图学习的思想,能够自适应地学习图的结构和特征,使信息流能够在孤立但相互依赖的节点之间流动,从而提高算法的性能和鲁棒性。

(2)潜在式图的性能:潜在式图效果也很明显, 有效提升了对长期周期性及时空关系的捕捉能力, 可以表示历史数据或人类洞察力的长期时空依赖 性,可广泛用于各种图神经网络。





Fig. 5 The ablation experiments conducted on the METR-LA dataset using the MSTGCN* model

3.6 对比试验

为了评估改进后模型 MSTGCN+AL 的有效性与 稳定性进行了对比试验。在相同的参数条件下,将 MSTGCN+AL 模型与 MSTGCN*模型各进行了多次 实验操作,取最大最小值以及均值的 MAE、RMSE、 MAPE 来进行评估,以此说明模型在数据上运行的 稳定性,结果见表4。 从表4的3个评价指标可以看出,在分布性和 收敛性方面2种算法的差别很大。本文中改进后的 MSTGCN+AL模型比原先论文中的MSTGCN*模型 有一定的优势。对于MAE、RMSE、MAPE指标来 说,MSTGCN+AL模型在数值上均优于MSTGCN*模 型,充分说明了MSTGCN+AL模型的有效性和稳 定性。

表 4 MSTGCN+AL 模型与 MSTGCN*模型稳定性评估 Table 4 Stability evaluation of MSTGCN+AL model and MSTGCN* model

	MSTGCN *			MSTGCN+AL		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Maximum	3.477 4	6. 871 4	0.0977	3.479 2	6.8872	0.097 5
Minimum	3.467 6	6.8506	0.097 5	3.437 6	6.713 2	0.0968
Mean	3.473 9	6.8623	0.097 6	3.460 4	6.8184	0.097 3

4 结束语

本文提出了一种针对长期时间序列数据预测框架,采用基于注意力机制的多图神经网络,预测路网 图上超前时间步长的交通状况。在实际数据上的大 量实验证明了所提出的模型在提高交通流量的预测能力方面的有效性。

未来,需要将影响交通预测的其他因素考虑进 来,如:经济因素、收费因素、其他运输方式等,也需 加强与高速公路工作人员的合作,以获取更多的高

63

速公路环路探测器测量的平均交通速度。同时,将 把多图建模的框架应用到其他基于图的时空预测任 务中,为使用图神经网络(GNN)处理多种非结构数 据奠定基础。

参考文献

- [1] 马蕴一,许明,金海波. 多通道自适应特征融合的城市路网交 通流量预测[J/OL]. 计算机工程与应用. [2024-05-28]. https://link.cnki.net/urlid/11.2127.TP.20240524.1035.002.
- [2] 李以诚,叶青.基于交通数据挖掘的交通流预测模型构建[J].
 自动化技术与应用,2023,40(3):39-41.
- [3] 李文娟. 基于图模型与向量自回归的时间序列预测算法研究 [D]. 桂林:桂林电子科技大学, 2023.
- [4] 李玲玲, 辛浩. 时序模型 ARIMA 在数据分析中的应用[J]. 福 建电脑, 2024, 40(4): 25-29.
- [5] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging ffeld of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains
 [J]. IEEE Signal Processing Magazine, 2013, 30(3): 83-98.
- [6] 尹恒,张凡,李天瑞.基于多邻接图与多头注意力机制的短期 交通流量预测[J].计算机科学,2023,50(4):40-46.
- [7] WU Zonghan, PAN Shirui, LONG Guodong, et al. Graph wavenet for deep spatialtemporal graph modeling [C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence(IJCAI'19). New York; ACM, 2019: 1907–1913.
- [8] SHAO Wei, JIN Zhiling, WANG Shuo, et al. Long-term spatiotemporal forecasting via dynamic multiple – graph attention [J]. arXiv preprint arXiv, 2204. 11008, 2022.
- [9] WANG Senzhang, CAO Jiannong, YU P S. Deep learning for spatio-temporal data mining: A survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2019,34(8): 3681-3700.
- [10] YE Jiaxia, ZHAO Juanjuan, YE Kejiang, et al. How to build a graph-based deep learning architecture in traffic domain: A survey
 [J]. arXiv preprint arXiv, 1911. 08451, 2020.
- [11] BUI K H N, CHO J, YI H. Spatial temporal graph neural network for traffic forecasting: An overview and open research issues[J]. Applied Intelligence, 2022, 52(3): 2763–2774.
- [12] JIANG Weiwei, LUO Jiayun. Graph neural network for traffic forecasting: A survey [J]. Expert Systems with Applications, 2022, 207: 117921-117921.
- [13] GAO Nan, XUE Hao, SHAO Wei, et al. Generative adversarial

networks for spatio – temporal data: A survey [J]. ACM Transactions on Intelligent Systems and Technology, 2022, 13 (2): 1–25.

- [14] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs [J]. arXiv preprint arXiv, 1312. 6203, 2013.
- [15] DEFFERRARD M, BRESSON X, VANDERGHEVNST P. Convolutional neural networks on graphs with fast localized spectral filtering [J]. Advances in Neural Information Processing Systems, 2016, 29: 3844–3852.
- [16] CHAI Di, WANG Leye, YANG Qiang. Bike flow prediction with multi-graph convolutional networks [C]//Proceedings of ACM Conference. New York: ACM, 2017: 1-9.
- [17] LV Mingqi, HONG Zhaoxiong, CHEN Ling, et al. Temporal multi-graph convolutional network for traffic flow prediction [J].
 IEEE Transactions on Intelligent Transportation Systems, 2021, 22 (6): 3337-3348.
- [18]黄琨. 基于多特征融合的图注意力股票走势预测研究[D]. 杭州:浙江科技大学, 2024.
- [19]杨国亮,习浩,龚家仁,等.基于多图融合的时空交通流预测 方法[J].传感器与微系统,2023,42(10):126-129.
- [20]徐东伟,朱宏俊,周磊,等.数据缺失下的交通流预测方法研究[J].武汉理工大学学报(交通科学与工程版),2024,48
 (2):211-217.
- [21] CHEN Weiqi, CHEN Ling, XIE Yu, et al. Multi-range attentive bicomponent graph convolutional network for traffic forecasting
 [J]. arXiv preprint arXiv, 1911. 12093, 2019.
- [22] OORD A, DIELEMAN S, ZEN H, et al. Wavenet: A generative model for raw audio[J]. arXiv preprint arXiv, 1609. 03499,2016.
- [23] YU Bing, YIN Haoteng, ZHU Zhanxing. Spatio-temporal graph convolutional networks: A deep learning framework for trafffc forecasting [C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI '18). Stockholm, Sweden ; AAAI, 2018; 3634–3640.
- [24] ZHENG Chuanpan, FAN Xiaoliang, WANG Cheng, et al. Gman: A graph multi-attention network for trafffcn prediction [J]. arXiv preprint arXiv, 1911. 08415, 2019.
- [25] YUE Zhihan, WANG Yujing, DUAN Juanyong, et al. TS2VEC: Towards universal representation of time series [J]. arXiv preprint arXiv,2106. 10466, 2021.
- [26] WANG L, BAI L, LI Z, et al. Correlated time series selfsupervised representation learning via spatiotemporal bootstrapping
 [J]. arXiv preprint arXiv,2306.06994, 2023.