

基于非线性交通流的深度时空图预测模型研究

田 芮¹, 徐志昊², 吕志强^{3*}

¹大连东软信息学院应用技术学院, 辽宁 大连

²青岛大学计算机科学技术学院, 山东 青岛

³青岛理工大学人工智能学院, 山东 青岛

收稿日期: 2026年7月3日; 录用日期: 2026年7月31日; 发布日期: 2026年8月10日

摘 要

在城市交通运营和管理中, 快速、准确的交通流预测至关重要。传统的预测方法往往在处理由交通流的高非线性和复杂性带来的预测挑战方面存在不足, 特别是在捕获交通流的动态空间关联性方面的缺失。图卷积网络被视为建立交通流的时空预测模型的关键技术之一。然而, 在深度交通流时空预测模型中, 这些模型往往面临过度平滑和梯度消失等问题。因此, 本研究提出了一种深度时空图预测模型, 重点解决了图卷积网络在交通流预测模型中的深度构建问题。该模型由一个结合了深度图卷积网络和时间卷积的时空预测模型组成, 通过构建一种新式的可微广义聚合函数和残差连接策略, 本研究实现了多层图卷积网络的信息聚合, 并建立了一个能够捕获时间相关性特征的时间卷积模块。通过在真实交通数据集上的一系列实证测试, 本研究验证了该模型的有效性超过了最新的基准方法, 总体平均预测误差降低了30.4%。

关键词

交通流, 预测方法, 时空模型, 图卷积网络, 信息聚合

Research on Deep Spatial-Temporal Graph Prediction Model Based on Non-Linear Traffic Flow

Rui Tian¹, Zhihao Xu², Zhiqiang Lyu^{3*}

¹School of Applied Technology, Dalian Neusoft University of Information, Dalian Liaoning

²College of Computer Science & Technology, Qingdao University, Qingdao Shandong

³School of Artificial Intelligence, Qingdao University of Technology, Qingdao Shandong

Received: July 3, 2026; accepted: July 31, 2026; published: August 10, 2026

*通讯作者。

文章引用: 田芮, 徐志昊, 吕志强. 基于非线性交通流的深度时空图预测模型研究[J]. 计算机科学与应用, 2026, 16(8): 48-59. DOI: 10.12677/csa.2026.168261

Abstract

In urban traffic operation and management, swift and accurate traffic flow prediction holds immense importance. Traditional prediction methods often fall short in dealing with the prediction challenges brought by the high non-linearity and complexity of the traffic flow, especially lacking in deciphering the dynamic spatial correlations of the traffic flow. Graph convolutional networks are seen as the key technology to establish spatial-temporal prediction models for traffic flow. However, in deep traffic flow spatial-temporal prediction models, these models often face issues such as over-smoothing and gradient vanishing. Therefore, this study proposes a deep graph spatial-temporal prediction model, focusing on addressing the issue of the depth construction of graph convolutional networks in traffic flow prediction models. This model consists of a spatial-temporal prediction model that combines deep graph convolutional networks and temporal convolution. By constructing a differentiable generalized aggregation function and a residual connection strategy, this study realizes the information integration of multiple layers of the graph convolutional network. Simultaneously, by introducing the convolution mechanism, it establishes a temporal convolution module capable of capturing time-related characteristics. Through a series of empirical tests on genuine traffic datasets, this study validates that the model effectiveness exceeds the latest benchmark methods, and on the whole, the predictive error is reduced within the range of 30.4%.

Keywords

Traffic Flow, Prediction Method, Spatial-Temporal Model, Graph Convolutional Network, Information Integration

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 概述

在交通流预测过程中,由于交通流常常表现出明显的空间关联性,因此把握交通网络的空间依赖性对完成交通流预测任务具有重要意义。例如,某一区域的交通流量突然增加,其周边区域的交通流量可能也会受到影响。图卷积网络(Graph Convolutional Network, GCN) [1]是一种处理非欧式拓扑结构数据的深度学习模型,被广泛应用于交通流预测任务。交通网络拓扑结构通常可以被表示为图[2],图中的节点通常代表交通网络中某一区域或交叉口,边则代表连接这些区域或交叉口的道路,每个节点的交通流数据(如交通量、速度等)通常被视为节点的属性。这种图可以用邻接矩阵来表示,邻接矩阵为图中节点间的关系提供了数学描述。在邻接矩阵中,任意两点之间边的属性代表了两者之间的距离。GCN能够直接对图结构执行卷积操作。GCN以每个节点及其相邻区域为基础,对每个节点执行相同的卷积操作,因此GCN具有捕获节点与相邻节点之间空间关系的能力。GCN能够有效捕捉交通网络的空间关联性,进一步提高了交通流预测模型的准确性。

然而,在面对错综复杂的城市级交通网络时,构建深度GCN模型具有不可替代的必要性。首先,从交通网络的空间特性来看,城市交通存在显著的远距离关联效应。局部区域的交通拥堵或突发事件其影响并不仅局限于相邻的几个交叉口,而是会像波纹一样顺着干道网络向外扩散,对数公里外的上游枢纽产生联动影响。浅层GCN由于感受野有限,只能捕捉到直接相邻节点的局部特征,唯有通过堆叠更多的网络层数,模型才能跨越更长的拓扑距离,捕获跨区域、大尺度的宏观空间关联。其次,实际的初步实

验结果也证实了这一观点的合理性：随着 GCN 层数的增加，模型由于感受野的扩大，其对全局交通趋势的预测性能在某一临界点前呈现出持续提升的趋势。这表明，更深的网络结构理论上能够带来更强的空间表征能力。然而，当研究人员试图进一步加深网络以追求更高的预测精度时，模型性能却往往遭遇断崖式下跌，这便引出了 GCN 在向深度延伸时难以绕过的过度平滑问题。

过度平滑是 GCN 处理图结构数据时常见的问题[3]。在 GCN 中，每一层都是计算图中节点及其邻近节点的特征。在经过多层图卷积后，节点间的初始特征差异变得模糊。图卷积的每一层都涉及对节点及其邻居特征的平均或加权平均，通过反复进行图卷积操作，不同节点的特征将趋于相同，从而导致不同节点之间的特征相似且难以区分，这种现象被称为过度平滑。随着 GCN 研究的快速发展，越来越多的科研成果被提出[4]。Li 等人[5]认为现有的基于 GCN 的研究大部分属于浅层 GCN 模型研究，且未深入考虑深度 GCN 训练的过度平滑问题。针对上述问题，部分研究人员尝试用不同的方法来训练深度 GCN，但这些研究的深度都限制在 10 层以内。当 GCN 的层数超过 10 层后，GCN 的模型性能明显下降。受深度卷积神经网络训练优势的启发[6]，DeepGCN 提出在 GCN 的基础上调整残差/密集连接，并进行膨胀卷积，从而训练深度 GCN [7]。在 S3DIS、PointNet 和 PPI 数据集上，DeepGCN 获得了良好的预测效果。例如，Bojchevski 等人[8]提出了一种基于 PageRank 的信息传递方案，该方案考虑了循环中的根节点。此外，DropEdge [9]设计了一种从图中随机删除边的方法来解决深度 GCN 训练的过度平滑问题。PairNorm 设计了一种归一化方法来解决深度 GCN 训练的过度平滑问题。然而，上述研究工作均没有从本质上消除 GCN 的过度平滑问题。

综上所述，虽然 GCN 在捕捉节点间空间特征方面具有独特的优势，但当 GCN 的层数增加时，过度平滑和梯度消失问题会阻碍其进一步的发展。在交通流预测领域，这些挑战性的问题表明 GCN 不能有效地学习交通网络中的深度空间模式及其依赖关系。因此，本研究创新地提出了一种图时空预测模型(Graph-based SpatioTemporal Prediction Model, GSTPM)。该模型专门用于解决深度 GCN 构建的复杂性问题，并且将其应用于交通流预测模型的建模。本研究的主要贡献如下：

- 本研究创新性地提出了 GSTPM，该模型具备捕捉和学习交通流数据时空特征的能力。GSTPM 采用 GCN 捕获交通节点之间的空间相关性特征。同时，该模型采用时间卷积来捕获交通流数据的时间依赖特征。
- 为了应对深度 GCN 的过度平滑和梯度消失的挑战，本研究提出了一种新的深度 GCN 特征聚合策略。该策略引入了一个可微分的广义聚合函数，旨在更全面地分析和聚合多层 GCN 的特征信息。本研究采用残差连接来解决梯度消失问题，以确保模型在深度增加的同时，仍能保持稳定的学习能力。
- 本研究使用由美国加州交通部和其他合作单位共同建立的性能测量系统(Performance Measurement System, PeMS) [10]提供的统一公共交通数据库来对 GSTPM 进行了实验，并与最先进的基准模型进行了对比。结果表明，GSTPM 显示出了模型的预测优势，其综合预测误差减少了 30.4%。同时，本研究通过实证方法证明，GSTPM 能够解决深度 GCN 的过度平滑和梯度消失的问题。

2. 模型准备工作

2.1. 交通图定义

为了充分利用交通流数据时序信息和空间结构信息，本研究将交通网络拓扑结构抽象为 $G=(V,E,W)$ ，其中， V 表示节点集， E 表示边集，本研究通过 V 和 E 来组成无向图 G 。针对特定的交通应用场景， E 中的边连接两个节点的距离信息被抽象为 d 。由现有研究可知，权重矩阵 W 的设计原则主要包含两个方面，分别是：① 在交通网络中，每个节点都可以通过直接或间接的路径与任何其他节点

连接。因此,从宏观角度看,任何节点的状态都受到所有其他节点的影响。② 基于节点之间的距离,直接相邻节点对当前节点的影响大于间接相邻节点对当前节点的影响。因此,权重矩阵 W 的计算过程可以被抽象为公式(1)。其中, σ^2 和 ϵ 表示用于调整权重矩阵 W 分布情况和稀疏程度的阈值,在本研究中分别被设置为 10 和 0.5。

$$w_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), i \neq j \text{ and } \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \geq \epsilon \\ 0, \text{ otherwise} \end{cases} \quad (1)$$

2.2. 时序交通流预测的定义

本研究根据当前观测到的历史信息序列来定义交通流预测任务,并预测未来的信息序列,如公式(2)所示。其中, $v_t \in \mathbb{R}^N$ 表示在第 t 个时间步下 N 个节点的交通流时序信息。因此,公式(2)可以被理解为通过前 M 个时间步的历史数据来获得未来 H 个时间步下概率最大的时序序列。

$$\hat{v}_{t+1}, \dots, \hat{v}_{t+H} = \arg \max \log P(v_{t+1}, \dots, v_{t+H} | v_{t-M+1}, \dots, v_t) \quad (2)$$

2.3. 不同层 GCN 信息传递定义

$$m_{vu}^l = \rho^l(h_v^l, h_u^l, h_{e_{vu}}^l), u \in N(v) \quad (3)$$

$$m_v^l = \zeta^l(\{m_{vu}^l | u \in N(v)\}) \quad (4)$$

$$h_v^{l+1} = \phi^l(h_v^l, m_v^l) \quad (5)$$

多层 GCN 之间信息传递问题是本研究重点关注的问题之一。节点 v 在第 l 层的信息传递过程如公式(3)和公式(5)所示。在第 l 层中, ρ^l 、 ζ^l 和 ϕ^l 分别表示用于信息构造、信息聚合和节点更新的可微函数。本研究的重点是更新各层节点的特征,同时将节点特征聚合为边特征。信息构造函数 ρ^l 根据节点 v 的特征 h_v^l 、其相邻节点的特征 h_u^l 以及对应的边特征 $h_{e_{vu}}^l$ 来为每个相邻节点 u 构建一个独立信息 m_{vu}^l 。同时, ζ^l 表示一种排列不变的聚合函数。该函数能够将一个无序的信息集合聚合为信息 m_v^l 。 ζ^l 的排列不变性能够在不同层 GCN 处理同构图时保证图信息的不变性和等价性,其中使用的 SoftMax 函数具有对称性。节点更新函数 ϕ^l 将原始节点特征 h_v^l 与聚合信息 m_v^l 进行融合,从而得到变换后的节点特征 h_v^{l+1} 。

3. 模型设计

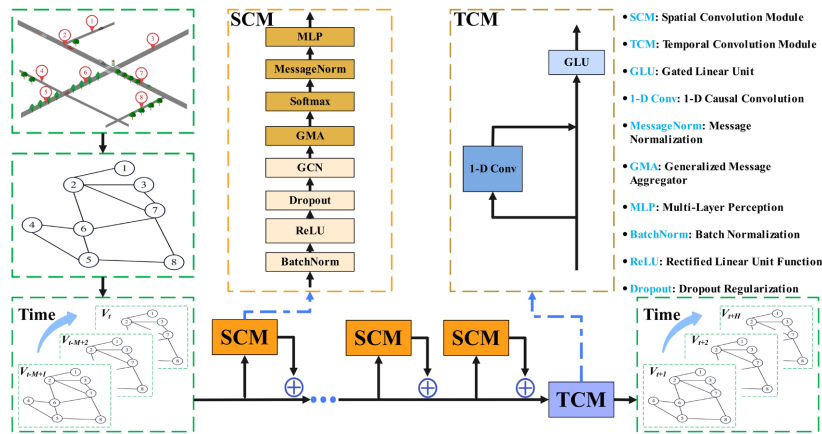


Figure 1. Diagram of GSTPM structure
图 1. GSTPM 架构图

图1展示了GSTPM的基本架构。首先, GSTPM将交通网络和交通流数据转化为图结构的空数据

3.1. 空间卷积模块

近年来,许多研究对GCN的排列不变性进行了深入探讨[11],尤其是在交通流预测领域。GCN的排列不变性赋予了它处理聚合型数据输入的能力,并对所有的连续不变函数展示出了极大的通用性。一系列相关的研究表明,通过结合使用GCN和不变线性层[12],可以有效地构建神经网络模型来完成交通预测任务。在保证GCN排列不变性的前提下,本研究基于GCN的排列不变函数来研究图同构等价性,定义如下:

图同构等价性:在交通流预测中,如果将信息聚合函数定义为 ζ ,该函数能对信息集 $\{m_{vu} | u \in N(v)\}$ 进行置换不变操作,则它将赋予GCN的信息传播算子 f 交通图同构等价性,即给定两个同构图 G_1 和 G_2 ,如果 G_2 是通过某种排列操作从 G_1 获取的,本研究将GCN对算子 F 的操作过程抽象为 $f(G_1) = \sigma * f(G_1)$,其中 $*$ 表示对图的排列操作。

广义信息聚合函数:本研究引入了一个参数化的广义信息聚合函数 $\zeta_x(\cdot)$,该函数以连续变量 x 为参数,输出一组排列不变集合函数,即对于任何给定的 x ,函数 $\zeta_x(\cdot)$ 在信息集合 $\{m_{vu} | u \in N(v)\}$ 的排序中显示出全局的排列不变性。这种方法可以在处理交通图中的节点信息时保持优先顺序的鲁棒性,从而提高预测的准确性。

SoftMax聚合:本研究引入了一种以SoftMax为核心的聚合方法。给定任何信息集合 $\{m_{vu} | u \in N(v)\}$, $m_{vu} \in \mathbb{R}^D$ 。SoftMax_{Agg β} ($\cdot$)表示一般的均值-最大聚合函数,其计算过程如公式(6)所示。

$$\text{SoftMax}_{\text{Agg}\beta}(\cdot) = \sum_{u \in N(v)} \frac{\exp(\beta m_{vu})}{\sum_{i \in N(v)} \exp(\beta m_{vi})} \quad (6)$$

β 表示一个连续变量,称为逆温度。SoftMax函数的主要优点是能够动态调整聚合方法。通过改变参数 β 的值,SoftMax函数可以在平均聚集和最大聚集之间切换。具体来说,当 β 数值较小时,即逆温度 β 较低时,SoftMax函数的计算过程偏重于平均聚合,从而使特征的影响更均匀地分布在图中的节点上。相反,当 β 较大时,即逆温度 β 较高时,SoftMax函数的计算过程偏重于最大聚合,即模型更加关注图中特征最突出的节点。此外,SoftMax函数可以被视为一种特殊的加权平均机制,该机制分配给节点的权重取决于 β 和元素本身的值。这种机制能够突出数据的重要特征,从而帮助模型从数据中提取更具代表性的特征。

广义信息聚合器:广义信息聚合器通过引入聚合函数来处理GCN中节点间的信息传输。在以信息传递为基础来构建有效的GCN时,广义信息聚合器能够确保所有消息特征都为正,即能够使用广义均值-最大值聚合函数SoftMax_{Agg β} ($\cdot$)。由公式(7)定义的信息构造函数 ρ' 描述了在消息传递过程中如何聚合和传输节点特征

$$m'_{vu} = \rho'(h'_v, h'_u, h'_{e_{vu}}) = \text{ReLU}(h'_u + \mathbf{1}(h'_{e_{vu}}) \cdot h'_{e_{vu}}) + \epsilon, u \in N(v) \quad (7)$$

信息归一化:在深度GCN领域中,归一化方法在网络训练中发挥着关键作用[13]。本研究采用批量

归一化(Batch Normalization, BatchNorm)方法来调整节点特征,这有助于缓解梯度消失或梯度爆炸的问题,加速训练收敛,并提高模型的泛化能力。同时,它有助于将网络的激活函数参数波动维持在较小的范围内,以防止特征值的剧烈变化,从而提高网络的稳定性。除了 BatchNorm 之外,本研究设计了一种信息归一化层(Message Normalization, MessageNorm)。MessageNorm 为调整聚合功能表现不佳的网络提供了一个新的视角。MessageNorm 整合了归一化聚合信息的特性以及其他特性,使得网络能够更好地理解和捕获不同节点的信息。在节点更新函数 $\text{MLP}(h_v^l + m_v^l)$ 中应用 MessageNorm 后,网络的学习能力明显提高,并且能够更有效地捕获图结构的复杂空间信息和潜在交互模式,从而提高网络的性能。节点更新函数如等式(8)所示。

$$h_v^{l+1} = \phi^l(h_v^l, m_v^l) = \text{MLP}\left(h_v^l + s \cdot \frac{\|h_v^l\|_2}{\|m_v^l\|_2} \cdot m_v^l\right) \quad (8)$$

通过使用可学习的缩放因子 s , 结合多层感知器 $\text{MLP}(\cdot)$ 和 $\mathcal{L}_2$ 标准化方法,可以有效地调整并平衡聚合信息 m_v^l 与输入 h_v^l 之间的关系。通过将 s 设置为可学习的标量并将其初始值设为 1, 系统能够自动学习并优化缩放因子,从而进一步提升模型的适应性和泛化能力。当缩放因子 s 满足特定条件时,节点更新函数能够以更简化的方式运行,从而进一步优化模型的性能和计算效率。

预激活残差连接: 深度 GCN 的性能通常受到网络深度和结构设计的限制。Li 等人的研究[14]证实了残差连接对训练深度 GCN 网络的有效性,其构建了残差 GCN 块,其中残差 GCN 块的结构顺序是: GCN→归一化→ReLU→求和。然而,此种方法构建的残差 GCN 块的结构顺序对模型性能的影响还有待研究。He 等人[15]提出,残差结构顺序对残差函数的有效性起重要作用,其提出残差函数的输出范围在 $(-\infty, +\infty)$ 之间,而求和之前的激活函数(如 ReLU)会限制模型的表示能力。因此,本研究提出了一种改进的 GCN 结构,即预激活残差连接,其结构顺序为: BatchNorm→ReLU→Dropout→GCN→广义信息聚合器→Softmax→MessageNorm→MLP→特征加权。通过验证,本研究发现预激活残差连接的性能更佳。预激活残差连接不仅考虑了残差函数的输出范围,还通过引入信息聚合器、信息标准化层和多层感知器等组件增强了特征表示能力。这有助于提高模型的泛化性能,并优化信息传输过程,从而使深度 GCN 能够有效学习交通图数据任务。

3.2. 时间卷积模块

本研究利用因果卷积和门线性单元(Gated Linear Unit, GLU)构建时间卷积模块。因果卷积仅利用过去和当前的时序信息进行计算,避免了未来信息对模型训练的影响,并且相关研究已经证明因果卷积在交通流量预测研究中的有效性。此外,当对时序信号进行量化时,因果卷积能够独立处理每个时序输入并确保时序序列的顺序不会被打乱。GLU 通过结合线性变换和非线性门控,实现了简单操作和复杂变换之间的平衡,从而增强了模型的表达能力。通过引入一个非线性门控, GLU 可以动态地控制每个时序特征对输出结果的影响,其在捕捉时间序列的复杂结构和动态差异方面显示出巨大优势。与 RNN 相比,基于线性变换组件的 GLU 能够更好地防止反向传播过程中出现梯度消失的问题,从而增强了模型的深度学习能力和鲁棒性。

本研究在时间卷积模块中将一维因果卷积和 GLU 相结合,这种方法能够在不需要填充的前提下探索交通网络中每个输入元素的 K_i 邻居,且每次序列长度都会减小 $K_i - 1$ 。因此,每个节点的时间卷积输入可以被视为具有 M 个时间步长度和 C_i 个特征通道的序列,即 $Y \in \mathbb{R}^{M \times C_i}$ 。本研究设计了一个卷积核 $\Gamma \in \mathbb{R}^{K_i \times C_i \times 2C_o}$ 来将输入数据映射到单一的输出元素 $[P, Q] \in \mathbb{R}^{(M-K_i+1) \times 2C_o}$ 。其中, P 和 Q 各占一半,并与通道的大小相同。因此,本研究将时间门控卷积过程定义为公式(9)。

$$\Gamma * Y = P \odot \sigma(Q) \in \mathbb{R}^{(M-K_t+1) \times C_o} \quad (9)$$

P 和 Q 表示 GLU 中的门输入。 $\odot$ 表示元素向哈达马积。 $\sigma(Q)$ 表示 Sigmoid 门操作，用于判断当前状态下输入 P 对挖掘时间序列的组成结构和动态差异的重要性。非线性门控能够通过堆叠时间层来充分分析所有输入字段的时间相关性。此外，本研究在不同时间卷积层之间引入残差连接，并通过使用相同的卷积核 Γ 将时间卷积泛化为三维变量，即 $Y_t \in \mathbb{R}^{M \times C_t}$ 。这种独特的方法对于解决交通流预测问题具有巨大的价值。本研究通过捕获交通流的时间动态特征来预测和应对交通信息变化。这对于智能交通系统发展和城市规划等各个方面都具有极其重要的意义。

4. 实验

4.1. 数据集介绍

本研究实验所用数据集为来自美国加州的统一公共高速公路交通数据库 PeMS。PeMS 是一个重要的交通类数据源，其提供了丰富的交通数据。这个实时数据收集系统采集了分布在加利福尼亚州的数万个传感器站点的数据。在这项研究中，本研究收集了 23 条高速公路上 2000 个节点的交通速度数据，涵盖了日常交通的各种模式和状况。这些节点的选择保证了数据具有代表性和多样性。本研究所选交通数据的时间跨度为 2024 年 1 月 1 日至 2024 年 4 月 1 日，其中包括了工作日，休息日和节假日，反映了不同交通模式和交通数据的变化特征。在数据建模之前，本研究进行了全面的数据预处理，以保证数据的质量和准确性。预处理步骤包括处理缺失数据，过滤噪声数据，以及对数据进行归一化处理。对于缺失的数据，本研究使用插值方法进行填补；对于噪声数据，本研究通过一系列的算法进行识别和过滤。同时，本文还对数据进行了归一化处理，用于减小不同属性之间的数值差异，从而提高后续模型建立和分析的准确性。

4.2. 实验设置

本研究使用数据集中前 80% 的数据作为训练样本，并用后 20% 的数据进行测试。训练样本和测试样本之间没有数据重叠。训练过程中使用的参数设置如表 1 所示。本研究以 5 分钟为一个时间步，因此一天可划分为 288 个时间步。模型预测的未来时间步长分别为 3、6 和 9，即预测未来 15 分钟、30 分钟和 45 分钟的交通数据。

Table 1. Training parameter configuration
表 1. 训练参数配置

设置项	数值	设置项	数值
批量大小	128	时间步	5 分钟
训练轮数	200	优化器	RMSprop
学习率	0.01	损失函数	L_2
批量大小	128	时间步	5 分钟
动态学习调整		StepLR [16]，步长为 5，gamma 为 0.7。	

本实验选择了基础模型 GCN、经典时空模型 STGCN、基于注意力的 ASTGCN 和 STAGCN，以及最新的基准模型 STCGCN、STJGCN 和 STRGCN。这些模型的性能评估有助于验证本研究中提出的新方法在交通领域的有效性[17]。

4.3. 实验对比与分析

4.3.1. 定量分析

在交通流预测的实验数据中可以观察到, 随着预测时间的延长(未来 15 分钟、30 分钟、45 分钟), MAE 和 RMSE 的值均呈上升趋势, 如图 2 和图 3 所示。这表明, 随着预测时间的延长, 预测误差也呈上升趋势, 这一现象已在长期交通流预测研究中得到证实。本实验中的所有模型均证实了这一趋势。在所有基准模型中, GCN 模型的预测精度最低, 其 MAE 和 RMSE 的误差值都相对较高。

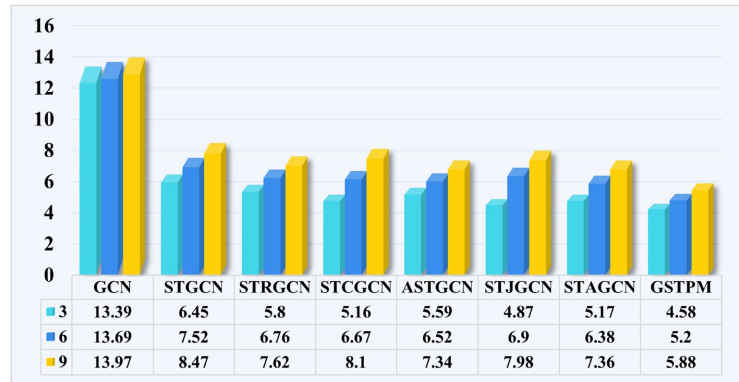


Figure 2. MAE statistical results for all benchmark models

图 2. 所有对比基准模型的 MAE 统计结果

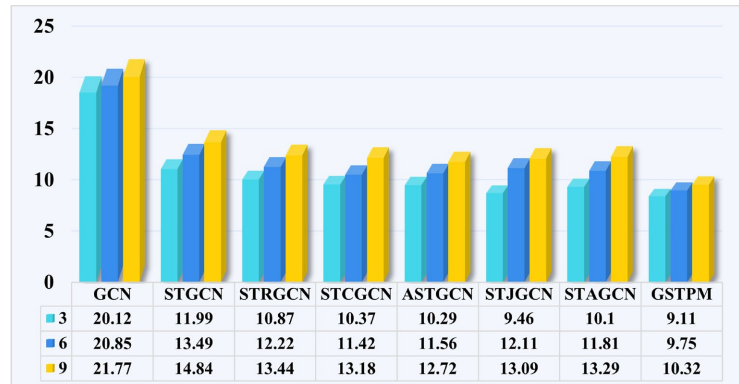


Figure 3. RMSE statistical results for all benchmark models

图 3. 所有对比基准模型的 RMSE 统计结果

这表明该模型在预测未来交通流数据时准确性和稳定性较差。虽然 STGCN 和 STRGCN 模型的预测效果相对较好, 但它们的预测误差仍然在不断增大。STGCN 模型的 MAE 从 6.45 增加到 8.47, RMSE 从 11.99 增加到 14.84。STRGCN 模型的 MAE 从 5.8 增加到 7.62, RMSE 从 10.87 增加到 13.44。这些结果表明, 随着预测时间的增加, 模型的预测精度和稳定性均呈下降趋势。在 STCGCN、ASTGCN、STJGCN 和 STAGCN 模型中, 较低的 MAE 和 RMSE 表明预测精度和稳定性相对较好, 特别是在预测未来 15 分钟内的交通速度时, STJGCN 模型的性能评估指标低于其他基准模型。GSTPM 模型(100 层 GCN 的信息聚合)的 MAE 和 RMSE 性能评估指标均最小, 并展示出最小的误差增加趋势。通过比较本文模型和所有对比模型的逐一对比和差值全局平均化, 本文模型的整体预测误差下降了 30.4%。GSTPM 模型的 MAE 从 4.58 增加到 5.88, 而 RMSE 从 9.11 增加到 10.32, 这表明 GSTPM 模型在维持预测精度和稳定性方面表现出了良好的控制力。

4.3.2. 定性分析

图 4(a)和图 4(b)展示了交通网络中某一节点在一天内交通速度的时序变化情况。在上午 6 点至 8 点的高峰时段, 以及下午 3 点至 5 点的繁忙时段, 车流量大幅度增加, 路段拥堵, 从而导致汽车的平均速度达到一天中的最低点。然而在深夜时间段, 由于车辆数量的减少, 车辆的平均速度相对增加。图 4(c)和图 4(d)展示了模型预测值与真实值之间的残差分析结果, 加深了对不同速度区间预测误差的理解。残差分析图展示了模型预测值与实际值之间的偏差, 以及误差的大小和方向。残差分析图中的散点距离零轴的距离越大, 偏差就越大, 这有助于评估模型预测的准确性。图 4 展示了各种基准模型的预测结果与真实交通速度数据的定性实验对比。在局部极值的预测以及非线性变化的处理方面[18], STRGCN 模型能够拟合交通速度的整体变化趋势, 但无法拟合真实交通数据的局部极值和非线性变化。这一问题在交通速度波动较大的区域尤为明显[19]。这是因为 STRGCN 在捕获交通数据的时序特征时重点关注交通数据的长期特征, 忽视了交通数据的短期波动, 从而导致模型在应对复杂和非线性的交通流变化时, 对局部变化适应性较差。

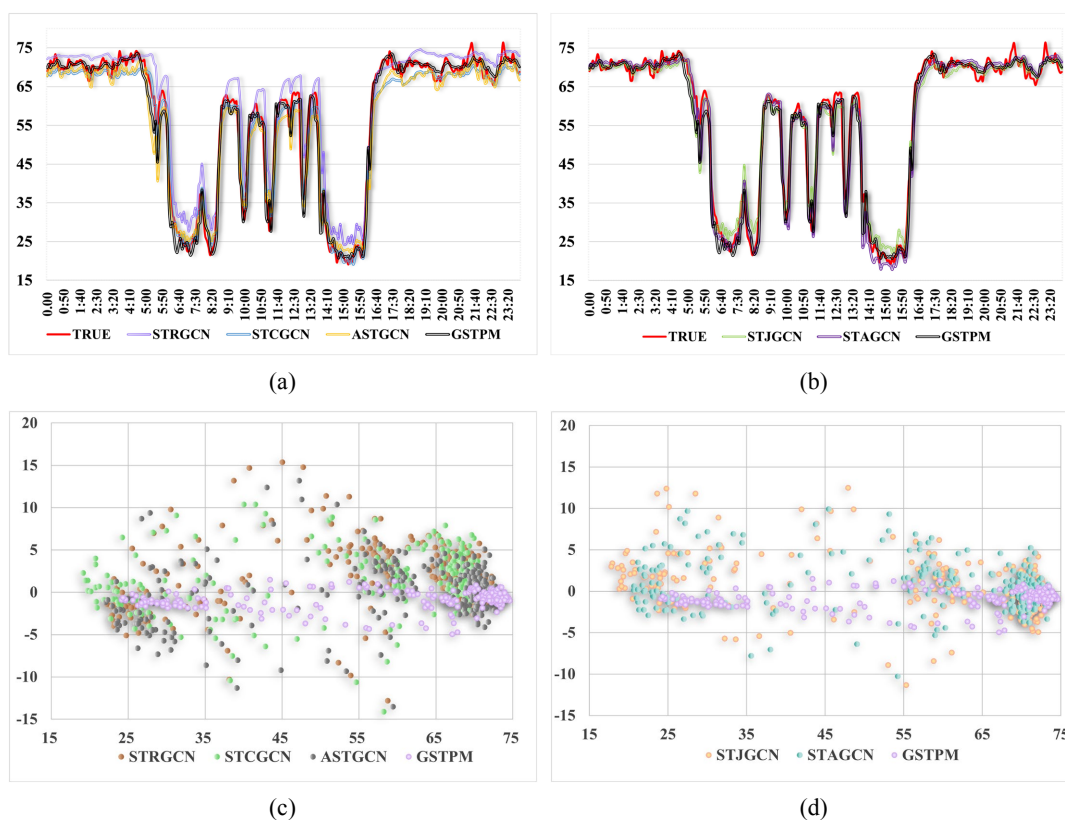


Figure 4. Comparison between the true values and predicted values of models

图 4. 模型真实值与预测值的比较

与 STRGCN 相比, STCGCN、ASTGCN 和 STAGCN 模型的预测结果更接近真实值。然而, 在处理局部极值和非线性变化时, 这些模型显示出了局限性。例如, STCGCN 通过设计更复杂的相关矩阵, 优化了时空特征和外部因素的建模。然而, 在处理高速波动区域的局部极值和非线性特征时[20], 该模型会受到外部因素的干扰, 从而影响其预测性能。另一方面, ASTGCN 和 STAGCN 中的注意力机制和动态图设计能够捕捉交通网络中的全局和潜在动态变化, 但是这种方法会降低模型对复杂局部变化的敏感性。与上述基准模型相比, STJGCN 模型在大多数场景下能够准确地完成交通速度预测任务。该模型通过引

入膨胀型因果时空联合图卷积层来捕捉不同范围内的时空依赖关系。虽然在预测局部极值时仍然存在误差，但其对交通数据非线性变化具有较强的分析能力。本研究所提出的 GSTPM 模型通过合理的聚合函数来捕获深度 GCN 空间相关特征，并通过时间序列卷积进一步学习长期时间依赖特性。因此，在基于局部极值和非线性变化的交通流预测任务中，GSTPM 表现出优越的预测性能。

4.3.3. 消融实验

在 GSTPM 模型中，GCN、聚合函数以及其他优化方法共同构建了 GSTPM 的空间卷积模块。本研究通过调整 GCN 的层数，即在空间卷积模块中控制残差连接的数量，来进行消融实验，并以此来测试 GCN 的层数对 GSTPM 整体效果的影响。相关的实验结果在表 2 中详细列出。此外，本研究通过调整参考模型中的 GCN 层数来对 GSTPM 在解决与 GCN 相关的过度平滑和梯度消失问题的有效性进行了验证，从而证明与基准模型相比，GSTPM 模型具有明显的优越性。

Table 2. The impact of the number of GCN layers on model prediction results
表 2. GCN 层数对模型预测结果的影响

模型	层数	MAE	RMSE	模型	层数	MAE	RMSE
GCN	1	13.97	20.77	STGCN	1	8.47	14.84
	5	14.91	22.18		5	9.03	15.55
	10	15.84	23.6		10	9.46	15.86
	20	17.71	26.43		20	10.68	17.08
	50	null	null		50	null	null
	100	null	null		100	null	null
STRGCN	1	7.62	13.44	STCGCN	1	8.1	13.18
	5	8.16	14.36		5	8.62	14.07
	10	8.71	15.28		10	9.15	14.95
	20	9.8	16.99		20	10.2	16.73
	50	null	null		50	19.4	30.21
	100	null	null		100	null	null
ASTGCN	1	7.34	12.72	STJGCN	1	7.98	13.09
	5	7.84	13.59		5	8.52	13.98
	10	8.34	14.47		10	9.07	14.86
	20	9.35	16.21		20	10.15	16.64
	50	18.63	32.01		50	26.42	32.36
	100	null	null		100	null	null
STAGCN	1	7.36	13.29	GSTPM	1	6.86	12.80
	5	7.87	14.18		5	6.82	12.72
	10	8.39	15.06		10	6.76	12.61
	20	9.42	16.83		20	6.66	12.51
	50	18.56	15.96		50	6.29	11.53
	100	null	null		100	5.88	10.32

除 GSTPM 模型外,随着 GCN 层数的增加,其他模型的 MAE 和 RMSE 会随之增加。这表明,GCN 层数过多可能会造成过度平滑问题,从而导致模型过于关注某些细微的细节,忽略了交通数据的主要特征,最终影响预测的精度。与所有基准模型相比,GSTPM 模型在深度 GCN 特征聚合方面展示了优越的性能。随着层数的增加,MAE 和 RMSE 基本保持稳定,在某些情况下甚至有所下降。特别是当层数达到 50 和 100 时,模型在生成准确预测结果的前提下,保证预测误差相对较低。由此可知,GSTPM 模型能够有效地处理更复杂(即层数更深)的模型结构,同时保持较低的预测误差。综上所述,GSTPM 模型在处理深度模型方面的表现优于其他模型。

GSTPM 模型中模块的消融:如图 5 所示,模型 T 仅采用时间卷积模块,主要侧重于时间序列数据的处理。本研究设计的时间卷积模块采用因果卷积来分析交通时序数据的时间相关性,并采用 GLU 来学习输入特征的复杂交互作用。该模块能够充分捕获交通数据的时间特征,但是,只考虑交通数据的时间特征会导致模型预测精度大幅度下降。

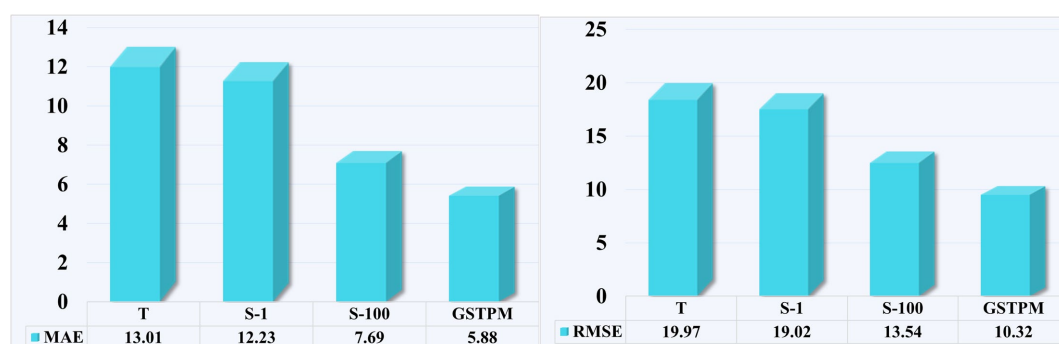


Figure 5. Ablation study results for the modules in GSTPM

图 5. GSTPM 中各模块的消融实验结果

S-1 和 S-100 模型分别通过不同层数的 GCN 来构建空间卷积模块,其中 S-1 模型中融入了单层 GCN, S-100 模型中融入了 100 层 GCN。多层 GCN 模型能够对多层次特征信息进行综合整理和归纳。与 T 模型相比,采用单层 GCN 的 S-1 模型略有改进。这说明空间信息的引入增强了模型预测能力,从而证明了采用 GCN 处理图结构数据并收集空间特征信息能够有效提升模型预测的准确性。与 S-1 模块相比,将 GCN 层数增加到 100 层可大幅度提高模型预测性能。由图 5 可知,深度 GCN 能够有效提取交通网络特征信息。通过分析消融实验中 GSTPM 模型与其他消融模型的实验结果可知,将时间信息和空间信息进行特征融合能够有效提高模型的预测精度,并获得准确的预测结果。本研究的实验结果证明了深度 GCN 特征聚合策略的有效性和必要性,以及 GSTPM 的优越性。在模型中融入时间卷积模块和深度空间卷积模块后,模型对交通数据的预测性能得到了显著提升,预测误差大幅度减小。

5. 结论

GCN 作为捕捉交通节点空间相关性的重要方法,在交通流预测领域中被广泛使用。虽然 GCN 在捕获交通网络空间特征方面展示了其优越性,但是随着 GCN 层数的增加,深度 GCN 会出现过度平滑和梯度消失问题。因此,本研究提出了一种基于深度 GCN 信息融合的深度图时空预测模型。针对深度 GCN 在模型训练过程中出现的过度平滑和梯度消失问题。虽然本研究已经成功克服了深度 GCN 训练中的过度平滑和梯度消失问题,并显著减少了 GSTPM 的预测误差,但深度 GCN 模型的参数量较高,且训练时间较长。首先,基于深度 GCN 构建的模型将包含大量的模型参数,这不仅增加了模型的存储和计算成本,还会导致模型在训练过程中出现过拟合问题,从而降低模型的通用性和鲁棒性。其次,基于深度 GCN 构

建的模型在带迭代训练的过程中对计算资源的需求不断增加。因此,模型的训练时间将相应地增长,从而影响模型预测的实时性和运行效率。因此,如何在提升模型性能和控制计算成本之间实现平衡,仍然需要进一步的研究。

参考文献

- [1] 叶宝林, 戴本岳, 张鸣剑, 等. 基于图卷积网络的交通流预测方法综述[J]. 南京信息工程大学学报, 2024, 16(3): 291-310.
- [2] 刘建敏, 林晖, 汪晓丁, 等. 基于图注意力机制的无地图场景轨迹预测方法[J]. 计算机工程, 2024, 50(7): 144-153.
- [3] ud din, A.M. and Qureshi, S. (2024) Limits of Depth: Over-Smoothing and Over-Squashing in GNNs. *Big Data Mining and Analytics*, 7, 205-216. <https://doi.org/10.26599/bdma.2023.9020019>
- [4] 王竟成, 张勇, 胡永利, 等. 基于图卷积网络的交通预测综述[J]. 北京工业大学学报, 2021, 47(8): 954-970.
- [5] Li, J., Zhang, Q.R., Liu, W.X., Chan, A.B. and Fu, Y.G. (2024) Another Perspective of Over-Smoothing: Alleviating Semantic Over-Smoothing in Deep GNNs. *IEEE Transactions on Neural Networks and Learning Systems*, 36, 6897-6910.
- [6] 崔建勋, 要甲, 赵泊媛. 基于深度学习的短期交通流预测方法综述[J]. 交通运输工程学报, 2024, 24(2): 50-64.
- [7] Gasteiger, J., Bojchevski, A. and Gunnemann, S. (2018) Predict Then Propagate: Graph Neural Networks Meet Personalized PageRank. arXiv: 1810.05997.
- [8] Bojchevski, A., Klicpera, J., Perozzi, B., et al. (2019) Is PageRank All You Need for Scalable Graph Neural Networks. *ACM KDD, MLG Workshop*, Anchorage, 5 August 2019, 1-7.
- [9] Cai, K.Q., Li, Y., Zhu, Y.W., Fang, Q., Yang, Y. and Du, W. (2023) A Geographical and Operational Deep Graph Convolutional Approach for Flight Delay Prediction. *Chinese Journal of Aeronautics*, 36, 357-367. <https://doi.org/10.1016/j.cja.2022.10.004>
- [10] Giraldo, M., Restrepo, J.C., Huertas, J., Agudelo, J.R. and Agudelo, A.F. (2024) Signal Synchronization Methods When Measuring Tailpipe Emissions with PEMs. *Transportation Research Part D: Transport and Environment*, 129, Article ID: 104154. <https://doi.org/10.1016/j.trd.2024.104154>
- [11] Chen, J., Shi, J. and Xu, R. (2024) Dual Subspace Manifold Learning Based on GCN for Intensity-Invariant Facial Expression Recognition. *Pattern Recognition*, 148, Article ID: 110157. <https://doi.org/10.1016/j.patcog.2023.110157>
- [12] Chen, K., Liu, S., Zhu, T., Qiao, J., Su, Y., Tian, Y., et al. (2023) Improving Expressivity of GNNs with Subgraph-Specific Factor Embedded Normalization. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Long Beach, 6-10 August 2023, 237-249. <https://doi.org/10.1145/3580305.3599388>
- [13] 刘建伟, 赵会丹, 罗雄麟, 等. 深度学习批归一化及其相关算法研究进展[J]. 自动化学报, 2020, 46(6): 1090-1120.
- [14] Li, G., Muller, M., Thabet, A. and Ghanem, B. (2019) DeepGCNs: Can GCNs Go as Deep as CNNs? 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 9266-9275. <https://doi.org/10.1109/iccv.2019.00936>
- [15] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Identity Mappings in Deep Residual Networks. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016*, Springer, 630-645. https://doi.org/10.1007/978-3-319-46493-0_38
- [16] Wen, X., Zeng, M., Chen, J., Maimaiti, M. and Liu, Q. (2023) Recognition of Wheat Leaf Diseases Using Lightweight Convolutional Neural Networks against Complex Backgrounds. *Life*, 13, Article 2125. <https://doi.org/10.3390/life13112125>
- [17] 何婷, 周艳秋, 辛春花. 基于时空位置关注图神经网络的交通流预测方法[J]. 计算机应用研究, 2024, 41(10): 2932-2938.
- [18] Edlinger, A., Garland, G., Banerjee, S., Degrune, F., García-Palacios, P., Herzog, C., et al. (2023) The Impact of Agricultural Management on Soil Aggregation and Carbon Storage Is Regulated by Climatic Thresholds across a 3000 km European Gradient. *Global Change Biology*, 29, 3177-3192. <https://doi.org/10.1111/gcb.16677>
- [19] Stiles, J., Kar, A., Lee, J. and Miller, H.J. (2023) Lower Volumes, Higher Speeds: Changes to Crash Type, Timing, and Severity on Urban Roads from COVID-19 Stay-at-Home Policies. *Transportation Research Record: Journal of the Transportation Research Board*, 2677, 15-27. <https://doi.org/10.1177/03611981211044454>
- [20] 史昕, 曹凤腾, 纪艺, 等. 基于多尺度时空特征和软注意力机制的交通流预测方法[J]. 计算机工程, 2024, 50(12): 346-357.