

基于 SA-GFSTCN 的高速公路交通拥堵预测研究

王庆荣¹, 高桓伊¹, 朱昌锋², 王俊杰¹

(1. 兰州交通大学电子与信息工程学院, 甘肃 兰州 730070; 2. 兰州交通大学交通运输学院, 甘肃 兰州 730070)

摘要: 针对现有交通拥堵预测方法中拥堵指数定义单一、静态-自适应图信息无法有效融合的问题, 设计一种创新的交通拥堵指数(TCI), 并提出基于静态-自适应图融合的交通拥堵预测模型——SA-GFSTCN。首先, 根据平均速度、交通流量和时间占有率 3 项指标反映的道路使用情况和交通流状况, 定义 TCI; 然后, 模型采用并行架构处理输入数据, 使用时空卷积和时空注意力模块对静态路网结构进行处理, 提取固定的结构性信息及其时空特征; 接着, 采用自适应图卷积和门控时间卷积处理自适应图数据, 并提取动态的时空关联特征; 最后, 通过交叉注意力机制将这两部分输出进行有效融合。在 2 个真实的交通数据集上的实验结果表明, SA-GFSTCN 模型在平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、均方根误差(RMSE) 3 项指标上相较于最优基线模型分别提升了 0.27 与 0.20、0.22 与 0.23 百分点、0.38 与 0.36, 验证了 SA-GFSTCN 模型的有效性。

关键词: 交通拥堵预测; 交通拥堵指数; 静态-自适应图融合; 自适应图卷积; 交叉注意力

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070026

Research on Highway Traffic Congestion Prediction Based on SA-GFSTCN

WANG Qingrong¹, GAO Huanyi¹, ZHU Changfeng², WANG Junjie¹

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China;

2. School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China)

【Abstract】 Existing traffic congestion prediction methodologies are based on simplistic definitions of congestion indices and fail to effectively integrate static-adaptive graph information. To address these issues, this paper proposes an innovative Traffic Congestion Index (TCI), and a novel traffic congestion prediction model based on static-adaptive graph fusion called SA-GFSTCN. The TCI is defined based on three metrics, namely average speed, traffic flow, and occupancy rate, which collectively reflect road usage and traffic conditions. The model employs a parallel architecture to process the input data using spatiotemporal convolution and spatiotemporal attention modules to model the static road network structure and extract fixed structural information along with the spatiotemporal characteristics. Concurrently, adaptive graph convolution and gated temporal convolution are used to process adaptive graph data and extract dynamic spatiotemporal associative features. Finally, a cross-attention mechanism effectively fuses the outputs of the adaptive graph convolution and gated temporal convolution. Experiments conducted on two real-world traffic datasets demonstrate that the SA-GFSTCN model outperforms the optimal baseline model in terms of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). Specifically, it achieves improvements of 0.27 and 0.20 in MAE, 0.22 and 0.23 percentage points in MAPE, and 0.38 and 0.36 in RMSE, respectively, across the datasets when compared to the baseline model. These results validate the effectiveness of the proposed model.

【Key words】 traffic congestion prediction; Traffic Congestion Index (TCI); static-adaptive graph fusion; adaptive graph convolution; cross attention

0 引言

近年来,随着城市化的加速及机动车数量的增加,城市交通系统面临严重拥堵问题,导致时间和能源的浪费,加重环境污染,并对居民生活质量及城市可持续发展产生负面影响。在此背景下,精确的交通拥堵预测成为智能交通系统(ITS)实施的核

心^[1]。有效地预测交通拥堵状态的关键在于精确地定义交通拥堵指数(TCI)以及将静态路网信息和自适应路网信息有效整合。

交通拥堵预测作为一个复杂的时间序列问题,主要使用统计学模型、传统机器学习和深度学习 3 种方法。统计学模型通过概率分布分析时间序列行为进行预测,这类模型包括灰色模型(GM)^[2]、自

基金项目: 国家自然科学基金(72161024);甘肃省教育厅“双一流”重大研究项目(GSSYLXM-04)。

作者简介: 王庆荣,女,教授,主研方向为智能交通、应急物流;高桓伊(通信作者),硕士研究生;朱昌锋,教授、博士;王俊杰,硕士研究生。

收稿日期: 2024-06-20

修回日期: 2024-08-26

E-mail: gaohuanyi1@163.com

回归差分移动平均 (ARIMA) 模型^[3]等。LIN 等^[4]结合 ARIMA 和广义自回归条件异方差 (GARCH) 模型, 提出基于 ARIMA-GARCH-M 模型的短期高速公路交通流量预测方法, 但其处理复杂非线性关系和突发事件的能力有限^[5]。为了克服上述问题, 研究者采用机器学习模型来提高预测的准确性和效率。机器学习方法包括 k 近邻 (k -NN)^[6]、贝叶斯网络 (BN)^[7] 和支持向量机 (SVM)^[8] 等。LIN 等^[9] 提出结合支持向量回归 (SVR) 和 k -NN 的方法, 用于基于最大信息系数 (MIC) 的短期交通流量预测, 尽管该方法改进了短期交通流量的预测精度, 但仍然面临非线性模式捕捉有限和数据依赖性的挑战^[10]。

近年来, 随着深度学习领域的快速发展, 越来越多的研究者开始探索其在交通预测问题中的应用。卷积神经网络 (CNN) 等深度学习模型已经显示出卓越的学习能力、精确稳定的预测性能, 并能有效提取深层次特征^[11]。ZHU 等^[12] 采用 CNN 结合城市交通的空间和时间数据, 提出自动化的早期交通拥堵检测方法, 该方法能有效捕捉和分析交通数据中的复杂空间-时间关系, 从而在交通拥堵形成前提供预警。

由于单一的深度学习模型难以同时捕捉交通流量数据中的时空相关性, 因此研究者们提出了能更全面分析这些特征的混合模型^[13]。YU 等^[14] 提出了时空图卷积网络 (STGCN) 用于交通流预测, STGCN 通过构建全卷积结构有效捕捉交通数据中的复杂空间-时间关系。WU 等^[15] 提出了 Graph WaveNet 模型, 该模型结合自适应矩阵和膨胀一维卷积, 有效捕捉数据中的隐藏空间依赖性及长序列时间依赖性。LI 等^[16] 提出了扩散卷积递归神经网络 (DCRNN), 该模型使用扩散卷积操作捕捉交通传感器间的空间依赖性, 并结合 RNN 捕捉时间动态特征。史昕等^[17] 提出了基于多尺度时空特征和软注意力机制的交通流预测 (MSTFSA) 方法。该方法结合图注意力网络和双向增强注意力门控循环单元 (GRU), 通过软注意力机制融合多时间尺度的特征, 显著提升了预测精度。

尽管交通拥堵预测领域取得了显著进展, 但现有方法仍存在诸多局限。首先, 静态图结构采用固定邻接矩阵, 单一固定邻接矩阵忽视实时道路动态, 影响预测准确性和模型适应性^[18-19]。其次, 自适应邻接矩阵在数据稀疏或噪声环境中表现不佳, 且忽略关键的地理和物理先验知识。此外, 具有先验知识的邻接矩阵虽考虑地理和交通流特性, 但缺乏灵活性, 可能忽略未知相关性。最终现有方法依赖静

态数据或动态调整机制, 未能提供一个综合静态与动态信息的框架。这些限制需通过新策略来解决。

针对以上问题, 本文提出了基于静态-自适应图融合时空卷积网络的交通拥堵预测模型——SA-GFSTCN, 主要贡献如下:

1) 提出了新的 TCI 计算方法, 该方法不仅考虑了传统的速度因素, 还整合了交通流量和时间占有率, 能够更加准确反映路段的拥堵程度。

2) 构建了融合静态带权有向图与无先验知识的自适应图的交通拥堵预测模型。带权有向图反映交通基础结构, 自适应图动态捕捉交通变化。这种双重图融合机制增强了模型对于复杂交通环境的适应能力, 提高了预测的准确性和可靠性。

3) 设计了基于交叉注意力机制的输出模块, 该机制通过识别并强化来自不同模块的关键信息, 有效实现深度特征融合, 提高模型对复杂交通状况变化的适应能力和预测准确性。

1 问题描述

交通拥堵预测问题可以定义为基于历史交通数据预测未来拥堵情况。针对拥堵程度的量化问题, 提出 TCI 的定义; 针对交通路网结构建模问题, 提出了静态路网结构图和自适应路网结构图的定义。

定义 1 (TCI) 受 WANG 等^[20] 研究的启发, 在原有仅受速度影响的 $TCI(I_{TCI})$ 计算的基础上提出了新的 TCI 计算公式。

单独考虑速度不足以全面评估交通拥堵的实际情况, 因此需要额外考虑交通流量和时间占有率对拥堵的影响, 即:

$$I_{TCI} = \begin{cases} I_{TCI,s} \cdot \exp\left(\alpha \cdot \frac{Q}{Q_{\max}} + \beta \cdot O\right), & \bar{v} < v_{\text{free-stream}} \\ 0, & \bar{v} \geq v_{\text{free-stream}} \end{cases} \quad (1)$$

式中: $I_{TCI,s}$ 是基于速度的 TCI; $\exp\left(\alpha \cdot \frac{Q}{Q_{\max}} + \beta \cdot O\right)$ 是调整因子, Q 与 $Q_{\max}$ 分别是当前流量和道路最大容量, O 是占有率, α 和 β 分别是通过随机森林分析交通流量和时间占有率对 TCI 的贡献, 表示间接确定的强度系数。

$I_{TCI,s}$ 的计算公式如下:

$$I_{TCI,s} = \begin{cases} 1 - \frac{\bar{v}}{v_{\text{free-stream}}}, & \bar{v} < v_{\text{free-stream}} \\ 0, & \bar{v} \geq v_{\text{free-stream}} \end{cases} \quad (2)$$

式中: $\bar{v}$ 是平均速度; $v_{\text{free-stream}}$ 定义为理想条件下车辆通过路段的速度, 即自由流速度, 通过滑动窗口法

计算得出。

将所有节点在过去 $\mathcal{h}$ 个时间步的 TCI 记作：

$$\mathbf{X}_{\mathcal{h}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1\mathcal{h}} \\ x_{21} & x_{22} & \cdots & x_{2\mathcal{h}} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{N\mathcal{h}} \end{bmatrix}, \mathbf{X}_{\mathcal{h}} \in \mathbb{R}^{N \times \mathcal{h}} \quad (3)$$

式中： $x_{N\mathcal{h}}$ 表示节点 N 在第 $\mathcal{h}$ 个时间步的 TCI。

定义 2(静态路网结构图) 将路网表示为带权有向图 $G_{di} = \{V, E, W\}$, 其中, V 是节点集合, E 是链路集合, $W \in \mathbb{R}^{N \times N}$ 是节点间的权重集合, W_{ij} 为节点 i 到节点 j 的权值, 即:

$$W_{ij} = \begin{cases} \exp\left(-\frac{\log_a(1+d_{v_i, v_j})}{|\rho(v_i, v_j)|^\beta}\right), & |\rho(v_i, v_j)| \geq \theta \\ 0, & |\rho(v_i, v_j)| < \theta \end{cases} \quad (4)$$

式中： d_{v_i, v_j} 表示节点 v_i 和 v_j 的实际距离； $\rho(v_i, v_j)$

v_j) 表示节点 v_i 和 v_j 处交通流的皮尔逊相关系数； θ 为相关性阈值, 设置为 0.05。

定义 3(自适应路网结构图) 自适应路网结构图定义为 $G_{adp} = (V, E)$, 其中, V 是节点集合, E 是基于节点间相互影响力的边集合。为学习 G_{adp} 中的边和权重, 引入自适应邻接矩阵 $A \in \mathbb{R}^{N \times N}$ 表示节点间的潜在关联性, 该矩阵不依赖于预定义的图结构。自适应邻接矩阵 A 通过模型中的参数化方法表示, 即:

$$A = \text{Softmax}(\text{ReLU}(\Theta_1 \Theta_2^T)) \quad (5)$$

式中： $\Theta_1, \Theta_2 \in \mathbb{R}^{N \times K}$ 是模型中可学习的节点嵌入矩阵； N 是图中的节点数； K 是嵌入维度。

2 交通拥堵预测模型构建

针对交通拥堵预测, 构建 SA-GFSTCN 模型, 模型总体架构如图 1 所示。

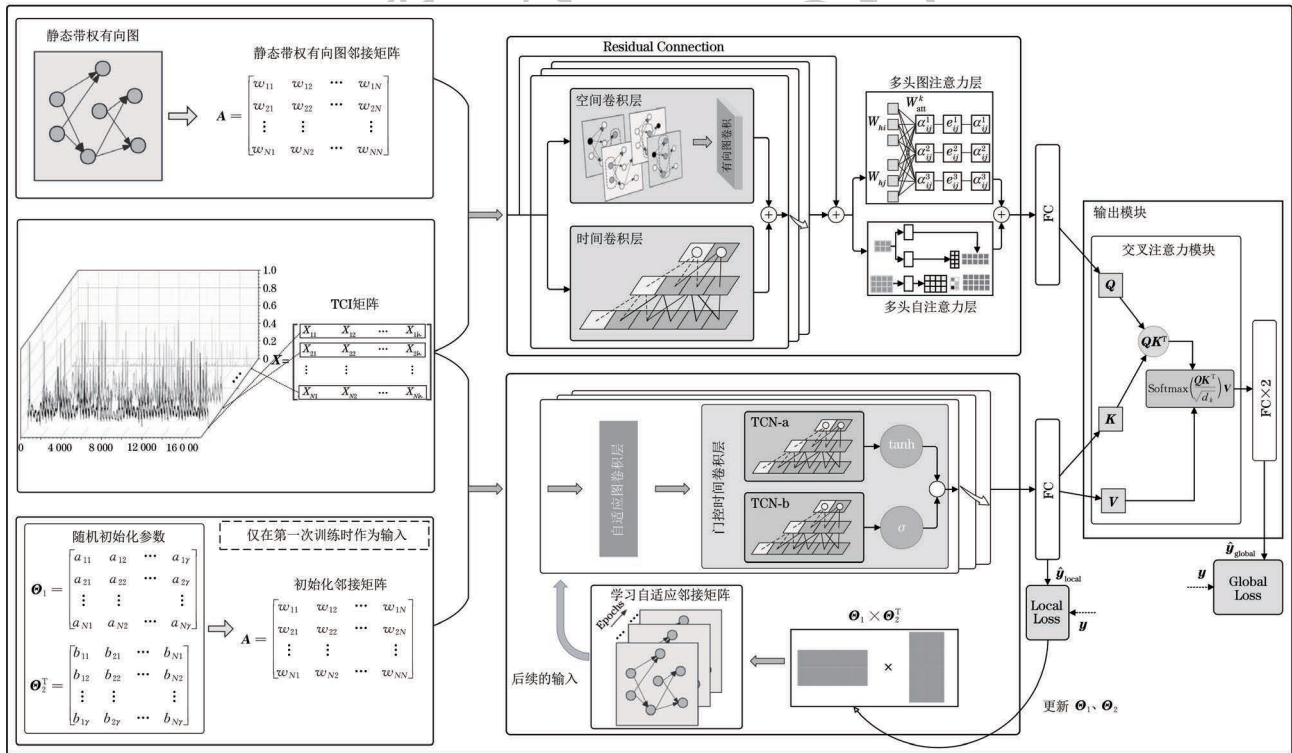


图 1 SA-GFSTCN 模型总体架构

Fig.1 Overall architecture of the SA-GFSTCN model

SA-GFSTCN 模型由三部分构成, 即静态图注意力时空卷积模块、自适应图时空卷积模块和输出模块。静态图模块利用带权有向图和特征矩阵初步分析时空相关性, 并通过时空注意力层深化此分析。自适应图模块以随机数初始化邻接矩阵, 并通过局部 Loss 更新, 同时利用门控时序卷积层捕捉时间依赖性。两模块输出通过交叉注意力层融合以形成最终预测结果。

2.1 静态图注意力时空卷积模块

2.1.1 时空卷积层

采用多层时空卷积堆叠来初步提取静态图的时空相关性, 其中有向图卷积层的输入为邻接矩阵与 TCI 数据, 在时空卷积层引入残差连接来增强网络表达能力, 如图 2 所示。

1) 有向图卷积层。

在静态图的时空卷积层中, 采用有向图卷积作

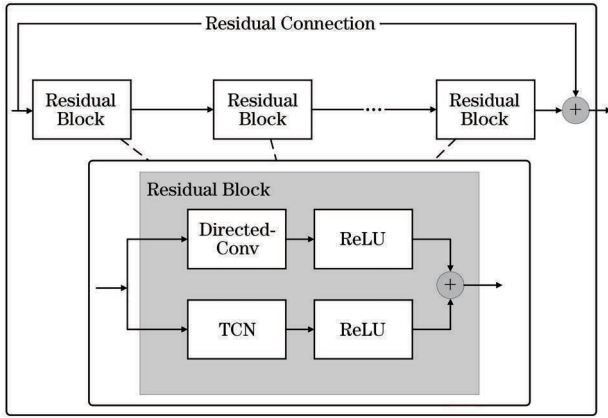


图 2 时空卷积层结构

Fig.2 Spatio-temporal convolutional layer structure

为空间卷积层,确保模型能够准确反映道路间单向或双向的交通流动特点。但是,仅考虑一阶的邻居节点不足以捕捉复杂的空间依赖性和方向性,因此有向图卷积采用多跳卷积的方式,同时考虑一阶邻居与二阶邻居的出度、入度邻接矩阵,从而捕捉节点更广泛的邻域信息。

首先构建邻接矩阵,其次执行图卷积,最后通过串联融合最终结果,如图 3 所示。

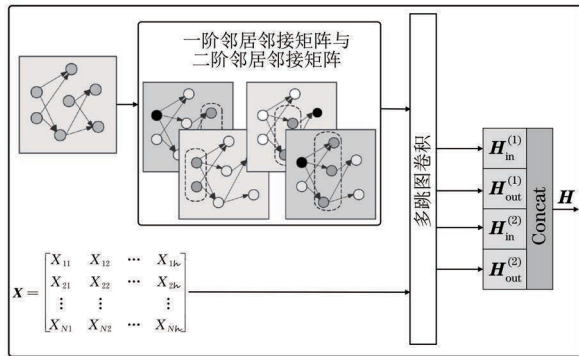


图 3 有向图卷积结构

Fig.3 Directed graph convolutional structure

一阶邻居的入度邻接矩阵定义如下:

$$\mathbf{A}_{in}^{(1)} = \mathbf{A} + \mathbf{I} \quad (6)$$

$$\mathbf{D}_{in,i} = \sum_j \mathbf{A}_{ij} \quad (7)$$

$$\text{Norm}(\mathbf{A}_{in}^{(1)}) = (\mathbf{D}_{in}^{(1)} + \mathbf{I})^{-1} \mathbf{A}_{in}^{(1)} \quad (8)$$

式中: $\mathbf{A}$ 代表原始邻接矩阵; $\mathbf{D}_{in}^{(1)}$ 代表一阶邻居的入度邻接矩阵; $\mathbf{I}$ 代表单位矩阵。

一阶邻居的出度邻接矩阵计算方式同上。

二阶邻居的入度邻接矩阵定义如下:

$$\mathbf{A}_{in}^{(2)} = \mathbf{A}_{in}^{(1)} \times \mathbf{A}_{in}^{(1)} \quad (9)$$

$$\mathbf{D}_{in}^{(2)} = \text{diag}\left(\sum_j \mathbf{A}_{in}^{(2)}[i, j]\right) \quad (10)$$

$$\text{Norm}(\mathbf{A}_{in}^{(2)}) = (\mathbf{D}_{in}^{(2)} + \mathbf{I})^{-1} \mathbf{A}_{in}^{(2)} \quad (11)$$

式中: $\mathbf{A}_{in}^{(2)}$ 代表二阶邻居入度邻接矩阵; $\mathbf{D}_{in}^{(2)}$ 代表二

阶邻居的入度邻接矩阵。

二阶邻居出度邻接矩阵计算方式同上。

4 种邻接矩阵并行输入多跳图卷积中进行计算,即:

$$\mathbf{H} = \sigma(\text{Norm}(\mathbf{A})\mathbf{X}\mathbf{W}) \quad (12)$$

式中: $\mathbf{H}$ 代表图卷积结果; $\mathbf{X}$ 代表输入的特征矩阵; $\mathbf{W}$ 代表可学习的权重。

将 4 个图卷积的结果进行拼接得到最终卷积结果,即:

$$\mathbf{H} = \text{Concat}(\mathbf{H}_{in}^{(1)} | \mathbf{H}_{out}^{(1)} | \mathbf{H}_{in}^{(2)} | \mathbf{H}_{out}^{(2)}) \quad (13)$$

式中: $\mathbf{H}_{in}^{(1)}$ 、 $\mathbf{H}_{out}^{(1)}$ 、 $\mathbf{H}_{in}^{(2)}$ 、 $\mathbf{H}_{out}^{(2)}$ 分别代表一阶、二阶邻居的入度、出度图卷积结果。

最终输出提取了空间相关性的中间状态数据。

2)时间卷积层。

时间卷积层采用时间卷积网络(TCN)^[21]处理输入的 TCI 序列,膨胀因果卷积能有效避免 LSTM 和 GRU^[22]在捕获时间序列长期依赖时可能出现的梯度消失或爆炸问题^[23],膨胀因果卷积如图 4 所示。

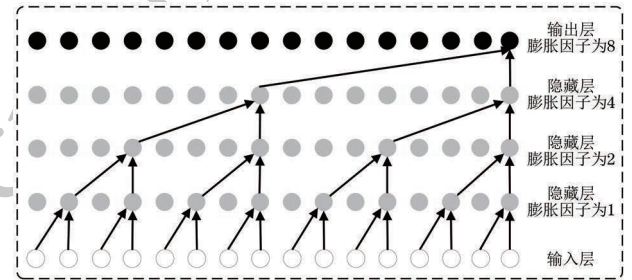


图 4 膨胀因果卷积

Fig.4 Dilated causal convolution

设 $\mathbf{x}_t$ 为时间步 t 的输入特征向量,残差块的输出向量 $\mathbf{y}_t$ 通过式(14)和式(15)计算得出:

$$\mathbf{H}(\mathbf{W}, \mathbf{x}_t) = \sum_{s=0}^{K-1} \mathbf{W}(s) \cdot \mathbf{x}_{t-d \times s} \quad (14)$$

$$\mathbf{y}_t = \mathbf{H}(\mathbf{W}', \mathbf{H}(\mathbf{W}, \mathbf{x}_t)) + \text{1DConv}(\mathbf{x}_t) \quad (15)$$

式中: $\mathbf{W}(s)$ 表示在卷积核 $\mathbf{W}$ 中索引 s 的权重; K 表示卷积核大小;膨胀因子 d 控制卷积核的膨胀。

残差块的输出为上述卷积结果与 $\mathbf{x}_t$ 通过一维卷积的相加结果, $\mathbf{W}$ 和 $\mathbf{W}'$ 表示两次卷积中的不同权重集,最终输出提取了时间相关性的中间状态数据。

2.1.2 时空注意力层

1)多头图注意力。

为考虑路网结构中各路段对指定路段的影响,并进一步捕捉图中节点之间的复杂关系,将初步提取空间相关性的数据输入到多头图注意力层中进一步处理。该机制通过对单一特征集进行分割,分别在多个注意力头中处理,使每个头专注于捕捉节点间不同方面的空间依赖性。该方法通过多角度分析

相同的特征,提高了特征提取的细致度,以及模型对复杂空间关系理解的深度和准确性。多头图注意力机制如图 5 所示。

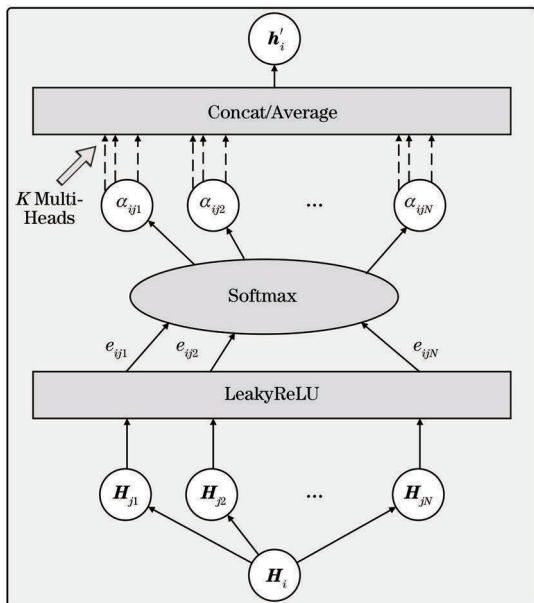


图 5 多头图注意力机制

Fig.5 Multi-heads graph attention mechanism

归一化的注意力系数计算如下:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{H}_i \parallel \mathbf{H}_j]))}{\sum_{k \in N_i} \exp(e_{ik})} \quad (16)$$

式中: $\mathbf{H} \in \mathbb{R}^{N \times D}$ 代表变换后的特征矩阵; $\mathbf{a}^T$ 代表共享的注意力机制参数; N_i 代表节点 i 的邻居节点集合。

利用归一化后的注意力系数,每个节点将其邻居的特征进行加权聚合,即:

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} \mathbf{H}_j \right) \quad (17)$$

式中: $\mathbf{h}'_i$ 是节点 i 的新特征向量。

将上述过程并行执行 K 次,每次使用不同的权重集 $\mathbf{W}^{(k)}$ 和注意力参数 $\mathbf{a}^{(k)}$,每个头计算的特征被拼接或平均处理。

对于中间层采用拼接操作:

$$\mathbf{h}'_i = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i} \alpha_{ij}^{(k)} \mathbf{H}_j^{(k)} \right) \quad (18)$$

对于输出层采用平均处理:

$$\mathbf{h}'_i = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^{(k)} \mathbf{H}_j^{(k)} \right) \quad (19)$$

最终输出进一步提取空间相关性的中间状态数据。

2) 多头自注意力。

为充分考虑同一路段历史数据中隐藏的时间模式,自注意力机制通过全局捕捉数据中的依赖关系和深刻理解整体结构,使其处理后的数据融合了跨

时间步的重要特征。多头自注意力机制通过并行化处理,使模型能够在不同的表示子空间中学习输入数据的内在相关性。每个头独立捕获数据的不同方面,从而增强模型对复杂数据结构的学习能力,该模块同样接收来自时空卷积层的输出,如图 6 所示。

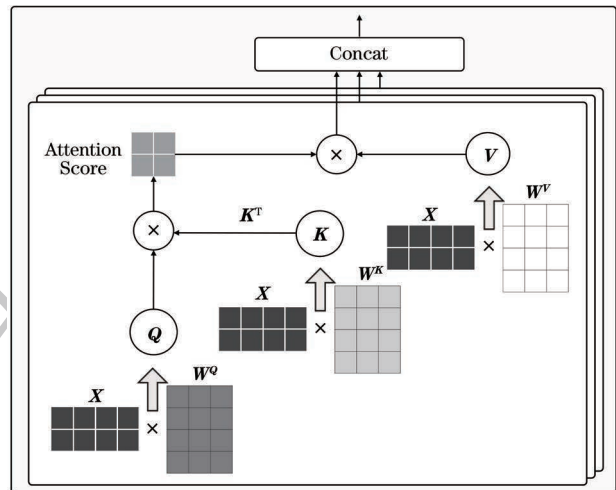


图 6 多头自注意力机制

Fig.6 Multi-heads self-attention mechanism

对于每个头,输入 $\mathbf{X}$ 首先经 3 个不同的线性变换生成查询矩阵 (Query, $\mathbf{Q}$)、键 (Key, $\mathbf{K}$) 和值 (Value, $\mathbf{V}$), 即:

$$\begin{aligned} \mathbf{Q} &= \mathbf{X} \mathbf{W}^Q \\ \mathbf{K} &= \mathbf{X} \mathbf{W}^K \\ \mathbf{V} &= \mathbf{X} \mathbf{W}^V \end{aligned} \quad (20)$$

式中: $\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V$ 是该头的参数矩阵。

使用点积注意力函数计算得分,通过 Softmax 函数计算权重分布。将注意力权重应用于值 $\mathbf{V}$,得到每个头的输出,即:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax} \left(\frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V} \quad (21)$$

式中: d_k 是键的维度。

各个注意力头的输出在最后一个维度上进行合并,融合每个头捕获的不同信息维度,即:

$$\mathbf{O}_{\text{output}} = \text{Concat}(\mathbf{o}_{\text{output},1}, \mathbf{o}_{\text{output},2}, \dots, \mathbf{o}_{\text{output},K}) \mathbf{W}^O \quad (22)$$

最终输出进一步提取时间相关性的中间状态数据。

2.2 自适应图时空卷积模块

尽管静态带权有向图解决了传统无向图的限制,但在适应交通网络动态变化上仍有局限。因此,引入自适应图时空卷积模块,自适应图通过从历史数据中学习,不断更新和调整图结构,捕捉动态变化,从而实现更准确的预测。

自适应图时空卷积模块由自适应图卷积层和门

控时间卷积层组成,输入仅为 TCI 数据,如图 7 所示。其中自适应图卷积层采用自适应图卷积网络(AGCN)处理学习得到的自适应图捕获动态空间特征,门控时间卷积层采用门控时间卷积网络(Gated-TCN)处理特征数据捕获动态时间特征。

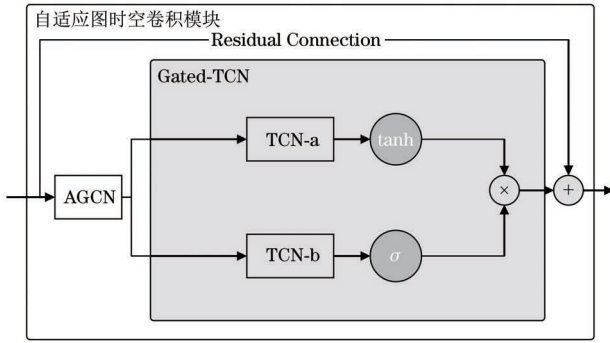


图 7 自适应图时空卷积模块结构

Fig. 7 Adaptive graph spatio-temporal convolution module structure

2.2.1 自适应图卷积层

自适应图卷积层接收 TCI 时间序列与交通图节点个数,AGCN 通过引入可学习的参数动态生成邻接矩阵 $\tilde{\mathbf{A}}$,而不依赖于静态、预定义的图连接。

具体计算过程为:首先,生成节点特征向量 Θ_1 , $\Theta_2 \in \mathbb{R}^{N \times K}$,将节点特征转换为中间表示。 Θ_1 、 Θ_2 通过反向传播和梯度下降方法进行更新,更新过程中采用局部损失函数 L_{local} ,即:

$$L_{\text{local}} = \frac{1}{N} \sum_{i=1}^N (g_i(\tilde{\mathbf{A}}, \mathbf{x}_i) - \mathbf{y}_i)^2 \quad (23)$$

式中: L_{local} 表示自适应图卷积层的局部损失函数; $g(\tilde{\mathbf{A}}, \mathbf{x}_i)$ 表示模型对给定输入 $\mathbf{x}_i$ 的预测输出,这一输出是基于自适应图结构 $\tilde{\mathbf{A}}$ 的; $\mathbf{y}_i$ 表示真实值; N 表示样本数量,代表在计算此损失函数时考虑的总数据点。

其次,计算自适应邻接矩阵 $\tilde{\mathbf{A}}$ 。最后,对 $\tilde{\mathbf{A}}$ 和特征矩阵 $\mathbf{X}$ 应用图卷积操作,即:

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{A}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (24)$$

式中: $\mathbf{H}^{(l)}$ 表示第 l 层的节点特征矩阵; $\mathbf{W}^{(l)}$ 表示第 l 层的特征转换权重矩阵; $\tilde{\mathbf{A}}$ 是根据 Θ_1 、 Θ_2 计算的自适应邻接矩阵,最终输出反映空间相关性的中间状态数据。

2.2.2 门控时间卷积层

门控时间卷积层采用 Gated-TCN 处理 AGCN 输出的空间相关性隐藏状态。Gated-TCN 的门控机制能够在图结构和交通流动频繁变化的情况下动态调整信息流,更有效地应对非线性和快速变化。

Gated-TCN 包含两个独立的 TCN,处理 AGCN 输出的空间相关性状态。TCN-a 通过 tanh

函数调制传输强度,而 TCN-b 通过 Sigmoid 函数控制信息流量。给定输入 $\mathbf{z}_t$,采用式(25)形式实现门控机制:

$$\mathbf{h}_t = \tanh(\mathbf{W}_f * \mathbf{z}_t + \mathbf{b}_f) \odot \sigma(\mathbf{W}_g * \mathbf{z}_t + \mathbf{b}_g) \quad (25)$$

式中: $\mathbf{W}_f$ 和 $\mathbf{b}_f$ 是信息变换通道的权重和偏置; $\mathbf{W}_g$ 和 $\mathbf{b}_g$ 是门控通道的权重和偏置; $\tanh(\cdot)$ 是双曲正切激活函数; σ 是 Sigmoid 激活函数。

最终输出反映时空相关性的中间状态数据。

2.3 输出模块

输出模块由交叉注意力层和全连接层整合静态与自适应图输出,优化信息流,提升预测准确性。

与自注意力机制不同,交叉注意力机制用于融合来自两个不同数据源的信息,查询 $\mathbf{Q}$ 由静态图处理模块的输出 $\mathbf{X}_1$ 生成,而键 $\mathbf{K}$ 和值 $\mathbf{V}$ 则来自自适应图处理模块的输出 $\mathbf{X}_2$,即:

$$\begin{aligned} \mathbf{Q} &= \mathbf{W}_Q \mathbf{X}_1 + \mathbf{b}_Q \\ \mathbf{K} &= \mathbf{W}_K \mathbf{X}_2 + \mathbf{b}_K \\ \mathbf{V} &= \mathbf{W}_V \mathbf{X}_2 + \mathbf{b}_V \end{aligned} \quad (26)$$

式中: $\mathbf{W}_Q$ 、 $\mathbf{W}_K$ 、 $\mathbf{W}_V$ 分别是不同数据源的权重矩阵; $\mathbf{b}_Q$ 、 $\mathbf{b}_K$ 、 $\mathbf{b}_V$ 是相应的偏置项。

交叉注意力的计算方式与式(21)一致,先对 $\mathbf{Q}$ 与 $\mathbf{K}$ 进行点积并按键向量维度的平方根进行缩放,再经 Softmax 得到注意力权重,最后用该权重对 $\mathbf{V}$ 加权求和得到输出。

最后,得到整个模型的预测结果。

3 实验

3.1 改进 TCI 的合理性验证

本节通过对比分析实际交通数据,旨在验证改进 TCI 公式在量化交通拥堵方面相比仅考虑速度的 TCI 公式具有优势。0 号节点前 2 016 个时间步的 TCI 数据如图 8 所示。

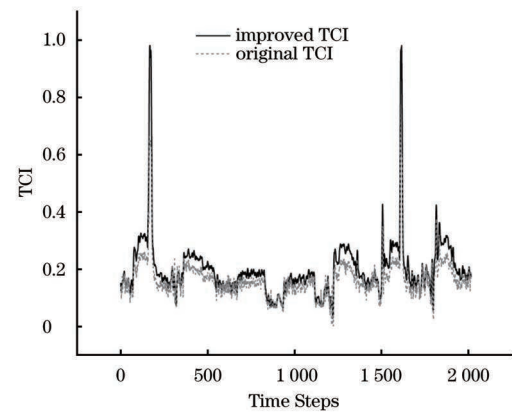


图 8 原始 TCI 与改进 TCI 计算结果对比

Fig. 8 Comparison of original TCI and improved TCI calculation results

由图 8 可见,虽然两种公式的计算结果在总体趋势上保持一致,但部分时间点的原始 TCI 计算值显著高于改进 TCI 计算值,相关 TCI 数据与实际交通情况的对比分析如表 1 所示。

表 1 TCI 计算结果与实际交通数据对比

Table 1 Comparison between TCI calculation results and actual traffic data

时间 步/个	原始 TCI	改进 TCI	平均速度/ (km·h ⁻¹)	交通流量/ (辆·h ⁻¹)	时间 占有率
307	0.182 18	0.183 45	62.4	104	0.054 7
308	0.146 79	0.172 52	65.1	67	0.033 9
309	0.112 71	0.172 45	67.7	52	0.034 5
310	0.220 18	0.166 19	59.5	10	0.015 8
311	0.213 63	0.156 63	60.0	6	0.016 0

尽管速度直观反映交通拥堵,但仅依此而忽视交通流量和时间占有率可能导致量化失真。由表 1 可见,从 307 至 311 个时间步,该路段车辆数及流密度持续下降,尤其在 310 和 311 个时间步,交通流量和时间占有率明显减少,速度虽略有下降,但未加剧拥堵。因此,改进的 TCI 公式在拥堵量化的准确性

上优于原计算方法。

3.2 实验数据准备

实验使用 PeMS08 和 PeMS04 美国加州高速公路数据集^[24],数据 5 min 记录 1 次,涵盖速度、流量和占有率,数据集另外包含节点连接性文件,直接描述了路网的空间连接情况,数据集信息如表 2 所示。数据集的缺失和异常值通过线性插值处理。预处理后的数据按 7:1:2 的比例划分为训练集、验证集和测试集。

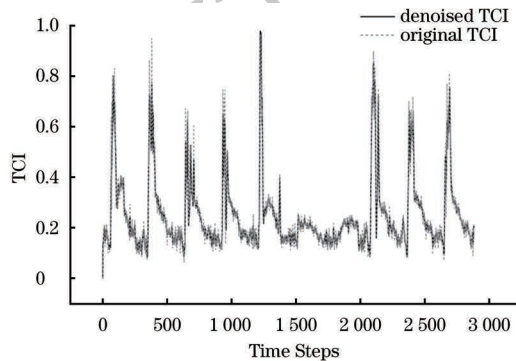
表 2 数据集统计信息

Table 2 Dataset statistical information 单位:个

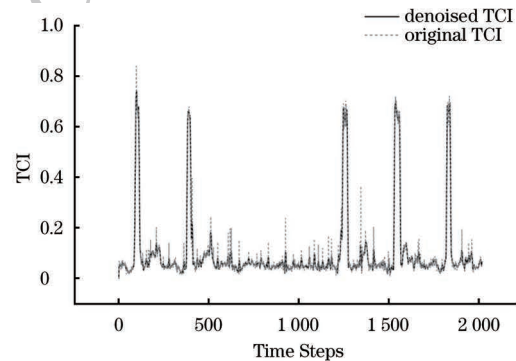
数据集	节点数	特征数	数据量
PeMS08	170	3	17 856
PeMS04	307	3	16 992

采用 SG (Savitzky-Golay) 滤波器进行数据去噪,它能在去除噪声的同时避免对重要特征造成太大影响^[25]。

由图 9(a)与图 9(b)可见,SG 滤波器去噪后的数据在减少噪声的同时保留了关键特征,提高了数据的可读性,并清晰展示了交通拥堵的周期性模式。



(a)PeMS08应用SG滤波器的效果



(b)PeMS04应用SG滤波器的效果

图 9 SG 滤波器去噪效果

Fig.9 SG filter denoising effect

3.3 评价指标

采用平均绝对误差(MAE, e_{MAE})、均方根误差(RMSE, e_{RMSE})、平均绝对百分比误差(MAPE, e_{MAPE}) 3 个评价指标,全面评估模型性能。为了让评价指标的数值更加直观并增强误差的敏感性,在 MAE 和 RMSE 公式上将预测值和目标值放大 100 倍,即:

$$e_{MAE} = \frac{1}{N} \sum_{i=1}^n |100y_i - 100y'_i| \quad (27)$$

$$e_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (100y_i - 100y'_i)^2} \quad (28)$$

$$e_{MAPE} = \frac{1}{N} \sum_{i=1}^n \frac{|y_i - y'_i|}{y_i} \times 100\% \quad (29)$$

式中: y_i 是真实值; y'_i 为预测值; N 为样本数量。

3.4 实验配置

模型的时间步长设置为 5 min,历史和预测时间段均为 6 个时间步,即 30 min。TCN 和 Gated-TCN 模块的膨胀因子依次为 2 的幂,卷积核大小固定为 2。图注意力层和自注意力层的头数设置为 8,有向图和自适应图卷积层的隐藏层维度分别设为 64 和 128。优化器使用 AdamW,训练轮数设为 500,早停阈值(patience)设为 15,批处理大小为 64。

3.5 参数分析

为评估批处理大小、注意力头数及图卷积隐藏层维度对模型性能的影响,通过实验分析这 3 个关

键参数对模型表现的具体影响。

1) 批处理大小。

批处理大小对于深度学习模型的训练效果有着重要影响,选用批处理大小为 16、32、64、128 进行实验,实验结果如图 10 所示。

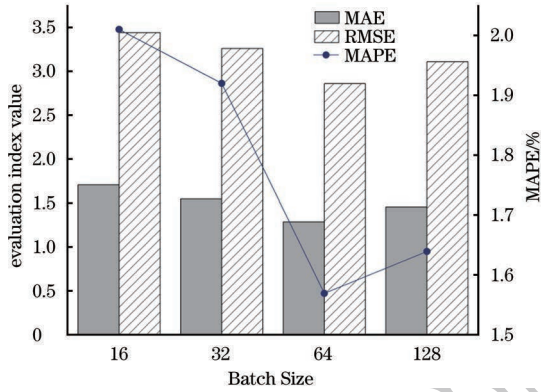


图 10 批处理大小对模型性能的影响

Fig. 10 Impact of Batch Size on model performance

由图 10 可知:模型性能在批处理大小为 64 时最优;当批处理大小为 16 和 32 时,由于梯度估计方差增大,性能较差;当批处理大小为 128 时,梯度稳定性虽提升,但限制了全局优化,导致性能下降。

2) 注意力头数。

在使用图注意力和自注意力组合时,头数的选择直接影响模型的性能。选用注意力头数组合(图

注意力/自注意力)为 4/4、6/6、8/8、10/10 进行实验,实验结果如图 11 所示。

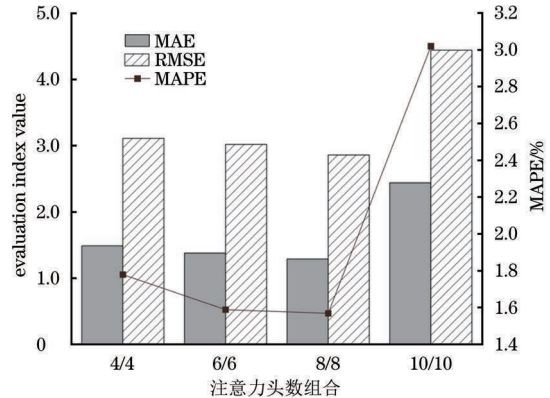


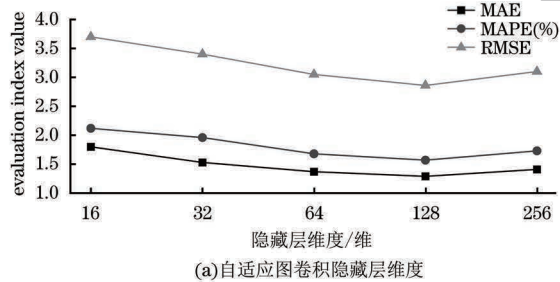
图 11 注意力头数对模型性能的影响

Fig. 11 Impact of number of attention heads on model performance

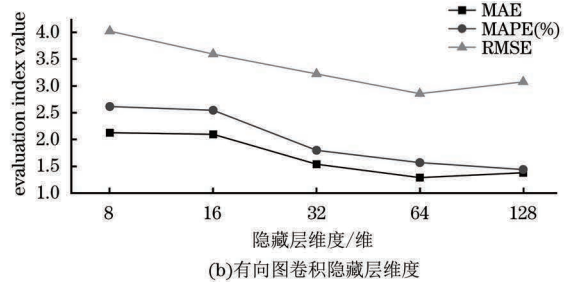
由图 11 可知:当注意力头数组合为 10/10 时,模型性能下降,因头数过多引入噪声;当头数为 4/4 和 6/6 时,尽管性能接近最优,但由于头数较少表达能力受限。

3) 图卷积隐藏层维度。

图卷积隐藏层维度的大小也对模型性能有较大影响。设置有向图卷积隐藏层维度为 8、16、32、64、128,设置自适应图卷积隐藏层维度为 16、32、64、128、256,实验结果如图 12 所示。



(a) 自适应图卷积隐藏层维度



(b) 有向图卷积隐藏层维度

图 12 图卷积隐藏层维度对模型性能的影响

Fig. 12 Impact of graph convolution hidden layer dimensions on model performance

由图 12 可知:有向图卷积模型在隐藏层维度为 64 时性能最佳,过小维度捕捉空间结构信息不足,过大则容易学习噪声;自适应图卷积模型的最佳隐藏层维度为 128,维度过小限制了动态信息的捕捉,过大则导致过度特化,降低适应性。

3.6 对比实验及结果分析

为了验证 SA-GFSTCN 模型的预测效果,选取 6 种模型进行对比,分别为历史平均 (HA)^[26]、SVR^[9]、TCN^[21]、DCRNN^[16]、STGCN^[14] 以及 Graph WaveNet^[15]。实验结果如表 3、表 4 所示。

结果表明:由于 HA 和 SVR 在处理非线性关系和特征提取方面的局限,其预测性能不如深度学

习模型;TCN 通过整合时间相关性,较传统模型显著提升性能,但忽略了空间相关性;相较而言,

表 3 各模型在 PeMS08 数据集上的表现

Table 3 Performance of various models on PeMS08 dataset

Model	MAE	MAPE/%	RMSE
HA	4.10	9.41	7.31
SVR	4.30	5.26	5.98
TCN	2.64	3.32	5.26
DCRNN	2.12	2.85	4.21
STGCN	1.89	2.44	3.97
Graph WaveNet	1.56	1.79	3.24
SA-GFSTCN	1.29	1.57	2.86

DCRNN 和 STGCN 利用复杂网络结构优化预测精度,DCRNN 在 PeMS04 数据集上通过图的拉普拉斯矩阵扩散过程精确模拟空间相关性,展现优越性能;Graph WaveNet 尽管在 PeMS08 和 PeMS04 数据集上表现良好,但忽略了重要的先验节点连接知识;SA-GFSTCN 模型在考虑静态图结构的固化问题和自适应图依赖性问题的同时有效地融合了这两者的信息优势,相较于最优基线模型 Graph WaveNet,在 MAE 指标上改进了 0.27 与 0.20,在 MAPE 指标上改进了 0.22 与 0.23 百分点,在 RMSE 指标上则改进了 0.38 与 0.36。

需要说明的是,选择与 SA-GFSTCN 性能相近的 Graph WaveNet 模型作为可视化对比对象,能更

加准确地评估模型的性能优越性。SA-GFSTCN 与 Graph WaveNet 预测对比如图 13 所示。

表 4 各模型在 PeMS04 数据集上的表现

Table 4 Performance of various models on PeMS04 dataset

Model	MAE	MAPE/%	RMSE
HA	5.07	14.03	9.36
SVR	3.77	4.62	5.56
TCN	3.12	3.47	6.11
DCRNN	2.37	2.90	4.19
STGCN	2.54	3.23	4.81
Graph WaveNet	1.73	1.95	3.47
SA-GFSTCN	1.53	1.72	3.11

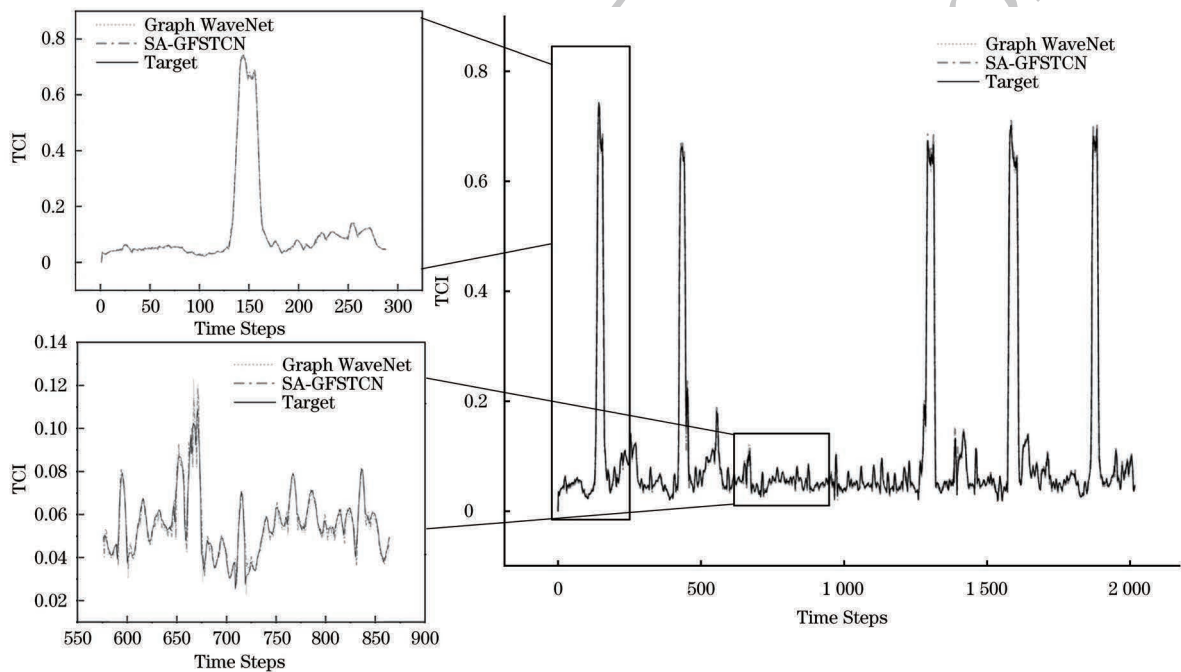


图 13 SA-GFSTCN 与 Graph WaveNet 预测对比

Fig.13 Comparison of prediction between SA-GFSTCN and Graph WaveNet

由图 13 可见,SA-GFSTCN 与 Graph WaveNet 在高值区间预测精度相近,但在低值区间,SA-GFSTCN 展现明显优势,凸显其在处理较小数值数据时的优越性能,并在整体性能上具有显著优势。

3.7 SA-GFSTCN 模型消融实验

为了进一步验证各个组件的有效性,本节采用 SA-GFSTCN 模型的 4 种变体在 PeMS08 数据集上进行一系列消融实验。

消融实验变体具体如下:

1)-STAttention: 移除静态图注意力时空卷积模块中的时空注意力模块。

2)-STConv: 移除静态图注意力时空卷积模块中的时空卷积模块。

3)-StaticGraphBlock: 移除静态图注意力时空

卷积模块。

4)-AdaptiveGraphBlock: 移除自适应图时空卷积模块。

由图 14 可见,SA-GFSTCN 模型在 MAE、RMSE、MAPE 3 项评价指标上均优于其他变体。移除静态图的时空注意力模块导致指标分别上升 0.16、0.48 和 0.05 百分点,表明时空相关性捕获减弱。移除时空卷积模块后,MAE、RMSE、MAPE 分别上升 0.34、0.68 和 0.25 百分点,影响时空依赖关系的提取。完全移除静态图注意力时空卷积模块使得模型仅依赖自适应图,导致 3 项指标分别提高 0.50、0.86 和 0.45 百分点。移除自适应图模块后,模型未能准确捕捉动态变化,3 项指标分别上涨 0.74、1.02 和 1.02 百分点。

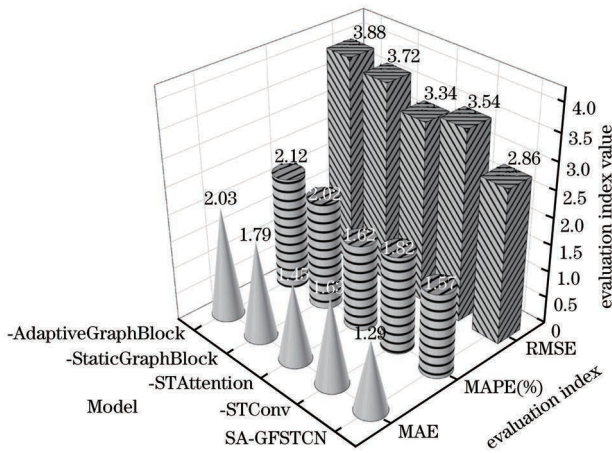


图 14 SA-GFSTCN 消融实验结果

Fig. 14 SA-GFSTCN ablation experimental results

4 结束语

本文针对 TCI 定义单一、图结构固化及静-动态信息融合难题,提出 SA-GFSTCN 模型。研究主要结论为:1)构建了融合平均速度、交通流量和时间占有率的 TCI 指标,通过平均速度直接指示拥堵程度,交通流量和占有率揭示了车辆累积与饱和状态,从而实现对拥堵的精准量化;2)采用了静态-自适应图并行处理结构,该双重机制有效维持了两种图结构的信息优势;3)设计了融合静态-自适应图输出的输出模块,通过交叉注意力机制模块能够精确整合静态与自适应处理模块的输出,最终提升了模型的预测精度。

尽管本文提出的 SA-GFSTCN 模型在预测交通拥堵方面取得了一定的精度改进,但仍存在局限性,特别是由于所使用的数据集未包含外部因素(如兴趣点、天气等),未能对这些因素进行深入分析。未来的工作将重点引入外部因素,并将聚类方法融合到交通拥堵的量化中。

参考文献

- [1] LI L C, DAI F, HUANG B, et al. AST3DRNet: attention-based spatio-temporal 3D residual neural networks for traffic congestion prediction[J]. *Sensors*, 2024, 24(4): 1261.
- [2] XIAO X P, DUAN H M. A new grey model for traffic flow mechanics [J]. *Engineering Applications of Artificial Intelligence*, 2020, 88: 103350.
- [3] SHAHRIARI S, GHASRI M, SISSON S A, et al. Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction [J]. *Transportmetrica A: Transport Science*, 2020, 16(3): 1552-1573.
- [4] LIN X F, HUANG Y Z. Short-term high-speed traffic flow prediction based on ARIMA-GARCH-M model[J]. *Wireless Personal Communications*, 2021, 117(4): 3421-3430.
- [5] ZHANG Z H, HAN Y, PENG T X, et al. A comprehensive spatio-temporal model for subway passenger flow prediction [J]. *ISPRS International Journal of Geo-Information*, 2022, 11(6): 341.
- [6] 梁艳平, 毛政元, 邹为彬, 等. 基于相似数据聚合与变 K 值 KNN 的短时交通流量预测[J]. *地球信息科学学报*, 2018, 20(10): 1403-1411.
LIANG Y P, MAO Z Y, ZOU W B, et al. Short-term traffic flow prediction based on similar data aggregation and KNN with varying K-value[J]. *Journal of Geo-Information Science*, 2018, 20(10): 1403-1411. (in Chinese)
- [7] LUO J. Short-term traffic flow prediction method in Bayesian networks based on quantile regression [J]. *Promet-Traffic&Transportation*, 2020, 32(6): 821-835.
- [8] TOAN T D, TRUONG V H. Support vector machine for short-term traffic flow prediction and improvement of its model training using nearest neighbor approach [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2021, 2675(4): 362-373.
- [9] LIN G C, LIN A J, GU D L. Using support vector regression and k-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient [J]. *Information Sciences*, 2022, 608: 517-531.
- [10] 姚俊峰, 何瑞, 史童童, 等. 基于机器学习的交通流预测方法综述[J]. *交通运输工程学报*, 2023, 23(3): 44-67.
YAO J F, HE R, SHI T T, et al. Review on machine learning-based traffic flow prediction methods[J]. *Journal of Traffic and Transportation Engineering*, 2023, 23(3): 44-67. (in Chinese)
- [11] 梁秀霞, 夏曼曼, 何月阳, 等. 基于时空多头图注意力网络的交通流预测[J]. *电子学报*, 2024, 52(2): 500-509.
LIANG X X, XIA M M, HE Y Y, et al. Traffic flow prediction based on spatio-temporal multi-head graph attention network[J]. *Acta Electronica Sinica*, 2024, 52(2): 500-509. (in Chinese)
- [12] ZHU L, KRISHNAN R, GUO F C, et al. Early identification of recurrent congestion in heterogeneous urban traffic[C]//*Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC)*. Washington D. C., USA: IEEE Press, 2019: 4392-4397.
- [13] 户佐安, 邓锦程, 韩金丽, 等. 图神经网络在交通预测中的应用综述[J]. *交通运输工程学报*, 2023, 23(5): 39-61.
HU Z A, DENG J C, HAN J L, et al. Review on application of graph neural network in traffic prediction[J]. *Journal of Traffic and Transportation Engineering*, 2023, 23(5): 39-61. (in Chinese)
- [14] YU B, YIN H T, ZHU Z X. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization, 2018: 3634-3640.
- [15] WU Z H, PAN S R, LONG G D, et al. Graph WaveNet for deep spatial-temporal graph modeling[C]//*Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao, China: International Joint Conferences on Artificial Intelligence Organization, 2019: 1907-1913.
- [16] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [EB/OL]. [2024-02-25]. <https://arxiv.org/abs/1707.01926>.
- [17] 史昕, 曹凤腾, 纪艺, 等. 基于多尺度时空特征与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.
SHI X, CAO F T, JI Y, et al. Traffic flow prediction method based on multi-scale spatio-temporal features and soft attention mechanism [J]. *Computer Engineering*, 2024, 50(12): 346-357. (in Chinese)
- [18] ZHANG C, WU Y, SHEN Y, et al. Adaptive graph convolutional recurrent network with transformer and whale

- optimization algorithm for traffic flow prediction [J]. Mathematics, 2024, 12(10): 1493.
- [19] 龚循强, 邱万锦, 吕开云, 等. 变分模态分解与自适应图卷积门控循环网络的交通流量组合预测模型[J]. 武汉大学学报(信息科学版), 2024, 49(12): 2329-2341.
- GONG X Q, QIU W J, LÜ K Y, et al. A combined traffic flow prediction model based on variational mode decomposition and adaptive graph convolutional gated recurrent network[J]. Geomatics and Information Science of Wuhan University, 2024, 49(12): 2329-2341. (in Chinese)
- [20] WANG X, ZENG R H, ZOU F M, et al. STTF: an efficient transformer model for traffic congestion prediction [J]. International Journal of Computational Intelligence Systems, 2023, 16(1): 2.
- [21] ZHANG Y L, SHANG K, CUI Z W, et al. Research on traffic flow prediction at intersections based on DT-TCN-attention[J]. Sensors, 2023, 23(15): 6683.
- [22] ZHANG Y Z, LIU S X, ZHANG P, et al. GRU- and Transformer-based periodicity fusion network for traffic forecasting[J]. Electronics, 2023, 12(24): 4988.
- [23] 胡杰, 龚永胜, 蔡世杰, 等. 基于多源数据的车流量时空预测方法[J]. 汽车工程, 2021, 43(11): 1662-1672.
- HU J, GONG Y S, CAI S J, et al. A spatio-temporal prediction method of traffic flow based on multi-source data [J]. Automotive Engineering, 2021, 43(11): 1662-1672. (in Chinese)
- [24] Caltrans performance measurement system[EB/OL]. [2024-02-25]. <http://pems.dot.ca.gov/>.
- [25] WANG Q R, CHEN X H. Short-term traffic flow prediction based on the SGA-KGCN-LSTM model [EB/OL]. [2024-02-25]. https://www.engineeringletters.com/issues_v31/issue_3/EL_31_3_36.pdf.
- [26] SMITH B L, WILLIAMS B M, OSWALD R K. Comparison of parametric and nonparametric models for traffic flow forecasting[J]. Transportation Research Part C: Emerging Technologies, 2002, 10(4): 303-321.

文字编辑 陆燕菲
栏目编辑 宋 圆