

基于低秩分解的时空图神经网络交通流量预测方法

乔毅晨*, 程泽生#

青岛大学计算机科学技术, 山东 青岛

收稿日期: 2025年5月6日; 录用日期: 2025年7月10日; 发布日期: 2025年7月22日

摘要

针对现有图神经网络(Graph Neural Network, GNN)模型在交通流量预测中存在的结构冗余与计算复杂度较高问题, 提出一种基于谱图理论与图信号处理的GNN优化方法。通过将时空图卷积网络模型与 k 阶奇异值分解相结合, 能够有效抑制预测过程中的高频噪声干扰, 在保持模型结构简洁性的同时显著提升预测精度。基于3个公开交通流数据集的实验结果表明, 所提模型在短时预测精度和长时预测稳定性上均达到或超越了当下先进模型的性能水平。

关键词

图卷积网络, 低通滤波器, 去噪, 交通流量预测

Low-Rank Decomposition Method for Spatiotemporal Graph Neural Network-Based Traffic Flow Prediction

Yichen Qiao*, Zesheng Cheng#

Computer Science and Technology, Qingdao University, Qingdao Shandong

Received: May 6th, 2025; accepted: Jul. 10th, 2025; published: Jul. 22nd, 2025

Abstract

To address the structural redundancy and high computational complexity issues inherent in existing

*第一作者。

#通讯作者。

文章引用: 乔毅晨, 程泽生. 基于低秩分解的时空图神经网络交通流量预测方法[J]. 交通技术, 2025, 14(4): 446-458.
DOI: 10.12677/ojtt.2025.144045

Graph Neural Network (GNN) models for traffic flow prediction, this study proposes an optimized GNN method based on spectral graph theory and graph signal processing. By integrating a Spatio-temporal Graph Convolutional Network (STGCN) model with k -order Singular Value Decomposition (k -SVD), the proposed approach effectively reduces high-frequency noise interference during predictions while maintaining structural simplicity and significantly enhancing prediction accuracy. Experimental evaluations on three publicly available traffic flow datasets demonstrate that the proposed model achieves or surpasses the performance of state-of-the-art models in both short-term prediction accuracy and long-term prediction stability.

Keywords

Graph Convolutional Network, Low-Pass Filter, Denoising, Traffic Flow Prediction

Copyright © 2025 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. 引言

全球机动车保有量激增导致城市交通超负荷运行, 传统交通管理系统的不足日益凸显。智能交通系统(Intelligent Transportation Systems, ITS)作为破解“拥堵-污染-能耗”恶性循环的关键技术, 其核心在于构建高效的交通流量预测体系[1]-[3]。精确的流量预测可提升路网通行效率, 并为交通规划者在道路通行能力设计、路面结构优化以及交叉口控制策略制定等方面提供科学依据[4]-[6]。现实场景中, 相关数据多通过政府及有关部门设置的传感器或监控摄像等设备获取。但数据采集面临多维约束: 固定式传感器存在空间覆盖盲区, 移动检测设备受恶劣天气影响显著, 导致原始数据存在时变噪声与结构化缺失[7]。为此, 研究者们借助机器学习和大数据分析技术, 结合成本较低的数据, 模拟并预测成本较高的交通流数据[8]-[10]。此方法已成为广泛认可的解决方案之一, 并持续得到优化和改进。交通流数据是普遍被视为难以在欧式空间中直接表示的非结构化数据类型[11]-[13]。GNN 凭借谱域分析框架与节点消息传递机制[14]-[16], 能够有效捕获路网节点间的高阶依赖关系。因此, GNN 已成为交通流量预测的核心技术范式。早期代表性工作如时空图卷积网络(STGCN [17]), 通过融合图卷积网络(Graph Convolutional Network, GCN [18])与时间卷积网络(Temporal Convolutional Network, T-CN [19]), 实现了对交通流数据时空特征的解耦提取, 其基于切比雪夫多项式的单跳邻域聚合机制兼具效率与精度优势。后续研究围绕提升时空建模能力展开: 时序图卷积网络(Temporal Graph Convolutional Network, T-GCN [20])引入门控循环单元(Gated Recurrent Unit, GRU [21])增强时序动态表征; 基于注意力机制的时空图卷积网络(Attention-based Spatiotemporal Graph Convolutional Network, ASTGCN [22])通过时空权重分配优化局部特征提取; 扩散卷积递归神经网络(Diffusion Convolutional Recurrent Neural Network, DCRNN [23])则借助无限阶扩散过程捕捉多跳传播效应。在实际的交通流量预测任务里, 多跳邻域模型的性能要优于单跳模型。其核心优势在于, 多跳邻域的消息传递机制能够更为精准地捕捉节点间的高阶依赖关系, 这对复杂交通流特征的建模起着关键作用。然而, 现有主流模型通过堆叠深度学习模块捕获多跳邻域消息, 虽取得一定进展, 却导致模型复杂度显著增加。如何平衡模型效率与特征提取能力, 仍是亟待解决的关键问题。针对上述问题, 本研究提出一种基于 k 阶奇异值分解(k -order Singular Value Decomposition, k -SVD)的谱域低通滤波方法, 将 k -SVD 的低频特征保留策略嵌入 STGCN 框架, 在无需依赖复杂动态图结构优化的前提下, 实现对高频噪声的抑制与核心时空特征的定向强化。

2. 图谱理论与图信号处理

2.1. 谱图理论基础

设无权图 $G = \{V, E\}$, 其中 V 代表节点集合, 而 E 代表节点之间边的集合。在交通流量预测任务中, 交通网络常被抽象为无向且无权的简单图。邻接矩阵 A , 其中 A_{ij} 表示上游节点 i 到下游节点 j 之间边的权重值(在无权图中为 1)。度矩阵 D 是一个对角阵, 其中 D_{ii} 表示节点 i 的度数, 即与其它相连边的权重值之和。拉普拉斯矩阵 L 定义为 $L = D - A$, 描述了相邻节点间的差异, 进而用于衡量图信号的平滑性。

谱图理论通过图的拉普拉斯矩阵特征分解实现图信号的谱域转换。通过特征值分解 $L = U\Lambda U^T$, 可将图信号 $x \in \mathbf{R}^N$ 转换至谱域: $\hat{x} = U^T x$ 。其中 U^T 是特征向量矩阵 U 的转置, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 特征值对角矩阵, 满足 $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ 。分量 $\hat{x}_i = u^T x_i$ 表示信号在频率 λ_i 上的投影强度。这一过程的实现有效地规避了图空间域中非结构化数据难以处理的问题。同时也可以通过逆变换 $x = U\hat{x}$ 则可将信号从谱域转换回空间域。原始 GCN 模型的核心公式为:

$$H^{n+1} = \sigma(Ug(\Lambda)U^{-1}H^n\omega) \quad (1)$$

其中, H^n 表示第 n 层的输入特征矩阵, U 表示拉普拉斯矩阵的特征向量矩阵, Λ 表示拉普拉斯矩阵的特征值对角矩阵, $\sigma(\cdot)$ 表示非线性激活函数。GCN 模型通过谱域变换函数 $g(\Lambda)$ 与卷积层权重矩阵 ω 学习对信号进行谱域处理。GCN 通过将信号从空间域变换至谱域, 在该域中进行处理后再转换回空间域, 这一过程使得 GCN 能够提取每一跳邻域节点间传递消息的共性特征。从数学角度, 拉普拉斯算子的二次型为:

$$\begin{aligned} x^T L x &= \begin{bmatrix} x_1^T & x_2^T & \dots & x_n^T \end{bmatrix} \begin{bmatrix} d_1 & -a_{12} & \dots & -a_{1n} \\ -a_{21} & d_2 & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1^T & x_2^T & \dots & x_n^T \end{bmatrix} \begin{bmatrix} d_1 x_1 - \sum_{1 \leq j \leq n, j \neq 1} a_{1j} x_j \\ d_2 x_2 - \sum_{1 \leq j \leq n, j \neq 2} a_{2j} x_j \\ \vdots \\ d_n x_n - \sum_{1 \leq j \leq n, j \neq n} a_{nj} x_j \end{bmatrix} \\ &= \sum_{1 \leq i \leq n} d_i x_i^T x_i - \sum_{i \leq i, j \leq n, i \neq j} a_{ij} x_i^T x_j = \sum_{i \leq i, j \leq n, i \neq j} a_{ij} x_i^T (x_i - x_j) \\ &= \sum_{1 \leq i, j \leq n, i \leq j} [a_{ij} x_i^T (x_i - x_j) + a_{ji} x_j^T (x_j - x_i)] = \sum_{(i,j) \in E} A_{ij} (x_i - x_j)^T (x_j - x_i) \end{aligned} \quad (2)$$

设 x 为拉普拉斯矩阵 L 的特征向量, λ 为对应的特征值, 则

$$x^T L x = x^T \lambda x = \lambda x^T x = \sum_{(i,j) \in E} A_{ij} (x_i - x_j)^T (x_j - x_i) \quad (3)$$

式(2)和(3)衡量了图信号的平滑度。由于二次型的非负性, 当 $\lambda_i \rightarrow 0$ 时, $x^T L x \rightarrow 0$, 即信号 x 与其相邻节点特征差异趋于零(低频信号)。交通流数据中, 区域路网的流量呈现空间连续性, 相邻路段流量相近, 即低频信号。相反, 高频信号对应相邻节点特征的剧烈波动, 通常由测量噪声(如传感器误差)或瞬时扰动(如交通事故)引起。

然而拉普拉斯算子特征值分解的时间复杂度高达 $O(n^3)$, 不适用于规模较大的图。因此, 研究者通过切比雪夫多项式近似的方法代替拉普拉斯算子特征值分解

$$H^{n+1} = \sigma \left((D+I)^{-\frac{1}{2}} (A+I) (D+I)^{-\frac{1}{2}} H^n \omega \right) \quad (4)$$

这使谱域内的训练仅需关注与每个卷积层权重矩阵 ω 的学习, 并直接利用包含自环的邻接矩阵传递

信息。从谱域的角度表示式(4)

$$(\mathbf{D} + \mathbf{I})^{-\frac{1}{2}} (\mathbf{A} + \mathbf{I}) (\mathbf{D} + \mathbf{I})^{-\frac{1}{2}} = \mathbf{Z} (\mathbf{I} - \tilde{\mathbf{\Lambda}}) \mathbf{Z}^T \left(\hat{\lambda}_i \in [0, 2) \text{ for } \hat{\lambda}_i \in \tilde{\mathbf{\Lambda}} \right) \quad (5)$$

其中, $\mathbf{Z}$ 表示归一化拉普拉斯矩阵 $\tilde{\mathbf{L}} = \mathbf{I} - (\mathbf{D} + \mathbf{I})^{-\frac{1}{2}} (\mathbf{A} + \mathbf{I}) (\mathbf{D} + \mathbf{I})^{-\frac{1}{2}}$ 的特征向量矩阵, $\tilde{\mathbf{\Lambda}}$ 表示 $\tilde{\mathbf{L}}$ 的特征值矩阵。特征值 $\hat{\lambda}_i \in [0, 2)$, 此时 GCN 在谱域的传播规则的本质是对谱域信号施加线性低通滤波 $g(\lambda_i) = 1 - \lambda_i$ 。同时 p 层 GCN 堆叠后的总频响函数为: $g_{\text{total}}(\lambda_i) = (1 - \lambda_i)^p$, 表示每层聚合一跳邻域信息, p 层堆叠等价于 p 跳邻域覆盖。当高频分量经过 p 次指数衰减 $(1 - \lambda_i)^p$, 例如 $\lambda_i = 1.5$ 时, 5 层 GCN 可使该频率分量衰减至 $0.5^5 \approx 0.03$, 即原始值的 6%。

2.2. k 阶奇异值分解(k -SVD)

奇异值分解是一种适用于任意大小矩阵的矩阵分解方法。设任意矩阵 $\mathbf{M}$ 解为左奇异矩阵 $\mathbf{P}$ 、奇异值矩阵 Σ 和右奇异矩阵 $\mathbf{Q}$

$$\mathbf{M}^T \mathbf{M} = (\mathbf{P} \Sigma \mathbf{Q}^T)^T \mathbf{P} \Sigma \mathbf{Q}^T = \mathbf{Q} \Sigma^T \Sigma \mathbf{Q}^T \quad (6)$$

$$\mathbf{M} \mathbf{M}^T = \mathbf{P} \Sigma \mathbf{Q}^T (\mathbf{P} \Sigma \mathbf{Q}^T)^T = \mathbf{P} \Sigma \Sigma^T \mathbf{P}^T \quad (7)$$

其中, 左奇异矩阵和右奇异矩阵分别对应于原始矩阵的行空间和列空间。将带自环的邻接矩阵定义为 $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, 其作用于信号 $\mathbf{x}$ 后得到

$$\tilde{\mathbf{A}}\mathbf{x} = \begin{bmatrix} x_1 + \sum_{j=2}^n a_{1j}x_j & x_2 + \sum_{j=1, j \neq 2}^n a_{2j}x_j & \cdots & x_n + \sum_{j=1}^{n-1} a_{nj}x_j \end{bmatrix}^T \quad (8)$$

若 $\mathbf{x}$ 是 $\tilde{\mathbf{A}}$ 的奇异向量, 对应的奇异值 σ 为

$$\tilde{\mathbf{A}}\mathbf{x} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 + \sum_{i=1}^n a_{1i}x_i \\ x_2 + \sum_{i=1}^n a_{2i}x_i \\ \vdots \\ x_n + \sum_{i=1}^n a_{ni}x_i \end{bmatrix} \Rightarrow (\tilde{\mathbf{A}}\mathbf{x})^T (\tilde{\mathbf{A}}\mathbf{x}) = \sum_{i=1}^n \left(x_i + \sum_{j=1}^n a_{ij}x_j \right)^2 \quad (9)$$

$$\left[(\tilde{\mathbf{A}}\mathbf{x})^T (\tilde{\mathbf{A}}\mathbf{x}) \right] \mathbf{x} = \sigma^2 \mathbf{x} \Rightarrow \sigma = \sqrt{1 + 2 \sum_{i,j=1}^n a_{ij}x_i x_j + \left(\sum_{i,j=1}^n a_{ij}x_j \right)^2} \quad (10)$$

假设 $\mathbf{x}$ 为低频信号, 相邻节点特征相似, 即满足 $x_i \approx x_j$, 此时

$$\sum_{i,j} a_{ij}x_i x_j \approx \sum_{i,j} a_{ij}x_i^2 = x_i^2 \sum_{i,j} a_{ij} = x_i^2 d_i \quad (11)$$

其中, d_i 为节点 i 的度数, 因此有

$$\sigma \approx \sqrt{1 + 2x_i^2 d_i + (x_i d_i)^2} = \sqrt{(1 + x_i d_i)^2} = 1 + x_i d_i \quad (12)$$

若 $\mathbf{x}$ 为高频信号, 相邻节点特征差异较大, 例如 $x_i = -x_j$, 此时

$$\sum_{i,j} a_{ij}x_i x_j = -\sum_{i,j} a_{ij}x_i^2 = -x_i^2 d_i \quad (13)$$

$$\sigma \approx \sqrt{1 - 2x_i^2 d_i + (x_i d_i)^2} = \sqrt{(1 - x_i d_i)^2} = |1 - x_i d_i| \quad (14)$$

观察式(12)和(14), 当 $x_i d_i > 1$ 或 $x_i d_i < 1$, 高频信号对应的 σ 都远小于低频信号。 k -SVD 通过保留前 k 个最大的奇异值(对应低频成分), 等效于将高频成分的权重置为 0, 从而实现理想低通滤波的效果。

在交通流量预测中, STGCN 模型的输出矩阵 $\mathbf{S}$ 兼具时间序列与图结构的双重特性。分从空间与时间

维度分析：节点间的空间相关性表现为矩阵 $\mathbf{S}$ 的列向量具有低秩性，即低频成分占据主导地位；而交通流量变化的周期性和连续性表现为矩阵 $\mathbf{S}$ 的行向量具有低秩性，即时序特征呈现平滑性。 k -SVD 通过保留前 k 个最大奇异值及其对应的奇异向量，对原矩阵进行低秩近似：

$$\mathbf{S}_k = \mathbf{P}_k \sum_k \mathbf{Q}_k^T \approx \mathbf{S} \quad (15)$$

上述过程在时空域同步实现去噪：在空间域，通过保留左奇异矩阵 $\mathbf{P}_k$ (对应空间低频分量)，有效抑制空间高频噪声(如孤立节点的异常波动、传感器误差导致相邻节点流量突变)；在时间域，通过保留右奇异矩阵 $\mathbf{Q}_k$ (对应时间主成分)，抑制时序预测中的误差积累(如突发事件引发流量异常波动、长期预测中的噪声扩散效应)。特别地，在长时预测场景中，噪声会随着时间步长逐渐积累，而 k -SVD 的低频截断特性会强制预测结果保持时空的平滑性，避免因高频成分干扰导致的预测发散问题，从而提升模型在长期预测中的稳定性。

3. 方法描述

3.1. 线性低通滤波器到理想低通滤波器

为了避免高维特征分解的计算代价，STGCN 采用切比雪夫多项式进行局部滤波近似，一阶切比雪夫多项式对应于一个线性低通滤波器。如图 1 所示，展示了在不同 p 值下对应的滤波函数图像。对于 p 阶多项式堆叠滤波函数表示为 $(1 - \tilde{\lambda})^p$ ，这种非线性衰减虽然增强了低频保留能力，但仍属于连续型滤波器。当消息传递次数趋近无穷时，系统将仅保留零频分量(全局均值)，导致过度平滑。因此，合理选择截断阶数 p 成为平衡特征提取与信息保真的关键。

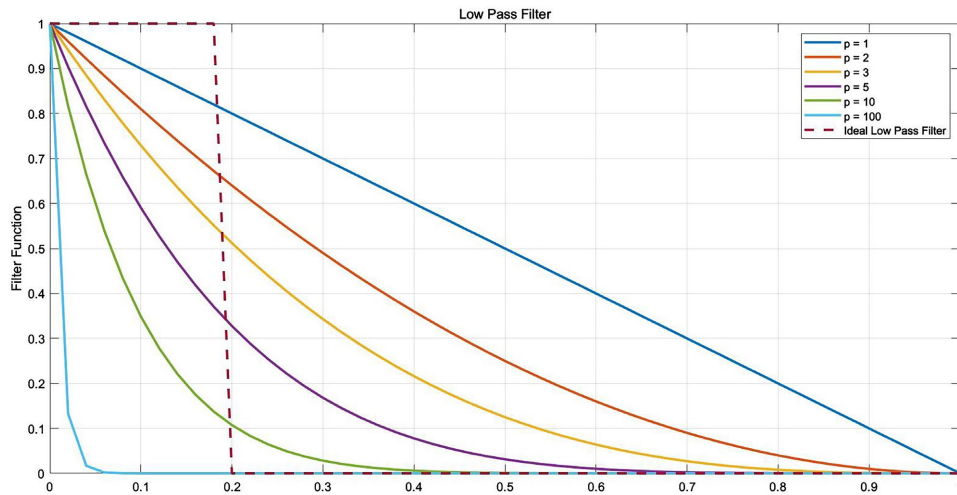


Figure 1. Filtering functions with different p values and the ideal low-pass filter
图 1. 具有不同 p 值的滤波函数与理想低通滤波器

在图信号处理中， k -SVD 过程等价于在谱域中仅保留 k 个最低频率分量，实现理想的频率截断效果。本研究设计了并行模块(Model-P)，通过并联两个子模块实现多级滤波：子模块 1 采用一阶切比雪夫多项式(线性低通滤波器)，子模块 2 采用 k -SVD (理想低通滤波器)，结构如图 2 所示。

通过可训练权重 α 和 β 动态融合子模块的输出，最终形成完整的图卷积操作：

$$\mathbf{y} = \alpha \cdot \text{ChebConv}(\mathbf{x}) + \beta \cdot k\text{-SVD}(\mathbf{x}) \quad (16)$$

其中， $\text{ChebConv}(\mathbf{x})$ 为线性低通滤波器函数， $k\text{-SVD}(\mathbf{x})$ 为理想低通滤波器函数。这一设计既保留了传

统图卷积的局部特征提取能力, 又通过 k -SVD 增强了对全局低频信息的捕捉, 从而在复杂交通场景中实现更精准的时空特征建模。

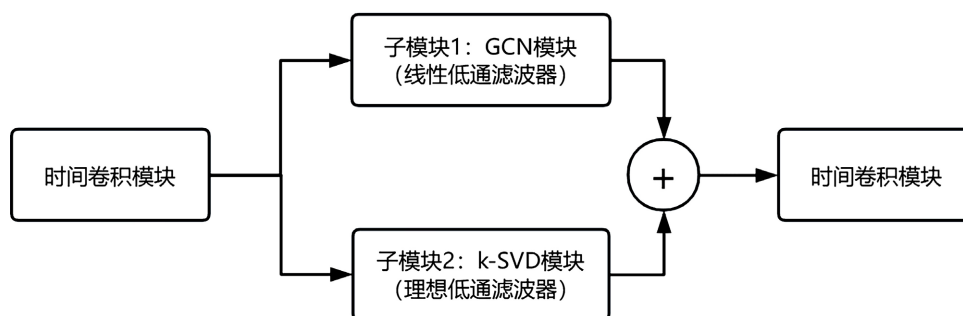


Figure 2. Parallel model Model-P
图 2. 并行模型 Model-P

3.2. 串行模型 Model-C

尽管 Model-P 并行模块在理论框架上具有创新性, 但存在两方面缺陷: 一是引入额外可训练参数导致模型复杂度上升; 二是需依赖复杂的后期处理才能实现理想预测精度。为此, 本研究提出串行模块 Model-C, 具体结构如图 3 所示。

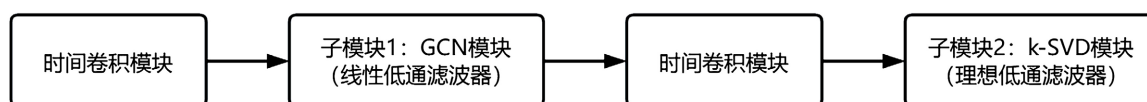


Figure 3. Serial model Model-C
图 3. 串行模型 Model-C

模型仅在训练周期后对输出矩阵执行 k -SVD, 此改进显著提升了性能, 达到与最新基准模型相当的水平。具体原因主要体现在以下三个方面:

首先, 权衡模型复杂性和计算效率。由于 Model-P 引入了额外的可训练参数(可训练权重 α 和 β), 需要在训练过程中反复执行 k -SVD 矩阵分解, 导致训练时间显著增加(时间复杂度为 $O(kn^2)$), 同时加剧了过拟合风险。此外, Model-P 的融合机制缺乏有效的约束, 导致 GCN 模块与 k -SVD 模块的梯度更新方向会产生冲突。相较之下, Model-C 通过保持原始 STGCN 的简洁架构, 避免引入额外可训练参数, 并且仅在训练周期结束时应用一次 k -SVD 操作, 有效得避免了梯度冲突, 额外计算开销可控(尤其是 k 远小于 n 时)。

其次, 从对原始模型的干扰程度分析。Model-P 的并行结构在训练中面临严格低频约束问题: k -SVD 的理想低通特性对特征空间施加严格低频限制, 可能滤除具有实际意义的中等频率成分(如早晚高峰过渡期的流量波动、区域性拥堵的传播特征)及部分高频有效信号, 导致特征多样性下降。这种过度平滑效应在长时预测中尤为明显, 会导致误差累积效应随时间步长放大, 限制了模型对复杂时空特征的学习能力。反观 Model-C, 其 k -SVD 作为独立后处理步骤, 允许 STGCN 自由学习时空特征, 仅对最终输出矩阵进行低秩优化以抑制噪声。实验结果表明 Model-P 在长时预测(如 60 分钟)任务中性能下降明显(如 BEIJING 数据集上 RMSE = 14.941 对比 Model-C 的 13.486)。

最后, 关于理论方面, 考虑到 STGCN 架构由两大核心组件构成: 两个 TCN 模块以及一个 GCN 模块。其输出矩阵的每一行代表图上所有节点在某一特定时间的流量数据, 每一列则反映了单个节点在所有时间段内的流量变化, 奇异值分解得到的左奇异矩阵和右奇异矩阵分别对应了 STGCN 输出矩阵的空

间域和时域。对于空间信息处理, 可将 STGCN 模型的输出矩阵 $\mathbf{S}$ 近似为 $\tilde{\mathbf{A}}\mathbf{x}$, 其中 $\tilde{\mathbf{A}}\mathbf{x}$ 体现了目标节点与其一跳相邻节点间的信息传递。式(10)表明, 空间域中 k -SVD($\tilde{\mathbf{A}}\mathbf{x}$) 可视为图信号的一种理想低通滤波处理。同时, 在时间信息处理上与空间信息类似, k -SVD 方法能够有效地保留时间信息的主要成分, 去除高频时序噪声。特别是在时间序列预测任务中, 噪声(或误差)随时间步长积累, 通过保留主成分也会抑制误差传播。由于时间模块并非本研究的核心创新点, 其具体实现细节将不在此详述。

4. 实验与分析

4.1. 数据集

实验使用 3 个数据集(表 1), 涵盖不同规模交通网络, 数据按照 7:1.5:1.5 的比例划分为训练集、验证集和测试集。实验中, 模型从过去 3、6 和 12 个时间片中提取特征信息, 进而预测未来的交通流量。为确保评估结果的可靠性, 所有模型在同一数据集上进行了 10 次重复的训练与测试。实验用服务器为 Intel Xeon W-2133@3.6Hz CPU 和 32 RAM 的 NVIDIA V100 GPU。

Table 1. Dataset

表 1. 数据集

数据集	节点数	时间片数	说明
BEIJING [24]	3126	21,576	Beijing, China
PEMSD7M [17]	228	12,672	District 7, the state of California
METR-LA [23]	207	34,272	Los Angeles Metropolitan

Model-P 作为 Model-C 的消融实验变体, 仅针对空间特征提取部分引入 k -SVD 并行模块, 其中 k 值取 3。STGCN 作为较早将 GCN 应用于交通流量预测任务的模型, 随着相关技术的不断更新, STGCN 模型逐渐失去了竞争力。因此, 若 k -SVD 这一简单流程能够取得优于结构更为复杂的模型的性能表现, 则更能充分彰显本研究提出方法的有效性。

4.2. 评价指标

使用 3 个评价指标来评价模型的性能, 均方根: $\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^n (h(x_i) - y_i)^2}$, 平均绝对误差: $\text{MAE} = \frac{1}{m} \sum_{i=1}^n |h(x_i) - y_i|$, 平均绝对百分比误差: $\text{MAPE} = \frac{100}{m} \sum_{i=1}^n \left| \frac{h(x_i) - y_i}{y_i} \right|$ 。

4.3. 实验结果

6 个基准模型和提出模型(Model-P 和 Model-C)的实验结果见表 2。Model-C 在 3 个数据集(各含 3 个时间片)的 9 项实验中, 7 项性能最优, 2 项为列前三, 仅在 METR-LA 数据集 15 分钟的预测结果排第四, 其中的一些结果甚至比基线模型高出 10%以上。

Table 2. Experimental results

表 2. 实验结果

数据集	模型	15 min			30 min			60 min		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
BEIJING	STGCN	10.257	5.133	9.61%	12.89	6.264	12.39%	16.533	8.042	16.10%

续表

BEIJING	ASTGCN	9.704	4.631	10.13%	12.653	6.069	12.38%	15.275	6.796	14.80%
	GMAN	9.597	4.855	14.04%	12.01	5.985	14.71%	13.77	6.338	16.60%
	DCRNN	9.02	4.389	9.69%	12.163	5.635	13.12%	14.918	6.558	15.90%
	MTGNN	8.624	<u>4.079</u>	9.03%	11.525	5.057	<u>11.77%</u>	13.756	5.954	14.30%
	DGCRN	<u>8.577</u>	4.309	<u>8.69%</u>	<u>11.181</u>	<u>4.911</u>	11.09%	<u>13.487</u>	5.79	<u>13.88%</u>
	Model-P	9.764	5.391	11.57%	11.429	6.207	13.71%	14.941	7.745	16.01%
	Model-C	8.273	3.97	8.28%	10.87	4.753	11.79%	13.486	<u>5.805</u>	13.72%
	STGCN	4.264	2.4	5.35%	5.599	3.143	7.05%	6.788	3.782	8.93%
PEMSD7M	ASTGCN	4.369	2.521	5.86%	5.513	2.989	7.70%	6.592	3.603	9.53%
	GMAN	5.736	2.909	7.34%	6.185	3.144	7.87%	7.993	3.993	10.03%
	DCRNN	4.433	2.343	5.57%	5.736	2.96	7.46%	7.223	3.679	9.85%
	MTGNN	<u>4.031</u>	<u>2.205</u>	5.04%	5.42	<u>2.735</u>	6.80%	<u>6.549</u>	3.124	<u>8.32%</u>
	DGCRN	4.069	2.218	5.20%	5.446	2.768	<u>6.78%</u>	6.734	3.336	8.57%
	Model-P	4.170	2.41	5.23%	<u>5.411</u>	2.914	6.96%	6.667	3.578	8.67%
	Model-C	3.984	2.132	<u>5.09%</u>	4.973	2.732	6.45%	6.321	<u>3.36</u>	8.29%
	STGCN	6.279	3.542	9.06%	8.155	4.12	10.70%	9.504	4.622	12.53%
METR-LA	ASTGCN	6.271	3.619	8.72%	7.555	4.178	10.33%	9.991	4.758	13.12%
	GMAN	8.805	5.591	10.03%	9.273	5.969	10.96%	10.245	7.165	12.78%
	DCRNN	6.292	3.54	8.70%	7.504	4.111	10.37%	9.661	4.757	12.31%
	MTGNN	<u>6.245</u>	3.239	9.04%	7.448	3.681	10.58%	9.052	4.221	12.21%
	DGCRN	6.045	3.141	7.95%	<u>7.292</u>	3.482	<u>9.69%</u>	<u>8.691</u>	4.133	<u>12.04%</u>
	Model-P	6.256	3.5	8.93%	8.009	4.06	10.51%	9.5	4.593	12.51%
	Model-C	6.273	<u>3.222</u>	<u>8.50%</u>	7.261	<u>3.478</u>	9.66%	8.666	<u>4.143</u>	11.80%

如图 4 所示, Model-C 与 STGCN 在 METRO-LA 数据集(60 分钟预测)的 250 周期训练损失对比, 观察结果可以得出 Model-C 在训练稳定性方面优于 STGCN。通过实验结果(表 2) Model-C 相较 STGCN 稳定提升约 5%, 结合图 4 的损失函数曲线可知, 两者共同验证了 Model-C 和 k -SVD 去噪方法的有效性, 同时证实了多跳邻域消息传递机制的重要性, 以及 k -SVD 作为低通滤波器的设计合理性。

实验结果表明, Model-P 的各项指标普遍劣于 Model-C。具体而言, 在 BEIJING 数据集 60 分钟预测中, Model-P 的 RMSE 为 14.941, 显著高于 Model-C 的 13.486。又如在 METR-LA 数据集 30 分钟预测中, Model-P 的 RMSE 为 8.009, 也高于 Model-C 的 7.261。深入分析表明, Model-C 的串行设计更契合交通流数据的时空多尺度特性: 首先通过 STGCN 的切比雪夫多项式卷积提取多跳邻域的局部特征(包含中等频率成分), 继而利用 k -SVD 实现全局高频噪声抑制, 形成“特征学习 - 噪声过滤”的分层处理机制。

反观 Model-P 的并行架构, 其强制融合线性低通滤波与理想低通滤波的设计存在内在缺陷: 一方面, 额外引入的可训练参数导致训练时间复杂度显著提升, 且双模块梯度更新方向的潜在冲突易引发优化目标紊乱; 另一方面, 训练阶段 k -SVD 的理想低通特性对特征空间施加严格低频约束, 可能误将早晚高峰过渡特征等中等频率有效信号及部分高频有用信息滤除, 造成特征多样性损耗。

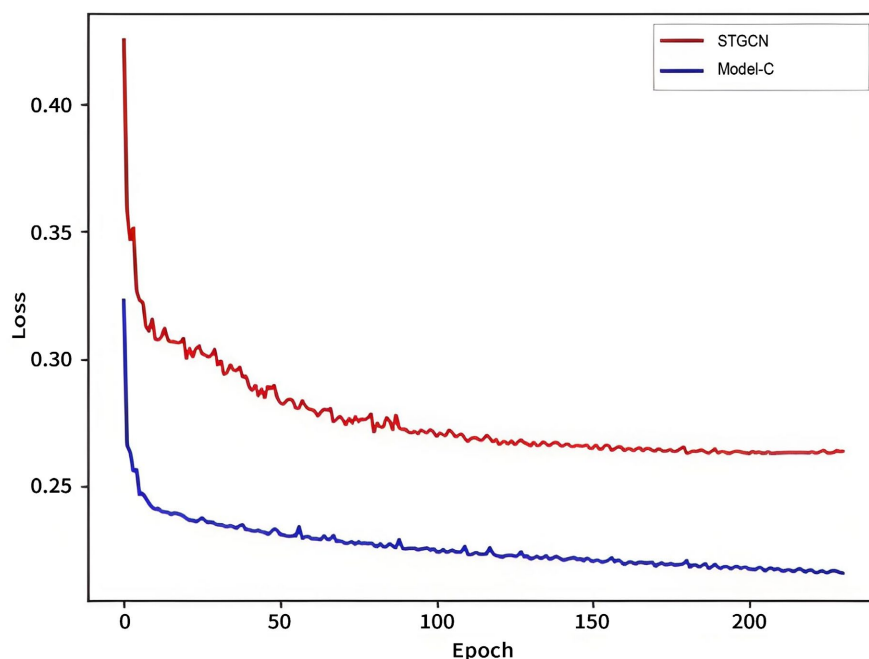


Figure 4. Comparison of loss function curves between Model-C and STGCN (METR-LA dataset, 60-minute prediction)

图 4. Model-C 与 STGCN 的损失函数曲线对比(METR-LA 数据集, 60 分钟预测)

4.4. 交通流量预测的分析

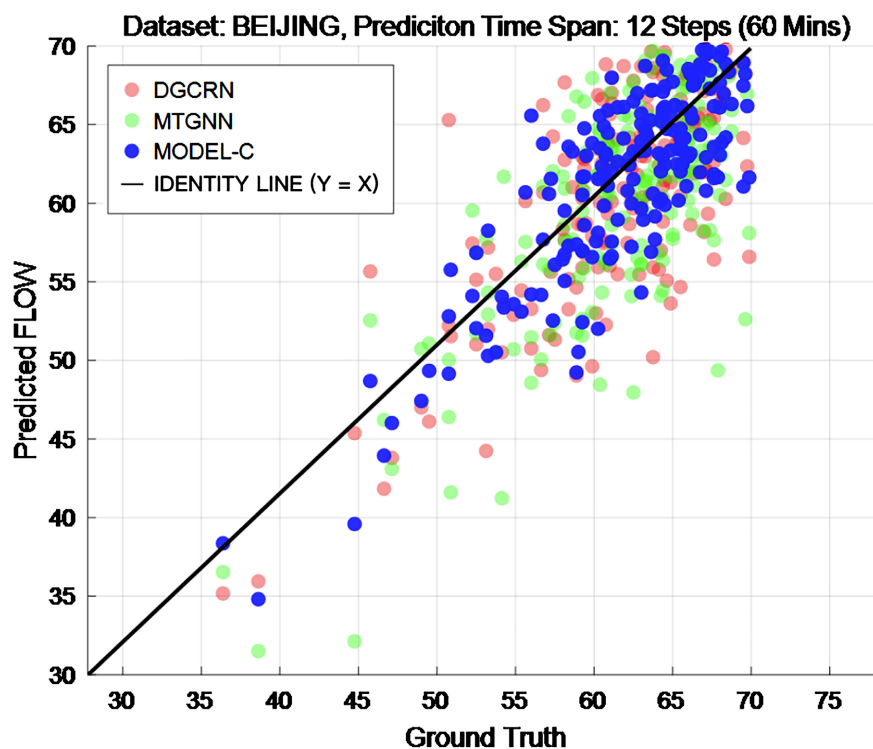


Figure 5. Predict result comparison on BEIJING dataset, 12 steps (60 mins)

图 5. 在 BEIJING 数据集 12 步(60 分钟)的预测结果对比

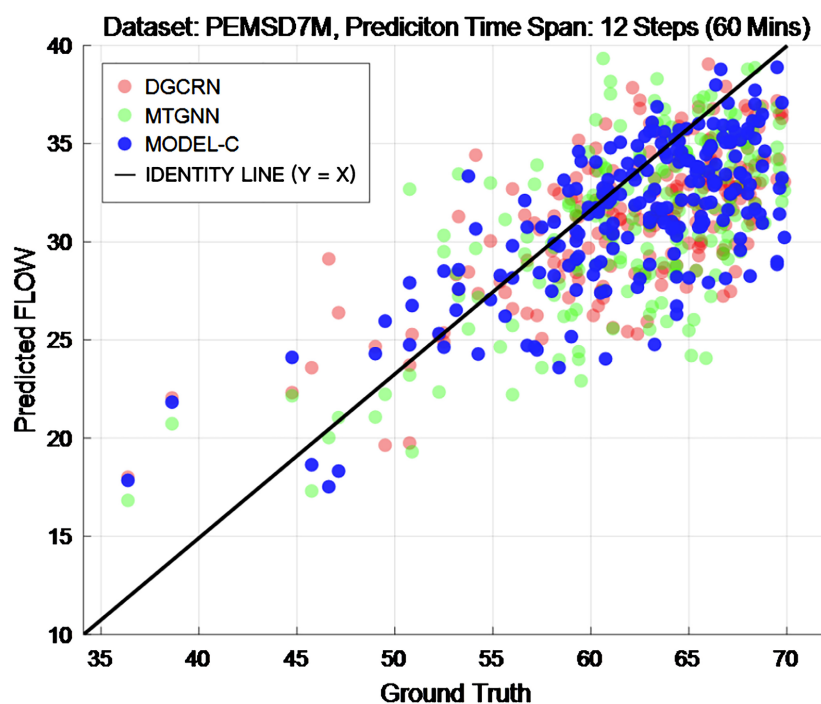


Figure 6. Predict result comparison on PEMS7M dataset, 12 steps (60 mins)

图 6. 在 PEMS7M 数据集 12 步(60 分钟)的预测结果对比

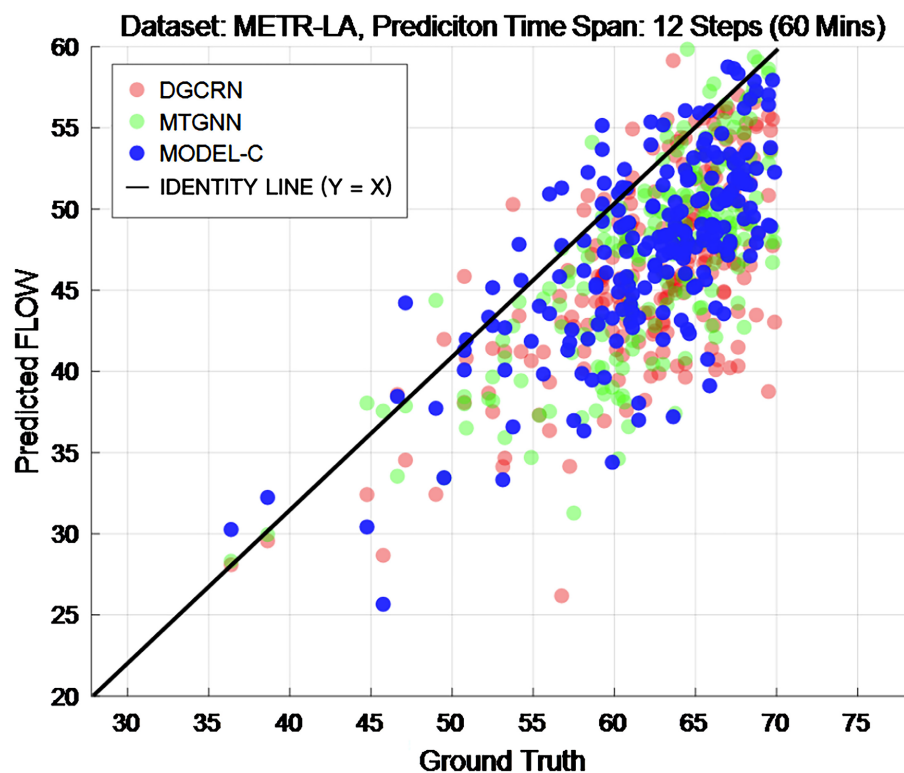


Figure 7. Predict result comparison on METR-LA dataset, 12 steps (60 mins)

图 7. 在 METR-LA 数据集 12 步(60 分钟)的预测结果对比

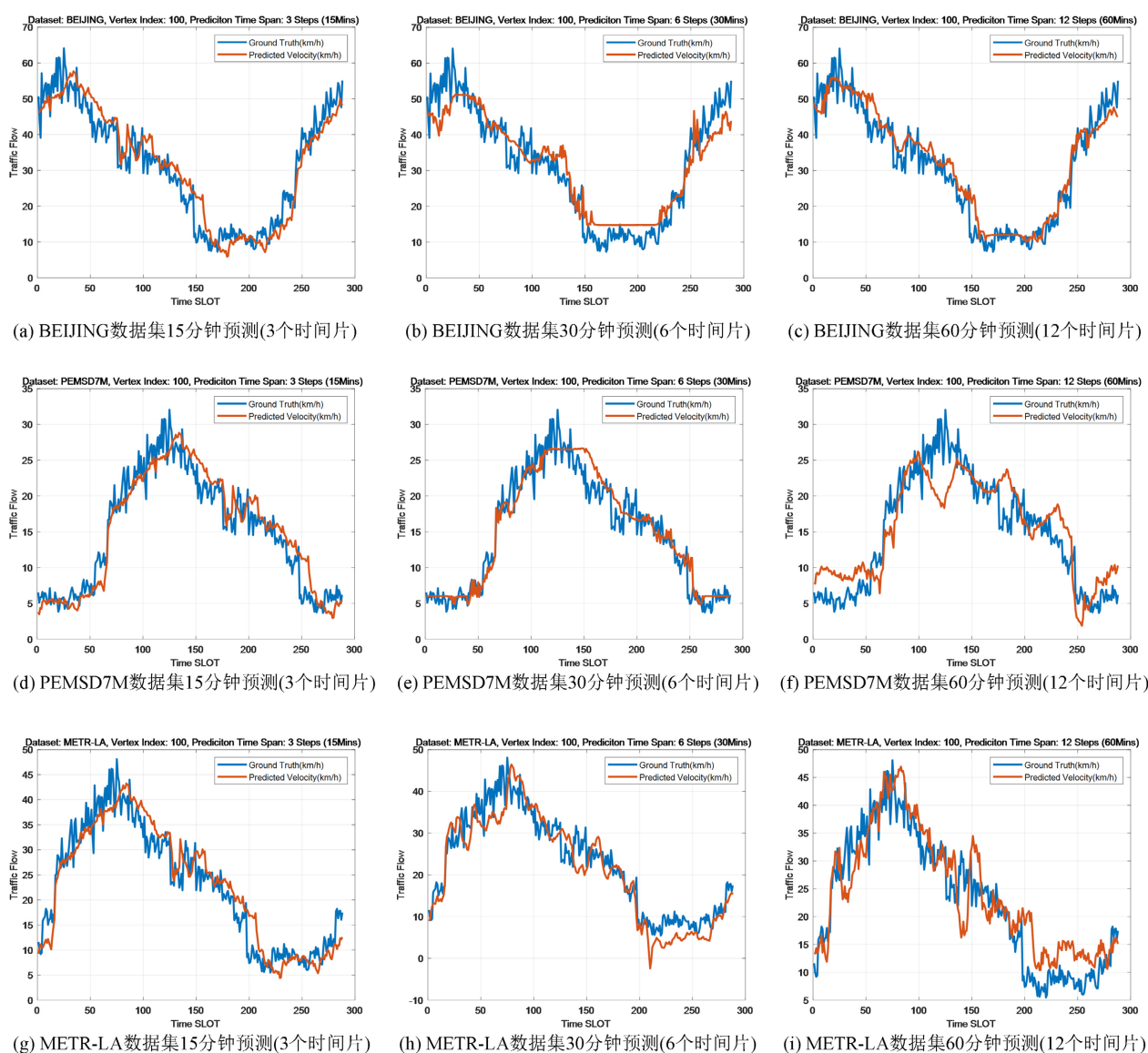


Figure 8. Prediction results for Model-C vs. ground truth

图 8. Model-C 的预测结果与真实值对比

本节通过实验评估模型在交通流量预测中的性能。如图 5~7 所示, 分别比较了同一数据集中不同模型在同一时间片上对随机选择 300 个顶点的预测结果(结果越收敛于直线 $y = x$, 则预测性能越好), Model-C 预测结果非常突出, 例如 BEIJING 数据集的预测误差显著低于基线(图 5)。

如图 8 所示, Model-C 在九个不同数据集上针对不同时间片的目标节点的预测结果与实际交通流结果的对比。预测总时长为 1 天, 即 $12 * 24 = 288$ 个时间片。实验结果表明, Model-C 的预测结果与真实值高度吻合。

时间序列分析(图 8)表明, Model-C 可以有效捕捉交通流量的高峰、低谷及稳态特征, 并做出合理的预测。然而, 但随着预测跨度越长(从 3 个时间片到 6 个时间片再到 12 个时间片), 模型对细微波动的捕捉能力下降, 预测结果就越平滑。此时, 模型对小幅波动的敏感度降低, 预测效果变差。这是由于时间模块本身存在的缺陷所致, 时间序列分析中的长程依赖问题一直难以有效解决。

此外, 从图 5~7 进一步分析发现, 与基线模型相比, Model-C 对高流量节点预测精度更优, 而对低流量节点误差较高。这可能是由于 Model-C 的基本原理是低通滤波, 即图信号通过 Model-C 后会变得更加平滑。在城市交通系统中, 高流量节点多位于连通热点区域, 节点密集且流量相似, 滤波后的平滑处理有助于精确预测。另一方面, 低流量节点和非热点区域则倾向于分散在热点区域周围, 且彼此之间仅通过少数节点连接, 这些节点的流量差异可能较大, 但低通滤波抑制了这类高频变化。这可能是 Model-C 呈现出图 5~7 所示特征的原因。此外, 影响 Model-C 性能的另一个因素是, 部分预测结果误差极大, 甚至成为异常值。若未来研究中能有效解决这一问题, 有望使整个模型在交通流量预测任务中表现更优。

5. 结束语

本文提出一种将 k -SVD 与 STGCN 相结合的交通流量预测方法, 通过在 STGCN 训练周期后对输出矩阵执行 k -SVD 的串行模块设计, 能在不引入额外可训练参数的前提下, 有效抑制预测过程中的高频噪声干扰, 在保持模型结构简洁性的同时显著提升预测精度。实验结果表明, 所提 Model-C 模型在 3 个公开交通流数据集上的短时预测精度和长时预测稳定性均达到或超越当下先进模型的性能水平, 验证了去噪机制对模型泛化能力的强化作用。但当前模型对低流量节点的预测误差较高, 未来研究将针对稀疏节点的高频噪声特性, 设计自适应低通滤波机制, 进一步提升模型在多尺度交通场景中的泛化能力与预测精度。

基金项目

RZ2400004850 强随机性下短时交通流量预测的图滤波与神经网络耦合模型研究; RZ2300000922 基于交通大数据的交通网络图属性推断研究。

参考文献

- [1] Boukerche, A., Tao, Y. and Sun, P. (2020) Artificial Intelligence-Based Vehicular Traffic Flow Prediction Methods for Supporting Intelligent Transportation Systems. *Computer Networks*, **182**, Article ID: 107484. <https://doi.org/10.1016/j.comnet.2020.107484>
- [2] Liu, Y., Wu, F., Liu, Z., Wang, K., Wang, F. and Qu, X. (2023) Can Language Models Be Used for Real-World Urban-Delivery Route Optimization? *The Innovation*, **4**, Article ID: 100520. <https://doi.org/10.1016/j.xinn.2023.100520>
- [3] Zhang, J., Wang, F., Wang, K., Lin, W., Xu, X. and Chen, C. (2011) Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, **12**, 1624-1639. <https://doi.org/10.1109/tits.2011.2158001>
- [4] Andreotti, E., Selpi and Boyraz, P. (2023) Potential Impact of Autonomous Vehicles in Mixed Traffic from Simulation Using Real Traffic Flow. *Journal of Intelligent and Connected Vehicles*, **6**, 1-15. <https://doi.org/10.26599/jicv.2023.9210001>
- [5] Lu, Q., Tettamanti, T., Hörcher, D. and Varga, I. (2019) The Impact of Autonomous Vehicles on Urban Traffic Network Capacity: An Experimental Analysis by Microscopic Traffic Simulation. *Transportation Letters*, **12**, 540-549. <https://doi.org/10.1080/19427867.2019.1662561>
- [6] Zhang, Z., Li, M., Lin, X., Wang, Y. and He, F. (2019) Multistep Speed Prediction on Traffic Networks: A Deep Learning Approach Considering Spatio-Temporal Dependencies. *Transportation Research Part C: Emerging Technologies*, **105**, 297-322. <https://doi.org/10.1016/j.trc.2019.05.039>
- [7] Lee, J., Hong, B., Lee, K. and Jang, Y. (2015) A Prediction Model of Traffic Congestion Using Weather Data. 2015 *IEEE International Conference on Data Science and Data Intensive Systems*, Sydney, 11-13 December 2015, 81-88. <https://doi.org/10.1109/dsdis.2015.96>
- [8] Coretti Sanchez, N., Martinez, I., Alonso Pastor, L. and Larson, K. (2022) On the Simulation of Shared Autonomous Micro-Mobility. *Communications in Transportation Research*, **2**, Article ID: 100065. <https://doi.org/10.1016/j.commtr.2022.100065>
- [9] Liu, Z., Liu, Y., Meng, Q. and Cheng, Q. (2019) A Tailored Machine Learning Approach for Urban Transport Network Flow Estimation. *Transportation Research Part C: Emerging Technologies*, **108**, 130-150.

- <https://doi.org/10.1016/j.trc.2019.09.006>
- [10] Xie, P., Li, T., Liu, J., Du, S., Yang, X. and Zhang, J. (2020) Urban Flow Prediction from Spatiotemporal Data Using Machine Learning: A Survey. *Information Fusion*, **59**, 1-12. <https://doi.org/10.1016/j.inffus.2020.01.002>
 - [11] Lv, M., Hong, Z., Chen, L., Chen, T., Zhu, T. and Ji, S. (2021) Temporal Multi-Graph Convolutional Network for Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, **22**, 3337-3348. <https://doi.org/10.1109/tits.2020.2983763>
 - [12] Peng, H., Du, B., Liu, M., Liu, M., Ji, S., Wang, S., *et al.* (2021) Dynamic Graph Convolutional Network for Long-Term Traffic Flow Prediction with Reinforcement Learning. *Information Sciences*, **578**, 401-416. <https://doi.org/10.1016/j.ins.2021.07.007>
 - [13] Sarker, I.H. (2021) Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, **2**, Article No. 160. <https://doi.org/10.1007/s42979-021-00592-x>
 - [14] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., *et al.* (2020) Graph Neural Networks: A Review of Methods and Applications. *AI Open*, **1**, 57-81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
 - [15] Xu, K., Hu, W., Leskovec, J., *et al.* (2018) How Powerful Are Graph Neural Networks?
 - [16] Scarselli, F., Gori, M., Hagenbuchner, M., Monfardini, G., *et al.* (2009) The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, **20**, 61-80. <https://doi.org/10.1109/tnn.2008.2005605>
 - [17] Yu, B., Yin, H. and Zhu, Z. (2018) Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 3634-3640. <https://doi.org/10.24963/ijcai.2018/505>
 - [18] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks.
 - [19] Lea, C., Flynn, M.D., Vidal, R., Reiter, A. and Hager, G.D. (2017) Temporal Convolutional Networks for Action Segmentation and Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 156-165. <https://doi.org/10.1109/cvpr.2017.113>
 - [20] Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., *et al.* (2020) T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, **21**, 3848-3858. <https://doi.org/10.1109/tits.2019.2935152>
 - [21] Chung, J., Gulcehre, C., Cho, K., *et al.* (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.
 - [22] Guo, S., Lin, Y., Feng, N., Song, C. and Wan, H. (2019) Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 922-929. <https://doi.org/10.1609/aaai.v33i01.3301922>
 - [23] Li, Y., Yu, R., Shahabi, C., *et al.* (2017) Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting.
 - [24] Cai, Z., Jiang, R., Yang, X., Wang, Z., Guo, D., Kobayashi, H.H., *et al.* (2023) MemDA: Forecasting Urban Time Series with Memory-Based Drift Adaptation. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Birmingham, 21-25 October 2023, 193-202. <https://doi.org/10.1145/3583780.3614962>