

基于优化图注意力网络的解耦式交通流预测仿真模型

马 驰, 韩天立

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2024年4月17日; 录用日期: 2024年5月13日; 发布日期: 2024年5月20日

摘 要

高效的交通流预测对提升智能交通系统的性能至关重要。其关键挑战在于精准地建模复杂的交通动态, 并全面地捕获交通流数据的时空相关性。大多数时空网络无法单独处理交通流时间序列中非平稳部分的分布变化, 并缺少在空间相关性方面建模的高效算法。为此, 本文设计一种新的基于优化图注意力网络的解耦式交通流预测仿真模型(D-EFGAT)。该模型设计了一种新的解耦-融合框架, 利用二次分解将复杂的交通流数据解耦为稳定趋势和波动事件序列, 通过双尺度时空编码网络分别对趋势和事件进行建模, 最后进行自适应融合。此外, 在图注意力网络中引入了注意力筛选机制和动态时空图编码, 更加高效地建模动态空间相关性。利用美国PeMS的交通流数据集进行仿真实验, 仿真结果表明D-EFGAT与基线模型相比具有最优的预测性能。

关键词

交通流预测, 解耦, 图注意力网络, 仿真实验

A Decoupled Traffic Flow Prediction Simulation Model Based on Enhanced Graph Attention Networks

Chi Ma, Tianli Han

School of Optical-Electrical Computer Engineering, University of Shanghai for Science & Technology, Shanghai

Received: Apr. 17th, 2024; accepted: May. 13th, 2024; published: May. 20th, 2024

Abstract

Efficient traffic flow prediction is crucial to improve the performance of intelligent transport sys-

tems. The main challenge is to accurately model complex traffic dynamics and comprehensively capture the spatio-temporal correlation of traffic flow data. Many spatio-temporal networks struggle to handle the distributional variations of the non-stationary part of the traffic flow time series and lack efficient algorithms for modeling spatial correlation. Therefore, a new decoupled traffic flow prediction simulation model (D-EFGAT) based on optimized graph attention networks was proposed. The model utilized a new decoupling-fusion framework that employed quadratic decomposition to separate complex traffic flow data into stable trends and fluctuating event sequences. The trends and events were then modeled separately using a dual-scale spatio-temporal coding network, and finally, adaptive fusion was performed. Furthermore, an attention screening mechanism and dynamic spatio-temporal graph encoding were introduced into the graph attention network to model dynamic spatial correlations more efficiently. Simulation experiments are conducted using the traffic flow dataset from PeMS in the U.S. The simulation results show that D-EFGAT has the optimal prediction performance compared to the baseline model.

Keywords

Traffic Flow Prediction, Decoupling, Graph Attention Networks, Simulation Experiment

Copyright © 2024 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

快速的城市化给现代城市交通管理带来了重大挑战。作为智慧城市的核心组成部分,智能交通系统(Intelligent Transportation Systems, ITS)被广泛应用于管理和改善交通状况。交通流预测作为 ITS 的重要技术受到了广泛关注与研究,其主要目的是通过建模路网传感器收集到的历史交通流数据来预测未来交通系统的流量。文献[1]指出,准确的交通流预测可在多种交通应用中发挥重要作用,如信号控制、线路规划与车辆调度。

鉴于交通流序列呈现出的复杂时间变化以及传感器之间动态的空间相关性,交通流预测面临着巨大的挑战。因此,为了准确预测未来的交通流量,需要高效捕获复杂而又动态的时空相关性。早期研究人员利用数理统计方法进行交通流预测,如历史平均(Historical Average, HA)、自回归综合移动平均(Autoregressive Integrated Moving Average, ARIMA)、向量自回归(Vector Autoregression, VAR)等模型。然而,这些方法受限于线性假设,难以捕捉交通流数据中的非线性相关性。随着计算资源的发展,支持向量回归和 k 近邻算法等机器学习方法被应用于交通流预测,虽然它们可以捕获更为复杂的时间依赖,但是人工提取的特征限制了它们的泛化能力。

随着深度学习在图像识别、目标检测和机器翻译等领域取得成功,其强大的能力也吸引了交通领域的研究人员。一系列交通预测方法开始利用深度学习技术,分别对每个传感器的交通数据中的时间模式进行建模。这些方法包括循环神经网络(Recurrent Neural Networks, RNN)、时空卷积网络(Time Convolutional Networks, TCN)和 Transformer。然而这些方法只考虑到交通流数据的时间依赖,忽略了路网中传感器之间复杂的空间相关性。为了捕获空间相关性,图神经网络(Graph Neural Networks, GNN)被提出。Li [2] 为了同时捕获时空相关性,提出 DCRNN (Diffusion Convolutional Recurrent Neural Network),该模型将空间依赖关系抽象为双向图随机游走,并引入了一种新的图卷积操作——扩散卷积,并在 Seq2Seq 框架中采用扩散卷积代替传统的矩阵乘法。然而,DCRNN 采用了预定义图,其基于交通网络的固定拓扑结构

构建图网络, 导致空间信息损失, 无法充分捕获路网传感器之间的动态变化和复杂关系。为了消除预定义图的影响, Wu 等人[3]提出了 GWN (Graph WaveNet), Bai 等人[4]提出 AGCRN (Adaptive Graph Convolutional Recurrent Network), 它们采用自适应图替代预定义图, 并结合节点嵌入进行学习, 通过反向传播捕获交通数据中准确的全局空间依赖。然而, 由于失去先验知识的指导, 模型容易过拟合或欠拟合。为此, Li 等人[5]提出的 STFGNN (Spatial-Temporal Fusion Graph Neural Network)通过时空融合图, 可以有效地利用交通路网中的拓扑结构和语义先验知识, 结合历史交通流数据, 构造更加精准的图架构, 以提高交通流预测的准确性和泛化能力。

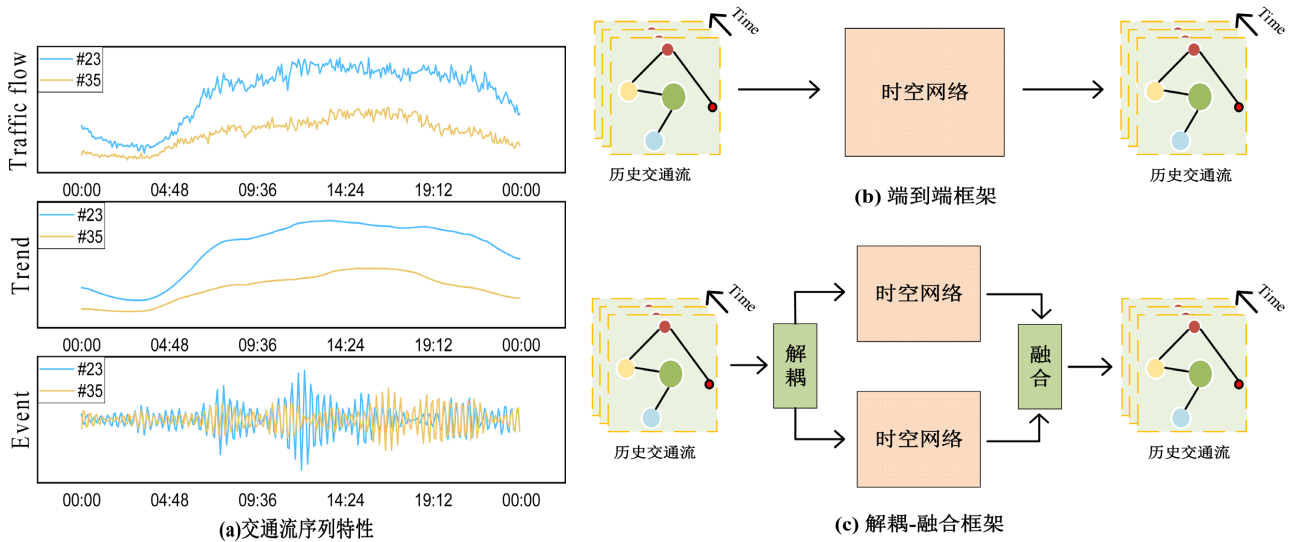


Figure 1. Traffic flow sequence characteristic and traffic forecasting framework map
图 1. 交通流序列特性和交通预测框架图

然而, 不论是预构建还是自适应生成的图, 模型在训练后的相邻节点的权重都是固定的。城市交通网络是一个动态变化的系统, 这种固定权重的方法在捕获交通节点之间的动态关联时存在限制。因此, 基于图注意力网络(Graph Attention Network, GAT)被广泛用于长期交通流预测。Guo 等人[6]提出的 ASTGCN (Attention Based Spatial-Temporal Graph Convolutional Networks)模型, 该模型利用图注意力网络分别捕获每个时间片相邻传感器之间的依赖关系。同时, 通过时空卷积模块动态调整权重。GAT 的核心问题在于它主要考虑了空间结构信息, 只捕获相邻节点间关联, 忽略了丰富的空间语义信息。例如, 相同功能的道路上或相同环境下的传感器可能具有高度相关性, 但这种相关性在 GAT 中未被充分利用。为了解决这个问题, Huang 等人[7]提出 LSGCN, 该模型在基于原始 GAT 的输入图的基础上进行了优化, 推导出完整的 GAT。完整 GAT 能够挖掘全局的空间信息。为了有效地将结构信息注入到完整 GAT 中 Feng 等人[8]提出 ASTTN (Adaptive Graph Spatial-Temporal Transformer Network), 该网络利用图拉普拉斯特征向量, 根据交通道路网络的拓扑结构生成图位置编码, 提升了模型精度。

尽管基于完整 GAT 的方法能够动态捕获全局空间信息并展现良好性能, 但在交通预测中仍存在以下限制: 1) 完整 GAT 的学习效率较低, 其模型训练的时间和空间复杂度为 $O(n^2)$, 计算成本很高, 难以在大规模数据集上的应用; 2) 完整 GAT 只考虑基于数值的空间语义信息, 缺乏先验结构知识, 容易导致图结构过度平滑。此外, 如图 1(a)所示为 PeMS08 数据集中 23# (23 号), 35#传感器采集的某一天数据, 发现交通流时间序列包含稳定趋势序列和波动事件序列, 二者相互交织在一起组成交通流序列。显然, 即使波动事件序列经历了分布的变化, 仍然可以基于稳定的趋势序列做出合理的预测。然而, 既往的模

型直接将交通数据输入时空网络, 通过端到端时空网络框架来捕获时空相关性, 如图 1(b)所示。这导致了模型难以处理波动事件序列上的分布变化, 而且由端到端时空网络捕获的时空依赖关系无法很好地推广到非平稳的交通时间序列。

为了解决上述问题, 本文提出一种基于优化完整图注意力网络的解耦式交通流预测模型(D-EFGAT)。该模型采用新的解耦 - 融合框架来缓解波动事件序列上的分布变化。如图 1(c)所示, 该框架首先利用二次分解将复杂的交通数据分解为稳定趋势和波动事件, 然后通过双尺度时空编码网络捕获时空相关性, 最后融合趋势和事件中的一些有用信息, 预测交通流量。此外, 为了克服完整图注意力网络在交通流预测中的限制, 本研究引入了一种新的注意力筛选机制, 并将动态时空图编码嵌入到完整的图注意力网络中, 以更有效地捕获动态空间相关性。经过在三个公共交通数据集上的大量仿真实验, 仿真结果显示该方法相较于基线模型具有最高的预测性能。

2. 整体设计

2.1. 符号和定义

定义 1 (交通网络) 本文将真实交通道路网络和为记录交通信息而部署的传感器抽象为一个有向图 $G = (V, E, A)$, 其中 $V = \{v_1, v_2, \dots, v_N\}$ 是一组 N 个节点, 代表道路上的传感器, E 代表有序节点对组成的有向边集合, 其中 (v_i, v_j) 代表一条从 v_i 到 v_j 的有向边, A 代表此交通网络形成的有向图的邻接矩阵, 定义如下:

$$A(i, j) = \begin{cases} 1 & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

其中 $i, j = 1, \dots, N$

定义 2 (交通状态) 本文将节点 i 在时刻 t 的交通状态记为 $x_t^i \in \mathbb{R}^C$, 其中 C 是交通流特征数目(如车辆速度、交通流量与车道占用率), 在本文中 $C = 1$, $X_t = [x_t^1, x_t^2, \dots, x_t^N] \in \mathbb{R}^{N \times C}$ 表示在时刻 t 所有节点的交通状态。

定义 5 (时间属性) 在交通网络中, 有 N 个节点, 传感器每天采样数据 N_d 次, 一周有 7 天。为了存储时间特征, 可以考虑使用两个独立的嵌入矩阵 $T^D \in \mathbb{R}^{N \times d}$, $T^W \in \mathbb{R}^{N \times d}$ 。 d 表示嵌入维数。

2.2. 问题定义

本文将交通流预测定义如下: 给定输入的历史周期数据 $X_{in} = [X_{t-T}, X_{t-(T-1)}, \dots, X_{t-1}] \in \mathbb{R}^{T \times N \times C}$, 通过模型训练的映射函数 f 来预测未来 P 时刻的交通流量, 其数学表达式如下:

$$[X_{in}; G] \xrightarrow{f} Y = [Y_{(t+1)}, Y_{(t+2)}, \dots, Y_{(t+P)}] \in \mathbb{R}^{P \times N \times C} \quad (2)$$

3. 基于优化完整图注意力网络的解耦式交通流预测模型

3.1. 总体框架

本文提出基于优化完整图注意力网络的解耦式交通流预测模型(D-EFGAT)来同步捕获时空相关性, 进行交通流预测。本研究提出的网络框架如图 2 所示, 主要有三个重要组成部分, 分别是解耦层、双尺度时空编码层和自适应融合解码层。该网络首先在解耦层应用二次分解将交通流分解为趋势 - 事件表示, 分别处理趋势的稳定特性和事件波动特性。然后在双尺度时空编码层对趋势采用时间注意力, 对事件采用因果卷积来分别捕获稳定和波动的变化。为了捕获动态空间依赖, 对趋势和事件序列同时采用优

化的完整图注意力网络(EFGAT), 该网络引入了一种新的注意力筛选机制, 并将动态时空图嵌入到完整的图注意力网络中。最后, 通过残差网络将原始交通流送到自适应融合解码层, 提取二次分解遗留的时空信息并和来自波动事件序列的有用信息一起融合到易于预测的趋势中, 充分捕获时空相关性并对交通流进行精准预测。

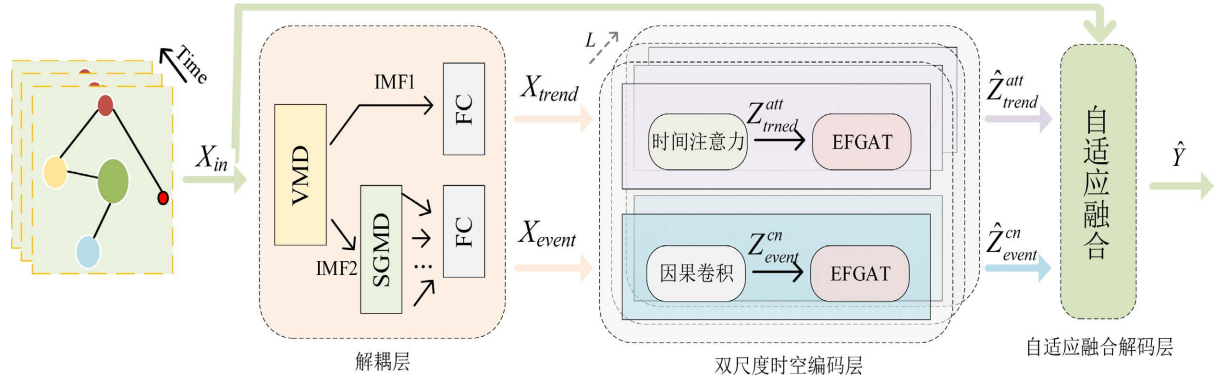


Figure 2. D-EFGAT overall framework diagram
图 2. D-EFGAT 总体框架图

3.2. 解耦层

文献[9]指出, 为了捕获复杂数据的优秀特征表示, 最佳做法是将其分解为不同的可解释来源, 以增强模型对结构丰富变量的处理能力。以交通流时间序列为例, 观察到它由两个主要成分组成: 稳定的趋势和突发的波动事件。这两个成分之间相互独立, 不会相互影响。因此, 当交通流时间序列中的一个部分发生变化, 例如, 由于事件发生分布变化, 趋势部分将不受影响。将交通时间序列分解为趋势和事件两个序列, 有助于模型更好地捕获非平稳的时间变化。为此, 本文将二次分解引入框架, 对交通流序列进行解耦操作, 其结构如图 3 所示。首先使用变分模态分解(Variable Mode Decomposition, VMD) [10]从原始交通流序列中提取低频分量和剩余高频分量, 然后使用辛几何模态分解(Symplectic Geometry Mode Decomposition, SGMD [11]对剩余的高频分量进行进一步的分解, 得到一系列干净且独立的辛几何分量。最后通过全连接层把分解得到的分量映射成趋势序列和事件序列。

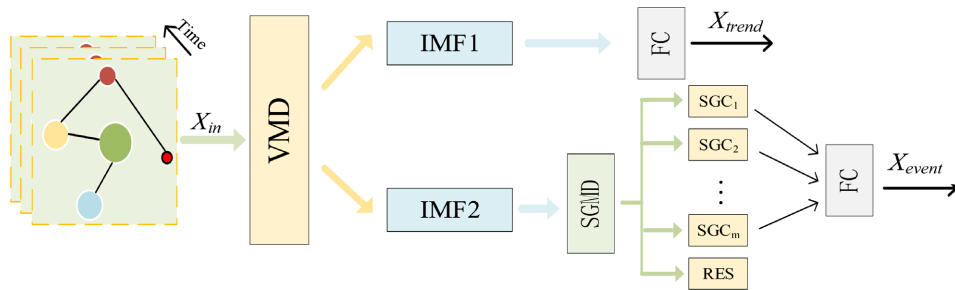


Figure 3. Decoupling layer diagram
图 3. 解耦层结构图

具体来说, 给定交通流序列 X_{in} , 输入到将分解层数设定为 2 的 VMD 后, 得到从原始序列中提取的低频分 IMF1, 包含序列中的稳定趋势或周期性成分。接着, 通过将原始序列与 IMF1 相减, 得到了一个高频分量 IMF2。然后, 利用 SGMD 对 IMF2 进行进一步分解, 其分解层数由交通流序列的特征和复杂

度自动确定, 得到 m 个辛几何分量 ($SGC_1, SGC_2, \dots, SGC_m$) 和一个残差。最终, 将低频分量视为趋势, 而将辛几何分量组成的结果视为事件。通过使用全连接层, 将趋势和事件转换成高维的表示 $X_{trend}, X_{event} \in \mathbb{R}^{T \times N \times d}$, 这样做有助于提升后续时空网络的表示能力。解耦层公式定义如下:

$$X_{trend} = W^l IMF1 + b^l \quad (3)$$

$$X_{event} = \sum_{i=1}^m (W^i SGC_i + b^i) \quad (4)$$

其中 $W^l, W^i \in \mathbb{R}^{C \times d}$, $b^l, b^i \in \mathbb{R}^d$ 都是可学习参数。经过解耦层处理, 得到了二次分解后的交通流数据的趋势和事件表示, 它们可以在下一步进行并行处理。

3.3. 双尺度时空编码层

本文精心设计了由因果卷积、时间注意力和 EFGAT 组成的双尺度时空编码层。通过堆叠编码层 L 次, 可以全面捕获交通流数据中稳定、波动的时间变化和空间相关性。

3.3.1. 时间相关性的捕获

既往模型往往直接采用单一顺序的方法来建模捕获时间相关性, 而本文提出一种新的结构, 将交通流序列拆分为趋势和事件来进行建模。从先前分析可知, 趋势和事件的时间变化是截然不同的。趋势的时间变化是稳定并且持续的, 而事件的时间变化却是波动且突然的。为此, 在处理趋势序列时, 要考虑到远处的时间片段与其紧密的时间联系, 而在处理事件序列时, 只需考虑连续邻近的时间片段。如图 4(a) 所示, 卷积核大小为 K , 空洞参数 $d=1$ 的因果卷积只能涵盖少量历史信息, 而图 4(b) 中时间注意力则能够与全局感受野中的所有历史信息进行交互。因此, 本文分别对事件和趋势使用了内核大小为 K 、空洞参数为 1 的因果卷积和时间注意力来捕获波动和稳定的时间变化。给定一个一维时序序列输入 $x \in \mathbb{R}^T$ 和一个滤波器 $f \in \mathbb{R}^K$, x 在时间步 t 上与 f 的因果卷积操作可以被定义为:

$$x * f(t) = \sum_{k=0}^{K-1} f(k) x_{t-dk} \quad (5)$$

在本文中, 事件序列 X_{event} 的因果卷积定义如下:

$$Z_{event}^{cn} = \text{Relu}(\theta * X_{event}) \quad (6)$$

其中 θ 是可学习参数, $\text{Relu}()$ 是非线性激活函数, $Z_{event}^{cn} \in \mathbb{R}^{T \times N \times d}$ 代表模型学习到的事件序列的特征表示。

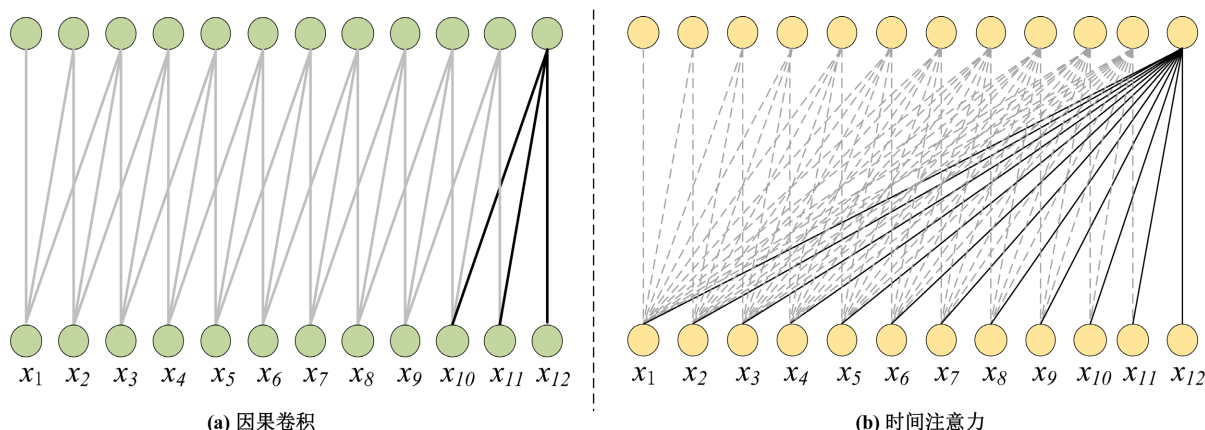


Figure 4. Causal convolution and temporal attention structures
图 4. 因果卷积和时间注意力结构图

由于趋势序列具有稳定持续的时间相关性, 所有历史时间片都与未来有很强的相关性, 所以本文利用时间注意力处理趋势序列 X_{trend} 。趋势序列的节点 n 在时间步 t 的时间注意力可定义为:

$$\alpha_{t,i}^n = \frac{\exp\left(\left(W^Q x_t^n\right)^T W^K x_i^n\right)}{\sum_{r=1}^t \exp\left(\left(W^Q x_t^n\right)^T W^K x_r^n\right)} \quad (7)$$

$$z_t^n = \sum_{i=1}^t \alpha_{t,i}^n \left(W^V x_i^n\right) \quad (8)$$

其中 x_t^n 代表趋势序列 X_{trend} 在节点 n 时间步 t 的输入, W^Q , W^K , $W^V \in R^{d \times d}$ 为可学习参数, $\alpha_{t,i}^n$ 表示节点 n 在时间步 t 与 i 之间的重要程度指数, $\exp()$ 是指数函数, $z_t^n \in Z_{trend}^{att}$ 表示模型学习到的趋势序列的特征。

3.3.2. 空间相关性的捕获

为了提高多变量交通流预测模型的性能, 许多研究已经证明捕获节点间的空间相关性是最有效的策略, 并提出了大量基于图神经网络的模型。文献[1]指出, 这些图模型大致可分为三类: 基于 GCN 的模型、基于 GAT 的模型和基于完整 GAT 的模型。然而, 基于 GCN 的模型无法捕获动态的空间相关性, 而基于 GAT 的模型仅能动态捕获相邻节点间的空间相关性。因此, 完整的 GAT 可能是交通流预测任务中最出色的空间建模技术, 因为它能够动态地捕获所有节点之间的空间相关性。在本文中, Z_{event}^{cn} , Z_{trend}^{att} 的完整 GAT 定义如下所示。为了简化公式, 删除掉 Z_{event}^{cn} , Z_{trend}^{att} 的上标和下标, 并在本小节中使用 Z 统一表示。

$$\beta_t^{n,i} = \frac{\exp\left(\left(W^Q z_t^n\right)^T W^K z_i^i\right)}{\sum_{r=1}^N \exp\left(\left(W^Q z_t^n\right)^T W^K z_r^r\right)} \quad (9)$$

$$\hat{z}_t^n = \sum_{i=1}^N \beta_t^{n,i} \left(W^V z_i^i\right) \quad (10)$$

其中 z_t^i 代表 Z_{event}^{cn} , Z_{trend}^{att} 在节点 i 时间步 t 的特征, W^Q , W^K , $W^V \in R^{d \times d}$ 为可学习参数, $\beta_t^{n,i}$ 表示在时间步 t 节点 n 与节点 i 之间的重要程度指数。 \hat{z}_t^n 是完整 GAT 输出的特征表示。但是, 正如前文所述, 原始的完整 GAT 存在两个主要限制。首先, 原始的完整 GAT 对于节点个数为 N 的网络的复杂度是二次的, 然而在真实数据集中, N 通常非常庞大, 这需要过高的计算资源。其次, 原始的完整 GAT 仅计算基于值的空间语义相关性, 缺乏先验结构知识, 容易导致图结构过度平滑, 模型过拟合。为了解决这些问题, 本文提出了一种新的 EFGAT 模型, 结合了注意力筛选机制和动态时空图编码。ESGAT 的体系结构如图 5(a) 所示。

路网区位图: 为了获取节点间先验结构知识, 许多模型用节点间距离来表示节点间空间依赖。然而, 仅仅使用距离来构建路网的拓扑结构存在缺陷。举例来说, 即使两个节点在地理上相距较远, 但如果它们位于相似的功能区域, 它们仍然很可能具有相似的交通状况。因此, 本文利用节点所在区域的相似性, 来构造图结构。其构造规则如下: 首先在 Caltrans Performance Measurement System (PeMS) 官网上获得相应传感器经纬度, 然后通过 FourSquare 将对应传感器映射到相应的路段并获取周围一定范围内十类兴趣点 (POI) 数量作为构造路网区位图的依据。POI 包括居住区、商场、停车场、加油站、饭店、客运站、公园、学校、艺术馆、工厂。最后, 将得到的路网 POI 数量矩阵 $E \in R^{I \times N}$ (I 代表兴趣点的类别数, N 代表路网节点数) 用改进的 TF-IDF 方法评估, 用以量化每个节点的每类路网兴趣点在整个兴趣点数量矩阵中

的重要程度, 具体公式如下:

$$\gamma_{i,j} = \frac{E(i,j)}{\sum_{r=1}^I E(r,j)} \times \log \frac{N}{1+|t_i \in N|} \quad (11)$$

其中 $\gamma_{i,j}$ 代表第 i 类别的 POI 在节点 j 处的重要程度, $E(i,j)$ 代表第 i 类别的 POI 在节点 j 处的数量, $|t_i \in N|$ 代表第 i 类别的 POI 在多少个节点中存在。前一项计算 j 节点的各类 POI 频率, 后项计算“逆节点”频率。“逆节点”频率代表各类 POI 对节点的重要程度。

在得到兴趣点的重要程度后, 利用余弦相似度定理计算节点之间在区位方面的相似度, 然后将该相似度作为构建路网区位图的边权重。公式如下所示:

$$A_{loc}(i,j) = \begin{cases} \frac{\sum_{r=1}^I \gamma_{r,i} \gamma_{r,j}}{\|\bar{\gamma}_i\| \times \|\bar{\gamma}_j\|} & \text{iff } \frac{\sum_{r=1}^I \gamma_{r,i} \gamma_{r,j}}{\|\bar{\gamma}_i\| \times \|\bar{\gamma}_j\|} \geq \sigma \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

其中 $\bar{\gamma}_i$ 代表节点 i 的兴趣点特征向量, σ 是用于稀疏化 A_{loc} 的阈值, 可以在不影响模型效果的情况下简化运算。

注意力筛选机制: 减小原始完整 GAT 复杂度的最直接方法是只从邻居节点处获取信息, 这种方法退化为 vanillaGAT, 失去了捕获全局信息的能力。为了捕获全局空间信息, 本文提出了一种新的注意力筛选机制, 该机制的灵感来自这样一个事实: 在交通系统中, 相同区位的节点具有相似的功能和流量。因此, 本文首先把 A_{loc} 作为稀疏查询矩阵, 筛选出与之具有较高相关性的 S 个节点, S 的筛选数量由筛选因子 e 控制, 本文设置 $S = \lceil e \log N \rceil$, 从 A_{loc} 筛选出 S 个与 n 节点最相关的节点, 记录其索引 $N_n^{loc} \in R^{S \times 1}$ 并与在交通路网中与 n 相邻的节点 N_n 合并组成新的图结构, 表示如下:

$$N_n^{att} = \text{concat}(N_n, N_n^{loc}) \quad (13)$$

其中, $\text{concat}()$ 代表合并操作。把 A_{loc} 中除了 N_n^{att} 索引中的系数作为注意力值输入到完整 GAT 中。做了此修改的完整 GAT 可表述为:

$$\beta_t^{n,i} = \frac{\exp\left(\left(W^Q z_t^n\right)^T W^K z_t^i\right)}{\sum_{r \in N_n^{att}} \exp\left(\left(W^Q z_t^n\right)^T W^K z_t^r\right)} \quad (14)$$

$$\beta_t^{n,j} = A_{loc}(n,j) \quad j \notin N_n^{att} \quad (15)$$

$$\hat{z}_t^n = \sum_{i=1}^N \beta_t^{n,i} \left(W^V z_t^i\right) \quad (16)$$

动态时空图编码: 为了有效地将时间信息注入到完整的 GAT 中, 本文提出了一种新的图位置编码方法。在普通 Transformer 架构[12]中, 序列的位置编码一直采用正弦和余弦函数, 这是自注意力机制用于区分时间片的重要组成部分。然而, 正弦曲线在图形中无法明确定义, 因为沿轴的位置没有明确的概念。为了在 GAT 中融入时间信息, 将对应时间属性的日嵌入 T^D 和周嵌入 T^W 引入时空图编码。首先将这些嵌入进行逐元素积运算, 得到新的时间嵌入 $p \in R^{N \times d}$, 其表达式为:

$$p = T^D \odot T^W \quad (17)$$

其中 \odot 代表哈达玛积, 代表对应位置元素相乘。

将新的时间嵌入与图表征相加, 完成时空图编码。最后, 本文提出的 EFGAT 可表示为:

$$\tilde{z}_t^i = z_t^i + p^i \quad (18)$$

$$\beta_t^{n,i} = \frac{\exp\left(\left(W^Q \tilde{z}_t^n\right)^T W^K \tilde{z}_t^i\right)}{\sum_{r \in N_n^{att}} \exp\left(\left(W^Q \tilde{z}_t^n\right)^T W^K \tilde{z}_t^r\right)} \quad (19)$$

$$\beta_t^{n,j} = A_{loc}(n, j) \quad j \notin N_n^{att} \quad (20)$$

$$\hat{z}_t^n = \sum_{i=1}^N \beta_t^{n,i} \left(W^V \tilde{z}_t^i\right) \quad (21)$$

在经过 EFGAT 捕获空间相关性后, 学习到的趋势和事件序列的特征表示 \hat{Z}_{event}^{cn} , \hat{Z}_{trend}^{att} 。

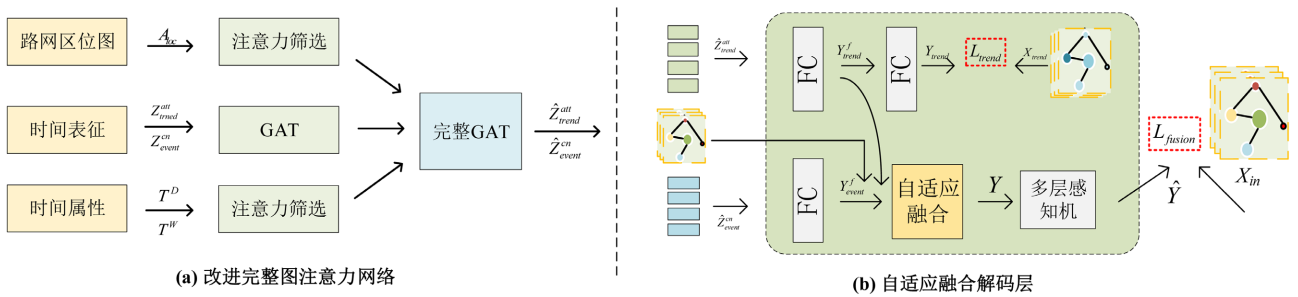


Figure 5. Structural diagram of EFGAT and adaptive fusion decoding layer

图 5. EFGAT 和自适应融合解码层结构图

3.3.3. 自适应融合解码层

为了将双尺度时空编码层学到的特征表示转化为预测的交通流, 本文提出自适应融合解码层, 其结构如图 5(b)所示。首先使用全连层在时间维度上对输入数据 \hat{Z}_{event}^{cn} , \hat{Z}_{trend}^{att} 进行操作, 以生成趋势和事件的未来表示 Y_{trend}^f , $Y_{event}^f \in R^{P \times N \times d}$ 。接着, 使用一个全连接层将趋势表征映射成 $Y_{trend} \in R^{P \times N \times 1}$, 接着进行监督学习, 以便在稳定的时间变化中获取信息。本文使用 L1 损失来实现监督:

$$L_{trend} = \sum_{i=1}^N \left| X_{trend}^i - Y_{trend}^i \right| \quad (22)$$

相较于大部分时间可以合理预测的稳定趋势不同, 波动的事件往往会导致预测结果出现分布变化。为此, 需要对事件进行合理取舍, 保留有意义的事件并剔除无用的事件。本文对趋势中每个时间片与有用事件进行加权求和, 权重由注意力机制计算, 并且可以通过反向传播学习。同时为了提取二次分解遗留的时空信息, 通过残差网络将原始交通流送到自适应融合解码层, 通过可训练参数学习。总而言之, 采用数据驱动的方法来动态评估事件预测的准确性并进行融合。自适应融合定义如下:

$$\gamma_{t,i}^n = \frac{\exp\left(\left(W^Q y_{tt}^n\right)^T W^K y_{et}^i\right)}{\sum_{r=T+1}^{T+t} \exp\left(\left(W^Q y_{tt}^n\right)^T W^K y_{er}^n\right)} \quad (23)$$

$$y_t^n = y_{tt}^n + \sum_{i=T+1}^{T+t} \gamma_{t,i}^n \left(W^V y_{ei}^n\right) + \sum_{i=T+1}^{T+t} W x_i^n \quad (24)$$

其中 $y_{tt}^n \in Y_{trend}$, $y_{et}^n \in Y_{event}$, 代表节点 n 时间步 t 的表征, $W \in R^{C \times d}$ 是可学习参数, $x_i^n \in X_{in}$ 是残差网络传

进来的原始输入, $Y \in R^{P \times N \times d}$ 是自适应融合的结果。

最后, 采用多层感知机对 Y 进行特征转换, 得到交通流预测值 $\hat{Y} \in R^{P \times N \times 1}$, 然后利用 L1 损失来监督交通预测:

$$L_{fusion} = \sum_{i=1}^N |X_{in}^i - Y^i| \quad (25)$$

全面考虑到交通流预测损失和趋势序列损失, D-EFGAT 总的损失函数定义如下:

$$L = L_{fusion} + L_{trend} \quad (26)$$

4. 仿真实验及分析

在本节中, 为了对 D-EFGAT 作定量与定性评估, 在三个真实的高速公路交通流数据集上做了大量仿真实验, 并通过消融实验验证 D-EFGAT 的各个模块的性能并分析超参数的影响。

4.1. 数据集

本文的仿真研究使用三个现实世界中开源的交通流数据集, 分别是 PEMS04、PEMS07 和 PEMS08。这 3 个数据集由 Caltrans Performance Measurement System (PeMS) 提供, 是美国加利福尼亚州主要都市区高速公路的交通流数据。数据集详细信息如表 1 所示。

Table 1. Dataset details

表 1. 数据集详情

数据集	节点数	时间步数	时间间隔	时间范围
PeMS04	307	16992	5 min	01/01/2018~01/28/2018
PeMS07	883	28224	5 min	05/01/2017~08/31/2017
PeMS08	170	17856	5 min	07/01/2016~08/31/2016

4.2. 基线模型

本文将 D-EFGAT 与以下 11 种模型进行对比:

- (1) HA: 简单历史平均方法, 将过去一时间段的平均值用于预测。
- (2) VAR [13]: 向量自回归模型是一种捕获时间相关性的时序模型。
- (3) SVR [14]: 支持向量回归是通过支持向量机对交通流预测的机器学习方法。
- (4) LSTM [15]: 具有门控机制的长短期记忆网络。
- (5) STGCN [16]: 将谱图卷积与一维卷积结合的时空图卷积模型。
- (6) DCRNN [2]: 扩散卷积循环神经网络是一种采用编码器-解码器结构, 将扩散 GCN 与 GRU 结合的模式。
- (7) GWN[3]: 一种结合门控 TCN 和空间 GCN, 采用自适应邻接矩阵的神经网络模型。
- (8) AGCRN [4]: 通过结合 GRU 和使用自适应图的 GCN, 提取时空相关性的时空模型。
- (9) STSGCN [17]: 通过构建多个局部时空图的图卷积神经网络。
- (10) ASTGCN [6]: 一种把时空注意机制引入图卷积网络的模型。
- (11) LSGCN [7]: 一种采用门控图模块来捕获长、短距离的空间依赖关系的图卷积网络。

4.3. 仿真实验参数设置

针对上述三个数据集中的缺失值, 使用线性插值进行填充, 并通过 Z-Score 归一化对数据进行标准

化, 以减少异常样本对模型的影响。所有数据集都按照 6:2:2 的比例进行划分, 分别形成训练集、验证集和测试集。

所有仿真在一台拥有 Intel(R) Core(TM) i9-12900H @ 2.50 GHz CPU 和 NVIDIA GeForce RTX 3070Ti GPU 的计算机上进行。本文采用历史时间步 $T=12$ 来预测 $P=12$ 交通流。在进行特征升维时, 输出特征维度 d 为 128, 因果卷积的核 K 大小为 2, 双尺度时空编码层的层数 L 设置为 2。本文使用 Adam 优化器训练模型, 初始学习率设置为 0.001, batch size 为 64, dropout 为 0.5, epoch 为 200。在进行 20 次训练后, 当损失不减少时, 学习率下降到 0.0001

4.4. 评价指标

本文采用三个标准评测指标来评估所有模型的性能, 分别为平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分比误差(MAPE)。其定义如下:

$$MAE = \frac{1}{N} \sum_{i=1}^N |X^i - \hat{Y}^i| \quad (27)$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{X^i - \hat{Y}^i}{X^i} \right| \quad (28)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X^i - \hat{Y}^i)^2} \quad (29)$$

其中 \hat{Y} 是模型输出预测交通流量值, X 是真实交通流量值。

4.5. 仿真结果对比与分析

本文模型 D-EFGAT 在三个真实交通流数据集上预测未来一小时(12 个时间步)的平均性能并与基线模型进行了比较, 结果如表 2 所示。观察结果可知, D-EFGAT 在数据集 PEMS04、PEMS07 和 PEMS08 上取得了最好的预测效果。特别是在数据集 PEMS08 上, D-EFGAT 模型相对于性能排名第二的模型, 其 MAE 提升了 9.87%, MAPE 也提升了 4%。

具体到各个基线模型, 数理统计方法(HA)在所有模型中表现最差, 因此可以被视为流量预测的最低水平。此外, 传统机器学习方法(VAR 和 SVR)的结果远不及基于神经网络的模型, 因为它们需要手动提取特征, 无法捕捉非线性依赖关系。神经网络模型 LSTM 性能也较差, 其根本原因是仅仅建模了交通流数据的时间相关性, 而没有考虑交通流数据的空间相关性。STGCN、DCRNN 和 STSGCN 同时考虑了时空相关性, 效果更好, 但它们在空间依赖方面仅捕获了局部信息, 只能捕获相邻节点空间相关性, 因此整体效果不及基于注意力机制的模型 ASTGCN。注意力机制能够处理较长的时间序列, 并且通过更多的参数提取不相邻节点间的空间依赖关系。GWN 与 AGCRN 由于引入自适应邻接矩阵, 也具备较强的动态时空相关性捕获能力, 模型性能较好。总而言之, D-EFGAT 在所有数据集上都取得了最佳性能, 主要有三个原因: 首先, D-EFGAT 将交通时间序列中的趋势和事件分开处理, 能够缓解分布变化; 其次, 本文模型设计了自适应融合模块和多监督功能, 有效地融合和利用了趋势和事件信息; 最后, D-EFGAT 提出了一种新动态图时空编码方法, 通过注入时间信息, 有效地捕获时空依赖。此外, 真实值与预测值之间的偏差与预测时间步长高度相关。图 6 将 D-EFGAT 与 STSGCN、GWN 和 ASTGCN 在 PeMS04、PEMS07 和 PeMS08 数据集上, 预测未来不同时间步(5~60 min)的交通流量误差变化进行对比。可以看到 D-EFGAT 在所有时间步上都比基线拥有更小的误差, 特别是在长期流量预测中。

Table 2. Comparison of average performance of different models predicting the next 12 time steps on the PeMS dataset
表 2. PeMS 数据集上不同模型预测未来 12 个时间步的平均性能比较

模型	PeMS04			PeMS07			PeMS08		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
HA	38.03	59.24	33.78%	45.12	65.64	24.51%	34.86	59.24	27.88%
VAR	24.54	38.61	17.24%	50.22	75.63	32.22%	19.19	29.81	13.10%
SVR	28.70	44.56	19.20%	32.49	50.22	14.26%	23.25	36.16	14.64%
LSTM	26.77	40.65	18.23%	29.98	45.94	13.20%	23.09	35.17	14.99%
STGCN	21.16	34.89	13.83%	25.33	39.34	11.21%	17.50	27.09	11.29%
DCRNN	21.22	33.44	14.17%	25.22	38.61	11.82%	16.82	26.36	10.92%
GWN	19.36	31.72	13.31%	21.22	34.12	9.07%	15.07	23.85	9.51%
AGCRN	19.83	32.26	12.97%	22.37	36.55	9.12%	15.95	25.22	10.09%
STSGCN	21.19	33.65	13.90%	24.26	39.03	10.21%	17.13	26.80	10.96%
ASTGNN	18.60	30.91	12.36%	20.62	34.00	8.86%	15.00	24.70	9.50%
LSGCN	21.53	33.86	13.18%	27.31	41.46	11.98%	17.73	26.76	11.20%
D-EFGAT	18.55	30.45	12.28%	20.12	33.65	8.38%	13.52	23.65	9.12%

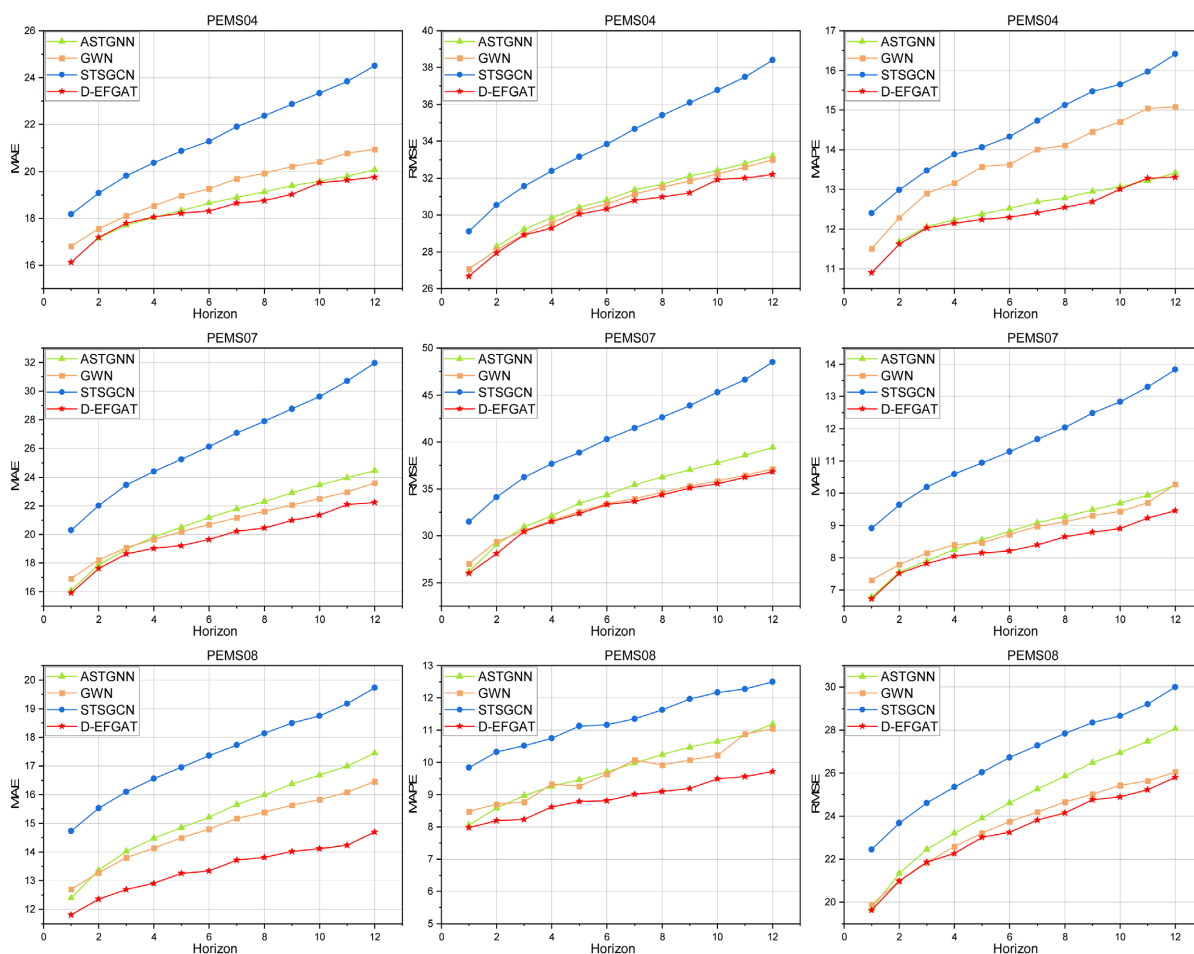


Figure 6. Comparison of prediction results of various models at different prediction time steps
图 6. 各模型在不同预测时间步长的预测结果对比图

4.6. 消融仿真实验

为了验证本文提出的解耦-融合框架在交通流预测中的有效性, 在 LSGCN 和 AGCRN 中增加了解耦-融合框架, 形成了 LSGCN+和 AGCRN+两个变体。原模型与变体的仿真结果如表 3 所示。由表 3 可知, 本文提出的解耦-融合框架在所有任务上都比端到端的框架取得可更好的效果, 验证了其可以有效地减轻事件序列的分布变化带来的影响。

为了研究 D-EFGAT 中不同组件的有效性, 本文将 D-EFGAT 与四种不同的变体进行了比较, 具体如下:

w/o DF: 在 D-EFGAT 基础上, 去除解耦层, 即遵循端到端框架, 直接将流交通流输入模型。

w/o EFGAT: 在 D-EFGAT 基础上, 把优化的完整图注意力网络替换为普通 GAT。

w/o TPM: 在 D-EFGAT 基础上, 去除 EFGAT 中的动态时空图嵌入。

w/o MS: 在 D-EFGAT 基础上, 去掉趋势事件的监督学习, 只使用交通流监督学习。

表 4 是 D-EFGAT 与四种不同的变体对比结果, 与其变体相比, 原始的 D-EFGAT 表现出最佳性能。具体而言“w/o EFGAT”的性能最差, 表明优化的完整图注意力网络至关重要, 它在全局感受野的情况下动态建模时空依赖, 能够大幅度提高模型性能。“w/o TPM”的结果表明把时间属性嵌入到 GAT 中有其存在价值, 它可以同步捕获时空相关性。“w/o DF”相较于 D-EFGAT 表现更差, 因为它忽略了在交通流序列中分离独立的分量。此外, “w/o MS”的性能不如 D-EFGAT, 这表明添加趋势事件的监督学习是正确操作。因此可以得出结论, 本文提出的解耦-融合框架和 EFGAT 对交通流预测是十分有效的。

Table 3. Performance comparison of LSGCN and AGCRN with and without decoupling-fusion framework
表 3. LSGCN 和 AGCRN 有无解耦-融合框架性能比较

模型	PeMS04			PeMS07			PeMS08		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
AGCRN	19.83	32.26	12.97%	22.37	36.55	9.12%	15.95	25.22	10.09%
LSGCN	21.53	33.86	13.18%	27.31	41.46	11.98%	17.73	26.76	11.20%
AGCRN+	19.02	31.42	12.80%	21.50	35.00	9.01%	15.30	25.01	9.93%
LSGCN+	21.00	33.12	13.56%	26.71	40.52	11.19%	17.03	26.12	10.72%

Table 4. Results of the comparison of D-EFGAT with four different variants on the PeMS dataset
表 4. PeMS 数据集上 D-EFGAT 与四种不同的变体对比结果

模型	PeMS04			PeMS07			PeMS08		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
w/o DF	19.61	32.01	13.15%	20.52	34.54	8.71%	14.21	24.69	9.65%
w/o EFGAT	20.09	33.86	14.12%	22.31	36.46	9.98%	15.73	26.76	10.30%
w/o TPM	19.49	31.92	13.03%	20.35	34.49	8.51%	14.01	24.42	9.43%
w/o MS	19.32	31.35	12.65%	20.34	34.12	8.46%	13.75	24.03	9.21%
D-EFGAT	18.55	30.45	12.28%	20.12	33.65	8.38%	13.52	23.65	9.12%

4.7. 超参数分析

为了进一步研究超参数设置对模型的影响, 本文在 PeMS04 和 PeMS08 数据集进行了仿真实验。图

7 展示了 D-EFGAT 模型随着超参数：双尺度时空编码层数 L 和特征表示维度 d 变化的性能变化。

分析误差折线图可知，随着双尺度时空编码层数增加，模型的误差先降低后升高，在 L 为 2 时，模型性能最优。其次，特征表示维度 d 为 128 时，D-EFGAT 可以达到最佳性能。显然，增加神经网络的规模可以提高表征能力，但过多的特征可能会在学习到的表征中引入噪声，导致模型性能下降。

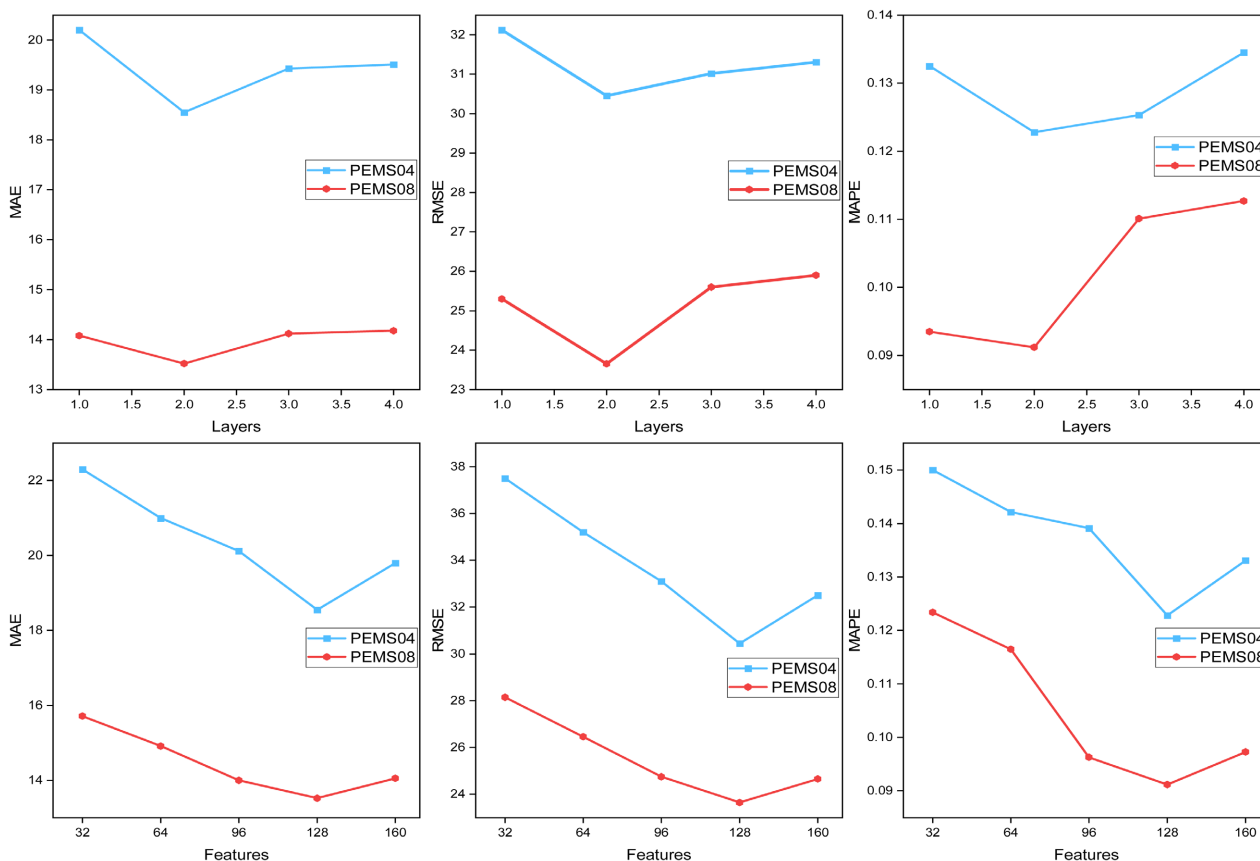


Figure 7. Prediction errors of hyperparameters d and L on PeMSD04 and PeMSD08 datasets

图 7. 超参数 d 与 L 在 PeMS04 和 PeMS08 数据集上的预测误差

5. 结束语

本文提出一种基于优化完整图注意力网络的解耦式交通流预测模型(D-EFGAT)，该模型采用一种新颖的解耦-融合框架，能够有效处理波动事件序列的分布变化问题。通过深入研究和实验验证，本文得出以下结论：首先，D-EFGAT 模型通过二次分解技术，成功地将交通数据分解为稳定趋势和波动事件两种序列。这种解耦方法有效地分化了数据的复杂性，并为后续的时空编码提供了更为清晰的数据基础。其次，在时空编码层中，结合了因果卷积、时间注意力和 EFGAT 这些组件的共同作用，全面捕获了交通数据的稳定趋势、波动分布变化和动态空间依赖，使得模型能够更加全面建模交通路网。此外，通过引入注意力筛选机制和动态时空图编码，克服了传统图注意力网络的局限，使得模型的效率更高。三个真实世界的公共数据集上的仿真实验结果表明，该模型在预测精度上显著优于现有的先进模型。通过消融实验证明了本文提出的创新点的有效性。本文提出解耦-融合框架来缓解波动事件序列上的分布变化，但在处理事件序列时采用的是因果卷积这种通用方法，还存在一些限制。在未来的研究中，将进一步探索事件序列的处理方法，优化模型适应能力并提高预测性能。

参考文献

- [1] Fang, Y., Qin, Y., Luo, H., *et al.* (2023) When Spatio-Temporal Meet Wavelets: Disentangled Traffic Forecasting via Efficient Spectral Graph Attention Networks. 2023 *IEEE 39th International Conference on Data Engineering (ICDE)*, Anaheim, 3-7 April 2023, 517-529. <https://doi.org/10.1109/ICDE55515.2023.00046>
- [2] Li, Y., Yu, R., Shahabi, C., *et al.* (2017) Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. arXiv preprint arXiv:1707.01926. <https://doi.org/10.48550/arXiv.1707.01926>
- [3] Wu, Z., Pan, S., Long, G., *et al.* (2019) Graph Wavenet for Deep Spatial-Temporal Graph Modeling. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, Macao, 10-16 August 2019, 1907-1913. <https://doi.org/10.24963/ijcai.2019/264>
- [4] Bai, L., Yao, L., Li, C., *et al.* (2020) Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting. *Advances in Neural Information Processing Systems*, **33**, 17804-17815.
- [5] Li, M. and Zhu, Z. (2020) Spatial-Temporal Fusion Graph Neural Networks for Traffic Flow Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 4189-4196. <https://doi.org/10.1609/aaai.v35i5.16542>
- [6] Guo, S., Lin, Y., Feng, N., *et al.* (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>
- [7] Huang, R., Huang, C., Liu, Y., *et al.* (2020) LSGCN: Long Short-Term Traffic Prediction with Graph Convolutional Networks. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, Yokohama, 7-15 January 2021, 2355-2361. <https://doi.org/10.24963/ijcai.2020/326>
- [8] Feng, A. and Tassioulas, L. (2022) Adaptive Graph Spatial-Temporal Transformer Network for Traffic Forecasting. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, Atlanta GA, 17-21 October 2022, 3933-3937. <https://doi.org/10.1145/3511808.3557540>
- [9] Bengio, Y., Courville, A. and Vincent, P. (2013) Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 1798-1828. <https://doi.org/10.1109/TPAMI.2013.50>
- [10] Fu, W., Zhang, K., Wang, K., *et al.* (2021) A Hybrid Approach for Multi-Step Wind Speed Forecasting Based on Two-Layer Decomposition, Improved Hybrid DE-HHO Optimization and KELM. *Renewable Energy*, **164**, 211-229. <https://doi.org/10.1016/j.renene.2020.09.078>
- [11] Pan, H., Yang, Y., Li, X., *et al.* (2019) Symplectic Geometry Mode Decomposition and Its Application to Rotating Machinery Compound Fault Diagnosis. *Mechanical Systems and Signal Processing*, **114**, 189-211. <https://doi.org/10.1016/j.ymsp.2018.05.019>
- [12] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [13] Zivot, E. and Wang, J. (2006) Vector Autoregressive Models for Multivariate Time Series. In: Zivot, E. and Wang, J., Eds., *Modeling Financial Time Series with S-PLUS®*, Springer, New York, 369-413.
- [14] Rucker, H., Burges, C.J., Kaufman, L., *et al.* (1996) Support Vector Regression Machines. *Proceedings of the 9th International Conference on Neural Information Processing Systems*, Denver, 3-5 December 1996, 155-161.
- [15] Hochreiter, S. and Schmidhuber J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [16] Li, C.L., Cui, Z., Zheng, W.M., *et al.* (2018) Spatio-Temporal Graph Convolution for Skeleton Based Action Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **32**. <https://doi.org/10.1609/aaai.v32i1.11776>
- [17] Song, C., Lin, Y.F., Guo, S.N., *et al.* (2020) Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial Temporal Network Data Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 914-921. <https://doi.org/10.1609/aaai.v34i01.5438>