

基于时序分解和注意力图神经网络的交通预测

杨永鹏¹, 杨震¹, 杨真真^{1,2}

(1. 南京邮电大学通信与信息工程学院, 南京 210023; 2. 南京邮电大学理学院, 南京 210023)

摘要: 如何有效挖掘隐藏在交通数据中的时空依赖信息、动态信息和空间异质信息一直是交通预测任务面临的关键问题。本文提出了一种基于时序分解和注意力图神经网络(Time-series decomposition and attention graph neural network, TDAGNN)的交通预测模型。采用双分支时序分解卷积神经网络(Dual time-series decomposition convolutional neural network, DTDCNN)从复杂的交通数据中挖掘时间依赖信息;采用多头交互注意力网络(Multi-head interactive attention, MIA)对原始交通特征和局部增强特征进行交互学习,以深入挖掘交通数据的异质信息和动态信息;引入自缩放动态扩散图神经网络(Self-scaling dynamic diffusion graph neural network, SDDGNN)在获取交通数据空间依赖信息的同时,避免图神经网络的尺度失真问题;将提出的TDAGNN应用于经典交通数据PEMS04、PEMS08、METR-LA和PEMS-BAY的交通预测实验中。实验结果表明,提出模型的平均MAE、RMSE和MAPE比其他经典算法最大可分别提高14.64、23.68和9.41%,从而证明其具有较高的交通预测精度。

关键词: 交通预测; 时序分解; 图神经网络; 注意力机制; 局部增强网络

中图分类号: TP391; TP183

文献标志码: A

Time-Series Decomposition and Attention Graph Neural Network Based Traffic Forecasting

YANG Yongpeng¹, YANG Zhen¹, YANG Zhenzhen^{1,2}

(1. School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2. College of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: In order to address challenges on how to accurately capture the spatial-temporal dependency, dynamic information and spatial heterogeneity information in traffic forecasting, we propose a time-series decomposition and attention graph neural network (TDAGNN) based traffic forecasting. Specifically, the model first adopts the dual time-series decomposition convolutional neural network (DTDCNN) to extract temporal dependency from traffic data. Secondly, the multi-head interactive attention (MIA) network is introduced to capture spatial heterogeneity and dynamicity from traffic data via the interactivity between the original features and the local augmentation features. Thirdly, the self-scaling dynamic diffusion graph neural network (SDDGNN) is introduced for capturing the spatial dependence and dynamicity from the traffic data. Finally, extensive experiments are carried out for some datasets. Experimental results demonstrate that the average MAE, RMSE and MAPE of the proposed model can be improved up to 14.64, 23.68 and 9.41% respectively, compared to other classic algorithms, proving its high prediction accuracy.

基金项目: 国家自然科学基金(62071242, 62171232); 江苏省研究生科研与实践创新计划项目(KYCX22_0955)。

收稿日期: 2024-05-13; **修订日期:** 2024-07-29

Key words: traffic forecasting; time-series decomposition; graph neural network; attention mechanism; local augmentation network

引言

近年来,随着我国经济的迅猛发展和城市化进程的不断加快,交通拥堵等问题已经成为影响我国经济发展的重要因素之一。为有效利用现有的交通基础设施来缓解交通资源与交通流量之间的冲突,智能交通系统(Intelligent transportation system, ITS)被提出并广泛应用于交通管理中。作为ITS的核心部分,交通预测^[1-3]通过分析历史交通数据,综合考虑交通多元要素,对交通速度或交通流量分别进行预测,从而制定有效的交通管制措施,缓解交通拥堵,直接或间接拉动经济增长,已成为交通领域研究的热点。

早期基于统计的交通预测模型包括自回归整合移动平均(Autoregressive integrated moving average, ARIMA)模型^[4-5]、卡尔曼滤波(Kalman filtering, KF)模型^[6]和向量自回归(Vector autoregressive, VAR)模型^[7]等。此类模型固有的线性特点违背了交通数据的复杂特征,使交通预测效果差强人意。针对上述模型存在的问题,基于传统机器学习的交通预测模型被提出,如支持向量回归(Support vector regression, SVR)模型^[8]和K最近邻(K-nearest neighbor, KNN)模型^[9]等。但此类模型的繁杂手动标注、浅层结构和时空依赖信息不敏感等问题导致其预测效果欠佳。

在此背景下,以卷积神经网络(Convolutional neural network, CNN)^[10]和循环神经网络(Recurrent neural network, RNN)^[11-12]为代表的基于深度学习的交通预测模型被陆续提出。此类模型将交通数据视为网格型数据,这种理念违背了交通数据是具有图状特性数据的宗旨,从而影响了交通预测的性能。近年来,基于Transformer的深度学习模型^[13-16]凭借其在长期时序预测的优势成为交通预测研究的另一个热点。例如,Informer模型^[13]引入了稀疏自注意力和生成式解码器等机制。频率增强分解Transformer(Frequency enhanced decomposed Transformer, FEDformer)模型^[16]将时序信号进行分解,并从频域的角度实施交通预测。但上述模型忽略了交通数据的空间依赖信息。

基于此,图神经网络(Graph neural network, GNN)^[17-21]应运而生并被广泛应用于交通预测中。例如,基于注意力的时空图卷积神经网络(Attention based spatial-temporal graph convolutional network, ASTGCN)的交通预测模型^[22]分别采用注意力机制和图卷积神经网络(Graph convolutional network, GCN)从时空域进行预测。基于图小波(Graph wavenet, GW)的交通预测模型^[23]采用门控时间卷积网络(Gated temporal convolution network, GTCN)^[24]和GCN捕获交通数据的时空依赖信息。基于时空同步图卷积神经网络(Spatial-temporal synchronous graph convolutional network, STSGCN)模型^[25]设计了一种时空同步模型,有效捕捉交通数据的局部时空依赖信息。动态图卷积循环网络(Dynamic graph convolutional recurrent network, DGCRN)模型^[26]采用动态滤波器构造提取动态信息的超网络,从而提高交通预测的性能。基于解耦动态时空图神经网络(Decoupled dynamic spatial-temporal graph neural network, D2STGNN)模型^[27]对扩散过程中产生的隐藏时间序列和独立于其他传感器的隐藏时间序列进行解耦,能够对交通数据的不同部分进行更精确的建模。

尽管上述模型已取得较好的交通预测效果,但交通数据独有的复杂特性使其仍面临诸多挑战,具体表现为如何有效挖掘交通数据中存在的时空依赖信息、动态信息和异质信息等,分别表述如下:

(1) 如何有效挖掘时空依赖信息。具体地说,在空间维度上,对于交通网络中某个特定的节点而言,不同位置的节点对其交通数据的未来变化趋势有着不同的影响。一般而言,与其邻近或相似度高的节点对其影响程度要高于距离较远或相似度低的节点。在时间维度上,对于交通网络中的每个节点

而言,某个特定时刻的交通数据依赖于历史交通数据的变化趋势。一般而言,邻近时间的相关性比其他时间的相关性大。基于此,通过综合考虑交通数据所特有的时空特性,深入挖掘交通数据中所蕴含的时空依赖性是该领域所要面对的第一个挑战。

(2) 如何有效挖掘动态信息。所谓动态信息,是指由于早晚高峰、恶劣天气、道路施工,以及节假日等突发因素的影响,相同位置的节点在不同时刻对某个固定节点的影响是不断变化的。同时,不同节点对某个固定节点的影响是随着时间的演进不断变化的。故如何从交通数据中挖掘对应的动态信息是交通预测面临的另一项挑战。

(3) 如何有效挖掘异质信息。所谓异质信息,是指处于交通网络中不同区域的两个结点在相同时刻的交通数据模式是不同的。例如,处于居民区域的节点交通数据与处于医院区域的交通数据是截然不同的。因此,从交通数据中挖掘对应的异质信息对于交通预测精度的提高是有益的。

针对上述问题,本文的主要贡献是提出了一种基于时序分解和注意力图神经网络(Time-series decomposition and attention GNN, TDAGNN)的交通预测模型。具体地说,该模型首先采用双分支时序分解卷积神经网络(Dual time-series decomposition CNN, DTDCNN)模块以有效挖掘隐藏在交通数据中的时间依赖信息;其次,采用多头交互注意力(Multi-head interactive attention, MIA)网络模块以深入挖掘交通数据中的动态信息和异质信息;最后,采用自缩放动态扩散图神经网络(Self-scaling dynamic diffusion GNN, SDDGNN)模块,有效地挖掘交通数据中的空间依赖信息。

与已有交通预测模型对比,本文模型的创新点如下:

(1) TDAGNN模型引入的DTDCNN模块采用时序分解方法将复杂的交通数据从时间维度上分解为两个简单的、易于处理的周期数据和趋势数据。之后,采用门控机制将两个并行时序分解的时间卷积神经网络分支进行信息合并为DTDCNN,从而通过控制信息流在不同层之间的流动来学习交通数据中更复杂的时间依赖性。

(2) TDAGNN模型引入的MIA模块从原始特征和局部增强特征两个角度交互获取节点间的注意力系数,深入挖掘交通数据中的动态信息和异质信息。

(3) TDAGNN模型引入的SDDGNN模块采用动态邻接矩阵捕获交通数据的空间动态性,同时采用扩散图神经网络捕获交通数据的空间依赖信息,另外引入自缩放网络有效地避免图神经网络中存在的尺度丢失等问题,最终增强模型挖掘交通数据中的空间依赖信息的性能。

(4) 通过大量的实验表明提出的TDAGNN模型具有较好的优越性。

1 相关工作

1.1 注意力机制

注意力机制^[28-30]通过提取数据中的重要信息而被广泛应用于交通预测、自然语言处理和多媒体处理等任务中。例如, Velickovic等^[28]将自注意力机制引入到图神经网络中,提出了图注意力网络; Brody等^[29]引入了动态注意力机制,推动其对图学习的能力; Shih等^[31]从频域出发,提出了一种时间序列预测注意力机制模型; Touvron等^[32]提出了基于自注意力的全局映射来增强卷积网络的功能。上述工作表明,基于自注意力机制的模型极大推动了机器学习在各领域中的应用。

1.2 图神经网络

GNN的核心思想是采用消息传递和聚合机制对存在于非欧氏空间的复杂信息进行学习,在人工智能领域取得了优异的成绩。例如, Kipf等^[19]对谱图卷积进行一阶近似逼近,提出了GCN的最初模型。Hamilton等^[33]提出了图抽样聚合(Graph sample and aggregation, GraphSAGE)神经网络,提高了GNN

的学习效率。Klicpera等^[34]将个性化网页排序与GCN相结合,提出了个性化传播神经预测(Personalized propagation of neural predictions, PPNP)模型。

1.3 时序分解

现有的交通预测模型大都将含有多种时间模式的复杂时间序列视为一个整体,其处理难度大,预测效果差。在此背景下,时序分解(Time-series decomposition, TD)^[15-16]通过将时间序列分解成表征周期性和趋势性的多个时间模式来提升时序预测的性能。例如自相关Transformer(Auto-correlation Transformer, Autoformer)^[15]和FEDformer^[16]皆将时序分解与Transformer有机结合,以增强时序预测的性能。

1.4 数据增强

数据增强通过扩充训练集、挖掘隐含节点信息等手段进一步增强模型的泛化能力。例如,Liu等^[35]提出了一个局部增强的GNN模型,该模型以中心节点的条件分布为理论基础,推导出增强GNN表征能力的局部数据增强方法。Kefato等^[36]提出了一个可联合学习数据增强的自监督GNN(Jointly learnable data augmentations for self-supervised GNN, JLDAS)模型用于自监督GNN的联合可学习数据扩充。

2 基于TDAGNN的交通预测

2.1 交通预测的基本概念

交通网络的定义:交通网络是一种复杂的图网络结构,常用 $G=(V, E, A)$ 表示,其中 V 表示包含 $N=|V|$ 个节点的交通网络的节点集合,图中每个节点为交通网络中的交通传感器,其功能为抓取对应区域的交通数据, E 为交通网络中节点之间边的集合, A 为反映图结构的邻接矩阵。

交通数据的定义:交通数据是由架设在城市路口的传感器以特定时间片抓取的包括交通流量和交通速度^[22]的数据,表示为 $\mathcal{X}=(X_{t-T+1}, X_{t-T+2}, \dots, X_t) \in \mathbb{R}^{T \times N \times F}$,其中 T 为抓取交通数据的时间, N 为路网中节点的个数, F 为每个节点所对应的交通特征数目,每个元素 $X_t=(X_t^1, X_t^2, \dots, X_t^N) \in \mathbb{R}^{N \times F}$ 为 t 时刻抓取的 N 个节点的特征。

交通预测问题的定义:交通预测的目的是根据过去 T 时间的交通数据 $\mathcal{X}$ 及交通路网 G ,预测未来的交通数据 $\hat{\mathcal{Y}}=(\hat{X}_{t+1}, \hat{X}_{t+2}, \dots, \hat{X}_{t+s})$,即

$$\hat{\mathcal{Y}} = f(\mathcal{X}, G) \quad (1)$$

式中: f 为预测模型,其性能决定于真实值 $\mathcal{Y}=(X_{t+1}, X_{t+2}, \dots, X_{t+s})$ 与预测值 $\hat{\mathcal{Y}}$ 的逼近程度。另外,该模型可针对交通数据中的交通流量或交通速度分别进行预测。

本文提出的TDAGNN模型包括两个多层感知机(Multi-layer perception, MLP)、 L 层堆叠的时空块和 L 个CNN模块,其中每个时空块包括双分支时序分解卷积神经网络(DTDCNN)、多头交互注意力网络(MIA)和自缩放动态扩散图神经网络(SDDGNN)三部分。其框架如图1所示。

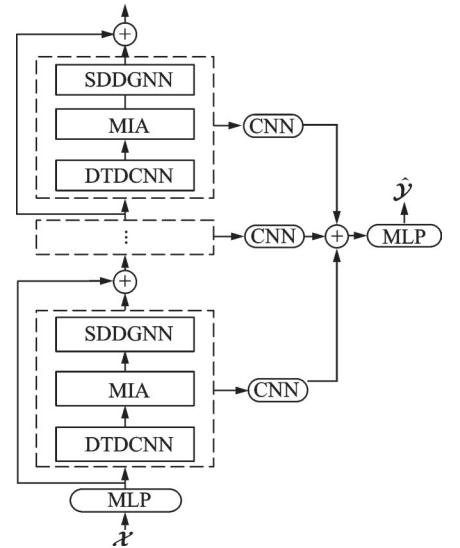


图1 基于时序分解和注意力图神经网络框架

Fig.1 Framework of TDAGNN

TDAGNN模型的数据处理流程为:首先,MLP将输入的 $\mathcal{X}=(X_{t-T+1}, X_{t-T+2}, \dots, X_t) \in \mathbb{R}^{T \times N \times F}$ 映射到高维空间得 $\tilde{\mathcal{X}} \in \mathbb{R}^{T \times N \times H}$,其中 H 为映射后的交通数据特征维度;其次,将得到的 $\tilde{\mathcal{X}}$ 依次经过 L 个堆叠的时空块,且每个时空块输出的结果经过CNN处理并累加后再经MLP得到最终预测结果 $\hat{\mathcal{Y}}=(\hat{X}_{t+1}, \hat{X}_{t+2}, \dots, \hat{X}_{t+S})$ 。TDAGNN模型第 l 个时空块的处理过程为:假设输入到第 l 个时空块的交通数据为 $\tilde{\mathcal{X}}^l$,其中 $\tilde{\mathcal{X}}^l$ 是经过第 $l-1$ 个时空块的输入和输出做残差连接后得到的。首先, $\tilde{\mathcal{X}}^l$ 经过DTDCNN模块得到 $\hat{\mathcal{X}}^l$;然后,将 $\hat{\mathcal{X}}^l$ 输入到MIA模块得到 $\tilde{\mathcal{X}}^l$;最后,将 $\tilde{\mathcal{X}}^l$ 输入到SDDGNN模块并得到 $\bar{\mathcal{X}}^l$ 。

2.2 双分支时序分解卷积神经网络

为了充分挖掘交通数据中的时间依赖信息,将时序分解与时间卷积网络(Temporal convolutional network, TCN)相结合提出了DTDCNN模型,其框架如图2所示。

DTDCNN首先引入两个并行的时序分解卷积神经网络(Time-series decomposition convolutional neural network, TDCNN)分别同时处理输入的交通数据 $\tilde{\mathcal{X}}^l$ 。随后,采用门控机制和全连接层(Fully connected layer, FC)处理上述两个并行TDCNN输出的交通数据得到DTDCNN最终的交通数据 $\hat{\mathcal{X}}^l$ 。其中,每一个TDCNN包括时序分解模块、分别处理周期数据和趋势数据的TCN和加法器。具体描述如下。

时序分解模块:TD-a和TD-b为两个具有相同功能TD模块。首先,该模块采用移动平均(Moving average, MA)算法^[37]提取交通数据在时间维度上的趋势数据;其次,TD模块将原始交通数据与趋势数据相减得到对应的周期数据,具体如下

$$\begin{cases} \tilde{\mathcal{X}}_t^l = \text{AveragePool}\left(\text{Pad}\left(\tilde{\mathcal{X}}^l\right)\right) \\ \tilde{\mathcal{X}}_s^l = \tilde{\mathcal{X}}^l - \tilde{\mathcal{X}}_t^l \end{cases} \quad (2)$$

式中:AveragePool($\cdot$)为移动平均池化函数,Pad($\cdot$)为时间序列填充函数, $\tilde{\mathcal{X}}_t^l$ 和 $\tilde{\mathcal{X}}_s^l$ 分别为趋势数据和周期数据。由上可得,交通数据经过TD-a和TD-b,最终得到两组并行的趋势数据和周期数据: $\tilde{\mathcal{X}}_{t-a}^l$ 和 $\tilde{\mathcal{X}}_{s-a}^l$ 、 $\tilde{\mathcal{X}}_{t-b}^l$ 和 $\tilde{\mathcal{X}}_{s-b}^l$ 。

时序分解卷积神经网络:如图2所示,TCN-a1、TCN-a2、TCN-b1和TCN-b2都为功能相同TCN。首先,采用TCN-a1和TCN-a2处理 $\tilde{\mathcal{X}}_{t-a}^l$ 和 $\tilde{\mathcal{X}}_{s-a}^l$,并将结果相加,采用TCN-b1和TCN-b2处理 $\tilde{\mathcal{X}}_{t-b}^l$ 和 $\tilde{\mathcal{X}}_{s-b}^l$,并将结果相加,具体如下

$$\begin{cases} \tilde{\mathcal{X}}_{c-a}^l = \text{TCN}\left(\tilde{\mathcal{X}}_{s-a}^l\right) + \text{TCN}\left(\tilde{\mathcal{X}}_{t-a}^l\right) \\ \tilde{\mathcal{X}}_{c-b}^l = \text{TCN}\left(\tilde{\mathcal{X}}_{s-b}^l\right) + \text{TCN}\left(\tilde{\mathcal{X}}_{t-b}^l\right) \end{cases} \quad (3)$$

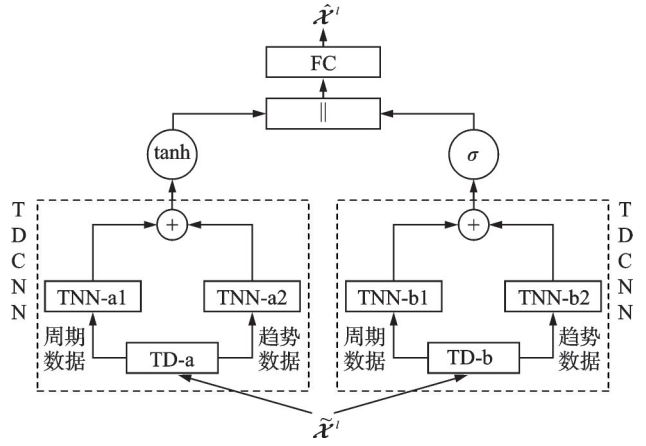


图2 双分支时序分解卷积神经网络框架

Fig.2 Framework of DTDCNN

随后,将 $\tilde{\mathbf{x}}_{c-a}^l$ 和 $\tilde{\mathbf{x}}_{c-b}^l$ 分别经过门控机制和FC模块的处理,得到DTDCNN模块的最终结果,具体如下

$$\hat{\mathbf{x}}^l = \text{FC}\left(\text{concat}\left(g\left(\tilde{\mathbf{x}}_{c-a}^l\right), \sigma\left(\tilde{\mathbf{x}}_{c-b}^l\right)\right)\right) \quad (4)$$

式中: $g(\cdot)$ 和 $\sigma(\cdot)$ 为激活函数,本文选择tanh函数和Sigmoid函数,concat($\cdot$)为对 $g\left(\tilde{\mathbf{x}}_{c-a}^l\right)$ 和 $\sigma\left(\tilde{\mathbf{x}}_{c-b}^l\right)$ 在时间维上的拼接操作,FC($\cdot$)为全连接函数。

2.3 多头交互注意力网络

为了充分挖掘交通数据中的动态信息和异质信息,提出了一种多头交互注意力(MIA)网络,如图3所示。

具体地说,MIA模块首先采用局部增强网络(Local augmentation network, LANet)^[35]对 $\hat{\mathbf{x}}^l$ 处理得到新交通数据 $\hat{\mathbf{x}}_a^l$,其中LANet是以中心图节点和服从标准正态分布的潜在表征为条件,采用最大似然估计更新图节点的特征,即

$$\hat{\mathbf{x}}_a^l = \text{MLP}\left(\text{concat}\left(Z, \hat{\mathbf{x}}^l\right)\right) \quad (5)$$

式中: $\hat{\mathbf{x}}^l$ 为MIA模块的输入,MLP($\cdot$)为多层感知机, $Z \sim \mathcal{N}(0, I)$ 为服从标准正态分布的可学习矩阵。

其次,MIA采用自注意力的交互策略,从而增强获取交通数据中动态信息和异质信息的能力,即

$$\hat{\mathbf{x}}^l = \text{softmax}\left(\frac{\alpha \mathbf{Q}_1 \mathbf{K}_2^T + (1-\alpha) \mathbf{Q}_2 \mathbf{K}_1^T}{\sqrt{d}}\right) \mathbf{V} \quad (6)$$

式中: $\mathbf{Q}_1$ 和 $\mathbf{K}_1$ 、 $\mathbf{Q}_2$ 和 $\mathbf{K}_2$ 分别为 $\hat{\mathbf{x}}^l$ 、 $\hat{\mathbf{x}}_a^l$ 经CNN处理后的值, $\mathbf{V}$ 为 $\hat{\mathbf{x}}^l$ 经FC处理后的值, α 为超参数, d 为数据的维度。

2.4 自缩放动态扩散图神经网络

为了增强模型挖掘交通数据中的空间依赖信息和缓解GNN的尺度丢失问题,引入了SDDGNN,主要包括动态扩散图神经网络(Dynamic diffusion GNN, DDGNN)和自缩放网络两个部分,如图4所示。

众所周知,GCN是在一种给定图拓扑结构条件下提取图节点特征的有效操作。从空间维度上,主要采用了信息的传递和聚合机制,其公式为

$$\hat{\mathbf{x}}_o^l = \tilde{\mathbf{A}} \mathbf{W} \hat{\mathbf{x}}^l \quad (7)$$

式中: $\tilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ 为归一化邻接矩阵, $\mathbf{A}$ 为图的邻接矩阵, $\mathbf{D}$ 为 $\mathbf{A}$ 的度矩阵, $\mathbf{W}$ 为学习参数矩阵。

近年来,Li等^[38]提出了一种有效的扩散卷积神经网络,采用了有限的 K 步来建模图信号的扩散过程。该方法已被证明是一种有效的时空建模方法。于是,本文首先将扩散卷积神经网络推广到式(7)中,提出了扩散图神经网络(Diffusion GNN, DGNN),表达式为

$$\hat{\mathbf{x}}_o^l = \sum_{k=0}^K (\tilde{\mathbf{A}})^k \hat{\mathbf{x}}^l \mathbf{W}_{k1} \quad (8)$$

由于交通网络中的两个节点的关系是双向的,故扩散过程是双向的。基于此,引入了前向邻接矩

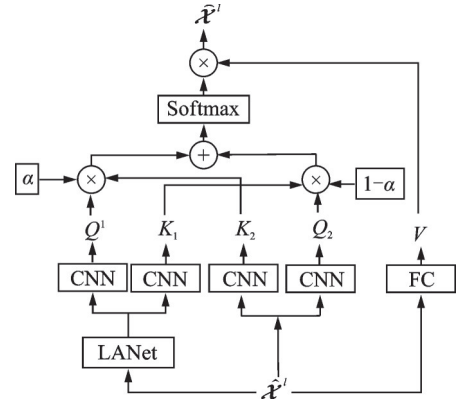


图3 多头交互注意力网络框架

Fig.3 Framework of MIA network

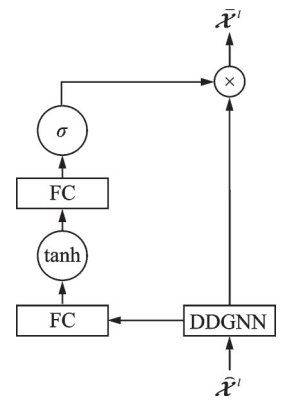


图4 自缩放动态扩散图神经网络框架

Fig.4 Framework of SDDGNN

阵和后向邻接矩阵,得到基于前向和后向邻接矩阵的扩散图神经网络,其格式为

$$\hat{\mathbf{x}}'_o = \sum_{k=0}^K (\tilde{\mathbf{A}})^k \hat{\mathbf{x}}'_o \mathbf{W}_{k1} + (\tilde{\mathbf{A}}^T)^k \hat{\mathbf{x}}'_o \mathbf{W}_{k2} \quad (9)$$

式中: $\mathbf{W}_{k1}$ 、 $\mathbf{W}_{k2}$ 为可学习的加权矩阵, $\tilde{\mathbf{A}}^T$ 为 $\tilde{\mathbf{A}}$ 的转置。

另外,传统 GNN 采用的邻接矩阵是归一化邻接矩阵 $\tilde{\mathbf{A}}$,其经特征值分解后的特征值范围为 $[-1, 1]$,其中的负特征值往往会导致训练的不稳定。为了避免上述问题,将自环和矩阵缩放^[39]引入到邻接矩阵的求解中,从而得到缩放归一化的邻接矩阵 $\hat{\mathbf{A}}$,具体为

$$\hat{\mathbf{A}} = \frac{\beta}{2} \left(\mathbf{I} + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \right) \quad (10)$$

式中: β 为归一化矩阵缩放因子,其取值范围为 $(0, 2]$,本文通过实验选取其值为1。

此外,由于天气、交通事故、早高峰、道路施工和节假日等因素影响,使得复杂交通网络对应的邻接矩阵具有动态性。为了解决该问题,将可学习的动态邻接矩阵引入到 DGNN 中,得到 DDGNN,于是得到

$$\hat{\mathbf{x}}'_o = \sum_{k=0}^K (\hat{\mathbf{A}})^k \hat{\mathbf{x}}'_o \mathbf{W}_{k1} + (\hat{\mathbf{A}}^T)^k \hat{\mathbf{x}}'_o \mathbf{W}_{k2} + (\mathbf{A}_a)^k \hat{\mathbf{x}}'_o \mathbf{W}_{k3} + (\mathbf{A}_a^T)^k \hat{\mathbf{x}}'_o \mathbf{W}_{k4} \quad (11)$$

式中: $\mathbf{A}_a = \text{softmax}(\text{ReLU}(\mathbf{E}_1 \mathbf{E}_2^T))$ 为可学习动态邻接矩阵, $\mathbf{W}_{k3}$ 、 $\mathbf{W}_{k4}$ 为可学习的加权矩阵。

最后,将自缩放网络引入到 DDGNN 中,得到最终的 SDDGNN 模型,具体为

$$\begin{aligned} G(\hat{\mathbf{x}}'_o) &= \sigma \left(\text{FC} \left(\tanh \left(\text{FC}(\hat{\mathbf{x}}'_o) \right) \right) \right) \\ \bar{\mathbf{x}}^l &= G(\hat{\mathbf{x}}'_o) \cdot \hat{\mathbf{x}}'_o \end{aligned} \quad (12)$$

式中: $\sigma(\cdot)$ 为 Sigmoid 激活函数, $G(\cdot)$ 为自缩放网络。

最后, L 个时空块输出的结果分别经过一层 CNN,然后累加,累加后的结果再经过一层 MLP 得到最终预测信号,具体公式为

$$\hat{\mathbf{y}} = \text{MLP} \left(\text{CNN}(\bar{\mathbf{x}}^1) + \text{CNN}(\bar{\mathbf{x}}^2) + \dots + \text{CNN}(\bar{\mathbf{x}}^L) \right) \quad (13)$$

式中: $\text{CNN}(\cdot)$ 为卷积算子。

3 实验结果与分析

3.1 实验设置

实验数据集:如表1所示,为了验证 TDAGNN 的有效性,以 PEMS04、PEMS08、METR-LA 和 PEMS-BAY 这4个公共交通数据集为实验对象,其中 PEMS04、PEMS08 为交通流量数据集, METR-LA 和 PEMS-BAY 为交通速度数据集。实验中对应的交通流量和交通速度预测是采用同模型分别对交通流量数据集和交通速度数据集进行学习的过程。

对比算法:实验所选用的对比方法包括 ARIMA^[4]、SVR^[8]、LSTM^[12]、DCRNN^[38]、STSGCN^[25]、GW^[23]、ASTGCN^[22]、Informer^[13]、FEDformer^[16]、DGCRN^[26]和 D2STGNN^[27],其中 ARIMA 为基于统计学方法,SVR 为基于传统机器学习方法,LSTM、Informer 和 FEDformer 为基于深度学习的方法,STSGCN、GW、ASTGCN、DGCRN 和 D2STGNN 为基于 GNN 的方法。

表1 交通数据集信息

数据集	节点数	时间点数	边数
PEMS04	307	16 992	340
PEMS08	170	17 856	295
METR-LA	207	34 272	1 515
PEMS-BAY	325	52 116	2 369

参数设置:将数据集按照6:2:2比例划分为训练集、验证集和测试集,实验皆用过去1 h的交通数据预测未来15、30和60 min的交通数据,选用的定量评价指标为平均绝对误差(Mean absolute error, MAE)、均方根误差(Root mean squared error, RMSE)和平均绝对百分比误差(Mean absolute percentage error, MAPE)。另外,通过大量实验为提出的TDAGNN方法选取合适的超级参数值提供依据。所涉及的超级参数选取情况为:多头交互注意力网络中的 $\alpha=0.2$, Epoch=100, Batch size=64, 自缩放动态扩散图神经网络的缩放因子 $\beta=1$, 权重衰减参数=0.000 1, 初始学习率=0.001, 扩散阶数 $K=2$, 映射的特征数 $H=64$ 。

3.2 实验结果与分析

针对不同数据集,分别展示了不同方法对15、30和60 min的交通预测性能指标,从量化角度验证了TDAGNN的有效性,每个定量评价指标的实验结果皆为执行10次实验取平均的结果,如表2所示。

表2 TDAGNN与其他对比模型在不同数据集上的交通预测性能对比

Table 2 Comparison of traffic forecasting performance of TDAGNN and other contrast models on different datasets

数据集	方法	15 min			30 min			60 min			平均值		
		MAE	RMSE	MAPE/ %	MAE	RMSE	MAPE/ %	MAE	RMSE	MAPE/ %	MAE	RMSE	MAPE/ %
PEMS04	ARIMA	21.98	35.21	16.52	25.38	39.21	21.03	26.67	40.74	22.43	24.68	38.39	19.99
	SVR	22.56	35.43	14.81	27.56	42.32	18.43	37.68	55.71	27.02	29.27	44.49	20.09
	LSTM	21.32	33.47	15.23	23.65	36.78	18.32	26.81	40.46	21.04	23.93	36.90	18.20
	DCRNN	20.34	31.94	13.65	23.21	36.15	15.70	24.65	38.12	17.05	22.73	35.40	15.47
	STSGCN	19.80	31.58	13.41	21.30	33.84	14.27	24.47	38.46	16.27	21.86	34.63	14.65
	GW	18.15	29.52	12.40	19.12	30.62	13.38	19.96	32.72	14.11	19.08	30.95	13.30
	ASTGCN	20.16	31.53	14.13	22.29	34.27	15.65	26.23	40.12	19.19	22.89	35.31	16.32
	Informer	17.89	28.75	12.54	18.81	30.24	14.00	20.23	32.30	14.02	18.98	30.43	13.52
	FEDformer	17.79	28.71	12.24	18.73	30.23	13.81	20.18	32.34	13.88	18.90	30.43	13.31
	DGCRN	18.27	28.97	12.47	19.39	30.86	13.42	21.09	33.59	14.94	19.58	31.14	13.61
	D2STGNN	18.43	29.42	12.89	19.63	31.19	13.79	21.64	33.87	15.30	19.90	31.49	13.99
PEMS08	TDAGNN	17.76	28.61	12.25	18.69	30.12	13.21	19.48	31.32	13.73	18.64	30.02	13.06
	ARIMA	19.56	29.78	12.45	22.35	33.43	14.43	26.27	38.86	17.38	22.73	34.02	14.75
	SVR	17.97	27.96	11.25	22.63	34.32	13.79	32.18	47.23	21.09	24.26	36.50	15.38
	LSTM	17.58	26.78	12.36	21.52	32.27	16.32	30.86	43.76	24.27	23.32	34.27	17.65
	DCRNN	16.62	25.48	10.04	17.88	27.63	11.38	22.51	34.21	14.17	19.00	29.11	11.86
	STSGCN	16.65	25.40	10.90	17.82	27.31	11.60	19.77	29.30	12.80	18.08	27.34	11.77
	GW	14.22	22.96	9.45	15.94	24.72	9.77	17.27	26.77	11.26	15.81	24.82	10.16
	ASTGCN	16.45	25.18	11.13	18.76	28.57	12.33	22.53	33.69	15.34	19.25	29.15	12.93
	Informer	14.51	22.46	9.44	15.34	24.01	9.98	16.59	26.38	10.73	15.48	24.28	10.05
	FEDformer	14.43	22.39	9.22	15.20	23.86	9.70	16.40	26.16	10.37	15.34	24.14	9.76
	DGCRN	13.89	22.07	9.19	14.92	23.99	9.85	16.73	26.88	10.84	15.18	24.31	9.96
	D2STGNN	14.29	22.43	9.90	15.42	24.40	10.61	17.37	27.33	11.81	15.69	24.72	10.77
	TDAGNN	13.78	22.01	8.75	14.48	23.44	9.45	15.66	25.61	10.02	14.64	23.68	9.41

续表

数据集	方法	15 min			30 min			60 min			平均值		
				MAPE/ %			MAPE/ %			MAPE/ %			MAPE/ %
		MAE	RMSE		MAE	RMSE		MAE	RMSE		MAE	RMSE	
METR- LA	ARIMA	3.99	8.21	9.60	5.15	10.45	12.70	6.90	13.23	17.40	5.35	10.63	13.23
	SVR	3.39	8.45	9.30	5.05	10.87	12.10	6.72	13.76	16.70	5.05	11.03	12.70
	LSTM	3.44	6.30	9.60	3.77	7.23	10.90	4.37	8.69	13.20	3.86	7.41	11.23
	DCRNN	2.77	5.38	7.30	3.15	6.45	8.80	3.60	7.60	10.50	3.17	6.48	8.87
	STSGCN	3.31	7.62	8.06	4.13	9.77	10.29	5.06	11.66	12.91	4.17	9.68	10.42
	GW	2.69	5.15	6.90	3.07	6.22	8.37	3.53	7.37	10.01	3.10	6.25	8.43
	ASTGCN	4.86	9.27	9.21	5.43	10.61	10.13	6.51	12.52	11.64	5.60	10.80	10.33
	Informer	2.73	5.24	7.16	3.14	6.32	8.54	3.63	7.44	10.14	3.17	6.33	8.61
	FEDformer	2.72	5.23	7.07	3.12	6.23	8.49	3.61	7.43	10.05	3.15	6.30	8.54
	DGCRN	2.69	5.08	6.71	2.99	6.09	8.11	3.44	7.35	9.78	3.04	6.17	8.20
	D2STGNN	2.70	5.10	6.73	3.02	6.07	8.14	3.46	7.21	9.83	3.06	6.13	8.23
	TDAGNN	2.63	5.02	6.64	2.98	5.97	8.01	3.43	7.10	9.67	3.01	6.03	8.11
PEMS- BAY	ARIMA	1.62	3.30	3.50	2.33	4.76	5.40	3.38	6.50	8.30	2.44	4.85	5.73
	SVR	1.85	3.59	3.80	2.48	5.18	5.50	3.28	7.08	8.00	2.54	5.28	5.77
	LSTM	2.05	4.19	4.80	2.20	4.55	5.20	2.37	4.96	5.70	2.21	4.57	5.23
	DCRNN	1.38	2.95	2.90	1.74	3.97	3.90	2.07	4.74	4.90	1.73	3.89	3.90
	STSGCN	1.44	3.01	3.04	1.83	4.18	4.17	2.26	5.21	5.40	1.84	4.13	4.20
	GW	1.30	2.74	2.73	1.63	3.70	3.67	1.95	4.52	4.63	1.63	3.65	3.68
	ASTGCN	1.52	3.13	3.22	2.01	4.27	4.48	2.61	5.42	6.00	2.05	4.27	4.57
	Informer	1.32	2.78	2.77	1.65	3.75	3.73	1.97	4.56	4.67	1.64	3.70	3.72
	FEDformer	1.31	2.77	2.76	1.64	3.72	3.71	1.95	4.49	4.65	1.63	3.66	3.71
	DGCRN	1.30	2.69	2.68	1.59	3.63	3.55	1.89	4.42	4.43	1.59	3.58	3.55
	D2STGNN	1.31	2.76	2.75	1.62	3.68	3.68	1.90	4.39	4.50	1.61	3.61	3.64
	TDAGNN	1.29	2.68	2.67	1.56	3.49	3.42	1.86	4.30	4.40	1.57	3.49	3.50

由表2可看出,提出的TDAGNN对交通流量数据集PEMS04和PEMS08在交通流量方面的预测都比其他经典的对比模型好。针对PEMS04交通流量的60 min预测,TDAGNN模型的MAE、RMSE和MAPE分别为19.48、31.32和13.73%,比其他对比模型最大分别提高18.2、24.39和13.29%,最小分别提高0.48、0.98和0.15%。同时,针对PEMS04交通流量的平均预测评价指标,TDAGNN模型的MAE、RMSE和MAPE分别为18.64、30.02和13.06%,比其他对比模型最大分别提高10.63、14.47和7.03%,最小分别提高0.26、0.41和0.24%。针对PEMS08交通流量数据集的60 min预测,TDAGNN模型的MAE、RMSE和MAPE分别为15.66、25.61和10.02%,比其他对比模型最大分别提高16.52、21.62和14.25%,最小分别提高0.74、0.55和0.35%。同时,针对PEMS08交通流量的平均预测评价指标,TDAGNN的MAE、RMSE和MAPE分别为14.64、23.68和9.41%,比其他对比模型最大分别提高9.62、12.82和8.24%,最小分别提高0.54、0.46和0.35%。

另外,针对交通速度数据集METR-LA和PEMS-BAY在交通速度方面的预测,提出的TDAGNN也比其他经典的对比模型好。针对60 min的METR-LA交通速度的预测,TDAGNN的MAE、RMSE和MAPE分别为3.43、7.10和9.67%,比其他对比模型最大分别提高3.47、6.66和7.73%,最小分别提高

0.01、0.11和0.11%。同时,针对METR-LA交通速度数据集,TDAGNN的平均预测MAE、RMSE和MAPE分别为3.01、6.03和8.11%,比其他对比模型最大分别提高2.59、5和5.12%,最小分别提高0.03、0.10和0.09%;针对60 min的PEMS-BAY交通速度的预测,TDAGNN模型的MAE、RMSE和MAPE分别为1.86、4.30和4.40%,比其他对比模型最大分别提高1.52、2.78和3.9%,最小分别提高0.03、0.09和0.03%。同时,针对PEMS-BAY交通速度数据集的平均预测评价指标,TDAGNN的MAE、RMSE和MAPE分别为1.57、3.49和3.5%,比其他对比模型最大分别提高0.97、1.79和2.27%,最小分别提高0.02、0.09和0.05%。

综上所述,与其他对比方法相比,提出的TDAGNN对交通流量和交通速度均具有最高的预测性能。此外,为了能够直观形象地评估提出的TDAGNN的性能,分别绘制了采用TDAGNN、GW和D2STGNN这3个交通预测模型对随机从PEMS04和PEMS08交通数据集中选取的第50个节点和第30个节点的前50 h的交通预测的真实值和预测值波形图,如图5所示,其中绿色表示真实值,蓝色表示预测值。

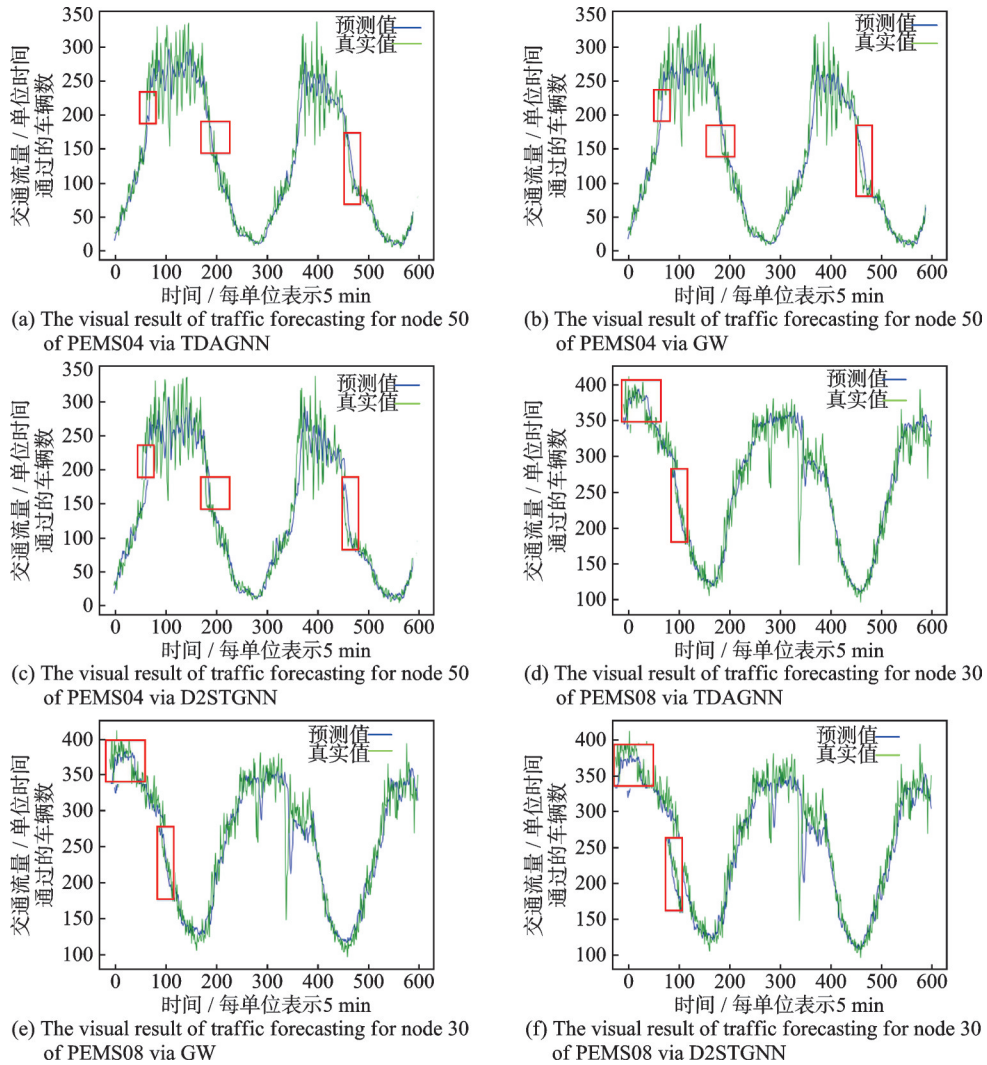


图5 3种交通预测模型对两种交通数据集不同节点的预测可视化结果比较

Fig.5 Comparison of visualization results of three traffic prediction models for different nodes in two different traffic datasets

由图 5(a~c)可看出,提出的 TDAGNN 模型总体上比其他两种模型的效果更优,尤其是图中用红色矩形框框出来的部分。例如,从图 5(a~c)中最左侧的矩形框框出的部分可看出,与 GW 和 D2STGNN 的效果图相比,TDAGNN 模型对应的两条曲线的差距最小。另外,从图 5(d~f)中最左侧的矩形框框出的部分可看出,与 GW 和 D2STGNN 的效果图相比,TDAGNN 模型对应的两条曲线基本重合,而 GW 和 D2STGNN 两个交通预测模型的预测值明显偏下。

综上所述,从视觉角度进一步验证了本文提出的 TDAGNN 交通预测模型的有效性。

3.3 消融实验

为验证提出的 TDAGNN 中每个模块的有效性,如表 3 所示,以 PEMS04、PEMS08 数据集的交通流量预测效果为例,从 9 个方面对其进行相应消融实验,具体描述如下。

TDANGNN:去掉 TDAGNN 中的 SDDGNN 模块以验证 SDDGNN 模块的有效性。

TDAGCN:采用 GCN 替换 SDDGNN 模块中的 DGNN 模块,以验证 DGNN 模块的有效性。

TDACGCN:采用 Chebyshev 图神经网络替换 SDDGNN 模块中的 DGNN 模块,以验证 DGNN 模块的有效性。

TDAGNN_NDA:去掉 SDDGNN 模块中的动态邻接矩阵模块,以验证动态邻接矩阵的有效性。

TDAGNN_NS:去掉 SDDGNN 模块中的自缩放功能,以验证自缩放功能的有效性。

TDAGNN_NMAT:去掉 TDAGNN 模型中的 MIA 模块,以验证 MIA 模块的有效性。

TDAGNN_OMAT:采用原始的多头自注意力机制替换 TDAGNN 中的 MIA 模块,以进一步验证 MIA 模块的有效性。

TDAGNN_NIMAT:将 TDAGNN 中的 MIA 模块采用非交叉模型的自注意力机制,以验证 MIA 模块的有效性。

TCNAGNN:不进行时序分解,直接采用 TCN 对交通数据在时间维上进行处理,以验证 DTDCNN 的有效性。

从表 3 可看出,当 TDAGNN 模型中去掉 SDDGNN 时,针对 PEMS04 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 9.21、12.37 和 7.72%,针对 PEMS08 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 4.13、4.49 和 2.76%,主要是因为没有整合动态邻接矩阵、自缩放功能和扩散图神经网络的 SDDGNN 的支撑,模型不能够有效地提取空间依赖信息、空间动态信息以及不能有效避免图神经网络中存在尺度丢失的问题;当分别采用 GCN 和 Chebyshev 图神经网络替换 SDDGNN 模块中的 DGNN 模块时,针对 PEMS04 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 8.25、10.75 和 7.14% 以及 7.13、9.47 和 5.33%,针对 PEMS08 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 3.69、4.24 和 2.13% 以及 3.42、3.87 和 2.44%,主要是因为 GCN 和 Chebyshev 图神经网络不能同时考虑信息在关联节点间的前后向传递,导致模型的学习能力降低;当从 SDDGNN 中去掉动态邻接矩阵模块时,针对 PEMS04 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 5.96、8.09 和 3.56%,针对 PEMS08 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 2.36、2.35 和 1.38%,主要是因为缺失动态邻接矩阵的支撑,模型不能够准确挖掘交通数据中的动态信息,从而使交通预测性能降低;当从 SDDGNN 中去掉自缩放功能时,针对 PEMS04 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 0.80、0.92 和 0.32%,针对 PEMS08 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 1.14、1.05 和 0.89%,主要是因为自缩放功能能够有效避免图神经网络中存在的尺度丢失问题,可增强模型的信息表征能力;当从提出的 TDAGNN 模型中去掉 MIA 模块以及采用原始的多头自注意力机制替换 TDAGNN 中的 MIA 模块时,针对 PEMS04 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 3.42、3.44 和 2.23% 以及 1.47、1.74 和 0.9%,针对 PEMS08 数据集预测的 MAE、RMSE 和 MAPE 分别下降了 3.4、3.49 和 3.01% 以及 1.13、1.48 和 1.70%,主要是因为 MIA 模块能够从原始特征和局部增强特征两个角度交互获取数据间的注意力系数,从而深度挖掘交通数据中的动态性信息和异质性信息,进而增强整个模型的有效性;当提出的 TDAGNN 中的 MIA 模块

表3 PEMS04和PEMS08数据集的消融实验

Table 3 Ablation studies for PEMS04 and PEMS08 datasets

数据集	方法	15 min			30 min			60 min		
		MAE	RMSE	MAPE/ %	MAE	RMSE	MAPE/ %	MAE	RMSE	MAPE/ %
PEMS04	TDAGNN	17.76	28.61	12.35	18.69	30.12	13.21	19.48	31.32	13.73
	TDANGNN	19.61	31.33	13.76	23.62	36.00	16.17	28.69	43.69	21.45
	TDAGCN	19.52	30.76	13.15	23.01	35.69	15.95	27.73	42.07	20.87
	TDACGCN	19.24	30.40	13.16	21.46	33.53	14.90	26.61	40.78	19.06
	TDAGNN_NDA	18.48	28.97	12.70	20.60	32.40	13.96	25.44	39.41	17.29
	TDAGNN_NS	17.77	28.62	12.37	19.52	31.21	13.51	20.28	32.24	14.05
	TDAGNN_NMAT	19.36	31.06	14.39	21.54	33.89	14.93	22.90	34.76	15.96
	TDAGNN_OMAT	18.15	29.08	13.04	19.97	31.76	13.98	20.95	33.06	14.63
	TDAGNN_NIMAT	18.22	29.15	12.69	19.68	31.38	13.79	20.80	32.85	14.94
	TCNAGNN	18.51	29.31	13.76	20.57	32.42	14.96	21.06	33.17	15.14
PEMS08	TDAGNN	13.78	22.01	8.75	14.48	23.44	9.45	15.66	25.61	10.02
	TDANGNN	15.25	24.20	9.95	17.15	26.68	11.03	19.79	30.10	12.78
	TDAGCN	14.96	23.34	9.40	16.79	26.16	10.35	19.35	29.85	12.15
	TDACGCN	14.74	23.06	9.33	16.28	25.58	10.39	19.08	29.48	12.46
	TDAGNN_NDA	14.45	22.61	9.31	15.88	24.90	10.04	18.02	27.96	11.4
	TDAGNN_NS	13.88	21.92	8.94	14.99	23.89	9.85	16.80	26.66	10.91
	TDAGNN_NMAT	14.81	23.02	10.45	16.33	25.34	11.54	19.06	29.10	13.03
	TDAGNN_OMAT	13.91	22.02	9.68	15.11	24.27	10.18	16.79	27.09	11.72
	TDAGNN_NIMAT	14.11	22.26	9.12	15.47	24.62	9.62	16.92	27.44	11.07
	TCNAGNN	14.58	22.65	11.47	15.76	24.76	11.31	18.03	27.99	12.06

采用非交叉的自注意力机制时,针对PEMS04数据集预测的MAE、RMSE和MAPE分别下降了1.32、1.53和1.21%,针对PEMS08数据集预测的MAE、RMSE和MAPE分别下降了1.26、1.83和1.05%,说明从多角度学习自注意力系数能够增强注意力机制的性能;最后,若模型不进行时序分解,直接采用TCN对交通数据在时间维上进行处理,针对PEMS04数据集预测的MAE、RMSE和MAPE分别下降了1.58、1.85和1.41%,针对PEMS08数据集预测的MAE、RMSE和MAPE分别下降了2.37、2.38和2.04%,说明将复杂的时间分解为多个时间模式单独处理的时序分解机制是一种有效的交通预测处理机制。

综上所述,消融实验证明了TDAGNN各个模块的有效性。

4 结束语

针对如何准确从交通数据中挖掘时空依赖信息、动态信息和空间异质信息的问题,提出了一种基于时序分解和注意力图神经网络的交通预测模型。该模型采用双分支时序分解卷积神经网络充分挖掘交通数据的时间依赖信息,采用多头交互注意力网络和自缩放动态扩散图神经网络充分挖掘交通数据中的空间依赖信息、动态信息和异质信息,从而有效地进行交通预测。实验结果表明,提出的TDAGNN对不同类型交通数据的预测效果皆最优。但是,本文涉及的工作仍然有进一步改善的地方。故未来工作中,将首先针对多头交互注意力网络在提高交通预测性能的同时也引入了较高的计算复杂

度的问题,持续改进自注意力机制来降低计算复杂度;其次,针对GNN存在的过平滑等问题,不断提出新的GNN模型,从而推动交通预测模型的空间特征学习能力;再者,考虑交通数据中可能存在的数据缺失、噪声干扰等问题,创新数据预处理技术进行数据补全和去噪;最后,将本文所涉及的方法推广并应用到金融、能源等其他时空预测任务中。

参考文献:

- [1] JIANG W, LUO J. Graph neural network for traffic forecasting: A survey[J]. *Expert Systems with Applications*, 2022, 207: 1-28.
- [2] YE J, ZHAO J, YE K, et al. How to build a graph-based deep learning architecture in traffic domain: A survey[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(5): 3904-3924.
- [3] 梅宇生,赵卓峰.基于Transformer的路网轨迹重建方法[J].*数据采集与处理*,2024,39(3): 678-688.
MEI Yusheng, ZHAO Zhuofeng. Map-constrained trajectory recovery mechanism based on Transformer[J]. *Journal of Data Acquisition and Processing*, 2024, 39(3): 678-688.
- [4] WILLIAMS B, HOEL L. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results[J]. *Journal of Transportation Engineering*, 2003, 129(6): 664-672.
- [5] KUMAR S, VANAJAKSHI L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data[J]. *European Transport Research Review*, 2015, 7(3): 1-9.
- [6] KUMAR S. Traffic flow prediction using Kalman filtering technique[J]. *Procedia Engineering*, 2017, 187: 582-587.
- [7] CHANDRA S, AL-DEEK H. Predictions of freeway traffic speeds and volumes using vector autoregressive models[J]. *Journal of Intelligent Transportation Systems*, 2009, 13(2): 53-72.
- [8] WU C, HO J, LEE D. Travel-time prediction with support vector regression[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2004, 5(4): 276-281.
- [9] MAY M, HECKER D, KORNER C, et al. A vector-geometry based spatial KNN-algorithm for traffic frequency predictions [C]//*Proceedings of IEEE International Conference on Data Mining Workshops*. Pisa: IEEE, 2008: 442-447.
- [10] JIANG W, ZHANG L. Geospatial data to images: A deep-learning framework for traffic forecasting[J]. *Tsinghua Science and Technology*, 2019, 24(1): 52-64.
- [11] LV Z, XU J, ZHENG K, et al. LC-RNN: A deep learning model for traffic speed prediction[C]//*Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. Stockholm: Morgan Kaufmann, 2018: 3470-3476.
- [12] HOCHREITER S, SCHMIDHUBER J. LoNG short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [13] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//*Proceedings of Association for the Advancement of Artificial Intelligence (AAAI)*. Virtual: AAAI, 2021: 1-14.
- [14] KITAEV N, KAISER L, LEVSKAYA A. Reformer: The efficient Transformer[C]//*Proceedings of International Conference on Learning Representations (ICLR)*. Addis Ababa: Ithaca, 2020: 1-12.
- [15] WU H, XU J, WANG J, et al. Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting [J]. *Neural Information Processing Systems*, 2021, 34: 22419-22430.
- [16] ZHOU T, MA Z, WEN Q, et al. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting [C]//*Proceedings of International Conference on Machine Learning*. Baltimore: ACM, 2022: 1-19.
- [17] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks And Learning Systems*, 2021, 32(1): 4-24.
- [18] ZHANG B, LUO S, WANG L, et al. Rethinking the expressive power of GNNs via graph biconnectivity[EB/OL]. (2024-02-11)[2024-07-28]. <https://arxiv.org/abs/2301.09505>.
- [19] KIPF T, WELING M. Semi-supervised classification with graph convolutional networks[C]//*Proceedings of International Conference on Learning Representations*. Sydney: ACM, 2017: 1-14.
- [20] ABBOUD R, CEYLAN I, GROHE M, et al. The surprising power of graph neural networks with random node initialization [C]//*Proceedings of International Joint Conference on Artificial Intelligence*. Virtual: Morgan Kaufmann, 2021: 2112-2118.
- [21] LIU X, JIN W, MA Y, et al. Elastic graph neural networks[C]//*Proceedings of International Conference on Machine*

Learning. Virtual: ACM, 2021: 6861-6871.

- [22] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting [C]//Proceedings of Association for the Advancement of Artificial Intelligence (AAAI). Honolulu: AAAI, 2019: 922-929.
- [23] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling[C]//Proceedings of International Joint Conference on Artificial Intelligence. Macao, China: Morgan Kaufmann, 2019: 1907-1913.
- [24] ZHAO L, SONG Y, ZHANG C, et al. T-GCN: A temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 99: 1-11.
- [25] SONG C, LIN Y, GUO S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]//Proceedings of Association for the Advancement of Artificial Intelligence (AAAI). New York: AAAI, 2020: 914-921.
- [26] LI F, FENG J, YAN H, et al. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution [J]. ACM Transactions on Knowledge Discovery from Data, 2023, 17(1): 1-21.
- [27] SHAO Z, ZHANG Z, WEI W, et al. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting[EB/OL]. (2022-09-05)[2024-07-28]. <https://arxiv.org/abs/2206.09112>.
- [28] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//Proceedings of International Conference on Learning Representations. Vancouver: Ithaca, 2017: 1-12.
- [29] BRODY S, ALON U, YAHAV E. How attentive are graph attention networks?[C]//Proceedings of International Conference on Learning Representations. Virtual: Ithaca, 2022: 1-26.
- [30] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[C]//Proceedings of Neural Information Processing Systems. [S.l.]: NIPS, 2017: 1-16.
- [31] SHIH S, SUN F, LEE H. Temporal pattern attention for multivariate time series forecasting[J]. Machine Learning, 2019, 108: 1421-1441.
- [32] TOUVRON H, CORD M, EL-NOUBY A, et al. Augmenting convolutional networks with attention-based aggregation[J]. Computer Vision and Pattern Recognition, 2021, 1: 1-14.
- [33] HAMILTON W, YING R, LESKOVEC J. Inductive representation learning on large graphs[J]. Advances in Neural Information Processing Systems, 2017, 30: 1-19.
- [34] KLICPERA J, BOJCHEVSKI A, GÜNNEMANN S. Predict then propagate: Graph neural networks meet personalized pagerank[C]//Proceedings of International Conference on Learning Representations (ICLR). New Orleans: Ithaca, 2019: 1-15.
- [35] LIU S, DONG H, LI L, et al. Local augmentation for graph neural networks[C]//Proceedings of International Conference on Machine Learning. Baltimore: ACM, 2022: 14054-14072.
- [36] KEFATO Z, GIRDZIJAUSKAS S, STÄRK H. Jointly learnable data augmentations for self-supervised GNNs[EB/OL]. (2021-08-23)[2024-07-28]. <https://arxiv.org/abs/2108.10420>.
- [37] ZENG A, CHEN M, ZHANG L, et al. Are Transformers effective for time series forecasting?[C]//Proceedings of Association for the Advancement of Artificial Intelligence (AAAI). Washington: AAAI, 2023: 1-15.
- [38] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[C]//Proceedings of International Conference on Learning Representations(ICLR). BC: Ithaca, 2018: 1-16.
- [39] SONG X, LIAN J, HUANG H, et al. Friend recommendations with self-rescaling graph neural networks[C]//Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). Washington: ACM, 2022: 1-11.

作者简介:



杨永鹏(1984-),男,博士研究生,研究方向:图神经网络及其应用等, E-mail: 2021010212@njupt.edu.cn。



杨震(1961-),通信作者,男,博士,教授,博士生导师,研究方向:图信号处理及其应用等, E-mail: yangz@njupt.edu.cn。



杨真真(1984-),女,博士,副教授,研究生导师,研究方向:图神经网络及其应用等, E-mail: yangzz@njupt.edu.cn。