



计算机科学

COMPUTER SCIENCE

基于多时间尺度时空图网络的交通流量预测模型

汪鸣, 彭舰, 黄飞虎

引用本文

汪鸣, 彭舰, 黄飞虎. [基于多时间尺度时空图网络的交通流量预测模型](#)[J]. 计算机科学, 2022, 49(8): 40-48.

WANG Ming, PENG Jian, HUANG Fei-hu. [Multi-time Scale Spatial-Temporal Graph Neural Network for Traffic Flow Prediction](#)[J]. Computer Science, 2022, 49(8): 40-48.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于注意力机制的医学影像深度哈希检索算法](#)

Deep Hash Retrieval Algorithm for Medical Images Based on Attention Mechanism

计算机科学, 2022, 49(8): 113-119. <https://doi.org/10.11896/jsjcx.210700153>

[基于非局部注意力生成对抗网络的视频异常事件检测方法](#)

Non-local Attention Based Generative Adversarial Network for Video Abnormal Event Detection

计算机科学, 2022, 49(8): 172-177. <https://doi.org/10.11896/jsjcx.210600061>

[基于双图神经网络信息融合的文本分类方法](#)

Text Classification Method Based on Information Fusion of Dual-graph Neural Network

计算机科学, 2022, 49(8): 230-236. <https://doi.org/10.11896/jsjcx.210600042>

[基于改进位置编码的谣言检测模型](#)

Rumor Detection Model Based on Improved Position Embedding

计算机科学, 2022, 49(8): 330-335. <https://doi.org/10.11896/jsjcx.210600046>

[全局信息引导的真实图像风格迁移](#)

Photorealistic Style Transfer Guided by Global Information

计算机科学, 2022, 49(7): 100-105. <https://doi.org/10.11896/jsjcx.210600036>

基于多时间尺度时空图网络的交通流量预测模型

汪 鸣 彭 舰 黄飞虎

四川大学计算机学院 成都 610041

(15680831992@qq.com)

摘 要 交通流预测在智能交通系统的建设中起着关键作用。但由于其复杂的时空依赖性和本身的不确定性使得研究变得极具挑战性。现有的一些方法主要是将单一的时间序列输入到循环神经网络以捕获时间依赖性,而且多数模型仅对时间模块和空间模块进行简单的堆叠,导致不能有效地融合时间和空间特征。为了解决以上问题,文中提出了一个多时间尺度时空图网络模型。模型先将序列数据划分为3种时间尺度序列,然后将序列输入到时空块(ST-Block)中提取数据的时空依赖性,最后进行预测。在时空块中使用图卷积网络和变体Transformer分别捕获数据中的时间和空间依赖性,并通过门控融合机制将两者提取到的特征进行融合。在两个真实的数据集上分别进行了短期和长期的预测实验,结果表明了MTSTGNN模型在交通流预测任务上的优秀性能。

关键词: 交通流预测;图卷积网络;注意力机制;Transformer;门控融合机制

中图法分类号 TP183

Multi-time Scale Spatial-Temporal Graph Neural Network for Traffic Flow Prediction

WANG Ming, PENG Jian and HUANG Fei-hu

College of Computer Science, Sichuan University, Chengdu 610041, China

Abstract Traffic flow prediction plays a key role in intelligent transportation system. However, due to its complex spatial-temporal dependence and its uncertainty, research becomes extremely challenging. Some existing methods mainly use a single time series and input it into a recurrent neural network to capture temporal dependency. Moreover, most models only simply stack temporal modules and spatial modules, resulting in ineffective feature fusion. To address these issues, this paper proposes a multi-time scale spatial-temporal graph neural network model. The model divides the sequence data into three time-scale sequences, then puts sequence data into the ST-Blocks to extract the spatial-temporal features of data, and finally makes the prediction. In ST-Block, graph convolutional network and variant Transformer are used to capture spatial dependency and temporal dependency respectively, and the output feature of the two sub-modules are fused through a gate mechanism. A large number of experiments are conducted on two real data sets in this paper, including short-term and long-term prediction, and the results show the excellent performance of MTSTGNN model on the task of traffic flow prediction.

Keywords Traffic flow prediction, Graph convolutional network, Attention mechanism, Transformer, Gate mechanism

1 引言

随着中国城市化进程的加快,人们对出行的需求日益增加,交通拥堵问题日益严重,制约着中国的城市发展。目前亟需建立一套智能交通系统^[1]来帮助人们合理规划交通路线以及缓解交通拥堵。而智能交通系统的核心在于准确预测未来的交通流状况,从而辅助系统中的任务。

目前交通流预测模型研究的主要问题是如何对交通数据进行特征提取。特征主要包括空间特征和时间特征,空间

特征包括交通路段的特点以及各路段之间的相关性,时间特征包括数据在时间维度上的周期性与趋势性。已有的研究通常使用基于统计理论的预测模型,如自回归差分滑动平均模型^[2]和卡尔曼滤波分析^[3]模型,以及基于机器学习方法的预测模型,如K近邻算法^[4]、支持向量机^[5]等。但这些方法难以处理交通流数据的时间和空间相结合的依赖关系,其中的时空相关性未能得到充分挖掘,交通流预测任务仍有很大的提升空间。

随着深度学习和神经网络的发展,对数据的时空依赖

到稿日期:2022-01-19 返修日期:2022-02-23

基金项目:国家重点研发计划(2020YFB0704502);四川省重点研发计划(2020YFG0089,2020YFG0308,2020YFG0304,22ZDYF3599)

This work was supported by the National Key R & D Program of China(2020YFB0704502) and Key R & D Program of Sichuan Province, China(2020YFG0089,2020YFG0308,2020YFG0304,22ZDYF3599).

通信作者:彭舰(jianpeng@scu.edu.cn)

关系建模方式愈发成熟,Fu等^[6]结合长短期记忆网络(Long Short Term Memory,LSTM)与门控循环单元(Gated Recurrent Unit,GRU)两种循环神经网络来预测交通流。然而这样的时序模型只适合捕获交通数据在时间维度上的特征,难以捕获空间维度上的特征。Liu等^[7]结合了卷积神经网络(Convolution Neural Networks,CNNs)与LSTM,提出了一种ConvLSTM模型用于提取交通流的时空特征。随着图神经网络的兴起,研究者将欧氏空间上的卷积操作迁移到了非欧氏空间,提出了图卷积神经网络GCN。Li等^[8]提出的DCRNN模型利用扩散图卷积对空间特征进行建模,并以基于GRU的编码解码结构对时间维度进行建模,但由于循环神经网络每一步的预测都依赖于之前时刻的预测结果,因此模型的效率十分低下。Yu等^[9]提出了STGCN,该算法用谱图卷积来捕获节点间的空间依赖关系,并利用一维扩张卷积对时间特征进行建模。随着注意力机制的发展,Xu等^[10]提出了STTN,该算法使用Transformer中的多头注意力机制分别对时间和空间相关性进行建模。可以看出,同时考虑数据的时间和空间相关性是研究的必然趋势。

大部分现有的交通预测模型均使用循环神经网络(Recurrent Neural Networks,RNNs)及其变体来建模时间序列数据中的特征。然而这种时间建模方式具有既定的序列性,并且网络每一步的预测都依赖于之前时刻的预测结果,一旦序列较长,就很难捕捉到其隐藏的周期和趋势关系。此外多数模型仅对时间和空间提取模块进行了简单的堆叠,并不能有效地融合空间特征和时间特征^[11]。

针对上述不足,本文从序列以及模型融合的角度对交通流预测模型进行改进,提出了一个多时间尺度时空图网络模型。该模型首先将输入划分为3种不同的时间尺度,再融合成一个新的合成时间序列,然后利用图卷积网络和变体Transformer分别捕获交通流数据的空间和时间特征,并引入门控融合机制模块有效结合时空模块。在网络构建方法上,将空间模块、时间模块和门控模块组合成一个时空块,通过叠加时空块的方式深化网络的层数,并利用残差网络来解决网络中的退化问题。本文的贡献如下:1)提出了一种多时间尺度建模方法,其目的是捕获时间序列中的长短期依赖关系,并且它对长期预测是有效的;2)提出了一种新的时空图建模方式,通过门控融合有效结合图卷积和变体Transformer,形成时空块,同时捕捉空间依赖性和时间依赖性。

2 相关理论知识

2.1 图卷积网络

卷积神经网络是以卷积操作为基础的前馈神经网络,可以高效处理图像、语音这类欧氏数据。但是实际生活中很多数据都是非欧氏数据,这类数据并不能直接使用卷积网络进行特征提取,因此提出了图卷积网络^[12-13](Graph Convolution Networks,GCNs)。GCNs是一种将卷积操作扩展到

图结构的神经网络,核心思想是中心节点对其邻域节点进行信息聚合。表1列出了进行图卷积操作涉及的一些符号定义。

表1 图相关定义
Table 1 Definition of graph

Symbol	Meaning
G	Graph
E	Edges of graph G
V	Vertices of graph G
A	Adjacency matrix of G
D	Degree matrix of adjacency matrix A
L	Laplacian matrix of graph G
I	An identity matrix
$H^{(l)}$	The Hidden state of layer l

其中,邻接矩阵反映图中节点之间的相邻关系,度矩阵反映与每个节点相连的边数量。最能反映图结构性质的图的拉普拉斯矩阵,相当于数学分析中的离散拉普拉斯算子,该矩阵是图研究领域中的核心研究对象,描述的是中心节点与邻居节点信号的差异,可以反映节点邻域的平滑性,具体定义为:

$$L = D - A \quad (1)$$

通常,正则化的拉普拉斯矩阵更为常用,定义为:

$$L^{\text{sym}} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2} \quad (2)$$

GCN模型的原理是利用卷积本质上是频域上的滤波这一特性,在频域上进行操作^[14]。将节点特征看成信号,通过傅里叶变换到频域空间,最后通过傅里叶逆变换得到图域卷积。每一层卷积仅处理一阶邻域信息,通过叠加若干卷积层可以实现多阶邻域的信息传递^[15]。每一个卷积层的传播规则如下:

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)}) \quad (3)$$

其中, $\sigma(\cdot)$ 表示非线性激活函数。简而言之,GCN的每一层通过邻接矩阵和特征矩阵相乘得到每个顶点邻居特征的汇总,然后再乘上一个参数矩阵,最后用激活函数得到聚合邻居节点特征的矩阵。

2.2 注意力机制

在序列相关的任务中,注意力机制最早被应用于机器翻译领域^[16],发展至今已为深度学习体系中的常用概念之一。注意力机制根据依赖对象选取的不同,可分为一般注意力机制与自注意力(Self Attention)机制。一般注意力机制刻画输出特征与各输入特征间的依赖关系,关联程度越强的输入特征权重值就越大。自注意力机制在输入序列内部进行权重计算,用于刻画输入特征间的依赖关系。自注意力机制对每个位置的输入特征计算其对应的查询向量与键值向量。利用这3个矩阵,计算注意力评分矩阵并进行归一化处理,得到特征权重系数。注意力机制与已有深度模型相结合,能够更精确有效地挖掘特征间的隐含关系,提升深度学习模型的预测性能。

谷歌团队的Vaswani等^[17]于2017年利用自注意力机制提出了Transformer模型,并在自注意力基础上提出了多头注意力机制,如图1所示。

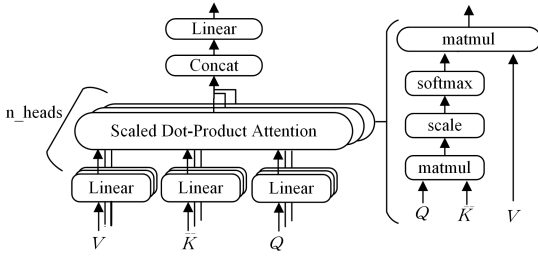


图1 多头注意力计算流程

Fig. 1 Calculation process of multi-head attention

由于注意力机制对位置不敏感,因此 Transformer 添加了位置编码(Positional Encoding)。通过使用不同频率的正弦、余弦函数生成正余弦位置编码,将编码与对应的词向量相加,从而使序列带有位置信息。如下式所示:

$$PE(pos, 2i) = \sin(pos/1000^{2i/d_{model}}) \quad (4)$$

$$PE(pos, 2i+1) = \cos(pos/1000^{2i/d_{model}}) \quad (5)$$

其中, pos 表示输入向量在整个序列中的绝对位置, d_{model} 表示向量的维度,通过奇偶性来区分。

Transformer 是非常有潜力的模型,在此基础上已经发展出了各种 X-former,其中最具有代表性的当属 Informer^[18] 以及 Autoformer^[19] 这类优秀的序列预测模型。在交通流预测方面,越来越多的研究者根据 Transformer 的注意力机制提取时间特征,这些研究中也有一部分会同时用注意力提取空间特征。例如,GMAN 模型^[20] 和 MRA-BGCN 模型^[21] 就使用了注意力模型来提取交通流数据中的时间和空间特征,而 ST-Transformer 模型^[10] 和 ST-GRAT 模型^[22] 在 Transformer 的基础上进行改进,提出了合适的交通流预测模型。

2.3 交通流预测任务

交通流预测任务是一个典型的时间序列预测任务,本文的研究只涉及交通流数据,因此只需保留交通状态数据中的一个属性,即 $\mathbf{X} \in \mathbf{R}^{N \times T \times 1}$ 。交通流预测问题是根据给定历史观测数据 $\mathbf{X} = \{x_{t-h+1}, x_{t-h+2}, \dots, x_t\} \in \mathbf{R}^{N \times H \times 1}$ 和对应时刻的路网结构 G (其中 H 为历史序列输入的窗口大小),来预测未来 T_p 时段所有节点的交通流数据 $\mathbf{Y} = \{y_{t+1}, y_{t+2}, \dots, y_{t+T_p}\} \in \mathbf{R}^{N \times T_p}$ (其中 $y_{t+T_i} = \{y_{t+T_i}^1, y_{t+T_i}^2, \dots, y_{t+T_i}^N\} \in \mathbf{R}^N$ 表示在未来 T_i 时间段中整个交通网络的流量状态)。简而言之,交通预测问题的本质是学习一个函数 F ,根据历史观测数据和对应的路网结构 G ,预测未来时段 T_p 内所有节点的交通流数据 Y ,具体公式如下:

$$(y_{t+1}, y_{t+2}, \dots, y_{t+T_p}) = F[\{x_{t-h+1}, x_{t-h+2}, \dots, x_t\}, G] \quad (6)$$

3 多时间尺度时空图网络模型

3.1 总体模型

多时间尺度时空图网络模型的总体架构如图 2 所示。该模型由 3 部分组成,依次为多尺度划分模块、时空特征提取模块以及预测输出模块。首先,多尺度划分模块将时间序列划分为若干类尺度的数据,使得节点的时间序列周期特征更加丰富;其次,通过时空特征提取模块中的图卷积网络和自注意力机制来提取复杂的时空特征;最后将所提取到的特征通过两层全连接层预测输出。

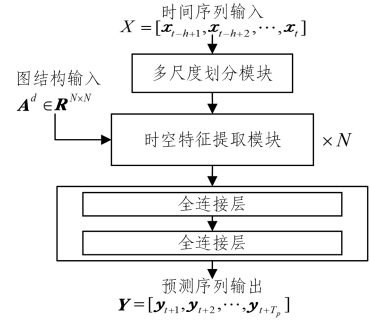


图2 整体模型架构

Fig. 2 Overall architecture of the model

3.2 多尺度划分模块

时间序列数据在宏观上具有一定周期和趋势。假设当前时刻为 t ,时间相关性为变量在 t 时刻与过去 n 个时刻的关系,即 $x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-n})$,也被称作自相关。由于 f 函数很难表示,本文使用深度学习来学习。我们认为不仅仅是 x_{t-1} 到 x_{t-n} 与 x_t 相关,而且 $x^{(2)}, x^{(3)}$ 等不同时间尺度的数据都与 x_t 相关。因此,为了更好地挖掘时间依赖关系,我们将输入的数据进行多尺度的划分,并将这些多尺度的数据融合起来,构成一个增强的时间序列,如图 3 所示,并将这个时间序列用作整个模型的输入。

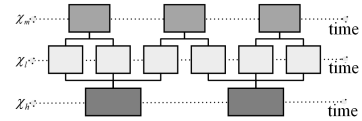


图3 多时间尺度划分示意图

Fig. 3 Schematic diagram of multi-time scale division

假设原始流量数据采样时间间隔为 t min,并且历史观测序列窗口为 T_w ,并且满足式(7):

$$T_w = q \times t \quad (7)$$

其中 q 为正整数。如图 2 所示,本文将原序列处理成 3 种不同尺度的序列,分别为高尺度、中尺度和低尺度,长度分别为 T_h, T_m, T_l 。计算式如下:

$$T_h = (q/3) \times (3 \times t) \quad (8)$$

$$T_m = (q/2) \times (2 \times t) \quad (9)$$

其中, T_h 和 T_m 的序列长度分别为 $q/3$ 和 $q/2$, T_l 与原序列长度一致即为 T_w 。设当前时间点为 t_0 ,原始历史观测序列大小为 T_w ,3 种时间尺度的序列各自体现的特性如下。

(1)低尺度段。即为原数据序列,其序列表示为 $\mathbf{X}_l = \{x_{t_0-T_l+1}^{(1)}, x_{t_0-T_l+2}^{(1)}, \dots, x_{t_0}^{(1)}\}$, $x^{(1)}$ 表示数据的低尺度单元。低尺度序列数据可以反映数据的局部变化趋势,对预测序列的前几个点有一定的影响。简而言之,它对短期预测的效果很好。

(2)中尺度段。中尺度段和低尺度段数据部分相同,但是每个单元所代表的时间长度不同。序列表示为 $\mathbf{X}_m = \{x_{t_0-T_m+1}^{(2)}, x_{t_0-T_m+2}^{(2)}, \dots, x_{t_0}^{(2)}\}$, $x_i^{(2)} = (x_i^{(1)} + x_{i+1}^{(1)})/2$ 为数据的中尺度单元。中尺度序列数据比低尺度数据更平滑,更能反映数据变化的整体趋势,更有利于中长期预测。

(3)高尺度段。与中尺度段一样,高尺度段和低尺度段属于同一段数据。其序列表示为 $\mathbf{X}_h = \{x_{t_0-T_h+1}^{(3)}, x_{t_0-T_h+2}^{(3)}, \dots,$

$\mathbf{x}_{t_0}^{(3)}$, 且 $\mathbf{x}_i^{(3)} = (\mathbf{x}_i^{(1)} + \mathbf{x}_{i+1}^{(1)} + \mathbf{x}_{i+2}^{(1)})/3$ 表示数据的高尺度单元。与中尺度序列数据一样,高尺度序列数据对预测也有类似的影响,都是为了反映数据变化的长期趋势。

为了更好地适用于模型,本文将3种尺度数据串联起来形成一个合成时间序列^[23],表示为 $\mathbf{X} = \{\mathbf{X}_l, \mathbf{X}_m, \mathbf{X}_h\}$ 。

3.3 时空特征提取模块

交通流量数据属于严格的时空数据,时间和空间相互纠缠,互相影响。单独捕获时空相关性再进行融合通常无法取得良好的效果。本文构建了一个时空特征提取块(ST-Block)来联合捕获图结构时间序列的时空相关性。更具体地说,一个时空特征提取块包含一个空间特征提取子模块和一个时间特征提取子模块,分别用于捕获空间和时间相关性,之后再利用门控融合机制来结合这两个模块,通过堆叠模块达到时空特征融合的效果。

叠加过多的神经网络模块会使模型的拟合效果变差。因此,在每一个网络层的输入和输出之间添加一个恒等映射,即不对原始输入做任何改变,对输出做加和,其表达式如下:

$$H(\mathbf{x}) = F(\mathbf{x}) + \mathbf{x} \quad (10)$$

其中, \mathbf{x} 表示残差块的输入, $H(\mathbf{x})$ 表示残差块的输出。在进行前向传播时,输入可以从任意低层直接传播到高层,这种残差网络很好地解决了深度神经网络的退化问题,同时加快了

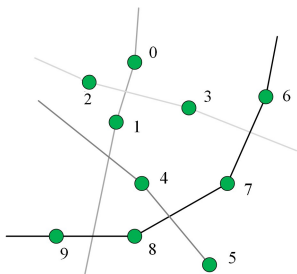


图5 路网中各节点的流量特征相似程度

Fig. 5 Similarity of traffic characteristics of each node in road network

从图5中可以看出,相通的节点之间的相关系数往往更大,反映出的空间相关性也越强。因此,使用邻接矩阵 \mathbf{A} 来体现,矩阵中元素的具体计算式如下:

$$A_{ij} = \begin{cases} (0,1), & \text{if } v_i \text{ and } v_j \text{ connected} \\ 0, & \text{others} \end{cases} \quad (11)$$

其中, v_i 代表节点 i , v_j 代表节点 j , $(0,1)$ 表示值的大小在 $0 \sim 1$ 区间, A_{ij} 代表两者之间的关联程度。

目前主流的空间特征提取方法是 GCN, 本文采用两层 GCN 模块和一层前馈神经网络模块来捕获交通数据流之间的空间依赖性。使用的 GCN 是基于谱图理论的图卷积,通过分析图的拉普拉斯矩阵的特征值和特征向量来研究图的性质。以在 t 时刻的交通网络中的交通流数据为例,整个图的信号可以表示为 $\mathbf{x} \in \mathbf{R}^{N \times C}$, 并且卷积核定义为 $g_\theta(\mathbf{L})$, 其定义在傅里叶域中卷积的公式为:

$$g_\theta(\mathbf{L}) * \mathbf{x} = \mathbf{U} g_\theta(\mathbf{U}^T \mathbf{x}) \quad (12)$$

其中, θ 为模型的参数, \mathbf{U} 为归一化拉普拉斯矩阵的特征向量, $\mathbf{U}^T \mathbf{x}$ 表示对 \mathbf{x} 的傅里叶变换。利用切比雪夫多项式的近似展开来求解这个公式,相当于通过卷积核 g_θ 提取图中每个节点的 0 到 $K-1$ 阶邻域的信息。因此最后的卷积公式为:

参数的优化收敛速度。本文在该模块的子模块中也使用了残差网络的构建方法。

通过以上若干神经网络的拼接,最终时空特征提取模块的结构如图4所示。

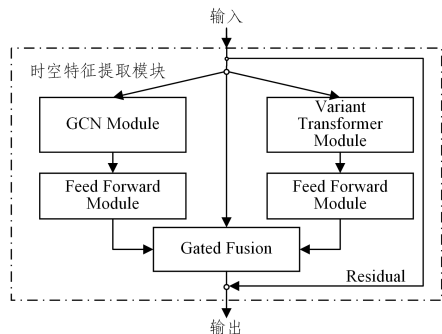
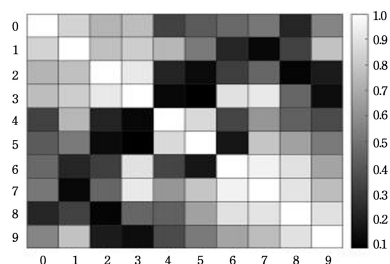


图4 时空特征提取块结构图

Fig. 4 Structure of ST-Block

3.3.1 空间特征提取子模块

空间特征的引入是因为考虑到节点之间的流量受路网连通的影响。路网结构即地图上节点(传感器)通过公路连接而抽象出来的图,相通即有边,有边则具有相关性。通过使用皮尔逊相关指数进行统计,部分节点之间的流量相似程度如图5所示。



$$g_\theta(\mathbf{A})\mathbf{x} = \sum_{k=0}^{K-1} \theta_k T_k(\hat{\mathbf{A}})\mathbf{x} \quad (13)$$

其中, $\hat{\mathbf{A}}$ 为递归公式 $T_k(\mathbf{x}) = 2T_{k-1}(\mathbf{x}) - T_{k-2}(\mathbf{x})$ 的最大特征值,其中 $T_0 = 1, T_1 = \mathbf{x}$ 。将这样的图卷积核作用在交通图上的每一个节点上,即可捕获到图中的一阶邻居信息。本文的图卷积模块由两层卷积层和一层 Add&Norm 共同组成,如式(14)和图6所示。

$$f(\mathbf{X}, \mathbf{A}) = \sigma(\hat{\mathbf{A}} \text{ReLU}(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}_0)\mathbf{W}_1) \quad (14)$$

其中, \mathbf{X} 代表特征矩阵, $\hat{\mathbf{A}}$ 为包含自连接的正则化拉普拉斯矩阵, $\sigma(\cdot)$ 和 $\text{ReLU}(\cdot)$ 为激活函数。

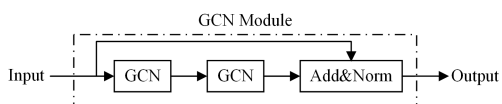


图6 图卷积模块结构

Fig. 6 Structure of graph convolutional module

为了提高模型的泛化能力,在图卷积模块之后添加了一个前馈神经网络模块。使用线性层将向量维度扩大,经过若干操作后,再用线性层将向量维度降回至节点数量的维度 N ,最后通过 ReLU 进行非线性激活。最终,空间模块的前馈

神经网络模块具体表示为:

$$FFN(\mathbf{x}_{in}) = \max(0, \mathbf{x}_{in}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \quad (15)$$

其中, $\mathbf{x}_{in} \in \mathbf{R}^{N \times T}$ 为输入矩阵。

3.3.2 时间特征提取子模块

在时间维度上,设计了一个变体 Transformer 网络,以此来有效捕捉时间序列中短期和长期的时间依赖性。其结构如图 7 所示。

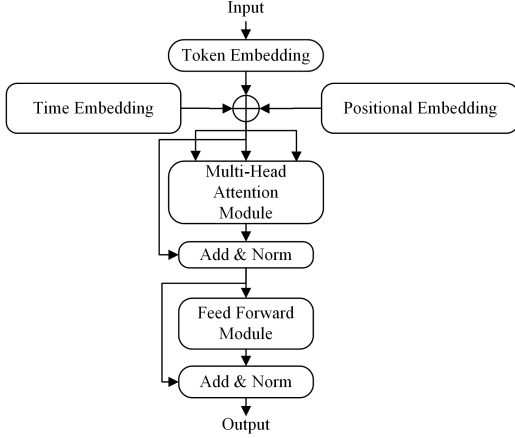


图 7 变体 Transformer 模块图

Fig. 7 Diagram of variant Transformer module

此模块由位置和时间嵌入模块与多头注意力模块两个部分组成。

与 Transformer 同理,在此模块中,本文对序列使用相对位置嵌入方法,生成一个位置矩阵 \mathbf{D}^p ,并将其整合到节点特征中。此外,本文还将时间嵌入为节点的特征。与相对位置编码类似,此模块对采样的绝对时间进行编码,将月、日、星期、小时和分钟分别编码为 $\mathbf{D}^m, \mathbf{D}^d, \mathbf{D}^w, \mathbf{D}^h, \mathbf{D}^m$ 。将它们相加成为最终的绝对时间嵌入表示,表达式如下:

$$\mathbf{D}^r = \mathbf{D}^m + \mathbf{D}^d + \mathbf{D}^w + \mathbf{D}^h + \mathbf{D}^m \quad (16)$$

其中, \mathbf{D}^r 为最终的绝对时间嵌入。联合之前的流量序列数据的编码 \mathbf{X} 和相对位置编码 \mathbf{D}^p ,将 3 种编码相加,得到一个含有位置信息的矩阵 \mathbf{X}^r 。至此,将位置和时间嵌入模块的输出作为下一个模块的输入,表达式如下:

$$\mathbf{X}^r = \mathbf{X} + \mathbf{D}^p + \mathbf{D}^r \quad (17)$$

对于多头注意力模块,多头注意力机制最初被应用在自然语言处理中,用于学习词向量与词向量之间的关系特征。对于交通流数据,在时间维度上,可以将一个时间片的空间的所有信息组成一个词向量,从而学习不同时间片之间的相关性;在形式上,将输入特征矩阵按照 $T \times N \times C$ 的形式输入,通过前馈神经网络映射到具有可学习的潜在子空间中。多头注意力将输入分成同源的 3 个部分,分别为 *Query*, *Key*, *Value*。将它们进行线性变换,表达式如下:

$$\mathbf{Q} = \mathbf{X}^r \mathbf{W}^q, \mathbf{K} = \mathbf{X}^r \mathbf{W}^k, \mathbf{V} = \mathbf{X}^r \mathbf{W}^v \quad (18)$$

其中, $\mathbf{W}^q \in \mathbf{R}^{NC \times d_q^*}$, $\mathbf{W}^k \in \mathbf{R}^{NC \times d_k^*}$, $\mathbf{W}^v \in \mathbf{R}^{NC \times d_v^*}$ 为 3 个线性变换,在自注意力中, $d_q^* = d_k^* = d_v^*$ 。然后将 3 个矩阵输入放缩点积注意力模块中,得到自注意力矩阵。具体公式为:

$$Atten(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k^*}}\right)\mathbf{V} \quad (19)$$

多头注意力即进行多次以上的操作,每一次只算一头,而且每次变换时参数 \mathbf{W} 均不同。整个多头注意力的表达式为:

$$\mathbf{h}_i = Attention(\mathbf{Q}\mathbf{W}_i^q, \mathbf{K}\mathbf{W}_i^k, \mathbf{V}\mathbf{W}_i^v) \quad (20)$$

$$\mathbf{O} = MultiH(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Concat(\mathbf{h}_1, \dots, \mathbf{h}_n)\mathbf{W}_o \quad (21)$$

其中, \mathbf{O} 为多头注意力输出矩阵, n 为注意力的头数, \mathbf{h}_i 表示每次进行注意力运算的结果。

3.3.3 门控融合机制模块

直接将交通流数据通过空间特征提取模块和时间特征提取模块的交错重复模块并不能有效捕获交通流的时空交融性。因此,本节使用了一个门控融合机制来结合时间和空间模块,使模型可以更有效地学习数据中的时空交融性。

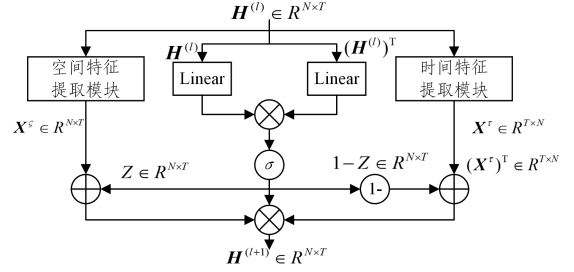


图 8 门控融合机制模块示意图

Fig. 8 Diagram of gate fusion mechanism module

与门控循环神经网络单元类似,门控融合机制可以控制神经网络中信息的传递。如图 8 所示,在时空模块的融合中,第一层的输入为上一层的输出 ($\mathbf{H}^{(l)} = \mathbf{X}$),然后依次叠加。首先输入矩阵 $\mathbf{H}^{(l)}$ 分别进入空间特征模块、时间特征模块以及两个 Linear 层。需要注意的是,进行空间和时间特征提取时,由于考虑的维度不同,需要将时间模块和时间模块对应的线性层的输入进行转置。从空间和时间两个模块输出的结果分别和权重矩阵进行点乘,然后将两者相加进行最后的输出。该步骤的计算式如下:

$$\mathbf{H}^{(l+1)} = z \circ \mathbf{X}^s + (1-z) \circ (\mathbf{X}^r)^T \quad (22)$$

其中, z 为用来控制门控融合中两者的比例。由于时间和空间处理的维度不同,这里将 $\mathbf{H}^{(l)}$ 复制两份分别从两个维度进行 Linear 运算,两个 Linear 层分别为 $Linear(N, N)$ 和 $Linear(T, T)$,接着通过激活函数进行非线性激活,最后进行按位相加得出门控融合权重 z ,具体表达式如下:

$$z = \sigma(\text{Linear}(\mathbf{H}^{(l)}) + \text{Linear}((\mathbf{H}^{(l)})^T)) \quad (23)$$

3.4 全连接预测模块

经过若干层的时空特征提取模块后,得到了一个最终的隐藏表示 $\mathbf{H}^{(l+1)}$ 。最后,将这个最终的隐藏表示通过一个两层的全连接前馈神经网络模块进行预测,表达式如下:

$$\hat{\mathbf{Y}} = \text{ReLU}(\mathbf{H}^{(l+1)}\mathbf{W}_0)\mathbf{W}_1 \quad (24)$$

其中, $\mathbf{H}^{(l+1)}$ 为最后一层的特征隐藏表示, \mathbf{W}_0 和 \mathbf{W}_1 为可训练的参数矩阵。

4 实验

4.1 数据集

本文利用两个真实的高速公路数据集 PeMSD4 和 PeMSD8 进行实验,数据集中的部分路网结构和传感器分布如图 9 所示。

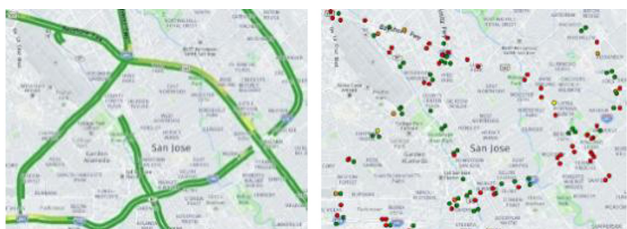


图9 公路结构以及该路网中的传感器分布

Fig. 9 Road structure and sensor distribution in road network

使用其中的车流量数据作为实验数据集,使用采集频率为 30 秒/次的原始数据汇总成时间间隔为 5 min 的数据样本,即每个点一个小时包含 12 个交通数据样本。使用历史 12 个样本作为输入对未来若干时间片的交通流状态进行预测。

本文按照 7:1:2 的比例将数据集划分成为训练集、验证集和测试集,并且利用训练集的均值和方差对划分后的 3 个数据集进行归一化处理,数据的详细信息如表 2 所列。

表 2 数据集描述

Table 2 Description of data sets

数据集	传感器数量	时间范围	样本间隔/min	训练样本数	验证样本数	测试样本数
PeMSD4	307	2018/01/01—2018/02/28	5	11 877	1 696	3 392
PeMSD8	170	2018/07/01—2018/08/31	5	12 482	1 783	3 566

4.2 对比模型和评价指标

实验通过两个回归任务常用的指标来评估模型的性能,分别为平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Squared Error, RMSE),这两个指标的评估效果较为客观,具体的计算式为:

$$MAE = \frac{1}{m} \sum_{i=1}^m |Y_i - \hat{Y}_i| \quad (25)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_i - \hat{Y}_i)^2} \quad (26)$$

本文选择结合了 MSE 和 MAE 的 Huber 作为损失函数,通过调节超参数来确定 MAE 和 MSE 的作用范围,具体的计算式为:

$$L_{\delta}(Y, \hat{Y}) = \begin{cases} \frac{1}{2} (Y - \hat{Y})^2, & \text{if } |Y - \hat{Y}| \leq \delta \\ \delta |Y - \hat{Y}| - \frac{1}{2} \delta^2, & \text{others} \end{cases} \quad (27)$$

为了验证本文提出的模型的有效性,设置了 6 种基准方法进行实验对比。

(1)HA:历史平均模型。HA 模型直接使用历史平均值作为未来时间点的预测结果。

(2)ARIMA:自回归差分滑动平均模型。ARIMA 模型是应用较为广泛的时间序列分析模型,其对单点的时间序列预测具有良好的效果。

(3)SVR:支持向量回归模型。SVR 是基于机器学习的预测方法,通过支持向量机对输入的序列数据进行回归预测,是单节点预测任务中的经典基准算法。

(4)GRU:门控循环单元网络。GRU 是一种特殊的循环

神经网络模型,实验中使用一层 GRU 和一层全连接的模型组合对网络中未来时间片的数据进行预测。

(5)STGCN:时空图卷积网络。STGCN 将 GCN 融入到一维卷积单元中,在预测过程中加入空间的影响,并用一层全连接的模型进行预测。

(6)DCRNN:扩散图卷积网络。DCRNN 利用扩散图卷积和 Seq2Seq 分别编码空间和时间信息,同时捕获数据的时间和空间关系,并依次解码出预测信息。

以上所选择的基准算法中,HA, ARIMA, SVR 和 GRU 是经典的时间序列预测算法,STGCN,DCRNN 是目前效果很好的深度模型预测算法,它们对时空预测任务均有较高的准确性。

4.3 实验及结果分析

表 3 和表 4 分别列出了本文模型和基准模型在 PeMSD4 和 PeMSD8 数据集上预测 15 min 和 60 min 时交通流量预测的误差表现。

表 3 PeMSD4 上各个模型的误差表现

Table 3 Error performance of our model and baseline on PeMSD4

Data	Method	15 min		60 min	
		MAE	RMSE	MAE	RMSE
PeMSD4	HA	28.22	41.93	38.29	56.76
	ARIMA	24.36	43.47	44.31	74.89
	SVR	22.43	33.36	30.14	43.95
PeMSD4	GRU	23.66	34.16	31.70	50.82
	STGCN	23.36	36.04	28.55	42.53
	DCRNN	21.22	32.45	26.54	40.07
	MTSTGNN	20.91	31.27	25.24	36.32

表 4 PeMSD8 上各个模型的误差表现

Table 4 Error performance of our model and baseline on PeMSD8

Data	Method	15 min		60 min	
		MAE	RMSE	MAE	RMSE
PeMSD8	HA	23.08	34.16	31.97	47.49
	ARIMA	18.25	29.58	33.87	60.15
	SVR	17.41	26.52	22.97	35.91
PeMSD8	GRU	19.87	26.34	24.87	44.40
	STGCN	17.83	25.53	21.18	31.47
	DCRNN	18.05	26.58	19.69	29.69
	MTSTGNN	16.11	24.51	18.12	27.95

可以看出,就两个评估指标而言,本节提出的模型在两个数据集上都接近最佳表现。本节将从 3 个方面来分析模型 Multi-Time scale Spatial-Temporal Graph Neural Network (MTSTGNN) 的优势。

(1)低误差优势。我们发现,一些基于神经网络并且充分考虑时间特征建模重要性的模型,如 MTSTGNN,DCRNN,STGCN,在预测精度方面普遍高于 HA,ARIMA 这类基于统计理论的预测模型。以数据集 PeMSD4 上对于未来 15 min 的预测任务为例,DCRNN 和 MTSTGNN 的均方误差比 ARIMA 的均方误差分别低了 11.2% 和 20.7%。主要原因在于,ARIMA 作为一个比较成熟的时间序列预测模型,其难以处理长时以及非平稳的序列数据,并且在计算误差时使用的是总体误差的一个平均值,如果某个节点数据波动较大,则会使最终的预测效果变差。

(2)同时考虑时间和空间关系时的预测优势。HA 作为

一个空间和时间均没有考虑的模型,其对交通流数据的预测误差高于其他模型,这说明,对时间和空间相关性的提取对于建模是非常有必要的。除此之外,STGCN,DCRNN 和本文提出的 MTSTGNN 都是时空数据预测的模型,均对时间和空间进行了建模并用于预测,平均绝对误差和均方误差大都低于 ARIMA,SVR 和 GRU 这类单纯的时间序列预测算法。这说明,在交通流预测任务下,正确建模空间关系有利于预测。

(3)模型对长时预测的优势。为了能够体现模型的长时预测能力的优势,本文绘制了根据预测步长的变化以及预测精度变化的趋势图,如图 10、图 11 所示。可以看出,没有考虑时空关系的模型 HA 与只考虑时间相关性的传统的时间

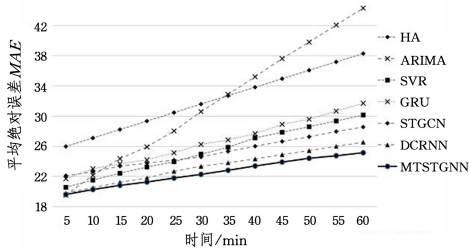


图 10 不同方法在 PeMSD4 上的误差变化趋势

Fig. 10 MAE trend of different methods on PeMSD4

为了能直观地说明时空相关性在交通预测方面的重要性,以及体现 MTSTGNN 的预测效果,在实验测试环节,本文选择了两个模型,分别为 HA 和 MTSTGNN,并在 PeMSD4 上随机选择了两个路段,对这两个路段提前一小时的交通流量预测情况进行可视化,对比预测和真实值之间的差距,如图 12 和图 13 所示。其中灰色虚线、灰色实线和黑色实线分别代表真实车流量数据、HA 预测结果和 MTSTGNN 预测结

果。从图中可以看出,MTSTGNN 对真实数据的拟合效果明显强于 HA,而且 HA 模型具有明显的滞后性。这说明考虑数据的时空相关性能够准确且及时地进行预测。另外,虽然 MTSTGNN 能够有效预测出未来的流量情况和趋势,但却不能有效预测某些突变的因素,例如传感器 1 所在路段在 12 点左右实际流量急剧下降,而预测的值却只是小幅度抖动。这在之后的研究中是可以改进的。

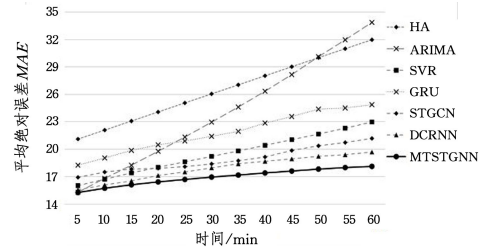


图 11 不同方法在 PeMSD8 上的误差变化趋势

Fig. 11 MAE trend of different methods on PeMSD8

果。从图中可以看出,MTSTGNN 对真实数据的拟合效果明显强于 HA,而且 HA 模型具有明显的滞后性。这说明考虑数据的时空相关性能够准确且及时地进行预测。另外,虽然 MTSTGNN 能够有效预测出未来的流量情况和趋势,但却不能有效预测某些突变的因素,例如传感器 1 所在路段在 12 点左右实际流量急剧下降,而预测的值却只是小幅度抖动。这在之后的研究中是可以改进的。

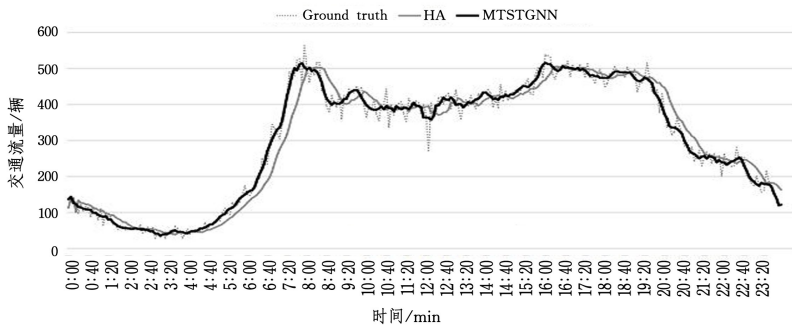


图 12 在传感器 1 上预测值与真实值的对比

Fig. 12 Comparison of prediction and ground truth on sensor 1

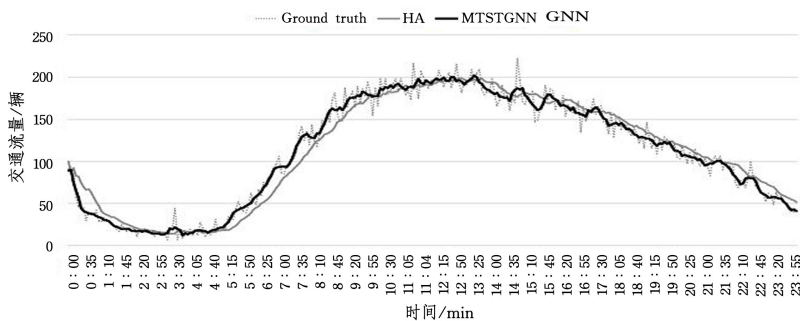


图 13 在传感器 2 上预测值与真实值的对比

Fig. 13 Comparison of prediction and ground truth on sensor 2

此外,本文对模型进行了消融实验,操作方法类似于控制变量法,目的是证明模型当中的各个子模块对于预测的提升能力。分别去除多时间尺度划分模块、门控融合机制模块、空间特征提取模块,其中去除门控融合模块直接改为堆叠两种子模块,去除空间特征提取模块改为只有时间特征提取模块。分别在 PeMSD4 和 PeMSD8 数据集上观察两种变体模型和本文提出的模型在 1-12 个时间步长上预测交通流量平均绝对误差(MAE)情况,结果如图 14 和 15 所示。不难发现,本文提出的模型比其余变体模型的预测效果更好(其中 NM 表示去除多时间尺度划分模块,NF 表示去除门控融合模块,NG 表示只有时间特征提取模块)。

具体来看,在 PeMSD8 数据集上,在去除多时间尺度划分模块后,预测前 3 个时间步长时,预测效果与本文模型不相上下,但随着预测步数的增加,本文模型效果逐渐明显。这说明,多尺度划分模块有利于预测步数较长的任务,而对于较短步数的预测,这个模块就显得不太重要。同时,在去除门控融合模块之后,此时时空块中的时间特征提取模块和空间提取模块变成了堆叠形式(串行)。去除门控机制后,说明模型不能有效地将两种模块的特征进行融合,使得模型效果无论是对于短期还是长期预测任务都比本文模型效果差。



图 14 PeMSD4 上的消融实验

Fig. 14 Ablation experiments on PeMSD4

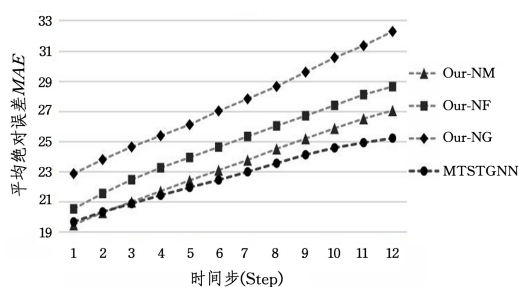


图 15 PeMSD8 上的消融实验

Fig. 15 Ablation experiments on PeMSD8

消融实验充分表明,多时间尺度划分模块和门控融合机制模块的引入能够有效提高模型的预测能力;而不考虑空间因素对整个交通网络未来交通流量的影响则不利于对整个网络的流量预测,再次说明了空间因素对预测的重要性。综上,本文提出的交通流预测模型是真实有效的。

结束语 为了更好地挖掘交通道路网络上交通流量数据复杂的时空相关性,准确预测未来交通流量数据,本文设计了一个

基于多时间尺度的时空图网络模型用于交通流预测工作。该方法首先通对时间序列进行多尺度划分,又将划分后的数据通过图卷积网络和多头注意力模型分别构建空间模块和时间模块,使用门控融合机制结合两个模块并组成时空块,通过堆叠时空块来捕获交通流数据的时空依赖性,最终在两个真实路网数据集上的实验结果表明,模型在 MAE 和 RMSE 两个指标值上均优于其他基线算法;然后分析了模型预测的局限性;最后通过消融实验证明了所提模型中的两个子模块对于提升模型预测性能的有效性,从多个角度说明了空间关系对交通流预测的重要性,证明了本文模型具有良好性能。未来的研究工作主要集中在以下两个方面:1)考虑将影响交通流的不确定因素引入模型中;2)考虑使用注意力机制结合多尺度数据,进一步提升模型的预测精度。

参考文献

- [1] XU H B, DENG H J. Application of artificial intelligence technology in urban intelligent transportation system [J]. Computer Products and Circulation, 2020(1): 167.
- [2] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process [J]. Dissertation Abstracts International, 1999, 60(1): 1-5.
- [3] OKUTANI I, STEPHANEDES Y J. Dynamic prediction of traffic volume through Kalman filtering theory [J]. Transportation Research Part B Methodological, 1984, 18(1): 1-11.
- [4] LINT H V, HINSBERGEN C V. Short-Term Traffic and Travel Time Prediction Models [J]. Transportation Research E-Circular, 2012, 45(7): 635-641.
- [5] DRUCKER H, BURGESS C J C, KAUFMAN L, et al. Support Vector Regression Machines [J]. Advances in Neural Information Processing Systems, 1997, 28(7): 779-784.
- [6] FU R, ZHANG Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction [C] // 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). 2016: 324-328.
- [7] LIU Y, ZHENG H, FENG X, et al. Short-term traffic flow prediction with Conv-LSTM [C] // 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP). 2017: 1-6.
- [8] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting [J]. arXiv: 1707. 01926, 2017.
- [9] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting [J]. arXiv: 1709. 04875, 2017.
- [10] XU M, DAI W, LIU C, et al. Spatial-temporal transformer networks for traffic flow forecasting [J]. arXiv: 2001. 02908, 2020.
- [11] ZHOU M. Research on short-term traffic flow prediction method based on road network spatiotemporal information [D]. Beijing: North China University of Technology, 2020.

- [12] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[C]// International Conference on Machine Learning, 2016; 2014-2023.
- [13] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4-24.
- [14] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[J]. arXiv:1312. 6203, 2013.
- [15] DEFFERRARD M, BRESSON X, VAND P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in Neural Information Processing Systems, 2016, 29: 3844-3852.
- [16] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409. 0473, 2014.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems, 2017; 5998-6008.
- [18] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]// Proceedings of AAAI, 2021.
- [19] WU H, XU J, WANG J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting [J]. Advances in Neural Information Processing Systems, 2021, 34: 22419-22430.
- [20] ZHENG C, FAN X, WANG C, et al. Gman: A graph multi-attention network for traffic prediction[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020; 1234-1241.
- [21] CHEN W, CHEN L, XIE Y, et al. Multi-range attentive bicomponent graph convolutional network for traffic forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020; 3529-3536.
- [22] PARK C, LEE C, BAHNG H, et al. ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed [C] // Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020; 1215-1224.
- [23] FAWAZ H I, FORESTIER G, WEBER J, et al. Data augmentation using synthetic data for time series classification with deep residual networks[J]. arXiv:1808. 02455, 2018.



WANG Ming, born in 1997, postgraduate, is a student member of China Computer Federation. His main research interests include graph neural network and time-series prediction.



PENG Jian, born in 1970, Ph.D, professor, Ph.D supervisor, is an outstanding member of China Computer Federation. His main research interests include big data and wireless sensor network.

(责任编辑:何杨)