

用于交通流预测的多分辨率时空注意力网络

赖俊龙

摘要

交通流预测是智能交通系统中的研究热点，一个高效的交通流预测模型能够有效地改善和缓解城市的交通问题。然而，由于交通路网拓扑结构的约束和交通流随时间动态变化的规律，即：空间依赖性和时间依赖性，交通流预测成为一个有挑战性的问题。为了解决该问题，提高预测精准度，本文首先复现了 ASTGCN 模型，然后在此基础上提出了多分辨率时空注意力网络，它是一种新的深度学习框架。具体来说，首先，在图卷积操作的基础上，引入了注意力机制，以便学习到交通流的动态空间依赖关系；其次，使用循环神经网络中的 GRU (Gated Recurrent Unit) 来学习交通流中的时间特征；最后，将经过上述处理后的近期交通数据与日交通数据、周交通数据进行融合，从而准确捕捉交通流的日周期性和周周期性，进而预测未来的交通状况。在多项交通预测任务上的实验证明，本文的方法与复现方法相比，在 MAPE (mean absolute percentage error)、MAE (mean absolute error) 和 RMSE (root mean squared error) 方面的性能分别提升了 1.235%、6.725% 和 6.200%。此外，在消融实验上的结果表明，本文使用的注意力机制和精心设计的多分辨率模块是有效的。

关键词：交通流预测；时空相关性；图卷积网络；注意力机制；循环神经网络

1 引言

近年来，随着城市的快速发展以及大量交通传感器的部署，城市的车辆及其带来的交通数据爆炸增长，使得我们进入交通大数据时代。智能交通系统 (Intelligent Transportation System, ITS) 应运而生，交通流预测是智能交通系统中的重要组成成分，能够有效地改善和缓解城市的交通问题。一个准确、有效和快速的交通流预测模型能够为人们的生活带来便利，例如：当发生交通事故时，有关部门能够及时地疏通道路；在日常出行中，导航能够根据未来交通流预测情况规划一些较为通畅的道路供出行人员选择等。

然而，交通流量预测具有挑战性。首先，它具有时间依赖性。在时间维度上存在波动性和不确定性，未来的交通流量会受到最近一段时间的交通流量（如图1所示）、天气、节假日等因素影响，而交通流量在较长时期内往往具有周期性，上个月、上周或昨天该时段的交通流量会有很大的概率与未来要预测的该时段的交通流量存在相似的“时间模式”。其次，它具有空间依赖性。道路与车辆在空间维度上的复杂关系也发挥重要作用，例如，在图1中，由 A_1 道路上的事故引起的交通流量的时间演变会影响到 A_2 、 A_3 、 A_4 处的道路。此外，由于道路交叉口或车道的复杂性，道路之间的相互作用不易辨别，导致交通流难以预测。

虽然现有的交通流量预测已存在很多有效的方法，然而如何更好地捕捉交通预测问题的时空相关性，进一步提高预测性能仍是一个值得研究的问题。为了解决这些问题，本文提出了一种新颖的深度学习模型：多分辨率时空注意力网络 (Multi-Resolution Spatio-Temporal Attention Network, MRSTAN)，它可以自适应地捕获交通网络的动态空间关系，也能够考虑到该网络的时空相关性。本文的主要贡献总结如下：

- 本文复现了上海交通大学 2019 年发表于 AAAI 会议的 ASTGCN 模型，该模型使用注意力机制对交通流的时空特征进行学习。
- 本文设计了一种预测交通流的新框架，具体而言，将注意力图卷积和循环神经网络应用到时空特征的学习中，并考虑交通流的相似性特征，以更好地挖掘交通流的特征。
- 在现实世界的高速公路交通数据集上进行的实验表明，本文的模型与 ASTGCN 相比取得了最佳的预测性能。

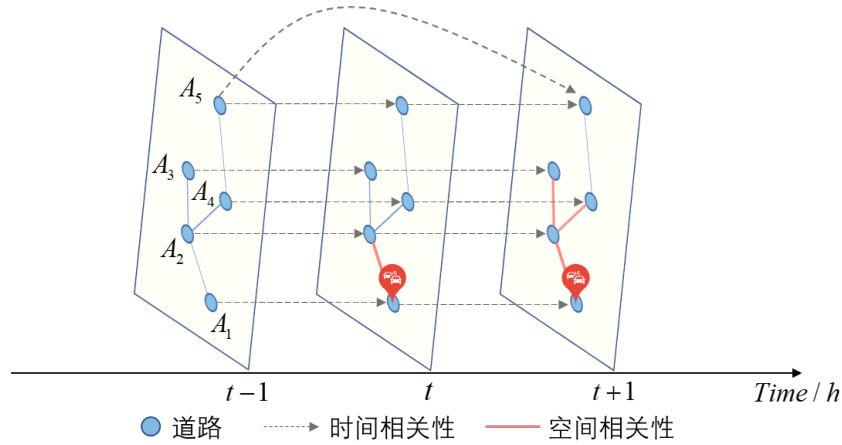


图 1: 交通流时空相关性示意图

2 相关工作

自 1970 年代研究人员首次在高速公路应用交通流量预测模型以来，该领域引起了越来越多的研究关注，并开发了许多新兴技术来提高预测性能。一般来说，现有的交通流量预测研究分为三类，本节详细介绍和分析。

2.1 传统的时间序列预测方法

传统的基于数理统计的方法往往是针对多个交通道路中的某个观测点进行单独预测的^[1]。比较典型的方法有卡尔曼滤波模型^[2]、ARIMA (Autoregressive Integrated Moving Average Mode)^[3]等。

这些方法具有复杂的数学过程，优点是能够较好地处理平稳的时间序列、具有较为直观的可解释性，但缺点是难以处理复杂的、具有较强不确定性的数据。因此，在实际应用中，由于交通流的时间序列受所多种因素影响而有较大的波动，这些模型的准确率、鲁棒性较差。

2.2 传统的机器学习预测方法

为了克服传统的预测方法性能不佳的问题，如 K 近邻 (K-NearestNeighbor, KNN) 模型^[4]、支持向量机 (Support Vector Regression, SVR) 模型^[5]和人工神经网络 (Artificial Neural Network, ANN)^[6]等一系列方法被提出。这些方法能够对高维和非线性特征的数据进行建模，并已在各种任务中取得不错的效果。但这些模型在捕获交通数据中复杂的时空相关性方面存在局限性，特别是在空间域中。

2.3 深度学习的预测方法

深度学习已经在各大领域取得了极大的关注，并在多种任务中取得了最先进的效果。为了获取空间特征，研究人员将卷积神经网络 (Convolutional Neural Networks, CNN) 引入到交通预测任务中。Jiang et al 提出将交通流网络视为图像并使用 CNN 来捕获交通流的空间特征^[7]；Cheng et al 提出将 CNN 技术与循环神经网络 (Recurrent Neural Network, RNN) 技术结合来获取交通数据的时空相关性，

从而预测交通阻塞状况^[8]。CNN 依赖于规则的拓扑结构，如：图像等，然而，交通网络是非欧几里得图结构。

为了克服 CNN 的不足，图卷积神经网络（Graph Convolutional Network, GCN）被提出。Diao et al 基于动态拉普拉斯矩阵设计了一种新颖的图卷积，结合门控 CNN 提出了 DGCNN（Dynamic Spatial-Temporal Graph Convolutional Neural Networks）模型^[9]；Zhao et al（2019）将 GCN 与 RNN 结合提出 T-GCN（Temporal Graph Convolutional Network）模型^[10]；Ali et al 基于 GCN 考虑了节点信息和边信息，提出了 TEST-GCN（Topologically Enhanced Spatial-Temporal Graph Convolutional Networks）模型^[11]等。

注意力机制（Attention Mechanism）在自然语言处理中获得巨大成功，许多研究者也将其引入到交通预测中。Guo et al 为了改进 GCN 捕获的是静态的空间邻接矩阵的缺点，引入了注意力机制，并综合考虑了日数据、周数据，建立了 ASTGCN（Attention Based Spatial-Temporal Graph Convolutional Networks）模型^[12]；Pan et al 基于注意力，考虑了地理、节点、边信息，以学习空间特征，并通过 RNN 获取时间信息，建立了 ST-MetaNet（Spatio-Temporal Meta Networks）模型^[13]；Zheng et al（2020）在空间和时间层面都采用注意力机制，并采用编码器-解码器架构，构建了 GMAN（Graph Multi-Attention Network）模型^[14]。以图作为输入，这些基于图神经网络的模型在交通道路车流量和速度预测等任务上表现出优于先前方法的效果。

在上述研究的基础上，考虑交通网络的图结构和交通数据的动态时空模式，同时采用图卷积和注意力机制对网络结构的交通数据进行建模。

3 复现的方法

此复现模型中提出了一种新的时空注意机制，以捕获交通网络上的动态时空相关性，同时考虑了交通流的不同尺度下的数据，包括日数据和周数据。模型的整体框架如下图2。一共包含三个组成部分，每个部分共享相同的网络结构，每一个都由多个时空块和一个全连通层组成。在每个时空块中都有一个时空注意模块和一个时空卷积模块。为了优化训练效率，每个组件中采用了残余学习框架。最后，根据参数矩阵将三个分量的输出进一步合并，得到最终的预测结果。该网络的整体结构是精心设计的，以描述交通流的动态时空相关性。

3.1 空间注意力

在空间维度上，不同位置的交通状况相互影响，相互影响具有很强的动态性。在这里，使用一种注意机制来自适应捕获空间维度中节点之间的动态相关性。以最近组件中的空间注意力为例：

$$\mathbf{S} = \mathbf{V}_s \cdot \sigma \left(\left(\mathcal{X}_h^{(r-1)} \mathbf{W}_1 \right) \mathbf{W}_2 \left(\mathbf{W}_3 \mathcal{X}_h^{(r-1)} \right)^T + \mathbf{b}_s \right) \quad (1)$$

$$\mathbf{S}'_{i,j} = \frac{\exp(\mathbf{S}_{i,j})}{\sum_{j=1}^N \exp(\mathbf{S}_{i,j})} \quad (2)$$

其中 $\mathcal{X}_h^{(r-1)} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{T_{r-1}}) \in \mathbb{R}^{N \times C_{r-1} \times T_{r-1}}$ 是第 r 层时空块的输入，第 r 层输入数据的通道数是 C_{r-1} ，当 $r = 1$ 时， $C_0 = F$ ，最近分量 $T_0 = T_h$ （日周期成分中 $T_0 = T_d$ ，周周期成分 $T_0 = T_w$ ）。 T_{r-1} 是第 r 层中的时间维度的长度。 $\mathbf{V}_s, \mathbf{b}_s \in \mathbb{R}^{N \times N}$, $\mathbf{W}_1 \in \mathbb{R}^{T_{r-1}}$, $\mathbf{W}_2 \in \mathbb{R}^{C_{r-1} \times T_{r-1}}$, $\mathbf{W}_3 \in \mathbb{R}^{C_{r-1}}$ 是可学习的参数。根据该层的当前输入动态计算注意矩阵 \mathbf{S} 。 \mathbf{S} 中的元素 $\mathbf{S}_{i,j}$ 的值在语义上表示节点 i 和节点 j 的相关强度，然后使用一个 softmax 函数来保证节点的注意权重和等于 1。在进行图卷积时，将

邻接矩阵 A 与空间注意矩阵 $\mathbf{S}' \in \mathbb{R}^{N \times N}$ 一起，动态调整节点之间的影响权重。

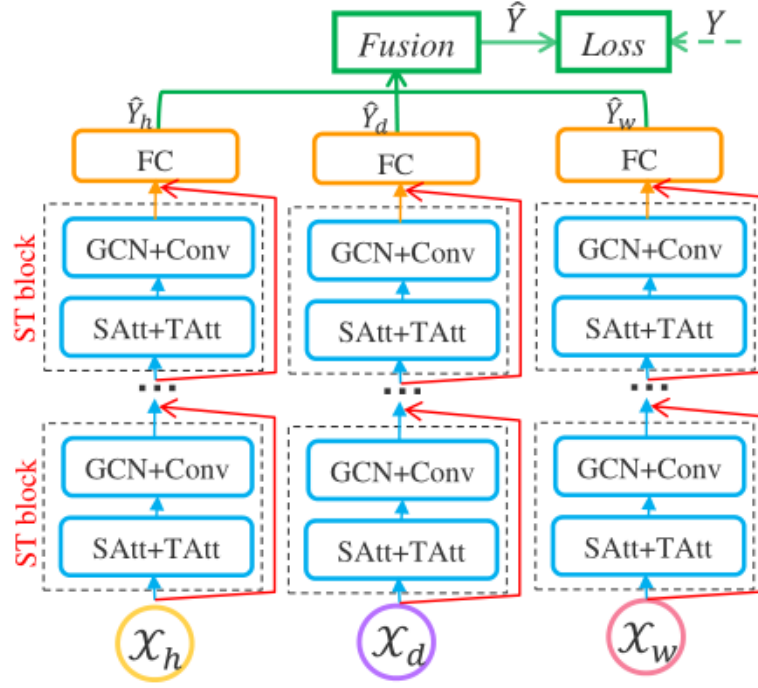


图 2: ASTGCN 的框架

3.2 时间注意力

在时间维度上，不同时间段的交通状况之间存在相关性，且在不同情况下相关性也不同。同样，使用一种注意机制来自适应地赋予数据不同的重要性：

$$\mathbf{E} = \mathbf{V}_e \cdot \sigma \left(\left(\left(\mathcal{X}_h^{(r-1)} \right)^T \mathbf{U}_1 \right) \mathbf{U}_2 \left(\mathbf{U}_3 \mathcal{X}_h^{(r-1)} \right) + \mathbf{b}_e \right) \quad (3)$$

$$\mathbf{E}'_{i,j} = \frac{\exp(\mathbf{E}_{i,j})}{\sum_{j=1}^{T_{r-1}} \exp(\mathbf{E}_{i,j})} \quad (4)$$

其中 $\mathbf{V}_e, \mathbf{b}_e \in \mathbb{R}^{T_{r-1} \times T_{r-1}}, \mathbf{U}_1 \in \mathbb{R}^N, \mathbf{U}_2 \in \mathbb{R}^{C_{r-1} \times N}, \mathbf{U}_3 \in \mathbb{R}^{C_{r-1}}$ 是可学习的参数。时间相关矩阵 E 由不同的输入决定。 E 中的元素 $\mathbf{E}_{i,j}$ 的值在语义上表示时间 i 和 j 之间的依赖强度。最后用 softmax 函数对 E 进行归一化处理，直接将归一化的时间注意矩阵应用于输入，得到 $\hat{\mathbf{X}}_h^{(r-1)} = (\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \dots, \hat{\mathbf{X}}_{T_{r-1}}) = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{T_{r-1}}) \mathbf{E}' \in \mathbb{R}^{N \times C_{r-1} \times T_{r-1}}$ 通过合并相关信息来动态调整输入。

3.3 时空图卷积

时空注意模块使网络自动对有价值的信息给予相对较多的注意，通过注意机制调整的输入输入到时空卷积模块中。复现的论文提出的时空卷积模块包括一个在空间维度上的图卷积，从邻域获取空间依赖性，以及一个沿时间维度的卷积，利用邻近时间的的时间依赖性。

空间维的图卷积：谱图理论将基于网格的数据的卷积运算推广到图结构数据。在本研究中，交通网络本质上是一个图结构，每个节点的特征可以看作是图上的信号。因此，为了充分利用交通网络的拓扑特性，复现的论文中在每个时间片上采用基于谱图理论的图卷积来直接处理信号，在空间维度上利用交通网络上的信号相关性。谱法将图转化为代数形式，分析图的拓扑属性，如图结构中的连通性。

在谱图分析中，用对应的拉普拉斯矩阵来表示图。通过分析拉普拉斯矩阵及其特征值，可以得到图结构的性质。定义了图的拉普拉斯矩阵为 $\mathbf{L} = \mathbf{D} - \mathbf{A}$ ，它的标准化形式是 $\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{N \times N}$ ，

其中 \mathbf{A} 为邻矩阵, \mathbf{I}_N 为单位矩阵, 度矩阵 $\mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{N \times N}$ 为对角矩阵, 由节点度组成, $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ 。拉普拉斯矩阵的特征值分解为 $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}$, 其中 $\mathbf{\Lambda} = \text{diag}([\lambda_0, \dots, \lambda_{N-1}]) \in \mathbb{R}^{N \times N}$ 是对角矩阵, \mathbf{U} 是傅里叶基。以 t 时刻的交通流为例, 整个图上的信号为 $x = \mathbf{x}_t^f \in \mathbb{R}^N$, 信号的图傅里叶变换定义为: $\hat{x} = \mathbf{U}^T x$ 。根据拉普拉斯矩阵的性质, \mathbf{U} 是一个正交矩阵, 因此对应的傅里叶反变换为 $x = \mathbf{U}\hat{x}$ 。图卷积是一种卷积运算, 通过使用在傅里叶域中对角化的线性算子来取代经典的卷积算子。在此基础上, 用核 g_θ 对图 \mathbb{R} 上的信号 x 进行滤波:

$$g_\theta *_G x = g_\theta(\mathbf{L})x = g_\theta(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)x = \mathbf{U}g_\theta(\mathbf{\Lambda})\mathbf{U}^T x \quad (5)$$

其中, $*_G$ 表示图卷积操作。由于图信号的卷积运算等于这些经过图傅里叶变换后进入谱域的信号乘积, 所以上面的公式可以理解为分别将 g_θ 和 x 进行傅里叶变换进入谱域, 然后将它们的变换结果相乘, 再进行傅里叶反变换得到卷积运算的最终结果。然而, 当图的尺度较大时, 直接对拉普拉斯矩阵进行特征值分解代价较大。因此, 采用 Chebyshev 多项式近似但有效地解决了该问题:

$$g_\theta *_G x = g_\theta(\mathbf{L})x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}})x \quad (6)$$

其中参数 $\theta \in \mathbb{R}^K$ 是一个多项式系数的向量, $\tilde{\mathbf{L}} = \frac{2}{\lambda_{\max}}\mathbf{L} - \mathbf{I}_N$, λ_{\max} 为拉普拉斯矩阵的最大特征值。切比雪夫多项式的递归定义是 $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, 其中 $T_0(x) = 1, T_1(x) = x$ 。用切比雪夫多项式的近似展开解该公式相当于用卷积核 g_θ 提取图中每个节点为中心的周围 0 到 $K-1$ 阶邻域的信息。图卷积模块使用整流线性单元 (ReLU) 作为最终激活函数, 即 $\text{ReLU}(g_\theta *_G x)$ 。

为了动态调整节点之间的相关性, 对于 Chebyshev 多项式的每一项, s 使用空间注意矩阵 $\mathbf{S}' \in \mathbb{R}^{N \times N}$ 来伴随 $T_k(\tilde{\mathbf{L}})$, 然后得到 $T_k(\tilde{\mathbf{L}}) \odot \mathbf{S}'$, 其中 \odot 为 Hadamard 乘积。因此, 上图卷积公式变为 $g_\theta *_G x = g_\theta(\mathbf{L})x = \sum_{k=0}^{K-1} \theta_k (T_k(\tilde{\mathbf{L}}) \odot \mathbf{S}')x$ 。

将这个定义推广到具有多通道的图信号。例如, 在最近的组件中, 输入为: $\hat{\mathbf{x}}_h^{(r-1)} = (\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_{T_{r-1}}) \in \mathbb{R}^{N \times C_{r-1} \times T_{r-1}}$, 其中每个节点的特征具有 C_{r-1} 通道。对于每个时间片 t , 在图形 $\hat{\mathbf{x}}_t$ 上执行 C_r 滤波器, 得到 $g_\theta *_G \hat{\mathbf{x}}_t$, 其中 $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_{C_r}) \in \mathbb{R}^{K \times C_{r-1} \times C_r}$ 为卷积核参数。因此, 每个节点都由该节点的 0 到 $K-1$ 邻居的信息更新。

时间维的卷积: 图卷积操作在空间维度上捕获图上每个节点的相邻信息后, 再在时间维度上叠加一个标准卷积层, 通过合并相邻时间片上的信息来更新节点的信号, 同样以最近组件中第 r 层的操作为例:

$$\mathcal{X}_h^{(r)} = \text{ReLU} \left(\Phi * \left(\text{ReLU} \left(g_\theta *_G \hat{\mathbf{x}}_h^{(r-1)} \right) \right) \right) \in \mathbb{R}^{C_r \times N \times T_r} \quad (7)$$

其中, $*$ 表示标准卷积操作, Φ 为时间维卷积核的参数, 激活函数为 ReLU。

综上所述, 时空卷积模块能够很好地捕捉交通数据的时空特征。时空注意模块和时空卷积模块构成时空块。将多个时空块进行叠加, 进一步提取更大范围的动态时空相关性。最后, 添加一个完全连通层, 以确保每个构件的输出与预测目标具有相同的维数和形状。最后一个全连接层使用 ReLU 作为激活函数。

3.4 多组件的融合

在本节中，将讨论如何集成这三个组件的输出。一些区域的交通流量在上午有明显的高峰期，因此日时段和周时段分量的输出更为关键。然而，在其他一些地方没有明显的交通周期模式，因此日周期和周周期组件可能是无助的。因此，当融合不同组件的输出时，三个组件对每个节点的影响权重是不同的，需要从历史数据中学习。所以融合后的最终预测结果为：

$$\hat{\mathbf{Y}} = \mathbf{W}_h \odot \hat{\mathbf{Y}}_h + \mathbf{W}_d \odot \hat{\mathbf{Y}}_d + \mathbf{W}_w \odot \hat{\mathbf{Y}}_w \quad (8)$$

其中 \odot 是阿达玛乘积。 $\mathbf{W}_h, \mathbf{W}_d$ 和 \mathbf{W}_w 为学习参数，反映了三个时间维度分量对预测目标的影响程度。

4 本文方法

ASTGCN 具有很多优点：通过注意力图卷积，邻接矩阵通过样本自动学习，而不再使用静态的交通拓扑结构；堆叠多个 ST block 可获得全局与局部特征；考虑了交通流的日数据、周数据，能够增强预测的稳定性。但在考虑时间相关性时没有考虑到时间的先后顺序，所以本文在 ASTGCN 的基础上，使用循环神经网络捕获交通流的时间特征，同时考虑日数据、周数据与最新交通流的相似性，而不是从三种交通流的重要性建模。本文提出了一种新颖的深度学习模型：多分辨率时空注意力网络 (Multi-Resolution Spatio-Temporal Attention Network, MRSTAN)，它可以自适应地捕获交通网络的动态空间关系，也能够考虑到该网络的时空相关性、交通流的周期性。模型的整体框架如下图3。具体来说，本文的模型包括三个主要组件，分别是用于空间相关性提取的空间注意力模块、用于时间特征学习的循环神经网络模块、用于多分辨率融合以及输出预测的融合模块，它们都是用深度网络构建的。

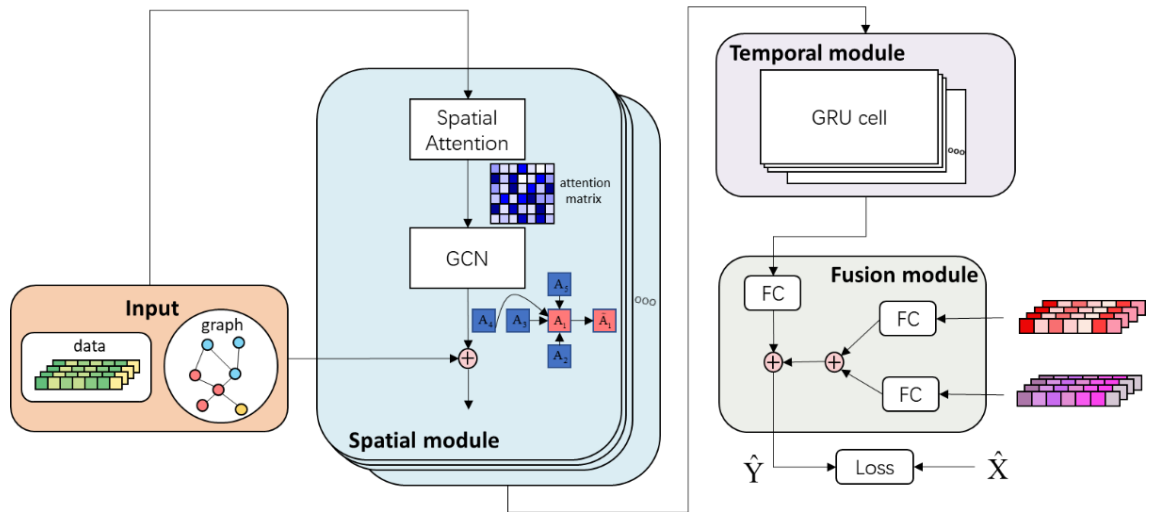


图 3: MRSTAN 的框架

4.1 空间依赖性

传统的 GCN 采用的图结构是启发式预定义的，也就是静态的，它仅代表道路网络的物理结构。因此，不能保证对流量之间的依赖关系进行最佳描述。例如，具有相似趋势但彼此远离的两个交通流之间的关系对于交通预测也很重要，因为这很可能说明它们共享相同的“时间模式”；对同一个节点来说，其邻居节点在不同时间步对其影响也是不同的。但是，这种依赖关系无法在预定义的基于道路拓扑的图中捕获。为了能够真实地动态地反映其空间结构，本文采用了复现的论文中提出的空间注意力

结构，以便改善 GCN 的不足之处，具体的空间注意力公式表示为：

$$S = V_s \cdot \sigma \left(\left(\tilde{X} \cdot W_1 \right) W_2 \left(W_3 \cdot \tilde{X} \right)^T + b_s \right) \quad (9)$$

其中的 $\tilde{X} \in R^{N \times C \times T_1}$ 是对输入矩阵 $X \in R^{T_1 \times N \times C}$ 重置后的结果， $V_s, b_s \in R^{N \times N}$ ， $W_1 \in R^{T_1}$ ， $W_2 \in R^{C \times T_1}$ ， $W_3 \in R^C$ 都是可学习的参数，而 σ 表示的是激活函数。接着对注意力得分矩阵采用 softmax 进行归一化得到矩阵 S' ，使得注意力权重总和为 1，其中的元素 $S'_{i,j}$ 表示节点 i 和节点 j 之间的相关强度，并满足：

$$S'_{i,j} = \frac{\exp(S_{i,j})}{\sum_{j=1}^N \exp(S_{i,j})} \quad (10)$$

上述的空间注意力能够动态地学习到节点间的空间依赖关系，本文将其与 GCN 结合，使得只能捕获静态拓扑结构的 GCN 能够自适应地获得动态的信息，从而更加使模型适应于真实的交通情况。加入空间注意力后的卷积操作公式变为：

$$y = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k \left(T_k(\tilde{L}) \odot S' \right) x \quad (11)$$

其中 \odot 表示哈达玛积。

通过串联地堆叠多层这样的图卷积，使得该卷积操作不仅可以学习到本来存在拓扑关系的节点的空间依赖关系，还可以学习到本来不存在拓扑关系但却存在隐藏关系的节点间的空间依赖关系。除此外，受残差网络的启发，本文在 GCN 中添加残差连接，以缓解过度平滑问题。

4.2 时间依赖性

循环神经网络通常应用于序列数据，因为它们具有记忆能力，可以学习序列部分之间的长期和短期依赖关系，其是最受欢迎的深度神经网络之一。由于梯度消失和梯度爆炸等缺陷，RNN 在长期预测方面存在局限性。为了解决这些问题，RNN 的变体 LSTM（Long Short-Term Memory）模型和 GRU 模型被提出，LSTM 与 GRU 相比，结构更加复杂，训练时间更长。因此，本文选择 GRU 模型来从交通数据中获取时间依赖性。如图4所示， h_{t-1} 表示在第 t-1 个时间步的隐藏状态； X_t 表示在第 t 个时间步的信息； \tilde{h}_t 是在第 t 个时间步存储的内存内容。GRU 有两个门，即一个重置门（reset gate）和一个更新门（update gate）。 r_t 表示重置门，用于控制忽略上一时刻状态信息的程度； z_t 表示更新门，用于选择性记忆第 t 个时间步存储的内容； $1 - z_t$ 表示遗忘第 t-1 个时间步的隐藏状态的程度， h_t 是时间 t 的输出状态， \otimes 表示的是哈达玛积。

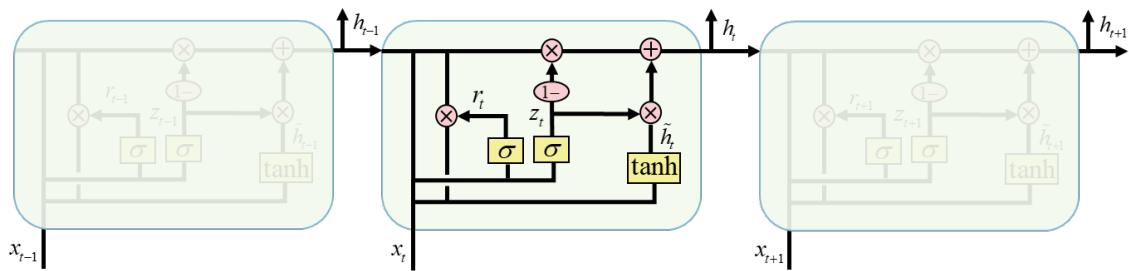


图 4: GRU 模型

对交通流预测而言，GRU 以 t-1 时刻的隐藏状态和当前交通信息为输入，获取 t 时刻的交通状况。该模型在捕获当前时刻的交通信息的同时，仍然保留了历史交通信息的变化趋势，并具有捕获时间依

赖性的能力。其具体公式如下：

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (12)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (13)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (14)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (15)$$

其中的 W_z 、 W_r 、 W_h 和 b_z 、 b_r 、 b_h 表示训练过程中的权重和偏差。对每个时间步使用 GRU 单元，从而使得它能够捕获到时间步之间的依赖关系。

4.3 多分辨率融合

交通流具有周期性，不同的分辨率下，交通流之间的相似性有助于模型更好地预测未来的交通流量。如下图5，2022年2月27日7:30-8:30为要预测的交通流的真实数据，其与昨天、上周同一时段的交通流量具有相似的趋势，例如：都在第5个观测点附近达到流量最低点，在接近第7个观测点达到最高峰，在这个时间段内交通流量都具有继续上升的趋势。

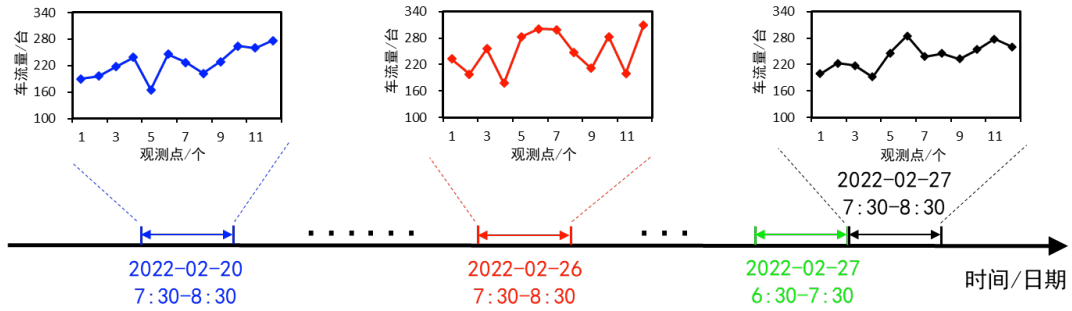


图 5: 交通流时间周期性

对于最近的交通流 $X \in R^{I_1 \times N \times C}$ （在上图中被表示为 2022 年 2 月 27 日 6:30-7:30 这段时间）、昨日交通流 $X_d \in R^{T_1 \times N \times C}$ 、上周交通流 $X_w \in R^{T_1 \times N \times C}$ 而言，多分辨率融合公式表示为：

$$\hat{Y} = W_r \cdot X_r + W_d \cdot X_d + W_w \cdot X_w \quad (16)$$

其中 $X_r \in R^{T_1 \times N \times C'}$ 是 X 经过时空依赖性处理以后的结果。 C' 是扩充的通道数， $\hat{Y} \in R^{T_2 \times N \times C}$ 表示未来交通流量的预测结果。

4.4 模型训练

本文采用 L1 范数作为损失函数，使得真实值 \hat{X} 与预测值 \hat{Y} 的绝对差值的总和最小化，即：

$$L_1(\theta) = \sum_{b=0}^{\text{size}-1} \sum_{n=1}^N \sum_{i=0}^{T_2} |\hat{x}_i^n - \hat{y}_i^n| \quad (17)$$

其中 θ 是模型参数集， \hat{x}_i^n 表示第 n 个结点在第 i 个时间步的真实值， \hat{y}_i^n 表示第 n 个结点在第 i 个时间步的预测值，size 表示每批次的样本数量， N 是节点总数， T 是要预测的时间步长。算法 1 概述了 MRSTAN 的训练过程。首先对输入数据进行预测（第 3-8 行），然后通过梯度下降迭代优化 MRSTAN（第 9 行），直到满足停止标准为止。

Procedure 1 Pseudo-Code for Training Procedure of MRSTAN.

Input: Traffic flow data: X, X_d, X_w , future ground truth: \hat{X}

Output: The trained model.

```
for each epoch do
    for each batch in training data do
        for layer = 0 to n do
            Spatial attention model:
            Attention GCN model:
        end
        GRU layer
        FC layer:
        Update the parameters;
    end
end
```

5 复现细节

5.1 与已有开源代码对比

本文引用了 ASTGCN^[12]的代码，其代码可在此获得：<https://github.com/wanhuaiyu/ASTGCN>。本文在其基础上进行了改进得到了模型 MRSTAN。与源代码相比，本文时间相关性方面采用了 GRU 模型，而不再是原论文中的注意力模型。其次，本文对日数据、周数据的处理方式也与原论文处理方式不相同。具体而言，原论文将日数据、周数据都输入到 ST block 中获取时空特征，最后与最近交通流进行融合，而本文中，日数据、周数据不再捕获时空特征，而是直接与捕获时空特征后的最近交通流进行直接融合，这样做的想法是希望学习出来的最近交通流与日数据、周数据足够相似。

5.2 实验环境与使用说明

代码运行所需环境：python ≥ 3.5 、mxnet $\geq 1.3.0$ 、pytorch、scipy、mxboard、tensorboard、pytest、pytest-xdist，其他没标明的表示版本没有特殊要求。

使用说明以 ASTGCN 模型为例，代码文件中可看到很多文件夹，接下来对文件夹简单进行介绍。configurations 中包含的是配置文件，可以直接在上面修改训练的代数、学习率、批训练数据大小等信息，也可以使用命令行进行修改；data 文件夹包含了 2 种数据，分别是 PEMS04 和 PEMS08 数据集，这两种数据集加州高速公路数据集，用于训练与测试；lib 文件中放置的 py 文件是一些对数据处理或者评估模型效果的工具类文件，其中的 data_preparation.py 文件需要首先运行，运行后将会生成由原数据处理后得到的用于后续训练和测试的数据集；model 文件中的 py 文件是模型框架的算法实现；根目录中的 train.py 是用于训练与测试，当实验环境已经配置好且数据集已经准备好时，就可以直接运行该文件或通过命令行运行该文件得到实验的结果。MRSTAN 模型的运行过程与之类似。

5.3 创新点

如前改进分析可知，本文创新点有：考虑了循环神经网络来获取时间特征，其相比于注意力的好处在于能够获得时间的先后顺序特征；将日数据、周数据从相似度的角度进行融合建模。具体效果见实验分析部分。

6 实验

6.1 数据集

本文使用加州运输局性能评估系统中的高速公路数据集 PeMS-04 和 PeMS-08。该数据集每隔 30s 收集一次，并在原数据的基础上聚合成每 5 分钟的间隔，包含总流量、平均速度、平均占有率三种交通特征数据。本文将数据集进行 Z-score 归一化并按随机顺序划分，60% 用于训练，20% 用于验证，20% 用于测试。

PeMS-04: 该数据集是旧金山湾的交通数据，时间跨度为 2018 年 1 月 1 日到 2018 年 2 月 28 日。该数据集包含 307 个节点，340 条边，16992 条数据。**PeMS-08:** 该数据集是圣贝纳迪诺的交通数据，时间跨度为 2016 年 7 月 1 日到 2016 年 8 月 31 日。该数据集包含 170 个节点，295 条边，17856 条数据。

6.2 评价指标

本文使用三种常用的性能指标来比较流量预测模型，分别是平均绝对误差（MAE）、均方根误差（RMSE）和平均绝对百分比误差（MAPE），具体公式如下。

$$MAE = \frac{1}{T_2 N E} \sum_{e=0}^{E-1} \sum_{n=1}^N \sum_{i=0}^{T_2} |\hat{x}_i^n - \hat{y}_i^n| \quad (18)$$

$$RMSE = \sqrt{\frac{1}{T_2 N E} \sum_{e=0}^{E-1} \sum_{n=1}^N \sum_{i=0}^{T_2} (\hat{x}_i^n - \hat{y}_i^n)^2} \quad (19)$$

$$MAPE = \frac{1}{T_2 N E} \sum_{e=0}^E \sum_{n=1}^N \sum_{i=0}^{T_2} \frac{|\hat{x}_i^n - \hat{y}_i^n|}{\hat{x}_i^n} \quad (20)$$

其中 E 是样本数，N 表示节点总数， T_2 表示要预测的时间步长， \hat{x}_i^n 表示第 n 个结点在第 i 个时间步的真实交通数据， \hat{y}_i^n 表示第 n 个结点在第 i 个时间步的交通流量预测值。

6.3 实验设置

所有实验均在 Ubuntu 服务器上进行（CPU: Intel(R) Xeon(R) Gold 5118, GPU: GeForce RTX 2080 Ti）。所有测试均采用 60 分钟作为历史时间窗口，即 12 个观测数据点，用于预测未来 60 分钟的交通流状况。所有基线方法都遵循他们论文中报告的最佳参数和结果。而对于本模型而言，所有隐藏维度都设置为 64，采用 3 阶的切比雪夫多项式进行图卷积操作，空间模块层数为 2，Adam 优化器的学习率为 0.001，一次训练所选取的样本数为 32，训练迭代次数设置为 250。

6.4 实验结果及分析

表1显示了 MRSTAN 和原方法在数据集 PeMS-04 和 PeMS-08 上提前 60 分钟预测的结果。

表 1: 算法实验对比

Methods	PeMS-04			PeMS-08		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ASTGCN	21.16	33.18	15.12%	18.22	28.44	11.49%
MRSTAN	20.45	32.38	13.29%	16.38	25.60	10.85%

从表中可以看到，改进的模型 MRSTAN 在所有指标中取得了最佳的效果，具体来说，在 PeMS-04

数据集上，本文方法在 MAPE 方面比原方法 ASTGCN 优于 1.83%，在 MAE 和 RMSE 方面的性能分别提升了 3.36% 和 2.41%；在 PeMS-08 数据集上，MAPE 比 ASTGCN 优于 0.64%，但 MAE 和 RMSE 分别获得了 10.09% 和 9.99% 的改进。这表明了改进的模型的有效性。

6.5 消融实验

为了验证 MRSTAN 不同模块的有效性，本文设计了 MRSTAN 的两个变体：1) “MRSTAN-A”：没有利用注意力机制的 MRSTAN；2) “MRSTAN-F”：没有利用融合机制的 MRSTAN。在 PeMS-04 和 PeMS-08 数据集上的消融实验结果见图6。

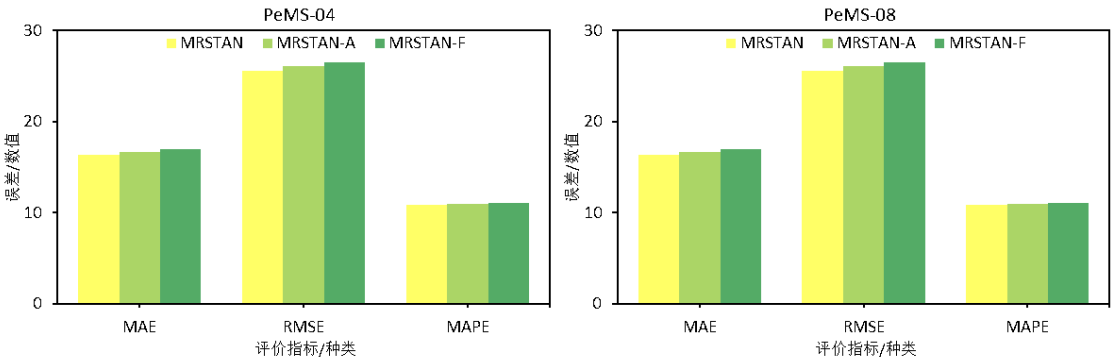


图 6: 总体误差

显然，这两个变体的误差更大，表现不如组件完整的 MRSTAN 模型，这表明了模型中的两种组件是有助于提高预测效果的。此外，可以观察到，当失去多分辨率融合模块后，模型的表现比失去注意力机制后的模型表现得更加糟糕，这可以看出，多分辨率融合模块对预测精度的重要性。本文还比较了这两种变体与原模型在不同预测步长下的误差，具体实验结果见图7。

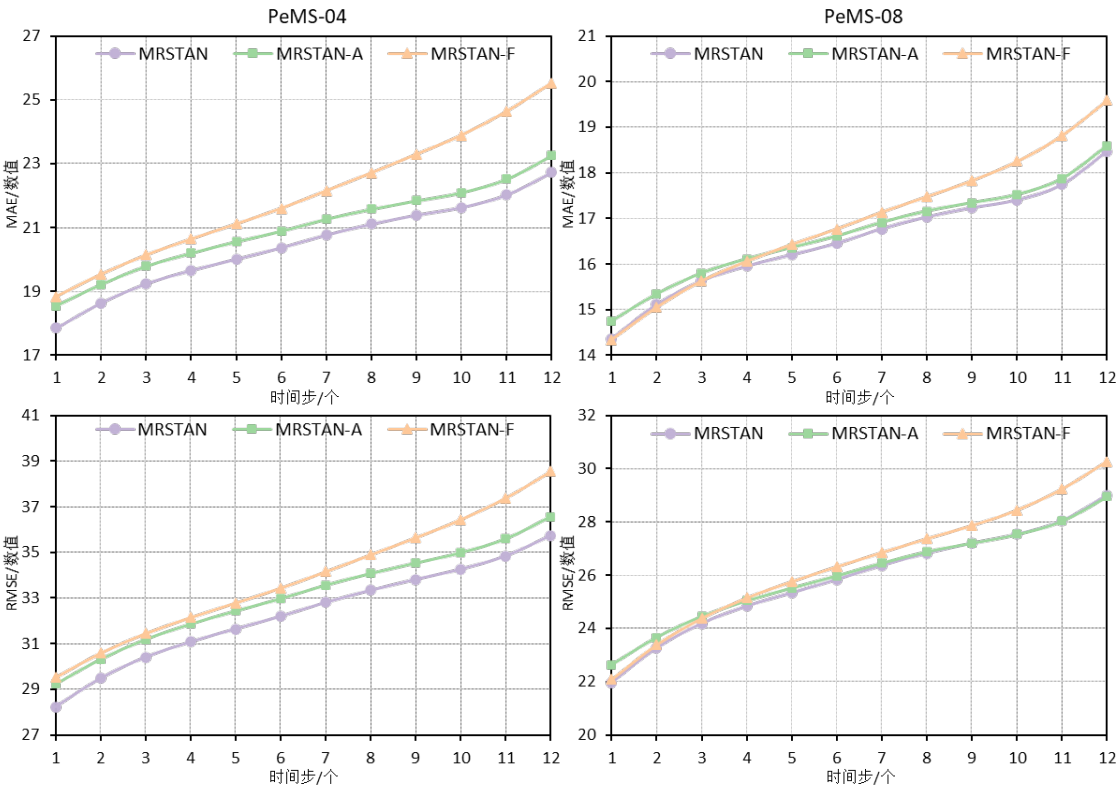


图 7: 时间步误差

可以观察到，在 PeMS-04 数据集上，原模型在每个时间步下的预测误差都是优于其两种变体，而在 PeMS-08 数据集上，原模型的各个时间步下的预测误差几乎都比其两种变体更小，只是在个别预测

步长上与其变体相差不大。这说明原模型在短期预测或者长期预测上的表现都优于这两种变体，表明了注意力机制和多分辨率融合模块在交通流长短期预测上的有效性。

7 总结与展望

在本文中，首先复现了用于交通流预测的 ASTCGN 模型，然后在此基础上提出了一种新颖的时空动态网络（MRSTAN）用于交通流的预测。该方法使用注意力机制和图卷积操作捕获不同节点之间的动态空间相关性，并通过循环神经网络捕获时间相关性，最后融合不同时期的交通流进行预测。对两个大规模数据集的评估表明，本文提出的模型明显优于原方法。此外，消融实验表明，模型中的组件时可以有效地挖掘交通流中隐藏的时空特征，这些组件结合在一起，可以获得更好的性能。

影响交通流的因素非常的多，包括：季节、交通事故、节假日与工作日、天气状况等，如何较好地结合这些因素建立一个更加全面的模型是未来工作的其中一个可行方向；深度学习对交通流模型的理解和可视化也是很重要的，交通流模型的可解释性有助于更好地理解交通流预测问题的本质。此外，交通流作为一种流体，已经有一些学者结合神经微分方程对其进行建模，这也是未来的研究方向之一。

参考文献

- [1] 王竟成, 张勇, 胡永利, 等. 基于图卷积网络的交通预测综述[J]. 北京工业大学学报, 2021, 47(8): 954-970.
- [2] 周欣荣, 沙海云, 高文宝, 等. 卡尔曼滤波在短时段交通预测上的应用[J]. 第一届中国智能交通年会论文集, 2005.
- [3] 徐翠翠. 基于 ARIMA 的城市交叉路口交通流预测建模[J]. 电子设计工程, 2022, 30(2): 20-23.
- [4] 张涛, 陈先, 谢美萍, 等. 基于 K 近邻非参数回归的短时交通流预测方法[J]. 系统工程理论与实践, 2010, 30(2): 376-384.
- [5] 赵亚萍, 张和生, 周卓楠, 等. 基于最小二乘支持向量机的交通流量预测模型[J]. 北京交通大学学报, 2011, 35(2): 114-117.
- [6] ZHU J Z, CAO J X, ZHU Y. Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections[J]. Transportation Research Part C: Emerging Technologies, 2014, 47: 139-154.
- [7] JIANG W, ZHANG L. Geospatial data to images: A deep-learning framework for traffic forecasting[J]. Tsinghua Science and Technology, 2018, 24(1): 52-64.
- [8] CHENG X, ZHANG R, ZHOU J, et al. Deeptransport: Learning spatial-temporal dependency for traffic condition forecasting[C]//2018 International Joint Conference on Neural Networks (IJCNN). 2018: 1-8.
- [9] DIAO Z, WANG X, ZHANG D, et al. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 33: 01.

2019: 890-897.

- [10] 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, 21(9): 3848-3858.
- [11] ALI M A, VENKATESAN S, LIANG V, et al. TEST-GCN: Topologically Enhanced Spatial-Temporal Graph Convolutional Networks for Traffic Forecasting[C]//2021 IEEE International Conference on Data Mining (ICDM). 2021: 982-987.
- [12] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 33: 01. 2019: 922-929.
- [13] PAN Z, LIANG Y, WANG W, et al. Urban traffic prediction from spatio-temporal data using deep meta learning[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 1720-1730.
- [14] 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: vol. 34: 01. 2020: 1234-1241.