

GA2T: 结合图注意力网络的交通流预测模型

祁舒畅, 刘起东, 刘超越, 徐明亮*, 邱紫鑫

(郑州大学计算机与人工智能学院 郑州 450001)
(iexumingliang@zzu.edu.cn)

摘要: 交通流预测是智能交通系统的核心组成部分. 针对当前交通流预测方法准确率低的问题, 提出一种交通流预测模型 GA2T. 通过构建具有融合式编解码器的 Transformer 架构对交通数据进行时间建模, 利用图注意力网络对交通数据进行空间建模, 从而捕获交通流复杂的时空依赖性. 在 2 个真实交通数据集 METR-LA 和 PEMS-BAY 上的实验结果表明, 与预测效果最佳的基线模型 DCRNN 相比, GA2T 在 3 个评价指标(MAE, MAPE 和 RMSE)上分别降低了 0.25, 0.38, 0.89 和 0.14, 0.34, 0.44, 在同类工作中处于领先的水平, 验证了该模型的可行性及有效性.

关键词: 交通流预测; 时空数据; 图注意力网络; Transformer
中图分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023.19758

GA2T: A Traffic Flow Prediction Model Combined with Graph Attention Networks

Qi Shuchang, Liu Qidong, Liu Chaoyue, Xu Mingliang*, and Qiu Zixin

(School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001)

Abstract: Traffic flow prediction is the core component of the intelligent transportation system. In view of the low accuracy of the current traffic flow prediction methods, a new traffic flow prediction model GA2T is proposed. By building a Transformer architecture with fused encoder and decoder to model traffic data temporally, and using graph attention networks to model traffic data spatially, the complex spatial-temporal dependencies of traffic flow are captured. The experimental results on two real traffic datasets METR-LA and PEMS-BAY show that compared with the best baseline model DCRNN, GA2T reduces the three evaluation metrics (MAE, MAPE, RMSE) by 0.25, 0.38, 0.89 and 0.14, 0.34, 0.44. This proves the effectiveness and feasibility of GA2T, and verifies its advances in peer works.

Key words: traffic flow prediction; spatial-temporal data; graph attention networks; Transformer

近年来, 我国经济高速发展, 汽车制造业的兴起和国民消费水平的提高使得私家车拥有率逐年上升, 全国汽车保有量不断增加. 在给人们生活带来便利的同时也引发了一些弊端, 如越来越严重的交通拥堵和交通事故严重影响人们的生活, 对

社会和个人经济造成巨大损失.

如何缓解和规避交通拥堵、提高出行效率是交通领域亟待解决的问题之一. 近年来, 随着相关技术的发展, 科学可行的调控方法——交通流预测成为解决该问题的一种有效手段. 交通流预测通

收稿日期: 2022-06-08; 修回日期: 2022-08-12. 基金项目: 国家自然科学基金(61906174, 62172371); 中国博士后科学基金(2020M672275); 河南省重点研发与推广专项(科技攻关)(222102210248). 祁舒畅(1995—), 女, 硕士研究生, 主要研究方向为时空数据分析、图神经网络、路径规划; 刘起东(1990—), 男, 博士, 副教授, 硕士生导师, 主要研究方向为多智能体强化学习、时空数据分析、图神经网络、运动规划; 刘超越(1998—), 男, 硕士研究生, 主要研究方向为时空数据分析、强化学习、路径规划; 徐明亮(1981—), 男, 博士, 教授, 博士生导师, CCF 会员, 论文通信作者, 主要研究方向为智能图形学、群组仿真; 邱紫鑫(1999—), 女, 硕士研究生, 主要研究方向为时空数据预测、强化学习.

过预知未来一段时间的城市道路交通状况, 达到及时规避和缓解交通事故和交通拥堵的目的. 但是, 由于道路网络的非欧几里得结构, 以及交通流数据难以捕获的、高度动态的时空依赖性, 准确地预测交通状况是困难的.

深度学习已在许多领域上获得了巨大成功, 特别是在自然语言处理和计算机视觉领域, 原因是这些领域所涉及的数据(如词句、图片等)都属于欧几里得空间; 而交通流数据的空间结构是非欧几里得的, 使得常见的深度学习模型, 如卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)都无法直接应用在交通流预测的空间依赖建模上. 近年来, 图神经网络(graph neural network, GNN)逐渐兴起, 可以有效地从非欧几里得数据中提取信息表示. 因此, 本文基于 GNN 提出一种深度学习模型 GA2T(graph attention network and Transformer), 根据历史交通信息对未来的交通状况进行预测, 为城市道路的交通问题提供解决方案.

GA2T 通过图注意力网络(graph attention networks, GAT)在非欧几里得空间中动态聚合交通流的空间特征, 通过时间注意力机制捕捉交通流时间序列的长期依赖性; 提出一个具有融合式编解码器的 Transformer 体系结构, 将解码器的功能融合到编码器中, 可以非自回归地进行多步预测. 在 2 个公开的数据集上进行实验, 通过对比 GA2T 与其他几种经典基线模型的预测性能, 证明 GA2T 在提取交通流时空依赖上的优越性.

1 相关工作

本文根据各类交通流预测方法的原理, 将其分为基于统计学、基于机器学习和基于深度学习的 3 种预测模型.

1.1 基于统计学的预测模型

早期工作是基于数理统计方法进行交通流预测. 历史平均(historical average, HA)模型是最早被用于交通流预测的统计学模型, Stephanedes 等^[1]首次将该模型用于预测未来交通流. 即便 HA 模型的计算速度很快, 但它只是简单地以历史时间片的均值作为预测值, 因此预测效果很差. 随后, Hamed 等^[2]使用差分自回归移动平均(autoregressive integrate moving average, ARIMA)模型预测城市交通道路的交通流状况, 该模型是一种参数型模型, 通过交通流的历史观测序列拟合模型参数, 实现对未来交

通状况的预测; 但该模型更适合处理平滑的时间序列, 难以捕获交通流高度非线性的时间特征. Okutani 等^[3]首次将卡尔曼滤波模型(Kalman filtering model, KFM)应用到交通流预测领域, 利用当前帧和上一帧的交通状况进行预测, 在一定程度上提高了预测精度. Ojeda 等^[4]改进了 KFM, 使其可以自适应地同时预测多个时间步的交通流状况; 但该模型擅长处理线性相关的时间序列, 对交通系统运行过程中各类突发状况的处理能力则较差.

1.2 基于机器学习的预测模型

近年来, 人工智能技术快速发展, 机器学习方法开始被应用到交通流预测领域.

支持向量机(support vector machine, SVM)模型作为一种经典的机器学习方法, 逐步被用于交通流的预测. 张本土^[5]利用 SVM 方法对城市道路未来的拥堵状况进行预测, 预测准确率和计算速度均大幅提升. 支持向量回归(support vector regression, SVR)模型是一种在 SVM 的基础上进行回归分析的改进模型. 姚智胜等^[6]利用 SVR 对道路中的交通流数据进行信息挖取, 进而对交通流进行有效的预测, 在提高短期预测准确率的同时, 也解决了过拟合和局部最优解问题. 与 SVM 和 SVR 不同, K 近邻(the K -nearest neighbor, KNN)模型是一种非参数的机器学习方法. Zhang 等^[7]提出一种基于 KNN 的交通流短期预测方法, 通过预先设定状态向量、KNN 值等非参数模型的相关因子, 从历史数据库中查找模式相似的近邻序列, 从而预测下一时刻的交通状况, 该方法同时提高了预测精度和运行效率.

1.3 基于深度学习的预测模型

深度学习作为机器学习的一个重要分支, 已经在很多领域得到广泛应用. 近年来, 越来越多的学者利用深度学习对交通流数据的时空特性进行分析预测.

Fu 等^[8]使用长短期记忆(long short-term memory, LSTM)网络和门控循环单元(gate recurrent unit, GRU)进行短期交通流量的预测, 这 2 个网络都属于 RNN, 具有良好的序列建模能力, 可以有效地捕获交通流的时间依赖性, 取得了良好的预测效果. 然而, 上述方法都没有考虑交通流的空间依赖性, 忽略了空间结构对交通流的约束信息. 为了解决这个问题, Zhang 等^[9]提出深度时空残差网络, 使用 CNN 捕获网格化区域的空间依赖和时间依赖性, 并分别使用 3 个残差网络从 period, closeness 和 trend 这 3 个角度建模交通流时空依赖; Yao 等^[10]

则结合了 CNN 和 LSTM 来联合建模交通流的空间和时间依赖性: 这些方法都将交通流的空间结构视作网格化的区域, 应用 CNN 提取空间特征. 由于交通流沿路网传播, 其空间结构与路网相似, 表现为非欧几里得图空间结构, 故无法直接被 CNN 处理.

近年来, GNN 飞速发展, 2016 年, Kipf 等^[11]提出图卷积网络(graph convolutional networks, GCN); 2017 年, Veličković 等^[12]提出 GAT 等. 文献[13-14]注意到交通流的非欧几里得图空间结构, 采用 GNN 聚合交通流邻域节点的空间特征. 2020 年, Zhao 等^[15]提出一种结合 GCN 和 GRU 的时间 GNN 模型, 分别利用 GCN 和 GRU 来捕获交通数据的空间依赖和时间依赖; Yu 等^[16]结合 GCN 和膨胀卷积的时空卷积块, 提出时空 GCN 模型来建模交通流数据的时空依赖性; Guo 等^[17]提出基于注意力机制的时空 GCN 模型, 是一种将 GCN 和注意力机制相结合的预测模型, 其分别利用时间和空间注意力机制对交通流数据的时间、空间相关性进行建模, 大大提高了预测精度; Cai 等^[18]则结合扩散卷积网络和 Transformer 模型, 可以有效地捕获交通流特征的时空依赖性; Zhang 等^[19]提出一种时空图扩散网络, 通过考虑全局地理信息提高了模型的预测能力.

本文使用 GAT 提取交通流动态的空间特征, 比已有的基于 GCN 的静态图方法具有更大的表示空间和更好的建模能力; 在时间建模方面, 本文采用基于注意力机制的 Transformer 架构, 不同于基于 RNN 或 CNN 的方法, 该模型还可以直接捕获交通流序列的长时依赖.

2 GA2T

2.1 问题定义

本文将交通流预测问题视作一种多维的时空序列预测问题, 通过道路网络中传感器的历史观测特征, 预测未来各个传感器节点的交通特征.

本文将道路网络定义为一张有向带权图 $G: (V, E, A)$, 其中, $V = \{v_1, v_2, \dots, v_N\}$ 表示图 G 中所有节点的集合; N 为路网节点的数量; E 表示边的集合; $A \in \mathbb{R}^{N \times N}$ 为邻接矩阵; A_{ij} 表示节点 v_i 与 v_j 的关联程度.

设 $\mathbf{X}_t \in \mathbb{R}^{N \times p}$ 表示在 t 时刻下 N 个节点的交通流特征, 其中, p 表示交通流信息的特征数. 交通流预测问题就是要建立一个模型 $f(\cdot)$, 在已知道

路拓扑图 G 和过去 T_h 个时间片的历史交通信息的基础上, 预测未来 T_p 个时刻的交通信息, 公式为

$$f(G; [\mathbf{X}_{t-T_h+1}, \dots, \mathbf{X}_t]) = [\mathbf{X}_{t+1}, \dots, \mathbf{X}_{t+T_p}].$$

2.2 空间依赖建模

2.2.1 构建交通图拓扑结构

与图片视频等欧几里得数据不同, 交通流数据在空间维度上呈现出非欧几里得的图式结构, GAT 会在这个稀疏的道路拓扑图上进行动态的空间依赖建模. 考虑到路网中道路节点间的相关性通常与距离成反比, 本文基于距离生成道路拓扑图 G , 并基于阈值高斯核函数^[20]稀疏化和归一化图 G . 图 G 邻接矩阵 A 第 i 行第 j 列的元素为

$$A_{ij} = \begin{cases} \exp\left(\frac{-\text{dist}(v_i, v_j)}{\sigma}\right), & \text{if } \text{dist}(v_i, v_j) \leq \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

其中, $\text{dist}(v_i, v_j)$ 表示 v_i 与 v_j 的欧几里得距离; σ 表示所有距离的标准差; ε 表示稀疏化的阈值, 当 $\text{dist}(v_i, v_j)$ 大于阈值 ε 时, 设置 $A_{ij}=0$.

2.2.2 采用 GAT 的空间层

交通流的空间模式与道路拓扑图 G 密切相关, 在不同时刻有很大的差异, 特别是当某些突发事件(如交通拥堵、交通事故等)发生时, 交通流的空间模式会发生突变. 本文采用 GAT 动态地捕获交通流各节点间的空间相关性, 根据交通模式的不同, 在聚合邻域特征时为各个节点分配不同的权重, 因此比 GCN 等基于静态图的方法具有更大的表示空间和更强的表示能力. 从图 1 可以看出, 中心节点 v_i 扫描所有邻域节点 v_j , 对它们的特征向量进行加权求和, 得到聚合后的特征向量. 其中, 加权求和的权重 α_{ij} 就是 GAT 中的注意力系数, 是由 v_i 对每个邻域节点 v_j 的注意力分数 e_{ij} 经过 Softmax 后得到的, 即

$$\alpha_{ij} = \text{Softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i \cup \{i\}} \exp(e_{ik})}.$$

其中,

$$e_{ij} = \text{leakyReLU}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_i \parallel \mathbf{W} \mathbf{h}_j]);$$

$\mathbf{h}_i \in \mathbb{R}^{d_h}$ 是第 i 个节点的输入特征向量; d_h 表示隐藏层的特征维度; $\mathbf{W} \in \mathbb{R}^{d_h \times d_h}$ 和 $\mathbf{a} \in \mathbb{R}^{2d_h}$ 表示可学习的参数; N_i 表示节点 v_i 的邻域节点集合; \cup 表示合并操作; \parallel 表示拼接操作.

数据的空间结构信息,与时间维度无关. TE 侧重于时序的连贯性,相邻时刻下的 TE 更相似;而 SE 更侧重于节点的特异性,这是因为节点在特征矩阵中的位置与其空间位置无关. 本文提出 SE, 其以一个节点级别的 One-Hot 向量组为输入, 通过一个 MLP 来学习路网节点的特异性信息, 公式为

$$SE = \text{MLP}(I_N, \theta_S).$$

其中, $SE \in \mathbb{R}^{N \times d_h}$; $I_N \in \mathbb{R}^{N \times N}$ 为单位矩阵, 即一个节点级别的 One-Hot 向量组; θ_S 为可学习的参数.

将 TE 与 SE 按位置相加即可得到 STE, 公式为

$$STE_{t,n} = TE_t + SE_n.$$

其中, $STE \in \mathbb{R}^{T \times N \times d_h}$, $STE_{t,n}$ 表示其在第 t 个时间片下在第 n 个路网节点上的分量.

2.5 模型框架

BERT^[21]是一个基于 Transformer 的预训练模型, 它仅使用 Transformer 的编码器, 并通过无监督任务来预训练模型参数, 可以提升下游任务的泛化能力和收敛速度. 图 2 所示为 BERT 采用的 2 类无监督任务中的一种: 文本序列在输入到编码器之前, 随机地选择一个单词 w_i 并对其执行掩蔽

操作, 最后以编码器输出序列的第 i 个分量来预测原来的单词 w_i .

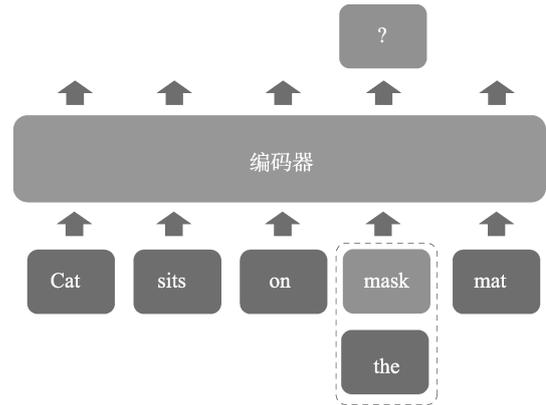


图 2 BERT 预训练模型的无监督任务

受该任务启发, 本文提出一种融合式编解码器结构, 将需要预测的时间片视为被掩蔽的对象, 并以整个历史-预测时间序列作为输入, 编解码器在掩蔽位置的输出直接用来生成预测结果. 这样, 解码器的功能被融入到编码器中, 非自回归地生成序列, 取代了 Transformer 自回归的序列生成过程. 图 3 所示为 GA2T 的整体结构.

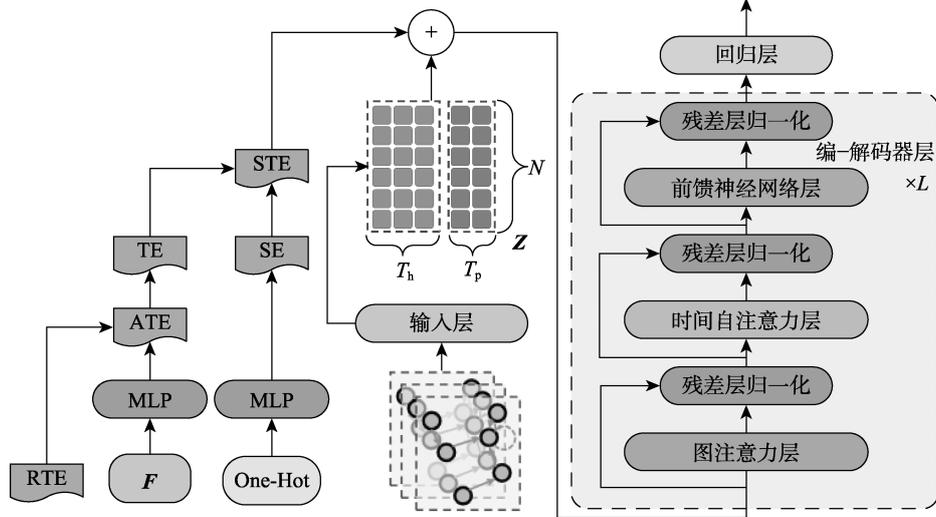


图 3 GA2T 的结构

(1) 首先使用一个全连接的输入层将交通流的历史特征序列映射到一个高维特征空间; 然后用一个零矩阵 $Z \in \mathbb{R}^{T_p \times N \times d_h}$ 表示被掩蔽的未来特征序列, 与历史特征序列进行拼接, 其中, T_p 表示预测时间片数; N 为路网节点数; d_h 表示隐藏层特征维度; 最后计算整个输入序列的 STE, 与输入序列进行加和, 以引入时空结构信息. 整个过程的计算公式为

$$H^0 = [(XW + b) \parallel Z] + STE.$$

其中, $X \in \mathbb{R}^{T_h \times N \times p}$ 表示原始交通流特征矩阵; $H^0 \in \mathbb{R}^{T \times N \times d_h}$; $T = T_h + T_p$ 表示时间窗口的总长度; $W \in \mathbb{R}^{p \times d_h}$ 和 $b \in \mathbb{R}^{d_h}$ 表示可训练的参数.

(2) H_0 作为输入被送入编解码器中. 编解码器由 L 层结构相同的编解码层组成, 每个编解码器层都包含 GAT 层、时间注意力层和前馈神经网络层 3 个模块. GAT 层和时间注意力层分别捕获空

间特征和时间特征, 前馈神经网络层用以提升模型建模能力, 各模块之间应用残差连接与层归一化. 考虑到未来的交通状况无法干涉过去的交通状况, 本文引入一个时序掩蔽矩阵 $M \in \mathbb{R}^{T \times T}$, 用于屏蔽未来时刻的影响, 公式为

$$M_{ij} = \begin{cases} -\infty, & \text{if } i < j \\ 0, & \text{otherwise} \end{cases}$$

(3) 使用一个全连接的回归层, 将编解码器在掩蔽位置的输出映射为交通流在未来时刻的预测值.

(4) 以预测值和真实值的平均绝对误差作为损失函数, 通过梯度下降方法优化模型参数.

3 实验

3.1 实验数据集

本文在 2 个公开的真实交通数据集 METR-LA^[22] 和 PEMS-BAY^[23] 上进行了大量实验.

(1) METR-LA 数据集. 交通数据由洛杉矶公路上安置的 2 070 个监测器在 2012-03-01—2012-06-30 采集得到.

(2) PEMS-BAY 数据集. 交通数据由加州运输部门安置的 325 个监测器在 2017-01-01—2017-05-31 采集得到.

3.2 实验参数设置

GA2T 需要调整的超参数如下:

(1) 注意力头数 K . GA2T 分别使用 GAT 层和时间注意力层提取交通流的空间特征和时间特征, K 直接影响模型的性能. 本文以 METR-LA 数据集为例, 测试不同 K 值对模型性能的影响. 表 1 所示为采用不同 K 值的 GA2T 在第 12 个预测时间片上的预测性能, 其中, 采用平均绝对误差(mean absolute error, MAE)、平均绝对百分比误差(mean absolute percentage error, MAPE)和均方根误差(root mean squared error, RMSE)评价 GA2T 的性能. 可以看出, 当 K 为 2 时, 模型的预测性能普遍优于 K 为 1 和 4 的情况, 这可能是因为当 K 过小时不能充分发挥多头注意力机制的建模能力, 而当 K 过大时则发生了过拟合现象.

(2) 表 2 所示为 $K=2$ 时, 隐藏层维度 d_h 对 GA2T 性能的影响. 可以看出, d_h 的变化会严重影响模

型的性能: 当 $d_h=32$ 时, MAE, MAPE 和 RMSE 的值分别为 3.36, 9.61 和 7.23, GA2T 的性能在所有对照组中最优; 而当 $d_h=64$ 时, MAE, MAPE, RMSE 的值变高, 这可能是因为过多的参数导致过拟合现象; 在 $d_h=8$ 时, GA2T 效果极差, 这是因为当 d_h 太小导致模型的拟合能力严重不足. 因此, 当 $K=2$, $d_h=32$ 时, GA2T 可以达到最佳性能.

表 2 d_h 值对 GA2T 性能的影响

d_h	MAE	MAPE/%	RMSE
8	3.78	11.19	7.78
16	3.39	9.89	7.34
32	3.36	9.61	7.23
64	3.46	9.95	7.36

(3) 其他超参数. 训练数据的 batch-size 设置为 64, 最大迭代数设定为 2 000, 激活函数采用 ReLU 函数. 使用 Adam 优化器进行训练, 初始学习率设定为 0.005. 将一个步长设定为 5 min, 以过去 60 min(12 个时间片)的交通数据预测未来 60 min(12 个时间片)的交通状况.

所有实验都在一台搭载 Ubuntu 18.04 操作系统, 一块 Geforce RTX 3080 GPU 的设备上进行.

3.3 实验结果分析

3.3.1 对比实验

选择一些经典的交通流预测模型作为基线方法, 并在 METR-LA 和 PEMS-BAY 这 2 个交通数据集中进行实验, 对比分析 GA2T 与基线方法的预测性能.

本文以 MAE, MAPE 和 RMSE 这 3 个指标来评价 GA2T 的预测性能. 表 3 和表 4 所示为本文模型与 FNN, FC-LSTM^[24], STGCN^[16], DCRNN^[22], HA^[1], ARIMA^[2], SVR^[25] 共 8 个模型在 2 个数据集上的预测性能. 方便起见, 本文只展示 15 min(短期预测), 30 min(中期预测)和 60 min(长期预测)的数据.

从表 3 和表 4 可以看出:

(1) 基于深度学习的模型(FNN, FC-LSTM^[24], STGCN^[16], DCRNN^[22])的预测误差明显低于基于统计学的预测模型(HA^[1], ARIMA^[2])和基于机器学习的预测模型(SVR^[25]), 说明数据驱动的深度学习方法在处理非线性的交通流数据时有很大优势.

(2) 同时建模时间依赖和空间依赖的深度学习方法(DCRNN, STGCN, GA2T)的预测性能普遍优于仅考虑时间依赖的深度学习方法(FNN, FC-LSTM), 说明空间依赖的重要性, 通过交通流的空

表 1 K 值对 GA2T 性能的影响

K	MAE	MAPE/%	RMSE
1	3.37	9.63	7.22
2	3.36	9.61	7.23
4	3.37	9.71	7.27

表 3 不同预测时长下 8 个模型在 METR-LA 数据集上的性能比较

模型	15 min			30 min			60 min		
	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE
HA ^[1]	4.16	13.00	7.80	4.16	13.00	7.80	4.16	13.00	7.80
ARIMA ^[2]	3.99	9.60	8.21	5.15	12.70	10.45	6.90	17.40	13.23
SVR ^[25]	3.88	9.50	8.45	5.21	12.30	10.65	6.78	16.90	12.73
FNN	3.89	9.60	7.39	4.38	11.90	8.67	6.53	14.60	9.94
FC-LSTM ^[24]	3.44	9.60	6.30	3.77	10.90	7.23	4.37	13.20	8.69
STGCN ^[16]	2.88	7.62	5.74	3.47	9.57	7.24	4.59	12.70	9.40
DCRNN ^[22]	2.77	7.30	5.38	3.15	8.80	6.45	3.60	10.50	7.60
GA2T	2.53	6.55	5.23	2.93	8.04	6.26	3.35	9.61	7.22

表 4 不同预测时长下 8 个模型在 PEMS-BAY 数据集上的性能比较

模型	15 min			30 min			60 min		
	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE
HA ^[1]	2.88	6.80	5.59	2.88	6.80	5.59	2.88	6.80	5.59
ARIMA ^[2]	1.62	3.50	3.30	2.33	5.40	4.76	3.38	8.30	6.50
SVR ^[25]	1.85	3.80	3.59	2.48	5.50	5.18	3.28	8.00	7.08
FNN	2.20	5.19	4.42	2.30	5.43	4.63	2.46	5.89	4.98
FC-LSTM ^[24]	2.05	4.80	4.19	2.20	5.20	4.55	2.37	5.70	4.96
STGCN ^[16]	1.36	2.90	2.96	1.81	4.17	4.27	2.49	5.79	5.69
DCRNN ^[22]	1.38	2.90	2.95	1.74	3.90	3.97	2.07	4.90	4.74
GA2T	1.33	2.79	2.84	1.66	3.68	3.77	1.93	4.46	4.40

间拓扑结构来融合和捕获空间特征,有助于提高交通流预测的性能和精度。

(3) 与其他的预测模型相比,GA2T 在 2 个数据集上都取得了最优的预测性能.与最优的基线模型相比,GA2T 的 MAE 指标在 2 个数据集上分别下降了 0.25 和 0.14, MAPE 指标分别下降了 0.89 和 0.44, RMSE 指标分别下降了 0.38 和 0.34, 这是因为该模型采用的 GAT 可以捕获交通流动态的空间依赖,而时间注意力层则可以直接捕获交通流数据的长时依赖。

综上所述,GA2T 无论在短时预测还是在中长时预测中都取得了更良好的预测效果,可以为城市交通管理提供更精准的预测服务,同时为出行者提供更可靠的数据支持。

3.3.2 消融实验

对 GA2T 进行消融实验,以验证各个模块在模型中的作用以及对模型整体性能的影响程度.为了验证 GAT 层作用,本文构建了 GA2T 的 2 个变体:GA2T_GCN 和 GA2T_NSL,前者将模型中建模空间依赖的 GAT 层替换为 GCN 层,后者则移除了 GAT 层,相当于不考虑空间依赖;为了验证 STE 模块对模型性能的影响,构建了 GA2T_NT(移除了 TE 模块),GA2T_NS(移除了 SE 模块)和 GA2T_NST(同时移除了 SE 和 TE 模块)3 个变体。

基于这些变体模型在 METR-LA 数据集上进行了大量实验,并记录每个模型在预测区间为 15 min, 30 min 和 60 min 时的预测效果,如图 4 所示。

从图 4a 可以看出,GA2T 的预测效果最好,GA2T_GCN 的预测效果次之,GA2T_NSL 的效果最差.这种现象产生的原因可能是:GA2T_GCN 采用 GCN 替代了 GA2T 中的 GAT,在空间建模时依据静态的先验的图结构,实际上是限制了模型的代表能力,导致预测效果变差;GA2T_NSL 则直接移除了 GAT 层,相当于完全不进行空间依赖建模,所以效果最差。

从图 4b 可以看出,STE 模块对模型性能的影响很大,如在 60 min 预测时,完全移除 STE 模块的 GA2T_NST 的 MAE, MAPE 和 RMSE 比 GA2T 上升了 0.08, 0.40, 0.20, 说明了 STE 模块的重要性;此外,只移除 SE 或 TE 的变体模型的预测性能也发生了不同程度的下降,其中,GA2T_NS 的预测误差比 GA2T_NT 更大,说明移除 SE 的影响比 TE 更大。

消融实验结果表明,GA2T 的 GAT 层和 STE 模块都发挥着重要的作用,有助于提升模型的预测能力和预测精度.尤其是空间类模块对模型的影响很大,说明建模空间依赖的重要性,也从侧面验证了 GA2T 的有效性。

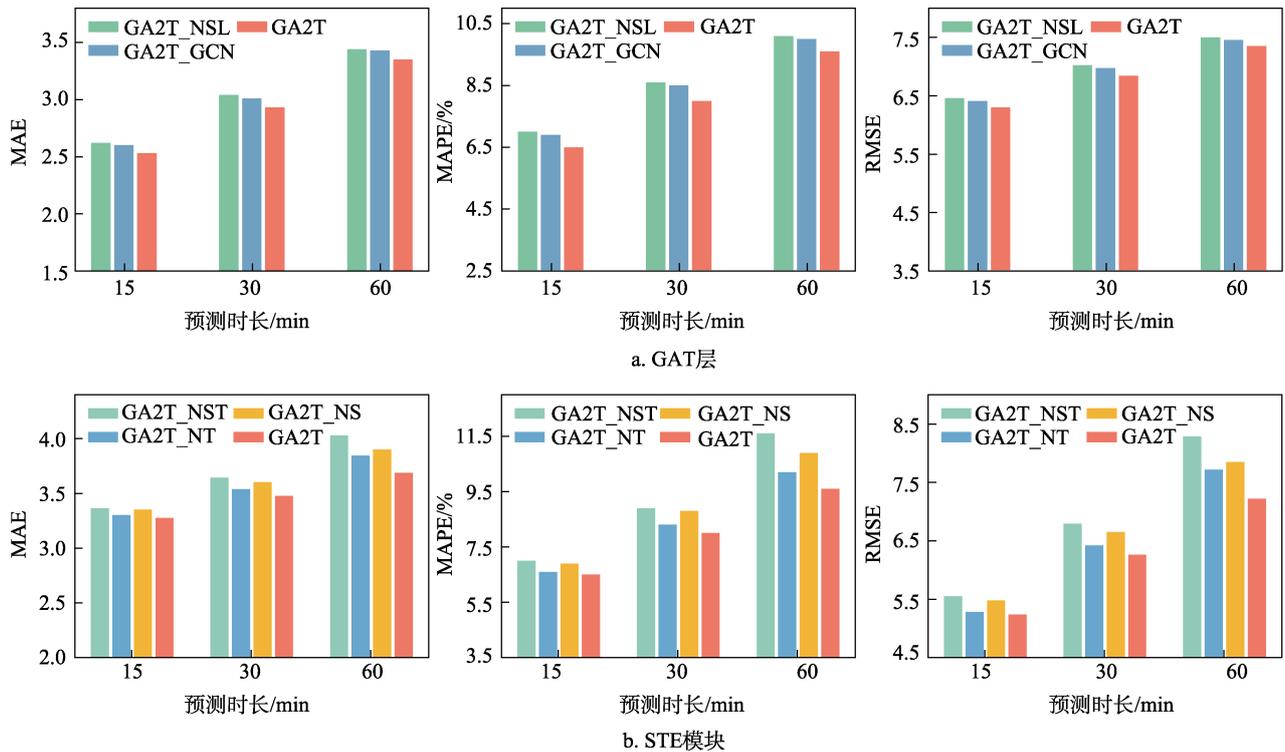


图 4 GA2T 及其变体的性能比较

4 结 语

本文基于 GAT 和 Transformer 提出一种交通流预测模型 GA2T. 该模型利用 GAT 对交通数据进行动态聚合的空间建模, 同时使用注意力机制捕获交通流序列的长期依赖性; 还设计了一种具有融合式编解码器的体系结构, 可以非自回归地预测未来的交通状况. 在 METR-LA 和 PEMS-BAY 这 2 个公开的真实交通数据集上进行了大量实验, 结果表明, 与预测效果最佳的基线模型相比, GA2T 取得了更优的预测性能. 最后通过大量的消融实验, 验证了 GA2T 中每个模块的重要程度.

目前, GA2T 仅应用在交通流预测领域, 理论上也可应用到其他具有时空特性的预测任务. 然而, 不同的时空预测任务各有特性, 如何使 GA2T 适应这些特性是一项巨大的挑战. 未来的工作将侧重于提高 GA2T 的泛化能力, 将其推广到更多的时空预测任务上.

参考文献(References):

[1] Stephanedes Y J, Michalopoulos P G, Plum R A. Improved estimation of traffic flow for real-time control[J]. Transportation Research Record, 1981(795): 28-39
 [2] Hamed M M, Al-Masaeid H R, Said Z M B. Short-term predic-

tion of traffic volume in urban Arterials[J]. Journal of Transportation Engineering, 1995, 121(3): 249-254
 [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] Ojeda L L, Kibangou A Y, de Wit C C. Adaptive Kalman filtering for multi-step ahead traffic flow prediction[C] //Proceedings of the American Control Conference (ACC). Los Alamitos: IEEE Computer Society Press, 2013: 4724-4729
 [5] Zhang Benshi. Urban traffic congestion estimation and prediction using floating car trajectory[D]. Dalian: Dalian University of Technology, 2014(in Chinese)
 (张本土. 基于浮动车轨迹的城市交通拥堵评估与预测[D]. 大连: 大连理工大学, 2014)
 [6] Yao Zhisheng, Shao Chunfu, Gao Yongliang. Research on methods of short-term traffic forecasting based on support vector regression[J]. Journal of Beijing Jiaotong University, 2006, 30(3): 19-22(in Chinese)
 (姚智胜, 邵春福, 高永亮. 基于支持向量回归机的交通状态短时预测方法研究[J]. 北京交通大学学报, 2006, 30(3): 19-22)
 [7] Zhang L, Rao Q, Yang W C, et al. An improved K-NN non-parametric regression-based short-term traffic flow forecasting model for urban expressways[C] //Proceedings of the 4th International Conference on Transportation Engineering. Reston: American Society of Civil Engineers, 2013: 1214-1223
 [8] Fu R, Zhang Z, Li L. Using LSTM and GRU neural network methods for traffic flow prediction[C] //Proceedings of the 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). Los Alamitos: IEEE Computer Society Press, 2016: 324-328
 [9] Zhang J B, Zheng Y, Qi D K. Deep spatio-temporal residual

- networks for citywide crowd flows prediction[C] //Proceedings of the 31st AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017: 1655-1661
- [10] Yao H X, Tang X F, Wei H, *et al.* Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction[C] //Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 5668-5675
- [11] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[OL]. [2022-06-08]. <https://arxiv.org/abs/1609.02907>
- [12] Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks[OL]. [2022-06-08]. <https://arxiv.org/abs/1710.10903>
- [13] Bui K H N, Cho J, Yi H. Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues[J]. *Applied Intelligence*, 2022, 52(3): 2763-2774
- [14] Jiang W W, Luo J Y. Graph neural network for traffic forecasting: A survey[J]. *Expert Systems with Applications*, 2022, 207: Article No.117921
- [15] Zhao L, Song Y J, Zhang C, *et al.* T-GCN: a temporal graph convolutional network for traffic prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(9): 3848-3858
- [16] Yu B, Yin H T, Zhu Z X. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C] //Proceedings of the 27th International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 3634-3640
- [17] Guo S N, Lin Y F, Feng N, *et al.* Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C] //Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 922-929
- [18] Cai L, Janowicz K, Mai G C, *et al.* Traffic transformer: capturing the continuity and periodicity of time series for traffic forecasting[J]. *Transactions in GIS*, 2020, 24(3): 736-755
- [19] Zhang X Y, Huang C, Xu Y, *et al.* Traffic flow forecasting with spatial-temporal graph diffusion network[C] //Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021: 15008-15015
- [20] Shuman D I, Narang S K, Frossard P, *et al.* The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains[J]. *IEEE Signal Processing Magazine*, 2013, 30(3): 83-98
- [21] Devlin J, Chang M W, Lee K, *et al.* BERT: pre-training of deep bidirectional transformers for language understanding[C] //Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics Press, 2019: 4171-4186
- [22] Li Y G, Yu R, Shahabi C, *et al.* Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[OL]. [2022-06-08]. <https://arxiv.org/abs/1707.01926>
- [23] Pan Z Y, Zhang W T, Liang Y X, *et al.* Spatio-temporal meta learning for urban traffic prediction[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(3): 1462-1476
- [24] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C] //Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 3104-3112
- [25] Wu C H, Ho J M, Lee D T. Travel-time prediction with support vector regression[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2004, 5(4): 276-281