

基于时间感知Transformer的交通流预测方法

刘起东, 刘超越, 邱紫鑫, 高志敏, 郭帅, 刘冀钊, 符明晟

引用本文

刘起东, 刘超越, 邱紫鑫, 高志敏, 郭帅, 刘冀钊, 符明晟. [基于时间感知Transformer的交通流预测方法](#)[J]. 计算机科学, 2023, 50(11): 88-96.

LIU Qidong, LIU Chaoyue, QIU Zixin, GAO Zhimin, GUO Shuai, LIU Jizhao, FU Mingsheng. [Time-aware Transformer for Traffic Flow Forecasting](#) [J]. Computer Science, 2023, 50(11): 88-96.

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)

[基于多粒度的Transformer目标检测算法](#)

Transformer Object Detection Algorithm Based on Multi-granularity

计算机科学, 2023, 50(11): 143-150. <https://doi.org/10.11896/jsjcx.230600028>

[基于多尺度Transformer融合多域信息的伪造人脸检测](#)

Forgery Face Detection Based on Multi-scale Transformer Fusing Multi-domain Information

计算机科学, 2023, 50(10): 112-118. <https://doi.org/10.11896/jsjcx.220900048>

[基于LpTransformer网络的手语动画拼接模型](#)

Sign Language Animation Splicing Model Based on LpTransformer Network

计算机科学, 2023, 50(9): 184-191. <https://doi.org/10.11896/jsjcx.221100043>

[融合音字特征转换的非自回归Transformer中文语音识别](#)

Non-autoregressive Transformer Chinese Speech Recognition Incorporating Pronunciation-Character Representation Conversion

计算机科学, 2023, 50(8): 111-117. <https://doi.org/10.11896/jsjcx.220600144>

[基于信息熵-切分概率模型的新词发现方法](#)

New Word Detection Based on Branch Entropy-Segmentation Probability Model

计算机科学, 2023, 50(7): 221-228. <https://doi.org/10.11896/jsjcx.220700074>

基于时间感知 Transformer 的交通流预测方法

刘起东^{1,2,3} 刘超越¹ 邱紫鑫¹ 高志敏^{1,2,3} 郭帅^{1,2,3} 刘冀钊⁴ 符明晟⁵

1 郑州大学计算机与人工智能学院 郑州 450001

2 智能集群系统教育部工程研究中心 郑州 450001

3 国家超级计算郑州中心 郑州 450001

4 兰州大学信息科学与工程学院 兰州 730000

5 电子科技大学计算机科学与工程学院 成都 611731

摘要 作为智能交通系统的关键一环,交通流预测面临着长时预测不准的难题,其主要挑战在于交通流数据本身具有复杂的时空关联。近年来,Transformer 的提出使得时序数据预测的研究取得了巨大进展,但将 Transformer 应用于交通流预测仍然存在以下两个问题:1)静态的注意力机制难以捕获交通流随时间动态变化的时空依赖关系;2)采用自回归的预测方式会引发严重的误差累积现象。针对以上问题,提出了一种基于时间感知 Transformer 的交通流预测模型。首先,设计了一种新的时间感知注意力机制,可以根据时间特征定制注意力计算方案,从而更精准地反映时空依赖关系;其次,在 Transformer 的训练阶段舍弃了 Teacher Forcing 机制,并采用非自回归的预测方式来避免误差累积问题;最后,在两个真实交通数据集上进行实验,实验结果表明,所提方法可以有效捕获交通流的时空依赖,相比最优的基线方法,长时预测性能提升了 2.09%~4.01%。

关键词: 交通流预测;时空建模;时间感知注意力机制;非自回归;Transformer

中图法分类号 TP391

Time-aware Transformer for Traffic Flow Forecasting

LIU Qidong^{1,2,3}, LIU Chaoyue¹, QIU Zixin¹, GAO Zhimin^{1,2,3}, GUO Shuai^{1,2,3}, LIU Jizhao⁴ and FU Mingsheng⁵

1 School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China

2 Engineering Research Center of Intelligent Swarm Systems, Ministry of Education, Zhengzhou 450001, China

3 National Supercomputing Center in Zhengzhou, Zhengzhou 450001, China

4 School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China

5 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract As a key part of intelligent transportation systems, traffic flow forecasting faces the challenge of long-term prediction inaccuracy. The key factor is that the traffic flow has complicated spatial and temporal correlations. Recently, the emerging success of Transformer has shown promising results in time series analysis. However, there are two obstacles when applying Transformer to traffic flow forecasting: 1) it's difficult for the static attention mechanisms to capture the dynamic changes of traffic flow along the space and time dimensions; 2) the autoregressive decoder in transformer could cause error accumulation problem. To address the above problems, this paper proposes a time-aware Transformer (TAformer) for traffic flow forecasting. Firstly, it proposes a time-aware attention mechanism that can customize attention calculation solution according to the time features, so as to estimate the spatial and temporal dependencies more accurately. Secondly, it discards the teacher forcing mechanism during the training phase and proposes a non-autoregressive inference method to avoid the problem of error accumulation. Finally, extensive experiments on two real traffic datasets show that the proposed method can effectively capture the spatial-temporal dependence of traffic flow. Compared with the state-of-the-art baseline method, the proposed method improves the performance of long-term prediction by 2.09%~4.01%.

Keywords Traffic flow Forecasting, Spatial-Temporal modeling, Time-aware attention, Non-autoregressive, Transformer

到稿日期:2022-10-24 返修日期:2023-05-11

基金项目:国家自然科学基金(62276238, 61906174, 62036010);中国博士后科学基金资助项目(2022T150590, 2020M672275);河南省自然科学基金(232300421095);河南省重点研发与推广专项(222102210248)

This work was supported by the National Natural Science Foundation of China(62276238, 61906174, 62036010), China Postdoctoral Science Foundation(2022T150590, 2020M672275), Natural Science Foundation of Henan Province, China(232300421095) and Henan Provincial Key Science and Technology Research Projects(222102210248).

通信作者:刘起东(ieqdliu@zzu.edu.cn)

1 引言

交通流预测是智能交通系统(Intelligent Transportation Systems, ITS)的核心组成部分,是实现高效交通管理的关键技术之一。准确预测道路未来的交通状况,不仅有助于交通管理部门更加高效地引导车流、疏散拥堵、降低交通事故的影响程度,还可以为车辆导航、路线规划,甚至自动驾驶提供额外的数据支持。然而,交通流的准确预测是一项非常具有挑战性的工作,因为一个路网节点未来的交通状况(例如交通速度、交通流量和交通密度)不仅取决于当前节点的历史数据,还取决于其邻居节点的信息。因此,如何在复杂的交通数据中高效地捕获节点之间的时空依赖非常重要。

随着人工智能技术的发展,越来越多的深度学习模型被应用于交通流预测并取得了里程碑式的突破^[1-2],例如图神经网络^[3-4]被用来捕捉空间依赖性,而循环神经网络(Recurrent Neural Network, RNN)^[5-6]则被用来捕捉时间依赖性。考虑到 RNN 存在长序列记忆问题并且无法并行化操作,基于 Attention 机制的 Transformer^[7]受到了越来越多研究者的青睐^[8-11]。本文对相关技术进行梳理,发现现有基于 Transformer 的工作仍然存在以下的一个或多个问题。1) Transformer 模型中采用的自回归生成方式引发了误差累积现象。如图 1 所示,Transformer 在训练时采用 Teacher Forcing 机制^[12](即解码器的初始输入是真实标签),而在推理过程中则自回归地逐步生成序列(即将上一步的输出作为下一步的输入)。这种训练和推理时的一致性导致了暴露偏差(Expo-

sure Bias)^[13],结合 Transformer 串行的推理过程,致使误差随预测步长的增加而逐步积累。虽然计划采样(Scheduled Sampling)^[14]可以在一定程度上缓解上述问题,但交通流预测这种回归型任务的输出(如交通流速)通常为连续型数值,比离散型的标签(如单词)对误差更敏感,结合 Transformer 的自回归生成方式,进一步助推了误差累积现象。2) 基于单一参数的注意力机制并不能捕捉交通流“时间敏感”的动态时空依赖关系。如图 2 所示,早上 8 点(上班时间)A 节点的流量大部分来自于节点 B(居民区),而下午 6 点(下班时间)A 节点的流量大部分来自于节点 C(高新技术开发区)。此外, A 节点的交通流速曲线在早上八点(早高峰)呈“V”字形,在中午 12 点(平稳时段)则十分平稳。总而言之,交通流的时空模式会随时间动态变化,并具有一定的周期性。现有的注意力机制在所有时刻往往都只采用同一套参数,即使我们可以将时间信息作为输入特征,但并不能从根本上解决时间敏感的注意力问题。

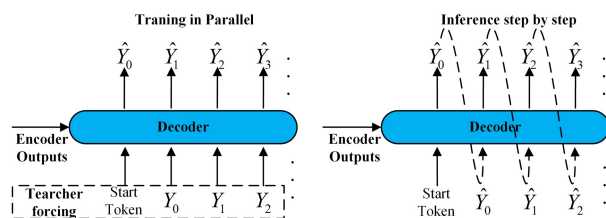


图 1 Transformer 的训练和推理过程

Fig. 1 Training and inference processes of Transformer

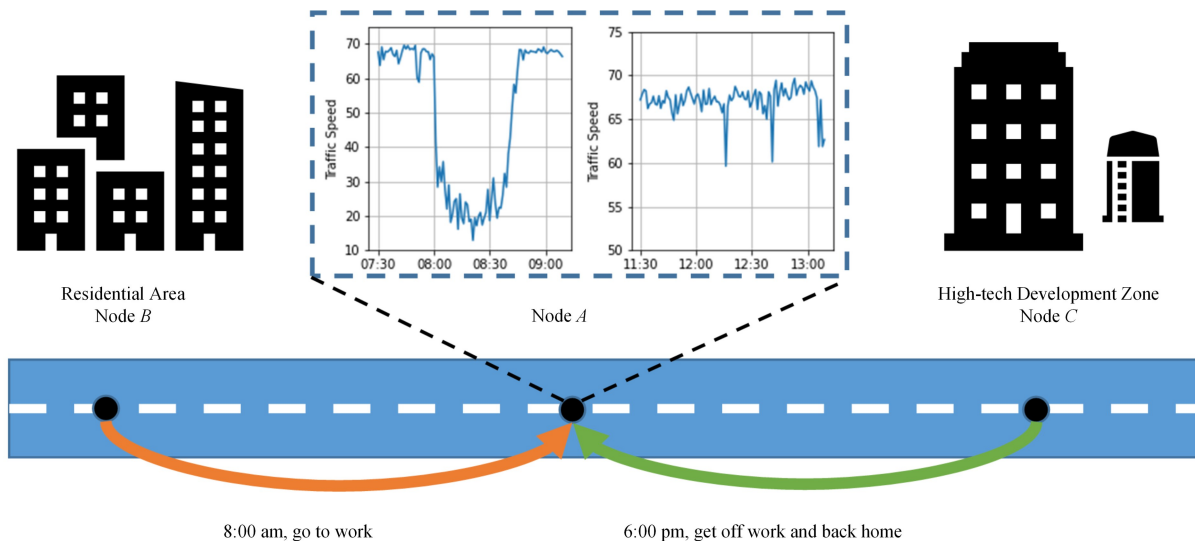


图 2 交通流注意力随时空动态变化的例子

Fig. 2 Example of traffic flow attention changing dynamically over time and space

针对上述问题,我们提出了一种时间感知的 Transformer 模型,用于交通流预测。本文的主要贡献如下:

1) 提出了一种时间感知的注意力机制(Time-Aware Attention, TAA),可以根据时序特征定制注意力计算方案,从而建模不同时刻下的时空依赖关系。

2) 基于时间感知注意力机制,提出了一种新的 Transformer 模型。通过采用非自回归的预测方式并行地生成目标序列,可以避免误差累积现象,从而具有更好的长时预测

性能,同时具有更快的推理速度。

3) 在两个真实的交通数据集上做了广泛的实验,相比最优的基线方法,本文方法在长时预测性能上提升了 2.09%~4.01%。

2 相关工作

近年来,图神经网络飞速发展,代表性的工作有 ChebNet^[3]、图卷积网络(Graph Convolutional Network, GCN)^[4]、

Graphsage^[15], 图注意力网络 (Graph Attention Network, GAT)^[16]等。这些方法可以在非欧几里得空间内进行信息融合, 并被广泛地应用在知识图谱、推荐系统等领域^[17-18]。由于道路网络也具有非欧的拓扑结构(图), 有工作开始将图神经网络引入到交通流预测任务中, 进行空间结构的建模^[19]。如 Yu 等^[20]使用 ChebNet 和时间门控卷积构建了 ST-Conv 块, 可以从历史交通流中联合提取时空特征。Zhao 等^[21]则结合图卷积网络和门控循环单元 (Gated Recurrent Unit, GRU)^[6]进行交通流预测。但这些方法都将道路网络看作一张无向图。而现实世界中往往有许多道路是单向通行的。针对此问题, Li 等^[22]将交通网络建模为有向图, 并结合扩散卷积和循环神经网络提出了扩散卷积循环神经网络 (Diffusion Convolutional Recurrent Neural Network, DCRNN)。

以上方法虽然可以在一定程度上提取空间特征, 但都依赖于预先构建好的道路拓扑图, 这种显式的静态图结构无法反映潜在的依赖关系和高度动态的时空模式。因此, 一些学者开始采用可学习的、动态的图结构来建模空间依赖。Wu 等^[23]提出了图波网络 (Graph WaveNet, GWN) 来建模交通流的时空依赖, 该网络结合了扩散卷积和时间膨胀卷积, 并引入了一种基于节点嵌入的自适应依赖矩阵, 可以提取到节点间隐含的依赖关系。Zhang 等^[24]则提出了一种灵活的图学习机制, 即结构学习卷积 (Structure Learning Convolution, SLC), 可以学习到潜在和动态的图结构信息, 大幅度地提升了交通流预测的精度。Li 等^[1]提出了一种新的时空融合图神经网络, 使用了一种数据驱动的时空图构建方法。此外, 还有一些基于 Attention 机制的动态图方法。例如 Zhang 等^[25]通过组合图注意力网络和门控循环网络, 成功捕获了交通流动态的空间相关性。Guo 等^[26]设计了基于 Attention 机制的时空图卷积网络 (Attention Based Spatial-Temporal Graph Convolutional Networks, ASTGCN), 在时空卷积的基础上额外引入了时空注意力机制, 并通过“近期、日周期、周周期”3 个并行的组件联合地提取交通流的时间模式。Zheng 等^[27]提出了图多头注意力网络 (Graph Multi-Attention Network, GMAN), 通过堆叠的时空注意力块 (ST-Attention Blocks) 来预测未来的交通流特征。

考虑到交通流预测是一种特殊的序列到序列 (Sequence to Sequence, Seq2Seq) 型任务, 最近的一些工作开始采用 Transformer 架构, 以利用其强大的序列建模能力。Cai 等^[8]先使用扩散卷积提取空间特征, 然后将其交由 Transformer 建模时间特征。Park 等^[9]采用类似于 Transformer 的编解码器结构, 通过节点级注意力机制和时间注意力机制捕获时空特征。类似地, Xu 等^[28]也采用时空注意力机制建模时空依赖, 仅构建了一个类 Transformer 的编码器, 以双层卷积代替编解码器生成目标序列。

在以上基于 Transformer 的工作中, 有一部分工作在生成目标序列时仍采用自回归方式, 这会导致较为严重的误差累积, 不利于长时预测。另一部分工作则以简单网络 (如卷积) 替代解码器, 不能充分发挥 Transformer 解码器的序列

生成能力。此外, 它们采用的注意力机制总以相同的参数建模时空依赖, 即使交通流在不同时刻会表现出不同的时空模式。以上问题制约了交通流预测效果的进一步提升, 为此, 我们采用非自回归生成来避免误差累积, 并提出了一种新的时间感知注意力机制, 来进行时间敏感的时空依赖建模。

3 时间感知 Transformer

3.1 问题定义

交通网络可以抽象为一张拓扑图 $\mathcal{G} = (V, \mathcal{E}, A)$, 其中 V ($|V| = N$) 表示所有节点 (传感器) 的集合, \mathcal{E} 为所有边的集合, $A \in \mathbb{R}^{N \times N}$ 表示加权邻接矩阵 (存储了任意两个传感器在位置上的接近程度)。

设输入 $X_t \in \mathbb{R}^{N \times d_s}$ 表示 N 个节点在 t 时刻观察到的交通流特征 (例如交通流速、交通流量、交通密度等), $Z_t \in \mathbb{R}^{d_z}$ 表示 t 时刻对应的时序特征 (例如周三 12 时 30 分)。其中, d_s 和 d_z 分别表示交通流特征和时序特征的维度。交通流预测就是在给定道路拓扑图 \mathcal{G} 以及时间特征 $Z = [Z_{t-T+1}, \dots, Z_t, \dots, Z_{t+T}]$ 的前提下, 通过挖掘过去 T 个时刻的数据 $X = [X_{t-T+1}, \dots, X_t]$ 来预测未来 T' 个时刻的交通状况, 即:

$$[X_{t+1}, \dots, X_{t+T'}] = f(X, \mathcal{G}, Z) \quad (1)$$

其中, $f(\cdot)$ 是要求解的映射函数。我们的目的是训练模型来近似函数 $f(\cdot)$, 从而使以下目标函数的值最小。

$$loss = \sum_{i=t+1}^{t+T'} \sum_{j=1}^N |X_{i,j} - \hat{X}_{i,j}| \quad (2)$$

其中, $X_{i,j}$ 代表第 j 个节点在 i 时刻的交通流真值, $\hat{X}_{i,j}$ 代表模型 $f(\cdot)$ 的预测值。不失一般性, 本文只考虑交通流速预测, 交通流量和交通密度的预测方式同理。

3.2 时间感知注意力机制

交通信息具有明显的时间规律, 例如早上 8 点与下午 6 点的交通流对周围邻居具有不同的依赖关系。因此, 对所有时间段采用同一套参数进行注意力计算会造成建模不准。虽然现有的一些工作^[29-30]已经开始解决这个问题, 例如 Liu 等^[31]提出了一种 Aspect-Aware 注意力机制, 其中每一种类型的边作为一个“Aspect”并设置独立的参数, 在异质图神经网络学习中获得比较理想 (State-of-The-Art, SOTA) 的结果。但是, 据我们所知, 类似时间这种连续特征的注意力建模在学术界仍然是一片空白。基于此, 本文提出了一种基于时间感知的注意力机制 (TAA), 可以根据时间特征“定制”注意力计算方案。

令 $Q \in \mathbb{R}^{N_q \times d_h}$, $K \in \mathbb{R}^{N_k \times d_h}$ 和 $V \in \mathbb{R}^{N_v \times d_h}$ 分别代表注意力机制的查询 (Query)、键 (Key) 和值 (Value) 矩阵, 其中 N_q 和 N_k 分别为查询和键序列的长度, d_h 代表模型的隐层特征维度, 则时刻 t 的注意力 Att 的计算式如式 (3) 所示:

$$Att(Q, K, V, Z_t) = \text{softmax} \left(\frac{[QW_q]C^T[KW_k]^T}{\sqrt{\sum_{i=1}^{d_h} [C_{t,i}^2]}} \right) VW_v \quad (3)$$

其中:

$$C^T = \text{diag}(MLP(Z_t, \theta)) \quad (4)$$

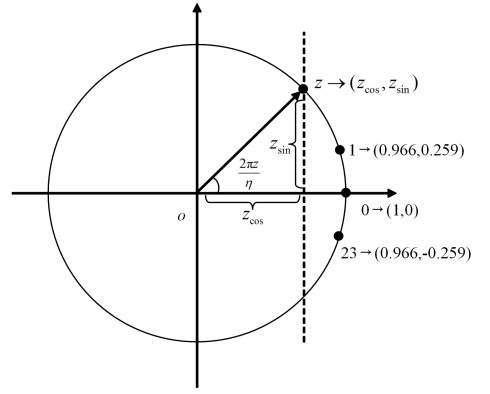
其中, $W_q \in \mathbb{R}^{d_h \times d_h}$, $W_k \in \mathbb{R}^{d_h \times d_h}$ 和 $W_v \in \mathbb{R}^{d_h \times d_h}$ 都是可学习的

参数; $diag(\cdot)$ 代表以对角线元素向量构建对角矩阵; MLP 为多层感知机, θ 为多层感知机的参数。式(4)的目的就是根据时序特征 $Z_t \in \mathbb{R}^{d_z}$ 得到对角矩阵 $C^t (C^t \in \mathbb{R}^{d_h \times d_h})$, 从而控制注意力系数的计算。式(3)的分母 $\sqrt{\sum_{i=1}^{d_h} [C_{i,i}^t]^2}$ 为缩放因子, 其作用是限制 softmax 的输入, 避免因为输入值过大而落入梯度极小的区间^[7]。

此外, 时间特征 Z_t 具有周期性, 例如 23 点之后是 0 点, 0 点之后是 1 点, 但 0 与 23 及 0 与 1 在数值上的差异却很大。为此, 引入空间转换函数 φ 来消除时间特征数值上的差异, 具体为:

$$(z_{\cos}, z_{\sin}) = \varphi(z, \eta) = \left(\cos\left(\frac{2\pi z}{\eta}\right), \sin\left(\frac{2\pi z}{\eta}\right) \right) \quad (5)$$

其中, z 为 Z_t 的一个因子, η 为 z 对应的周期。例如, 当 z 代表“小时”时, $\eta=24$ 并且 $z \in [0, 23]$ 。当 z 代表“分钟”时, $\eta=60$ 并且 $z \in [0, 59]$ 。空间转换函数 φ 就是将输入特征 z 映射到 (z_{\cos}, z_{\sin}) 这两个分量上, 使时间特征的周期性衔接更加平滑, 如图 3 所示。



注: 在新空间中, 早上 1 点和晚上 23 点到凌晨的欧几里得距离相同。

图 3 空间转换

Fig. 3 Spatial transformation

3.3 模型架构

针对 Transformer 进行交通流预测时存在的误差累积问题, 本文设计了一种新的交通流预测模型 (Time-Aware Transformer, TAformer), 如图 4 所示。

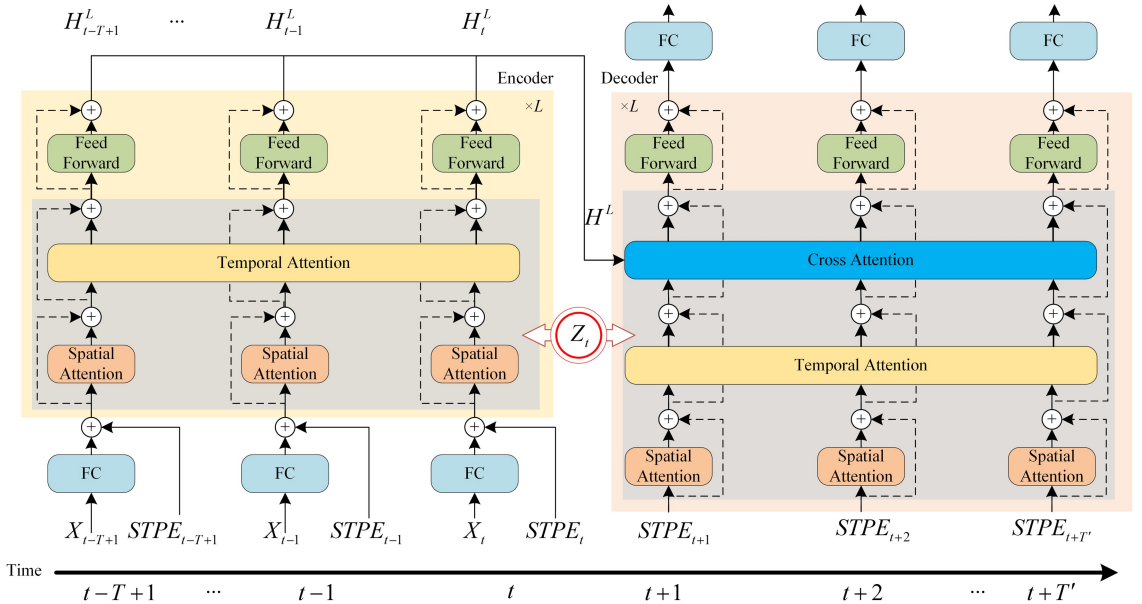


图 4 TAformer 的整体架构

Fig. 4 Overall architecture of TAformer

3.3.1 编码器

编码器的目的是将窗口大小为 T 的历史序列数据映射到一个特殊维度的空间, 以便解码器推理和预测。与 Transformer^[7] 类似, TAformer 也会在输入序列中加上位置编码以识别各分量间的顺序关系, 具体为:

$$X_t' = FC(X_t) + STPE_t \quad (6)$$

其中, $STPE_t \in \mathbb{R}^{N \times d_h}$ 表示 N 个节点在 t 时刻的时空位置编码; FC 表示一个全连接层, 目的是将原始特征 $X_t \in \mathbb{R}^{N \times d_x}$ 映射到与时空位置编码相同的特征空间。令 $STPE_{t,i}$ 表示第 i 个节点在 t 时刻的时空位置编码, 则有:

$$STPE_{t,i} = [TPE_t \parallel SPE_i]W \quad (7)$$

其中, $SPE_i \in \mathbb{R}^{d_h}$ 表示第 i 个节点的空间位置编码, $TPE_t \in$

\mathbb{R}^{d_t} 表示 t 时刻的时间位置编码, \parallel 表示拼接 (Concatenate) 操作, $W \in \mathbb{R}^{(d_t+d_h) \times d_h}$ 为可学习的参数。关于 SPE_i 和 TPE_t , 本文将在下面进行详细介绍。

在传统 Transformer 模型中, 位置编码与时刻 t 相关, 通过正余弦函数生成位置编码^[7]。令 $TPE_t \in \mathbb{R}^{d_t}$ 表示时间位置编码, 则该位置编码的第 $k \in \{1, \dots, d_t\}$ 个因子 $TPE_{t,k}$ 可表示为:

$$TPE_{t,k} = \begin{cases} \sin\left(\frac{t}{10000^{k/2}}\right), & k \bmod 2 = 0 \\ \cos\left(\frac{t}{10000^{(k-1)/2}}\right), & k \bmod 2 = 1 \end{cases} \quad (8)$$

交通流数据在任意时刻 t 的输入不是一个点, 而是一张图, 因此在建模时不仅要考虑时间顺序信息, 还要考虑

空间位置信息。本文采用 Node2Vec^[32] 在路网拓扑图 G 上进行无监督学习,从而得到所有节点的空间位置编码 $SPE \in \mathbb{R}^{N \times d_s}$ 。

$$SPE = \text{Node2vec}(G) \quad (9)$$

编码器由 L 个串联的结构相同的层组成。令 $H^l = [H_{t-T+1}^l, \dots, H_t^l]$ 代表编码器第 l 层的特征表示, $H_t^l \in \mathbb{R}^{N \times d_h}$ 代表第 t 时刻 N 个节点的特征表示, $H_t^0 = X_t'$ 。如图 4 所示,通过组合空间注意力模块和时间注意力模块来联合建模时空依赖。编码器(解码器)堆叠了 L 个结构相同的编码器层(解码器层),每层的空间注意力模块、时间注意力模块以及交叉注意力模块均采用本文提出的时间感知注意力机制,其中 Z_t 为 t 时刻的时间特征,模块与模块之间均采用残差连接,如图 4 中虚线所示。前馈层是一个采用 Relu 作为激活函数的两层全连接网络。为了缓解模型较深时可能出现的梯度弥散和网络退化问题^[33],各模块之间都添加了残差连接^[34]。

接下来将简要介绍时间注意力模块(记作 Att^t),空间注意力模块(记作 Att^s)将在 3.3.3 节中介绍。

令 $S_t = Att^s(H_t^l)$,时间注意力模块 Att^t 包括以下 3 个步骤:

1) 合并过去时间片段的 S_t 为一个三维张量,其记作 $\mathbf{S} \in \mathbb{R}^{T \times N \times d_h}$,即令 $\mathbf{S} = [S_{t-T+1}, \dots, S_t]$ 。

2) 对三维张量 \mathbf{S} 进行转置: $\mathbf{S} \in \mathbb{R}^{T \times N \times d_h} \rightarrow \mathbf{S} \in \mathbb{R}^{T \times N \times d_h}$ 。

3) $\forall i \in \{1, \dots, N$,采用式(3)的时间感知注意力机制在时间维度上执行聚合操作,即 $U_i = Att^t(Q_i, K_i, V_i, Z_t)$ 。其中, $Q_i = K_i = V_i = S_i$ 。

将上文得到的 $U_i (\forall i \in \{1, \dots, N$) 合并并转置之后输入前馈层中,以获得第 $l+1$ 层的隐藏表示 H^{l+1} 。编码器第 L 层的输出 $H^L \in \mathbb{R}^{T \times N \times d_h}$ 将用于后面解码器的推理和预测。

3.3.2 解码器

解码器与编码器的结构相似。令 $\tilde{H}^l = [\tilde{H}_{t+1}^l, \dots, \tilde{H}_{t+T}^l]$ 代表解码器第 l 层的特征表示。为了避免暴露偏差,本文在训练过程中舍弃了 Teacher Forcing 机制,在测试阶段采用非自回归的预测方式。因此,解码器的初始输入仅包含相应的时空位置编码,即:

$$\tilde{H}^0 = [STPE_{t+1}, \dots, STPE_{t+T}] \quad (10)$$

为了更精准地提取编码器中隐藏的历史信息,在解码器中采用时间感知的交叉注意力模块(记作 Att^c)来进行预测推理。具体步骤如下:

1) 依次通过空间注意力模块和时间注意力模块获得 $\tilde{Q} \in \mathbb{R}^{T \times N \times d_h}$,即:

$$\tilde{Q} = Att^s[Att^t(\tilde{H}_{t+1}^l), \dots, Att^t(\tilde{H}_{t+T}^l)] \quad (11)$$

2) 将编码器的输出 H^L 作为 Att^c 的键和值,即令 $\tilde{K} = \tilde{V} = H^L$ 。

3) 对 \tilde{Q}, \tilde{K} 和 \tilde{V} 进行转置操作,即:

$$\tilde{Q} \in \mathbb{R}^{T \times N \times d_h} \rightarrow \tilde{Q} \in \mathbb{R}^{T \times N \times d_h} \quad (12)$$

$$\tilde{K}, \tilde{V} \in \mathbb{R}^{T \times N \times d_h} \rightarrow \tilde{K}, \tilde{V} \in \mathbb{R}^{T \times N \times d_h}$$

4) $\forall i \in \{1, \dots, N$,基于式(3)的时间感知注意力机制,

从交通流的历史特征中提取有用信息,即 $\tilde{U}_i = Att^c(\tilde{Q}, \tilde{K}_i, \tilde{V}_i, Z_t)$ 。

我们将解码器第 L 层的输出 \tilde{H}^L 输入到一个全连接网络 FC 中,获得最终的预测值 \hat{X} ,即:

$$\hat{X} = FC(\tilde{H}^L) \quad (13)$$

值得注意的是,TAformer 在训练阶段舍弃了 Teacher Forcing 机制,在推理阶段采用非自回归的方式生成目标序列,该方式的优点如下:1)可以减少暴露偏差,避免出现误差累积现象,从而提高长时预测能力;2)高度并行的序列生成方式大大加快了模型的推理速度。4.5.2 节的实验结果验证了本文的观点。

3.3.3 空间注意力模块

一般来说,道路的邻接矩阵 \mathbf{A} 蕴藏着丰富的信息,并将 $D^{-1/2} \mathbf{A} D^{-1}$ 作为核函数。然而,越来越多的研究表明只考虑道路的拓扑结构是不够的。通过对全局信息进行自适应的空间建模可以补充邻接矩阵 \mathbf{A} 的有限表达,有助于提升模型性能^[35]。基于以上分析,本文的空间注意力模块(Att^s)将由两部分组成:拓扑信息和补充信息。令 $h_i^l \in H_t^l$ 表示第 i 个节点在 t 时刻 l 层的特征表示,那么空间注意力模块层与层之间的更新方式如式(14)所示:

$$h_i^{l+1} = \sum_{j \in \mathcal{N}_i} \alpha_{i,j}^{\text{top}} h_j^l W_1 + \sum_{j \in \mathcal{V}} \alpha_{i,j}^{\text{sup}} h_j^l W_2 \quad (14)$$

其中, \mathcal{V} 为全部节点集; \mathcal{N}_i 代表节点 i 的邻域节点集合; $W_1 \in \mathbb{R}^{d_h \times d_h}$ 和 $W_2 \in \mathbb{R}^{d_h \times d_h}$ 为第 l 层的可学习参数; $\alpha_{i,j}^{\text{top}}$ 和 $\alpha_{i,j}^{\text{sup}}$ 则分别表示拓扑信息注意力系数和补充信息注意力系数,其计算式如下:

$$\alpha_{i,j}^{\text{top}} = \frac{\exp(\beta_{i,j}^{\text{top}})}{\sum_{j \in \mathcal{N}_i, U_i} \exp(\beta_{i,j}^{\text{top}})} \quad (15)$$

$$\alpha_{i,j}^{\text{sup}} = \frac{\exp(\beta_{i,j}^{\text{sup}})}{\sum_{j \in \mathcal{V}} \exp(\beta_{i,j}^{\text{sup}})} \quad (16)$$

其中,

$$\beta_{i,j}^{\text{top}} = \frac{[h_i^l W_q^{\text{top}}] C [h_j^l W_k^{\text{top}}]^T}{\sqrt{\sum_{d=1}^{d_h} [C_{d,d}]^2}} + \gamma^{\text{top}} \tilde{L}_{i,j} \quad (17)$$

$$\beta_{i,j}^{\text{sup}} = \frac{[h_i^l W_q^{\text{sup}}] C [h_j^l W_k^{\text{sup}}]^T}{\sqrt{\sum_{d=1}^{d_h} [C_{d,d}]^2}} + \gamma^{\text{sup}} \tilde{L}_{i,j} \quad (18)$$

其中, $W_q^{\text{top}}, W_k^{\text{top}}, W_q^{\text{sup}}, W_k^{\text{sup}} \in \mathbb{R}^{d_h \times d_h}$ 及 γ^{top} 和 γ^{sup} 都是可学习的参数; 对角矩阵 $\mathbf{C} \in \mathbb{R}^{d_h \times d_h}$ 通过式(6)计算得到; 而 $\tilde{\mathbf{L}}$ 为图 G 重归一化的拉普拉斯矩阵^[4], 即 $\tilde{\mathbf{L}} = \mathbf{D}_0^{-1/2} \mathbf{A} \mathbf{D}_1^{-1/2}$, \mathbf{D}_0 和 \mathbf{D}_1 分别表示 \mathbf{A} 的出度矩阵和入度矩阵。通过引入 $\tilde{L}_{i,j}$ 可以进一步捕获道路拓扑图 G 中边的权重信息和方向信息^[9]。

4 实验

4.1 数据集描述

为了验证本文方法的有效性,本文在两个公开的真实交通数据集^[22]上进行了实验,这两个数据集的描述如下。

METR-LA 数据集包含了从 2012 年 3 月 1 日至 2012 年 6 月 30 日这 4 个月内,由洛杉矶县高速公路上 207 个传感器

收集的交通流速数据。

PEMS-BAY 数据集包含了从 2017 年 1 月 1 日至 2017 年 5 月 31 日这 5 个月内,由海湾地区的 325 个环路传感器收集的交通流速数据。

METR-LA 和 PEMS-BAY 数据集都以 5min 的时间间隔汇总交通流速数据。表 1 列出了这两个数据集的详细统计信息。

表 1 本文数据集的统计信息

Table 1 Statistics of datasets used in this paper

Datasets	Nodes	Edges	Samples
METR-LA	207	1515	34249
PEMS-BAY	325	2369	52093

4.2 基线方法

本文选择了一批具有代表性的算法作为基线,这些方法已经被广泛应用于交通流预测任务。

1) HA(Historical Average Model)^[36]:历史平均模型。

2) ARIMA_{kal}(Autoregressive Integrate Moving Average Model with Kalman filter)^[37]:基于卡尔曼滤波器的差分自回归移动平均模型。

3) FC-LSTM(Fully Connected Long Short-Term Memory unit)^[38]:全连接长短时记忆单元构成的循环神经网络,仅建模时间依赖。

4) DCRNN(Diffusion Convolutional Recurrent Neural Network)^[22]:扩散卷积循环神经网络,集成了扩散卷积和门控循环单元。

5) STGCN(Spatio-Temporal Graph Convolutional Network)^[20]:时空图卷积网络,结合了图卷积和一维时间卷积。

6) GWN(Graph WaveNet)^[23]:图波网络,采用扩散卷积和时间膨胀卷积。

7) MRA-BGCN(Multi-Range Attentive Bicomponent GCN)^[39]:多范围注意的双分量图卷积网络。

8) ST-GRAT(Spatio-Temporal Graph Attention)^[9]:采用自回归式 Transformer 架构的时空图注意力神经网络。

4.3 实验环境和参数设置

本文按照 7:1:2 的比例划分训练集、验证集和测试集,采用 z-score 标准化,并以过去 $T=12$ 个时间片(60 min)的历史交通流来预测未来 $T'=12$ 个时间片(60 min)的交通流速。采用 DCRNN^[22]中基于阈值高斯核的方式生成道路拓扑图 \mathcal{G} 。这些设置与其他用于对比的基线方法保持一致。所有的实验都在一台具有一块 NVIDIA Quadro RTX 4000 GPU 和 16 个 Core i9-11900K CPU 核的设备上进行。本文还使用了混合精度训练^[40]和梯度检查点^[41]技术来减小训练过程中的显存占用。本文采用 Adam 优化器^[42]进行训练,初始学习率设为 0.005,并在训练 20 轮后衰减为原来的 1/10。权重衰减设为 1×10^{-5} ,批次大小设置为 64,共训练 30 轮。编解码器层堆叠的层数 $L=3$,模型的特征维度 $d_h=d_t=d_s=32$ 。

4.4 实验结果与分析

本文选择平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Squared Error, RMSE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)作为预测性能的评价指标,这些指标越低代表模型的预测性能越好。表 2 和表 3 分别列出了各类方法在 METR-LA 数据集和 PEMS-BAY 数据集上的预测性能。

表 2 METR-LA 数据集上的实验结果

Table 2 Experimental results on METR-LA Dataset

Model	15 min			30 min			60 min		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
HA	4.16	7.80	13.0	4.16	7.80	13.0	4.16	7.80	13.0
ARIMA _{kal}	3.99	8.12	9.6	5.15	10.45	12.7	6.90	13.23	17.4
FC-LSTM	3.44	6.30	9.6	3.77	7.23	10.9	4.37	8.69	13.2
DCRNN	2.77	5.38	7.3	3.15	6.45	8.8	3.60	7.60	10.5
STGCN	2.88	5.74	7.6	3.47	7.24	9.6	4.59	9.40	12.7
GWN	2.69	5.15	6.9	3.07	6.22	8.4	3.53	7.37	10.0
MRA-BGCN	2.67	5.12	6.8	3.06	6.17	8.3	3.49	7.30	10.0
ST-GRAT	2.60	5.07	6.6	3.01	6.21	8.2	3.49	7.42	10.0
TAformer	2.52	5.17	6.5	2.90	6.14	7.9	3.35	7.13	9.5

表 3 PEMS-BAY 数据集上的实验结果

Table 3 Experimental results on PEMS-BAY Dataset

Model	15 min			30 min			60 min		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
HA	2.88	5.59	6.8	2.88	5.59	6.8	2.88	5.59	6.8
ARIMA _{kal}	1.62	3.30	3.5	2.33	4.76	5.4	3.38	6.50	8.3
FC-LSTM	2.05	4.19	4.8	2.20	4.55	5.2	2.37	4.69	5.7
DCRNN	1.38	2.95	2.9	1.74	3.97	3.9	2.07	4.74	4.9
STGCN	1.46	3.01	2.9	2.00	4.31	4.1	2.67	5.73	5.4
GWN	1.30	2.74	2.7	1.63	3.70	3.7	1.95	4.52	4.6
MRA-BGCN	1.29	2.72	2.9	1.61	3.67	3.8	1.91	4.46	4.6
ST-GRAT	1.29	2.71	2.7	1.61	3.69	3.6	1.95	4.54	4.6
TAformer	1.30	2.78	2.7	1.60	3.66	3.6	1.87	4.33	4.3

1) TAformer 在 30 min 和 60 min 时的预测性能优于所有

的基线方法,特别是在提前 60 min 预测时,TAformer 相比最

优的基线方法,在两个数据集上的平均绝对误差分别降低了 2.09% 和 4.01%。这说明 TAformer 擅长进行长时预测,并且可以有效地建模复杂的时空依赖。

2) TAformer 在 15 min 时的预测结果只比最优的基线方法稍差,这是非自回归的预测方式造成的,通过略微牺牲短时预测精度,来换取中长时预测性能的提升。

3) 基于深度学习的方法普遍优于基于传统统计学的方法 (HA, ARIMA_{kal}),这说明深度学习方法在建模高度非线性的交通流时更有效。同时建模时空依赖的方法又普遍优于只建模时间依赖的方法 (FC-LSTM),说明了空间依赖的重要性。

4) 特别地,在交通状况更加复杂的 METR-LA 数据集上,动态建模空间依赖的方法 (ST-GRAT, TAformer) 明显优于其他基于静态图结构的方法 (DCRNN, STGCN, GWN),这是因为动态图方法可以学习到动态变化的图结构,相比静态图方法拥有更大的表示空间,因此可以支撑更加复杂的时空依赖建模。

4.5 消融实验

本节针对 TAformer 中的时间感知注意力机制、非自回归生成和其他关键组件 (空间注意力层的各个组件以及时空

编码等)进行了消融实验,通过观察模型在替换或删除组件后的性能变化,验证了这些组件的有效性。

4.5.1 时间感知注意力机制的消融实验

为了评估 TAA 以及空间转换函数 φ 对模型性能的影响,我们测试了 3 种注意力机制: TAA, SDPA 以及 TAA-no- φ 。具体来说, TAA 即第 3.2 节中描述的时间感知注意力机制,也是 TAformer 所采用的方法; TAA-no- φ 是移除空间转换函数 φ 后的 TAA 机制; SDPA (Scaled Dot Product Attention) 则是经典的缩放点积注意力机制^[7]。

3 种注意力机制的实验结果如图 5 所示,可以看到在 3 个预测时段下, TAA 和 TAA-no- φ 的预测性能都显著优于 SDPA,例如提前 60 min 预测时, TAA 和 TAA-no- φ 在 METR-LA 数据集上的平均绝对误差分别比 SDPA 小 0.08 和 0.05,在 PEMS-BAY 数据集上则分别小 0.05 和 0.04,这说明时间感知注意力机制可以更好地捕获交通流随时间动态变化的时空模式。此外,我们还观察到 TAA 的性能总是略优于 TAA-no- φ ,这说明引入空间转换函数 φ 确实有利于提升预测性能,因为它可以较好地平滑时序特征的周期性衔接。

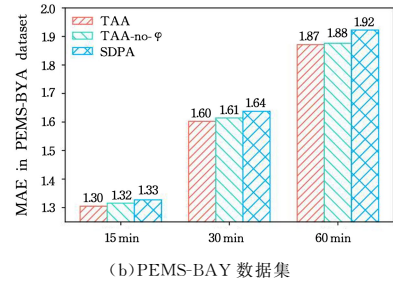
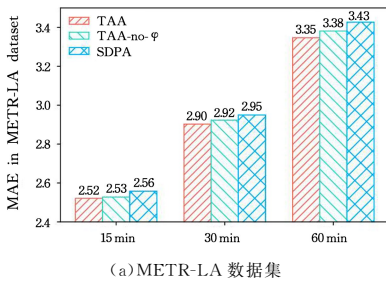


图 5 3 种注意力机制在两个数据集上的性能对比

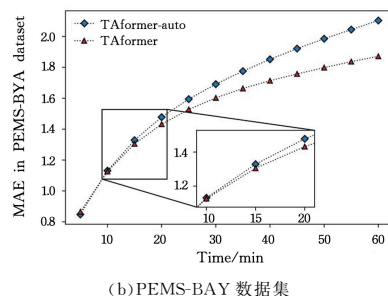
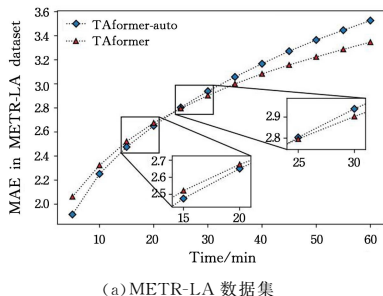
Fig. 5 Performance of three attention mechanism on two datasets

4.5.2 非自回归生成的消融实验

为了研究非自回归生成在模型中的作用,我们保持 TAformer 的组件和结构不变,换用自回归方式预测交通流序列,自回归的 TAformer 记作 TAformer-auto。图 6 给出了两种方法的平均绝对误差在不同预测时段下的变化曲线。

首先观察到,模型的预测误差会随着预测时段的增加而增大,这是合理的,因为长期预测的难度明显大于短期预测。其次, TAformer-auto 在中长期 (大于 15 min) 的预测性能明显

差于 TAformer,并且随着预测时段的增加,这种差异越发明显。这说明前者在预测过程中发生了误差累积现象,严重影响了长期预测的性能,而后者由于采用了非自回归生成,可以避免误差累积,对长时预测更加友好。我们还注意到, TAformer-auto 在短期预测 (15 min) 下的效果相对较好,这可能是因为训练时采用了 Teacher Forcing 机制,模型总在知晓前一时间刻真实标签的基础上进行预测,相当于其优化目标专注于单步预测,且短期预测的误差累积现象相对不明显。



(a) METR-LA 数据集

(b) PEMS-BAY 数据集

图 6 TAformer 和 TAformer-auto 的性能对比

Fig. 6 Performance comparison between TAformer and TAformer-auto

此外,与自回归方式相比,非自回归方式在推理时也能并行地生成预测序列,因此可以显著加快推理速度。表 4 列出了

TAformer 和 TAformer-auto 在 PEMS-BAY 数据集上训练和推理时的速度差异,可以发现:两者在训练时的运行速度接

近,但在推理时,TAformer 相比 TAformer-auto 运行时间缩短了 75.09%。这表明非自回归的 TAformer 在实际应用时会展现出更好的实时性。

表 4 在 PEMS-BAY 数据集上的运行速度

Table 4 Running speed on PEMS-BAY dataset

(单位:(s/epoch))

Model	Training	Inference
TAformer	608.44	46.77
TAformer-auto	615.33	187.73

4.5.3 其他组件的消融实验

本文还设计了以下几种 TAformer 的变体,以验证模型中其他组件的有效性。

1) w/o Topology: 移除空间注意力模块中局部建模的拓扑信息。

2) w/o Supplement: 移除空间注意力模块中全局建模的

补充信息。

3) w/o Laplace: 移除空间注意力模块中引入的先验拉普拉斯矩阵 \tilde{L} 。

4) w/o SPE: 移除空间位置编码。

5) w/o TPE: 移除时间位置编码。

6) w/o STPE: 同时移除时间和空间位置编码。

除上述差异外,所有变体都与 TAformer 具有完全相同的设置。表 5 列出了这些变体在 METR-LA 数据集上的性能。在表 5 的 1—3 行,我们研究了空间注意力模块中拓扑信息、补充信息和先验拉普拉斯矩阵 \tilde{L} 的作用,三者对模型性能都起着促进作用。先验的道路拓图不能完整地反映空间依赖关系,通过全局建模捕获潜在的空间特征,在一定程度上补充了拓扑信息的不足。引入 \tilde{L} 则可以导入边的权重和方向信息,可略微增强模型的性能。

表 5 METR-LA 数据集上的消融实验结果

Table 5 Ablation experimental results on METR-LA dataset

Model	15 min			30 min			60 min		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
w/o Topology	2.57	5.30	6.7	2.95	6.25	8.1	3.38	7.20	9.7
w/o Supplement	2.54	5.21	6.5	2.95	6.22	8.0	3.40	7.24	9.7
w/o Laplace	2.53	5.18	6.5	2.91	6.17	7.9	3.36	7.14	9.6
w/o SPE	2.63	5.38	6.8	3.10	6.47	8.5	3.70	7.65	10.7
w/o TPE	2.81	5.93	7.7	2.89	6.11	8.0	3.56	7.54	10.0
w/o STPE	2.80	5.76	7.7	3.14	6.53	8.7	3.97	8.21	11.1
TAformer	2.52	5.17	6.5	2.90	6.14	7.9	3.35	7.13	9.5

从表 5 的第 4—6 行可以发现,移除任意一种位置编码后,模型的性能都会急剧下降。这是因为在移除时间和空间位置编码后,注意力机制在建模时将分别丢失时间序列的顺序特征和节点的空间位置信息。虽然在 15 min 的预测结果上看,时间位置编码的影响更大,但从整体看,空间位置编码更为重要,而同时移除两者后,模型效果最差。

结束语 本文提出了一种新的时间感知的 Transformer 模型(Time-Aware Transformer, TAformer),并将其应用于交通流预测任务。一方面,TAformer 采用时间感知注意力机制,可以有效提取交通流随时间动态变化的时空模式。另一方面,TAformer 通过非自回归生成目标序列,避免了误差累积,在长时预测上表现出了巨大的优势。在两个交通数据集上进行实验,结果表明,本文方法可以有效捕获交通流复杂的时空依赖,并具有十分优越的预测性能。

值得注意的是,非自回归的引入虽然提高了模型的预测速度和长时预测能力,但也牺牲了一定的短时预测性能。相反,自回归对短时预测更有利。为了弥补这一缺陷,我们下一步的工作将考虑寻找一种折中的方案,将这两种方式结合,同时兼顾长时和短时预测。在未来,我们还会尝试将 TAformer 用于其他具有时空图式数据的任务,比如空气质量监测和区域疫情预警,以验证 TAformer 在其他任务上的适应性和泛化能力。

参考文献

[1] LI M, ZHU Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press. 2021:

4189-4196.

- [2] 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 the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020: 914-921.
- [3] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]//Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 3837-3845.
- [4] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [5] GRAVES A. Long short-term memory[J]. Supervised Sequence Labelling with Recurrent Neural Networks, 2012, 385: 37-45.
- [6] CHO K, VAN MERRIËNBOER B, BAHDANAU D, et al. On the properties of neural machine translation: Encoder-decoder approaches[J]. arXiv:1409.1259, 2014.
- [7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach, CA, 2017: 5998-6008.
- [8] CAI L, JANOWICZ K, MAI G, et al. Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting[J]. Transactions in GIS, 2020, 24(3): 736-755.
- [9] 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. Ireland, 2020: 1215-1224.
- [10] CHEN C, LIU Y, CHEN L, et al. Bidirectional Spatial-Temporal

- Adaptive Transformer for Urban Traffic Flow Forecasting[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(10): 6913-6925.
- [11] YAN H, MA X, PU Z. Learning dynamic and hierarchical traffic spatiotemporal features with transformer[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 22386-22399.
- [12] WILLIAMS R J, ZIPSER D. A learning algorithm for continually running fully recurrent neural networks[J]. Neural Computation, 1989, 1(2): 270-280.
- [13] SCHMIDT F. Generalization in generation: A closer look at exposure bias[J]. arXiv:1910.00292, 2019.
- [14] BENGIO S, VINYALS O, JAITLEY N, et al. Scheduled sampling for sequence prediction with recurrent neural networks[C]// Advances in Neural Information Processing Systems. Montreal, Quebec, Canada, 2015: 1171-1179.
- [15] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C]// Advances in Neural Information Processing Systems. Long Beach, CA, 2017: 1024-1034.
- [16] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]// 6th International Conference on Learning Representations. Vancouver, 2018.
- [17] ARORA S. A survey on graph neural networks for knowledge graph completion[J]. arXiv:2007.12374, 2020.
- [18] WU S, SUN F, ZHANG W, et al. Graph neural networks in recommender systems: a survey[J]. ACM Computing Surveys (CSUR), 2022, 55(5): 1-37.
- [19] JIANG W, LUO J. Graph neural network for traffic forecasting: A survey[J]. Expert Systems with Applications, 2022, 207: 117921.
- [20] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[J]. arXiv:1709.04875, 2017.
- [21] ZHAO L, SONG Y, ZHANG C, et al. T-gen: A temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(9): 3848-3858.
- [22] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[J]. arXiv:1707.01926, 2017.
- [23] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling[J]. arXiv:1906.00121, 2019.
- [24] ZHANG Q, CHANG J, MENG G, et al. Spatio-temporal graph structure learning for traffic forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020: 1177-1185.
- [25] ZHANG J, SHI X, XIE J, et al. GAAN: Gated attention networks for learning on large and spatiotemporal graphs[J]. arXiv:1803.07294, 2018.
- [26] 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. Honolulu, Hawaii, USA: AAAI Press, 2019: 922-929.
- [27] 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. New York: AAAI Press, 2020: 1234-1241.
- [28] XU M, DAI W, LIU C, et al. Spatial-temporal transformer networks for traffic flow forecasting[J]. arXiv:2001.02908, 2020.
- [29] REN Y, ZHANG J. Fake news detection on news-oriented heterogeneous information networks through hierarchical graph attention[C]// 2021 International Joint Conference on Neural Networks (IJCNN). Shenzhen: IEEE, 2021: 1-8.
- [30] HU Z, DONG Y, WANG K, et al. Heterogeneous graph transformer[C]// Proceedings of The Web Conference 2020. Taipei: 2020: 2704-2710.
- [31] LIU Q, LONG C, ZHANG J, et al. Aspect-Aware Graph Attention Network for Heterogeneous Information Networks[J/OL]. IEEE Transactions on Neural Networks and Learning Systems, 2022. <https://ieeexplore.ieee.org/abstract/document/9930625>.
- [32] GROVER A, LESKOVEC J. node2vec: Scalable feature learning for networks[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016: 855-864.
- [33] ORHAN A E, PITKOW X. Skip connections eliminate singularities[J]. arXiv:1701.09175, 2017.
- [34] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society, 2016: 770-778.
- [35] LU B, GAN X, JIN H, et al. Spatiotemporal adaptive gated graph convolution network for urban traffic flow forecasting[C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Ireland: ACM, 2020: 1025-1034.
- [36] LIU J, GUAN W. A summary of traffic flow forecasting methods[J]. Journal of Highway and Transportation Research and Development, 2004, 21(3): 82-85.
- [37] XU D, WANG Y, JIA L, et al. Real-time road traffic state prediction based on ARIMA and Kalman filter[J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(2): 287-302.
- [38] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]// Advances in Neural Information Processing Systems. 2014: 3104-3112.
- [39] 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. New York: AAAI Press, 2020: 3529-3536.
- [40] MICIKEVICIUS P, NARANG S, ALBEN J, et al. Mixed precision training[J]. arXiv:1710.03740, 2017.
- [41] CHEN T, XU B, ZHANG C, et al. Training deep nets with sub-linear memory cost[J]. arXiv:1604.06174, 2016.
- [42] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980, 2014.



LIU Qidong, born in 1990, Ph.D, associate professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include swarm intelligence, motion planning and spatio-temporal data analysis.