计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(5):87−95 [doi: 10.15888/j.cnki.csa.009834] [CSTR: 32024.14.csa.009834]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

面向交通预测的窗口时空注意力建模方法①

王嘉俊[1], 李　勇[2], 欧丽云[2]

[1](福州大学 计算机与大数据学院, 福州 350108)
[2](福建警察学院 治安系, 福州 350007)
通信作者: 李　勇, E-mail: yl112@fjpsc.edu.cn

摘　要: 在目前交通预测中, 基于深度学习的时空分离建模方法难以有效地表达数据中的时空耦合相关性, 时空联合建模方法虽然可以一定程度上弥补时空分离建模的缺点, 但是其在构建时空超图时存在表达能力不足、计算复杂度高等问题, 为此提出一种改进的窗口时空注意力网络 (window spatial-temporal attention network, W-STANet) 的时空联合建模方法. W-STANet 主要由数据嵌入层、时空相关性建模层、预测头 3 部分组成. 时空相关性建模层通过堆叠多个时空注意力块实现对交通数据时空相关性特征的学习; 同时, 通过引入局部窗口计算方式以及数据移位和交换操作, 大幅度降低了建模过程中的计算复杂度, 并实现了在时空图中局部视角和全局视角的建模. 在 5 个真实交通公共数据集上的实验结果表明, 相比其他时空联合建模方法有着更好的预测性能. 与其他时空分离建模方法相比, 在大规模路网的数据集上的预测性能较优.

关键词: 时空相关性; 交通预测; 自注意力机制; 深度学习; 时空耦合相关性

引用格式: 王嘉俊,李勇,欧丽云.面向交通预测的窗口时空注意力建模方法.计算机系统应用,2025,34(5):87–95. http://www.c-s-a.org.cn/1003-3254/9834.html

Window Spatio-temporal Attention Modeling Approach for Traffic Prediction

WANG Jia-Jun[1], LI Yong[2], OU Li-Yun[2]

[1](College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China)
[2](Public Security Department, Fujian Police College, Fuzhou 350007, China)

Abstract: Currently, in traffic prediction, deep learning-based spatio temporal separation modeling methods have difficulty in expressing spatio-temporal coupling correlations in data effectively. Although spatio-temporal joint modeling methods can compensate for the shortcomings of spatio-temporal separation modeling to some extent, there are deficiencies such as insufficient express ability and high computational complexity in constructing spatio-temporal hypergraphs. To address these issues, this study proposes an improved spatio-temporal joint modeling method, window spatial-temporal attention network (W-STANet). W-STANet mainly comprises three parts: a data embedding layer, a spatio-temporal correlation modeling layer, and a prediction head. The spatio-temporal correlation modeling layer learns spatio-temporal correlation features of traffic data by stacking multiple spatio-temporal attention blocks. Meanwhile, by introducing the local window calculation method, data shifting and permutation operations, the computational complexity in the modeling process is greatly reduced, and the modeling from both local and global perspectives within the spatio temporal graph is achieved. Experimental results on five real traffic public datasets demonstrate superior prediction performance compared to other spatio-temporal joint modeling methods. Compared with spatio-temporal separation modeling methods, it has superior prediction performance on large-scale road network datasets.

① 基金项目: 国家自然科学基金 (62301159); 福建省自然科学基金面上项目 (2023J01229)
收稿时间: 2024-10-17; 修改时间: 2024-11-08; 采用时间: 2024-11-28; csa 在线出版时间: 2025-02-26
CNKI 网络首发时间: 2025-02-27

© 中国科学院软件研究所　http://www.c-s-a.org.cn

Key words: spatio-temporal correlation; traffic prediction; self-attention mechanism; deep learning; spatio-temporal coupling correlation

交通预测作为智能交通系统 (intelligent transportation system, ITS) 中的重要部分, 其主要目标在于通过对历史数据的学习, 预测未来交通参数的变化, 进而辅助交通管理及决策. 交通数据通常包含时间维度信息和空间维度信息两部分, 如何提高预测结果的可用性, 难点在于对数据中复杂的时空相关性的变化进行合理的建模.

随着深度学习技术的迅速发展, 相较于传统的时序预测方法如 ARIMA[1]等, 其优势在于能够考虑数据中时间和空间相关性, 因此在性能上具有绝对的优势. 基于深度学习的交通参数预测方法的核心是构建一个时空表征学习模块, 时空表征学习模块建模时空相关性的方式可以分为两种: 时空分离建模和时空联合建模. 时空分离建模是指在建模的过程中, 分别单独地对数据中的时间相关性和空间相关性进行建模. 具体做法是通过图神经网络 (graph neural network, GNN) 或注意力机制 (如 self-attention[2]) 等方法对数据中的空间相关性进行建模, 再结合时序建模方法如: 循环神经网络 (recurrent neural network, RNN) 及变体[3]、一维卷积神经网络 (1D-CNN) 等提取时间相关性, 最后将时间特征和空间进行融合[4,5]. 在现有的工作中, 大部分的工作集中于对交通数据的时空相关性采取分离建模. 时空联合建模是指同时对数据中的时空相关性进行建模, 这类方法通常需要将不同位置的序列信息展开成时空超图, 即将不同时刻的网络拓扑图之间的节点建立虚拟连接, 在同一维度上建立时间维度和空间维度之间信息的相互关系. 最后使用 GNN 等模型进行超图的时空表征学习, 如 STFGNN[6]、FourierGNN[7]等.

时空分离建模的缺点在于建模中将数据中时间相关性和空间相关性分离开来, 忽视了交通数据中时空耦合相关性, 时空耦合相关性是指交通流的时间相关性和空间相关是并不总是独立的[8]. 时空联合建模相较于时空分离建模的优势在于在建立数据中时空相关性的基础上, 并将时空耦合相关性考虑在内. 然而时空联合建模所面临的主要问题是对时空超图建模所带来的计算复杂度高以及如何构建合适的时空超图. 其中时空超图的构建尤为重要, 因为不同时刻的不同节点之间的相关性是复杂多样的, 时空超图构建的好坏直接影响了交通数据中时空相关性表示, 进而影响最终的预测结果. 但现有的时空联合建模方法为了计算效率所构建的时空超图通常难以合适的表达数据中原本的时空相关性.

综上, 为解决上述问题, 本文设计了一种基于注意力机制的联合时空建模方法, 通过构建全连接时空超图, 并借助注意力机制计算“语义相似度”的方式来表示时空超图中各个节点之间动态的时空相关性, 避免了前期时空超图构建表达能力不足的问题. 另外, 为降低建模过程中的计算复杂度, 采用窗口局部计算注意力的方式, 并通过添加额外的数据平移和交换操作, 使得模型能够在局部时空相关性的基础上拥有全局的时空感受视野. 实验的结果表明, 本文提出的模型能够有效地捕获数据中的动态时空相关性, 并且在计算效率上不落后于现有的模型.

1 相关工作

在相关工作的介绍中, 本文继续按照不同的时空建模方式进行阐述. 为了更直观地理解时空联合建模和时空分离建模, 图 1 简要展示了二者在时空建模中的主要区别.

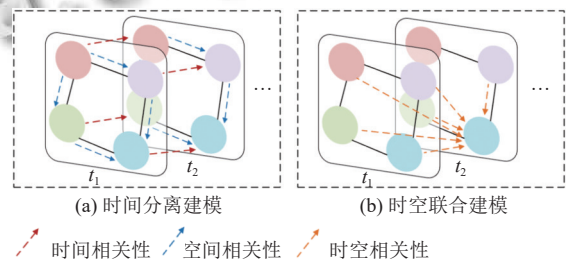


图 1 不同时空建模方式示意图

1.1 时空分离建模

时空分离建模方式的工作可以进一步的分为基于 GNN 和注意力机制两类, 其中 GNN 使用最多的变体是图卷积网络 (graph convolution network, GCN[9]). 如 STGCN[10]、Graph WaveNet[11]及 TGC-LSTM[12]中都使用 GCN 对空间相关性进行建模, 不同的是在时间相关性上, STGCN 和 Graph WaveNet 使用的是 1D-CNN,

© 中国科学院软件研究所 http://www.c-s-a.org.cn

而后者是 LSTM[13]. 基于注意力机制的方法如 GMAN[14]、ST-GRAT[15]等, 二者在空间相关性和时间相关性上的建模上都采用注意力机制进行建模. GCN 和注意力机制在空间相关性建模的区别在于, GCN 对路网中的节点的空间相关性进行建模的方式是通过预定义的邻接矩阵来实现, 体现的是“静态”的空间相关性变化. 而注意力机制是通过计算各个观测节点之间的“语义相关性”来衡量空间相关性, 体现的是“动态” 的空间相关性变化. 由于交通数据中的时空相关性通常是动态变化的[16], 因此基于注意力机制的方法通常要比基于 GCN 的方法有着更好的预测性能. 同样也有结合 GCN 和注意力机制两种方法的工作, 如 ASTGNN[17]、LSGCN[18]等, 这样可以同时考虑数据中空间相关性的“静态”和“动态”两个方面, 提高模型的建模能力.

1.2 时空联合建模

时空联合建模的核心在于如何定义时空超图中不同时刻空间拓扑之间的联系. 例如 FC-STGNN[19]中通过节点之间的相似度来确立不同节点在不同时间戳上联系, 同时基于时间距离引入 Decay Matrix 建立同一节点在不同时间戳上的联系, 从而构建最终的图结构数据, 并使用全连接的 GCN 实现建模. STSGCN[8]通过考虑相邻时间步之间空间拓扑图之间的关系, 实现的是局部时空拓扑图的构建, 并基于此同样使用 GCN 完成数据的建模. 在 STFGNN[6]中则是通过基于时间序列相似性以及卷积网络, 将部分空间拓扑融合为时空拓扑, 同时通过引入门控扩张卷积兼顾局部和全局相关. 与前面提到的工作不同的是, 在 FourierGNN[7]中, 作者假设时空超图中所有的节点都是全连接的, 避免了前期超图的构建, 但是这带来了高计算复杂度的问题, 以致于无法直接用 GCN 等模型进行建模. 因此作者设计一个新的傅里叶图算子, 一定程度上降低建模过程中的计算复杂度. 因此, 在时空联合建模方法中, 定义合适的时空超图可以一定程度上避免建模过程中的计算开销, 但也有可能忽略了数据中部分时空信息. 而通过全连接方式的构图, 虽然保留了数据原始时空信息, 但也会不可避免地在数据中引入一些噪声, 并且需要合理地设计建模方法来降低计算开销, 使得能够在实际中运用.

从上述相关工作中可以得出, 虽然时空联合建模方式在理论上对时空数据有着更好的表达能力, 但在实际中由于时空超图构建以及计算开销等问题, 使得在预测结果上并没有比时空分离建模更好, 相反还可能因为不恰当地定义时空超图, 适得其反. 因此, 目前大部分性能比较好的工作集中于时空分离建模的方式.

2 模型方法

模型 W-STANet 整体结构如图 2(a) 所示, 主要包含 Embedding 层, 时空相关性建模层以及预测头 3 部分: (1) Embedding 层, 对模型的输入 (例如: 交通流量、时间信息、空间位置信息) 进行初步的编码表示. (2) 时空相关性建模层, 通过叠加多个时空注意力块对数据的时空动态相关性进行联合建模, 学习在复杂道路网络下交通数据的特征表示. (3) 预测头, 主要包含 2 层全连接 (fully connection, *FC*) 网络, 激活函数为 GELU. 此外, 模型还通过引入 *ReVIN*[20]操作来缓解数据分布不一致的问题.

2.1 问题定义

定义 1. 交通路网表示. 对于交通路网可以表示为如下图结构数据$G=(V,E,A)$, 其中$V=(v_1,v_2,\cdots,v_N)$和$E=(e_1,e_2,\cdots,e_K)$分别表示图中的节点集合以及边集合. V满足($|V|=N$), 为交通路网中观测节点数量.

定义 2. 交通数据表示. 对于 t 时刻交通路网中 N 个观测节点的数据可以表示为如下$\boldsymbol{X}_t\in\mathbb{R}^{N\times C}$, C表示交通数据的维度. 进一步, 对于 T 个时间步交通数据可以表示为$\boldsymbol{M}=[\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_T]\in\mathbb{R}^{T\times N\times C}$.

交通参数预测任务可以描述为通过对数据的学习得到一个函数映射, 其将 T 个时间步长的历史数据和路网结构作为输出, 输出T'个时间步长的预测值. 具体如式 (1) 所示:

$$[\boldsymbol{X}_{t+T+1},\boldsymbol{X}_{t+T+2},\cdots,\boldsymbol{X}_{t+T+T'}]=F([\boldsymbol{X}_{t+1},\boldsymbol{X}_{t+2},\cdots,\boldsymbol{X}_{t+T}],G) \tag{1}$$

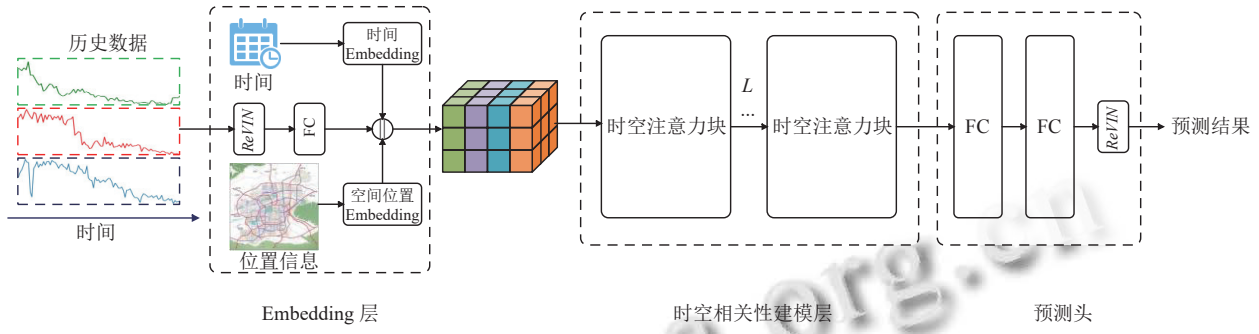
2.2 数据预处理及 Embedding

首先将不同位置的交通序列数据展开为时空超图, 如图 3 所示. 为提高模型对不同时刻、不同位置交通模式的感知能力, 如早高峰等时间信息. 将时间信息及观测点位置信息通过嵌入编码 (Embedding) 的方式引入模型建模过程中, 如图 2(a) 中 Embedding 层所示. 具体操作如下, 对于时间信息包含“分钟”和“天” 两种频率尺度, 分别表示为$\boldsymbol{E}_{\text{time}}\in\mathbb{R}^{T\times N\times d_1}$、$\boldsymbol{E}_{\text{day}}\in\mathbb{R}^{T\times N\times d_2}$. 同理, 对交通网络中的每个观测点位做相同的编码操作, 即$\boldsymbol{E}_S\in\mathbb{R}^{T\times N\times d_3}$. 对于交通数据, 先经过 *ReVIN* 操

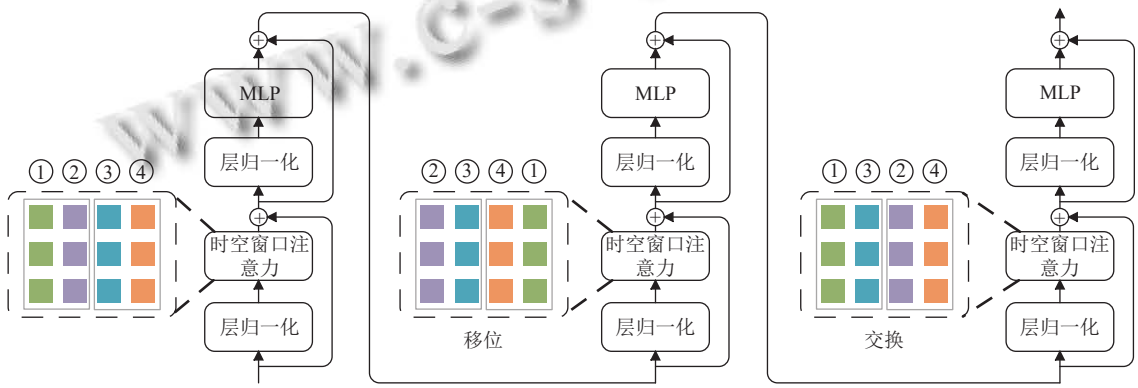
© 中国科学院软件研究所 http://www.c-s-a.org.cn

作, 再通过一层全连接网络, 得到高维向量特征表示 $\boldsymbol{E}_{\text{traffic}} \in \mathbb{R}^{T\times N\times d_4}$, 如式 (2). 最后, 将这些特征编码进行拼接操作得到模型的输入 $Input = (E_{\text{time}} \| E_{\text{day}} \| E_S \| E_{\text{traffic}}) \in \mathbb{R}^{T\times N\times D}$, 其中, $D = d_1 + d_2 + d_3 + d_4$ 为模型的隐藏层的维度.

$$\boldsymbol{E}_{\text{traffic}} = FC(ReVIN([\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_T])) \tag{2}$$



(a) W-STANet 模型整体框架



(b) 时空注意力块计算流程示意图

图 2 模型结构

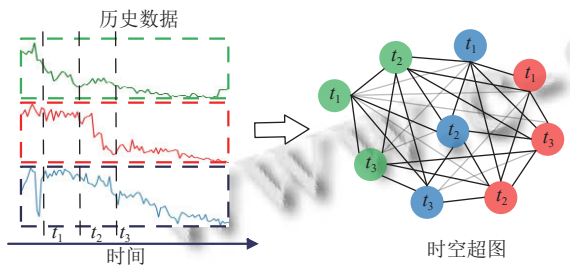


图 3 时空超图构建示意图

2.3 时空相关性建模

为了缓解时空超图建模中使用多头自注意力 (multi-head self-attention, MSA)[2]方式带来的较大计算复杂度问题, 本文通过对超图进行窗口划分, 分别对窗口内的子图使用 MSA 计算相关性, 并通过对数据块的移位和交换操作, 达到直接对整个时空超图使用 MSA 建模的效果. 整个时空相关性建模过程由多个时空注意力块组成, 如图 2 所示, 其中每个块的具体计算过程由第 2.3.1 和 2.3.2 节给出.

2.3.1 多头自注意力 (MSA)

MSA 将输入分别划分成 query、key 和 value, 并转化为 h 组不同的线性投影, 并通过注意力聚合操作并行得到 h 组输出, 随后将这 h 组输出拼接并通过另一个可学习的线性投影进行变换, 得到最终输出. 每一组的输出代表一个头. 具体运算过程如式 (3) 和式 (4) 所示.

$$\boldsymbol{h}_i = Softmax\left(\frac{\boldsymbol{H}\boldsymbol{W}_i^{(q)}\left(\boldsymbol{H}\boldsymbol{W}_i^{(k)}\right)^{\mathrm{T}}}{\sqrt{d_{\text{model}}}}\right)\left(\boldsymbol{H}\boldsymbol{W}_i^{(v)}\right) \tag{3}$$

$$\boldsymbol{O} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_i]\boldsymbol{W}_O \tag{4}$$

其中, $\boldsymbol{W}_i^{(v)}, \boldsymbol{W}_i^{(k)}, \boldsymbol{W}_i^{(q)} \in \mathbb{R}^{d\times d_{\text{model}}}$, $\boldsymbol{W}_O \in \mathbb{R}^{id_{\text{model}}\times d_o}$ 分别为可学习的参数矩阵, $\boldsymbol{H} \in \mathbb{R}^{p\times d}$ 为输入矩阵.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

2.3.2 窗口注意力计算

由于每个窗口的计算无法具有全局的感受野范围, 为了实现路网中所有观测节点之间的时空信息交互, 本文将每个时空注意力块的计算过程分成 3 部分.

如图 2(b) 所示, 本文先将输入按照窗口大小划分为非重叠的块, 图 2(b) 给出的是空间维度的划分示例, 即将路网中的观测节点进行划分. 具体来说, 对道路网络中的观测点采取对半分策略, 而在时间维度上, 同样采取划分窗口的方式, 时间维度窗口的大小在本文中设置为 4. 具体流程如图 2(b) 中所示, 在完成第 1 次时空注意力计算后, 依次对数据进行移位和交换操作再进行时空注意力计算. 其中移位操作的目的是实现不同窗口之间的信息交互, 其时空信息的传递具有间接性. 单通过移位操作无法满足路网中部分节点之间信息直接传递的需求, 为此需要额外增加交换操作, 其时空信息的传递是直接性. 至此, 通过数据的移位及交换操作, 实现了对数据中局部时空相关性和全局时空相关性的建模. 具体计算过程如式 (5) 和式 (6) 所示:

$$F_{\mathrm{STWA}}(\boldsymbol{X}):\begin{cases}\hat{\boldsymbol{X}}=STW\text{-}MSA(LN(\boldsymbol{X}))+\boldsymbol{X}\\ \boldsymbol{Y}=MLP(LN(\hat{\boldsymbol{X}}))+\hat{\boldsymbol{X}}\end{cases}\tag{5}$$

$$\begin{cases}\boldsymbol{Z}_1^l=F_{\mathrm{STWA}}(\boldsymbol{Z}^{l-1})\\ \boldsymbol{Z}_2^l=F_{\mathrm{STWA}}(Shift(\boldsymbol{Z}_1^l))\\ \boldsymbol{Z}^l=F_{\mathrm{STWA}}(Permute(\boldsymbol{Z}_2^l))\end{cases}\tag{6}$$

其中, *STW-MSA* 代表时空窗口多头注意力计算; *MLP* 为多层感知机; *LN* 为层归一化; $\boldsymbol{Z}^l$ 为第 l 个时空注意力块的输出; $Shift(\cdot)$ 和 $Permute(\cdot)$ 分别为数据移位和交换操作.

通过窗口注意力计算, 可以进一步提高计算效率, 具体分析如下. 首先, 通过式 (3)、式 (4) 可以得出注意力机制在时空超图上的一次计算复杂度为式 (7) 所示, 其中 N、T、d 分别表示观测点数量、时间步长度、特征维度. 同理, 对于一个窗口内的注意力计算复杂度可以表示为 $\mathrm{O}(4ptd^2+2(pt)^2d)$, 其中 p、t 表示一个窗口内观测点数量和时间步长度. 因此, 将原始的数据划分为 $(N/p)\times(T/t)$ 个窗口, 总的计算复杂度如式 (8) 所示. 根据本文中 $p=N/2$、$t=4$, 一个时空注意力块的计算复杂度为 $\mathrm{O}(12NTd^2+12N^2Td)$.

$$\mathrm{O}(MSA)=4NTd^2+2(NT)^2d\tag{7}$$

$$\mathrm{O}(STWA)=4NTd^2+2ptNTd\tag{8}$$

3 实验分析

为了充分验证本文所提出方法的有效性, 本文分别在交通流量预测和交通速度预测两种任务, 总共 5 个真实数据集上进行验证.

3.1 数据集介绍

本文中所使用的交通流量数据集包括 PEMS04、PEMS07 以及 PEMS08[8]. 交通速度数据集包括 PEMS-BAY、METR-LA. 上述数据集采样间隔均为 5 min, 并且在实验过程中, 所有数据集的交通网络均被视为无向图. 为了与先前的工作进行一致性比较, 本文的预测任务形式是用历史 1 h (12 步长×5 min) 数据预测未来 1 h 结果, 即预测出 12 个步长变化, 对此可以进一步根据实际需要得到 10–15 min 的预测结果. 需要注意的是在 PEMS04、PEMS07 以及 PEMS08 数据集中, 与各个基准模型之间的比较采用的是预测结果与实际结果之间 12 个步长的平均误差, 而在 PEMS-BAY 与 METR-LA 中, 比较方式具体详见第 3.5 节. 数据集具体参数如表 1 所示.

表 1 数据集详情

数据集	观测点数	时间步	时间间隔 (min)	时间范围
PEMS04	307	16 992	5	2018/01/01–2018/02/28
PEMS07	883	28 224	5	2017/05/01–2017/08/31
PEMS08	170	17 856	5	2016/07/01–2016/08/31
PEMS-BAY	325	52 116	5	2017/01/01–2017/05/31
METR-LA	207	34 272	5	2012/03/01–2012/06/30

3.2 基准模型设置

为验证本文所提出模型的有效性, 本文选择的基准模型可以分成两大类: 基于图神经网络模型和基于自注意力模型. 基于图神经网络的模型: DCRNN[21]、STGCN[10]、MTGNN[22]、STSGCN[8]、STFGNN[6]、STGODE[23]、STG-NCDE[24]. 基于自注意力的模型: GMAN[14]、ASTGNN[17]、PDFormer[25]. 基准模型的超参数均按照原始论文中所提供的超参数设置.

3.3 实验设置及评价指标

为与第 3.2 节中所提的基准模型保存一致, 本文将 PEMS04、PEMS07 及 PEMS08 数据集按照 6:2:2 比例划分为训练集、验证集和测试集. 将 PEMS-BAY 和 METR-LA 按照 7:1:2 比例划分.

在评价指标方面, 本文使用 mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), 以及 root mean squared error (*RMSE*) 这 3 个评价指标, 分别如式 (9)–式 (11) 所示. 最后, 在 PEMS04、PEMS07 以

© 中国科学院软件研究所 http://www.c-s-a.org.cn

及 PEMS08 数据集上使用 *HuberLoss* 作为损失函数, 而在 PEMS-BAY 和 METR-LA 数据集上使用 *MSE* Loss 作为损失函数, 具体形式如式 (12)、式 (13) 所示:

$$RMSE(y,\hat{y}) = \frac{1}{N}\sum_{i\in N}(y_i-\hat{y}_i)^2 \tag{9}$$

$$MAE(y,\hat{y}) = \frac{1}{N}\sum_{i\in N}|y_i-\hat{y}_i| \tag{10}$$

$$MAPE(y,\hat{y}) = \frac{1}{N}\sum_{i\in N}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{11}$$

$$\begin{cases} HuberLoss = \dfrac{1}{m}\sum\limits_{i=1}^{m} l_i \\ l_i = \begin{cases} 0.5(y_i-f(x_i))^2, & |y_i-f(x_i)| < delta \\ delta\times(|y_i-f(x_i)|-0.5\times delta), & |y_i-f(x_i)| \geqslant delta \end{cases} \\ \text{其中}, delta = 1.0 \end{cases} \tag{12}$$

$$MSE = \frac{1}{m}\sum_{i}^{m}(y_i-f(x_i))^2 \tag{13}$$

3.4 模型超参数设置

对于不同数据集的超参数设置如表 2 所示. 其中隐藏层维度可以是 32、64、128, 多头数量可以是 1、2、4、8, 学习率设置范围为[0.0001, 0.001]. 模型训练过程中使用 Adam 优化器进行梯度更新, 同时使用 ExponentialLR 算法来动态调整学习率, gamma 为 0.98. 此外, 模型训练轮数 epoch 均为 300.

表 2 不同数据集的模型超参数设置

数据集	Batch size	Hidden dimension	Weight decay	学习率
PEMS04	64	128	0.001	0.0008
PEMS07	16	128	0.003	0.001
PEMS08	64	128	0.003	0.0015
PEMS-BAY	64	128	0.0001	0.001
METR-LA	128	128	0.001	0.0015

3.5 实验结果及分析

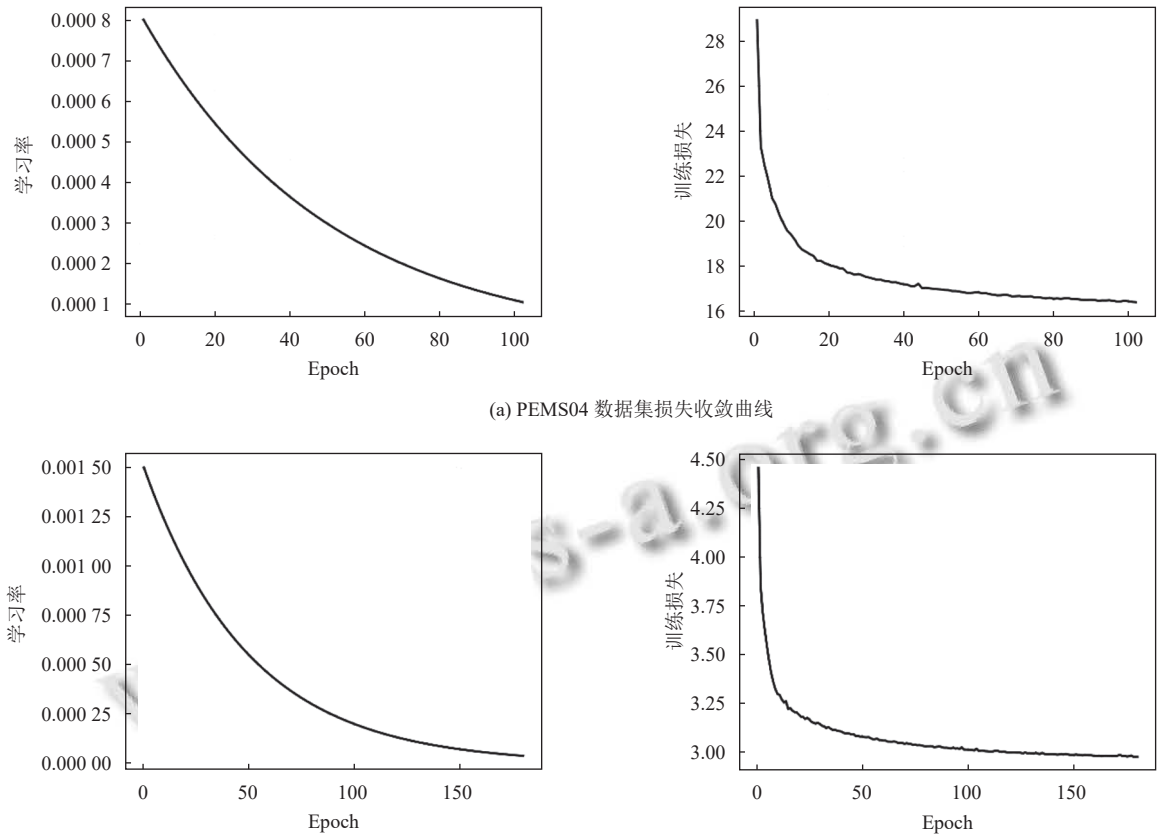
本文所提出的模型在交通速度预测任务的实验结果如表 3 所示, 图 4 给出了 PEMS04 和 METR-LA 两个数据集训练过程的损失收敛曲线. 与现有的预测模型相比, 本文的模型在 PEMS-BAY 数据集上有着明显的性能提升. 而在 METR-LA 数据集上, 由于是高速道路数据并且该路网结构相对单一, 大部分传感器处于同一条直线道路上, 因此各个方法之间并没有较大的性能差距.

表 3 模型在 PEMS-BAY 和 METR-LA 数据集上的实验结果 (车辆平均速度预测)

数据集	Horizon (min)	Metric	DCRNN	STGCN	GWNet	MTGNN	GMAN	PDFormer	W-STANet
PEMS-BAY	15	*MAE*	1.31	1.36	1.30	1.33	1.35	1.32	**1.13**
		MAPE (%)	2.73	2.86	2.71	2.81	2.87	2.78	**2.32**
		RMSE	2.76	2.88	2.73	2.80	2.90	2.83	**2.27**
	30	*MAE*	1.63	1.70	1.63	1.66	1.65	1.64	**1.36**
		MAPE (%)	3.73	3.79	3.73	3.75	3.74	3.71	**2.93**
		RMSE	3.73	3.80	3.73	3.77	3.82	3.79	**2.90**
	60	*MAE*	1.99	2.02	1.99	1.95	1.92	1.91	**1.60**
		MAPE (%)	4.71	4.72	4.71	4.62	4.52	4.51	**3.60**
		RMSE	4.60	4.63	4.60	4.50	4.49	4.43	**3.67**
MTRE-LA	15	*MAE*	**2.67**	2.75	2.69	2.69	2.80	2.83	2.73
		MAPE (%)	**6.85**	7.11	6.99	6.89	7.41	7.77	6.89
		RMSE	5.16	5.29	**5.15**	5.16	5.55	5.45	5.61
	30	*MAE*	3.12	3.15	3.08	3.07	3.12	3.21	**3.05**
		MAPE (%)	8.42	8.62	8.46	8.16	8.73	9.18	**7.93**
		RMSE	6.27	6.35	6.21	**6.13**	6.49	6.46	6.60
	60	*MAE*	3.54	3.60	3.51	3.47	**3.44**	3.62	3.51
		MAPE (%)	10.32	10.35	9.96	9.70	10.05	10.92	**9.31**
		RMSE	7.47	7.43	7.28	**7.21**	7.35	7.47	8.07

具体从表 3 的结果可知, 在数据集 PEMS-BAY 上本文的模型的评价指标均高于其他所有模型, 证明本文模型在交通速度预测任务上的有效性. W-STANet 相比于基准模型 PDFormer 在 15 min 的预测中, *MAE* 误差降低了 14%, *RMSE* 降低了 19%. 在 60 min 的预测中, *MAE* 降低了 15%, *RMSE* 降低了 17%. 从整体结果来看, 本文模型的联合时空建模方式有效提高了复杂道路的时空动态相关性建模能力, 主要是因为基于注意力机制的时空建模方式相较于基于 GNN 方法, 能更好地捕获复杂交通网络中信息传递的动态特性.

© 中国科学院软件研究所 http://www.c-s-a.org.cn



(a) PEMS04 数据集损失收敛曲线

(b) METR-LA 数据集损失收敛曲线

图 4　部分数据集训练损失收敛曲线

表 4 呈现的是在交通流量预测任务上的实验结果. 总体来看, 本文所提出的模型相较于基准模型都有着较好的性能优势. 从表 4 中可以看到, 本文的模型相较于 ASTGNN 在 PEMS04 数据集上的 *MAE*、*RMSE* 指标分别降低了 0.8% 和 3%, 在 PEMS08 数据集上分别降低 9.5% 和 6.8%. 相较于 PDFormer 而言在性能上的提升并没有那么显著, 但从 PEMS07 数据集的结果可以得出, 与 PDFormer 和 ASGTNN 这类时空分离建模方法相比, 在面对较大范围的交通网络场景中, 本文的模型仍然表现出更优秀的性能, 从侧面反映了本文的联合时空建模在大范围交通道路网络下能更好地对复杂时空相关性变化进行建模. 而从 STSGCN 的结果中可以看到, 虽然也采用的是时空联合的建模方式, 但是 STSGCN 在建立时空超图的过程中只考虑了邻近时间步之间的时空相关性, 难以较好地表示交通数据中复杂的时空相关性, 因此在性能上还不如其他时空分离建模的方法, 进一步表明了时空联合建模中时空超图构建的重要性.

另外, 为了比较本文所提出的时空联合建模方法与其他时空分离建模方法的计算效率, 在同一数据集 PEMS04 上并固定 batch size 大小为 16 进行测试, 计算每个 epoch 在训练过程以及推理过程所需要的平均时间, 单位为 epoch/s, 如表 5 所示, 其中 STANet 表示本文提出的模型不使用窗口方式计算. 从表 5 结果可以看到, 本文模型与性能比较相近的模型 ASTGNN 和 PDFormer 相比, 在训练过程中分别快了 2.32 倍和 1.13 倍. 同时使用窗口方式的计算效率比不用窗口方式快了 72%.

此外, 本文进一步探究了 Shift 和 Permute 操作对性能的影响, 结果如图 5 所示. 其中, w/o Shift 表示不使用移位操作; w/o Permute 表示不使用交换、w/o Both 表示同时不使用移位和交换操作; Origin 表示模型原始的结构. 从图 5 中的结果可以看到, 在不使用移位或者交换操作的情况下, 模型性能均有不同程度的下降, 其中 w/o shift 的性能结果下降较为明显. 这是因为模型建模过程中只使用数据的交换操作会给模型带来一定程度的干扰, 从而导致了性能下降. 从 w/o Both 的结果中可以看到, 只采取局部窗口计算相关性的建模方

© 中国科学院软件研究所 http://www.c-s-a.org.cn

式就能得到不错的性能, 主要原因是在本文的短时交通预测任务中, 交通数据的局部时空相关性特征为主要部分, 这是由交通数据本身的特性所决定的. 而交通数据除了局部时空相关性外, 还具有全局时空相关性, 从侧面反映了本文所设计的带有移位和交换的局部窗口时空相关性建模方式能够较好地考虑到数据中的全局时空相关性, 进一步提升模型对交通数据表征的学习能力.

表 4 模型在 PEMS04、PEMS07 和 PEMS08 数据集上实验结果 (交通流量预测)

模型	PEMS04			PEMS07			PEMS08		
	MAE	*MAPE* (%)	*RMSE*	*MAE*	*MAPE* (%)	*RMSE*	*MAE*	*MAPE* (%)	*RMSE*
DCRNN[21]	21.18	14.17	33.92	22.39	10.65	35.33	16.70	10.72	26.17
STGCN[10]	20.66	13.66	33.07	22.32	10.61	35.36	16.96	10.91	26.26
Graph WaveNet[11]	18.94	13.09	30.82	20.84	8.84	33.79	14.73	9.36	24.12
MTGNN[22]	19.12	13.17	31.63	20.86	9.02	34.07	15.29	10.19	24.59
STSGCN[8]	20.52	14.46	34.14	24.26	10.20	39.03	17.88	11.71	27.36
GMAN[14]	19.14	13.19	31.60	20.97	9.05	34.09	15.31	10.13	24.92
STFGNN[6]	19.82	13.15	31.74	22.07	9.22	35.81	16.64	10.55	26.21
STGODE[23]	20.84	13.77	32.82	22.99	10.14	37.54	16.81	10.62	25.97
STG-NCDE[24]	19.21	12.76	31.09	20.53	8.80	33.84	15.45	9.92	24.81
ASTGNN[17]	18.44	12.37	31.02	20.62	8.86	34.02	14.97	9.49	24.71
PDFormer[25]	18.32	12.10	**29.97**	19.83	8.53	32.87	13.58	9.04	23.50
W-STANet	**18.29**	**11.95**	30.06	**19.66**	**8.24**	**32.67**	**13.54**	**9.02**	**23.03**

表 5 模型在 PEMS04 数据集上的计算效率比较

模型	Training	Inference
ASTGNN[17]	208.69	52.11
PDFormer[25]	133.87	7.93
STTN[26]	100.21	11.67
STSGCN[8]	89.58	2.84
W-STANet	62.70	6.04
STANet	108.12	10.77

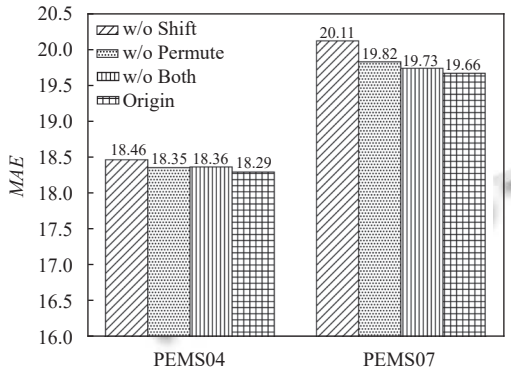


图 5 移位和交换操作对性能的影响

4 结论与展望

本文针对目前已有的时空联合建模过程在计算效率上的不足以及超图构建带来的影响, 提出一种在全连接时空超图上建模的时空联合建模模型 W-STANet. W-STANet 模型通过使用窗口方式分别先对超图中的子图进行建模, 再通过引入移位和交换操作, 将模型的感受视野从局部延伸到全局, 有效提升了模型在交通参数预测任务中的准确性, 同时在计算效率上有着较好的提升. 在 5 个公开数据集上的实验结果表明, 本文提出的模型能够有效建模交通参数数据中复杂的时空相关性, 在较大路网的情况下的预测性能要优于基准模型. 后续将继续探索时空联合建模的时空表达能力, 进一步探究性能和效率平衡的建模方法; 同时, 进一步研究全连接时空超图所带来的额外数据噪声问题, 降低对短期交通参数预测的影响.

参考文献

1 崔建勋, 要甲, 赵泊媛. 基于深度学习的短期交通流预测方法综述. 交通运输工程学报, 2024, 24(2): 50–64. [doi: 10.19818/j.cnki.1671-1637.2024.02.003]

2 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010. [doi: 10.5555/3295222.329534]

3 包广斌, 刘晨, 张波, 等. 基于 CEEMDAN-SBiGRU-OMHA 的短期电力负荷预测. 计算机系统应用, 2024, 33(10): 124–132. [doi: 10.15888/j.cnki.csa.009661]

4 张志远. 基于时空图卷积神经网络的管网产量模拟计算. 计算机系统应用, 2024, 33(6): 169–176. [doi: 10.15888/j.cnki.csa.009524]

5 夏英, 石栀琦. 面向交通流量预测的多头注意力时空卷积图网络模型. 计算机应用研究, 2023, 40(3): 766–770. [doi: 10.19734/j.issn.1001-3695.2022.07.0362]

© 中国科学院软件研究所 http://www.c-s-a.org.cn

6 Li MZ, Zhu ZX. Spatial-temporal fusion graph neural networks for traffic flow forecasting. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021. 4189–4196. [doi: 10.1609/aaai.v35i5.16542]

7 Yi K, Zhang Q, Fan W, *et al*. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. Proceedings of the 37th Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2023. 69638–69660. [doi: 10.48550/arXiv.2311.06190]

8 Song C, Lin YF, Guo SN, *et al*. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 914–921. [doi: 10.1609/aaai.v34i01.5438]

9 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017.

10 Yu B, Yin HT, Zhu ZX. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: AAAI, 2018. 3634–3640. [doi: 10.24963/ijcai.2018/505]

11 Wu ZH, Pan SR, Long GD, *et al*. Graph WaveNet for deep spatial-temporal graph modeling. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: ijcai.org, 2019. 1907–1913. [doi: 10.24963/ijcai.2019/264]

12 Cui ZY, Henrickson K, Ke RM, *et al*. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(11): 4883–4894. [doi: 10.1109/TITS.2019.2950416]

13 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]

14 Zheng CP, Fan XL, Wang C, *et al*. GMAN: A graph multi-attention network for traffic prediction. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 1234–1241. [doi: 10.1609/aaai.v34i01.5477]

15 Park C, Lee C, Bahng H, *et al*. ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed. Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020. 1215–1224. [doi: 10.1145/3340531.3411940]

16 Yao HX, Tang XF, Wei H, *et al*. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu: AAAI, 2019. 5668–5675. [doi: 10.1609/aaai.v33i01.33015668]

17 Guo SN, Lin YF, Wan HY, *et al*. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(11): 5415–5428. [doi: 10.1109/TKDE.2021.3056502]

18 Huang RZ, Huang CY, Liu YB, *et al*. LSGCN: Long short-term traffic prediction with graph convolutional networks. Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence. Huawei: ijcia.org, 2021. 326. [doi: 10.24963/ijcai.2020/322]

19 Wang YC, Xu YC, Yang JF, *et al*. Fully-connected spatial-temporal graph for multivariate time-series data. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2024. 15715–15724. [doi: 10.1609/aaai.v38i14.29500]

20 Kim T, Kim J, Tae Y, *et al*. Reversible instance normalization for accurate time-series forecasting against distribution shift. Proceedings of the 10th International Conference on Learning Representations. OpenReview.net, 2021.

21 Li YG, Yu R, Shahabi C, *et al*. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018.

22 Wu ZH, Pan SR, Long GD, *et al*. Connecting the Dots: Multivariate time series forecasting with graph neural networks. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020. 753–763. [doi: 10.1145/3394486.3403118]

23 Fang Z, Long QQ, Song GJ, *et al*. Spatial-temporal graph ode networks for traffic flow forecasting. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. ACM, 2023. 364–373. [doi: 10.1145/3447548.3467430]

24 Choi J, Choi H, Hwang J, *et al*. Graph neural controlled differential equations for traffic forecasting. Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI, 2022. 6367–6374. [doi: 10.1609/aaai.v36i6.20587]

25 Jiang JW, Han CK, Zhao WX, *et al*. PDFormer: Propagation delay-aware dynamic long-range Transformer for traffic flow prediction. Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington: AAAI, 2023. 4365–4373. [doi: 10.1609/aaai.v37i4.25556]

26 Xu MX, Dai WR, Liu CM, *et al*. Spatial-temporal Transformer networks for traffic flow forecasting. arXiv:2001.02908, 2021.

(校对责编: 张重毅)

© 中国科学院软件研究所 http://www.c-s-a.org.cn