

# 基于空间信息延迟感知的时空Transformer交通流预测

孔文翔

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2023年3月8日; 录用日期: 2023年4月12日; 发布日期: 2023年4月23日

## 摘要

准确可靠的交通流预测对城市规划、居民出行、公共安全等具有重要现实意义, 目前对于交通流预测的方法主要集中在RNN等深度学习模型, 仍有以下问题: 一是没有考虑长、短距离动态空间依赖, 二是忽视了空间信息传播情况。基于以上情况本文提出一种基于空间信息延迟感知的时空Transformer模型。首先, 该模型采用多头自注意力机制捕捉城市交通流的时空特性; 其次, 在编码层使用空间与时间自注意力机制关注长、短距离动态空间影响与时间动态特性, 同时也模拟了空间信息传播延迟情况; 最后, 时空编码层通过跳跃连接处理, 从而进一步提高模型的性能。实验在两个真实的交通流数据集上进行, 对比主流的交通预测算法, 结果表明所提出的DAST-Transformer模型具有更好的性能和预测精度。

## 关键词

交通流预测, 时空特性, Transformer, 自注意力机制, 延迟感知

# Spatial-Temporal Transformer Traffic Flow Prediction Based on Spatial Information Delay Awareness

Wenxiang Kong

School of Optical-Electrical Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 8<sup>th</sup>, 2023; accepted: Apr. 12<sup>th</sup>, 2023; published: Apr. 23<sup>rd</sup>, 2023

## Abstract

Accurate and reliable traffic flow prediction has important practical significance for urban planning,

residents' travel, public safety, etc. At present, the methods for traffic flow prediction are mainly integrated in the RNN and other in DL models, but there are still the following problems: first, the dynamic spatial dependence of long and short distance is not considered, and second, the spatial information dissemination is ignored. Based on the above situation, this paper proposes a spatial-temporal Transformer model based on spatial information delay awareness. First of all, the model uses the multi-attention mechanism to capture the space-time characteristics of urban traffic flow; Secondly, in the encoder layer, the spatial-temporal self-attention mechanism is used to focus on the long and short distance dynamic spatial impact and temporal dynamic characteristics, and the spatial information transmission delay is also simulated; Finally, the spatial-temporal encoder layer can further improve the performance of the model through jump connection processing. The experiment is carried out on two real traffic flow data sets. Compared with the mainstream traffic flow prediction algorithms, the results show that the proposed DAST-Transformer model has better performance and prediction accuracy.

## Keywords

Traffic Flow Prediction, Spatial-Temporal Characteristics, Transformer, Self-Attention Mechanism, Delay Awareness

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近些年来, 飞速增长的机动车数量给城市发展和居民出行带来巨大的影响, 不仅对城市的道路交通规划带来巨大的挑战, 同时给城市管理添加了压力。为缓解道路交通对城市发展的影响, 智慧交通系统 [1] 已经被用于诊断、改善城市道路交通状况中来, 作为该系统的核心部分, 交通流预测被广泛研究与应用, 其精度影响到智慧交通系统的稳定运行。交通流预测基于历史交通流数据来研究未来一定时间段内的交通流变化情况, 准确的交通流预测能够有效地缓解道路拥堵状况以及应用于路线规划等方面。

交通流预测早期的研究方法基于统计学习方法, 例如历史平均模型(Historical Average, HA) [2]、自回归移动平均模型(Auto Regressive Integrated Moving Average, ARIMA) [3]、向量自回归模型(Vector Autoregression, VAR)等方法, 这些方法通常需要良好的模型结构以及大量的数据集作为支撑, 一些交通流数据中关键信息并没有得到深度挖掘, 同时非线性特征表征能力不足。随着后来机器学习的发展, 支持向量机回归(Support Vector Regression, SVR) [4]、K 邻近(K-Nearest Neighbor, KNN) [5]、贝叶斯网络(Bayesian Networks) [6]等机器学习方法能够处理交通流数据中非线性问题, 较统计学习方法具有良好的性能, 由于一些方法依靠其中的核函数影响到模型的性能, 并不能够构建合理特征去提升预测精度。深度学习作为非参数的学习网络, 其在信息处理上和结构上更具有优势, 能够更好的处理时间序列数据, 一般处理时间序列的神经网络包括循环神经网络(Recurrent Neural Networks, RNN) [7]以及其变种长短期记忆网络(Long Short-Term Memory, LSTM)、门控循环单元结构(Gated Recurrent Unit, GRU), Smith 等人使用 BP (Backpropagation) [8]神经网络处理动态信息, 相比较于传统统计学习模型以及非参数回归模型, 其在短时预测精度上有很大提高。

对于交通流预测而言最主要是能够提取数据中复杂的动态空间与时间相关性, 卷积神经网络(Convolutional Neural Networks, CNN) [9]被应用于捕获交通流数据的空间依赖性, RNN 被应用于捕获交

通流数据的动态时间相关性。随着图神经网络(Graph Neural Networks, GNN)在图像领域的广泛应用,基于 GNN 的方法同样应用在交通流预测中。Yu 等人[10]运用 GCN 捕捉高阶的时空信息,但局限在交通流的走向仅为单向。Zhao 等人[11]考虑到交通网络的复杂拓扑结构和影响交通流因素的多样性,交通模型通常是复杂多变的,为捕捉空间和时间相关性,提出时间图卷积网络 T-GCN,该模型与图卷积网络和门控递归单元相结合。Hu 等人[12]提出了一种动态图生成模块,该模块预先收集关于节点之间的地理邻近性和空间异质性信息。

基于注意力机制的预测模型 Sequence2Sequence [13]被应用于交通流预测中,在长时间预测下具有良好的效果,2017 年,由完全注意力机制所构成的 Transformer 模型[14]被提出, Li 等人[15]将 Transformer 模型应用于交通流预测中,能够预测任意时段的交通量情况,但未能适大规模、多时段交通预测问题。

为解决上述的一些问题,设计中提出基于空间信息延迟感知的时空 Transformer 模型(DAST-Transformer)用于预测交通流,采用 Transformer 的基本结构,本文主要的贡献如下:

- 1) 提出将周期数据、位置信息等融合至嵌入层,更好捕获时空动态相关性;
- 2) 设计时空编码层,考虑到时空相关性的建立,针对时空动态关系,建立时空自注意力机制,利用 mask 矩阵将位置空间与语义空间邻域信息结合到自注意力机制中去,提升对当前节点影响较大区域的关注,兼顾邻接节点与较远节点的影响因素;
- 3) 建立延迟感知模块模拟空间信息传播情况,将得到的交通流序列融合到空间信息自注意力机制中。

## 2. 整体设计

### 2.1. 问题定义

**定义 1.**交通流预测中通常将每个传感器定义为一个单独节点,传感器之间的距离以及关系被视作节点之间的边,交通网络可以定位为无向图  $G = (v, \varepsilon, A)$ , 其中  $v = \{v_1, v_2, \dots, v_N\}$  作为节点的集合,  $N$  为节点的数量,  $\varepsilon$  为节点边的集合,  $A \in \mathbb{R}^{N \times N}$  为无向图  $G$  的邻接矩阵,反映节点之间的连通关系和强度。

**定义 2.**交通流张量定义,定义  $X_t \in \mathbb{R}^{N \times C}$  表示在  $t$  时刻道路交通网路中  $N$  个节点的交通流数据,其中  $C$  表示交通流的特征维度,那么在过去时间长度为  $T$  的交通流可以表示为

$$X = (X_{(t-T+1)}, X_{(t-T+2)}, X_{(t-T+3)}, \dots, X_{(t)}) \in \mathbb{R}^{T \times N \times C}.$$

对于交通流量的预测,其主要目的是能够依靠历史时间交通流数据  $X$  预测未来  $T'$  个时间步长的交通流数据,预测值组成  $Y = (X_{(t+1)}, X_{(t+2)}, X_{(t+3)}, \dots, X_{(t+T')}) \in \mathbb{R}^{T' \times N \times C}$ , 其中存在函数映射关系  $f(\cdot)$  如式(1)。

$$\{X; G\} \xrightarrow{f(\cdot)} Y \quad (1)$$

### 2.2. 模型建立

预测模型主要是由数据嵌入层、堆叠  $L$  层的时空编码器以及输出层组成,如下图 1 所示,接下来详细介绍其中的结构和各个部分的组成。模型基本采用 Transformer 模型的结构与形式,多头注意力机制捕捉交通流数据时空相关性,对原始数据与路网信息处理后分别建立空间自注意力机制和时间自注意力机制,其中空间注意力机制为了能够考虑到动态关系与较远节点的影响问题,故空间注意力机制分为两部分分别捕捉其中的关系,包括语义空间自注意力机制与位置空间自注意力机制,另外位置空间注意力机制为模拟空间信息传播的时间延迟情况,设计延迟感知模块更新其中的参数信息,由此组合成基本的时空编码层。经过跳跃连接处理后的交通流序列沿用 transformer 模型中的位置全连接前馈网络的处理方式,得到最终的预测结果。

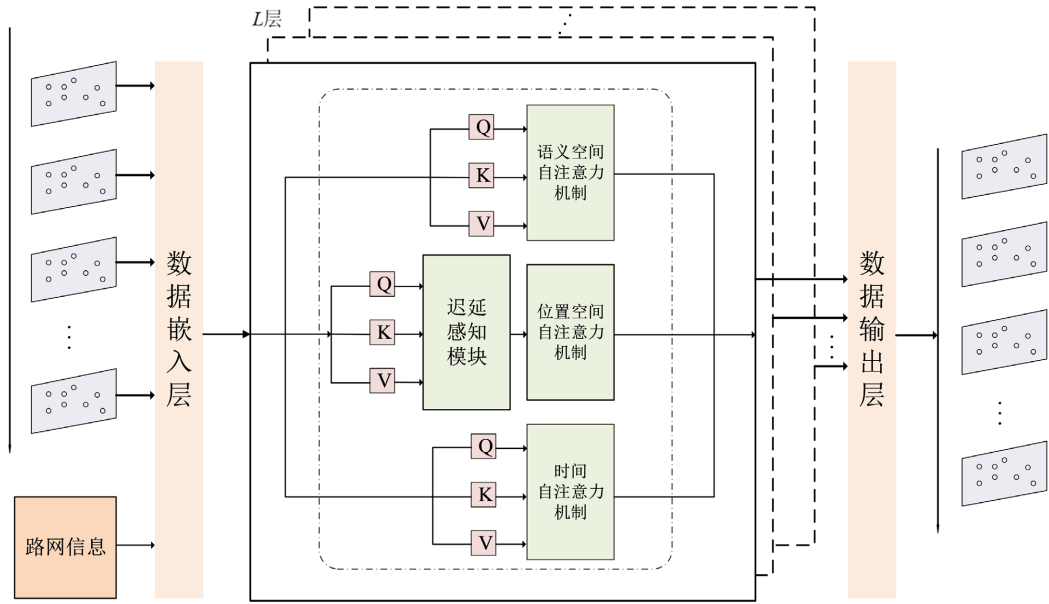


Figure 1. Schematic diagram of model structure  
图 1. 模型结构示意图

### 2.3. 数据嵌入层

将交通流的原始数据通过全连接层输入  $X \in \mathbb{R}^{T \times N \times C}$  转化为高维显示  $X_{CL} \in \mathbb{R}^{T \times N \times d}$ , 其中  $d$  表示数据嵌入需求的维度, 之后进一步将空间信息和时间信息嵌入其中。

首先对于空间特征使用图拉普拉斯变换, 能够尽可能保留原始的道路空间结构情况, 计算无向图  $G$  的归一化拉普拉斯矩阵  $\Delta$ , 如式(2)所示, 其中  $I$  为单位矩阵,  $D$  为邻接矩阵  $A$  的度矩阵, 通过特征分解得到特征向量矩阵和特征值矩阵, 如式(3)所示, 其中  $U$  为特征向量矩阵,  $\Lambda$  为特征值矩阵, 利用其中  $k$  个不为 0 的特征向量通过投影方式生成空间拉普拉斯嵌入矩阵  $X_{space} \in \mathbb{R}^{N \times d}$ 。

$$\Delta = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{2}$$

$$\Delta = U^T \Lambda U \tag{3}$$

其次对于时间特征而言, 考虑到城市道路交通周期性的出行规律和特性, 将周周期数据分量和日周期分量融和其中,  $w(t)$  和  $d(t)$  作为转换函数将原始的时刻  $t$  转换为周周期索引和分钟索引, 连接历史时间长度为  $T$  的数据片段构成周周期数据分量  $X_w \in \mathbb{R}^{T \times d}$  和日周期数据分量  $X_d \in \mathbb{R}^{T \times d}$ 。另外受到其他 Transformer 模型位置编码方法的启发, 在这里同样采用时间位置编码  $X_{td} \in \mathbb{R}^{T \times d}$  以导入数据序列的位置数据信息。

如下式(4)所示对上述多组数据融合相加得到  $X_{embed} \in \mathbb{R}^{T \times N \times d}$ , 其将得到的  $X_{embed}$  作为时空编码层的输入。

$$X_{embed} = X_{td} + X_w + X_d + X_{space} + X_{CL} \tag{4}$$

### 2.4. 时空编码层

综合考虑到交通流时空特性, 使用 transformer 结构设计基于自注意力机制的时空编码层, 主要是用于提取动态时空特征, 首先是时间自注意力机制, 用于捕捉历史时间动态特性和长期依赖特性, 其次是空间注意力机制, 包括语义空间与位置空间部分, 重点捕获数据中的长、短距离动态空间依赖, 其中迟

延感知模块,进一步巩固空间注意力机制,同时建立延迟模式模拟空间信息传播中的时间延迟情况。由于输入到时空编码层的数据三维数据,在各个模块中需要从不同的维度处理数据,故可以使用切片表示法,简化明确需要处理的数据段。

### 2.4.1. 空间自注意力机制

道路交通空间相关性不仅仅受到当前节点的邻接节点的影响,还可能受到周边节点的影响,对当前节点的影响也随时间动态变化,在这里空间自注意力机制用于捕获交通数据的空间动态相关性,在时间片  $t$  中可以得到自注意力机制中的  $Q$ 、 $K$ 、 $V$  矩阵,计算过程如下式(5)所示,其中  $W_I^{SSA}$ 、 $W_{II}^{SSA}$ 、 $W_{III}^{SSA}$  是可学习的参数,在空间维度应用自注意力机制,如式(6)所示通过计算得到注意力分数,即可得到在时间片  $t$  中节点之间的空间依赖性矩阵  $A_t^{SSA}$ 。

$$\begin{cases} Q_t^{SSA} = X_t W_I^{SSA} \\ K_t^{SSA} = X_t W_{II}^{SSA} \\ V_t^{SSA} = X_t W_{III}^{SSA} \end{cases} \quad (5)$$

$$A_t^{SSA} = \frac{(Q_t^{SSA})(K_t^{SSA})^T}{\sqrt{d'}} \quad (6)$$

上式中  $d'$  是  $Q_t^{SSA}$ 、 $K_t^{SSA}$ 、 $V_t^{SSA}$  矩阵的维度,通过计算得到  $A_t^{SSA}$  表征了时间片  $t$  中各个节点的空间影响程度,同时在不同的时间片中节点之间相互影响的程度也不同,意味着空间依赖性的动态变化性,因此上述空间自注意力机制能够捕捉到空间动态性,将得到的空间依赖性矩阵  $A_t^{SSA}$  与  $V_t^{SSA}$  矩阵相乘得到模块的输出,如式(7)所示:

$$Head^{SSA} = \text{softmax}(A_t^{SSA})V_t^{SSA} \quad (7)$$

不仅需要考虑到空间动态性,同时还要考虑到邻接节点以及更远节点对当前节点的影响,因此在这里引入两个 mask 矩阵  $M_I$ 、 $M_{II}$ ,进一步引入两个空间自注意力机制,其中矩阵  $M_I$  根据节点之间的距离进行定义,若是节点之间的距离小于阈值,则定义为 1 否则为 0。矩阵  $M_{II}$  使用 DTW 算法计算节点之间的交通流相似度,为每个节点选择具有最高相似度的  $k$  个节点作为语义相近节点,将当前的节点与语义相近节点之间的权重设置为 1 否则为 0,从而构造矩阵  $M_{II}$ 。将得到的矩阵  $M_I$  和  $M_{II}$  融入到模块的空间依赖性矩阵中,排除一些对当前节点影响较小的节点依赖关系,以捕获交通数据中邻接节点及更远节点的有效空间相关性,其中  $Head_t^{SSA-G}$ 、 $Head_t^{SSA-S}$  分别为位置空间和语义空间自注意力输出。

$$Head_t^{SSA-G} = \text{softmax}(A_t^{SSA} \odot M_I)V_t^{SSA} \quad (8)$$

$$Head_t^{SSA-S} = \text{softmax}(A_t^{SSA} \odot M_{II})V_t^{SSA} \quad (9)$$

### 2.4.2. 延迟感知模块

为了能够模拟当前历史时间段中各个时间片的空间信息传播延迟情况,需要整合给定时间段内各个时间片的历史交通流信息,将最终的结果合并到位置空间自注意力模块中,以明确空间信息传播时间延迟情况。

首先在历史交通流数据集中选取具有代表性的一组交通流数据,其代表了某种交通模式,通过大小为  $F$  的滑动窗口对历史交通流数据进行分割得到这组数据,然后对这组数据使用 k-Shape 聚类算法,其能够保留时间序列的形状,同时不受缩放等的影响,聚合的结果使用集合  $P = \{p_i | i \in [1, \dots, N_p]\}$  来表示,其中  $N_p$  代表聚类的总数,据此能够将  $P$  视为一组交通流数据。

相似的交通流数据意味着节点之间存在着类似的影响，故将数据集中的每个节点的历史交通流序列于提取到的集合  $P$  进行比较，最终将相似的数据信息融合到每个节点的交通流序列表示中。在给定节点中从  $(t-F+1)$  到  $t$  时刻，历史交通流序列  $X_{(t-F+1:t),n}$  首先将其通过嵌入矩阵中  $W^o$  得到高维形式  $o_{t,n} = X_{(t-F+1:t),n} W^o$ ，同时通过另一个嵌入矩阵  $W^m$  将集合  $P$  中的每个交通序列转换为记忆向量  $m_{i,n} = p_i W^m$ ，将高维形式的历史交通流序列与记忆向量比较得到相似性向量  $w_i = \text{softmax}(o_{t,n}^T m_{i,n})$ 。根据相似性向量  $w_i$  对集合  $P$  加权求和从而获得综合历史交通流序列  $e_{t,n} = \sum w_i (p_i W^c)$ ，其中  $W^c$  作为可学习的参数，经过完整的计算从而获得单个节点的综合历史交通流序列，通过累计得到  $N$  个节点的综合历史交通流序列  $E_t$ ，通过  $E_t$  的补充更新式(5)中  $K_t^{SSA}$  值，如式(10)所示。

$$\bar{K}_t^{SSA} = K_t^{SSA} + E_t \tag{10}$$

通过上述方法获得整合后的从  $(t-F+1)$  到  $t$  时刻的综合交通流序列，从而更新空间依赖性矩阵  $A_t^{SSA}$  以模拟空间信息传播的时间迟延情况，如下图 2 演示了单节点情况。

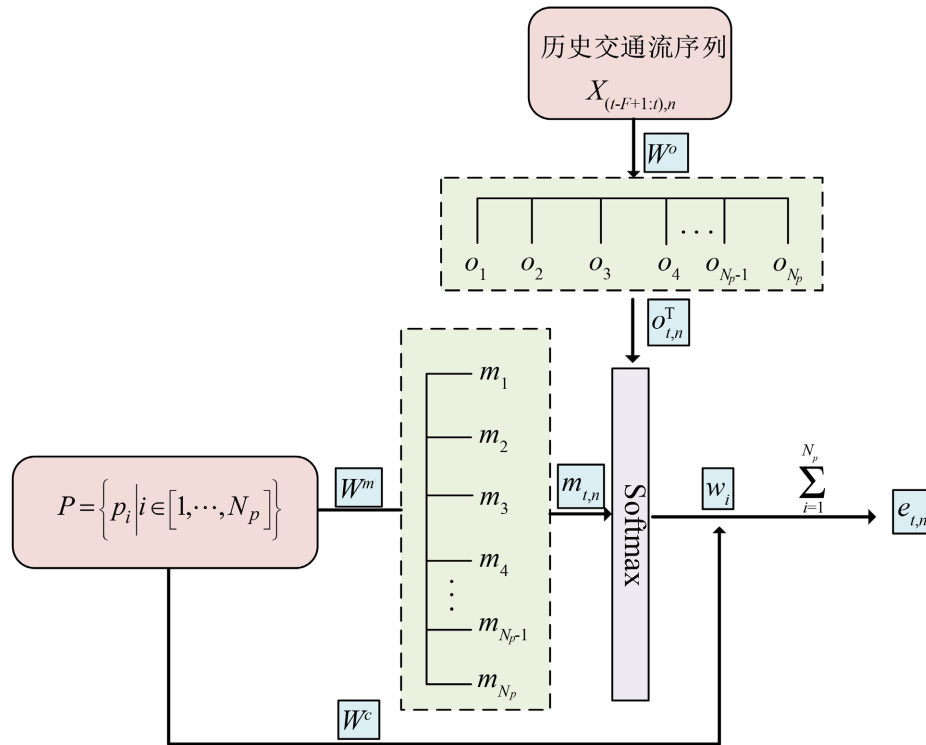


Figure 2. Delay awareness single node  
图 2. 单节点延迟感知

### 2.4.3. 时间自注意力机制

对于不同的时间片中的交通流数据存在时间动态相关性，不同节点在不同的情况下呈现出不同的形式，因此这里使用到时间自注意力机制用于挖掘时间动态相关性，与空间自注意力类似，通过式(11) (12) 计算当前的节点  $n$  计算时间依赖性矩阵  $A_n^T$ ，其中  $W_I^T$ 、 $W_{II}^T$ 、 $W_{III}^T$  为可学习的参数。

$$\begin{cases} Q_n^T = X_n W_I^T \\ K_n^T = X_n W_{II}^T \\ V_n^T = X_n W_{III}^T \end{cases} \tag{11}$$

$$A_n^T = \frac{(Q_n^T)(K_n^T)^T}{\sqrt{d'}} \quad (12)$$

从上式中可以看出时间自注意力机制能够发现不同节点中的不同动态时间相关性，另外时间自注意力机制具有全局性，将得到的时间依赖性矩阵  $A_n^T$  与  $V_n^T$  矩阵相乘得到模块的输出，如式(7)所示。

$$Head_n^T = \text{softmax}(A_n^T)V_n^T \quad (13)$$

## 2.5. 数据输出层

时空编码层最后主要将三种自注意力机制融合到一个多头自注意力模块，降低计算复杂性后进行预测，注意力头主要包括语义空间、位置空间以及时间自注意力头，将这些计算得到的结果连接输出得到整合时空信息后的输出，如下式(14)所示：

$$Head = \oplus \left( Head_{1-z_{ssa-g}}^{SSA-G}, Head_{1-z_{ssa-s}}^{SSA-S}, Head_{1-z_i}^T \right) W \quad (14)$$

式(14)中  $z_{ssa-g}$ 、 $z_{ssa-s}$ 、 $z_i$  分别表示各个注意力机制头数量， $W \in \mathbb{R}^{d \times d}$  作为可学习的参数矩阵，另外在后续处理中还用来 Transformer 模型中的位置全连接前馈网络的处理方式，获得输出  $X_l \in \mathbb{R}^{T \times N \times d}$ 。

经过多层连接，其中每个时空编码层通过  $1 * 1$  的卷积构成的跳跃连接与输出  $X_l$  相加得到隐藏状态矩阵  $X_{hid} \in \mathbb{R}^{T \times N \times d_{skip}}$ ，其中跳跃连接矩阵  $X_{hid} \in \mathbb{R}^{T \times N \times d_{skip}}$ ， $d_{skip}$  为跳跃维度。为了实现多步预测的目的，将隐藏状态矩阵转换为所需要的维度，经过  $1 * 1$  卷积运算，得到  $T'$  步预测结果  $\bar{X} \in \mathbb{R}^{T' \times N \times C}$ 。

$$\bar{X} = \text{Conv}(\text{Conv}(X_{hid})) \quad (15)$$

## 3. 实验设计与验证

### 3.1. 实验数据与对比基线

实验选取的数据集采用公开数据集 PEMS04 与 PEMS08，实验数据由加州高速路网采集，路网中的传感器超过 39000 个，采样周期为 5 分钟，每小时由 12 组不同的数据，如表 1 所示为数据的基本情况。

**Table 1.** Dataset description

**表 1.** 数据集描述

数据集	节点数量	时间跨度	数据长度
PeMS04	307	2018/1/1~2018/2/28	16,992
PeMS08	170	2016/7/1~2016/8/31	17,856

另外选取多组基线模型与现有模型对比，其中包括 HA、SVR、VAR 基本预测方法，同时还包括一些近年文献中的模型：

- 1) HA: 历史平均模型，相同时序位置的数据均值作为模型预测值；
- 2) SVR: 支持向量回归，通过核函数映射到特征空间，进行回归处理；
- 3) VAR: 向量自回归模型，多变量序列的自回归方法，大幅降低信息长范围丢失问题；
- 4) DCRNN (Li et al. 2018) [16]: 扩散图卷积神经网络，引入编码器 - 解码器结构，通过循环神经网络捕捉时间依赖；
- 5) MTGNN (Wu et al. 2020) [17]: 多元时间序列预测图神经网络，首次使用 GNN 应用于多元序列预测；
- 6) GMAN (Zheng et al. 2019) [18]: 图多头注意力网络，采用编码器 - 解码器结构，由多个时空块组成；

7) Traffic Transformer (Yan et al. 2021) [19]: 提出全局编码器和全局-局部解码器提取和融合交通流的空间依赖性。

### 3.2. 实验设计与环境

为了能够评价模型的性能, 采用以下三种度量函数作为评测模型性能的依据。

平均绝对误差(Mean Absolute Error, MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (16)$$

均方根误差(Root Mean Square Error, RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (17)$$

平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (18)$$

其中  $\hat{y}_i$ 、 $y_i$  分别为预测值与观测值,  $n$  作为表示样本数量。

实验的环境均在 Windows 操作系统下, 系统配置 Inter(R) Core(TM) i9-10900X @ 3.70GHz GPU: NVIDIA GeForce RTX 3090, 在 Pycharm 集成开发环境下编译, 基于 Pytorch 框架和 Python3.9 实现模型的预测。

### 3.3. 实验结果与分析

为了验证模型的有效性, 根据现有的两组公开数据集测试模型的性能, 同时与选取的基线模型继续对比, 验证模型的可靠性。实验中对于数据集按照训练集、验证集和测试集进行划分, 其中各占 60%、20%、20%, 采用 Adam 优化器训练, 同时学习率设置为 0.001, 批量大小为 16, 利用过去 12 个步长的历史时间数据预测未来 12 个时间步长(1 小时)的交通流量。

实验结果如下表 2 所示, 分别在 PeMS04 和 PeMS08 两个数据集上验证了模型的性能, 实验中采用的三个指标作为评价指标, 通过与其他基线模型的对比结果如表 1 中所示, 其中从表中可以看到 HA 模型在预测表现上较差, 模型仅仅考虑到时间上的相关性, 其次 SVR、VAR 模型无法捕捉到非线性时空动态依赖关系。

**Table 2.** Prediction effect under different models on PeMS04

**表 2.** PeMS04 和 PeMS08 在不同模型下的预测效果

模型	PeMS04			PeMS08		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
HA	38.03	59.24	27.88	38.03	59.24	27.88
SVR	28.66	44.6	19.15	23.23	36.12	14.65
VAR	23.64	36.56	18.06	22.32	33.82	14.47
DCRNN	22.74	36.58	14.75	18.19	28.18	11.24
MTGNN	19.09	31.55	12.95	15.38	24.94	10.17



Continued

GMAN	19.14	31.62	13.19	15.31	24.92	10.14
Traffic Transformer	18.92	31.35	<b>12.72</b>	15.19	<b>24.88</b>	9.93
DAST-Transformer	<b>18.44</b>	<b>31.30</b>	12.92	<b>14.58</b>	24.90	<b>9.01</b>

\*注意：每列最大值用粗体标注。

对于基于神经网络的模型来说，其中 MTGNN 优于 DCRNN，DCRNN 虽然考虑到扩散的过程情况，但是对于时空信息的捕捉没有 MTGNN 更完善，忽视了动态特性，另外的 GMAN 与 MTGNN 差距较小，模型都是基于图神经网络所展开设计的。对于 Traffic Transformer 模型而言，虽然其在 PeMS04 数据集上的 MAPE 以及在 PeMS08 数据集上的 RMSE 的值优于 DAST-Transformer，但是其并未考虑到空间传播的延迟情况以及长、短距离节点的影响，且当前模型在数据集 PeMS04 和 PeMS08 上的表现较 Traffic Transformer 有提升，显著优于其他方法。

### 3.4. 消融分析

为了验证所提出模型中不同组成部分的有效性，故将其中的一些部分网络去除，形成该模型的变种，并在数据集 PeMS04 上进行验证，消融实验包括以下的变种：

1) 无 mask 矩阵：在模型的基础之上，去除其中的 mask 矩阵参与的语义与位置空间自注意力机制，保留空间自注意力机制；

2) 无延迟感知模块：在模型的基础之上，去除延迟感知模块(Delay Awareness)。

图 3 中展示了两种情况下的模型性能，其中 mask 矩阵保证在编码曾计算注意力机制时，仅对有效序列长度进行计算，进一步提升预测的精度和有效性。另外延迟感知模块对当前预测性能影响较大，该模块实现空间信息在节点之间传播的情况，有益于节点的时空依赖的完善，从而提升模型的预测性能。

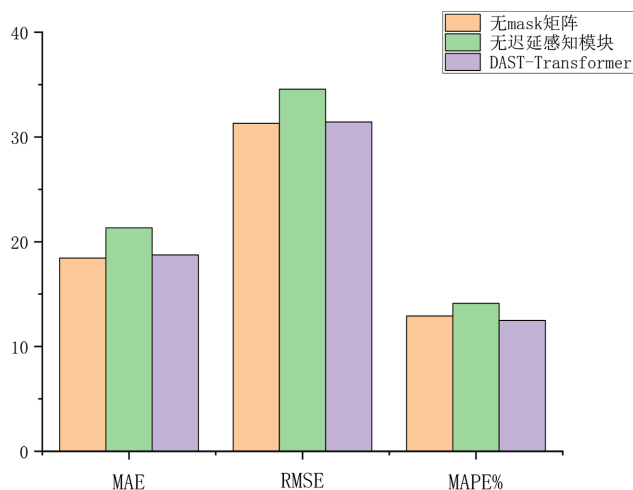


Figure 3. Comparison resultson PeMS04

图 3. PeMS04 数据集对比结果

## 4. 结束语

本文提出一种融合时空动态相关性以及长短距离内的区域交通流变化和空间信息传播延迟影响的 Transformer 模型，其中以时空数据作为嵌入层，将历史交通流数据作为输入，经过时空编码层的处理，

其中时空编码层围绕时间自注意力机制与空间自注意力机制，最后通过跳跃连接和输出层处理，从而构建完整的预测模型。模型在两个不同的真实数据集上验证 DAST-Transformer 的性能，将模型与其他基线模型相比较，从相关指标可以得出该模型在不同的数据集下能够表现出良好的预测性能。

本文设计的模型仍有不足之处，在未来的工作中将进一步完善，增加一些外部因素或者扰动，提升模型的预测精度。同时在其他一些时空预测任务中也可以将 DAST-Transformer 应用其中。

## 参考文献

- [1] Yin, C.T., Xiong, Z., Chen, H., *et al.* (2015) A Literature Survey on Smart Cities. *Science China Information Sciences*, **58**, 1-18. <https://doi.org/10.1007/s11432-015-5397-4>
- [2] Nagy, A.M. and Simon, V. (2018) Survey on Traffic Prediction in Smart Cities. *Pervasive and Mobile Computing*, **50**, 148-163. <https://doi.org/10.1016/j.pmcj.2018.07.004>
- [3] Zivot, E. and Wang, J. (2006) *Modeling Financial Time Series with S-PLUS*. Springer, New York.
- [4] 傅贵, 韩国强, 遯峰, 许子鑫. 基于支持向量机回归的短时交通流预测模型[J]. 华南理工大学学报(自然科学版), 2013, 41(9), 71-76.
- [5] Djenouri, Y., Belhadi, A., Lin, J.C.-W. and Cano, A. (2019) Adapted K-Nearest Neighbors for Detecting Anomalies on Spatio-Temporal Traffic Flow. *IEEE Access*, **7**, 10015-10027. <https://doi.org/10.1109/ACCESS.2019.2891933>
- [6] Ben-Gal, I. (2008) Bayesian Networks. In: Ruggeri, F., Kenett, R.S. and Faltin, F.W., Eds., *Encyclopedia of Statistics in Quality and Reliability*, Wiley, Hoboken. <https://doi.org/10.1002/9780470061572.eqr089>
- [7] Vinayakumar, R., Soman, K.P. and Poornachandran, P. (2017) Applying Deep Learning Approaches for Network Traffic Prediction. 2017 *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, 13-16 September 2017, 2353-2358. <https://doi.org/10.1109/ICACCI.2017.8126198>
- [8] Smith, B.L. and Demetsky, M.J. (1994) Short-Term Traffic Flow Prediction Models—A Comparison of Neural Network and Nonparametric Regression Approaches. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, San Antonio, 2-5 October 1994, 1706-1709. <https://doi.org/10.1109/ICSMC.1994.400094>
- [9] Ranjan, N., Bhandari, S., Zhao, H.P., Kim, H. and Khan, P. (2020) City-Wide Traffic Congestion Prediction Based on CNN, LSTM and Transpose CNN. *IEEE Access*, **8**, 81606-81620. <https://doi.org/10.1109/ACCESS.2020.2991462>
- [10] Yu, B., Yin, H. and Zhu, Z. (2018) Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 3634-3640. <https://doi.org/10.24963/ijcai.2018/505>
- [11] Zhao, L., Song, Y., Zhang, C., *et al.* (2019) T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, **21**, 3848-3858. <https://doi.org/10.1109/TITS.2019.2935152>
- [12] Hu, J., Lin, X. and Wang, C. (2022) DSTGCN: Dynamic Spatial-Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Sensors Journal*, **22**, 13116-13124. <https://doi.org/10.1109/JSEN.2022.3176016>
- [13] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate. <https://arxiv.org/abs/1409.0473>
- [14] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [15] Li, G.Y., Zhong, S.H., Chan, G., *et al.* (2021) A Lightweight and Accurate Spatial-Temporal Transformer for Traffic Forecasting. <https://arxiv.org/abs/2201.00008>
- [16] Li, Y., Yu, R., Shahabi, C. and Liu, Y. (2017) Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. <https://arxiv.org/abs/1707.01926>
- [17] Wu, Z., Pan, S., Long, G., *et al.* (2020) Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Virtual Event, 6-10 July 2020, 753-763. <https://doi.org/10.1145/3394486.3403118>
- [18] Zheng, C., Fan, X., Wang, C. and Qi, J. (2020) GMAN: A Graph Multi-Attention Network for Traffic Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 1234-1241. <https://doi.org/10.1609/aaai.v34i01.5477>
- [19] Yan, H., Ma, X. and Pu, Z. (2021) Learning Dynamic and Hierarchical Traffic Spatiotemporal Features with Transformer. *IEEE Transactions on Intelligent Transportation Systems*, **23**, 22386-22399. <https://doi.org/10.1109/TITS.2021.3102983>