

时空特征变分学习的交通流预测模型

欧阳毅[†], 汤文燕, 邵泳博, 黎晏伶

(浙江工商大学 管理工程与电子商务学院, 浙江 杭州 310018)

摘要: 交通流量时空预测是智能交通系统的关键任务. 针对城市交通流序列的非线性和多模态特性, 本文提出了一种基于时空特征融合的变分学习模型(ST-FVAE), 采用局部时空特征融合和全局特征融合两个阶段对具有图空间拓扑特性的交通流数据进行预测. 局部特征融合模块由时间卷积残差单元和图卷积神经网络(GCN)构成, 提取交通节点的局部时间特征信息, 并利用GCN将空间拓扑信息嵌入到局部时间特征信息中. 通过基于局部时空图特征融合的变分自编码器交通流预测模型, 学习全局时空相关性特征. 在全局时空特征融合变分自编码器的学习过程中, 为使 Q 分布能够逼近实际数据 P 分布, 通过最大化似然函数的变分推断证据下界(ELBO)使得两个分布之间的KL散度最小化, 提出了计算分布期望的KL差异构造训练损失函数的方法, 进一步提高预测准确率. 通过对交通流数据集和交通速度数据集的预测实验结果表明: 本文提出的方法在交通流量和速度的预测方面都具有较好的预测特性, 对于30 min和60 min的预测鲁棒性更好.

关键词: 交通流预测; 时空融合; 变分自编码器; 图卷积

引用格式: 欧阳毅, 汤文燕, 邵泳博, 等. 时空特征变分学习的交通流预测模型. 控制理论与应用, 2025, 42(1): 158 – 166

DOI: 10.7641/CTA.2023.21057

Spatial-temporal feature variational inference model for traffic flow forecasting

OUYANG Yi[†], TANG Wen-yan, SHAO Yong-bo, LI Yan-ling

(School of Management Engineering and E-Commerce, Zhejiang Gongshang University, Hangzhou Zhejiang 310018, China)

Abstract: Traffic flow spatial-temporal data forecasting is a crucial task for intelligent transportation systems. This paper proposes a variational learning model based on the spatio-temporal feature fusion (ST-FVAE). The model aims to address the nonlinear and multi-modal features of urban traffic flow sequences by utilizing local spatiotemporal feature fusion and global feature fusion. It also takes into account the graph spatial topological characteristics to predict traffic flow data. The local feature fusion module is composed of a temporal convolutional residual unit and a graph convolutional neural network model (GCN). It extracts the local temporal feature information of traffic nodes and uses GCN to embed the spatial topological information into the local temporal feature information. We learn global spatial-temporal correlation features via this variational auto-encoder traffic flow prediction model of local spatial-temporal graph feature fusion. During the learning process of global spatio-temporal graph feature fusion variational auto-encoder, to make the variational Q distribution approximate the actual data P distribution, we use the variational inference ELBO (evidence lower bound) that maximizes the likelihood function to minimize the KL scatter between the two distributions. Meanwhile, we construct the training loss function using the KL function property. We perform prediction evaluation on three different large-scale traffic datasets. Experiments show that the model proposed in this paper has better prediction performance in both traffic flow and speed. Moreover, our method is more robust for 30 and 60-minute forecasting.

Key words: traffic flow prediction; spatio-temporal fusion; variational autoencoder; graph convolution

Citation: OUYANG Yi, TANG Wenyan, SHAO Yongbo, et al. Spatial-temporal feature variational inference model for traffic flow forecasting. *Control Theory & Applications*, 2025, 42(1): 158 – 166

收稿日期: 2022-12-03; 录用日期: 2023-11-29.

[†]通信作者. E-mail: oyy@mail.zjgsu.edu.cn; Tel.: +86 13666695906.

本文责任编辑: 季飞.

浙江工商大学“数字+”学科建设管理项目(SZJ2022C004), 浙江省智能交通工程技术研究中心项目(2015ERCITZJ-KF1)资助.

Supported by the Zhejiang Gongshang University “Digital+” Discipline Construction Management Project (SZJ2022C004) and the Zhejiang Intelligent Transportation Engineering and Technology Research Center Project (2015ERCITZJ-KF1).

1 引言

随着城市化持续发展, 交通已成为城市运行中的关键因素. 然而, 道路容量和私家车使用量的不匹配加剧了交通拥堵、不均衡出行、安全风险提高、环境污染加剧等交通问题. 面对这些问题, 交通流预测成为提升交通流转效率的重要措施. 精准的交通流预测可以根据历史或实时数据, 按照设定的预测算法, 为城市交通系统的规划和管理提供未来一定步长的交通信息状况, 帮助城市管理者 and 交通规划者及时采取协调, 降低交通问题带来的负面影响.

目前, 国内外交通流量预测模型可大致分为3类: 线性模型、非线性模型以及神经网络模型. 线性模型包括历史平均模型(history average model, HA)、时间序列模型以及卡尔曼滤波模型(Kalman filtering, KF). 历史平均模型取交通流量数据的平均值作为交通流的预测值. 但是, 道路传感器数据随时间动态变化并在时间序列中呈现周期相关性, 相关学者便将交通流预测问题转换为时间序列预测问题, 主要采用与先前周期信息相结合的统计模型, 通过选择适当的回归模型进行流量建模预测. 这类模型对交通流序列数据的平稳性要求很高, 当交通流因道路维护或节假日存在异常值时, 模型的预测性能就会有所降低. 卡尔曼滤波模型和差分整合移动平均自回归模型(autoregressive integrated moving average model, ARIMA)不同, 这类模型对于非平稳时间序列数据的适应力强. Kumar^[1]提出并评估了一种基于卡尔曼滤波技术(Kalman filter technique, KFT)的预测方案, 该方案只需要有限的输入数据. 杨高飞等^[2]将ARIMA和卡尔曼滤波模型组合对交通流进行预测, 和单一模型相比, 该方法达到了更优的预测精度. 之后, 卡尔曼滤波模型的变种自适应卡尔曼滤波模型^[3]、扩展卡尔曼滤波算法^[4]均被广泛应用于短时预测、结构识别、协同估计等领域. 这些模型是对卡尔曼滤波方法在非线性领域的扩展, 但是这种方法当环境很大时, 特征点会非常多, 状态向量维度会变得非常大, 容易出现维度爆炸的问题. 而且线性模型算法结构依赖于统计学知识, 面对交通流复杂的时空动态性, 预测性能仍需进一步提升.

非线性模型利用机器学习对数据特征自主学习, 不仅对交通流时序数据具有较好的拟合效果, 而且非线性模型可移植性高的特点, 促使算法可以应用于不同的交通道路, 其中, K 近邻(K -nearest neighbor, K-NN)算法和支持向量机(support vector machine, SVM)比较具有代表性. Davis等^[5]于1991年首次提出用KNN算法进行交通流量预测, 可通过寻找最优 K 值调整模型的预测性能. SVM采用核函数方法解决了非线性可分的问题, 具有良好的泛化能力和保证全局最小值的优势使其被广泛用于短时交通流量预测^[6], 这种

模型性能取决于核函数的选择. 另外, Cetiner等^[7]初步尝试利用人工神经网络(artificial neural network, ANN)对交通历史数据建模以改变场景进行预测. 然而, 非线性模型并不能完全适用于实际的交通路网数据. 大规模交通流数据及其高维度特征会影响模型寻找最优参数, 降低模型运作的效率.

深度神经网络(deep neural network, DNN)的应用开辟了交通流预测的新篇章. 交通流数据信息通过在神经网络层级之间进行提炼, 可以获取深层次的特征信息, 再利用深层网络结构逼近交通流的变化规律, 拟合时空非线性关系, 从而提高模型预测性能^[8]. 循环神经网络(recurrent neural network, RNN)较强的记忆功能和时序建模能力, 通常被用作对交通流的流量序列建模^[9], 但实时的交通流数据易受道路状态、传感器、天气温度等外部事件的影响, 这说明交通流数据在空间上同样呈现复杂的动态性. RNN并不能有效对交通流的空间特征进行建模, 仅对交通流时序特征建模存在局限性. 因此, 相关学者开始对交通路网的拓扑结构进行建模. 学者Ma等^[10]则提出了一种深度卷积神经网络(convolutional neural network, CNN), 通过二维时空矩阵转换为交通流量的时空关系图并对交通特征进行提取. 由于考虑的问题规模较大, 该方法并未对局部空间特征进行分析, 对此, STGCN模型^[11]设计了一种包含图卷积网络与门控卷积神经网络的时空卷积模块, 对道路上流量速度进行预测.

交通路网是一个复杂系统, 当前区域的交通流量预测值不仅受到该道路历史交通流数据的影响, 相邻区域的数据、神经网络的设置及潜在特征(如: 车辆特征、传感器等), 同样是重要影响因素. 为提高模型的准确率, 短时交通流量预测模型逐渐衍生出基于算法或模型的组合预测模型, 利用神经网络克服交通路网复杂的非线性特征, 同时通过算法优化核心参数, 加快训练速率. Tan等^[12]将移动平均(moving average, MA)、指数平滑(exponential smoothing, ES)、ARIMA和神经网络模型聚合后对交通流量预测. Yang等^[13]对模拟退火算法、ARIMA和反向传播神经网络(back propagation in neural network, BPNN)进行组合, 以此增强BPNN的全局优化能力. 但是这些组合模型尚未考虑到其他影响因素对模型造成的随机性. 如何细化时空特征的提取融合又是一项重要的研究课题^[14-15]. Zhao等^[16]提出了时间图卷积网络(temporal graph convolutional network, T-GCN)模型, 该方法将图卷积网络(GCN)和门控递归单元(gate recurrent unit, GRU)相结合, 从而捕捉时空相关性. Duan等^[17]设计了空间图注意力分量(spatial graph attention component, S-GAT)、时间图注意力分量(temporal graph attention component, TGAT)和融合层3个组件动态整合空间维

度、时间维度和周期特征的相关性. 现有研究大多数分别提取交通流的时间和空间特征, 再将两种特征信息进行融合, 从而挖掘出交通流的时空相关性. 这种方法相较于单方面建模时间或空间特征, 预测性能会更加优越. 预测性能取决于神经网络的选择以及模型参数的设置.

本文考虑交通路网的时空特征信息以及潜在表示的影响, 提出了一种基于时空特征融合的变分自编码器交通流量预测模型 (spatio-temporal feature fusion variational autoencoder, ST-FVAE). 本文主要贡献如下: 1) 构建了局部时空特征融合模块 (local spatio-temporal feature, Local STF), 利用时间卷积残差单元提取交通节点的局部时空特征信息, 并利用图卷积神经网络 (GCN) 将空间拓扑信息嵌入局部时间特征信息中, 再采用多层感知机 (multi-layer perceptron, MLP) 和掩码操作将时空特征进行融合, 保留影响预测值的关键区域特征; 2) 提出了全局特征变分自编码器 (global variational autoencoder, GVAE), 将融合后的特征信息和极大特征编码器共同强化学习交通路网的特征信息, 进一步提高了模型的预测准确性和泛化性; 3) 预测实验表明, ST-FVAE 在各项评价指标方面均优于基准模型, 并且在 30 min 和 60 min 预测中具有更好的鲁棒性.

2 时空特征融合

本文将交通路网定义为 $G(V, E)$, 其中: $v_i \in V$ 为传感器节点, $e_{ij} \in E$ 表示节点 i 与节点 j 对应的边, 邻接矩阵采用 $A \in \mathbb{R}^{N \times N}$ 表示. 图中的每个节点, 其边的数量和流量都不一样, 考虑到中心节点的特征值会大于边缘节点, 对道路拓扑结构信息进行归一化处理 $\hat{A} = D^{-1/2}(A+I)D^{-1/2}$, 其中: D 为度矩阵, $I \in \mathbb{R}^{N \times N}$ 为单位矩阵. $T = \{t_1, t_2, \dots, t_n\}$ 为观察数据的时间点, $L^t = D^t - \hat{A}^t$, L^t 是时刻 t 的交通图的拉普拉斯矩阵.

2.1 基于图卷积的局部空间特征嵌入

针对不同区域及道路之间的交通流量之间存在相关性, 本文采用图卷积神经网络获取其中空间特征相关性信息. t 时刻空间卷积单元的输入为 $\hat{X}^t \in \mathbb{R}^N$ (N 为传感器节点个数) 和归一化的图邻接矩阵 \hat{A} . 为获取局部空间拓扑信息, 将当前时刻的空间拓扑信息嵌入 X^t 特征, 即: $X_g^t = \sum_{k=0}^K \theta_k L^k X^t$, 其中 θ 为卷积核参数. K 值越大对周围节点的信息聚合能力越强, 包含的空间特征信息越多. 但实际道路交通情况, 随着 K 的增大, 其他道路对所选取的中心道路的影响会变小, 因此的取值不易过大 (实验中 $K = 3$). 本文将此作为局部空间特征嵌入, 再利用两层的 GCN^[18] 对交通路网中时刻节点 t 的空间信息聚合如下:

$$S(X^t) = \sigma(\hat{A} \text{ReLU}(X_g^t W_0) W_i), \quad (1)$$

其中: W_i 为第 i 层神经网络的权重参数矩阵, σ 为 Sigmoid 激活函数.

2.2 基于时间卷积残差单元的特征嵌入和融合

在交通领域, GRU 的引入可以显著改善受时间特性影响的交通流预测模型^[16]. 时间特征提取单元的输入是 $t-1$ 与 t 时刻的交通流数据 X^{t-1} 和 X^t . 利用 σ 激活函数控制上一时刻的信息输入, 得到 $t-1$ 时刻交通流势能数据, 与 t 时刻流量信息进行融合操作得到 $\hat{X}^t = X^t \oplus \sigma(X^{t-1})$. 与 GRU 方式不同, 本文采用时间维度上的卷积操作提取时间相关性特征, 卷积核的大小决定了融合时间范围的深度. 利用该方法得到 t 时刻时间融合特征信息; 为了防止在神经网络训练过程出现网络退化现象, 将 t 时刻时间融合特征与 X^t 进行残差连接; 将融合后的特征通过 σ 激活函数输出, 得到时刻的时间特征 $T(X^t)$, 其形式化的定义如下:

$$T(X^t) = \sigma(W * X^t \oplus \sigma(X^{t-1})), \quad (2)$$

式中: \oplus 表示时间特征提取单元中的两个张量 (element-wise) 之和, $*$ 为时间维度上的卷积操作, $W \in \mathbb{R}^{K \times C}$ 为时间维度上的卷积核权重矩阵. 利用局部空间特征嵌入单元 (local spatial feature cell, LSC) 和局部时间特征卷积残差单元 (local temporal feature cell, LTC) 的输出, 通过矩阵求和操作, 以迭代的方式实现时空特征的融合.

交通运输网络通常是一个复杂的大规模网络, 这种网络中的数据量庞大且呈高度非线性. 利用传统模型处理输入数据, 一方面运算能力差、计算缓慢; 另一方面预测性能低. 因此, 对于时空特征的融合, 本文采用多层感知网络 (MLP) 实现数据的并行式处理, MLP 的每个节点均含有一个非线性激活函数的神经元, 能够有效获取时空特征的依赖性. 可以将时间和空间特征进行综合训练, 训练速度比较快且具有非线性数据处理能力. 但是, MLP 神经元之间的全连接式会导致处理图数据时效率不高, 导致特征学习不够充分. 基于 MLP, 采用掩码操作弱化该局限性. 对影响预测值较小的值进行屏蔽, 从而使该特征值的正向传播失活, 最终达到掩码的目的.

考虑到变分自编码的自适应性, 本文将局部时空特征融合的学习任务分解为两个阶段, 首先采用时间卷积残差单元在时间维度上提取交通节点数据的时间特征信息; 再利用变分自编码器进行时空特征融合处理. 在第 1 阶段, 利用多层感知神经网络 (MLP), 对空间特征提取单元和时间特征提取单元的输出进行融合处理. 通过利用 MLP 和特征掩码操作, 保留对预测值影响较大的时空特征信息, 可进一步提高预测精度. 为简化符号描述, 令 S^t, T^t 分别表示 t 时刻提取

的空间特征 $S(X^t)$ 和时间特征 $T(X^t)$. 首先将图邻接矩阵 \hat{A} 和 $t-1$ 时刻的交通流数据 X^{t-1} 作为空间特征提取单元的输入, 获取 S^{t-1} , 同时 X^t 经过时间卷积残差单元获取 T^t , 在训练过程中对时空特征提取单元进行归一化处理, 保持网络中每一层中的时空特征为相同的变化区间. 通过元素维度的矩阵求和获取时空特征的预融合和终融合, MLP表示多层感知网络(实验中选取了3个全连接层), \oplus 表示相对应元素加操作, 其中 w 的取值为0.5, 表示时间和空间特征各取一半. 后面是将 $S(X^t)$ 和 $F(X^t)$ 拼接. 未经过MLP直接进行矩阵求和部分为预融合, 而经过MLP和非线性激活层的输出作为终融合第 $K+1$ 层的输入, 具体结构如图1所示. 时空特征预融合和终融合处理分别如式(3)和式(5)所示:

$$H_{\text{pre}}^k = w(S^{t-1}) \oplus (1-w)T^t, \quad (3)$$

$$H_{\text{fin}}^k = \Gamma(\sigma(\text{MLP}(S^{t-1}) \oplus \text{MLP}(T^t))), \quad (4)$$

$$F^t = \text{MLP}(H_{\text{fin}}^k \| H_{\text{pre}}^k), \quad (5)$$

其中: $\Gamma(\cdot)$ 代表mask操作, $\|$ 表示张量(tensor)拼接操作. 局部时空特征具体融合过程为: 首先分别对时间局部特征 T^t 和空间局部特征 S^{t-1} 进行批次归一化处理, 一方面通过 \oplus 对时空特征直接进行预融合, 另一方面通过MLP并行式处理时空特征, 利用 \oplus 矩阵求和后进行mask操作, 以此屏蔽影响力度小的特征信息. 最终将第 K 层时刻的输入 H_{fin}^k 与图拓扑结构信息构建的时空特征融合单元进行模块堆叠, 对于每一个时空融合特征模块的输出都进行了归一化处理, 加快网络训练过程中的收敛速度与网络稳定性. 对于 K 层的时空特征输出作为第 $K+1$ 层的时空特征模块的输入(实验中 $K=3$), 通过在网络中叠加这样模块达到特征融合增强学习的目的, 从而获取局部时空特征融合信息.

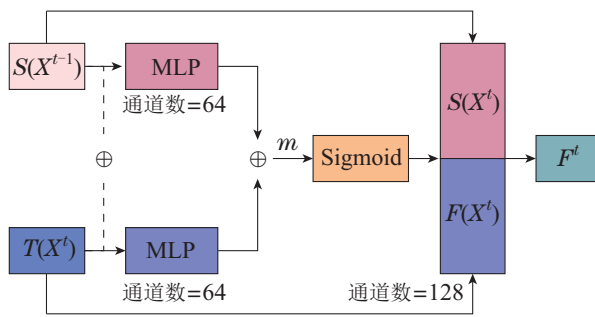


图 1 时空特征融合

Fig. 1 Spatio-temporal feature fusion

3 基于时空特征融合变分自编码器的交通流预测模型(ST-FVAE)

数据样本的分布 $p(X)$ 很难直接获取, 假设对高维信息数据的样本 X 预测 \hat{X} 是由不可直接观测的连续随机变量 Z 按照条件分布生成的, 即: $\hat{X} \sim p_\theta(\hat{X}|Z)$. 本文采用一种满足均值场假设条件的 Q 分布 $q_\phi(Z|\hat{X})$

(其中 $\hat{X} = (X, F)$), 利用变分推断的思想通过 q_ϕ 逼近 $p_\theta(\hat{X}|Z)$. 具体处理如下: 首先构造编码器(Encoder), 对 X 和时空特征融合后的 F 进行混合编码(为简化符号描述用 F 表示 F^t)获取隐变量 Z , 通过解码器(Decoder)输出预测值 \hat{X} . 由于交通流序列的非线性和多模态特性, 为了使 $q_\phi(Z|\hat{X})$ 能够逼近 $p(X)$, 设 Z 的分布为多元高斯分布, 编码器的输出为多元高斯分布的均值 μ 和方差 σ , 通过最大化似然函数的证据下界(evidence lower bound, ELBO), 使得两分布间的KL散度(kullback-leibler divergence)最小化, 实现变分自编码器^[19](variational autoencoder, VAE)对交通流数据的预测.

引理 1^[20] 从分布 $p(z)$ 到分布 $q(z)$ 的KL散度, 表示为 $\text{KL}(q(Z)||p(Z))$, 在对 $(q(z), p(z))$ 上是凸函数. 有

$$\text{KL}(E_{\psi}q(z|\psi)||p(z)) \leq E_{\psi}\text{KL}(q(z|\psi)||p(z)). \quad (6)$$

定理 1 变分自编码器中, $q_\phi(Z|X)$ 向 $p_\theta(Z|X)$ 的逼近过程, 等价于最小化

$$E_{F \sim q_\phi(F)}[\text{KL}([q_\phi(Z|\hat{X})||p_\theta(Z)])]. \quad (7)$$

此定理的证明可以通过以下步骤形成:

步骤 1 用 $\text{KL}(q_\phi||p_\theta)$ 散度衡量分布 Q 与实际数据分布 P 之间的差异, 定义如下:

$$\text{KL}(q_\phi(Z|X)||p_\theta(Z|X)) = \int q_\phi(Z|X) \log \frac{q_\phi(Z|X)}{p_\theta(Z|X)} dZ. \quad (8)$$

步骤 2 优化目标是最小化KL散度, 则有

$$\begin{aligned} \log p_\theta(X) - \text{KL}(q_\phi||p_\theta) &= \\ E_{Z \sim q_\phi(Z|\hat{X})}[\log p_\theta(X|Z)] - \\ \text{KL}(q_\phi(Z|X)||p_\theta(Z)). \end{aligned} \quad (9)$$

步骤 3 由 $\log p_\theta(X)$ 通常为常量, 最小化 $\text{KL}(q_\phi(Z|X)||p_\theta(Z|X))$ 散度, 等价于最大化ELBO. 因此

$$\begin{aligned} \text{ELBO} &= E_{Z \sim q_\phi(Z|X)}[\log p_\theta(X|Z)] - \\ \text{KL}(q_\phi(Z|X)||p_\theta(Z)) &= \\ E_{Z \sim q_\phi(Z|X)}[\log p_\theta(X|Z)] - \\ \text{KL}(E_{F \sim q_\phi(F)}[q_\phi(Z|\hat{X})||p_\theta(Z)]. \end{aligned}$$

结合引理1, 可得

$$\begin{aligned} \text{KL}_E(Q|P) &= (E_{F \sim q_\phi(F)}[q_\phi(Z|\hat{X})||p_\theta(Z)] \leq \\ &E_{F \sim q_\phi(F)}[\text{KL}([q_\phi(Z|\hat{X})||p_\theta(Z)])]. \end{aligned}$$

根据定理1, 可获取时空特征融合条件下, 变分自编码器KL损失的上界. 通过减少上界达到 Q 分布向分布 P 逼近. 将变分自编码器预测模型中训练 Q 分布的目标损失函数定义为

$$L_{\text{KL}} = E_{F \sim q_{\phi}(F)}[\text{KL}([q_{\phi}(Z|\hat{X})||p_{\theta}(Z))]]. \quad (10)$$

ST-FVAE的训练过程如图2所示,其中Encoder由

特征编码器和local STF构成,而Decoder模块由隐变量 $Z = 32$ 维和256个隐含层节点个数构成。

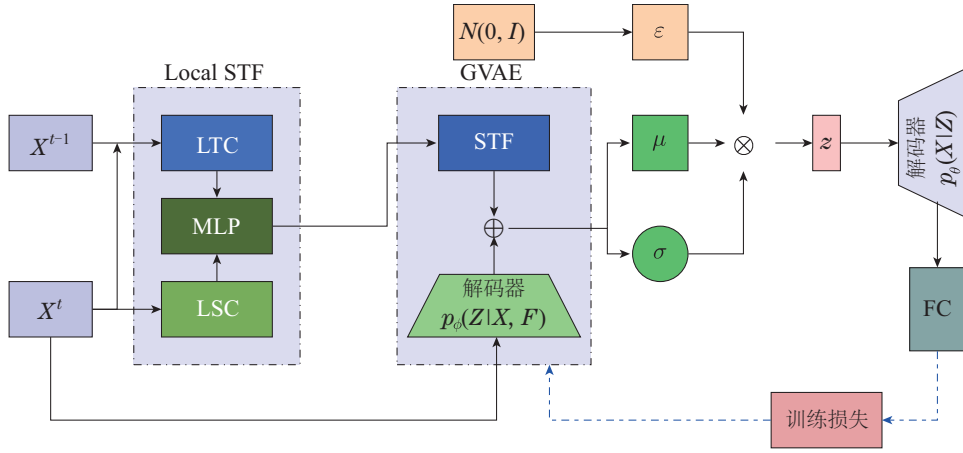


图2 ST-FVAE预测模型框架图

Fig. 2 Overview of the architecture of ST-FVAE model for traffic flow forecasting

4 实验分析

本文ST-FVAE模型算法,在公开数据集PeMS-BAY, PeMS04和PeMS08上进行车速和流量预测,用于验证算法性能.在实验中,将数据集60%数据划分为训练集,20%数据为测试集,20%数据为验证集.

4.1 评价指标

本文分别采用平均绝对误差(mean absolute error, MAE),均方根误差(root mean squared error, RMSE)和平均绝对百分比误差(mean absolute percentage error, MAPE)等评价指标^[21-23]与各种基准模型量化预测值和实际数据之间的误差水平.为进一步评价本模型与其它方法预测值的绝对差异,本次使用最大绝对误差(absolute error, AE)指标进一步比较模型性能.

AE:度量预测结果与真实数据间绝对偏差的最大值,最大值越小说明预测值和观测值越接近.

$$\text{AE} = \max(|\hat{y}_i - y_i|). \quad (11)$$

4.2 环境及参数设置

本文的实验均是在Python3.6, PyTorch1.8的环境中进行的,并采用Xavier方法对ST-FVAE进行网络参数初始化设置,采用批归一化方法加快收敛速度.ST-FVAE超参数包括批量大小(batch size)、学习率(learning rate)、训练迭代次数(epoch)、节点时间步长、节点特征通道容量(channels)、各模块的神经网络层数、时空特征提取模块的卷积层数,以及隐变量的参数维度.迭代次数是所有训练样本实验的次数,节点时间步长影响历史时间窗口长度和预测未来时间窗口的设置.经过反复实验的调整, Batch size 设置为64, Epoch 设置为300, MLP和VAE的Learning rate 设置为0.001.在15 min预测实验中,历史时间窗口长度设置为12,预测未来时间窗口长度为3(即用过去60 min信息,来预测接下来15 min的数值).在尺度为30 min和60 min预

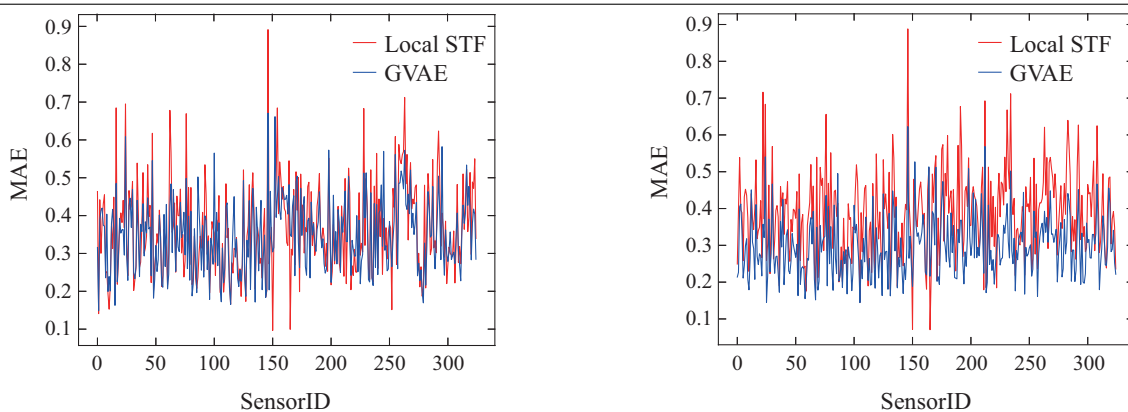
测中,未来时间窗口长度分别为6和12.

ST-FVAE中时间特征处理模块中卷积核大小设置为(1, 3),有3层2D卷积层,而时空特征提取模块的神经网络层数为2.隐变量 Z 的参数维度设置为 $N \times 32$ 维(N 为传感器个数).特征通道容量表征输入的特征数量信息,局部空间特征嵌入单元的全连接层和空间卷积的容量分别是128和256,局部时间特征卷积残差单元的通道容量为64,时空特征经过多层感知机的通道均为64,融合后的特征的通道变为128.

4.3 实验结果对比

为验证实验结果数据的有效性和对比的可用性,本文采用文献作者提供的算法代码(默认的超参数设置)和标准函数库(ARIMA和SVR方法)进行仿真实验复现.训练和测试环节,所有模型所使用的实验环境、PeMS数据集、超参数的初始化设置以及评价指标的选择均一致,这样可以确保所有方法在相似的实验环境下进行比较.同时,在实验过程中,本文对所有模型采用相同的Epochs进行多次迭代,并计算平均结果以减少随机性的影响.通过与各基准文献的原数值结果对比情况对模型进行调试,以验证结果数据的准确性.

首先利用交通流车速数据集PeMS-BAY,针对15 min, 30 min, 60 min这三个不同步长对模型进行训练预测.由于ST-FVAE的隐变量 Z 值是由均值 μ 和方差 σ 共同输出的,在模型迭代学习过程中,对隐变量 Z 的均值 μ 成份进行了分析.实验发现在100次迭代后, μ 值的分布比较随机,同时差异较大,而经过1000次迭代后, μ 值整体是比较平滑的.因此,对于控制变量预测输出的差异,可以通过多次迭代均值 μ 优化减少.图3(a)-(b)分别描述了100次和800次 epoch 迭代, Batch 内12组数据的在单个传感器上的MAE值差异.



(a) 100次迭代Local STF单元与GVAE单元预测差异

(b) 800次迭代Local STF单元与GVAE单元预测差异

图 3 消融性实验

Fig. 3 Ablation experiment

图3中红色线条是仅经过Local STF后的预测值与真实样本的MAE值, 蓝色线条为Local STF单元的输出去往GVAE单元后的预测值与真实样本的MAE值. 这个误差值是 $325 \times 64 \times 12$ 的矩阵相减后, 按照列相加再取绝对值均值(其中325是PeMS-BAY数据集的传感器个数). 从图3(a)的蓝色和红色曲线对比可看出, 仅经过Local STF模块的预测输出的误差明显比经过

GVAE单元的预测误差高, 即红色误差曲线明显要高于蓝色误差曲线.

GVAE可以使误差值维持在一个较小的区间, 对噪声数据具有抑制作用. 经过多次迭代学习后, 从图3(b)可以看出对噪声数据具有抑制作用更加明显.

表1为ST-FVAE与8种基准模型^[11,22-29]在数据集PeMS-BAY上的车速预测结果.

表 1 PeMS-BAY数据集车速预测对比

Table 1 Performance comparison on the PeMS-BAY dataset with different timespan.

算法对比	15 min				30 min				60 min			
	AE	MAE	RMSE	MAPE%	AE	MAE	RMSE	MAPE%	AE	MAE	RMSE	MAPE%
ARIMA ^[24]	64.76	1.62	3.30	3.50	63.12	2.33	4.76	5.40	67.53	3.38	6.50	8.30
STGCN ^[11]	60.47	1.64	3.21	3.04	60.92	1.76	3.57	3.31	64.57	2.01	4.28	3.86
DCRNN ^[22]	47.17	1.38	2.95	2.90	54.95	1.74	3.97	3.90	62.46	2.07	4.74	4.90
GMAN ^[25]	58.77	1.57	2.96	2.77	61.76	1.66	3.23	2.93	62.22	2.56	4.64	4.23
GraphWaveNet ^[27]	50.13	2.04	3.18	3.66	58.13	2.27	3.66	4.06	58.21	2.66	4.48	4.73
ASTGCN ^[23]	62.63	1.60	2.63	2.59	64.09	1.67	3.59	3.39	70.06	2.20	4.92	4.76
AutoFormer ^[28]	69.06	1.80	3.68	4.56	70.15	1.95	4.04	4.78	87.02	2.19	4.61	6.81
ESG ^[29]	51.05	1.51	2.38	3.52	60.63	2.48	4.25	3.98	64.99	2.16	3.87	4.13
ST-FVAE	41.55	1.42	2.43	2.75	48.05	1.68	2.83	2.94	52.34	1.97	3.72	3.84

ST-FVAE在不同步长的预测实验中, AE, MAE, RMSE和MAPE这4种评价指标均优于较多基准模型, 在30 min和60 min预测中, 度量指标会明显弱于其它基准模型. 而ARIMA模型各项性能指标值大多数高于其它基准模型, 尤其在60 min预测中更加明显, 说明ARIMA模型对于60 min历史数据并不理想. 这主要是由于ARIMA模型通过对呈现季节性规律的时间序列建模捕捉交通流时间相关性, 更适用于具有较长历史信息的预测问题. 此外, 深度卷积循环神经网络(diffusion convolutional recurrent neural network, DCRNN)、演化多尺度图神经网络(evolutionary multi-scale

graph neural network, ESG)和自适应时空图卷积网络(attention based spatial-temporal graph convolutional networks, ASTGCN)等预测模型在15 min步长的预测实验会得到较低的度量指标值, 在一定程度上能够提升预测性能, 但是在30 min和60 min步长的预测中, 其性能和其它基准模型相比, 并没有表现出很卓越的预测精度. 例如, DCRNN模型在15 min预测实验中MAE指标值最低, 但是在60 min的预测中其MAE值比STGCN指标值高. 这主要由于基于卷积的STGCN不仅考虑了空间上的邻域, 还对时间的邻域进行建模, 可以捕获交通网络演化过程中复杂的时空相关性. 而

基于递归的DCRNN容易出现梯度消失的问题,因此难以捕捉长时依赖性.但STGCN只能捕捉最基本的时间特征和有用的空间特征,并未对时间和空间特征的融合展开研究,因此性能仍需提升.本文所建模型ST-FVAE和基准模型图多注意力网络(graph multi-attention network, GMAN)均采用编码器-解码器的架构,其中GMAN利用转换注意机制来缓解错误传播,由于注意力机制会对节点赋予相同权重,本文采用基于卷积的变分自编码器获取交通流的时空相关性.VAE变分推断的预测方式促使其具有较好的平稳特性,因此ST-FVAE在各个步长均具有较优越的预测精度,所有度量指标都保持在一个较小的级别,且从15 min拓展至60 min预测过程中,各项指标的增长趋势维持在较小的变化区间.这不仅说明ST-FVAE具有较好的稳健性,也体现了该模型在30 min和60 min数据预测方面的优势.

除了对交通流速度的预测之外,为验证ST-FVAE模型的泛化特性,使用了含流量和速度的数据集PeMS04和PeMS08进行了训练预测.表2-4分别呈现了在PeMS04上15, 30 min和60 min的流量预测对比情况,表5展示了在PeMS08上进行60 min的流量和速度预测指标值.

表2 PeMS04数据集交通流量预测对比(15 min)
Table 2 Performance comparison of 15mins traffic flow forecasting on PeMS04 dataset

算法对比	PeMS04			
	AE	MAE	RMSE	MAPE%
SVR ^[26]	683.43	33.40	36.66	20.75
FC-LSTM ^[21]	641.44	27.61	41.07	20.17
DCRNN ^[22]	624.31	20.37	33.12	18.93
GraphWaveNet ^[27]	569.03	24.04	35.82	15.38
ASTGCN ^[23]	804.98	23.90	36.69	58.18
AutoFormer ^[28]	627.28	19.56	30.82	15.85
ST-Norm ^[30]	335.68	35.39	50.13	16.58
ESG ^[29]	660.91	18.82	30.57	19.42
ST-FVAE	401.54	24.53	37.59	17.40

从表2-4可以看出,传统SVR在数据集PeMS04上的预测性能与基于神经网络的基准模型相比较,性能指标较弱.与近年来提出的其他交通流预测模型相比,循环递归网络模型FC-LSTM^[21]能获取部分时间相关性,因此性能优于SVR^[26],但由于缺乏对时空特征的融合,其性能仍有足够的提升空间.深度神经网络模型DCRNN和GraphWaveNet由于考虑了时空相关性建模,取得了更好的性能提升.ASTGCN引入注意力机制来同时获取空间和时间相关性,进一步优化了时空融合的研究.AutoFormer和ESG模型进化了时间序列的分解建模,例如,AutoFormer具有自相关机制的

新型分解框架,提升了对复杂时间序列的渐进分解能力.ESG采用特定于尺度的图结构学习器对时间序列间相关性进行建模.这两种方法在30 min和60 min均具有优越的预测性能.本文模型ST-FVAE除了进行时空特征融合处理,还考虑了时间周期性的影响,在15 min和30 min预测步长中与其它基准模型的性能指标并无明显差异,但是在60 min预测的RMSE和MAPE相比达到了更好的性能指标,其中RMSE的数值最小,针对异常值更敏感,而ST-FVAE模型的预测值与真实值相差不会很大.这主要是由于变分自适应编码器的解码器模块所产生的数据,经过训练能更好地拟合历史数据趋势.

表3 PeMS04数据集交通流量预测对比(30 min)
Table 3 Performance comparison of 30mins traffic flow forecasting on PeMS04 dataset

算法对比	PeMS04			
	AE	MAE	RMSE	MAPE%
SVR ^[26]	732.12	28.60	44.29	18.30
FC-LSTM ^[21]	661.59	29.67	43.92	21.60
DCRNN ^[22]	624.31	22.67	34.92	19.01
GraphWaveNet ^[27]	613.25	28.90	42.17	17.94
ASTGCN ^[23]	804.98	22.93	35.22	16.56
AutoFormer ^[28]	629.39	20.29	31.85	15.75
ST-Norm ^[30]	348.26	27.74	41.13	22.47
ESG ^[29]	689.85	20.03	32.43	19.44
ST-FVAE	445.32	25.79	39.23	17.35

表4 PeMS04数据集交通流量预测对比(60 min)
Table 4 Performance comparison of 60mins traffic flow forecasting on PeMS04 dataset

算法对比	PeMS04			
	AE	MAE	RMSE	MAPE%
SVR ^[26]	762.35	39.97	60.54	27.01
FC-LSTM ^[21]	692.80	34.46	50.45	25.10
DCRNN ^[22]	704.25	27.20	41.35	23.12
GraphWaveNet ^[27]	623.80	37.46	53.65	22.85
ASTGCN ^[23]	635.87	22.93	35.22	21.56
AutoFormer ^[28]	656.81	21.97	34.16	21.89
ST-Norm ^[30]	696.66	24.48	36.38	25.87
ESG ^[29]	690.34	27.10	35.52	24.62
ST-FVAE	602.07	27.90	32.51	18.83

如表5可知,ST-FVAE在PeMS08的实验结果中MAE和MAPE均表现出较好的预测能力.另外,可以观察到交通流量60 min预测中ST-Norm模型AE指标明显小于ESG和AutoFormer等模型,但MAE指标却较大.为进一步分析这一现象,本文绘制了各模型在PeMS08数据集下各传感器60 min预测数据的AE和

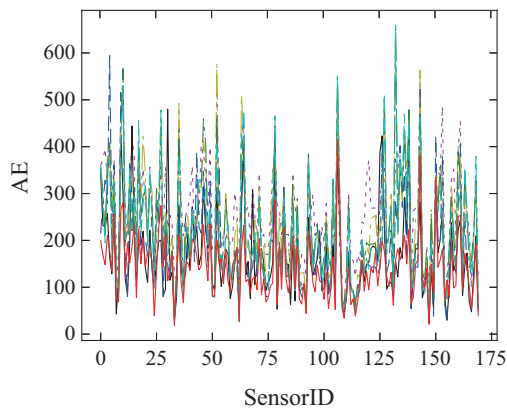
MAE曲线, 如图4. 图4(a)–(b)分别描述了在不同传感器上的AE 和MAE 值差异, 其中横坐标均代表传感器的个数. 从曲线上可以看出部分模型虽然个别传感器

预测AE偏差较大, 但整体MAE指标值处于较低的稳定区间(如ESG蓝色虚线), 原因是AE指标的随机性波动较大, 并不能很好反映整体性能.

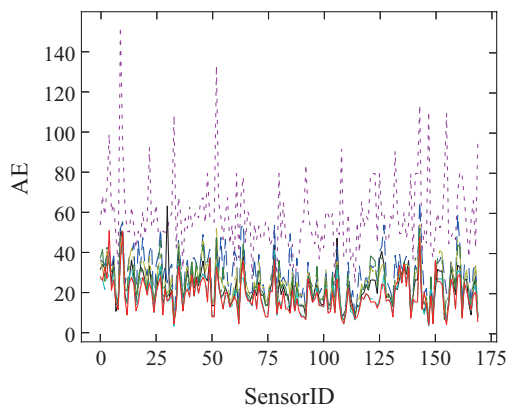
表 5 PeMS08数据集流量及车速预测对比(60 min)

Table 5 Performance comparison of 60mins traffic flow and speed forecasting for PeMS08

算法对比	交通流: 60 min				车速: 60 min			
	AE	MAE	RMSE	MAPE%	AE	MAE	RMSE	MAPE%
SVR ^[26]	683.43	33.40	46.66	44.75	65.12	5.73	9.41	8.78
FC-LSTM ^[21]	508.64	29.09	41.95	22.90	62.36	3.73	6.95	5.81
DCRNN ^[22]	560.14	16.65	25.11	16.19	63.99	1.61	3.77	3.09
GraphWaveNet ^[27]	654.47	25.14	38.68	12.33	62.97	1.78	3.74	2.98
ASTGCN ^[23]	714.25	26.94	38.74	72.87	75.39	1.73	3.83	5.30
AutoFormer ^[28]	661.14	18.91	29.52	18.94	65.36	1.77	4.06	3.44
ST-Norm ^[30]	446.71	22.23	33.07	21.57	55.80	1.57	3.18	2.63
ESG ^[29]	594.13	19.40	31.34	12.93	69.97	1.73	3.74	2.98
ST-FVAE	415.61	18.58	29.71	12.91	62.12	1.49	3.43	2.55



(a) 各模型的AE差异对比曲线



(b) 各模型的MAE差异对比曲线

--- FC-LSTM — ST-Norm - - - ESG
 — DCRNN - - - Autoformer - - - ST-FVAE
 — GraphWaveNet

图 4 PeMS08数据集交通流量60 min预测AE与MAE指标对比

Fig. 4 Performance comparison of AE and MAE for 60 min traffic flow forecasting on PeMS08 datasets

5 总结

本文对交通流量的空间相关性进行分析, 提出了一种时空特征融合的交通流预测方法. 根据交通流量存在的时空相关性设计了LSC和LTC单元来分别提取空间特征和时间特征并融合为时空特征, 构建极大特征编码器来学习数据的原始分布, 利用变分自编码器的结构实现对交通流的预测. 通过减少上界达到 Q 分布向分布 P 逼近, ST-FVAE模型利用变分自编码器的对于时空特征与预测值之间的分布关系进行了自适应学习, 使模型具有较好的泛化能力, 并提出了一种时空数据融合的处理方式. 该模型也适用于处理医疗、气象、金融等领域的时空依赖性相关任务, 并具自适应过滤异常数据能力. 同时, 本文模型在训练速度方面较DCRNN, ESG等基准模型表现更快, 这为处理大规模数据和实时应用提供了潜在优势. 但模型复杂度的提高导致收敛速度下降, 笔者在未来将进一步探索学习率的设置、参数优化算法等, 使该模型的收敛速度提高, 从而节省计算资源并增强模型的泛化性.

参考文献:

- [1] KUMAR S V. Traffic flow prediction using Kalman filtering technique. *Procedia Engineering*, 2017, 187: 582 – 587.
- [2] YANG Gaofei, XU Rui, QIN Ming, et al. Short-term traffic forecasting based on ARMA and Kalman filter. *Journal of Zhengzhou University (Engineering Edition)*, 2017, 38(2): 36 – 40. (杨高飞, 徐睿, 秦鸣, 等. 基于ARMA和卡尔曼滤波的短时交通预测. 郑州大学学报: 工学版, 2017, 38(2): 36 – 40.)
- [3] REN P, WANG S, CHEN X, et al. Fusion estimation strategy based on dual adaptive Kalman filtering algorithm for the state of charge and state of health of hybrid electric vehicle Li-ion batteries. *International Journal of Energy Research*, 2022, 46(6): 7374 – 7388.
- [4] XU W, WANG S, JIANG C, et al. A novel adaptive dual extended Kalman filtering algorithm for the Li-ion battery state of charge

- and state of health co-estimation. *International Journal of Energy Research*, 2021, 45(10): 14592 – 14602.
- [5] DAVIS G A, NIHAN N L. Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering*, 1991, 117(2): 178 – 188.
- [6] ZHANG Y, LIU Y. Traffic forecasting using least squares support vector machines. *Transportmetrica*, 2009, 5(3): 193 – 213.
- [7] CETINER B G, SARI M, BORAT O. A neural network based traffic-flow prediction model. *Mathematical and Computational Applications*, 2010, 15(2): 269 – 278.
- [8] BRAZ F J, FERREIRA J, GONCALVES F, et al. Road traffic forecast based on meteorological information through deep learning methods. *Sensors*, 2022, 22(12): 4485.
- [9] WANG Y, WU H, ZHANG J, et al. Predmn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(2): 2208 – 2225.
- [10] MA X, DAI Z, HE Z, et al. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 2017, 17(4): 818.
- [11] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *ArXiv Preprint*, 2017, arXiv: 1709.04875.
- [12] TAN M C, WONG S C, XU M C, et al. An aggregation approach to short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2009, 10(1): 60 – 69.
- [13] YANG H, LI X, QIANG W, et al. A network traffic forecasting method based on SA optimized ARIMA-BP neural network. *Computer Networks*, 2021, 193(3): 108102.
- [14] HE Wenwu, PEI Boyu, MAO Guojun, et al. Road network traffic flow prediction based on lightweight spatio-temporal graph convolution model. *Journal of Railway Science and Engineering*, 2022, 19(9): 2552 – 2562.
(贺文武, 裴博琰, 毛国君, 等. 基于轻量时空图卷积模型的路网交通流预测. *铁道科学与工程学报*, 2022, 19(9): 2552 – 2562.)
- [15] WANG T, NI S, QIN T, et al. TransGAT: A dynamic graph attention residual networks for traffic flow forecasting. *Sustainable Computing: Informatics and Systems*, 2022, 36: 100779.
- [16] ZHAO L, SONG Y, ZHANG C, et al. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21(9): 3848 – 3858.
- [17] DUAN Y, CHEN N, SHEN S, et al. FDSA-STG: Fully dynamic self-attention spatio-temporal graph networks for intelligent traffic flow prediction. *IEEE Transactions on Vehicular Technology*, 2022, 71(9): 9250 – 9260.
- [18] KIPFT N, WELLING M. Semi-supervised classification with graph convolutional networks. *ArXiv Preprint*, 2016, arXiv: 1609.02907.
- [19] KINGMA D P, WELLING M. Auto-encoding variational bayes. *ArXiv Preprint*, 2013, arXiv: 1312.6114.
- [20] YIN M, ZHOU M. Semi-implicit variational inference. *International Conference on Machine Learning*. Jinan, China: PMLR, 2018: 5660 – 5669.
- [21] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*. Montreal, Canada: MIT Press, 2014, 27(2): 3104 – 3112.
- [22] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *ArXiv Preprint*, 2017, arXiv: 1707.01926.
- [23] BAI L, YAO L, LI C, et al. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in Neural Information Processing Systems*. Vancouver, BC, Canada: Curran Associates, 2020: 17804 – 17815.
- [24] MAKRIDAKIS S, HIBON M. ARMA models and the Box-Jenkins methodology. *Journal of Forecasting*, 1997, 16(3): 147 – 163.
- [25] ZHENG C, FAN X, WANG C, et al. Gman: A graph multi-attention network for traffic prediction. *AAAI Conference on Artificial Intelligence*, 2020, 34(1): 1234 – 1241.
- [26] WILLIAMS B M, DURVASULA P K, BROWN D E. Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record*, 1998, 1644(1): 132 – 141.
- [27] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling. *ArXiv Preprint*, 2019, arXiv: 1906.00121.
- [28] WU H, XU J, WANG J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 2021, 34: 22419 – 22430.
- [29] YE J, LIU Z, DU B, et al. Learning the evolutionary and multi-scale graph structure for multivariate time series forecasting. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Washington, DC, USA: Association for Computing Machinery, 2022: 2296 – 2306.
- [30] DENG J, CHEN X, JIANG R, et al. ST-Norm: Spatial and temporal normalization for multi-variate time series forecasting. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. New York, USA: Association for Computing Machinery, 2021: 269 – 278.

作者简介:

欧阳毅 副教授, 硕士生导师, 目前研究方向为模式识别、智能控制、图像处理、人体姿态识别等, E-mail: ooy@mail.zjgsu.edu.cn;

汤文燕 硕士研究生, 目前研究方向为智慧物流与智能决策, E-mail: twy15298547169@163.com;

邵泳博 硕士研究生, 目前研究方向为智慧物流与智能决策, E-mail: 964286719@qq.com;

黎晏伶 硕士研究生, 目前研究方向为智慧物流与智能决策, E-mail: 1457588359@qq.com.