

基于动态自适应时空图的多元时序预测模型

乔少杰¹⁾ 薛骥¹⁾ 杨国平¹⁾ 韩楠²⁾ 李贺³⁾

袁冠⁴⁾ 黄江涛⁵⁾ 毛睿⁶⁾

¹⁾(成都信息工程大学软件工程学院 成都 610225)

²⁾(成都信息工程大学管理学院 成都 610225)

³⁾(西安电子科技大学计算机科学与技术学院 西安 710071)

⁴⁾(中国矿业大学计算机科学与技术学院 江苏 徐州 221116)

⁵⁾(南宁师范大学广西人机交互与智能决策重点实验室 南宁 530100)

⁶⁾(深圳大学计算机与软件学院 广东 深圳 518060)

摘要 深度学习模型在多元时间序列预测、智能驾驶、图像识别等多个领域广泛应用,其中多元时间序列预测是学者们关注的重点之一,多元时间序列预测是典型的回归任务,旨在通过海量的历史数据构建模型以预测未来状态,被广泛运用于交通、电力、金融等领域.多元时间序列数据具有复杂的时空依赖性,现有模型大多仅能捕获序列数据中的时间特征,难以捕获空间特征,而图神经网络解决了这一问题.图神经网络能够自然地建模实体间的复杂关系,可以很好地处理拓扑数据,而多元时序数据大多可以构造为拓扑图,因此图神经网络可以很好地学习多元时序数据中的空间特征.基于图神经网络的多元时间序列预测模型受到广泛关注并取得了一定的成果,但现有基于图神经网络的模型仍存在诸多不足.首先,现有方法大多分别捕获和建模多元时间序列数据中的空间特性和时间特性,未充分考虑多元时间序列的时空统一性,导致模型的次优建模;其次,现有方法主要基于静态预定义图或动态自适应图,其中静态预定义图通常根据监测节点之间的空间相关性进行构造且不会随着时间而改变,基于预定义图的研究忽略了时间序列数据中的时间特征,即忽略了数据模式随时间发生的改变;而自适应图通常由模型自主学习并不包含监测节点间的固有属性,基于自适应图的研究忽略了大量有效的领域知识,如道路的连通性和道路间的属性.为了解决上述问题,提出基于动态自适应时空图的多元时序预测模型MTP-Graph(Multivariate Time series Prediction model based on dynamic adaptive spatio-temporal Graph),利用时空融合模块将时空信息进行统一处理,避免了分开捕获时间特性与空间特性而导致的次优建模问题,提出图结合模块将静态预定义图和动态自适应图进行动态融合,获取时空信息的同时充分考虑领域知识,使模型可以更好地学习多元时间序列中的时空特性.在PeMSD3、PeMSD7和PeMSD8数据集上的大量实验结果表明,MTP-Graph预测性能优于其他基准方法,验证了MTP-Graph的可用性和有效性.

关键词 多元时序预测;时空数据库;图神经网络;注意力机制;机器学习

中图分类号 TP311

DOI号 10.11897/SP.J.1016.2024.02925

收稿日期:2024-03-12;在线发布日期:2024-09-11. 本课题得到国家自然科学基金(62272066)、四川省科技计划资助(2023YFG0027, 2024YFFK0413)、网络空间安全教育部重点实验室及河南省网络空间态势感知重点实验室开放基金课题(KLCS20240106)、成都市区域科技创新合作项目(2023-YF11-00020-HZ)、成都市技术创新研发项目(2024-YF05-01217-SN)、CCF-蚂蚁科研基金项目(CCF-AFSG RF20240106)资助. 乔少杰,博士,教授,中国计算机学会(CCF)杰出会员,主要研究领域为时空数据库、人工智能数据库. E-mail: sjqiao@cuit.edu.cn. 薛骥,博士研究生,主要研究领域为时空数据库. 杨国平,硕士生,主要研究领域为数据库. 韩楠(通信作者)博士,副教授,主要研究领域为数据库、数据挖掘. E-mail: hannan@cuit.edu.cn. 李贺,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为时空数据库、数据挖掘. 袁冠,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为数据库. 黄江涛,博士,副研究员,中国计算机学会(CCF)高级会员,主要研究领域为数据库. 毛睿,博士,教授,中国计算机学会(CCF)杰出会员,主要研究领域为数据库.

A Multivariate Time Series Forecasting Model Based on Dynamic Adaptive Spatio-Temporal Graphs

QIAO Shao-Jie¹ XUE Qi¹ YANG Guo-Ping¹ HAN Nan² LI He³
YUAN Guan⁴ HUANG Jiang-Tao⁵ MAO Rui⁶

¹(School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225)

²(School of Management, Chengdu University of Information Technology, Chengdu 610225)

³(School of Computer Science and Technology, XiDian University, Xi'an 710071)

⁴(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116)

⁵(Guangxi Key Lab of Human-machine Interaction and Intelligent Decision, Nanning Normal University, Nanning 530100)

⁶(College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060)

Abstract Deep learning models have been extensively used in many fields such as multivariate time series prediction, intelligent driving, image recognition, and so on. Among them, multivariate time series forecasting is one of the focuses from scholars. Multivariate time series forecasting is a typical regression task, which aims to construct a model to predict the future states through a huge number of historical data, and it is widely used in the fields of transportation, electricity and finance. Multivariate time series data have complex spatio-temporal dependencies, the existing models can only capture temporal features, but it is difficult to capture spatial features. The graph neural network emerges to solve this problem. Graph neural network can naturally model the complex relationship between entities and is very good at dealing with topological data, and most of the multivariate temporal data can be constructed as a topological graph, so the graph neural network can better learn the spatial features from multivariate temporal data. Multivariate time series prediction models based on graph neural networks have received widespread attention and achieved certain success. However, there are some drawbacks in the existing methods based on graph neural networks. Firstly, the existing methods only capture and model the spatial features or temporal features from multivariate time series data separately, without considering the spatio-temporal unity of multivariate time series data, leading to the suboptimal modeling problem. Secondly, existing methods are mainly based on static predefined graphs or dynamic adaptive graphs, the static predefined graphs are usually constructed based on the spatial correlations among the monitored nodes and do not change over time, and the research based on static predefined graph often ignore temporal features of time series data, which ignores the changes in data patterns over time. Whereas, adaptive graphs are usually learnt autonomously by the model and do not contain intrinsic properties among the monitored nodes, so the research based on adaptive graphs ignore a large number of effective domain knowledge, such as road connectivity and properties between roads. In order to solve the above problems, a Multivariate Time series Prediction model based on dynamic adaptive spatio-temporal Graph called MTP-Graph is proposed, which unifies the spatio-temporal information by using spatio-temporal fusion module, avoiding sub-optimal modelling problem caused by capturing temporal and spatial features separately, and the proposed graph combination module dynamically integrates combines the predefined graphs with adaptive graphs, and takes into full consideration domain knowledge while obtaining spatio-temporal information from data, in order to better learn the spatio-temporal features from multivariate time series data. Extensive experiments are conducted on the PeMSD3, PeMSD7 and PeMSD8 datasets, and the results show that MTP-Graph outperforms other benchmark methods in prediction performance, which shows the usability and effectiveness of MTP-Graph.

Keywords multivariate time series prediction; spatio-temporal database; graph neural networks; attention mechanism; machine learning

1 引言

多元时序预测被广泛运用于现实生产和生活场景中,如电力预测^[1]、交通预测^[2]。随着计算机和模型架构的不断发展,深度学习技术相比传统统计方法有着更好的预测性能和应用前景。其中,基于时空图的多变量时间序列预测是近年来出现的研究热点,它将每个收集信息的传感器视作一个节点,其节点特征与内含信息随着时间的变化不断变化,且节点之间存在信息传递性等节点相关性。时空图神经网络(Spatial Temporal Graph Neural Networks, STGNNs)假设拓扑结构中节点的未状态取决于其邻居及其自己的历史状态,即节点具有时空依赖性。因此,如何捕获时间和空间的内在依赖性成为影响预测性能的关键问题。而本文主

要将研究重点聚焦在多元时序预测任务中的交通预测领域。

交通预测^[3]是多元时序预测中的重要研究领域,其目标是预测城市交通网络的交通状况,包括交通流量、速度和通行时间等,对交通状况进行准确的预测,有助于构建智能交通系统(Intelligent Transport System, ITS)。其中交通数据具有两种很强的时空特性,第一种是时间周期性,第二种是空间异质性。其周期性表现为交通流量在一段时间内具有重复的模式,如图1(a)所示;空间异质性表现为不同监视器观测到的流量区间可能不同,如图1(b)所示,监视器A的流量区间为40~450,而监视器B的流量区间为0~160。此外,时间周期性与空间异质性是相互关联的,例如工业区的周期性与居民区的周期性是不同的,将时间特征与空间特征进行统一是提高时空预测准确性的有效途径。

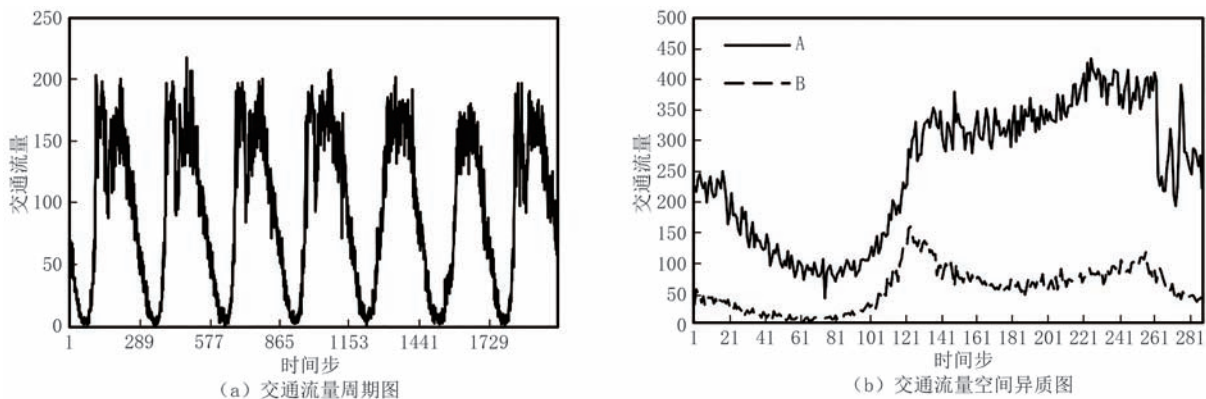


图1 交通流量时空特性图

按照消息传递方式的不同可以将用于多元时序预测的STGNNs分为三类,第一类是基于卷积神经网络(Convolutional Neural Network, CNN)^[4-5];第二类是基于循环神经网络(Recurrent Neural Network, RNN)^[6];第三类是基于注意力机制^[7-9]。基于CNN的STGNNs通过堆叠时间卷积层和空间卷积层来捕获局部时空特征。基于RNN的STGNNs可以有效地将拓扑结构的时空特征融合到时间特征中。基于注意力机制的STGNNs结合空间注意力与时间注意力进行建模,利用注意力机制对信息进行聚合。本文提出的MTP-Graph基于RNN方法进行消息的传递。

然而现有方法主要存在两点不足之处:

(1) STGNNs通常将时间和空间依赖分别建模,因为邻接图在设计时仅用于处理节点特征,即捕获空间依赖性。而时间依赖性通常利用不同的消息传递方式进行捕获。例如,在交通领域,不同路口在不同时间或相同时间具有不同的流量特征,而流量特征与时间和空间是密不可分的,现有方法却割裂了时间与空间的统一性。这种分开捕获时间依赖和空间依赖的建模方式破坏了多元时间序列的时空相关性,导致了次优建模。虽然STSGCN^[10]提出将时间图与空间图进行拼接以学习时空特征,STFGNN^[11]在STSGCN的基础上提出一种并行处理操作,可以通过学习更多的时空相关性,但这两种方法的拼接方式不佳,产生了大量的冗余信息,进而

影响了模型的性能。

(2) 现有的时空图神经网络使用两种邻接图构建方式,第一种是基于领域知识的静态预定义图,如 STGCN (Spatial-temporal Graph Convolutional Networks)^[12], ASTGCN (Attention based Spatial-Temporal Graph Convolutional Networks)^[8];第二种是利用自适应函数构建的动态自适应图,如 AGCRN (Adaptive Graph Convolutional Recurrent Network)^[13]通过自适应矩阵来捕获多变量序列之间的关系。这两种邻接图构建方式均存在不足:预定义图由于缺乏动态性,随着时间推移可能导致节点信息的丢失;而自适应图由模型自主生成,会包含很多冗余边,例如交通路网中节点 A 和 B 不连通,但在自适应图中连通。这些冗余边中包含的无用信息会损害模型性能。如何将预定义图和自适应图相结合成为提升预测性能的关键。

为了解决上述两个问题,本文提出了 MTP-Graph 模型,主要贡献包括:

(1) 提出一种新的基于时空图的多元时序预测模型 MTP-Graph,通过生成自适应时间嵌入和自适应空间嵌入同时学习序列的时空特征,使用时空融合模块 STFMM (Spatio-temporal Fusion Model) 对时空特征进行融合,统一了时空特性,弥补了单一因素考虑带来的不足,时空邻接矩阵中的冗余信息少于 STFGNN 与 STSGCN;

(2) 提出一种将静态预定义图和动态自适应图结合的模块 PCM (Pattern Combination Module),通过线性注意力机制将生成的动态自适应图与静态预定义图动态结合,既能从数据集中自适应地学习时空特征,也可以从预定义图中学习领域知识;

(3) MTP-Graph 利用线性注意力机制对静态图与动态图之间的信息进行融合,相较于普通的注意力机制,应用线性注意力机制可以使图结合模块的时间复杂度从 $O(n^2)$ 下降到 $O(n)$;

(4) 在三个真实数据集 PeMSD3、PeMSD7 和 PeMSD8 上进行了大量实验,实验结果表明本文提出的 MTP-Graph 模型具有较好的预测性能,且预测性能优于其他模型。

2 预备知识及相关工作

本节从注意力机制、图神经网络以及多元时序预测任务介绍与本文相关的代表性工作。

2.1 注意力机制

注意力机制已成功应用于各种任务,如计算机视觉^[14]、自然语言处理^[15]、交通流量预测^[16]。注意力机制通过嵌入查询向量 Q 、关键向量 K 和值向量 V 来计算输入序列中元素之间的注意力得分。Transformer^[15]是一种经典的注意力机制实现方式,其使用编码器-解码器架构,通过自注意力模块和前馈神经网络构建。注意力机制与 RNN 和 CNN 模型完全不同,它有效捕捉了序列的长距离依赖性,注意力分数可用公式(1)表示:

$$S = \text{softmax}(Q \cdot K / d^{1/2}) \cdot V \quad (1)$$

其中, d 表示向量嵌入维度。

然而,注意力机制在提升模型感受域的同时极大地增加了计算复杂度。近期研究开始尝试使用简单的线性层来实现注意力机制,这些方法有效降低了计算复杂度,同时阐明了线性层和注意力机制之间的联系。文献[17]使用多层感知器 (Multilayer Perceptron, MLP) 架构来实现类似注意力机制的效果。此外,文献[18]使用简单的线性层和归一化层来构建注意力机制,将注意力的计算复杂度降低到了线性水平。

2.2 图神经网络

图神经网络主要基于图卷积神经网络 (Graph Convolutional Network, GCN)^[19]以及图注意力机制^[20]进行消息的传递与汇聚,其中图卷积网络依赖于图拉普拉斯矩阵的分解实现,主要分为空域方法^[21]及谱域方法^[22]。

标准的图神经网络假设输入节点特征是静态的,仅考虑空间内的信息流,并不能普遍适用于多变量时间序列任务。而 STGNNs 解决了这一问题,在考虑空间信息流的同时也考虑了时间信息流,即考虑随时间而动态变化的节点特征。时空图神经网络按照其消息传递方法可以分为基于 RNN、基于 CNN 和基于注意力机制的 STGNNs。

如图 2(a) 所示,基于 RNN 的 STGNNs^[23]在当前时间 $T+2$ 节点 N_2 接受邻居 N_1 的当前步信息以及其前一步 $T+1$ 时间的信息。基于 RNN 的方法通常使用 GCN 捕获节点的空间特征,然后使用 RNN 捕获对应的时间特征,具体形式如公式(2):

$$Y = \text{RNN}(\text{GCN}(X, A; \Theta)) \quad (2)$$

其中, Y 表示模型输出, X 为输入序列, A 表示邻接矩阵, Θ 表示模型中所有训练参数。

如图 2(b) 所示,基于 CNN 的 STGNNs^[2, 12]在当前时间 $T+2$ 节点 N_2 接受邻居 N_1 的当前步和一定时

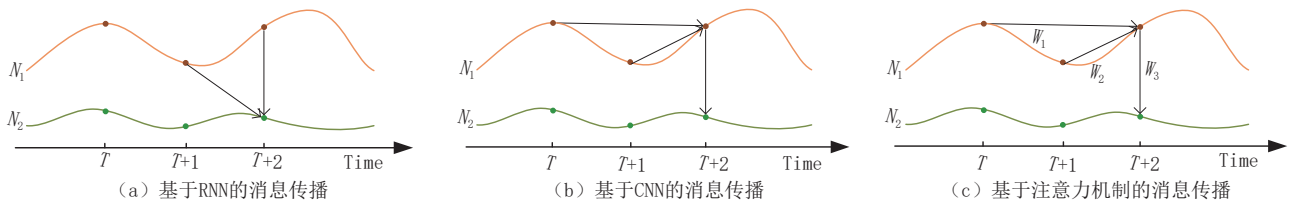


图2 STGNNs的三种消息传播方式

间窗口内的信息. 基于CNN的方法与RNN的不同之处在于其首先使用时间卷积网络(Temporal Convolutional Network, TCN)来捕获一段时间内的时间特征,之后再使用GCN来捕获节点的空间特征,具体形式如公式(3)所示:

$$Y = \text{GCN}(\text{TCN}(\|_{t=1}^T X_t, A; \Theta); \Psi) \quad (3)$$

其中, X_t 表示当前 t 时刻输入序列, $\|$ 表示向量拼接操作.

如图2(c)所示,基于注意力机制的STGNNs^[8-9]在当前时间 $T+2$ 节点 N_2 接受邻居 N_1 的当前步和之前一段时间窗口内信息的加权和, W_1, W_2, W_3 分别表示在 $T, T+1, T+2$ 时刻的信息的权重.

基于注意力机制的方法与前述方法完全不同,它们利用两种不同的注意力结构来捕获节点的时空特征,具体形式如公式(4)所示:

$$Y = \text{SA}(\|_{t=1}^T \text{TA}(X_t, A; \Theta); \Psi) \quad (4)$$

其中, SA表示时间注意力层, TA表示空间注意力层.

基于CNN的图神经网络架构虽然通过并行计算加快训练速度,但受到核大小的限制难以捕捉长期的序列相关性;基于注意力机制架构虽然可以捕捉到长期的序列相关性,但极大地增加了运算成本;而基于RNN的图神经网络架构在捕捉长期序列相关性的同时通过遗忘门机制减少了计算复杂度,因此本文提出的MTP-Graph采用基于RNN的实现方法.

2.3 多元时序预测任务

多元时序预测任务广泛运用于各个领域,其中交通流量预测是多元时序预测任务中一个基础且重要的应用,引起了广泛的关注. 早期的工作主要围绕着统计分析方法,例如自向量回归(Vector Autoregressive, VAR)^[24],自回归-滑动平均混合模型(Autoregressive Integrated Moving Average, ARIMA)^[25]. 然而,由于交通数据变化的复杂性导致多元时间序列之间的关系通常是非线性的,基于统计分析的方法性能不佳. 支持向量回归算法(Support Vector Regression, SVR)^[26]等机器学习方

法的提出,极大地缓解了这一困境. 但是,传统机器学习方法在成功建模复杂时空依赖关系的同时也极度依赖高质量的人工标注,这导致了模型构建的人工花费过高,难以在真实世界中应用.

随着计算机技术的发展以及算力的增加,数据驱动的深度学习方法可以有效地提取交通数据的特征. 其中文献[27]利用CNN来预测一定区域内的交通流量,这类方法将空间划分为单元,并在结构化的二维或三维空间中进行卷积运算. 在此基础上,文献[28]提出基于残差卷积单元的模型来预测人群流量. 文献[29]利用时空残差CNN捕获交通数据的空间和时间依赖性. 文献[30]利用结合了条件随机场的GCN捕获全局时空依赖性.

然而,上述方法无法捕获道路中的拓扑结构,即空间特性. STGNNs的提出使得交通数据可以直接以图的形式表达并进行学习. Zhao等人^[31]结合了GCN和门控循环单元(Gate Recurrent Unit, GRU)来捕捉输入序列的时空特征.

基于注意力机制,文献[32]利用多关系同步注意力机制捕获复合空间关系. Kong等人^[7]借助Gumbel-Softmax技术构造离散图,利用图注意力机制进行流量预测. 文献[9]提出自适应时空变换器对复杂时空因素进行模拟. 文献[14]利用Node2Vec得到静态空间矩阵,然后将空间矩阵与时间矩阵进行拼接,并通过时间注意力模块与空间注意力模块来分别捕获时间特征与空间特征,但该方法难以捕捉路网的动态变化特征. Li等人^[33]提出了一种神经网络结构搜索算法,利用元学习的方式生成空间图卷积网络中的邻接矩阵和时间卷积网络中的卷积核,并将边元知识分别映射为每个空间图卷积网络的邻接矩阵,同时用另一组全连接网络将点元知识分别映射为每个节点的时间卷积核,从而建模属性图对时空相关性的影响,虽然AutoST降低了模型设计过程所需要的人力成本,但在预测性能上表现一般.

与基于预定义图的方法相比,文献[2]设计了空间自注意模块来捕捉动态的空间依赖关系,并通

过两个图掩码矩阵来捕获不同粒度的空间依赖性. 文献[4]提出使用图学习模块从数据中提取邻接图, 然后使用混合跳跃传播层和膨胀起始层提取交通流的时空特征. 文献[13]使用自适应参数矩阵学习各个节点上序列的特定模式.

与上述方法的不同之处在于, 本文的工作将多元时序数据的时空信息统一, 并且融合了静态预定义图和动态自适应图, 既能从数据中学习潜在模式, 也能结合已有领域知识进行动态预测.

3 问题定义

定义1. 属性图 G . 属性图 $G=(U, M, A)$, 其中 U 表示属性图中节点的集合, M 表示边的集合, $A \in R^{N \times N}$ 表示邻接矩阵, N 表示 G 中节点的数量.

定义2. 多元时序预测任务. 多元时序预测任务的目标是通过对观察到的历史数据序列 $X_{1:T}^i \in R^{T \times d}$, $X_{1:T}^i = \{X_1^i, X_2^i, X_3^i, \dots, X_i^i, \dots, X_T^i\} \in R^{T \times N \times d}$ 建模得到预测模型 F , 利用 F 预测未来一段时间内的数据 $X_{T+1:T+L}^i$, 其中 i 表示第 i 个节点, N 表示节点数量, 包含 N 个节点的任务称为 N 元时序预测, T 表示观测时间长度, d 表示数据嵌入维度. 预测任务形式化如

公式(5)所示:

$$X_{T+1:T+L}^i = F(X_{1:T}^i, G, \theta) \quad (5)$$

其中, L 表示预测时间长度, θ 表示模型 F 中所有参数.

定义3. 图卷积网络 GCN. GCN 可以将 G 中的空间信息进行汇聚, 其形式化如公式(6)所示:

$$Z_{(t)} = X_t \otimes g_\alpha = U \times \sum_{k=0}^{K-1} \gamma_k \times U \times \xi_k \times U^T \times X_t \quad (6)$$

其中, $Z_{(t)}$ 表示 t 时刻卷积输出, X_t 表示输入序列, g_α 表示多项式卷积核, \otimes 表示卷积操作, U 表示特征向量, K 表示卷积核数量, γ 表示可学习参数, ξ 表示特征值对角矩阵.

4 多元时序预测模型

MTP-Graph 采用端到端的学习框架, 如图3所示, 由三个模块组成, 其中时空融合模块将多元时间序列中时间信息与空间信息进行融合构建动态时空自适应图, 以确保时空信息的统一性; 图结合模块将动态自适应图和静态预定义图相结合, 在数据驱动的模型基础上增加了领域知识; 最后利用时空循环网络并结合图中的时空信息对动态时空自适应图进行更新.

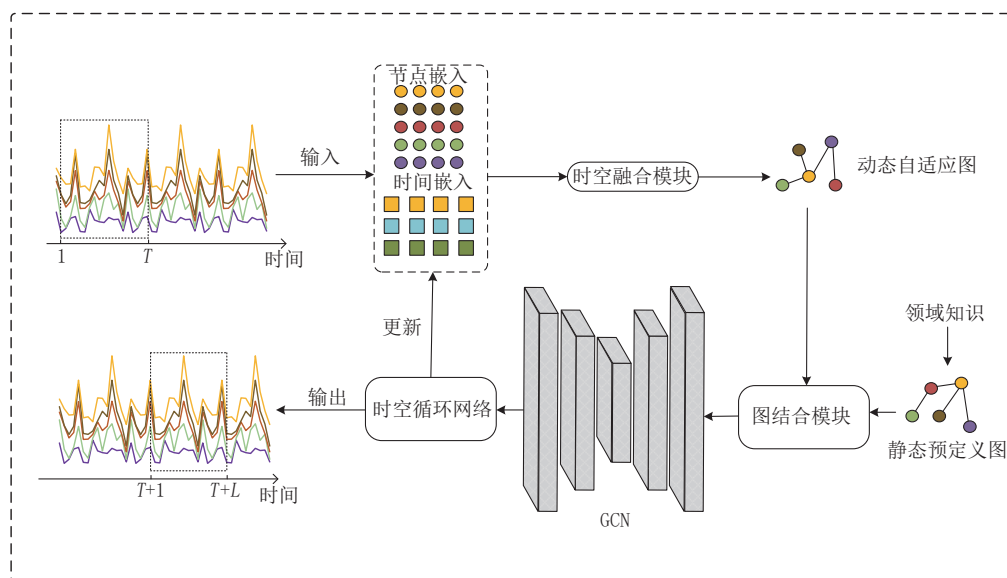


图3 MTP-Graph内部结构图

4.1 时空融合模块

邻接矩阵是图神经网络进行多元时序预测的基础. 邻接矩阵的生成函数共有3种:(1)距离函数, 基于节点之间的地理或空间距离构建邻接矩阵;

(2)相似性函数, 通过计算节点之间属性的相似性来构建邻接矩阵;(3)自适应函数, 通过可学习的节点嵌入构建邻接矩阵, 可以从数据中动态学习隐藏的空间特征.

上述三种方法仅关注了邻接矩阵在空间上的信息,忽略了时间信息.对于多元时序预测,时间信息是至关重要的,本文提出时空融合模块STFM,在自适应函数的基础上,增加了对时间信息的嵌入.

时空融合模块由公式(7)和(8)构成,其输入为两个自适应矩阵,其中 $E_{node} \in R^{N \times d}$ 表示节点嵌入参数矩阵, $E_{time} \in R^{T \times d}$ 表示时间嵌入参数矩阵, d 表示嵌入维度, E_{node} 和 E_{time} 使用相同的嵌入维度以便后续融合操作.时空融合模块在每个时间步将 E_{node} 和 E_{time} 进行融合,时空融合模块利用矩阵乘法或加法进行了融合操作,通过实验发现利用矩阵加法的性能更优,具体融合操作如公式(7)和(8)所示:

$$E = E_{node} + E_{time} \quad (7)$$

$$A^0 = E \times E^T \quad (8)$$

其中, E 表示时空自适应矩阵, l 表示第 l 个时间步, A^0 是时空邻接矩阵.例如在数据集PeMSD3中包含358个节点,进行12步预测,即生成形状为 358×10 维的 E_{node} 和形状为 12×10 维的 E_{time} ,并且MTP-Graph的图结构是动态的,与输入时间步无关.

4.2 图结合模块

多元时间序列中的每个序列中包含多种信息,例如交通流量预测任务中,每个节点的信息为交通流量,而交通流量受到位置、时间、附近路况等多种复杂因素的影响.这些复杂因素可以分为3类:时间依赖、空间依赖、外部因素.前两类因素可以通过STFM模块中生成的时空邻接矩阵进行捕捉,而外部因素只能由领域知识进行定义,例如道路的车道数,附近是否有居民区.具有相同外部因素的节点即使没有道路相连但仍然可能拥有相似的模式.因此本文提出了图结合模块PAM将数据驱动的时空自适应图与领域知识预定义图进行结合,使得模型可以同时捕获到这三类因素,进而提高时序预测性能.PAM具体架构由静态预定义矩阵提取模块以及动态结合注意力网络构成.

由于空间相邻信息很容易从数据集或地图中获取,因此本文选用它们作为领域知识来提取静态预定义矩阵 A^1 .例如在PeMSD3数据集中,将其节点连通信息作为静态预定义矩阵.

与文献[17]中工作原理类似,使用切比雪夫多项式对公式(6)中的 $U \times \xi_k \times U^T$ 进行截断操作从而实现图卷积,得到公式(9):

$$X_t \otimes g_a = \sum_{k=0}^{K-1} \gamma_k \times T_k(L) \times X_t \quad (9)$$

$$T_k(L) = 2L \times T_{k-1}(L) - T_{k-2}(L) \quad (10)$$

其中, L 表示规范化拉普拉斯矩阵, $T_k(L)$ 表示切比雪夫多项式, $T_0(L) = I$, $T_1(L) = L$, I 表示单位矩阵.为了进一步简化运算,令 $K=2$, L 为对称归一化的拉普拉斯矩阵,即 $L = I - D^{-1/2} A D^{-1/2}$,其中 D 表示节点度矩阵,可得到公式(11):

$$Z = \gamma_0 \times X - \gamma_1 \times D^{-1/2} \times A^0 \times D^{-1/2} \quad (11)$$

令 $\beta = \gamma_0 = -\gamma_1$,并推广到高维数据可得到公式(12):

$$Z = (I + D^{-1/2} \times A \times D^{-1/2}) \times X \times \beta + b \quad (12)$$

其中, β 和 b 均为可训练参数.

A^0 和 A^1 的切比雪夫多项式展开形式如公式(13)和公式(14)所示:

$$Z^0 = (I + D^{-1/2} \times A^0 \times D^{-1/2}) \times X \times \beta_0 + b_0 \quad (13)$$

$$Z^1 = (I + D^{-1/2} \times A^1 \times D^{-1/2}) \times X \times \beta_1 + b_1 \quad (14)$$

本文使用注意力机制计算节点权重,将预定义图和时空自适应图进行结合,形式如公式(15):

$$J(X_t) = f_b((f_a(Z^0; \theta_a) + Z^1); \theta_b) \quad (15)$$

其中, $f_a(\cdot)$ 和 $f_b(\cdot)$ 表示注意力网络, θ_a 、 θ_b 分别为 $f_a(\cdot)$ 和 $f_b(\cdot)$ 网络中参数,其中 $f_a(\cdot)$ 对 A^0 中的节点进行权重分配,减少 A^0 中冗余边的权重;而 $f_b(\cdot)$ 对结合图再次进行权重分配,以获得结合图中节点间的相似模式.

在文献[17]的基础上,本文利用线性层构建了注意力网络 $f_a(\cdot)$ 和 $f_b(\cdot)$ 进行动态组合,如公式(16)至公式(18)所示:

$$\Gamma_1 = \text{pool}(A) \quad (16)$$

$$\Gamma_2 = \sigma(\text{Linear}_1(\Gamma_1)) \quad (17)$$

$$\Gamma_3 = S = \sigma(\text{Linear}_2(\Gamma_2)) \quad (18)$$

其中, pool 为池化层, σ 为激活函数, Γ_2 为注意力亲和系数, Γ_3 为最终注意力分数, Linear_1 和 Linear_2 均为线性层.

4.3 时空循环网络

基于文献[13],本文使用GCN替换GRU中的MLP层构建了时空图卷积模块来学习多元时序数据的时空特征.然后,通过叠加多个GRU来学习流量序列的时空特征,最终输出预测结果 y_t .替换后的单个GRU实现如公式(19)-(22)所示:

$$z_t = \sigma(J([X_t, h_{t-1}]) \times E_A \times W_z + E_A \times b_z) \quad (19)$$

$$r_t = \sigma(J([X_t, h_{t-1}]) \times E_A \times W_r + E_A \times b_r) \quad (20)$$

$$h'_t = \tanh(J([X_t, r_t \odot h_{t-1}]) \times E_A \times W_h + E_A \times b_h) \quad (21)$$

$$h_t = r_t \odot h_{t-1} + (1 - z_t) \odot h'_t \quad (22)$$

其中, $[\cdot]$ 表示拼接操作, z 表示重置门, r 表示更新门, h_t 和 X_t 分别表示 t 时刻的输出和输入, h'_t 表示 t

时刻的隐藏状态, $W_\varepsilon, b_\varepsilon, W_r, b_r, W_h, b_h$ 均表示可学习参数, \odot 表示哈达玛积.

模型训练的损失函数为平滑 L1 损失 (Smooth L1 Loss), 具体形式如公式(23):

$$Smooth\ L1(x, y) = \begin{cases} |x_i - y_i| - 0.5, & \text{否则} \\ \frac{1}{2(x_i - y_i)}, & \text{if } |x_i - y_i| < 1 \end{cases} \quad (23)$$

其中, x_i 表示第 i 个时间步的真实值, y_i 为第 i 个时间步的预测值.

本文提出的预测模型如算法 1 所示.

算法 1. MTP-Graph 预测模型

输入: 多元时间序列数据集 X , 观测时间步 T , 预定义图 A^1

输出: 训练好的 MTP-Graph 模型

1. 根据 X 、 T 构建训练集 D_i ;
2. 根据 T 生成点嵌入参数矩阵 E_{node} 和时间嵌入参数矩阵 E_{time} ;
3. 初始化 MTP-Graph 参数空间;
4. DO
5. 从训练集 D_i 中选择批处理样本 D_p ;
6. 根据公式(7)-(8)计算得到 A^0 ;
7. 根据公式(19)至公式(22)得到输出 y_i ;
8. 最小化目标函数(公式(23))获得最优参数空间 Θ ;
9. WHILE ($epoch \neq 0$);

算法 1 描述了 MTP-Graph 模型的训练过程, 首先从输入的时序数据中构建训练数据集 (第 1 行), 然后生成可训练参数矩阵 E_{node} 和 E_{time} 并根据配置文件初始化参数空间 (第 2-3 行), 再根据训练条件对模型进行训练 (第 4 至 7 行), 最后通过反向传播和 AdamW 优化器^[34]对模型进行迭代优化直到训练终止 (第 8-9 行), 通过上述训练方式可以获取并学习到历史时序数据中的统一时空特征, 从而获得更加准确的预测结果.

时间复杂度分析: 步骤 1 中构建训练集的时间复杂度为 $O(L_x - T)$, 其中 L_x 表示输入数据集 X 的样本数量; 步骤 2 中参数矩阵生成的时间复杂度为 $O(1)$; 步骤 3 中初始化参数空间的时间复杂度为 $O(1)$; 步骤 5 中的选择批处理样本的时间复杂度为 $O(1)$; 步骤 6 中生成 A^0 的时间复杂度为 $O(H * T)$, 其中 H 表示模型中 GRU 单元的数量; 步骤 7 的时间复杂度为 $O(|M|)$, 其中 $|M|$ 表示邻接图 A 的边数量; 步骤 9 最小化目标函数的时间复杂度为 $O(1)$. 模型共训练 P 次, 即循环执行 P 次步骤 5 到 8, 因此算法 1 的时间复杂度为 $O(L_x - T + P * (H * T * |M| + 1))$. 算法 1 的时间复杂性是随着节点数量线性变化的, 因此具

有线性时间复杂性, 相比于其他算法, 具有较好的性能优势, 将在 5.4 节实验中进行验证.

5 实验与分析

5.1 数据描述及预处理

实验基于 PeMSD3、PeMSD7 以及 PeMSD8 数据集 (<https://github.com/wanhuaayu/ASTGCN#datasets>), 上述数据集均采自美国加州地区, 其中 PeMSD3 数据集收集了 358 个监视器在 2018 年 9 月至 11 月的交通流量数据; PeMSD7 数据集收集了 883 个监视器在 2017 年 3 月到 8 月的交通流量数据; PeMSD8 数据集收集了 170 个监视器在 2016 年 7 月到 8 月的交通流量数据.

对于数据中存在的缺失值, 本文使用线性插值法来进行填充. 本文使用 12 个历史时间步作为输入, 并将未来 12 个时间步的输出作为预测结果, 即使用一个小时的历史数据来预测未来一个小时的数据. 此外, 本文通过标准归一化对输入数据进行预处理, 使得数据平稳化.

5.2 实验设置

MTP-Graph 基于 PyTorch 框架实现, 按 6:2:2 的比例将数据集划分为训练集、验证集和测试集.

对于所有数据集, MTP-Graph 将其节点连通信息作为静态预定义矩阵 A^1 , 实验中相同部分的网络参数设置一致, 如表 1 所示: 初始学习率为 0.0003, epoch 为 100, 在验证集中若连续 15 个 epoch 没有提升则终止训练, 嵌入维度为 10, GRU 单元个数为 64, 批大小为 64. SVR 中多项式核函数的次数为 3, 容忍停止标准为 $1e-3$, 松弛变量 ϵ 为 0.1; LSTM 中编码器和解码器均包含两个循环层, 每个循环层包含 256 个 LSTM 单元; STGCN 中空间卷积核大小为 64, 时间卷积核大小为 16; ASTGCN 中图卷积核大小均为 64, 切比雪夫多项式阶数为 2; DCRNN 中最大随机游走步数为 3; STFGNN 中快速 DTW 算法的最大搜索长度为 12, 空间图的稀疏度为

表 1 实验参数设置

| 参数 | 值 |
|------------|--------|
| 初始学习率 | 0.0003 |
| epoch | 100 |
| 提前终止 epoch | 15 |
| 嵌入维度 | 10 |
| GRU 单元个数 | 64 |
| 批大小 | 64 |

0.01. STSGCN 中时空同步图卷积层个数为4,滤波器大小序列为[64, 64, 64]; AGCRN 中切比雪夫多项式阶数为2; MTGNN 中混合跳跃传播层层数为3,保留比率为0.05.

对所有模型,采用 MAPE、MAE、RMSE 三个指标进行性能评估,具体如公式(26)所示:

$$MAE = \frac{1}{P} \sum_{i=1}^P |x_i - y_i| \quad (24)$$

$$MAPE = \frac{100\%}{P} \sum_{i=1}^P \left| \frac{x_i - y_i}{x_i} \right| \quad (25)$$

$$RMSE = \sqrt{\frac{1}{P} \sum_{i=1}^P (x_i - y_i)^2} \quad (26)$$

其中, P 为采样总数, x_i 为第 i 个采样的真实值, y_i 表示第 i 个采样的预测值.

5.3 基准方法

本节对基准对比算法进行介绍:(1) LSTM (Long Short-Term Memory Network)^[35],一种特殊的 RNN 模型,可以对序列信息选择性地存储和遗忘;(2) STGCN^[12],通过两种类型的 GCN 捕获交通

流的时空相关性;(3) ASTGCN^[8],使用叠加的多重注意力机制来分别捕捉不同时期交通流的空间和时间特征;(4) STFGNN^[11],通过融合时间图和空间图有效地学习交通流的时空特征;(5) STSGCN^[10],将时间特征分为三类,并对每一类时间特征使用卷积运算和注意力机制来捕捉交通流的时空相关性;(6) DCRNN^[6],使用扩散卷积运算来捕捉交通流的时空相关性,并利用随机游走来进行节点信息的扩散;(7) AGCRN^[12],使用自适应参数来捕捉节点模式,利用图卷积操作捕捉空间相关性;(8) MTGNN^[36],利用混合跳跃传播层和扩展初始层来捕获时间序列中的时空依赖关系;(9) STG-NRDE^[37],使用对数变换将时序样本转换为相对较短的特征向量序列,并通过两个神经粗糙微分方程来捕获空间特征与时间特征.

5.4 实验结果

表2给出本文提出的 MTP-Graph 模型和9种基准方法在 PeMSD3、PeMSD7 和 PeMSD8 数据集上的12个时间步的平均预测性能.

表2 MTP-Graph 和基准方法在 PeMSD3、PeMSD7 和 PeMSD8 数据集上的预测误差

| 方法 | PeMSD3数据集 | | | PeMSD7数据集 | | | PeMSD8数据集 | | |
|-----------|--------------|--------------|--------------|--------------|--------------|-------------|--------------|--------------|-------------|
| | MAE | RMSE | MAPE(%) | MAE | RMSE | MAPE(%) | MAE | RMSE | MAPE(%) |
| LSTM | 21.33 | 35.11 | 23.33 | 29.98 | 45.84 | 13.20 | 22.20 | 34.06 | 14.20 |
| STGCN | 17.42 | 28.78 | 16.85 | 25.33 | 39.34 | 11.21 | 18.14 | 27.92 | 11.13 |
| ASTGCN | 17.69 | 29.66 | 19.40 | 28.05 | 42.57 | 13.92 | 18.61 | 28.16 | 13.08 |
| STFGNN | 16.77 | 28.34 | 16.30 | 22.07 | 35.80 | 9.21 | 16.64 | 26.22 | 10.60 |
| STSGCN | 17.48 | 29.21 | 16.78 | 24.26 | 39.03 | 10.21 | 17.13 | 26.80 | 10.96 |
| DCRNN | 17.66 | 28.95 | 17.08 | 24.70 | 38.12 | 17.12 | 18.36 | 28.36 | 11.55 |
| AGCRN | 16.17 | 28.51 | 15.05 | 19.85 | 32.69 | 12.91 | 16.66 | 26.04 | 10.46 |
| MTGNN | 16.46 | 28.56 | 16.46 | 20.82 | 34.09 | 9.03 | 15.71 | 24.62 | 10.03 |
| STG-NRDE | 15.50 | 27.06 | 14.90 | 20.45 | 33.73 | 8.65 | 15.32 | 24.72 | 8.90 |
| MTP-Graph | 15.03 | 26.29 | 14.87 | 19.78 | 33.18 | 8.37 | 14.73 | 24.36 | 9.51 |

可以发现:MTP-GRAPH除了在 PeMSD8 的 MAPE 指标略逊于 STG-NRDE 外,其余指标均为最优.其中 LSTM 性能最差,因为 LSTM 仅考虑了交通流量的时间特征,并未捕获空间特征.基于图神经网络的模型预测性能更好,因为拓扑图结构可以有效地捕获交通流的空间特征.其中 DCRNN 虽然通过扩散卷积操作来捕获交通流的特征,但不能捕获交通流数据的时空相关性,因此预测性能较差; STGCN 提出的时空卷积模块可以对数据并行处理,保留了非线性建模能力,从而提高了预测性能; STSGCN 对每个时间片构造局部时空图,并通过裁剪操作消除图中冗余信息,进一步提高预测的准确

性; STFGNN 将多个空间图和时间图进行拼接,可以有效地学习隐藏的时空关联,但由于拼接操作的劣势,邻接矩阵的非对角线元素包含冗余信息,损害了模型的性能; MTGNN 通过混合跳跃传播层捕捉长序列的相关特征,但 MTGNN 在时间特征的提取上存在不足; AGCRN 对自适应图中的每个节点使用单独的权重以便学习每个节点特有的交通模式,提高了预测性能; STG-NRDE 利用神经粗糙微分方程来分别捕获空间特征与时间特征,进一步提高了预测性能.

图4展示了 DCRNN、STGNN、AGCRN 和 MTP-Graph 模型在 PeMSD8 数据集上 60 分钟内

12个时间步的预测误差,时间步1表明模型预测5分钟后的交通流量状况,时间步12表明模型预测60分钟后的交通流量状况.MTP-Graph在任意时间步的预测误差均小于其他方法,且在长期预测的

效果上更佳,实验结果证明了MTP-Graph在短期预测和长期预测均更加准确.主要原因在于MTP-Graph统一了时空信息,并考虑了领域知识,使预测的结果更加精准.

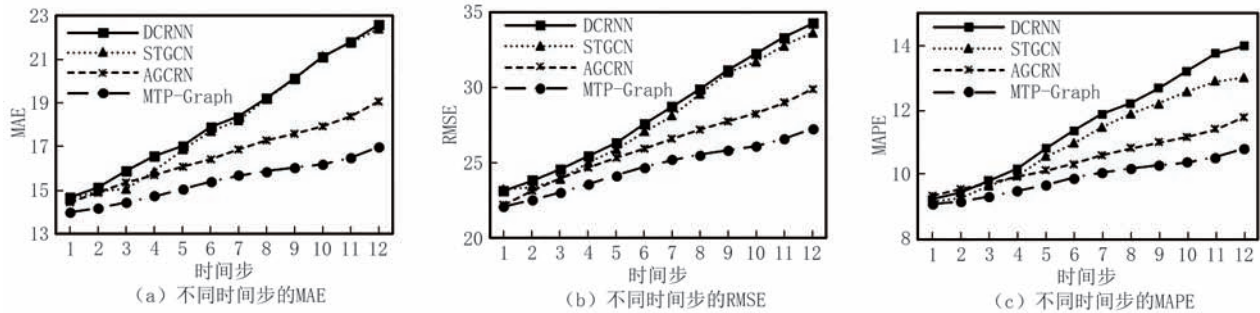


图4 不同方法在PeMSD8数据集上不同时间步的预测误差对比

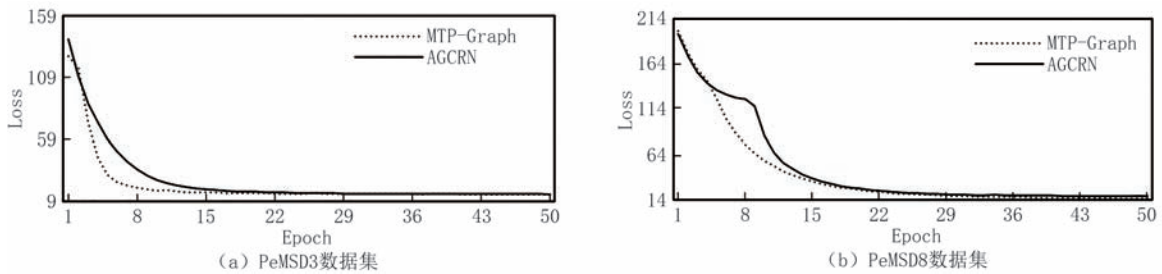


图5 MTP-Graph和AGCRN在不同数据集的收敛速度

通过图5可以观察到,MTP-Graph相较于AGCRN收敛速度更快,值得注意的是,AGCRN也使用自适应函数构造邻接矩阵,验证了本文提出的时空融合模块和图结合模块有利于模型的训练和收敛速度.其主要原因在于时间特征与空间特征的融合有利于加快模型的收敛速度,并减少训练的迭代次数.

表3给出了MTP-Graph在PeMSD3、PeMSD7和PeMSD8数据集上的平均训练时间及预测时间,可以发现,随着数据集中监测节点的增加,模型的训练时间与预测时间也随之增加,例如:PeMSD7数据集的监测节点数量为883个,远多于PeMSD8数据集监测节点数量170个,因此PeMSD7数据集上的训练时间与预测时间明显高于PeMSD8数据集.此外,可以发现MTP-Graph具有较好的鲁棒性:随着数据集节点的增加,训练时间并没有爆炸性增长,与

监测节点数量线性相关,与算法1的时间复杂性分析结果一致.此外,随着节点数量增加,算法预测时间变化不大,进而证明MTP-Graph算法的预测时间性能优势.

5.5 消融实验

以PeMSD8数据集为例,对MTP-Graph进行消融实验,共设计了两个变体:(1)不包含时空融合模块的MTP-Graph-V1;(2)不包含图结合模块的MTP-Graph-V2.表4展示了MTP-Graph及其变体的在12个时间步的平均评估指标,可以看出MTP-Graph性能最佳,MTP-Graph-V2其次说明了本文提出的融合两个模块算法的有效性,并且时空统一模块的作用更大,这也符合预期,即对时空特性统一建模更能提升算法预测的准确性,避免了次优建模问题.

图6展示了在不同时间步MTP-Graph及其变体的性能,可以观察到在任意时间步MTP-Graph

表3 MTP-Graph在不同数据集上的训练时间及预测时间

| 数据集 | 训练时间/s·epoch ⁻¹ | 预测时间/s |
|--------|----------------------------|--------|
| PeMSD3 | 187.8 | 15.98 |
| PeMSD7 | 228.6 | 17.81 |
| PeMSD8 | 69.83 | 15.32 |

表4 消融实验

| 模型 | MAPE/% | MAE | RMSE |
|--------------|-------------|--------------|--------------|
| MTP-Graph | 9.91 | 15.41 | 24.75 |
| MTP-Graph-V1 | 10.34 | 16.03 | 25.40 |
| MTP-Graph-V2 | 10.15 | 15.75 | 25.10 |

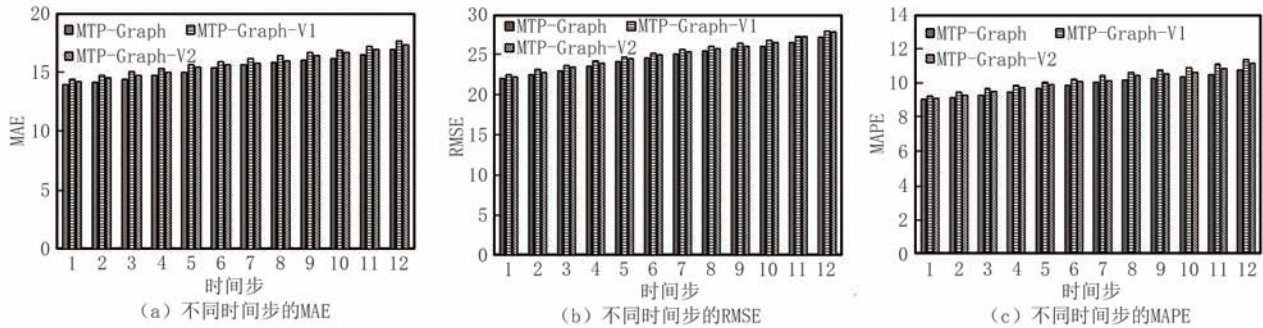


图6 MTP-Graph及其变体在PeMSD8数据集上不同时间步的预测性能

均为最优, MTP-Graph-V2次之, 进一步验证了对时空特性统一建模的重要性. 上述消融实验在PeMSD3数据集上的结果类似.

此外, 为了验证线性注意力机制对MTP-Graph性能的影响, 表5展示了MTP-Graph与MTP-Graph-A在PeMSD8数据集上的平均评估指标结果, 可以看到MTP-Graph的预测性能最佳, MTP-Graph优于MTP-Graph-A模型的原因在于: 采用线性注意力机制可以更为准确地为结合图中节点分配权重.

表6展示了MTP-Graph与MTP-Graph-A在PeMSD8数据集上每个epoch的平均训练时间, 可以观察到: MTP-Graph的平均训练时间相较于

MTP-Graph-A缩短了32.9%. 主要原因在于: 相比于普通的注意力机制时间复杂度 $O(n^2)$, 线性注意力机制的时间复杂度可以降低到 $O(n)$.

表6 使用不同注意力机制的训练时间

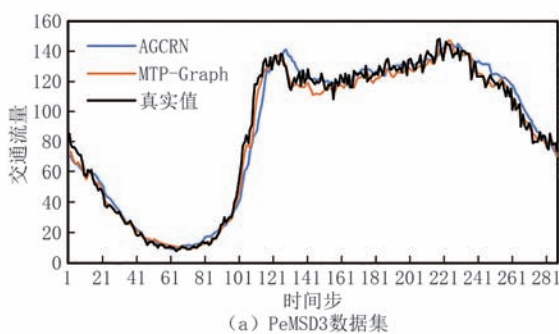
| 模型 | 训练时间/s·epoch ⁻¹ |
|-------------|----------------------------|
| MTP-Graph | 69.83 |
| MTP-Graph-A | 104.06 |

5.6 结果可视化

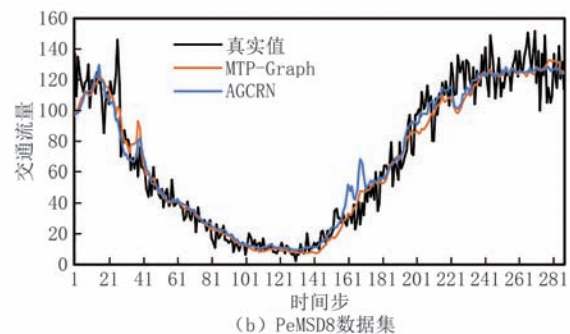
通过图7可以观察到, 在PeMSD3和PeMSD8数据集随机节点的24小时可视化图中MTP-Graph的预测结果与实际结果都更加吻合, 表明MTP-Graph可以更好地学习多元时间序列的模式并应用于真实场景. 主要原因在于: MTP-Graph同时学习时空序列的时间信息与空间信息, 避免了次优建模, 在数据驱动基础上充分利用领域知识.

表5 注意力机制对比实验

| 模型 | MAPE/% | MAE | RMSE |
|-------------|-------------|--------------|--------------|
| MTP-Graph | 9.91 | 15.41 | 24.75 |
| MTP-Graph-A | 10.09 | 15.77 | 25.06 |



(a) PeMSD3数据集



(b) PeMSD8数据集

图7 MTP-Graph在PeMSD3和PeMSD8数据集上的可视化结果

6 结论与展望

本文提出一个基于融合时空图的多元时序预测模型MTP-Graph, 将时空信息进行统一建模, 将静态预定义图和动态自适应图进行融合, 可以更好地学习多元时间序列中的时空特性. 在PeMSD3、

PeMSD7和PeMSD8数据集上进行大量实验, 结果表明, MTP-Graph相较于主流基准模型具有更优的预测性能, 验证了该模型可以同时从动态自适应图和静态预定义图中学习到多元时序中潜在的时空依赖关系. 未来的工作对自适应图及预定义图进行结构优化, 减少训练参数以节省训练时间并适应更复杂的多元时序预测.

参 考 文 献

- [1] Alsharekh M F, Habib S, Dewi D A, et al. Improving the efficiency of multistep short-term electricity load forecasting via R-CNN with ML-LSTM. *Sensors*, 2022, 22(18): 6913
- [2] Jiang J, Han C, Zhao W X, et al. PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(4): 4365-4373
- [3] Yin D, Jiang R, Deng J, et al. MTMGNN: Multi-time multi-graph neural network for metro passenger flow prediction. *GeoInformatica*, 2023, 27(1): 77-105
- [4] Wang Y, Jing C. Spatiotemporal graph convolutional Network for multi-scale traffic forecasting. *ISPRS International Journal of Geo-Information*, 2022, 11(2): 1025
- [5] Zhang Q, Chang J, Meng G, et al. Spatio-temporal graph structure learning for traffic forecasting //Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020, 34(1): 1177-1185
- [6] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting//Proceedings of the International Conference on Learning Representations. Ithaca, USA, 2018: 1-16
- [7] Kong X, Zhang J, Wei X, et al. Adaptive spatial-temporal graph attention networks for traffic flow forecasting. *Applied Intelligence*, 2022, 52(4): 4300-4316
- [8] Yu Li, Li Hui-Yuan, Jiao Chen-Lu, et al. Trajectory prediction in complex scenes based on multi-head attention adversarial mechanism. *Chinese Journal of Computers*, 2022, 45(6): 1133-1146(in Chinese)
(余力, 李慧媛, 焦晨璐等. 基于多头注意力对抗机制的复杂场景行人轨迹预测. *计算机学报*, 2022, 45(6): 1133-1146)
- [9] Huang B, Dou H, Luo Y, et al. Adaptive spatiotemporal transformer graph network for traffic flow forecasting by IoT loop detectors. *IEEE Internet of Things Journal*, 2023, 10(2): 1642-1653
- [10] Song C, Lin Y, Guo S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting //Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020, 34(01): 914-921
- [11] Li M, Zhu Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting //Proceedings of the AAAI Conference on Artificial Intelligence. Virtual, 2021, 35(5): 4189-4196
- [12] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting //Proceedings of the International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 3634-3640
- [13] Bai L, Yao L, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in Neural Information Processing Systems*, 2020, 33: 17804-17815
- [14] Huang Ying, He Zi-Fen, Yang Hong-Kuan, et al. Multi-scale segmentation of episodic video instance through polarized self-attention manipulation. *Chinese Journal of Computers*, 2022, 45(12): 2605-2618(in Chinese)
(黄莹, 何自芬, 杨宏宽等. 极化自注意力调控的情景式视频实例多尺度分割. *计算机学报*, 2022, 45(12): 2605-2618)
- [15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017: 5998-6008
- [16] Zheng C, Fan X, Wang C, et al. GMAN: A graph multi-attention network for traffic prediction//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(01): 1234-1241
- [17] Tolstikhin I O, Housley N, Kolesnikov A, et al. MLP-mixer: an all-MLP architecture for vision //Proceedings of the Annual Conference on Neural Information Processing Systems. Virtual, 2021, 34: 24261-24272
- [18] Guo M H, Liu Z N, Mu T J, et al. Beyond self-attention: External attention using two linear layers for visual tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(5): 5436-5447
- [19] Kipf T, Welling M. Semi-supervised classification with graph convolutional networks//Proceedings of the International Conference on Learning Representations, Washington, USA, 2017: 1-14
- [20] Mo X, Huang Z, Xing Y, et al. Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(7): 9554-9567
- [21] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 2017, 30: 1024-1034
- [22] Parashar A, Rhu M, Mukkara A, et al. SCNN: An accelerator for compressed-sparse convolutional neural networks. *ACM SIGARCH Computer Architecture News*, 2017, 45(2): 27-40
- [23] Seo Y, Defferrard M, Vandergheynst P, et al. Structured sequence modeling with graph convolutional recurrent networks //Proceedings of the Neural Information Processing. Siem Reap, Cambodia, 2018: 362-373
- [24] Li Y, Wood J, Ji L, et al. Fitting multilevel vector autoregressive models in Stan, JAGS, and Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 2022, 29(3): 452-475
- [25] Williams B M, Hoel L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, 2003, 129(6): 664-672
- [26] Lin G, Lin A, Gu D. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Information Sciences*, 2022, 608: 517-531
- [27] 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
- [28] Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction //Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco,

- USA, 2017, 31(1): 1655-1661
- [29] Bao Y X, Shi Q, Shen Q Q, et al. Spatial-temporal 3D residual correlation network for urban traffic status prediction. *Symmetry*, 2022, 14(1): 33
- [30] Diao C, Zhang D, Liang W, et al. A novel spatial-temporal multi-scale alignment graph neural network security model for vehicles prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 24(1): 904-914
- [31] 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
- [32] Huang J, Luo K, Cao L, et al. Learning multispect traffic couplings by multirelational graph attention networks for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(11): 20681-20695
- [33] Li J, Zhang S, Xiong H, et al. Autost: Towards the universal modeling of spatio-temporal sequences. *Advances in Neural Information Processing Systems*, 2022, 35: 20498-20510
- [34] Loshchilov I, Hutter F. Decoupled weight decay regularization // *Proceedings of the International Conference on Learning Representations*. Louisiana, USA, 2019: 1-19
- [35] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks // *Proceedings of the Annual Conference on Neural Information Processing Systems*. Montreal, Canada, 2014: 3102-3112
- [36] Wu Z, Pan S, Long G, et al. Connecting the dots: Multivariate time series forecasting with graph neural networks // *Proceedings of the International Conference on Knowledge Discovery & Data Mining*. New York, USA, 2020: 753-763
- [37] Choi J, Park N. Graph neural rough differential equations for traffic forecasting. *ACM Transactions on Intelligent Systems and Technology*, 2023, 14(4): 1-27



QIAO Shao-Jie, Ph. D., professor. His research interests include spatio-temporal databases, artificial intelligence for databases.

XUE Qi, Ph. D., candidate. His current research interest is spatio-temporal databases.

YANG Guo-Ping, M. S. His research interest is databases.

HAN Nan, Ph. D., associate professor. Her current interests include databases and data mining.

LI He, Ph. D., associate professor. His research interests include databases and data mining.

YUAN Guan, Ph. D., professor, engineer. His current research interest is databases.

HUANG Jiang-Tao, Ph. D., associate researcher. His research interest is databases.

MAO Rui, Ph. D., professor. His research interest is databases.

Background

Multivariate time series prediction is widely used in fields, e. g., transportation, meteorology and electricity, and is a critical task in many applications. The researches focus on designing accurate multivariate time series prediction model. As we know, traditional statistical methods are difficult to model the nonlinear features of multivariate time series. Deep learning models have been proposed in recent years for multivariate time series forecasting with higher prediction accuracy. However, uniformly modeling spatial and temporal information of time series data is difficult and challenging.

In this paper, we propose the MTP-Graph model, a multivariate spatio-temporal prediction model based on dynamic adaptive spatio-temporal graphs. The MTP-Graph can uniformly model spatio-temporal features, preserve the integrity of spatio-temporal data, and integrate dynamic adaptive graphs with static adaptive graphs and take into consideration domain knowledge. Extensive experiments have been conducted on two real datasets by comparing MTP-Graph with 10 popular models, and the

results show that MTP-Graph outperforms other algorithms.

This paper is supported by the NSFC project named "Research on Context Awareness Driven Multiple-motion-pattern Trajectory Prediction Models". This project focuses on the research of multiple-motion-pattern trajectory prediction. Aiming to analyze the mobile databases, we apply the big data processing, context-aware computing and trajectory data mining techniques. The difficulties and essential problems in this project contain: (1) using big data processing techniques to retrieve complex contexts, reduce the dimensionalities of trajectory big data and propose a new hybrid spatio-temporal indexing structure in order to effectively index the historical and future locations of moving objects; (2) proposing a trajectory points of interesting mining algorithm based on Hausdorff distance by integrating the context information; and (3) exploring the fitness of context and spatio-temporal data and formalize complex context information, and employing the state-of-the-art prediction models, for example, the Gaussian mixture model and Kalman filtering approach, to accurately predict continuous multiple-motion-pattern trajectories.