计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202440113
61(8): 2020−2029, 2024

基于图提示微调的交通流量预测

赖培源[1,2] 李 程[2,3] 王增辉[2] 王昌栋[3] 廖德章[2]
[1](广东工业大学信息工程学院 广州 510006)
[2](广东省华南技术转移中心有限公司 广州 511458)
[3](中山大学计算机学院 广州 510006)
(peiyuanlai@hotmail.com)

Traffic Flow Prediction Based on Graph Prompt-Finetuning

Lai Peiyuan[1,2], Li Cheng[2,3], Wang Zenghui[2], Wang Changdong[3], and Liao Dezhang[2]
[1](*School of Information Engineering*, *Guangdong University of Technology*, *Guangzhou* 510006)
[2](*Guangdong Provincial South China Technology Commercialization Center Co.*, *Ltd.*, *Guangzhou* 511458)
[3](*School of Computer Science and Engineering*, *Sun Yat-sen University*, *Guangzhou* 510006)

Abstract Traffic flow prediction is a crucial foundational aspect in the development of smart cities, holding significant implications for urban traffic management and user travel planning. The expanding dimensions of both time and space contribute to the increasing volume of traffic flow data, with real-time updates presenting a challenge in deploying cost-effective forecasting models to meet diverse traffic prediction demands. Inspired by the success of the emerging paradigm of graph-based finetuning in downstream tasks of natural language processing, we introduce, for the first time to our best knowledge, graph based finetuning to enhance the generalization capabilities of traditional traffic flow prediction models, called TPGPF (traffic flow prediction based on graph prompt-finetuning), which could enhance the generality and effectiveness of self-supervised learning. In the context of spatiotemporal multidimensional traffic flow prediction models, our approach involves pretraining the model based on historical datasets and introducing learnable prompt vectors. With the pretrained model solidified, a self-supervised learning model is guided by the introduced prompt vectors to adapt to new data prediction tasks, thereby enhancing the generality and effectiveness of traffic flow prediction models. Extensive experiments on five real-world public datasets validate the effectiveness of our work, which provides an effective method to overcome the cost challenge brought by quasi real-time training of traffic flow data.

Key words graph prompt; traffic flow prediction; finetuning; pre-trained model; self-supervised learning

摘 要 交通流量预测是建设智慧城市重要的基础功能,对城市的交通管理和用户出行规划具有重要意义.由于时间维度和空间维度的扩展,交通流量的数据具有规模大、增长快速、实时更新等特征,传统的训练模型通常需要将大量的历史数据进行训练预测,导致较长的计算时间和较高的算力成本,因此,如何使用低计算成本的预测模型来满足广泛的流量预测需求是重要的技术挑战.近年来兴起的提示微调范式在自然语言处理的下游任务推广中取得了较好的效果,受其启发,提出利用少量的实时数据来微调优化大规模历史数据预训练的模型,为交通流量模型预测的优化应用提出了一种新的思路.通过引入图提示微

收稿日期: 2024-03-15; **修回日期**: 2024-04-24
基金项目: 国家自然科学基金项目(62276277); 广东省技术转移智能匹配工程技术研究中心项目(2022A175)
This work was supported by the National Natural Science Foundation of China (62276277) and the Guangdong Provincial Engineering Research Center of Intelligent Matching for Technology Commercialization (2022A175).
通信作者: 王昌栋(changdongwang@hotmail.com)

调的交通流量预测（traffic flow prediction based on graph prompt-finetuning, TPGPF）模型的泛化能力，在时空多维度下的交通流量图预测模型中，基于历史数据集进行预测模型的预训练，并引入可学习的提示向量，在预训练模型固化的情况下指导预训练的自监督学习模型，以适应新的数据预测任务，提升交通流量预测模型的通用性和有效性. 通过在 5 个公开数据集上进行了大量的实验，证明了 TPGPF 的有效性.

关键词 图提示；交通流量预测；微调；预训练模型；自监督学习

中图法分类号 TP181

图提示学习是通过对输入的微调，引导预训练模型更高效地适应新任务，被认为是当前通用人工智能（artificial general intelligence, AGI）的研究前沿[1-2]，基于提示学习的微调范式在自然语言处理的下游任务推广中取得了较好的效果，在图学习领域的提示学习范式正成为新的研究热点[3]. 随着城市的快速普及和人口的快速增长，智能交通成为当前的研究热点. 交通流量预测是建设智慧城市重要的基础功能，对城市的交通管理[4]、智慧交通[5]以及个人出行规划[6]具有重要的意义. 目前的研究主要聚焦于对交通流量的数据挖掘，从时间[7]、空间数据[8]的角度，或者它们的综合使用，在两者的主次关系[9]、互监督[10]，以及时空相关性[11]等角度开展了相关研究，而图卷积神经网络和注意力机制成为了近来的主要训练方法[12-14]. 然而，交通流量预测模型仍面临着诸多挑战，首先，由于时间维度和空间维度的扩展，交通流量的数据非常大，如何使用低计算成本的预测模型来满足广泛的流量预测需求，是成本角度的重要技术挑战. 其次，由于交通流量数据实时更新，且最新的数据对预测效果有重要的影响，如何利用最新的实时数据来优化预训练模型，是工程化角度的重要挑战. 最后，由于时空数据存在不确定性的相互影响，如何协同不同维度数据的预测权重，是模型研究角度的重要挑战.

提示学习源自自然语言处理，是近年来新兴起的一种方法. “预训练，提示”的范式在将预训练语言模型推广到各种下游语言任务中已经取得显著的成效[15]. 有一些研究也开始转向图神经网络的提示学习研究，利用历史数据进行预训练，然后微调以适应下游任务，从而节省标记数据的成本[16-17]. 受其启发，本文提出将较大规模的历史交通流量数据进行模型的预训练，在此基础上将最新的交通流量数据作为下游任务进行微调预测，不仅能够解决实时交通数据更新与大规模数据训练昂贵成本带来的挑战，也能让预训练模型和下游任务紧密结合. 本文提出 TPGPF（traffic flow prediction based on graph prompt-finetuning）模型，通过引入图提示学习来提升传统交通流量预测模型的泛化能力，也可应用于其他预测任务，如智慧农业中的植物、动物养殖，天气预测等.

本文的主要贡献包括 3 个方面：

1）提出了一种新的交通流量预测方法 TPGPF，据我们所知，这项工作尝试将“预训练-提示-微调”范式引入交通流量预测模型，提高了自监督学习模型的通用性和有效性；

2）利用一种通用的图提示特征（graph prompt feature, GPF）调整方法，使得预训练的自监督学习模型能够更高效地适应下游交通流量预测任务，这为克服交通流量预测准实时训练带来的成本挑战提供了一种有效的方法；

3）在 5 个公开的交通流量预测数据集上进行了广泛的实验，结果表明 TPGPF 优于最先进的基线.

1 相关工作

1.1 交通流量预测相关研究

随着城市不断扩张和道路车辆数量增加，很多学者开展了面向交通流量预测的相关研究工作. 早期的研究主要基于对历史数据的统计预测，例如，Castro-Neto 等人[18]提出了一种称为在线支持向量机回归（OL-SVR）的有监督统计学习技术，用于预测典型和非典型条件下的高速公路短期交通流量. 作为最直接的角度，基于时间维度的预测模型设计是交通流量预测的研究重点. 例如，Kumar 等人[19]提出了一种季节性 Box-Jenkins 差分自回归移动平均模型（SARIMA），用于克服其他模型对健全数据库的依赖问题，通过有限输入数据对短期的交通流量进行预测. 当观察到相邻区域之间的交通流量数据具有较强的相关性后，后续的研究则将注意力转向了空间数据，关注跨地域之间交通流量的相互影响. 在空间数据的研究方面，很多学者使用图神经网络方法来构建预测模型. 例如 Zhang 等人[20]提出了一种基于时空序列的流量预测算法，称之为 ST-ResNet，该模型

将时空数据和深度学习结合起来，利用时空残差网络来预测城市各个区域的人群流入和流出数量. 在时空数据结合中，有相关研究进一步引入深度学习技术，例如 Yu 等人[21] 提出了一种新的深度学习框架，用图网络来解决交通领域的时序预测问题，称为时空图卷积网络(ST-GCN). 该框架旨在解决传统模型经常忽略空间和时间依赖性带来的挑战，特别是在中长期预测任务中较好地捕获了交通流量数据中的时空相关性. 随着注意力机制在图神经网络中表现出良好的性能，相关研究也被引入到交通流量预测中. 例如，Zheng 等人[22] 引入了一种名为 GMAN 的图多注意网络，该网络采用具有多个时空注意力模块的编码器-解码器架构来模拟时空因素对交通状况的影响，取得了良好的性能.

此外，围绕交通流量数据的时空依赖关系也有相关的研究. 例如，Wang 等人[6] 提出了 ST-GCN 的分层交通流量预测方法. 该方法利用相邻路段之间的时空依赖关系，实现更准确的交通流量预测. 近年来，人们的关注点已经从主要解决交通流量预测中的时空依赖性更广泛地转向时空异质性. 例如，Bai 等人[23] 认为，学习节点的模式对于交通预测至关重要，并采用了学习特定节点的模式，而不是使用预定义的图来捕捉，从而避免了使用预定义图. 他们引入了一种自适应图卷积递归网络(AGCRN)，旨在自动捕获交通序列中细粒度的空间和时间相关性.

此外，Song 等人[24] 提出了一种名为时空同步图卷积网络(STS-GCN)的新模型，用于预测时空网络数据. 该模型的开发是为了克服复杂的时空相关性和异质性带来的挑战. Ji 等人[9] 提出了一种时空自监督学习(ST-SSL)框架用于交通流量的预测，该框架对交通模式的图表征进行了增广，并通过辅助的自监督学习方法来捕获空间和时间异质性，取得了明显的性能提升. 然而，ST-SSL 重点解决了预训练模型中对时空数据的分析预测优化，并未考虑实时交通数据与历史交通流量大数据的协同，历史数据的预训练模型在实时预测应用中仍存在时间滞后性.

综上所述，与交通流预测相关的模型已经获得了广泛的关注，并产生了丰富的研究成果. 然而，这些模型中的大多数都没有解决对实时数据变化的适应以及由重复训练带来的计算成本问题，如何使用少量的实时交通流量数据来优化预训练模型，是本文的主要研究内容.

1.2 图提示技术相关研究

最近的研究已经证明了图“预训练，提示”范式在适应下游任务方面的有效性[25]. 例如，Liu 等人[26] 提出了图提示框架，将预训练和下游任务整合到一个公共的任务模板中，使用可学习的提示向量来缩小下游任务和预训练模型之间的差距，指导下游任务以特定任务的方式从预训练模型中获取最相关的知识. Chen 等人[27] 提出了一种统一的图混合预训练框架，称之为多任务图双提示(ULTRA-DP)，通过提示机制将任务和位置识别注入到图神经网络(graph neural network, GNN)中，从而提高预训练性能. 由于微调需要对原本模型的参数进行调整，由此会带来较大的额外开销，而使用提示学习则可以在不改变原本神经网络参数的情况下，让已有的预训练模型对下游任务进行高效适配，因此提示学习受到了越来越多的关注. Sun 等人[15] 提出了一种新的迁移学习范式，即图预训练和提示调优(GPPT)，它引入了图提示功能，将独立节点转换为令牌对，通过任务令牌和结构令牌的组合将下游节点分类转化为连接预测问题，并通过实验证明了 GPPT 的有效性. 通常，预训练模型在较大的数据集上进行训练，以保持内在固有的图属性，然后在较小的数据集上进行微调，以调整模型并最大限度地减少下游任务损失. 例如，Chen 等人[28] 提出了一种新颖的框架 GPL-GNN，用于图神经网络的图提示学习，旨在缩小不同任务之间的差距. GPL-GNN 利用无监督方法，在不需要标记数据的情况下，将无监督的预训练结构表示作为提示信息纳入下游任务. 值得说明的是，“预训练，微调”是让预训练模型“迁就”各种下游任务，而“预训练，提示”则是让各种下游任务来“迁就”预训练模型. 尤其是针对交通预测的图提示相关研究成果非常少.

本文提出的基于图提示微调的交通流量预测方法是基于历史数据集进行预测模型的预训练，该方法引入可学习的提示向量，在预训练模型固化的情况下，指导预训练的自监督学习模型适应新的数据预测任务，旨在提升交通流量预测模型的通用性和有效性.

2 TPGPF 模型

在本节中，我们主要介绍提出的基于图提示微调的交通流量预测模型 TPGPF 和具体算法.

2.1 模型框架

如图 1 所示，TPGPF 模型主要包括 4 个组件：图构建模块、图模型预训练模块、图提示微调模块、预测模块.

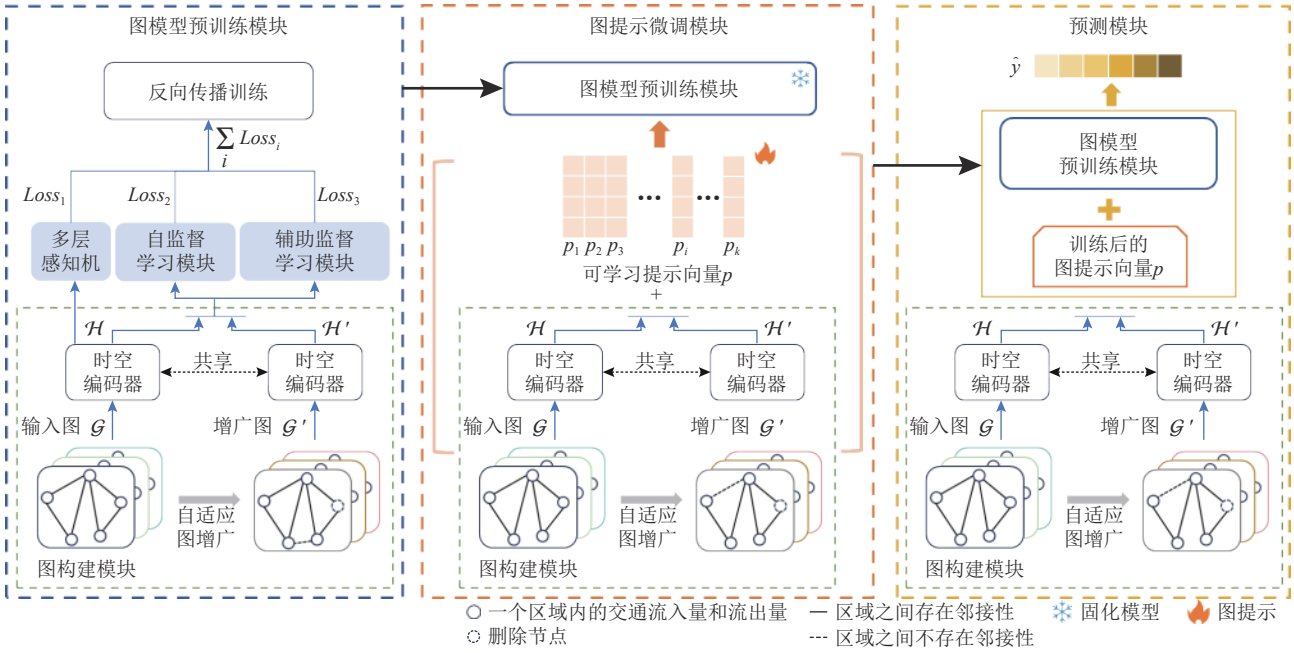


Fig. 1 Framework diagram of TPGPF model

图 1 TPGPF 模型框架图

第 1 个组件主要用于构建交通流量图，对单个时间段内的区域流量进行构图，其中将区域作为节点，流入和流出流量作为节点的属性，节点之间的边表示空间上的邻接属性. 将各个时间段的交通流量子图合并，构建出整体的交通流量图.

第 2 个组件是图模型多任务预训练模块，主要包括 3 个学习任务：1)通过 MLP 预测下一个时刻的交通流量；2)通过伪类标自监督学习模块捕获交通流量图的空间信息；3)通过构建正负样本进行对比学习捕获时间信息. 将这 3 个任务通过一个多任务学习框架进行训练，构成预训练模块的核心框架.

第 3 个组件是图提示微调模块，通过对时间子图和针管后的子图的原始表征，经过编码器后分别增加可学习的提示向量，在固化的预训练模型中进行提示向量的训练.

第 4 个组件是预测模块，将固化的预训练模型和提示向量融合进行未来交通流量的预测. 接下来介绍每个模块的具体实现细节.

2.2 图构建模块

交通流量数据通常记录为一个时间段 t 内任一给定区域 r 的流入量(表示为 f_{in})和流出量(表示为 f_{out}). 为了方便后面的讨论，这里我们先定义空间数据. 定义预测对象的地区可以分为 M 个区域，例如把一个城市分为若干个 10×10 平方公里的区域，其中第 i 个区域记为 r_i，则区域集合 $R=\{r_1, r_2, \cdots, r_M\}$，这里的 M 是划分的区域数量，注意划分过程中这些区域的面积可能不是严格相等的. 由于空间之间的相邻关系对交通流量预测有较大的影响，我们用一个 $M\times M$ 的矩阵 $\boldsymbol{A}$ 来表示区域之间的连接关系. 具体来说，第 i 个区域 r_i 和第 j 个区域 r_j 之间的连接关系可以表示为

$$a_{i,j}=\begin{cases}1, & \text{如果区域}r_i\text{和区域}r_j\text{存在邻接关系},\\ 0, & \text{其他情况},\end{cases} \tag{1}$$

这里 $a_{i,j}=1$ 表示区域 r_i 和区域 r_j 之间存在邻接关系，不存在邻接关系则记为 $a_{i,j}=0$.

此外，我们对时间数据进行定义，将统计数据的时间范围平均分为若干个时间段，如将 1 个月分为 30 天，每一天记为 1 个时间段，则第 k 个时间段我们记为 t_k，时间数据集 $T=\{t_1, t_2, \cdots, t_N\}$，$N$ 是划分的时间段数量.

基于以上定义，交通流量图(traffic flow graph, TFG)的构建过程如图 2 所示，将同一个时间段内的区域流量作为区域节点的数据，节点的连接线代表空间区域之间的邻接性. 基于时间维度，我们构造了一个交通流量图，表示为 $\boldsymbol{G}=(\boldsymbol{R}, E, \boldsymbol{A}, F)$，其中 $\boldsymbol{R}\in\mathbb{R}^{M\times S}$ 是 M 个空间区域的集合，S 为区域 r_i 的嵌入维度，E 是 2 个空间的连接边集合，$\boldsymbol{A}\in\mathbb{R}^{M\times M}$ 是邻接关系矩阵，F 是每个时间间隙中的交通流入流量和流出流量.

2.3 图模型预训练模块

针对交通流量图的编码表征，我们采用了 ST-

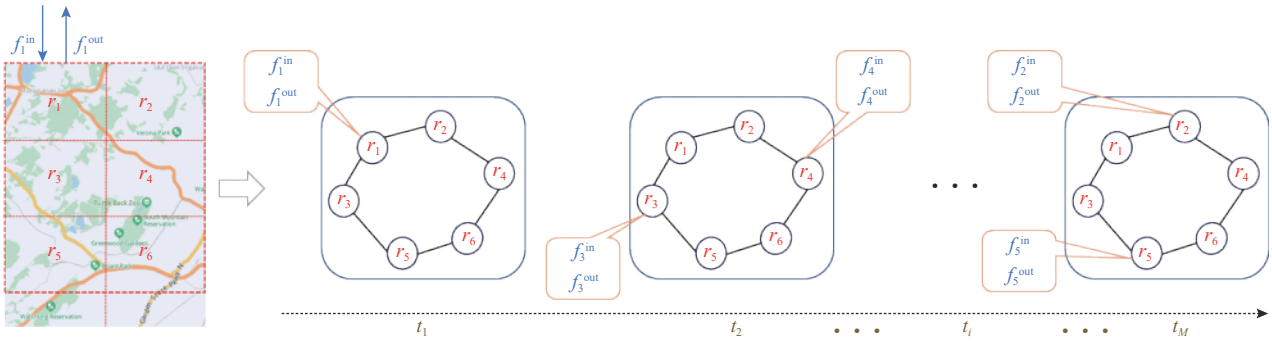


Fig. 2 Flow chart of the graph construction of TPGPF model

图 2 TPGPF 模型的图构建流程图

SSL 中的时空编码器[9]，通过有辅助的自监督学习方法来捕获空间数据和时间数据的异质性. 首先，通过一个基于门控机制的一维因果卷积时间编码器和空间卷积编码器，分别对交通流量图进行时间和空间上的编码，生成图表征 $\boldsymbol{H}\in\mathbb{R}^{M\times D}$，其中，$D$ 表示每个交通流量图在一个时间段内的表征维度. 为了进一步获取交通流量图中的特征信息，利用服从伯努利分布的掩码概率，分别从异质性引导层面、交通流量层面和图拓扑层面进行图增广[9]，得到增广后的交通流量图 $\boldsymbol{G}'$，并将流量图 $\boldsymbol{G}'$ 通过上述时空编码器生成表征 $\boldsymbol{H}'$. 以交通流量图 $\boldsymbol{G}'$ 和增广图 $\boldsymbol{G}'$ 生成的表征 $\boldsymbol{H}$ 和 $\boldsymbol{H}'$，在多任务框架中同时训练以下 3 个任务.

第 1 个任务是将区域 r_i 的表征 $\boldsymbol{h}_i$ 通过一个多层感知机(multilayer perceptron, *MLP*)进行传递，预测下一个未来时间 t+1 中的交通流量 $\hat{\boldsymbol{f}}_{t+1,i}\in\mathbb{R}^{M\times 2}$，包括 $\hat{f}^{\text{in}}_{t+1,i}$ 和 $\hat{f}^{\text{out}}_{t+1,i}$，

$$\hat{\boldsymbol{f}}_{t+1}=MLP(\boldsymbol{h}_i). \tag{2}$$

定义模型训练的损失函数为：

$$L_1=\sum_{i=1}^{M}\left[\lambda\left|\boldsymbol{f}^{\text{in}}_{t+1,i}-\hat{\boldsymbol{f}}^{\text{in}}_{t+1,i}\right|+(1-\lambda)\left|\boldsymbol{f}^{\text{out}}_{t+1,i}-\hat{\boldsymbol{f}}^{\text{out}}_{t+1,i}\right|\right], \tag{3}$$

其中，$\boldsymbol{f}^{\text{in}}_{t+1,i}$ 和 $\boldsymbol{f}^{\text{out}}_{t+1,i}$ 分别表示流入流量和流出流量的真实值，λ 表示是平衡交通流量影响的调整参数.

第 2 个任务是对交通流量图的增广图进行聚类，并将聚类的结果作为标签，指导基于原始交通流量图的聚类预测训练. 在区域 r_i 上采用基于软聚类的自监督学习方法，参考文献 [9] 的方式，首先生成 K 个簇表征 $\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_K\}$ 作为每个区域簇的潜在因素，计 c_K 算区域 r_i 的表征 $\boldsymbol{h}_i$ 和第 k 个簇表征 c_K 之间的相关性系数 $\tilde{z}_{i,k}=\boldsymbol{c}_k^{\mathrm{T}}\tilde{\boldsymbol{h}}_i$，这里 $\tilde{\boldsymbol{h}}_i\in\mathbb{R}^{1\times D}$ 是区域 r_i 的增广图 $\boldsymbol{G}'$ 通过编码器得到的表征. 因此，可以得到每一个区域 r_i 的簇标签 $\tilde{\boldsymbol{z}}_m=(\tilde{z}_{m,1},\cdots,\tilde{z}_{m,k})^{\mathrm{T}}$. 类似地，可以得到区域 r_i 原始图表征 $\boldsymbol{h}_i$ 的预测簇标签相关系数 $\hat{z}_{i,k}=\boldsymbol{c}_k^{\mathrm{T}}\boldsymbol{h}_i$. 对于所有区域而言，损失函数定义为

$$L_2=-\sum_{i=1}^{M}\sum_{k=1}^{K}\tilde{z}_{i,k}\log\frac{\exp(\hat{z}_{\mathrm{i,k}}/\omega)}{\sum_{\mathrm{j=1}}^{\mathrm{K}}\exp(\hat{z}_{\mathrm{j,k}}/\omega)}, \tag{4}$$

其中，ω 是用于控制归一化输出平滑程度的温度参数. 参考文献 [9] 的方法，采用最大熵原理的正则化分布方法来解决簇分配概率之和为 1 的问题.

第 3 个任务是通过对比学习来获取交通流量图的表征向量信息. 首先将原始交通流量图与增广交通流量图通过编码器得到的区域表征 $\boldsymbol{h}_{t,i}$ 和 $\tilde{\boldsymbol{h}}_{t,i}$ 进行融合，

$$\boldsymbol{v}_{t,i}=\gamma_1\otimes\boldsymbol{h}_{t,i}+\gamma_2\otimes\tilde{\boldsymbol{h}}_{t,i}, \tag{5}$$

其中⊗是元素积，γ_1，γ_2 是可学习的参数. 其次，将来自同一时间段的区域交通流量图表征和全集合交通流量图(城市级)表征作为正样本对；反之，不同时间段的区域交通流量图表征和全集合交通流量图表征则作为负样本对. 采用这个对比学习方法，能进一步捕捉交通流量图的时空异质性. 定义损失函数：

$$L_3=-\sum_{i=1}^{M}\log g(\boldsymbol{v}_{t,i},\boldsymbol{s}_t)-\sum_{i=1}^{M}\log(1-g(\boldsymbol{v}_{t',i},\boldsymbol{s}_t)), \tag{6}$$

其中 t 和 t' 分别表示 2 个不同的时间段，g 是一个标准函数，定义为：

$$g\left(\boldsymbol{v}_{t,i},\boldsymbol{s}_t\right)=\sigma\left(\boldsymbol{v}_{t,i}\boldsymbol{W}\boldsymbol{s}_t\right), \tag{7}$$

其中，$\boldsymbol{W}\in\mathbb{R}^{M\times M}$ 为可学习的变换矩阵，σ 是 sigmoid 函数，$\boldsymbol{s}_t$ 定义为

$$\boldsymbol{s}_t=\sigma\left(\frac{1}{N}\sum_{i=1}^{M}\boldsymbol{v}_{t,i}\right). \tag{8}$$

最后，将式(3)(4)(6)的 3 个损失函数进行联合训练，构建成新的损失函数：

$$L=L_1+L_2+L_3. \tag{9}$$

根据文献 [9]，我们采用反向传播训练算法来预训练 ST-SSL 模型. 这个预训练阶段获得图模型的最

优参数 Φ 和原始交通流量图的表征向量 $\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_M$, 这些参数在随后的提示微调训练阶段保持固定.

2.4 图提示微调模块

图模型预训练模块训练得到固化模型后, 我们将通过图提示微调模块, 采用嵌入提示向量的方式来提升预训练的固化模型和下游预测任务的适应性. 为了实现这一目标, 我们引入了图提示向量(GPF)[29], 这需要固定预训练模型中的参数 Φ, 并为每个区域的嵌入向量 $\boldsymbol{h}_i$ 分配一个独立的可学习提示向量 $\boldsymbol{p}_i \in \mathbb{R}^{d\times 1}$. 具体来说, 通过将可学习的提示向量 $\boldsymbol{p}_i$ 添加到初始交通流量图表征向量 $\boldsymbol{h}_i$ 中, 生成向量 $\boldsymbol{h}_i$ 的提示特征向量 $\boldsymbol{h}_i^*$,

$$\boldsymbol{h}_i^* = \boldsymbol{h}_i + \boldsymbol{p}_i. \tag{10}$$

如果按照节点数量来配置提示向量 $\boldsymbol{p}_i$ 的数量, 会增加提示微调模型的计算负担. 为了进一步提升实时交通流量数据的提示微调性能, 降低学习多个提示向量的计算负担, 我们减少了提示向量的数量, 并引入注意力机制来协同多个提示向量的作用. 具体地, 我们采用 k 个独立的基础提示向量进行训练, 并把基础提示向量标记为 $\boldsymbol{p}_1^{\text{basic}}, \boldsymbol{p}_2^{\text{basic}}, \cdots, \boldsymbol{p}_k^{\text{basic}} \in \mathbb{R}^{d\times 1}$, k 是可以根据下游任务进行调整的超参数. 并为每个节点 $\boldsymbol{h}_i$ 定义一组系数 $\boldsymbol{a}_i=(a_{i,1}, a_{i,2}, \cdots, a_{i,k})$, 因此, 每个节点对应的提示向量 $\boldsymbol{p}_i$ 都可以通过这些基础向量的注意力聚合得到, 定义为:

$$\boldsymbol{p}_i = \sum_{j=1}^{k} \alpha_{i,j} \boldsymbol{p}_j^{\text{basic}}, \ \alpha_{i,j} = \frac{\exp\left(\boldsymbol{a}_j^{\mathrm{T}} \boldsymbol{h}_i\right)}{\sum_{l}^{k} \exp\left(\boldsymbol{a}_l^{\mathrm{T}} \boldsymbol{h}_i\right)}, \tag{11}$$

其中 $\boldsymbol{a}_j \in \mathbb{R}^{d\times 1}$, $\forall j = 1,2,\cdots,k$. 通过将原始图表征 $\boldsymbol{h}_i$ 和增广图表征 $\boldsymbol{h}_i'$ 分别替换为增加提示向量后的提示训练表征 $\boldsymbol{h}_i^*$ 和 $\boldsymbol{h}_i'^*$, 可以在 ST-SSL 模型的预训练架构中进行 $\boldsymbol{p}$ 参数训练, 并得到提示特征 $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_M\}$.

2.5 预测模块

在获取固化的预训练图模型参数矩阵 $\boldsymbol{\Phi}$ 和训练得到的提示特征 $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_M\}$后, 我们可以利用带有图提示的预测模型来开展未来时间的交通流量预测[29].

3 实验与结果

在本节中, 为了验证所提的图提示学习方法在交通流量预测的有效性, 我们在 5 个公开的真实数据集上进行了大量的实验, 并进行了综合分析. TFGP 的所有代码和数据集可以从 https://pan.baidu.com/s/199zDBu3b-FHZKhMHrojpvw 下载. 实验主要围绕 3 个问题:

1)本文提出的 TPGPF 模型在交通流量预测上是否比其他模型具有优越性.

2)在图提示层面的超参数设置对 TPGPF 模型的影响如何.

3)在 TPGPF 模型中不同组件的作用影响如何.

3.1 实验设置与数据集

为了评估 TPGPF 的有效性和通用性, 我们在 NYCBike1[20], NYCBike2[30], NYCTaxi[20], CHIBike[31] 和 TDrive[31] 等 5 个不同的公开数据集上进行了实验, 数据集信息如表 1 所示. 时间段间隔有 30 min 和 1 h 两种设置, 车辆数量包括自行车的数量或者出租车的数量 2 种, 为了验证模型在不同规模场景下的预测性能, 选用了节点数量跨度比较大的数据集, 从 6 800 多辆到 4 亿多辆, 代表不同规模的数据. 数据主要在 2014—2020 年采集.

Table 1 Statistics of the Experimental Datasets

表 1 实验数据集统计信息

数据集	时间段长度/h	区域面积	车辆数量	时间范围
NYCBike1	1	16×8	6.8×10^3	04/01/2014—09/30/2014
NYCBike2	0.5	10×20	2.6×10^6	07/01/2016—08/29/2016
NYCTaxi	0.5	10×20	22×10^6	01/01/2015—03/01/2015
CHIBike	0.5	15×18	1.2×10^6	07/01/2020—09/30/2020
TDrive	1	32×32	493×10^6	02/01/2015—06/30/2015

在所有数据集的实验中, 我们采用预测时间之前的 2 h 和 3 天的历史交通流量数据来预测下一个时间段的流量, 并采用滑动窗口方法来进行交通数据的采样. 对于每个数据集, 按照 7∶1∶2 的比例来划分预训练模型训练数据集、图提示向量训练数据集和测试集. 在实验评估中, 我们使用了 2 个常见的指标: 平均绝对误差(mean absolute error, MAE)损失和平均绝对百分比误差(mean absolute percentage error, MAPE), 分别评估流入量和流出量的预测性能. 此外, 所有的实验都是在一台装载 NVIDIA Tesla V100 的云服务器上进行.

3.2 与基线的对比实验

为了更好地评估 TPGPF 模型的预测性能, TPGPF 和 8 个基线模型进行了性能对比, 其中包括 4 个时空融合预测方法, 即 ST-ResNet, ST-GCN, GMAN 和 PDFormer[31]; 3 个基于异质性的时空预测方法, 即 STS-GCN, STF-GNN 和 ST-SLL.此外, 我们还引入了一个基于大语言模型和知识图谱相结合的预测模型, 主要用于改进 KP4SR 模型[32], 即将交通节点的流量数

据作为图谱的属性节点,并基于大语言模型进行交通流量的预测.

在对比实验中,将每个交通流量图在一个时间段内的表征维度 D 设置为 64. 编码器的时间和空间卷积核大小都设置为 3. 在交通流量图的增广图中,我们把流量增广和拓扑增广的扰动比都设置为 0.1. 训练阶段使用 Adam 优化器,批大小设置为 32. 对于基线,使用相应作者提供的代码,且所有的超参数和初始化策略都根据作者的建议进行设置,部分基线实验直接采用了已公开的实验结果[9,31].

如表 2 所示,相比于 4 个时空融合预测模型,TPGPF 总体性能明显领先,尤其是在 CHIBike 数据集中,流入流量预测的 MAPE 提升高达 7.73 个百分点. 如表 3 所示,在各种不同规模的数据集中,大部分考虑异质性的方法比未考虑异质性的方法更好,尤其是 ST-SLL 方法在 5 个数据集上都全面领先于其他基线方法. 这说明在交通流量预测中,充分考虑空间信息的关联性和时空异质性,有助于取得更好的预测性能. TPGPF 都表现出整体领先的性能. 其中,对于车辆规模最小的 NYCBike1 数据集,有全面的提升,在 MAPE 方面的表现上提升明显,尤其是流入方向上,MAPE 指标达到 23.69%,优于排名第 2 的 ST-SLL(24.06%). 流入和流出方向的 MAE 指标也分别达到 4.88 和 5.19,优于性能排名第 2 的 ST-SSL(MAE 指标分别为 4.90 和 5.21). 在车辆节点数量最大的 TDrive 数据集中,TPGPF 也有整体领先的性能表现. 在 CHIBike 数据集中,流入方向的 MAPE 指标也提升明显,从到 23.25% 提升到 22.84%. 虽然有 2 个性能结果略低于 ST-SLL 模型,但基本处于相同水平,这说明通过提示学习的方式,在固化大规模数据预训练的基础上,通过实时的小数据集进行下游任务的适配性训练,具有较好的可行性.

Table 2 Performance Comparison of TPGPF Model and Spatial-Temporal Fusion Models on Five Datasets in Terms of MAE and MAPE

表 2 TPGPF 模型与时空融合模型在 5 个数据集上的 MAE 和 MAPE 性能比较

数据集	指标	类型	ST-ResNet	ST-GCN	GMAN	PDFormer	TPGPF(本文)
NYCBike1	MAE	流入	5.53±0.06	5.33±0.02	6.77±3.42	5.36	**4.88**
		流出	5.74±0.07	5.59±0.03	7.17±3.61	5.71	**5.19**
	MAPE/%	流入	25.46±0.20	26.92±0.08	31.72±12.29	24.51	**23.69**
		流出	26.36±0.50	27.69±0.14	34.74±17.04	25.42	**24.49**
NYCBike2	MAE	流入	5.63±0.14	5.21±0.02	5.24±0.13	4.91	**5.07**
		流出	5.26±0.08	4.92±0.02	4.97±0.14	4.59	**4.73**
	MAPE/%	流入	32.17±0.85	27.73±0.16	27.38±1.13	27.25	**22.32**
		流出	30.48±0.86	26.83±0.21	26.75±1.14	25.87	**21.39**
NYCTaxi	MAE	流入	13.48±0.14	13.12±0.04	15.09±0.61	12.76	**12.15**
		流出	10.78±0.25	10.35±0.03	12.06±0.39	10.56	**9.93**
	MAPE/%	流入	24.83±0.55	21.01±0.18	22.73±1.20	19.99	**16.66**
		流出	24.42±0.52	20.78±0.16	21.97±0.86	20.12	**17.08**
CHIBike	MAE	流入	4.767	4.212	4.115	**3.950**	3.99
		流出	4.627	4.148	4.090	**3.837**	3.98
	MAPE/%	流入	31.382	31.224	31.150	30.214	**22.84**
		流出	30.571	30.782	30.662	29.914	**22.72**
TDrive	MAE	流入	19.636	21.373	19.244	16.12	**15.76**
		流出	19.616	20.913	18.964	16.03	**15.76**
	MAPE/%	流入	17.831	17.539	17.110	16.46	**16.00**
		流出	18.502	16.984	15.788	16.39	**16.00**

注:黑体数值表示最高得分,下划线数值表示次优得分.

3.3 图提示向量参数分析

为进一步评估 TPGPF 模型中图提示向量对整个模型的预测性能影响,我们对提示向量的数量进行参数分析. 如图 3 所示,MAE 的性能基本不受提示向量数量的影响,在参数 k=1~10 情况下基本保持稳定. TPGPF 模型的 MAPE 性能随着提示向量数量的增加而增加;然而达到一定数量后,比如 k=8 以后,性能曲线出现拐点开始下降,说明图提示数量并非越多

Table 3 Performance Comparison of TPGPF Model and Spatial-Temporal Heterogeneity Models in Terms of MAE and MAPE

表 3 TPGPF 模型与时空异质模型的 MAE 和 MAPE 性能比较

数据集	指标	类型	STS-GCN	STF-GNN	ST-SSL	TPGPF（本文）
NYCBike1	MAE	流入	5.81±0.04	6.53±0.10	4.90	**4.88**
		流出	6.10±0.04	6.79±0.08	5.21	**5.19**
	MAPE/%	流入	26.51±0.32	32.14±0.23	24.06	**23.69**
		流出	27.56±0.39	32.88±0.19	24.60	**24.49**
NYCBike2	MAE	流入	5.25±0.03	5.80±0.10	**5.06**	5.07
		流出	4.94±0.05	5.51±0.11	**4.73**	**4.73**
	MAPE/%	流入	29.26±0.13	30.73±0.49	22.57	**22.32**
		流出	28.02±0.23	29.98±0.46	21.55	**21.39**
NYCTaxi	MAE	流入	13.69±0.11	16.25±0.38	12.22	**12.15**
		流出	10.75±0.17	12.47±0.25	10.00	**9.93**
	MAPE/%	流入	22.91±0.44	24.01±0.30	17.10	**16.66**
		流出	22.37±0.16	23.28±0.47	17.11	**17.08**
CHIBike	MAE	流入	4.256	4.234	**3.96**	3.99
		流出	4.265	4.264	4.03	**3.98**
	MAPE/%	流入	32.991	32.222	23.25	**22.84**
		流出	32.612	32.321	22.73	**22.72**
TDrive	MAE	流入	23.825	22.144	15.80	**15.76**
		流出	24.287	22.876	15.79	**15.76**
	MAPE/%	流入	18.094	18.094	16.07	**16.00**
		流出	18.987	18.987	16.09	**16.00**

注：黑体数值表示最高得分，下划线数值表示次优得分.

越好，而与具体的数据场景有关，对输入的优化必须根据实际的情况来确定图提示的数量规模，从而得到较好的预测性.

3.4 消融实验分析

为评估 TPGPF 模型中各个模块对模型的整体影响，我们通过消融实验来进行分析. 首先我们移除预训练固化模型和提示微调模型中时空编码器中的结构增广，将其替换为随机边增删的方式，并作为第 1 个变体来分析结构增广方式对模型的影响，称之为 TPGPF-sa.第 2 个消融变体移除了预训练固化模型和提示微调模型中时空编码器中的流量增广，增加了随机增广方式，该变体称之为 TPGPF-sa. 表 4 给出了消融实验结果，可以看出，在车辆数量规模最小的 NYCBike1 和车辆数量最大的 TDrive 数据集中，TPGPF 模型保持着领先的性能，但提升的性能不显著，甚至在随机的图增广方式中，2 个变体模型在个别性能上超过了 TPGPF 模型. 相对而言，变体模型 TPGPF-ta 的性能优于 TPGPF-sa，可见流量增广模块在模型中的作用比结构增广的作用更大. 与此同时，ST-SSL 也可以作为图提示向量的一个消融实验的变体，从表 3 中可以看出，TPGPF 模型整体性能上全面领先于 ST-SSL，说明图提示微调模块在模型中起到了重要的预测性能提升作用，例如在 NYCTaxi 数据集中，流入方向的预测指标 MAPE 比 ST-SSL 提升了 0.5 个百分点.

4 总结与展望

在本文工作中，我们提出了一种名为 TPGPF 的交通流量预测方法，首次引入基于图提示来增强传统交通流预测模型的泛化能力. 多个不同公开数据集的实验结果表明，所提出的模型优于多种最先进的基线. 在未来，我们将探索更多的图构建模式，例如以区域为维度构建基于时间维度的子图，观察单一区域在较长时间中的潜在信息，代替现在使用时间段内的区域子图方式. 此外，随着大语言模型的广泛应用和研究的深入，其在预测任务方面也表现出较好的性能预期. 未来，基于大语言模型的提示微调，甚至直接基于大语言模型进行交流流量的预测，都是可以进一步探索的研究方向.

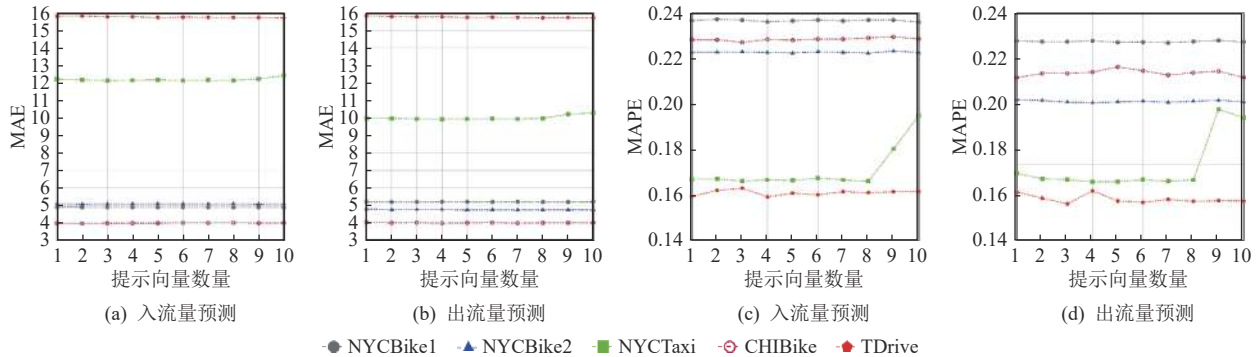


Fig. 3 Parameters comparison of graph prompt feature

图 3 图提示向量参数比较

Table 4 Ablation Study Results for TPGPF Model

表 4 TPGPF 模型的消融实验结果

数据集	指标	类型	TPGPF	TPGPF-sa	TPGPF-ta
NYCBike1	MAE	流入	**4.88**	4.89	4.91
		流出	**5.19**	5.21	5.26
	MAPE/%	流入	23.69	23.67	**23.56**
		流出	24.49	**24.45**	24.59
NYCBike2	MAE	流入	**5.07**	5.11	5.08
		流出	4.73	4.74	**4.71**
	MAPE/%	流入	**22.32**	22.85	22.42
		流出	21.39	21.38	**21.23**
NYCTaxi	MAE	流入	12.15	12.22	**12.08**
		流出	9.93	**9.78**	9.95
	MAPE/%	流入	16.66	16.93	**16.57**
		流出	**17.08**	17.27	17.11
CHIBike	MAE	流入	**3.99**	4.07	4.11
		流出	**3.98**	4.06	4.14
	MAPE/%	流入	**22.84**	22.99	**22.63**
		流出	**22.72**	22.74	22.97
TDrive	MAE	流入	**15.76**	16.02	15.84
		流出	**15.76**	15.98	15.82
	MAPE/%	流入	**16.00**	16.51	16.12
		流出	**16.00**	16.09	16.03

注：黑体数值表示最优值.

作者贡献声明：赖培源提出了算法思路、实验方案并撰写论文；李程、王增辉负责完成算法编程及实验；王昌栋提出指导意见并修改论文；廖德章负责实验数据采集和预处理.

参 考 文 献

[1] Bernardo G, José C, Helena A. A survey on traffic flow prediction and classification[J]. Intelligent Systems with Applications, 2023, 20: 200268

[2] Wang Meng, Wang Haofen, Li Bohan, et al. Survey on key technologies of new generation knowledge graph[J]. Journal of Computer Research and Development, 2022, 59(9): 1947−1965 (in Chinese)
(王萌, 王昊奋, 李博涵, 等. 新一代知识图谱关键技术综述[J]. 计算机研究与发展, 2022, 59(9): 1947−1965)

[3] Chen Rui, Wang Zhanquan. Uni-LSDPM: A unified online learning session dropout prediction model based on pre-training[J]. Journal of Computer Research and Development, 2024, 61(2): 441−459 (in Chinese)
(陈芮, 王占全. Uni-LSDPM: 基于预训练的统一在线学习会话退出预测模型[J]. 计算机研究与发展, 2024, 61(2): 441−459)

[4] Cheng Zeyang, Lu Jian, Zhou Huajian, et al. Short-term traffic flow prediction: An integrated method of econometrics and hybrid deep learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 5231−5244

[5] Cao Shuqin, Wu Libing, Wu Jia, et al. A spatio-temporal sequence-to-sequence network for traffic flow prediction[J], Information Sciences, 2022, 610: 185−203

[6] Wang Hanqiu, Zhang Rongqing, Cheng Xiang, et al. Hierarchical traffic flow prediction based on spatial-temporal graph convolutional network[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 16137−16147

[7] Wang Ming, Peng Jian, Huang Feihu. Multi-time scale spatial-temporal graph neural network for traffic flow prediction[J]. Computer Science, 2022, 49(8): 40−48 (in Chinese)
(汪鸣, 彭舰, 黄飞虎. 基于多时间尺度时空图网络的交通流量预测模型[J]. 计算机科学, 2022, 49(8): 40−48)

[8] Zhao Wenzhu, Yuan Guan, Zhang Yanmei, et al. Multi-view fused spatial-temporal dynamic graph convolutional network for urban traffic flow prediction[J]. Journal of Software, 2024, 35(4): 1751−1773 (in Chinese)
(赵文竹, 袁冠, 张艳梅, 等. 多视角融合的时空动态图卷积网络城市交通流量预测[J]. 软件学报, 2024, 35(4): 1751−1773)

[9] Ji Jiahao, Wang Jingyuan, Huang Chao, et al. Spatio-temporal self-supervised learning for traffic flow prediction[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2023, 37(4): 4356−4364

[10] Lin Guancen, Lin Aijing, Gu Danlei, et al. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient[J]. Information Sciences, 2022, 608: 517−531

[11] Ni Qingjian, Peng Wenqiang, Zhang Zhizheng, et al. Spatial-temporal graph neural network for traffic flow prediction based on information enhanced transmission[J]. Journal of Computer Research and Development, 2022, 59(2): 282−293 (in Chinese)
(倪庆剑, 彭文强, 张志政, 等. 基于信息增强传输的时空图神经网络交通流预测[J]. 计算机研究与发展, 2022, 59(2): 282−293)

[12] Yan Biwei, WangGuijuan, Yu Jiguo, et al. Spatial-temporal chebyshev graph neural network for traffic flow prediction in IoT-based ITS[J]. IEEE Internet of Things Journal, 2022, 9(12): 9266−9279

[13] Zheng Haifeng, Lin Feng, Feng Xinxin, et al. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(11): 6910−6920

[14] Du Shengdong, Li Tianrui, Yang Yan, et al. A sequence-to-sequence spatial-temporal attention learning model for urban traffic flow prediction[J]. Journal of Computer Research and Development, 2020, 57(8): 1715−1728 (in Chinese)
(杜圣东, 李天瑞, 杨燕, 等. 一种基于序列到序列时空注意力学习的交通流预测模型[J]. 计算机研究与发展, 2020, 57(8): 1715−1728)

[15] Sun Mingchen, Zhou Kaixiong, He Xin, et al. GPPT: Graph pre-training and prompt tuning to generalize graph neural networks[C]//Proc of the 28th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2022: 1717−1727

[16] Hu Ziniu, Dong Yuxiao, Wang Kuansan, et al. GPT-GNN: Generative

pre-training of graph neural networks[C]//Proc of the 26th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2020: 1857−1867

[17] Lu Yuanfu, Jiang Xunqiang, Fang Yuan, et al. Learning to pre-train graph neural networks[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2021, 35(5): 4276−4284

[18] Castro-Neto M, Jeong Y S, Jeong M K, et al. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions[J]. Expert Systems with Applications, 2009, 36(3): 6164−6173

[19] Kumar S V, Vanajakshi L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data[J]. European Transport Research Review, 2015, 7(3): 1−9

[20] Zhang Junbo, Zheng Yu, Qi Dekang. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2017, 31(1): 1655−1661

[21] Yu Bing, Yin Haoteng, Zhu Zhanxing. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting [C]//Proc of Int Joint Conf on Artificial Intelligence Stockholm, Sweden: Ijcai.org, 2018: 3634−3640

[22] Zheng Chuanpan, Fan Xiaoliang, Wang Cheng, et al. GMAN: A graph multi-attention network for traffic prediction[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2020, 34(1): 1234−1241

[23] Bai Lei, Yao Lina, Li Can, Wang Xianzhi, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17804−17815

[24] Song Chao, Lin Youfang, Guo Shengnan, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2020, 34(1): 914−921

[25] Sun Xiangguo, Zhang Jiawen, Wu Xixi, et al. Graph prompt learning: A comprehensive survey and beyond[J]. arXiv preprint, arXiv: 2311.16534, 2023

[26] Liu Zemin, Yu Xingtong, Fang Yuan, et al. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks[C]//Proc of the ACM Web Conf 2023. New York: ACM, 2023: 417−428

[27] Chen Mouxiang, Liu Zemin, LiuChenghao, et al. ULTRA-DP: Unifying graph pre-training with multi-task graph dual prompt[J]. arXiv preprint, arXiv: 2310.14845, 2023

[28] Chen Zihao, WangYing, Ma Fuyuan, Yuan Hao, et al. GPL-GNN: Graph prompt learning for graph neural network[J]. Knowledge-Based Systems, 2024, 286: 111391

[29] Fang Taoran, Zhang Yunchao, Yang Yang, et al. Universal prompt tuning for graph neural networks[J]. Advances in Neural Information Processing Systems, 2023, 36: 52464−52489

[30] Yao Huaxiu, Tang Xianfeng, Wei Hua, et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2019: 5668−5675

[31] Jiang Jiawei, Han Chengkai, Zhao W X, et al. PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2023, 37(4): 4365−4373

[32] Zhai Jianyang, Zheng Xiawu, Wang Changdong, et al. Knowledge prompt-tuning for sequential recommendation[C]// Proc of the 31st ACM Int Conf on Multimedia. New York: ACM, 2023, 6451−6461



Lai Peiyuan, born in 1984. PhD. Senior engineer. Student member of CCF. His main research interests include data mining, recommendation algorithm, and graph neural network.

赖培源, 1984 年生. 博士研究生, 高级工程师, CCF 学生会员. 主要研究方向为数据挖掘、推荐算法、图神经网络.



Li Cheng, born in 1995. Master candidate. His main research interests include graph data mining and AI for medicine.

李　程, 1995 年生. 硕士研究生. 主要研究方向为图数据挖掘、AI 医疗.



Wang Zenghui, born in 1998. Bachelor, software engineer. His main research interests include data mining, recommendation algorithm, and graph clustering.

王增辉, 1998 年生. 本科, 软件工程师. 主要研究方向为数据挖掘、推荐算法、图聚类.



Wang Changdong, born in 1984. PhD, associate professor, PhD supervisor. Distinguished member of CCF and IEEE senior member. His main research interests include graph data mining, AI for medicine, multi-view clustering, and AI for agriculture.

王昌栋, 1984 年生, 博士, 副教授, 博士生导师. CCF 杰出会员, IEEE 高级会员. 主要研究方向为图数据挖掘、AI 医疗、多视图聚类、AI 农业.



Liao Dezhang, born in 1993. Bachelor, software engineer. His main research interests include data mining, recommendation algorithm, and graph clustering.

廖德章, 1993 年生. 本科, 软件工程师. 主要研究方向为数据挖掘、推荐算法、图聚类.