

## 城市场景下基于域适应迁移学习的交通流预测方法

夏振厂<sup>1,2,3</sup>, 班一杰<sup>2</sup>, 田有亮<sup>1</sup>, 刘冰艺<sup>2</sup>, 崔建群<sup>4</sup>, 常亚楠<sup>4</sup>

(1. 贵州大学公共大数据国家重点实验室, 贵州 贵阳 550025; 2. 武汉理工大学计算机与人工智能学院, 湖北 武汉 430070;  
3. 吉林大学符号计算与知识工程教育部重点实验室, 吉林 长春 130012; 4. 华中师范大学计算机学院, 湖北 武汉 430079)

**摘要:** 针对在数据与算力受限场景下交通流预测模型跨城市迁移能力的问题, 提出了一种基于域适应的可迁移交通流预测方法, 并设计了迁移学习模型及轻量化时空序列预测网络。首先, 设计一种基于多核均值最大差异的节点级域适应方法去捕获多元尺度特征, 提高方法的跨域特征表示能力。其次, 采用多元信息联合学习城市区域特征, 并增加特征映射层, 通过非线性变换的特性将源域与目标域的特征分布对齐, 从而提升迁移能力。最后, 设计一种轻量化时空序列预测网络来满足交通流预测方法在硬件算力受限情况下的高实时性要求。实验结果表明, 相较于基准方法, 所提方法性能提高 9.03%。

**关键词:** 交通流预测; 迁移学习; 域适应; 轻量化预测网络

**中图分类号:** TP391

**文献标志码:** A

**DOI:** 10.11959/j.issn.1000-436x.2025128

## Traffic prediction method based on domain adaptive transfer learning in urban scenarios

XIA Zhenchang<sup>1,2,3</sup>, BAN Yijie<sup>2</sup>, TIAN Youliang<sup>1</sup>, LIU Bingyi<sup>2</sup>, CUI Jianqun<sup>4</sup>, CHANG Yanan<sup>4</sup>

1. State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China

2. School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China

3. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

4. School of Computer Science, Central China Normal University, Wuhan 430079, China

**Abstract:** To address the issues of cross-city transfer capability of traffic prediction models in scenarios with limited data and computing resources, a domain adaptation transfer learning-based traffic prediction method was proposed, along with a transfer learning model and a lightweight spatiotemporal sequence prediction network. Firstly, a node-level domain adaptation method based on the maximum mean discrepancy of multiple kernels was designed to capture multi-scale features, thereby improving the method's cross-domain feature representation. Secondly, multi-information joint learning was adopted to learn urban regional features, and an additional feature mapping layer was added, aligning the feature distribution of the source domain and target domain through nonlinear transformation characteristics, thereby enhancing transferability. Finally, a lightweight spatiotemporal sequence prediction network was designed to meet the high real-time requirements of traffic prediction methods when hardware computing power was limited. Experimental results show that the proposed method has achieved a performance improvement of 9.03% compared to benchmark methods.

**Keywords:** traffic prediction, transfer learning, domain adaptation, lightweight prediction network

收稿日期: 2025-02-08; 修回日期: 2025-07-07

通信作者: 常亚楠, yncchang@mail.ccnu.edu.cn

基金项目: 国家自然科学基金资助项目(No.62372206); 湖北省自然科学基金资助项目(No.2023AFB697); 公共大数据国家重点实验室基金资助项目(No.PBD2024-0503); 中央高校基本科研业务费专项资金资助项目(No.93K172024K21)

**Foundation Items:** The National Natural Science Foundation of China (No.62372206), The Natural Science Foundation of Hubei Province(No.2023AFB697), The Foundation of State Key Laboratory of Public Big Data (No.PBD2024-0503), The Fundamental Research Funds for the Central Universities (No.93K172024K21)

## 0 引言

随着经济的发展,越来越多的家庭拥有私家车。汽车保有量的增加,给城市交通控制和管理带来巨大挑战。在此背景下,高精度的交通流预测在智慧城市中起着至关重要的作用。及时准确地进行交通流预测不仅有助于城市从业者的决策和资源配置,也有助于城镇居民的出行规划。然而,基于深度学习的交通流预测模型需要大量的数据来训练模型,而许多城市缺乏足够数据和算力来训练高复杂度的深度学习模型(尤其是图神经网络模型通常包含海量参数,训练它们需要大量数据支撑),直接应用预训练模型又存在数据特征差异大、适应性差、预测精度低等问题。因此,研究基于迁移学习的交通流预测通过引入源城市的数据和知识,可以在目标城市数据有限的情况下获得良好的训练效果,解决跨域知识,弥补目标域数据和模型复杂度之间的矛盾,提高交通流的预测精度。

近年来,由于交通流预测具有较高的应用价值,受到了学术界和工业界的关注,并提出了很多有效的解决方法。早期的交通流预测模型大多数是基于自回归综合移动平均(ARIMA, autoregressive integrated moving average)<sup>[1]</sup>、支持向量回归(SVR, support vector regression)<sup>[2]</sup>、卡尔曼滤波<sup>[3]</sup>等统计或传统机器学习的思想,例如,Feng等<sup>[4]</sup>提出了基于具有时空相关性的自适应多核支持向量机(AMSVM-STC, adaptive multi-kernel support vector machine with spatial-temporal correlation)方法,用于短期交通流预测,并考虑了时间和空间的相关性。何祖杰等<sup>[5]</sup>通过改进灰狼算法(IGWO, improved grey wolf optimization algorithm)优化支持向量机(SVM, support vector machine),有效解决了传统灰狼算法收敛速度慢及容易陷入局部最优解的问题。然而,这些方法作为浅层模型,无法有效捕捉交通数据中的复杂非线性时空相关性。

为了提高交通流预测精度,研究人员提出了大量基于深度学习的方法。深度学习方法凭借其强大处理能力逐渐应用于交通流预测<sup>[6]</sup>,其中图神经网络<sup>[7]</sup>(GNN, graph neural networks)可有效捕捉交通数据中的时空依赖关系,为交通流量预测提供新的思路。Bai等<sup>[8]</sup>提出时空图到序列模型(STG2

Seq, spatial-temporal graph to sequence model),通过图结构捕捉城市多级客流特征;Pan等<sup>[9]</sup>通过深度元学习提升不同区域的交通预测性能。本文发现,虽然现有基于图神经网络的交通流预测方法在一定程度上能够提升预测精度,但仍然存在以下2个问题。

1) 现有基于图神经网络的方法虽然在交通流预测中展示了强大的特征提取能力,但是其固有的复杂网络结构往往伴随大量的参数,这意味着需要丰富的数据资源支持进行训练。然而,在实际城市环境中存在数据资源不均衡性,许多城市由于数据监测基础设施不完善,缺乏足够的支撑,严重限制了模型的泛化能力<sup>[10]</sup>。

2) 尽管深度学习方法在交通流量预测中展示了巨大潜力,但它们高度依赖大规模训练数据集,这在算力资源有限的实际应用中构成挑战。

现有基于图神经网络的交通流预测方法的这2个不足将直接影响交通流预测精度的进一步提高。

基于以上分析,本文提出一种基于域适应的可迁移交通流预测(DATE, domain adaptive transfer learning)方法。为了提升不同城市数据的跨域特征表达能力,本文方法采用多核最大均值差异损失函数进行节点级的域适应,从而提高模型的适应性和健壮性。本文的主要贡献包含以下3个方面。

1) 设计了一种基于多核最大均值差异损失函数进行节点级的域适应,从而提升域适应过程中的跨域特征表示能力和增强模型在未知数据上的泛化性能,减少过拟合风险。

2) 为解决源城市与目标城市特征分布差异,构建了一种特征映射网络,该网络在特征提取网络后端,能够通过前馈神经网络结构学习非线性映射,从而对齐不同城市特征分布和增强迁移学习能力。

3) 设计了一种轻量化时空序列预测网络去提高预测模型的训练效率及其在低算力场景下的适配性。该网络在空间特征提取模块中应用深度可分离式卷积(DSC, depthwise separable convolution)核以高效提取空间特征,引入注意力机制的门控循环单元(GRU, gated recurrent unit)进行时空序列特征捕捉以减少处理时间序列信息时的计算负担,从而提高训练效率。

## 1 相关工作

随着机器学习技术的发展,深度学习技术已经在交通流预测方面取得了巨大成功,成为相关研究领域的首选工具。从模型构建的角度来看,图神经网络方法和基于迁移学习的方法是当前深度学习在交通流预测领域的2个重要分支。

### 1.1 基于图神经网络的交通流预测方法

图神经网络作为处理图结构的强大工具,广泛用于交通流预测中。在图神经网络的应用中,Bai等<sup>[8]</sup>提出一种基于图神经网络的STG2Seq模型,通过使用图神经网络来捕捉城市的多级客流特征,从而提高交通流预测的精度。Lu等<sup>[11]</sup>利用空间和语义邻域理解城市交通动态,构建基于图结构的交通流预测模型。Pan等<sup>[9]</sup>将深度元学习方法应用到提升不同区域的交通预测性能中。Lu等<sup>[12]</sup>提出了一种模型无关的少样本时空图学习(ST-GFSL, spatio-temporal graph few-shot learning)框架,旨在通过跨城市知识增强特征提取。ST-GFSL框架设计了基于节点级元知识的非共享参数,使特征提取定制化,并通过参数匹配在相似的时空特征中传递知识。Jin等<sup>[13]</sup>则设计了名为基于域选择的跨城市迁移(CrossTReS, cross-city transfer via region selection)的选择性跨城市迁移学习框架,通过自适应地对源城市区域重新加权,改善目标城市的微调效果。Wang等<sup>[14]</sup>在交通流预测中,通过采用深度可分离卷积构建轻量化时空网络,从而去验证其在高维数据中的有效性。尽管基于图神经网络的交通流预测方法在交通流量预测中展现出巨大潜力,但它们高度依赖大规模训练数据集,这在算力资源有限或数据获取困难的实际应用中构成挑战。因此,如何在这些约束条件下设计和优化网络模型,以适应不同规模的数据和计算要求,仍是一个亟待解决的问题。

### 1.2 基于迁移学习的交通流预测方法

迁移学习作为应对数据稀缺和资源分配不均的有效方法,在交通流预测任务中得到了广泛关注与拓展应用。现有的基于迁移学习的交通流预测方法一般以微调为基础,使用监督学习或元学习在丰富的源数据上训练模型,然后对设计算法在目标数据上进行微调。模型无关的元学习(MAML, model-agnostic meta-learning)<sup>[15]</sup>和爬行动物模型(Rep-tile)<sup>[16]</sup>等基于元学习方法提出多源参数初始化,

通过少量迭代快速适应新任务,减少目标域微调需求,为时空序列预测提供了新视角。时空元学习(MetaST, meta-learning spatial-temporal)<sup>[17]</sup>方法构建了多源城市迁移学习的长期时空流量预测模型,相比单源模型大幅提高了性能,突显出多源数据迁移的潜力。然而,现有基于迁移学习的交通流预测方法忽略了源训练数据和目标数据微调之间的差距。因此,在丰富的源数据上学习到的模型可能包含噪声,甚至是对目标城市有害的源知识,从而对迁移学习性能产生负面影响。最近,一些研究人员采用了基于图神经网络的迁移学习方法来预测数据稀缺的交通网络中的图结构交通流数据。Huang等<sup>[18]</sup>首先采用图聚类方法将交通网络划分为多个子图,并将子图与图卷积网络(GCN, graph convolutional network)相结合在数据丰富的源领域训练模型。然后,将训练好的模型应用于数据稀缺的目标领域。Tang等<sup>[19]</sup>提出了一个可转移框架,该框架利用图表示学习和对抗性领域适应来学习域不变特征,然后进一步将GRU网络纳入模型以学习时间交通数据。Zhao等<sup>[20]</sup>提出了一个基于GCN和长短期记忆网络(LSTM, long short-term memory network)的领域对抗网络,用于解决数据稀缺问题的城市交通流量预测,并设置了域分类器来区分源域和目标域的交通数据,以帮助学习域不变特征。然而,上述研究中的域鉴别器/分类器在区分数据表示的原始域时,会出现梯度消失问题,造成预测模型的精度不高。

## 2 DATE方法设计

### 2.1 问题定义

为了对交通流进行预测,本文将城市中的道路连通性进行简化处理,道路视为无向边,道路的交点视为节点,通过将图 $G=(V,E,A)$ 表示一个城市 $C$ 的道路网,其中, $V=\{v_1,\dots,v_N\}$ 是一个包含 $N$ 个节点的集合( $|V|=N$ ),每个节点初始特征向量维度为 $d=16$ ,分别包括节点空间位置编码、历史交通均值特征、该区域主兴趣点(POI, point of interest)类型ID的嵌入向量等融合信息。 $E\subset V\times V$ 是边的集合, $A$ 是网络 $G$ 的邻接矩阵。这里,无向图 $G$ 表示区域间的交通联系, $N$ 表示图中的节点数量,需要注意的是无向图无法体现交通流的方向差异。

城市 $C$ 交通数据的时间范围又可以分为长度相

等不相邻的时间间隔  $t = \{1, \dots, T_C\}$ 。对于具体的交通数据, 为了和城市道路网对应的节点相关联, 本文将城市  $C$  中的每个节点  $v$  和一个向量  $\mathbf{x}_v = [\mathbf{x}_v^{(t)}]_{t=1}^{T_C}$  相关联, 该向量描述了时间间隔  $1 \sim T_C$  期间  $v$  的交通数据 (如出租车接载次数)。

除此之外, 本文还给出了基于时空关系构建多元信息的符号化定义, 包括道路连接图  $G_r^C$ , POI 信息图  $G_{V_{\text{vec}}}^C$  和人员流动图  $G_S^C$ 、 $G_D^C$ 。其具体定义如下。

1) 道路连接图  $G_r^C$ 。节点  $v_i \in V$  表示城市交通路网中每一个交叉路口或道路网格单元; 若 2 个节点  $(v_1, v_2)$  对应的区域之间存在高速公路或主干道连接, 则构建无向边  $(v_1, v_2) \in E$ 。边是无向的且不带权重, 表示连接关系的存在性, 节点之间构成道路连接图。为了展示城市道路网络结构中节点 (代表区域/交叉口) 之间如何通过无向边表示连接关系, 反映真实道路的空间邻接性, 本文给出了图 1 所示的道路连接示意图展示城市道路网络结构中节点连接情况。在图 1 中, 例如节点  $v_1$  与  $v_6$  之间的边用虚线表示, 这说明虽然 2 个节点在空间中相连接, 但区域之间并非高速公路或主干道, 因此在实验中不予考虑。

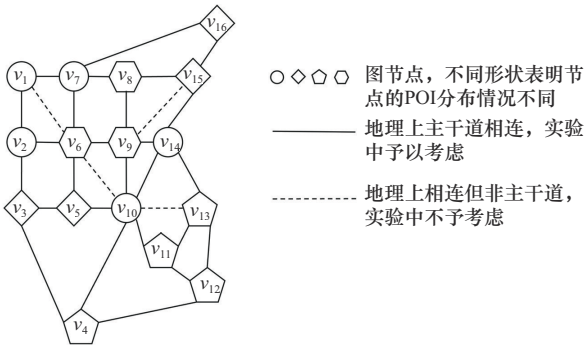


图 1 道路连接示意

2) POI 图  $G_{V_{\text{vec}}}^C$ 。由于 POI 分布相似的城市区域内节点表明城市功能相似, 为 POI 给予不同的类型编号, 给定节点  $v$ , 将其和与之余弦相似度前  $M$  的节点连接起来。

$$\text{POISim}(v, v') = \text{CosSim}(\text{Vec}_v, \text{Vec}_{v'}), v, v' \in C \quad (1)$$

其中,  $\text{Vec}_v$  是  $v$  的 POI 向量,  $\text{Vec}_v$  的第  $i$  维  $\text{Vec}_{(v,i)}$  表示节点  $v$  中第  $i$  类 POI 的数量,  $\text{CosSim}(\cdot)$  表示余弦相似度计算, 范围是  $[0, 1]$ , 本文选取与每个节

点相似度 Top- $k$  的节点建立边连接 ( $k=8$ )。

总体而言, 在 POI 图中, 每个节点都涵盖多个 POI 类别。以 20 个节点为例, 假设每个节点包含 14 个 POI 类别, 那么 POI 特征向量的维度即  $[20, 14]$ 。基于此, 在节点连接过程中, 需针对这些 POI 类别分别执行连接操作。如图 1 所示, 不同形状标识出各节点的 POI 分布情况。节点  $v_5$  与  $v_6$  尽管在地理上相邻, 但由于其 POI 相似度较低, 在 POI 图中并未建立连接; 相反,  $v_5$  却会与地理上并不相邻的  $v_{15}$  相连。

3) 人员流动图  $G_S^C$ 、 $G_D^C$ 。人员流动图是基于历史出行数据构建的, 节点表示所在城市区域功能, 边由人员迁徙或交通出行量决定。例如, 住宅区与写字楼之间的联系程度通常高于住宅区之间的联系。

给定人类从出发地到目的地的流动区域节点对  $M_C = (v_s, v_d)$ , 计算区域  $v, v' \in C$  之间的权重为  $w(v, v')$ , 并进一步计算节点  $v$  的出发地和目的地分布, 将其流动模式建模为

$$s(v_i | v) = \frac{w(v_i, v)}{\sum_{v'} w(v', v)}, d(v_i | v) = \frac{w(v, v_i)}{\sum_{v'} w(v, v')} \quad (2)$$

为了将相似度最高的  $M$  个节点连接起来, 本文选择每个节点  $v$  与源分布  $s$  中库尔贝克-莱布勒 (KL, Kullback-Leibler) 散度最高的节点, 形成  $G_S^C$ 。

在给出城市网络的定义之后, 针对跨城市转移进行的时空预测任务, 本文给出以下形式化问题定义。分别给定源城市  $S$  和目标城市  $T$ , 以及 2 个城市交通数据的形式化表示  $X_S = \{x_{vS}, vS \in S\}$ ,  $X_T = \{x_{vT}, vT \in T\}$ , 研究目标是利用目标数据和丰富的源数据学习一个预测模型  $N_\theta$ , 要求  $N_\theta$  满足以下要求。

$$\min L_\theta(X_T, \theta) = \sum_{v \in T, t=1}^{T_T} L(\hat{x}_v^{(t)}, x_v^{(t)}) \quad (3)$$

其中,  $\hat{x}_v^{(t)}$  和  $x_v^{(t)}$  是目标节点  $v$  在时间间隔  $t$  时的预测值和真实值,  $L$  是误差指标, 根据实际应用要求, 可以取平均绝对误差、均方根误差等。

## 2.2 DATE 方法总体框架

本文在目标城市算力和数据资源受限的问题场景下, 设计一种广泛适用的迁移学习模型。首先, 为了提高域适应过程中的跨域特征表示能力

和未知数据上的泛化性能，提出将多核均值最大差异方法应用于节点适应过程，同时构建一个特征映射网络置于特征提取网络后端，以进一步对齐特征分布；其次，为了应对计算资源受限的场景问题，通过在预测模型的网络结构中应用可分离式卷积和引入注意力机制的GRU来实现预测网络的轻量化。

DATE方法的总体框架如图2所示，主要部分由一个特征网络 $N_{\theta_f}$ 、一个加权网络 $N_{\theta_w}$ 和一个预测网络 $N_{\theta}$ 这三个部分组成。其中，特征网络用于捕捉城市间共享的空间特征，以发现共同的城市功能；加权网络则在学习过程中选择性地为源节点分配高权重，从而改善目标微调；通过预测网络 $N_{\theta}$ 来选择性地使用 $\lambda_{v_s}$ 进行训练，从而初始化目标微调。通过基于域适应的选择性源训练方法和微调操作，形成了从特征提取、重要性加权到最终预测的完整迁移学习。

### 2.3 多元信息图特征提取网络

为了提取源区域和目标区域的通用空间特征，本文使用特征网络去识别共享的城市功能并进行选择性源训练。为了更好地捕捉这些特征，DATE方法通过使用节点和边缘级域适应目标来训练特征网络，本文 $N_{\theta_f}$ 的具体实现受文献[21]启发，对多元信息融合添加特征映射层以进一步对齐特征分布。

特征网络的输入是将多元城市数据信息建模为图，

得到一个多元信息图 $\{G_{Vcc}^C = (C, E_v^C)\}$ ， $v \in \{r, p, s, d\}$ ， $C \in \{S, T\}$ （其中， $E_v^C$ 表示 $G_{Vcc}^C$ 的边集）， $v$ 是节点的集合，每个节点初始特征向量维度为 $d=16$ ，分别包括节点空间位置编码、历史交通均值特征、该区域主POI类型ID的嵌入向量等融合信息。然后，利用特征网络 $N_{\theta_f}$ 将 $G_{Vcc}^C$ 投影为节点空间特征 $\Psi_C \in R^{|C| \times d_{emb}}$ ， $C \in \{S, T\}$ ，其中 $d_{emb}$ 是输出节点特征的维度。

特征网络的重要功能是融合城市区域功能、环境因素以及人类活动等多个视图数据，从多个维度的数据中学习具有代表性的城市节点特征表示，提升对城市区域功能中节点关联的理解和下游任务的性能。因此，本文设计的特征网络将以单信息表示学习为基础，采用联合学习模块进行特征融合以学习综合嵌入表示。

#### 2.3.1 单信息表示学习

单信息表示学习利用单一视图来学习区域表征。通过给定单一区域相关性，并将相关性信息构建成图后，采用图注意力网络（GAT, graph attention network）来学习节点的表征。GAT利用注意力机制来处理图结构数据，通过向邻近节点传播信息来更新节点的表示，而邻近节点的权重使用注意力机制进行自动学习。

具体来说，给定输入节点特征 $h = \{h_1, h_2, \dots, h_n\}$ ，

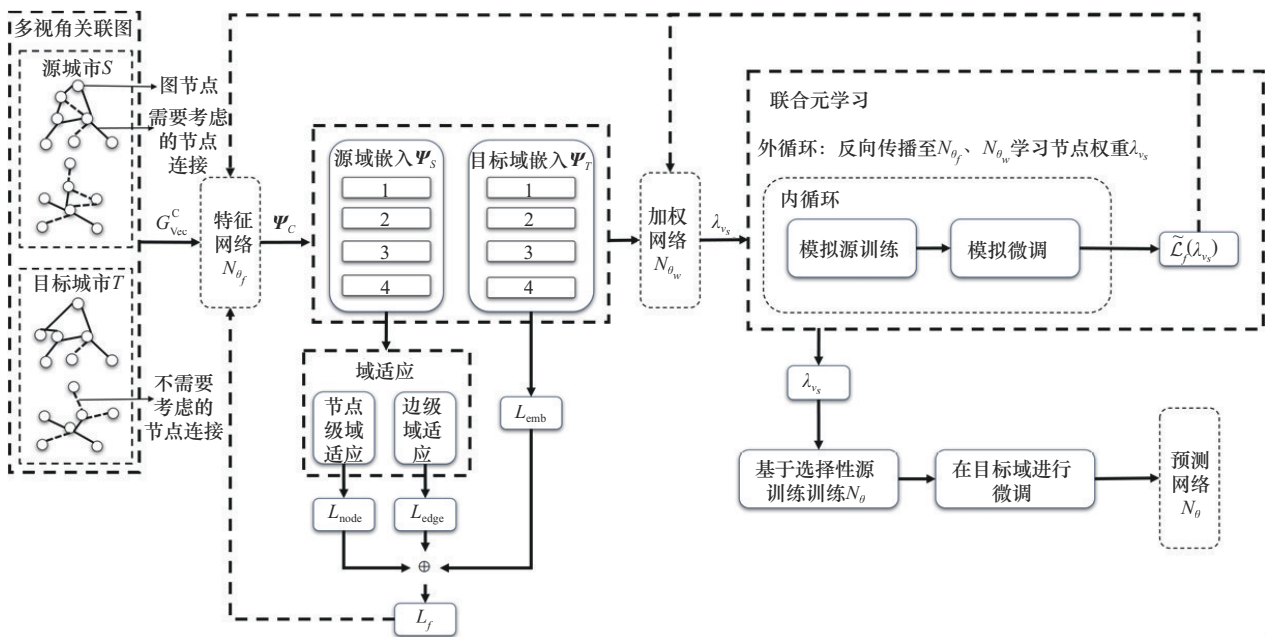


图2 DATE方法的总体框架

$\mathbf{h}_i \in \mathbf{R}^F$ ,  $\mathbf{h}_i$ 是第*i*个区域节点的向量,  $F$ 是该向量特征的维数。GAT层通过下列步骤更新节点表示。

$$e_{ij} = \exp\left(\text{ReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]\right)\right)$$

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}$$

$$\mathbf{h}'_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}\mathbf{h}_j\right) \quad (4)$$

其中,  $\mathbf{W}$ 与 $\mathbf{a}$ 均为可学习的参数矩阵,  $\parallel$ 是连接操作。为了提高特征表达性能, 在每个GAT层应用多头注意力机制。在具体操作中, 本文将2个GAT层堆叠在一起作为一个GAT块, 同时在图 $G_r^C$ 、 $G_{\text{Vcc}}^C$ 、 $G_S^C$ 和 $G_D^C$ 上应用GAT块, 并将每个视图的输出节点表示分别记为 $\Psi_r$ 、 $\Psi_{\text{Vcc}}$ 、 $\Psi_S$ 和 $\Psi_D$ 。

### 2.3.2 联合学习模块

为了解决单信息表示学习仅仅学习到单一区域节点的相关性问题, 本文使用一个联合学习模块进行学习, 在学习过程中实现多元信息之间的协作, 并有效地融合多元信息表征。所使用的框架由2个部分组成, 一部分是自注意力层, 能够实现所有信息之间的信息共享; 另一部分是融合层, 通过学习自适应权重来融合多元信息表征。从多个维度观察到的各区域节点相关性具有多样性和高度相关性。以人员流动相关性和区域POI相关性为例, 在早高峰时段, 大多数人从有居住POI的区域向学校或者商业POI的区域流动, 而在晚高峰时段, 人们则向相反的方向流动。这种高相关性意味着将多元的信息整合在一起, 可以增强每个维度信息的学习过程。

为了进一步增强联合学习模块的可理解性, 本文设计了多元信息联合学习模块的详细流程, 如图3所示。在联合学习模块中, 输入来自多个维度的区域结构信息(如道路连接图、POI图、人员流动图), 分别形成3个结构化特征向量, 作为模型的多元视图输入。然后利用自注意力机制计算视图间的注意力权重矩阵, 并将不同视图信息映射为共享的全局向量, 提升视图间的信息交互能力。每个特征经过查询/键(Query/Key)映射后执行跨视图注意力加权, 获得表征全局重要性的表示向量。最后, 各视图生成的全局表示通过线性融合层进行加权融合。该模块能够有效整合来自城市多维信息源

的异质表征, 并通过端到端的注意力机制实现信息间的深度协同, 为跨域预测任务提供稳定且泛化能力强的输入特征。

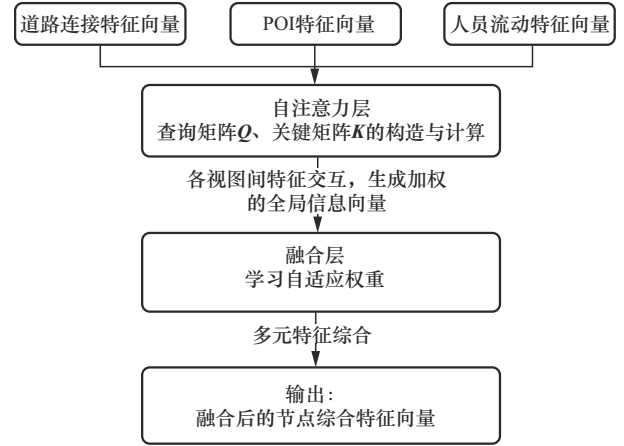


图3 多元信息联合学习模块的详细流程

具体来说, 基于上述场景, 本文采用自注意力机制在不同信息的表征中传播知识。给定来自*M*个不同数据的表征为 $\{\Psi_i \in \mathbf{R}^{n \times d}\}_{i=1}^M$ , 其中*d*为嵌入大小。对于每个表征 $\Psi_i$ , 将其关联为一个键矩阵 $\mathbf{K}_i \in \mathbf{R}^{n \times k}$ 和一个查询矩阵 $\mathbf{Q}_i \in \mathbf{R}^{n \times k}$

$$\mathbf{K}_i = \Psi_i \mathbf{W}_k, \mathbf{Q}_i = \Psi_i \mathbf{W}_q \quad (5)$$

其中,  $\mathbf{W}_k$ 和 $\mathbf{W}_q$ 是可学习的参数矩阵。对于每个维度信息, 在多元信息之间传播信息的方式为

$$[\mathbf{A}_i]_{i=1}^M = \text{softmax}\left(\left[\frac{\mathbf{Q}_i \mathbf{K}_i^\top}{\sqrt{k}}\right]_{i=1}^M\right), \hat{\Psi}_i = \sum_{i=1}^M \mathbf{A}_i \Psi_i \quad (6)$$

其中,  $\mathbf{A}_i$ 是注意力权重矩阵,  $\hat{\Psi}_i$ 可以视为第*i*个维度的相关全局信息。为了将这些信息纳入学习过程, 计算

$$\Psi'_i = \alpha \hat{\Psi}_i + (1 - \alpha) \Psi_i, 0 \leq \alpha \leq 1 \quad (7)$$

其中,  $\Psi'_i$ 是第*i*个视图的全局信息表示, 可以作为全局信息的权重。

如果想在各种应用中采用学习到的节点特征表示, 需要一种能保留多元信息相关性的综合节点特征表示法。为了融合多元信息表征, 在联合学习中还需要使用一个融合层, 它可以学习不同维度信息的自适应权重, 具体为

$$\Psi = \sum_i^M w_i \Psi'_i, w_i = \sigma(\Psi_i \mathbf{W}_f + \mathbf{b}_f) \quad (8)$$

其中,  $w_i$  是第  $i$  个维度信息的权重, 由单个多层感知机 (MLP, multilayer perceptron) 网络以第  $i$  个嵌入数据为输入进行学习,  $W_f$  与  $b_f$  均为网络的可学习参数矩阵,  $\sigma$  是 sigmoid 函数。

为了实现多元信息融合层的学习, 需要  $\Psi$  参与每个维度信息的学习目标, 因此可以将每个维度信息的表示更新为

$$\bar{\Psi}_i = \frac{\hat{\Psi}_i + \Psi}{2} \quad (9)$$

通过将基础模型的输出与输入进行联合学习, 可以得到在学习节点相关性更新后的节点嵌入  $\bar{\Psi}_i$ 、 $\bar{\Psi}_{Vcc}$ 、 $\bar{\Psi}_S$  和  $\bar{\Psi}_D$ 。

### 2.3.3 综合嵌入表示特征映射

针对本文特征网络的下游任务, 还需要一个融合后的城市区域节点特征, 因此利用式(8)的结果  $\Psi$  作为融合后的综合嵌入表示, 记为  $\Psi_f$ 。为了使融合后的城市区域节点特征更好地衔接下游任务, 本文在更新图嵌入表示的训练过程中, 在特征网络获取综合区域节点特征后添加特征映射层。该层由 MLP 和批量归一化组成, 用于深度处理从多元信息联合学习模块获得的的城市区域节点综合特征  $\Psi_f$ , 从而在调整特征维度的同时, 通过非线性变换增强特征的表达力, 使特征表示更符合后续任务的需求。

### 2.3.4 基于多核 MMD 度量的节点级适应

作为跨领域适应的重要技术, 节点级域适应专注于调整不同领域内样本特征的分布, 以便模型能够学习到领域间共享的表示。DATE 方法通过使用多核均值最大差异, 利用线性组合能够更好地刻画源域和目标域之间的共同特征空间, 以此来更全面地捕捉数据的复杂结构, 允许模型在不同尺度和特征上进行节点级适应, 实现迁移学习。它的源域和目标域的区域特征通过式(10)来描述最大均值差异 (MMD, maximum mean discrepancy) 方法。

$$L_{node} = \text{MMD}(\phi_S, \phi_T) \quad (10)$$

其中,  $\phi_S$  和  $\phi_T$  分别代表源城市和目标城市区域的特征矩阵, 每个区域的特征向量通过特征网络映射得到。通过最小化源城市  $S$  和目标城市  $T$  的区域特征向量  $\phi_S$  和  $\phi_T$  之间的差异  $L_{node}$ , 促使模型学习领域不变特征。直观地说,  $L_{node}$  对齐了 2 个城市的节点 (即区域) 特征分布。多核 MMD 通过引入多个核函数的线性组合扩展了传统 MMD 在特征分布对

齐中的表达能力。不同核函数具有不同的感受野与分布刻画能力, 当多个核函数通过凸组合被引入后, 可在更广泛的特征空间尺度上度量源域与目标域间的嵌入分布差异, 从而提升域对齐的鲁棒性和适应性。

### 2.3.5 特征网络损失定义

本文将模型训练的目标定义为减少区域节点关系重建和源城市、目的地预测 2 个部分的损失。在区域节点关系重建部分, 为了让学习到的节点嵌入根据不同的区域属性保留区域节点相似性, 根据相应的内嵌重建区域节点相关性。使用  $e_p^i$  来表示视图  $p$  中第  $i$  个区域的嵌入表示, 学习目标定义为

$$L_p = \sum_{ij} (C_p^{ij} - e_p^i \top e_p^j)^2 \quad (11)$$

其中,  $C_p^{ij}$  表示 POI 的相关性矩阵,  $e_p^i$  与  $e_p^j$  表示对应区域节点的嵌入向量。该计算过程表示将实际相关性得分与嵌入向量内积的差距作为损失求和, 从而得到整个 POI 属性相关性重构任务的目标函数  $L_p$ 。

在源城市、目的地预测部分, 需要根据节点表示  $\tilde{\Psi}_s = \{e_s^i\}_{i=1}^n$  和  $\tilde{\Psi}_d = \{e_d^i\}_{i=1}^n$  来从源节点预测目的节点, 或者从目的节点来预测源节点。给定源节点  $v_i$ , 可以对目的节点  $v_j$  的分布进行如下建模。

$$\hat{p}_s(v_j | v_i) = \frac{\exp(e_s^i \top e_d^j)}{\sum_j \exp(e_s^i \top e_d^j)} \quad (12)$$

相应地, 将给定目的节点  $v_j$  的源节点  $v_i$  的分布情况建模为

$$\hat{p}_d(v_j | v_i) = \frac{\exp(e_d^i \top e_s^j)}{\sum_j \exp(e_d^i \top e_s^j)} \quad (13)$$

然后, 给定包含现实世界中源节点和目的节点对的人类流动数据集  $M$ , 学习目标函数被定义为预测分布的负对数似然。形式上, 目标函数可以表示为

$$L_{mob} = \sum_{(v_i, v_j) \in M} -\ln \hat{p}_s(v_j | v_i) - \ln \hat{p}_d(v_i | v_j) \quad (14)$$

最终, 特征网络的损失被定义为

$$L_{emb} = L_{mob} + L_p \quad (15)$$

## 2.4 自适应加权网络

加权网络组件的设计目标是让模型可以自适应地对源节点进行重新加权。例如,如果目标城市交通顺畅,那么拥堵严重的源路段对于学习的结果将是不利的,因此应降低这类路段对应的节点的权重。为了实现这一目标,本文通过源-目标联合元学习来训练特征网络 $N_{\theta_f}$ 和加权网络 $N_{\theta_w}$ 。通过在联合元学习的内循环中进行模拟源训练和目标微调,使2个网络都能学会对源节点重新加权,从而改善目标微调。

加权网络 $N_{\theta_w}$ 需要将节点空间特征 $\Psi_S$ 、 $\Psi_T$ 转换为源节点权重 $\lambda_{v_s} \geq 0$ 。具体的加权网络计算流程是将源节点嵌入向量 $\mathbf{h}_i$ 与目标域上下文向量 $\boldsymbol{\mu}$ 映射到统一嵌入空间后,计算两者间的相似性作为权重分配依据。权重计算式为

$$w_i = \tanh(\mathbf{h}_i^T \cdot \boldsymbol{\mu}) \quad (16)$$

其中, $\mathbf{h}_i$ 表示经过变换后的第 $i$ 个源节点嵌入向量, $\boldsymbol{\mu}$ 为目标域节点表示的均值池化向量, $T$ 为转置操作, $\tanh(\cdot)$ 是非线性压缩函数,用于稳定梯度并抑制异常激活值。权重 $w_i$ 的输出范围在 $(-1,1)$ ,体现源节点与目标域语义的匹配程度。

为了进行权重转换,首先,进行嵌入表示的变换,从而将原始的高维嵌入向量映射到一个更合适的新空间中,即用 $N_{\theta_w}$ 对 $\Psi_S$ 和 $\Psi_T$ 进行变换,得到变换后的特征。

$$\hat{\Psi}_{v_s} = N_{\theta_w}(\Psi_{v_s}), \hat{\Psi}_{v_t} = N_{\theta_w}(\Psi_{v_t}) \quad (17)$$

其次,进行目标域的上下文向量计算,对目标域中处于有效位置的数据进行均值池化,从而得到目标城市特征 $\hat{\Psi}_T$ ,该特征能够较好地概括整个目标域的特征。最后,进行源域数据匹配度的计算,通过向量内积和 $\tanh$ 函数激活得到源节点权重 $\lambda_{v_s}$ 。

$$\lambda_{v_s} = \max\left(\tanh\left(\hat{\Psi}_{v_s} \hat{\Psi}_{v_t}\right), 0\right) \quad (18)$$

其中, $\hat{\Psi}_{v_c}$ 表示节点 $v_c \in C, C \in \{S, T\}$ 的特征。

在网络结构的实现上,该加权网络通过一个2层MLP来转换源域和目标域的嵌入表示,在第一层中进行特征的降维,第二层保持维度并给出变换后的特征。由此便可将源域和目标域的数据转换为一个统一的嵌入空间,空间中的数据表示将更有利于模型自适应地对源节点重新加权。

## 2.5 轻量化时空序列预测网络

在整个迁移学习框架中,预测网络 $N_{\theta}$ 将使用权重 $\lambda_{v_s}$ 对源城市进行选择训练,从而达到初始化目标微调的目的。具体来说,将源数据 $X_S$ 、目标数据 $X_T$ 和学习到的 $\lambda_{v_s}$ 作为输入,并利用权重 $\lambda_{v_s}$ 进行源训练,权重将根据目标数据 $X_T$ 进行微调。

在预测网络的构建上,本文针对复杂的时空序列预测任务中的交通流量预测,设计了一个高度集成时空特征提取与序列建模机制的轻量化神经网络模型。该模型通过整合深度可分离卷积和门控递归单元等高效组件,优化计算效率并保留模型的表达能力,同时引入残差结构以确保深度网络的稳定训练,从而增强对长期依赖关系的学习。模型的轻量化旨在减少计算负担,使模型在资源受限的情况下能够适合大规模或实时预测场景。

模型的输入和处理流程如下。首先,设置预测范围 $\tau = 6$ ,即输入数据由前6个时间区间的观测数据组成,预测输出为下一个时间区间的交通流数据。然后,在逐个分析每个时间滞后阶段时,将数据按不同时间滞后间隔切分,确保每次只专注处理一个特定滞后时段的信息。这些切分后的数据会经历由深度可分离卷积和残差网络层组成的处理流程,以高效提取空间特征。经过这一系列提炼后,数据会被有序堆叠并调整其维度,使时间滞后步数转化为序列维度的一部分,同时原始批次大小与空间特征维度被整合为一个新的统一维度,变换后的向量作为后续时序处理单元的输入,最终输出预测值。

### 2.5.1 深度可分离式卷积核应用

受到文献[22]的启发,本文将深度可分离卷积应用于时空序列预测任务的神经网络,在资源受限环境下有效提取空间特征的同时提高计算效率。深度可分离卷积将传统的卷积操作分解为深度卷积(DC, depthwise convolution)和逐点卷积(PC, pointwise convolution)2个部分。其工作流程是在特征维度上进行分组卷积,对每个通道进行独立的深度卷积,并在输出前使用一个 $1 \times 1$ 的逐点卷积将所有通道进行聚合,得到最终的输出特征图,从而达到减小计算量、提升计算效率的目的。

在使用深度卷积处理交通流数据过程中,普通卷积仅仅能够在空间维度上进行分离,增加网络宽度并提取丰富特征。深度卷积的卷积核为单通道模

式, 能够对每个输入通道进行独立卷积, 得到与输入特征图通道数一致的输出特征图。在提取空间特征过程中假设卷积核尺寸为  $CS_K \times CS_K \times 1$ , 而卷积核个数与输入通道数相同即为  $M$ , 则参数量为  $CS_K \times CS_K \times M$ , 由于每个卷积核都要做  $CS_F \times CS_F$  次乘法运算, 计算量为  $CS_K \times CS_K \times M \times CS_F \times CS_F$ 。

在将深度卷积应用到交通流预测的过程中, 输入特征图通道数、卷积核个数和输出特征图个数都是相等的, 这样会导致输出的特征图个数过少, 从而影响信息的有效性。因此, 需要对流量特征进行逐点卷积, 将  $1 \times 1$  的卷积核进行升维, 所有通道合并为输出特征图。逐点卷积的卷积核尺寸为  $1 \times 1 \times M$ , 卷积核个数为  $N$ , 因此参数量为  $M \times N$ , 而每个卷积核都要做  $CS_F \times CS_F$  次乘法运算, 计算量为  $M \times N \times CS_F \times CS_F$ 。

综合以上 2 步的计算资源消耗, 当采用深度可分离卷积核来捕获空间特征时的总参数量  $P_{DSC} = CS_K \times CS_K \times M + M \times N$ , 总计算量  $C_{DSC} = P_{DSC} \times CS_F \times CS_F$ 。相比传统的不进行进一步处理, 而是直接统一提取空间特征的普通卷积方法, 其参数量  $P_{TC} = CS_K \times CS_K \times M \times N$ , 则计算量  $C_{TC} = P_{TC} \times CS_F \times CS_F$ 。综上, 最终两者计算量的比值为

$$\frac{C_{DSC}}{C_{TC}} = \frac{P_{DSC} \times CS_F \times CS_F}{P_{TC} \times CS_F \times CS_F} = \frac{1}{N} + \frac{1}{CS_K^2} \quad (19)$$

在假定使用  $3 \times 3$  的卷积核情况下, 若通道数  $N$  较多, 则前一项比值趋近于 0。由此可见, 使用深度可分离卷积的参数量和计算量大幅下降, 从而在资源受限的情况下, 减少模型的计算量。

### 2.5.2 残差层构建

在应用深度可分离卷积之后, DATE 方法能够获得包含空间特征的向量  $\mathbf{z}$ , 本文通过由一系列残差单元组成的残差层来增强学习能力。在组成上每个残差单元由带有批量归一化 (BN, batch normalization) 和 ReLU 激活函数的两层卷积组成, 然后将这些层的输出与未经处理的输入相加, 从而确保网络能够学习残差 (即新特征与原特征之差) 而非绝对特征, 进而缓解 DATE 方法中的深度网络训练存在的梯度消失问题, 促进了更深层次特征的学习。除此之外, 本文还使用了一种“跳连”方法, 通过“跳连”, 输入  $\mathbf{z}$  在经过一些变换后, 其结果会与原始输入相加来保证即使在深层网络中也能顺

利传播梯度, 从而帮助解决梯度消失问题。最后每个残差块的输出  $\mathbf{z}'$  是其内部卷积操作的结果与输入  $\mathbf{z}$  的和, 即

$$\mathbf{z}' = \mathbf{z} + F(\mathbf{z}) \quad (20)$$

其中,  $F(\mathbf{z})$  代表了网络中从  $\mathbf{z}$  到输出的残差或额外贡献。通过这样的设计, 网络可以更容易地学习到恒等映射, 即  $F(\mathbf{z}) = 0$ , 这使得即使在极深的网络中也能保证信息的顺畅流动, 避免了梯度消失或爆炸的问题, 从而促进训练的稳定性与深度网络的优化。在具体实现上, 本文的残差层由 3 个输入和输出通道大小为 64, 卷积核大小定义为  $3 \times 3$  的残差层构成。

### 2.5.3 注意力机制的门控循环单元方法

为了衡量不同时刻交通流的流量特征对预测流量的影响, DATE 方法通过引入注意力机制来生成含有注意力概率分布的交通特征表示, 也就是在不同时间步的隐藏状态中分配不同的权重, 从而强调对最终预测更重要的时间步信息。

首先, DATE 方法使用一个全连接层和激活函数 (这里使用的是 Tanh 函数) 从隐藏状态中提取特征, 生成分数  $\mathbf{e}_t$ 。

$$\mathbf{e}_t = \tanh(\mathbf{W}_a \mathbf{h}_t + \mathbf{b}_a) \quad (21)$$

其中,  $\mathbf{W}_a$  和  $\mathbf{b}_a$  是注意力层的权重矩阵和偏置项。

其次, 通过一个线性变换和 softmax 函数来归一化这些分数, 得到注意力权重  $\alpha_t$ 。

$$\alpha_t = \frac{\exp(\mathbf{v}^T \mathbf{e}_t)}{\sum_{k=1}^T \exp(\mathbf{v}^T \mathbf{e}_k)} \quad (22)$$

其中,  $\mathbf{v}$  是另一个权重向量。

最后, 利用这些权重对隐藏状态进行加权求和, 得到加权后的全局特征  $\mathbf{h}_w$ 。

$$\mathbf{h}_w = \sum_{t=1}^T \alpha_t \mathbf{h}_t \quad (23)$$

DATE 方法的最后阶段是将输出的特征  $\mathbf{h}_w$  通过一个全连接层将学到的高维时空特征转化为单个预测值, 即最终的预测值。此外, 若交通流量预测任务需要输出值在 0~1 (如概率或比率), 可以使用 sigmoid 函数对全连接层的输出进行约束。

### 2.5.4 模型训练过程

DATE 方法对应的迁移学习模型的训练过程整体按照上文所述的选择性源训练方法和微调进行。

在选择性源训练的回合中,通过双级域适应和联合元学习策略来优化特征网络 $N_{\theta_f}$ 和加权网络 $N_{\theta_w}$ ,以获得当前节点权重 $\lambda_{v_s}$ ,用于在一个回合的源数据上选择性地训练预测模型 $\theta$ ,权重 $\lambda_{v_s}$ 会在训练过程中不断更新,最后在目标域上执行微调。

### 3 实验评估

本节将通过多组实验来验证DATE方法的有效性,首先对实验数据集、对比算法、评估指标和参数设置进行描述,在此基础上,为验证所设计的迁移学习框架的有效性,在纽约(NY, New York)、芝加哥(CHI, Chicago)、华盛顿(DC, Washington DC)的3个城市的出租车数据集上进行实验,通过对比实验验证选择性迁移学习方法的优势和轻量化方面的优势,最后进行消融实验,并对结果进行分析。

#### 3.1 数据集与评价指标

##### 3.1.1 数据集

本文采用交通流量预测中常用的纽约、芝加哥、华盛顿的出租车数据作为数据集<sup>[18,23]</sup>。数据收集的时间范围为一年,在节点数上,纽约有458个节点,华盛顿有207个节点,芝加哥有351个节点;在时间划分上,将每个城市的数据集按照1h的时间间隔进行划分。从数据集中可以看出,不同城市节点空间密度的差异,纽约的网格密度显著高于华盛顿,中心城区的节点间距更紧密,且流量极值主要集中在曼哈顿岛一带;而芝加哥则呈现双核心(环区与机场区)布局。相比之下,华盛顿的出租车出行主要集中在商业核心区及国会山区域,城市边缘区域流量稀疏,且晚间低频区域显著增多。将数据较多的纽约和芝加哥作为源城市,并选取华盛顿作为目标城市。对于源城市,将最后2个月用于测试,之前2个月用于验证,其余8个月用于训练。对于目标城市,采用相同的测试和验证数据设置,并分别使用验证前最后30天、7天和3天的数据进行训练,以评估不同目标数据量下模型的性能。构建多元图的数据采用从开源地图数据项目OpenStreetMap中收集的POI和道路数据,以及从出租车数据集中提取人类流动性数据。

需要注意的是在交通流预测中,模型的泛化性从根本上受到数据集之间分布的影响,如城市布局差异、交通动态和传感器覆盖率。NY、CHI和DC

这样的出租车数据集提供了有价值的基准,而且它们代表了一组具有相对同质交通模式的特定城市环境。然而,滴滴数据集(DiDi)和北京出租车数据集(TaxiBJ)在城市环境、道路网络拓扑、交通行为、流动模式以及时间交通波动等呈现出更多具有不同特征的异构环境。当应用基于一个城市数据训练的模型来预测另一个城市的交通时,这些差异带来了挑战。如果不考虑城市的特定特征,模型的泛化能力可能会受到损害,从而导致性能的负转移。

##### 3.1.2 评估指标

本文采用均方根误差(RMSE, root mean square error)和平均绝对误差(MAE, mean absolute error)作为交通流预测准确率的评价指标。

1)均方根误差能够提供关于数据集中误差的整体大小以及这些误差的分布情况的信息, RMSE的定义为

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (24)$$

其中, $y_i$ 表示第*t*个观测的实际值, $\hat{y}_i$ 是对应的预测值, $n$ 是样本数量。RMSE通过计算预测值与实际值之差平方和的均值后再开方得到,因此对较大误差给予更大惩罚。这使得RMSE在误差较大时比MAE增长更快,适用于对大误差更为敏感的应用场景。

2)平均绝对误差是预测值与实际值之差的绝对值的平均数,它直接反映了预测误差的平均大小,不考虑误差的方向,MAE的定义为

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |X_i - \hat{X}_i| \quad (25)$$

相比RMSE,MAE对异常值的敏感度较低,因为其计算过程中没有平方操作。因此,在处理含有极端值的数据集时,MAE更稳健。MAE易于理解和解释,尤其适用于需要直观了解平均偏差的场合。

### 3.2 实验环境与参数设置

#### 3.2.1 实验环境

本次实验所设计的模型均在Pytorch框架下实现,在训练过程中通过使用计算统一设备架构(CUDA, compute unified device architecture)和利用图形处理器(GPU, graphics processing unit)的计算能力进行加速训练,具体用到的硬件及软件配置如表1所示。

表1 部分参数设置

配置名称	型号或版本
操作系统	Windows
GPU	NVIDIA GeForce GTX 1660 Ti
CPU	AMD Ryzen 7 4800H with Radeon Graphics 2.90 GHz
内存	6.0 GB
编程语言	Python 3.8
深度学习框架	Torch 1.13.0
CUDA	11.1

### 3.2.2 参数设置

本文 DATE 方法的部分参数设置如表 2 所示。具体为：源训练的回合数设置为 100，微调的回合数设置为 80，模型的批处理数设置为 16，学习率是  $2 \times 10^{-3}$ ，权重衰减系数是  $5 \times 10^{-5}$ ，元学习内部和外部循环的学习率分别设置为  $5 \times 10^{-5}$  和  $1 \times 10^{-4}$ 。此外，模型的源域权重移动平均数则用于在训练过程中计算源域权重，对源域知识在模型训练过程中的平滑程度和稳定性产生重要影响，本文将源域权重的移动平均数设置为 0.6。本文还使用 Adam 优化器执行参数更新。

表2 部分参数设置

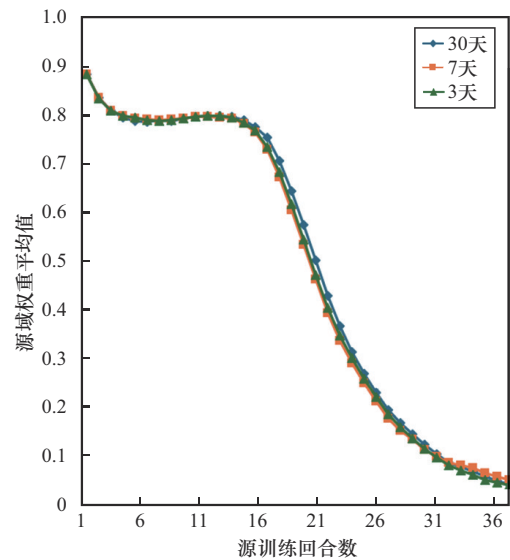
参数名称	值
源训练回合数	100
微调回合数	80
批处理数	16
学习率	$2 \times 10^{-3}$
权重衰减	$5 \times 10^{-5}$
元学习内部循环学习率	$5 \times 10^{-5}$
元学习外部循环学习率	$1 \times 10^{-4}$
节点级域适应权重	2
边级域适应权重	2
源域权重的移动平均数	0.6

## 3.3 实验结果与分析

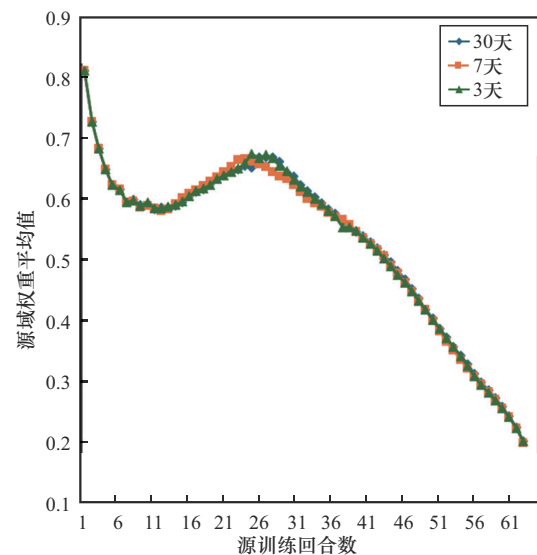
### 3.3.1 训练过程分析

为了进一步增强模型的可信度与解释力，本文对 DATE 方法训练过程的动态行为进行了可视化分

析。图 4 给出了选择性源训练过程中源域权重的变化。从图 4 可以看到，2 个迁移任务在中间的一些回合中的权重均出现了一定程度的回升情况，这是因为在训练中 DATE 方法引入的训练策略存在课程学习效应，即因为采用渐近式训练策略，模型会在特定阶段重新引入源域数据以巩固基础特征，导致权重回升。但随着选择性源训练的进行，2 个迁移任务中源域的平均权重最终在经过一定回合数的训练后均呈现逐渐下降的趋势，这表明在源训练过程中，模型逐渐学习到与目标城市无关的来源知识，从而导致权重下降，证明了模型具备良好的知识学习能力和稳定性。



(a) 芝加哥到华盛顿迁移任务中的权重变化



(b) 纽约到华盛顿迁移任务中的权重变化

图 4 训练过程中源域权重的变化

### 3.3.2 对比实验

为了检测DATE方法的整体迁移性能, 本文将该方法和其他迁移学习方法进行对比实验。本文不仅给出不同目标数据规模下的迁移性能, 还对不使用迁移学习方法, 仅使用预测网络的情况下在目标城市的预测结果进行性能比较。本文的迁移任务分别从纽约迁移到华盛顿和从芝加哥迁移到华盛顿, 使用的数据样本为出租车交通数据。参与对比的方法包括经典的微调(Fine-Tuning)迁移方法、Wang等<sup>[23]</sup>提出的跨城市区域迁移学习框架(RegionTrans, region-based transfer learning framework), RegionTrans通过区域级别的匹配和卷积LSTM实现跨城市的时空特征对齐, 适用于区域相似性的目标城市, 能有效解决城市间分布差异问题。Yao等<sup>[17]</sup>提出的时空元学习(MetaST, meta-learning spatial-temporal)采用自适应迁移策略, 能够在不同城市间进行知识迁移。Wang等<sup>[24]</sup>提出的时空动态对抗自适应网络(ST-DAAN, spatial-temporal dynamic adversarial adaptation network)通过结合深度域适应和全局空间注意力机制, 能在异构城市间实现有效的知识迁移。Chen等<sup>[25]</sup>提出的基于交通流量矩阵的图卷积与注意力机制(TFM-GCAM, traffic flow matrix-based graph convolution and attention mechanism)(对比的模型为经过迁移处理的TFM-GCAM(TTEM-GCAM, transferable TFM-GCAM)), TFM-GCAM结合图神经网络和交通引导的全局注意力机制, 能高效捕捉复杂时空依赖关系, 适用于动态交通流的预测任务。Sun等<sup>[26]</sup>提出快速动态时间图卷积网络(FD-TGCN, fast and dynamic temporal graph convolution network)(对比的模型为经过迁移处理的FD-TGCN(TFD-TGCN, transferable FD-TGCN)), FD-TGCN利用快速时间卷积网络和动态图卷积网络, 提升实时性和计算效率。Ganin等<sup>[27]</sup>提出域对抗神经网络(DANN, domain-adversarial neural network)<sup>[27]</sup>(对比的模型为与域适应的可迁移交通流预测模型相结合的DATE-DANN), DANN利用训练数据和测试数据来自相似但不同的分布特征实现高效迁移。

表3为纽约到华盛顿迁移任务的对比结果。为了真实反映模型的性能, 本文的性能指标包括最大提升比例和最小提升比例, 以及性能提升的平均值。由表3可知, 在使用迁移学习的情况下, DATE方法

在所有目标时间窗口(特别是短期预测)下优于基准方法, 与表现最差和最佳的基准方法相比, DATE方法分别实现了5.26%~8.81%(最小提升)和9.70%~15.55%(最大提升)的改善, 平均提升效果稳定在7.54%~12.76%。特别地, 当目标数据仅为7天和3天时, DATE方法展现了更加显著的性能, 证明了在极度有限数据资源下, DATE方法相比其他方法具有更高的适应性和效率, 这是因为DATE方法通过设计多元信息联合学习模块充分挖掘了影响交通流的多维因素(如道路结构、功能区域、人口流动), 比只考虑单一信息的基线方法拥有更全面的视角, 提升了预测精度。

表3 纽约到华盛顿迁移任务的对比结果

方法	30天		7天		3天	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Fine-Tuning	4.289	1.517	4.497	1.614	4.756	1.765
RegionTrans	4.185	1.511	4.494	1.704	4.86	1.856
MetaST	4.171	1.489	4.50	1.633	4.753	1.750
ST-DAAN	4.325	1.522	4.537	1.590	4.982	1.729
预测网络	4.925	1.775	5.301	1.80	5.248	1.839
TTFM-GCAM	4.623	1.553	4.764	1.733	4.801	1.834
TFD-TGCN	4.215	1.533	4.523	1.694	4.756	1.755
DATE-DANN	4.436	1.401	4.62	1.683	4.815	1.803
DATE方法	<b>3.957</b>	<b>1.331</b>	<b>4.211</b>	<b>1.466</b>	<b>4.368</b>	<b>1.612</b>
最大提升比例	12.10%	15.18%	9.70%	15.55%	14.04%	15.14%
最小提升比例	5.40%	5.26%	6.72%	8.46%	8.81%	7.26%
平均提升比例	7.91%	12.36%	7.54%	12.76%	10.35%	10.20%

表4为芝加哥到华盛顿迁移任务的对比结果。本文把方法的各部分提升比例的平均数作为一个综合提升的衡量标准, DATE方法在芝加哥到华盛顿的迁移任务中有9.03%的综合提升效果, 最大提升比例达到13.82%~25.04%, 最小提升也有5.00%~8.81%, 平均提升效果为8.67%~13.09%。这表明无论源城市与目标城市之间的地理或文化差异如何, 本文所实现的迁移学习框架都能有效地促进知识迁移, 并在数据量较少时尤为有效。这是因为DATE方法能够借助源城市中类似模式的学习去更好地预测, 从而提高预测精度。

表 4 芝加哥到华盛顿迁移任务的对比结果

方法	30天		7天		3天	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Fine-Tuning	4.222	1.500	4.434	1.614	4.646	1.742
RegionTrans	4.236	1.632	4.424	1.780	4.710	1.956
MetaST	4.150	1.477	4.428	1.646	4.697	1.750
ST-DAAN	4.326	1.511	4.486	1.595	4.786	1.721
预测网络	4.925	1.774	5.301	1.300	5.249	1.840
TTFM-GCAM	4.610	1.521	4.655	1.634	4.766	1.754
TFD-TGCN	4.542	1.632	4.574	1.685	4.783	1.768
DATE-DANN	4.453	1.421	4.703	1.649	4.833	1.741
DATE方法	<b>3.929</b>	<b>1.393</b>	<b>4.118</b>	<b>1.474</b>	<b>4.311</b>	<b>1.583</b>
最大提升比例	16.83%	25.04%	13.82%	18.56%	14.04%	15.14%
最小提升比例	5.40%	5.00%	6.72%	7.80%	8.81%	7.26%
平均提升比例	9.80%	13.09%	8.67%	12.43%	10.54%	10.20%

此外，为了验证 DATE 方法在轻量化处理方面的性能，本文针对预测网络，将可分离式卷积核替换为传统卷积核，并分别将 GRU 替换为 LSTM 与双向长短期记忆网络 (BiLSTM, bidirectional long short-term memory network) 进行性能实验，实验数据设定与纽约到华盛顿的迁移任务相同，并将其模型参数和计算量进行分析，实验结果如表 5 所示。其中，模型 1 表示使用本文轻量化处理的预测网络迁移学习模型；模型 2 表示使用模块替换为传统卷积核和 LSTM 的预测网络模型；模型 3 表示使用模块替换为传统卷积核和 BiLSTM 的预测网络模型。经过计算，模型 2、3 与模型 1 在指标上的平均性能差距分别为 1.50% 与 1.99%。除此之外，从表 5 中还可以看出，在不同规模的目标域数据集上三者展现出几乎等同的学习效能，且性能差距微乎其微。实验结果表明，尽管结构有所差异，两者都能有效应对目标任务，维持了高水平的预测性能。尽管性能差距不大，但是模型 1 在保持卓越预测能力的同时，显著提升了资源效率。具体而言，相较于

对照模型 2，轻量化模型在训练速度上也取得了 10.1% 的增长，对照模型 3，轻量化模型在训练速度上也取得了 19.8% 的增长。这一提升直接关乎模型部署与迭代的效率，对资源有限或实时应用场景尤为重要。此外，模型尺寸和计算复杂度也得到了有效控制，模型 1 与模型 2 相比分别减少了 7.89% 的参数量与 7.87% 的计算需求，与模型 3 相比分别减少了 23.24% 的参数量与 23.07% 的计算需求。这对于移动设备、边缘计算等对内存和计算力有严格限制的环境来说，具有更高的适配性。这些提升主要得益于 2 个方面，一方面，正如上文所述，深度可分离卷积的引入显著降低了模型的参数量与计算量；另一方面，GRU 相较于 LSTM 与 BiLSTM 的较少参数量与更快训练速度也为性能提升提供了有力支持。因此，DATE 方法在轻量化预测网络设计是有效的，特别是在保持现有预测精度前提下，显著优化了模型的效率和资源占用，能够支持其在实际应用中资源敏感型场景下的部署和应用。

### 3.3.3 消融实验

本节将通过消融实验来检验 DATE 方法中通用节点特征学习和节点再加权 2 个部分的功能有效性。为了评估模型在 2 个领域数据上的泛化能力和学习能力，在进行节点分类的时候，将每个城市中节点的上下车总数进行排序，分别按照 25% 的比例给定类别 0、1、2、3，并进行 80 个回合的预训练嵌入，通过使用逻辑回归分类器对预训练得到的嵌入表示进行分类任务评估从而得到交叉验证得分，实验结果如表 6 所示。CVSCORE\_S 和 CVSCORE\_T 分别代表源域和目标域上预训练嵌入表示的交叉验证平均得分，它们用来评估模型在 2 个领域数据上的泛化能力。由于交叉验证利用了未参与模型训练的数据进行测试，因此可以提供模型在新数据上的表现预期，从而验证模型的泛化能力。从表 6 中可以看出，当单独移除节点级或边级的域适应策略时，尽管这可能促进源域上特征嵌入的学

表 5 模型性能的对比结果

模型	30天		7天		3天		平均训练时间/s	模型参数量	计算量
	RMSE	MAE	RMSE	MAE	RMSE	MAE			
1	3.957	1.331	4.211	1.466	4.368	1.612	28 789	297 099	11.94×10 <sup>9</sup>
2	3.896	1.32	4.121	1.441	4.332	1.581	32 024	322 561	12.96×10 <sup>9</sup>
3	3.872	1.315	4.098	1.433	4.318	1.572	35 891	387 073	15.52×10 <sup>9</sup>

习,但却不可避免地导致了目标域适应性能的衰减。考虑到迁移学习的核心目标在于提升模型在未见目标领域上的泛化能力,维持或增强目标域的表现至关重要。因此,结合节点级与边缘级的域适应机制在本文方法中显得尤为必要,它确保了模型能够学到既具有领域通用性又适应目标场景的特征表示。

表6 通用特征学习模块消融实验结果

方法	CVSCORE_S	CVSCORE_T
使用MK-MMD方法节点适应且包含特征映射层	0.698 6	0.632 9
使用MMD节点适应且不包含特征映射层	0.681 2	0.609 1
去除节点级域适应 ( $\alpha=0$ )	0.694 2	0.608 6
去除边缘级域适应 ( $\alpha=0$ )	0.698 7	0.603 7

除此之外,从表6中还可以看出,在引入多核最大均值差异(MK-MMD, multi-kernal maximum mean discrepancy)方法进行节点级域适应,并整合特征映射层后,源域和目标域的交叉验证性能分别显著提升了2.55%和3.90%,从而验证了所提出的域适应策略的有效性和在促进模型学习更泛化且领域无关特征方面的作用,这是因为MK-MMD方法域适应机制有效地缩小了源域与目标域特征分布差异,使模型能够在目标域上得到更准确的预测。因此,向DATE方法的域适应设计中加入MK-MMD方法的应用与特征映射层,在促进通用特征提取和提升模型跨域适应性方面表现出一定的正面效应。

### 3.3.4 抽样实验

本节通过抽样实验,探究了DATE方法在不同目标域数据完整度下的迁移性能。为评估模型在该场景下的训练效果,实验在30天时间窗口内开展,在训练前,分别按50%、20%、10%的比例对原始数据进行随机采样,以采样后的数据训练模型,实验结果如表7所示。从表7中可以看出,当源城市数据的完整度为50%时,模型的预测效果未受到明显影响。这表明,此时源域数据仍保留了较多有效信息,对模型训练的支撑作用较强,足以使模型保持较好的性能。然而,当数据完整度进一步降低至10%时,模型性能出现显著下滑:RMSE增长了31.31%,MAE增长了23.30%。这一结果说明,当源域数据完整度降至过低水平,数据量过于稀缺

时,模型因缺乏足够的信息输入,其迁移任务的效果会受到较大限制,难以维持较高的预测精度。因此,数据完整度对DATE方法的迁移性能仍有影响。在实际应用中,应尽可能确保源域数据的较高完整度,以保障模型的迁移效果。

表7 抽样实验结果

数据完整度	RMSE	MAE
100%	3.957	1.331
50%	3.964	1.407
20%	4.345	1.566
10%	5.196	1.641

## 4 结束语

为了满足智慧城市应用对交通流高精度预测的要求,本文提出了一种基于域适应的可迁移交通流预测方法,以满足不同应用的性能需求。为了迁移学习域适应过程中更好地捕捉跨域特征表示,本文提出基于节点级的MK-MMD方法去最小化源域和目标域差异,通过结合多种核函数捕捉不同尺度特征模式的方式,赋予模型更强的适应性和稳健性,并减少过拟合风险。同时,本文设计了一种轻量化时空序列预测网络去提高模型的训练效率和在低算力场景下的适配性,通过采用引入了注意力机制的GRU,以减少训练时的计算负担,并构建残差结构确保深度网络的稳定训练,从而增强对长期依赖关系的学习。最后,本文在纽约、华盛顿、芝加哥的城市数据上进行大量对比实验。实验表明,在不同目标域数据量级下,DATE方法相较于基准模型综合效果提升9.03%,并在参数量和计算量上分别减少了7.89%和7.87%,验证了本文方法的有效性。

虽然相比其他交通流预测模型,DATE方法具有较高的预测精度,但当目标城市图结构与源城市差异极大时,可能出现迁移效果显著下降的情况。这主要是由于方法设计过程中假设了源域与目标域之间的结构相似性,当这种假设不能满足时,域适应模块所捕获的跨域特征分布差异可能过大,导致迁移性能降低。针对这一局限性,在实际应用中,应优先考虑选择结构相似或功能相近的城市作为迁移源城市。在未来的研究中,一方面,笔者将考虑从图重构的角度出发,引入图的重构损失和将有向图或带权有向边纳入模型,以更真实地模拟交通流

在节点间的非对称特征, 设计更精简和高效的交通流预测模型; 另一方面, 笔者计划后续收集多模态的数据集 (如公共交通、骑行等) 和在 DiDi、TaxiBJ 数据集上进行实验分析, 探索在迁移过程中加入结构自适应模块和尝试将 DATE 方法应用于更多样化的交通数据, 扩大模型的适用性和泛化能力, 降低结构差异带来的迁移性能衰减问题。

### 参考文献:

- [1] LI X L, PAN G, WU Z H, et al. Prediction of urban human mobility using large-scale taxi traces and its applications[J]. *Frontiers of Computer Science*, 2012, 6(1): 111-121.
- [2] ZHANG N, ZHANG Y L, LU H T. Seasonal autoregressive integrated moving average and support vector machine models[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2011, 2215(1): 85-92.
- [3] SHEKHAR S, WILLIAMS B M. Adaptive seasonal time series models for forecasting short-term traffic flow[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2007, 2024(1): 116-125.
- [4] FENG X X, LING X Y, ZHENG H F, et al. Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(6): 2001-2013.
- [5] 何祖杰, 吴新焱, 刘中华. 基于改进灰狼算法优化支持向量机的短期交通流预测[J]. *厦门大学学报(自然科学版)*, 2022, 61(2): 288-297.  
HE Z J, WU X Y, LIU Z H. Optimized SVM model for short-term traffic flow prediction based on improved gray wolf optimizer[J]. *Journal of Xiamen University (Natural Science)*, 2022, 61(2): 288-297.
- [6] BENGIO Y. Learning deep architectures for AI[J]. *Foundations and Trends® in Machine Learning*, 2009, 2(1): 1-127.
- [7] JIANG W W, LUO J Y. Graph neural network for traffic forecasting: a survey[J]. *Expert Systems with Applications*, 2022, 207: 117921.
- [8] BAI L, YAO L N, KANHERE S S, et al. STG2Seq: spatial-temporal graph to sequence model for multi-step passenger demand forecasting[J]. *arXiv Preprint, arXiv: 1905.10069*, 2019.
- [9] PAN Z Y, ZHANG W T, LIANG Y X, et al. Spatio-temporal meta learning for urban traffic prediction[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(3): 1462-1476.
- [10] LI M X, TANG Y H, MA W. Few-sample traffic prediction with graph networks using locale as relational inductive biases[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(2): 1894-1908.
- [11] LU B, GAN X Y, JIN H M, et al. Spatiotemporal adaptive gated graph convolution network for urban traffic flow forecasting[C]//*Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. New York: ACM Press, 2020: 1025-1034.
- [12] LU B, GAN X Y, ZHANG W N, et al. Spatio-temporal graph few-shot learning with cross-city knowledge transfer[C]//*Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2022: 1162-1172.
- [13] JIN Y L, CHEN K, YANG Q. Selective cross-city transfer learning for traffic prediction via source city region re-weighting[C]//*Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2022: 731-741.
- [14] WANG D J, GUO G M, OUYANG T P, et al. A lightweight spatio-temporal neural network with sampling-based time series decomposition for traffic forecasting[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2025, 26(6): 8682-8693.
- [15] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[J]. *arXiv Preprint, arXiv: 1703.03400*, 2017.
- [16] NICHOL A, ACHIAM J, SCHULMAN J. On first-order meta-learning algorithms[J]. *arXiv Preprint, arXiv: 1803.02999*, 2018.
- [17] YAO H X, LIU Y D, WEI Y, et al. Learning from multiple cities: a meta-learning approach for spatial-temporal prediction[C]//*Proceedings of the World Wide Web Conference*. New York: ACM Press, 2019: 2181-2191.
- [18] HUANG Y J, SONG X Z, ZHANG S Y, et al. Transfer learning in traffic prediction with graph neural networks[C]//*Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. Piscataway: IEEE Press, 2021: 3732-3737.
- [19] TANG Y, QU A, CHOW A H F, et al. Domain adversarial spatial-temporal network: a transferable framework for short-term traffic forecasting across cities[J]. *arXiv Preprint, arXiv: 2202.03630*, 2022.
- [20] ZHAO H, YANG H, WANG Y, et al. Domain-adversarial-based temporal graph convolutional network for traffic flow prediction problem[C]//*Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. Piscataway: IEEE Press, 2021: 1365-1370.
- [21] ZHANG M Y, LI T, LI Y, et al. Multi-view joint graph representation learning for urban region embedding[C]//*Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization*, 2020: 4431-4437.
- [22] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. *arXiv Preprint, arXiv: 1704.04861*, 2017.
- [23] WANG L Y, GENG X, MA X J, et al. Cross-city transfer learning for deep spatio-temporal prediction[J]. *arXiv Preprint, arXiv: 1802.00386*, 2018.
- [24] WANG S Z, MIAO H, LI J Y, et al. Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(5): 4695-4705.
- [25] CHEN J, ZHENG L, HU Y Z, et al. Traffic flow matrix-based graph neural network with attention mechanism for traffic flow prediction[J]. *Information Fusion*, 2024, 104: 102146.
- [26] SUN L J, LIU M Z, LIU G F, et al. FD-TGCN: fast and dynamic temporal graph convolution network for traffic flow prediction[J]. *Information Fusion*, 2024, 106: 102291.
- [27] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. *The Journal of Machine Learning Research*, 2016, 17(1): 2096-2030.

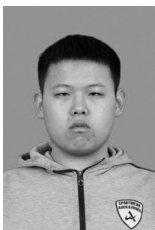
## [作者简介]



夏振厂 (1987-), 男, 河南周口人, 博士, 武汉理工大学讲师、硕士生导师, 主要研究方向为物联网、网络拥塞控制、强化学习等。



刘冰艺 (1990-), 男, 湖北武汉人, 博士, 武汉理工大学副教授、博士生导师, 主要研究方向为无线网络、车载自组织网络、物联网等。



班一杰 (2001-), 男, 山东临沂人, 武汉理工大学硕士生, 主要研究方向为交通流预测。



崔建群 (1974-), 女, 江苏南通人, 博士, 华中师范大学教授、博士生导师, 主要研究方向为机会网络、物联网、移动网络、应用层组播等。



田有亮 (1983-), 男, 贵州盘州人, 博士, 贵州大学教授、博士生导师, 主要研究方向为博弈论、密码学与安全协议、大数据隐私保护等。



常亚楠 (1984-), 女, 河南漯河人, 博士, 华中师范大学副教授、硕士生导师, 主要研究方向为无线 mesh 网络、延迟容忍网络、社交网络等。