

面向数据稀缺场景的智能交通流量预测^{*}

李 云^{1,2}, 高 雅¹, 姚枝秀^{2,3}, 夏士超³, 吴广富³

¹(重庆邮电大学 计算机科学与技术学院, 重庆 400065)

²(重庆邮电大学 通信与信息工程学院, 重庆 400065)

³(重庆邮电大学 软件工程学院, 重庆 400065)

通信作者: 夏士超, E-mail: xiashichao@cqupt.edu.cn



摘 要: 交通流预测是智能交通系统 (intelligent transportation system, ITS) 中交通管理的重要基础和热门研究方向。传统的交通流预测方法通常需要借助大量高质量历史观测数据进行预测, 而针对更为普遍的数据稀缺的交通路网场景预测精度则急剧下降。针对这一问题, 提出一种基于时空图卷积网络的迁移学习模型 (transfer learning based on spatial-temporal graph convolutional network, TL-STGCN), 结合数据充足的源路网的交通流特征, 辅助预测数据稀缺的目标路网未来交通流。首先, 采用基于时间注意力的时空图卷积网络学习源路网和目标路网交通流数据的时空特征表示; 其次, 结合迁移学习方法, 提取两个路网特征表示的域不变时空特征; 最后, 利用这些域不变时空特征对目标路网未来交通流做出预测。为了验证模型的有效性, 在真实世界数据集上进行实验。结果表明, 与现有方法对比, TL-STGCN 在平均绝对误差、均方根误差以及平均绝对百分比误差指标中均取得最高精度, 证明对于数据稀缺的交通路网预测任务, TL-STGCN 具有更好的预测性能。

关键词: 智能交通; 交通流预测; 图卷积网络; 注意力机制; 迁移学习

中图法分类号: TP311

中文引用格式: 李云, 高雅, 姚枝秀, 夏士超, 吴广富. 面向数据稀缺场景的智能交通流量预测. 软件学报, 2025, 36(8): 3787–3801. <http://www.jos.org.cn/1000-9825/7239.htm>

英文引用格式: Li Y, Gao Y, Yao ZX, Xia SC, Wu GF. Intelligent Traffic Flow Prediction for Data Scarcity Scenarios. Ruan Jian Xue Bao/Journal of Software, 2025, 36(8): 3787–3801 (in Chinese). <http://www.jos.org.cn/1000-9825/7239.htm>

Intelligent Traffic Flow Prediction for Data Scarcity Scenarios

LI Yun^{1,2}, GAO Ya¹, YAO Zhi-Xiu^{2,3}, XIA Shi-Chao³, WU Guang-Fu³

¹(School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

²(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

³(School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Traffic flow prediction is an important foundation and a hot research direction for traffic management in intelligent transportation systems (ITS). Traditional methods for traffic flow prediction typically rely on a large amount of high-quality historical observation data to achieve accurate predictions, but the prediction accuracy significantly decreases in more common scenarios with data scarcity in traffic networks. To address this problem, a transfer learning model is proposed based on spatial-temporal graph convolutional networks (TL-STGCN), which leverages traffic flow features from a source network with abundant data to assist in predicting future traffic flow in a target network with data scarcity. Firstly, a spatial-temporal graph convolutional network based on time attention is employed to learn the spatial and temporal features of the traffic flow data in both the source and target networks. Secondly, domain-invariant spatial-

* 基金项目: 国家自然科学基金 (62301099, 62071077); 中国博士后科学基金 (2023MD734137); 重庆市自然科学基金创新发展联合基金 (2022NSCQ-LZX0191); 重庆市教委科学技术研究项目 (KJQN202300638)

收稿时间: 2023-07-12; 修改时间: 2024-03-15, 2024-05-15; 采用时间: 2024-06-17; jos 在线出版时间: 2024-11-18

CNKI 网络首发时间: 2024-11-20

temporal features are extracted from the representations of the two networks using transfer learning techniques. Lastly, these domain-invariant features are utilized to predict the future traffic flow in the target network. To validate the effectiveness of the proposed model, experiments are conducted on real-world datasets. The results demonstrate that TL-STGCN outperforms existing methods by achieving the highest accuracy in mean absolute error, root mean square error, and mean absolute percentage error, which proves that TL-STGCN provides more accurate traffic flow predictions for scenarios with data scarcity in traffic networks.

Key words: intelligent transportation; traffic flow prediction; graph convolutional network (GCN); attention mechanism; transfer learning

随着经济的发展,科技的进步,汽车保有量不断增加,交通系统不断扩张且变得愈加复杂,城市道路变得拥堵,人们的出行变得愈发困难.智能交通系统(intelligent traffic system, ITS)用于规划管理和智能调度交通运行已成为未来的建设趋势^[1].准确地预测未来交通流是智能交通系统的基础和核心任务,有助于提前预知路况,缓解道路拥堵,同时为行程规划、地区功能划分等上层任务提供基础.

传统的交通预测模型包括历史平均模型、自回归平均移动模型等统计学习方法.这些模型算法简单,但在实际应用中,对高度非线性交通流的预测能力有限.近年来随着深度学习研究的发展,深度学习的方法如循环神经网络(recurrent neural network, RNN)和门控卷积网络(gated convolutional neural network, Gated CNN)等被广泛应用于交通预测任务中^[2].这类方法可以利用自循环机制和数据的时间依赖性进行预测,与统计学习的方法相比预测精度更高.然而,交通流是一种典型的时空数据,在进行预测时既需要考虑道路网络的时间依赖性,同时也需要考虑其空间依赖性.对此,一些研究^[3,4]采用RNN等提取交通流数据的时间依赖性,并结合卷积神经网络(convolutional neural network, CNN)提取空间依赖性,进而得到交通流数据完整的时空依赖性.但传统的CNN只适用于规则的欧几里得空间结构数据,这些方法仅适用于网格结构的交通路网.然而,大部分的交通路网本质上是图结构的,因此一些研究^[5-8]引入了图神经网络(graph neural network, GNN)来提取此类非欧几里得数据的空间特征,同时结合RNN或Gated CNN方法提取交通流数据时间特征,并结合注意力机制从而更好地关注到关键信息.这些方法在交通路网预测任务中取得了良好的效果.

然而,上述研究工作大多基于大量完备的历史观测数据来进行有效的预测,这对于一些交通路网是不现实的.例如,一些交通路网遭受天气、停电、机器故障等影响会导致路网数据丢失或异常;或由于隐私问题,历史交通数据不可使用;还有部分地区发展水平低,传感器数量少,导致数据稀缺.一般而言,数据稀缺指的是交通路网中一段时间或某些时间的交通流数据不连续,丢失,难以获取,或者收集到的数据样本少.在这些情况下,缺少历史观测数据,传统方法往往从大量历史数据中提取时空相关性,在数据稀缺时这些方法提取到的特征不准确,因此无法对未来交通流进行准确的预测.迁移学习是解决交通预测中数据稀缺问题的有效方法,将数据丰富的交通路网作为源域,数据稀缺的交通路网作为目标域,通过迁移学习从源域中学习可转移知识,以辅助目标域进行交通流预测.近年来,一些研究^[9-12]已经将迁移学习的方法用于交通预测任务中,然而这些方法基于CNN网络将交通路网划分成网格结构,所以不适用于图结构的交通数据.部分研究^[13]将图嵌入和迁移学习结合起来,但是交通流数据有强烈的时间相关性,这种方法忽略了数据间的长期依赖关系,对于时间特征的提取不够深入.因此,如何在历史观测数据稀缺的图结构交通路网上准确地预测其未来交通流,仍然是值得解决的问题.

针对上述问题,本文提出一种基于时空图卷积网络的迁移学习模型(transfer learning based on spatial-temporal graph convolutional network, TL-STGCN),该模型利用时空图卷积网络结合迁移学习来预测目标域的未来交通流.本文的贡献如下.

(1) 提出一种基于时空图卷积网络的迁移学习模型 TL-STGCN,从数据充足的源路网中学习可转移的知识,用以解决数据稀缺的目标交通路网的未来交通流预测问题.

(2) 采用时间注意力机制,给予关键信息更多关注,对时序数据的长期依赖关系进行处理,从而捕捉交通数据不同时间之间的动态相关性.进一步结合 GNN 与 Gated CNN 构成时空图卷积网络,以提取源路网和目标域交通路网的时空特征表示.将时空图卷积网络与对抗性域适应相结合,以减小两个域间特征表示分布的差异,促进知识的迁移.

(3) 在真实世界的交通数据集中进行了实验,并与多种基线方法进行对比,采用了平均绝对误差、均方根误差

以及平均绝对百分比误差这3种指标对模型进行评估.实验结果表明在这3种指标上,TL-STGCN的精度均高于基线方法,验证了本文的TL-STGCN的预测效果优于基线模型.

本文第1节介绍相关工作.第2节介绍本文的问题定义.第3节介绍本文提出的TL-STGCN模型.第4节通过对比实验验证了所提模型的有效性.最后总结全文和展望下一步工作.

1 相关工作

1.1 交通流预测

早期的交通流预测算法主要基于统计方法或者机器学习,如历史平均模型(history average, HA)、自回归积分滑动平均模型(autoregressive integrated moving average model, ARIMA)等,但是这些模型很难提取交通路网中高度非线性的时空特征,且在大规模数据上表现较差.随着深度学习的快速发展,一些研究采用RNN来提取交通流数据的时间相关性.Mou等人^[14]提出一种基于长短期记忆网络(long short-term memory, LSTM)的模型用于预测单个路段的交通流,该模型利用LSTM捕捉交通流数据在时间上的相关性.Poonia等人^[15]采用LSTM设计更深层次的网络模型,用以学习交通流数据的时间依赖关系及非线性特征,从而预测瞬时交通流.Zhang等人^[16]考虑天气条件,将天气特征和交通数据结合,提出一种基于门控循环单元(gated recurrent unit, GRU)的循环神经网络模型,GRU作为模型隐藏层中的神经元,以提取数据在时间上的相关性.然而交通流数据还具有空间相关性,因此越来越多的研究结合RNN和CNN等深度学习方法对交通数据进行研究,旨在建立复杂的时空数据模型,从而捕获交通路网的时空依赖关系.Ma等人^[17]把交通路网的时空信息转化为二维时空矩阵,将该矩阵作为描述交通流时空关系的图像,利用CNN提取该图像中的时空特征.Cheng等人^[3]提出了一种端到端的框架,结合注意力机制调整时空信息,用CNN来获取交通路网的空间相关性,RNN用来提取时间相关性.Zhuang等人^[4]将经过预处理的数据作为CNN模型的输入以提取交通流的空间特征,空间特征作为LSTM模型的输入,再提取时间上的特征.结合RNN和CNN的方法能够同时捕捉交通流的时间特征和空间特征,这些方法适用于标准的网格化数据.然而现实中的交通流数据都基于路网,路网本质上是更为复杂的非欧氏图结构^[18],CNN无法处理这种结构.

为了解决CNN对复杂的图数据特征提取不足的问题,图卷积网络(graph convolutional network, GCN)^[19]被提出,由于GCN与交通路网的图结构高度契合,因此被广泛应用于交通预测任务中.Guo等人^[5]基于GCN和GRU,提出一种时空神经网络模型用于交通预测,将交通数据表示为图形,道路为边,十字路口为节点,使用GCN和GRU来提取时空特征.Cui等人^[6]提出了一个交通图卷积算子以适应交通路网的物理特性,并提取综合特征,设计一种基于GCN和LSTM的深度学习框架,以学习空间和时间特征.冯宁等人^[7]将空间维度图卷积、时间维度卷积和全连接层构成一个组件以捕捉数据的时空相关性,通过3个相同的组件分别建模交通流数据的近期、日周期、周周期特性,从而对交通流进行预测.Yu等人^[8]提出了时空图卷积网络,该模型堆叠多个时空块,每个时空块串联一维卷积和GCN层,利用GCN获取交通数据的空间相关性,门控卷积用来提取交通数据的时间相关性.Zhao等人^[20]提出基于时间图卷积网络交通流预测算法,GCN用于学习复杂的拓扑结构捕获空间依赖性,GRU用于捕获时间依赖性.

注意力机制在深度学习中得到了广泛应用,注意力机制旨在忽略不重要的信息,对关键信息给予更多关注.一些研究结合了注意力机制,从而得到更准确的时空特征.Zhang等人^[21]提出了注意力图卷积序列模型,利用图卷积运算捕获交通流数据的空间特征,采用注意力机制为每个时间步长添加注意力向量以捕捉信息间的时间相关性.Guo等人^[22]提出一种基于注意力的时空图卷积网络来解决交通预测问题,该模型对交通流的临近、每日和每周的依赖关系建模,采用时空注意力机制有效地捕捉交通数据中的动态时空关联,用GCN提取数据在空间上的相关性,用标准卷积提取时间上的相关性.赵文竹等人^[23]结合局部卷积多头注意力机制来获取时间相关性,使用动态图卷积学习节点在静态空间拓扑和动态流量模式下的空间特性,从而进行交通流预测.

1.2 迁移学习

迁移学习旨在从数据充足的源域中学习可迁移的知识,以辅助数据稀缺的目标域完成预测任务,是解决数据

稀缺的目标域预测问题的一种有效方法^[24]. Wang 等人^[9]提出了一种城市人流预测的跨域时空知识转移的深度注意适应网络, 源域和目标域通过共享的 Conv-LSTM 层学习时空数据表示, 结合注意机制捕获更广泛的空间相关性, 学习到的表示映射到一个共同的特征空间, 通过最大平均差异进行知识转移. Li 等人^[10]为了解决数据不足的问题, 提出一个基于迁移学习的通用流量预测框架, 使用堆叠的 LSTM 构建网络模型以捕捉数据中的高维非线性关系, 再应用 3 种不同的迁移策略, 结合交通流模式和地理属性进行迁移流量预测, 提高了模型预测精度. Wan 等人^[11]针对网络流量预测中样本量小的问题, 提出了一种基于 LSTM 和迁移学习的神经网络模型, 构建 LSTM 网络, 并对其进行训练, 通过迁移学习将执行源任务的神经网络模型参数转移到执行目标任务的神经网络模型中, 用少量的目标域数据构建出性能良好的预测模型. Wang 等人^[12]提出一种深度时空预测任务的跨城市迁移学习方法, 该方法学习一个城市区域匹配函数将目标区域与相似的源区域连接起来, 将 Conv-LSTM 和 Conv2D 叠加以提取时空依赖关系, 利用匹配函数将学习到的源域知识传递给目标域. Tang 等人^[13]提出一种域对抗时空网络, 利用图嵌入和 GRU 相结合的方式得到节点嵌入信息, 采用对抗性域适应的方式将源域的嵌入表示迁移到目标域中, 得到一种可转移的交通预测模型, 用来提升交通流预测的精度. Ren 等人^[25]等提出一种深度迁移学习的交通预测模型, 模型采用向量化的方式表达时空数据, 采用深度学习的方式提取交通特征表示, 将其他外部因素赋予相应的权重, 通过迁移学习的方式辅助完成预测任务.

尽管已有大量工作对交通流预测进行研究, 并取得了一定的成果. 然而, 交通流预测仍然面临着如下挑战: 对于历史观测数据稀缺的图结构的交通路网, 无法准确预测其未来一定时间段的交通流. 对此, 本文提出 TL-STGCN 模型, 基于结合注意力机制的时空图网络提取交通流时空特征, 采用对抗性域适应的方式进行知识迁移, 以准确地预测目标路网交通流.

2 问题定义

2.1 交通路网

交通流数据一般通过分布在路网上的多个传感器所获取, 本文中交通路网由无向图 $G = (V, E, A)$ 表示, 其中 V 表示节点的集合, $|V| = N$ 表示图中节点的个数, 每个节点代表路网中的一个传感器, 传感器通常布置在交通路网的路段中; E 为图 G 中节点间边的集合, 表示节点之间的连通性, 若两个节点之间有道路相连, 则两个节点间存在一条边; $A \in \mathbb{R}^{N \times N}$ 是图 G 的邻接矩阵, 表示图 G 中节点的连接情况, 若节点 i 与节点 j 有边, 则邻接矩阵中 (i, j) 的值为 1, 反之则为 0. 定义特征矩阵 $x_t = (x_{t,1}, \dots, x_{t,i}, \dots, x_{t,N}) \in \mathbb{R}^N$ 表示在 t 时刻图 G 的交通流信息, 其中 $x_{t,i}$ 表示 i 节点在 t 时刻的特征值. 因此, 交通数据的结构如图 1 所示.

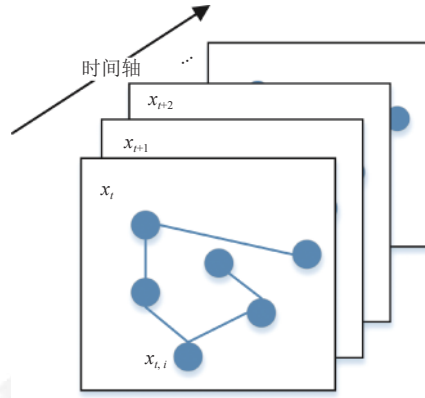


图 1 交通数据的图结构

2.2 源域图

源域图定义为无向图 $G^s = (V^s, E^s, A^s)$, 表示具有充足历史交通数据的源路网. 其中 V^s 为源域节点的集合; E^s

为源域图边的集合; $A^S \in \mathbb{R}^{N^S \times N^S}$ 为源域图的邻接矩阵; $x_t^S \in \mathbb{R}^{N^S}$ 为源域图在 t 时刻的特征矩阵.

2.3 目标域图

目标域图定义为无向图 $G^T = (V^T, E^T, A^T)$, 表示历史观测数据稀缺的目标交通路网. 其中 V^T 为目标域节点的集合; E^T 为目标域图边的集合; $A^T \in \mathbb{R}^{N^T \times N^T}$ 为目标域图的邻接矩阵; 目标域图在 t 时刻的特征矩阵表示为 $x_t^T \in \mathbb{R}^{N^T}$.

2.4 问题定义

假设源域和目标域具有相同的特征空间, 但特征的边缘分布不同, 例如从不同城市收集到的交通流数据. 此外, 源域有丰富的交通数据, 而目标域面临着交通数据稀缺的问题. 本文旨在学习一个预测函数 $f(\cdot)$, 通过从源域迁移知识, 以辅助预测目标域未来一段时间内的交通流信息.

本文中所用到的符号及其含义如表 1 所示.

表 1 文中符号及描述

符号	含义
$G = (V, E, A)$	无向图
V	无向图的节点
E	无向图的边
A	无向图的邻接矩阵
$G^S = (V^S, E^S, A^S), G^T = (V^T, E^T, A^T)$	源域图, 目标域图
x_t^S, x_t^T	源域图、目标域图在 t 时刻的特征矩阵
$\mathcal{X}^S, \mathcal{X}^T$	源域和目标域的数据集
p, q	历史观测时间步长, 预测时间步长
$\mathcal{X}_p$	过去 p 时间段的历史观测数据
$\mathcal{X}_{p+q}$	未来 q 时间段的交通数据
T^S, T^T	源域和目标域的时间序列样本数
$f_g(\cdot), f_w(\cdot), f_{\text{pre}}(\cdot)$	时空图卷积网络, 域判别器, 交通预测网络
$\theta_g, \theta_w, \theta_{\text{pre}}$	时空图卷积网络、域判别器、交通预测网络的训练参数
$\mathbf{H}_g, \mathbf{H}_g^S, \mathbf{H}_g^T$	时空特征表示, 源于时空特征表示, 目标域时空特征表示
γ, λ	梯度惩罚平衡系数, 交通预测模块和特征表示学习的平衡系数
$\mathcal{L}_{\text{src}}, \mathcal{L}_{\text{tgt}}$	源域预测损失, 目标域预测损失
$\mathcal{L}_{\text{wd}}, \mathcal{L}_{\text{gard}}$	域判别器损失, 梯度惩罚

不失一般性, 本文根据过去 p 时段历史交通流数据, 预测未来 q 时段内目标域交通流. 定义源域数据集和目标域数据集分别如下:

$$\mathcal{X}^S = \left\{ (x_{t+1}^S, \dots, x_{t+p}^S), (x_{t+p+1}^S, \dots, x_{t+p+q}^S) \right\}_{t=0}^{T^S-1} \quad (1)$$

$$\mathcal{X}^T = \left\{ (x_{t+1}^T, \dots, x_{t+p}^T), (x_{t+p+1}^T, \dots, x_{t+p+q}^T) \right\}_{t=0}^{T^T-1} \quad (2)$$

其中, $\mathcal{X}_p = (x_{t+1}, \dots, x_{t+p})$ 为过去 p 时间段内交通流数据的历史观测值, 作为模型的输入, $\mathcal{X}_{p+q} = (x_{t+p+1}, \dots, x_{t+p+q})$ 为未来 q 时段的交通流数据, 是模型的输出, T^S 和 T^T 为时间序列样本的数量, 且 T^S 的数量远大于 T^T 的数量. 综上, 本文的交通流预测问题定义如下:

$$\min_f \text{error}(\mathcal{X}_{p+q}^T, \hat{\mathcal{X}}_{p+q}^T) \quad (3)$$

其中, $\hat{\mathcal{X}}_{p+q}^T = f(\mathcal{X}^S, G^S, \mathcal{X}^T, G^T)$ 表示交通流的预测值, $f(\cdot)$ 表示学习到的预测函数.

3 TL-STGCN 模型

3.1 总体框架

本文提出的 TL-STGCN 模型主要由时空卷积网络、域判别器模块和交通流预测模块构成,其总体框架如图 2 所示。

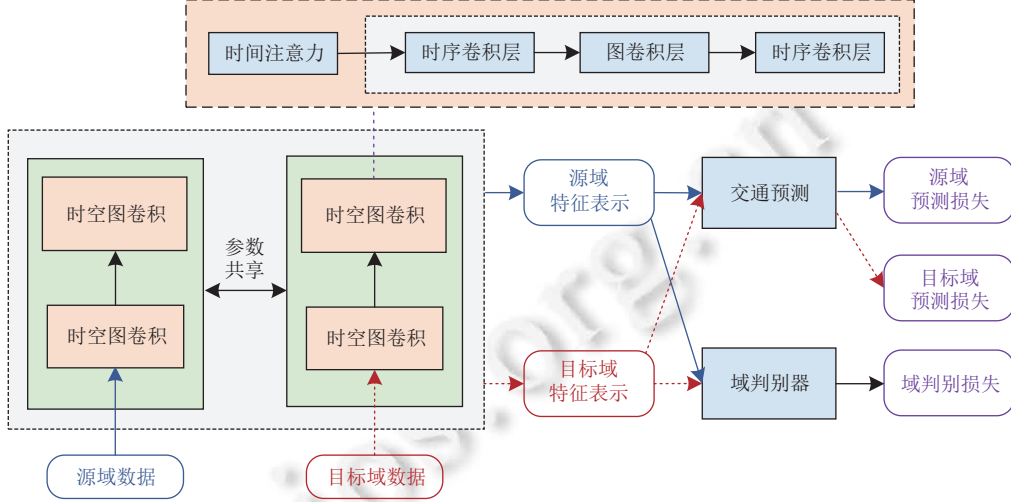


图 2 TL-STGCN 模型框架

首先,将源域和目标域的历史观测数据作为输入,分别送入到时空图卷积网络中以学习交通流数据的时空特征表示。该网络先采用时间注意力机制计算出不同时间序列间的相关性,以自适应赋予数据相应的权重。接着将经过注意力机制调整后的数据输入到 Gated CNN 和 GCN 中提取源域和目标域的时空特征表示。

然后,将学习到的特征表示发送到域判别器中。域判别器用来度量源域和目标域特征表示分布之间的距离。模型结合域判别器,通过对抗性预适应的方式训练特征提取网络(即时空图卷积网络)以最小化源特征表示分布和目标特征表示分布之间的距离,最终学习到域不变的特征。

最后,结合交通预测模块,将源域和目标域的时空数据合并到表示学习的过程中,进一步学习域不变和更具鉴别性的特征,并通过学习到的特征来预测目标域未来的交通流数据。

3.2 时空图卷积网络

在 TL-STGCN 的时空图卷积网络中,采用时间注意力机制动态调整输入,以更好地提取交通流数据的时空特征表示,并采用 GCN 构成图卷积层以提取空间特征、Gated CNN 构成时序卷积层来提取时间特征。

3.2.1 时间注意力机制

在时间维度上,交通路网中不同时间段的交通状况之间存在相关性,不同情况下的相关程度也不同,例如,相邻时间段之间的交通状况关联更大。注意力机制能够寻找需要关注的信息并根据信息的重要程度进行加权,使模型对重要的信息给予更多的关注,以提升时间维度上的特征提取能力。因此,TL-STGCN 模型中采用了如下所示的时间注意力机制^[22],自适应地对数据赋予不同的权重:

$$\mathbf{E} = \mathbf{V}_e \cdot \sigma \left(\left((\mathbf{X}_p^{(r-1)})^T \mathbf{U}_1 \right) \mathbf{U}_2 \left(\mathbf{U}_3 \mathbf{X}_p^{(r-1)} \right) + \mathbf{b}_e \right) \quad (4)$$

$$\mathbf{E}'_{i,j} = \frac{\exp(\mathbf{E}_{i,j})}{\sum_{j=1}^{T_{r-1}} \exp(\mathbf{E}_{i,j})} \quad (5)$$

其中, $\mathbf{X}_p^{(r-1)} = \mathbf{X}_p \in \mathbb{R}^{N \times C_{r-1} \times T_{r-1}}$ 是第 r 层特征提取模块的输入; C_{r-1} 是第 r 层中输入数据的通道数,当 $r=1$ 时, $C_0 = F$

表示输入特征的个数; T_{r-1} 表示第 r 层中时间序列的长度, $r=1$ 时 $T_0=p$; $V_e, b_e \in \mathbb{R}^{T_{r-1} \times T_{r-1}}$, $U_1 \in \mathbb{R}^N$, $U_2 \in \mathbb{R}^{C_{r-1} \times N}$, $U_3 \in \mathbb{R}^{C_{r-1}}$ 是可学习的参数; σ 为 Sigmoid 激活函数. 时间注意力矩阵 $\mathbf{E}$ 由变化的输入决定, $\mathbf{E}$ 中的元素 $\mathbf{E}_{i,j}$ 的值表示时刻 i 和 j 之间的依赖性强度, 最后利用 Softmax 函数对 $\mathbf{E}$ 进行归一化.

本文直接将归一化后的时间注意力矩阵 $\mathbf{E}'$ 与输入相乘, 动态调整输入, 得到注意力机制调整后的数据, 即:

$$\tilde{\mathbf{X}}_p^{(r-1)} = (\tilde{x}_{t+1}, \dots, \tilde{x}_{t+p}) = \mathbf{X}_p \mathbf{E}' \in \mathbb{R}^{N \times C_{r-1} \times T_{r-1}} \quad (6)$$

3.2.2 空间特征提取

图卷积层采用了 GCN 来提取空间特征. GCN 中包括谱域卷积和空域卷积, 为了充分利用交通路网的拓扑性质, 本文采用谱域卷积的方式处理空间特征. 无向图的拉普拉斯矩阵定义为 $\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{N \times N}$, 其中 $\mathbf{I}_N$ 为单位矩阵, $\mathbf{D} \in \mathbb{R}^{N \times N}$ 为对角矩阵, $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{A}_{ij}$ 表示节点的度, $\mathbf{A}$ 为图的邻接矩阵. 根据实对称矩阵正半定的性质, 归一化的拉普拉斯矩阵可以分解为 $\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$, 其中 $\mathbf{\Lambda} \in \mathbb{R}^{N \times N}$ 是特征值的对角矩阵, $\mathbf{U} \in \mathbb{R}^{N \times N}$ 是特征向量矩阵. 给定时间 t 的特征矩阵 $\tilde{x}_t \in \mathbb{R}^N$, 对于信号 $\tilde{x}_t$ 的傅里叶变换为 $\hat{x}_t = \mathbf{U}^T \tilde{x}_t$, 其反傅里叶变换为 $\tilde{x}_t = \mathbf{U} \hat{x}_t$, 则图卷积操作定义如下:

$$\Theta *_{\mathbf{G}} \tilde{x}_t = \Theta(\mathbf{L}) \tilde{x}_t = \mathbf{U} \Theta(\mathbf{\Lambda}) \mathbf{U}^T \tilde{x}_t \quad (7)$$

其中, $\Theta \in \mathbb{R}^N$ 是卷积核, $*_{\mathbf{G}}$ 表示图卷积操作, $\mathbf{U}^T \tilde{x}_t$ 为图信号 $\tilde{x}_t$ 的傅里叶变换.

为了解决计算代价太大的问题, 本文采用了切比雪夫多项式^[8], 表示如下:

$$\Theta *_{\mathbf{G}} \tilde{x}_t = \Theta(\mathbf{L}) \tilde{x}_t \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}}) \tilde{x}_t \quad (8)$$

其中, $\tilde{\mathbf{L}} = (2\mathbf{L}/\lambda_{\max}) - \mathbf{I}_N$ ($\lambda_{\max}$ 表示 $\mathbf{L}$ 的最大特征值), $\theta \in \mathbb{R}^K$ 为切比雪夫多项式系数向量, 切比雪夫多项式使用递归的方式进行定义: $T_0(\tilde{x}_t) = 1$, $T_1(\tilde{x}_t) = \tilde{x}_t$, $T_k(\tilde{x}_t) = 2\tilde{x}_t T_{k-1}(\tilde{x}_t) - T_{k-2}(\tilde{x}_t)$, $K-1$ 为多项式的系数.

在公式 (8) 的基础上引入一阶近似, 假定 $K=1$, 通过堆积的多层图卷积网络建立 K 阶邻居的依赖, 降低运算代价. 假设 $\lambda_{\max} \approx 2$, 公式 (8) 可以简化为:

$$\Theta *_{\mathbf{G}} \tilde{x}_t \approx \theta_0 \tilde{x}_t + \theta_1 \left(\frac{2\mathbf{L}}{\lambda_{\max}} - \mathbf{I}_N \right) \tilde{x}_t \approx \theta_0 \tilde{x}_t - \theta_1 (\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}) \tilde{x}_t \quad (9)$$

其中, θ_0 和 θ_1 是卷积核的两个共享参数, 通过对参数进行约束来避免过拟合. 使 $\theta = \theta_0 = -\theta_1$, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$, $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$, 因此本文中的图卷积表示为公式 (10):

$$\Theta *_{\mathbf{G}} \tilde{x}_t = \theta (\mathbf{I}_N + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}) \tilde{x}_t = \theta (\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}) \tilde{x}_t \quad (10)$$

3.2.3 时间特征提取

TL-STGCN 设计了时序卷积层, 在时间轴上采用 Gated CNN 的方式来捕捉时间特征. Gated CNN 能够并行处理时序数据, 具有训练速度快、结构简洁的优势. 时间卷积层包含内核宽度为 K_t 的 1-D 因果卷积, 以及一个非线性的门控线性单元 (gated linear unit, GLU). 对于图中的每个节点, 时间卷积不加填充地搜索输入元素的 K_t 邻域, 从而使序列长度每次缩短 $K_t - 1$. 每个节点处的时间卷积输入 $\tilde{x} \in \mathbb{R}^{p \times C_t}$ 可以被视为长度为 p 的序列, C_t 为信道数. 输入沿着时间维度进行一维卷积, 卷积核为 $\Gamma \in \mathbb{R}^{K_t \times C_t}$, 个数为 $2C_0$, 将输入 $\tilde{x}$ 映射到单个输出元素, 得到 $[PQ] \in \mathbb{R}^{(p-K_t+1) \times 2C_0}$, 然后进行 GLU 激活. 因此, 时间门控卷积可以定义为:

$$\Gamma *_{\mathbf{T}} \tilde{x} = P \odot \sigma(Q) \in \mathbb{R}^{(p-K_t+1) \times C_0} \quad (11)$$

其中, P, Q 都是 GLU 的输入, $*_{\mathbf{T}}$ 为时间卷积核, $\odot$ 是 Hadamard 乘法, σ 为 Sigmoid 激活函数.

3.2.4 时空图卷积

时空图卷积模块由两个时序卷积层和一个空间卷积层构成, 并在每个时空图卷积模块中进行归一化的处理来防止过拟合. 第 1 个时序卷积层捕捉时间特征后, 通过空间卷积层来提取空间特征, 再次利用时序卷积层提取时间特征, 这样的结构能够更深入地提取时间维度上的信息, 并且避免在处理空间特征过程中时间信息丢失^[8]. 第 l 个时空图卷积块的输入为 $\tilde{x}^l \in \mathbb{R}^{p \times N \times C^l}$, 输出为 $\tilde{x}^{l+1} \in \mathbb{R}^{(p-2(K_t-1)) \times N \times C^{l+1}}$, 计算过程如下:

$$\bar{x}^{l+1} = \Gamma_1^l *_{\Gamma} ReLU(\Theta^l *_{\Gamma} (\Gamma_0^l *_{\Gamma} \bar{x}^l)) \quad (12)$$

其中, $\Theta^l *_{\Gamma}$ 表示图卷积核, Γ_0^l 和 Γ_1^l 分别为块中第 1 个和第 2 个门控卷积核, $ReLU(\cdot)$ 为线性激活函数.

综上所述, 令本文的时空卷积网络表示 $f_g(\cdot)$, 该网络用于学习源域和目标域的时空特征表示. 给定任意域的交通流数据 $\mathcal{X}_p$ 及其邻接矩阵 $\mathbf{A}$, 先通过注意力机制对输入进行调整, 再通过时空卷积模块学习到该域的时空特征表示 $\mathbf{H}_g$ 如下:

$$\mathbf{H}_g = f_g(\mathcal{X}_p, \mathbf{A}; \theta_g) \quad (13)$$

其中, θ_g 表示网络中所有可训练参数.

3.3 域适应

将上述的时空图卷积网络 $f_g(\cdot)$ 作为特征提取器, 假设其学习到源域和目标域的特征表示分别为 $\mathbf{H}_g^S$ 和 $\mathbf{H}_g^T$. 然而, 由于两个域的特征表示分布不同, 用源域数据训练的模型不能够直接迁移到目标域, 因此本文采用对抗性域适应的方式减小域间差异, 学习域不变的特征表示, 以促进知识迁移^[26]. 在 TL-STGCN 中添加一个域判别器模块, 将域判别器和特征提取器相结合, 其中域判别器用来分辨学习到的特征表示来自源域还是目标域, 通过对抗的方式训练域判别器和特征提取器以学习到更接近源域的特征表示.

具体来说, 给定来自任意域的特征表示 $\mathbf{H}_g$, 域判别器学习一个函数 $f_w(\mathbf{H}_g; \theta_w)$, 其中 θ_w 为可训练的参数, 该函数将特征表示从多维表示映射到实数. 源域和目标域之间的 Wasserstein 距离 W_1 计算如下:

$$W_1 = \sup_{\|f_w\|_L \leq 1} \mathbb{E}_{\mathbb{P}_{H^S}} [f_w(\mathbf{H}_g; \theta_w)] - \mathbb{E}_{\mathbb{P}_{H^T}} [f_w(\mathbf{H}_g; \theta_w)] \quad (14)$$

其中, $\mathbb{P}_{H^S}$, $\mathbb{P}_{H^T}$ 分别为源域和目标域的时空特征表示分布. $\|f_w\|_L \leq 1$ 为 Lipschitz 约束, 函数 f_w 中的参数都满足 1-Lipschitz. 接着, 通过最大化如下域判别器损失来近似估算 Wasserstein 距离:

$$\mathcal{L}_{wd}(\mathcal{X}_p^S, \mathcal{X}_p^T) = \frac{1}{N^S T^S} \sum_{i=1}^{N^S} \sum_{t=0}^{T^S-1} f_w([f_g(\mathcal{X}_p^S, \mathbf{A}^S; \theta_g)]_i; \theta_w) - \frac{1}{N^T T^T} \sum_{j=1}^{N^T} \sum_{t=0}^{T^T-1} f_w([f_g(\mathcal{X}_p^T, \mathbf{A}^T; \theta_g)]_j; \theta_w) \quad (15)$$

其中, $[f_g(\mathcal{X}_p^S, \mathbf{A}^S; \theta_g)]_i$ 表示源域图中第 i 个节点的交通数据特征表示, $[f_g(\mathcal{X}_p^T, \mathbf{A}^T; \theta_g)]_j$ 表示目标域图中第 j 个节点的交通数据特征表示.

为了避免梯度爆炸或消失, 本文给参数 θ_w 添加一个梯度惩罚 $\mathcal{L}_{gard}$ 强制执行 Lipschitz 约束:

$$\mathcal{L}_{gard}(\hat{\mathbf{H}}) = (\|\nabla_{\hat{\mathbf{H}}} f_w(\hat{\mathbf{H}}; \theta_w)\|_2 - 1)^2 \quad (16)$$

其中, $\hat{\mathbf{H}}$ 表示源域和目标域特征表示对之间沿直线的随机点.

接着, 通过求解如下最大化问题来估计 Wasserstein 距离:

$$\max_{\theta_w} \mathcal{L}_{wd} - \gamma \mathcal{L}_{gard} \quad (17)$$

其中, γ 为梯度惩罚的平衡系数.

在训练时首先将域判别器训练到最优, 然后保持域判别器的参数不变, 通过最小化 Wasserstein 距离的估值, 使特征提取网络学习到域差异较小的特征表示. 根据上述, 域不变的特征表示学习通过解决如下最大最小问题实现:

$$\min_{\theta_g} \max_{\theta_w} \mathcal{L}_{wd} - \gamma \mathcal{L}_{gard} \quad (18)$$

公式 (18) 通过迭代的方式学习具有较小域差异的特征表示, 最终学习到域不变的特征表示.

3.4 交通预测

本文的目标是学习域不变的和有鉴别性的特征表示, 从而预测数据稀缺的目标域的未来交通流. 因此 TL-STGCN 中设计一个交通预测模块, 将源域和目标域的未来交通流数据整合到表示学习的过程中.

交通预测模块旨在训练一个神经网络模型 $f_{pre}(\cdot)$, 给定任意域的特征表示 $\mathbf{H}_g$, 通过 $f_{pre}(\mathbf{H}_g; \theta_{pre})$ 得到该域的预

测值, 其中 θ_{pre} 为可训练的神经网络参数.

交通预测模块的输出为预测到的未来交通流数据, 得到源域预测值 $f_{\text{pre}}(\mathbf{H}_g^S; \theta_{\text{pre}})$ 为 $\hat{\mathbf{X}}_{p+q}^S = (\hat{x}_{t+p+1}^S, \dots, \hat{x}_{t+p+q}^S)$, 目标域预测值 $f_{\text{pre}}(\mathbf{H}_g^T; \theta_{\text{pre}})$ 为 $\hat{\mathbf{X}}_{p+q}^T = (\hat{x}_{t+p+1}^T, \dots, \hat{x}_{t+p+q}^T)$, 其中 $\mathbf{H}_g^S$ 和 $\mathbf{H}_g^T$ 分别为源域和目标域的时空特征表示.

交通预测模块采用平均平方误差损失函数进行性能评估, 则源域和目标域的损失函数分别定义如下:

$$\mathcal{L}_{\text{src}}(\mathbf{H}_g^S) = \frac{1}{TS} \sum_{t=0}^{T^S-1} \sum_{k=p+1}^{p+q} \|\hat{x}_{t+k}^S - x_{t+k}^S\|^2 \quad (19)$$

$$\mathcal{L}_{\text{tgt}}(\mathbf{H}_g^T) = \frac{1}{T^T} \sum_{t=0}^{T^T-1} \sum_{k=p+1}^{p+q} \|\hat{x}_{t+k}^T - x_{t+k}^T\|^2 \quad (20)$$

综上, 本文提出的 TL-STGCN 模型最终损失函数如下:

$$\min_{\theta_g, \theta_{\text{pre}}} \mathcal{L}_{\text{src}} + \mathcal{L}_{\text{tgt}} + \lambda \max_{\theta_w} (\mathcal{L}_{\text{wd}} - \gamma \mathcal{L}_{\text{gard}}) \quad (21)$$

其中, λ 为交通预测模块和特征表示学习之间的平衡系数.

4 实验分析

4.1 数据集

本文选取真实数据集 PEMS04、PEMS08、PEMS-BAY、METR-LA^[7,22] 来对模型的有效性进行验证. PEMS、METR-LA 分别是加州高速公路和洛杉矶高速公路上的交通流数据, 数据集中包括了车流量、行驶速度、密度等信息. 本文选用其中的车流量数据, 单位为辆/h. 交通流数据由传感器采集, 将采集频率为 30 s/次的原始数据汇总成时间间隔为 5 min 的数据样本, 即每个传感器每小时收集 12 个数据样本, 每个传感器每天包含 288 个测量值, 原始数据集的描述如表 2 所示.

本文在 PEMS04、PEMS08、METR-LA、PEMS-BAY 中构建源域数据集 S-PEMS04、S-PEMS08、S-METR-LA、S-PEMS-BAY. 具体来说, 分别从 PEMS04、PEMS08 中选取前 56 天的数据为训练集, 后 3 天的数据为测试集来构建 S-PEMS04、S-PEMS08; 从 METR-LA、PEMS-BAY 分别选取前 100 天的数据为训练集, 后 3 天的数据为测试集以组成 S-METR-LA、S-PEMS-BAY.

从原始数据集中构建数据稀缺的目标域数据集 T-PEMS04、T-PEMS08、T-METR-LA、T-PEMS-BAY. 目标数据集分别选取 4 个原始数据集中最后 14 天的交通流数据, 其中前 11 天为训练集, 后 3 天为测试集. 此外, 定义一个数据稀缺率来模拟目标域数据缺失的百分比, 稀缺率越高表示数据质量越差^[10,12]. 在实验中, 若没有说明, 则默认将目标域中的稀缺率设置为 20%, 即从训练数据中随机选取 20% 的数据将其替换为 0. 目标域数据集和源域数据集描述如表 3 所示.

表 2 原始数据集描述

数据集	传感器数量	时间范围
PEMS04	307	2018.01.01–2018.02.28
PEMS08	170	2017.07.01–2018.08.31
PEMS-BAY	325	2017.01.01–2017.05.31
METR-LA	207	2012.03.01–2012.06.30

表 3 目标域、源域数据集描述

数据集	训练集天数	测试集天数
S-PEMS04, S-PEMS08	56	3
S-METR-LA, S-PEMS-BAY	100	3
目标域数据集	11	3

4.2 实验设置及评估指标

为了评价 TL-STGCN 模型的有效性, 分别采用平均绝对误差 (mean absolute error, MAE) 函数、均方根误差 (root mean square error, $RMSE$) 函数以及平均绝对百分比误差 (mean absolute percentage error, $MAPE$) 函数, 定义如下:

$$MAE = \frac{1}{MNq} \sum_{i=1}^N \sum_{t=0}^{M-1} \sum_{k=p+1}^{p+q} |x_{t+k,i} - \hat{x}_{t+k,i}| \quad (22)$$

$$RMSE = \sqrt{\frac{1}{MNq} \sum_{i=1}^N \sum_{t=0}^{M-1} \sum_{k=p+1}^{p+q} (x_{t+k,i} - \hat{x}_{t+k,i})^2} \quad (23)$$

$$MAPE = \frac{1}{MNq} \sum_{i=1}^N \sum_{t=0}^{M-1} \sum_{k=p+1}^{p+q} \frac{|x_{t+k,i} - \hat{x}_{t+k,i}|}{x_{t+k,i}} \quad (24)$$

其中, $x_{t+k,i}$ 为真实的未来交通流, $\hat{x}_{t+k,i}$ 为交通流的预测值, M 为时间样本, N 为节点数.

本文使用 1 h ($p = 12$) 的历史观测时间来预测未来 30 min ($q = 6$) 的交通流. 在实验中, 模型的时空特征提取模块 $f_g(\cdot)$ 由两层堆叠的时空图卷积模块组成. 域判别器 f_w 、流量预测模块 $f_{pre}(\cdot)$ 分别由两个全连接层组成. 系数 γ 和 λ 分别设定为 10 和 1. 域判别器的学习率 α_1 、流量预测模块的学习率 α_2 设置为 0.000 1, 以每 50 个训练步 0.8 的速率衰减. 批大小设置为 64, 迭代次数为 500.

为验证 TL-STGCN 模型的性能, 选择以下基线方法进行对比.

- HA: 使用目标域的历史流量平均值进行预测.
- GRU^[27]: 采用门控机制基于时间序列来捕获交通流的长期依赖性, 本文在目标域数据集上执行 GRU 模型来预测交通流.
- GCN^[28]: GCN 模型能够提取空间特征来进行预测, 本文在目标域数据集上执行 GCN 模型来预测交通流.
- STGCN^[8]: 将 GCN 和门控卷积结合, 联合提取时空特征. 本文只在目标域数据集上训练 STGCN 模型, 并对目标域进行交通流预测.
- DastNet^[13]: 结合图嵌入和 GRU, 在源域上学习节点的图嵌入, 并且将学习到的图嵌入迁移到目标域中, 辅助目标域进行预测任务.
- MGAT^[29]: 针对数据稀疏的交通路网, 结合 LSTM 和 CNN 与空间注意力机制, 对目标城市和源城市相似的多粒度特征编码进行迁移, 自适应地将元知识应用于目标路网.
- S-only: 在源域数据集上执行本文中基于时间注意力机制的时空图卷积网络模型, 并在目标域测试集上直接进行预测.
- TL-STGCN-noTA: 跟 TL-STGCN 相比, 消除本文的注意力机制, 利用在源域上学习到的特征, 在目标域数据集上进行预测.

4.3 结果及分析

(1) 与基线方法对比

将本文提出的 TL-STGCN 和基线模型在目标域数据集上进行未来 30 min 的交通流预测任务. 对于没有迁移学习的预测模型 (HA、GRU、GCN、STGCN), 直接将目标域数据集用于训练和预测. 对于其他有迁移学习的模型, 本文通过迁移学习任务对其性能进行评价. TL-STGCN 模型与基线模型性能的评估 MAE 、 $MAPE$ 、 $RMSE$ 的结果如表 4 所示. 其中“—”表示误差过大, 说明模型性能极差.

从表 4 中可得知, 就 4 个目标域数据集的所有评价指标来看, 在未来多个时间步内, 对比现有的深度学习模型, 本文提出的 TL-STGCN 模型在数据稀疏的交通路网预测任务中拥有更好的性能.

对于没有迁移学习的方法来说, 基于传统统计方法的 HA 模型性能最差, 这是因为其直接采用历史平均值来进行预测, 没有考虑交通路网数据中存在的复杂关系. 基于深度学习的 GRU 方法虽然能提取交通流数据间的时间关系, 但是忽略了其中的空间关系; GCN 的方法能提取交通流的空间特征, 但是没有对时间特征进行处理. STGCN 模型能够同时捕获交通数据的时空特征, 但是这种方法依赖于大量的历史观测数据来训练模型, 因此, 在数据稀疏的交通路网中预测性能仍不佳.

对于结合了迁移学习的方法, DastNet 采用图嵌入和迁移学习的方式从源域获取知识, 利用到了源域丰富的知识, 因此该方法性能相对较好. 但是仍低于 TL-STGCN, 这是因为 TL-STGCN 利用了时间注意力机制能够更好地提取特征. MGAT 方法也结合了迁移学习的方法, 且采用了空间注意力机制, 但这种方法将采用普通卷积获取空

间特征, 提取到的特征不够准确, 因此精度仍低于 TL-STGCN 模型.

对于 TL-STGCN 模型的两个变种 S-only 和 TL-STGCN-noTA 而言, S-only 将本文的时空图卷积网络直接用于迁移学习任务, 由于源域和目标域间特征表示分布的差异较大, 因此在迁移学习任务中, 该方法性能最差. TL-STGCN-noTA 加入了本文的域适应模块, 没有加入时间注意力机制, 其预测精度较高, 这是因为通过域适应的方式能减小域间特征表示的差异, 从而更好地将源域中丰富的知识用于目标域中, 说明了本文域适应模块的有效性. TL-STGCN 结合了时间注意力机制, 能够更好地提取数据间的时间相关性, 由上述可知, TL-STGCN 模型对于数据稀缺的交通路网的预测问题, 有更好的预测效果.

表 4 TL-STGCN 与基线在评价指标上的性能对比

评价指标	目标域	源域	HA	GRU	GCN	STGCN	DastNet	MGAT	S-only	TL-STGCN-noTA	TL-STGCN
MAE	T-PEMS04	S-PEMS08					35.73	36.53	—	38.51	34.93
		S-METR-LA	74.91	43.86	43.72	42.99	29.05	28.82	35.32	28.95	28.41
		S-PEMS-BAY					26.98	26.63	32.47	27.01	26.35
	T-PEMS08	S-PEMS04					23.59	23.51	27.22	23.41	22.80
		S-PEMS-BAY	56.82	26.80	35.03	26.51	26.03	25.78	30.58	25.90	25.36
		S-METR-LA					26.81	25.48	27.95	25.61	25.15
	T-PEMS-BAY	S-PEMS04					4.68	4.62	4.97	4.72	4.50
		S-PEMS08	18.85	6.88	7.71	6.60	6.85	6.83	7.86	6.51	6.03
		S-METR-LA					6.42	6.43	8.58	6.30	6.19
	T-METR-LA	S-PEMS-BAY					7.86	7.90	7.93	8.11	7.74
		S-PEMS04	19.54	12.15	10.04	7.91	7.69	7.61	8.36	7.84	7.50
		S-PEMS08					7.90	7.57	7.98	7.59	7.42
MAPE (%)	T-PEMS04	S-PEMS08					32.01	32.36	—	31.61	31.46
		S-METR-LA	—	39.59	36.23	38.81	31.06	30.94	32.65	30.68	30.01
		S-PEMS-BAY					23.85	23.53	24.61	23.64	23.16
	T-PEMS08	S-PEMS04					36.76	36.89	36.74	37.98	36.08
		S-PEMS-BAY	—	38.36	40.73	40.48	38.63	38.12	38.67	37.93	37.61
		S-METR-LA					38.41	38.36	38.94	37.99	37.95
	T-PEMS-BAY	S-PEMS04					8.71	8.93	9.80	9.03	8.44
		S-PEMS08	—	15.00	16.29	13.96	11.12	10.98	12.64	10.86	10.69
		S-METR-LA					13.12	13.06	16.18	13.09	12.87
	T-METR-LA	S-PEMS-BAY					27.01	26.93	27.58	26.73	26.68
		S-PEMS04	—	35.41	33.37	26.91	26.03	25.83	27.69	25.81	25.38
		S-PEMS08					25.13	24.98	26.95	25.03	24.81
RMSE	T-PEMS04	S-PEMS08					51.25	50.95	—	55.02	50.16
		S-METR-LA	113.23	60.73	60.91	60.23	56.85	55.98	64.61	57.82	55.42
		S-PEMS-BAY					53.92	54.26	60.63	55.96	53.51
	T-PEMS08	S-PEMS04					36.98	36.72	43.60	37.20	36.14
		S-PEMS-BAY	80.39	42.29	58.16	41.98	40.21	39.78	46.88	39.87	39.04
		S-METR-LA					40.79	40.67	42.54	40.98	40.23
	T-PEMS-BAY	S-PEMS04					6.78	6.83	7.92	6.98	6.55
		S-PEMS08	16.90	9.61	10.76	9.21	7.03	6.98	7.35	6.99	6.68
		S-METR-LA					8.86	8.98	11.67	8.79	8.34
	T-METR-LA	S-PEMS-BAY					11.43	11.39	11.21	11.61	11.17
		S-PEMS04	24.14	14.20	13.47	11.19	10.69	10.56	10.99	10.39	10.26
		S-PEMS08					10.53	10.78	11.65	10.93	10.17

(2) 路网数据稀缺率的影响

为了进一步证明本文所提出的模型的有效性, 以及数据稀缺率对预测结果的影响, 在此对 30 min 的交通流预

测任务进行了目标稀缺率为 10%–40% 的实验. 为便于对比, 该实验选择 STGCN 模型和 GCN 模型、GRU 模型, 以及结合迁移学习的模型 DastNet 和 MGAT 来进行对比. 其中, STGCN、GCN 和 GRU 模型只在 T-PEMS08 目标域数据集上进行预测, 而 DastNet、MGAT 以及 TL-STGCN 在 S-PEMS04→T-PEMS08 通过迁移学习来预测. 实验结果如图 3 所示.

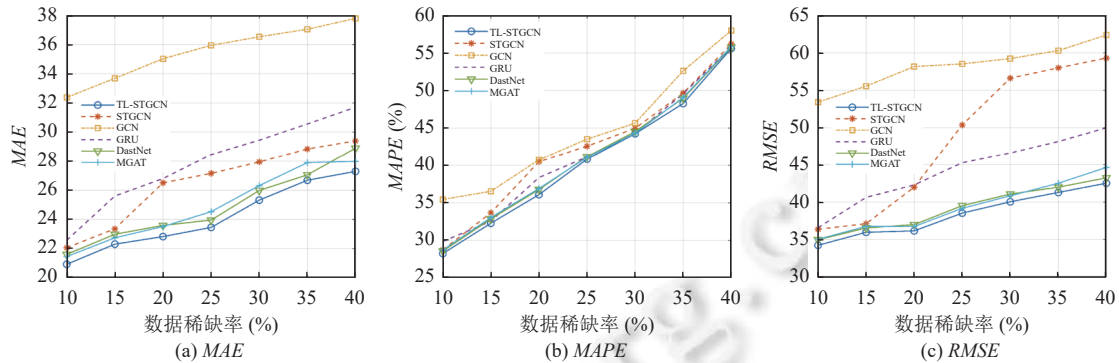


图 3 不同数据稀缺率下的 MAE , $MAPE$, $RMSE$

从图 3 中可以看出, 几种模型的 MAE 、 $RMSE$ 和 $MAPE$ 曲线随着数据稀缺率的上升而上升, 这说明历史观测数据质量越低, 其预测性能越差. 此外, 在数据稀缺率不同的情况下, 相较于 STGCN, TL-STGCN 的 MAE 、 $MAPE$ 、 $RMSE$ 分别平均下降了 8.9%、3.6%、20.9%; 和 GCN 相比, TL-STGCN 的 MAE 、 $MAPE$ 、 $RMSE$ 分别平均降低了 32.1%、8.6%、34.1%; 和 GRU 相比, TL-STGCN 的 3 个误差指标分别降低了 13.5%、2.1%、13.2%. TL-STGCN 相比 DastNet, 其 MAE 平均降低了 3.0%, $MAPE$ 平均降低了 1.1%, $RMSE$ 平均下降了 2.1%. 相较于另一种针对数据稀缺场景的预测方法 MGAT 而言, TL-STGCN 的 MAE 、 $MAPE$ 、 $RMSE$ 指标分别平均降低了 3.2%、1.2%、1.9%. 因此, 可得知随着数据稀缺率的上升, TL-STGCN 的性能仍旧高于基线实验.

(3) 训练集大小影响

为了研究训练集大小对模型性能的影响, 本文在训练集条件不同的情况下进行迁移学习任务 S-PEMS04→T-PEMS08 以预测目标域未来 30 min 交通流.

首先研究源域训练集大小对预测结果的影响.

此时, 将目标域 T-PEMS08 的训练集大小固定为 11 天, 数据稀缺率定为 20%. 为了比较不同源域训练集大小对模型性能的影响, 本文选择源域训练集天数分别为 10 天、20 天、30 天、40 天、50 天, 采用 MAE 和 $RMSE$ 指标对模型进行评估. 图 4 为不同源域训练集大小下的 MAE 和 $RMSE$ 结果, 可知源域训练集越大, TL-STGCN 的性能越好. 这是因为随着源域训练集增大, 可以从源域迁移更多域不变的知识.

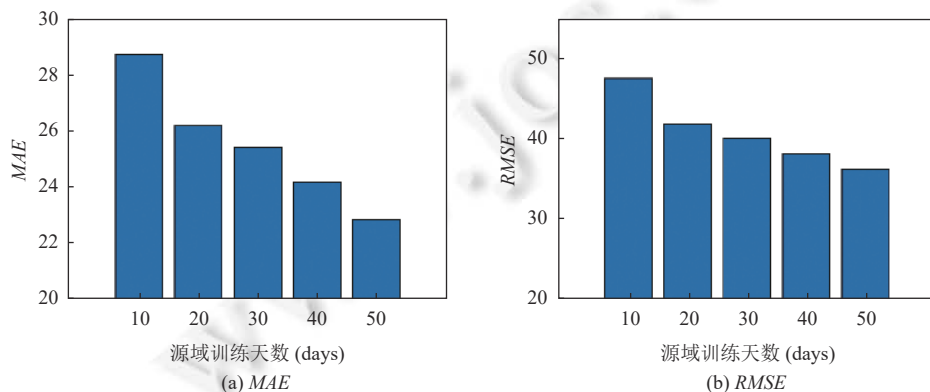


图 4 不同源域训练集大小的 MAE , $RMSE$

然后研究目标域训练集大小对预测结果的影响. 在此将源域 S-PEMS04 的训练集大小设置为 56 天, 目标域 T-PEMS08 的数据稀缺率固定为 20%. 将目标域训练集大小分别设置为 2 天、5 天、8 天、11 天来观察目标域训练集大小对模型性能的影响. 图 5 为 MAE 和 RMSE 的评价结果. 从图 4 中可以看出, 目标训练集越大, TL-STGCN 模型的性能就越好. 这是因为随着目标域数据的增多, 可以得到更多目标域数的时空特征, 并将其整合到迁移学习过程中, 更好地学习域不变的特征表示.

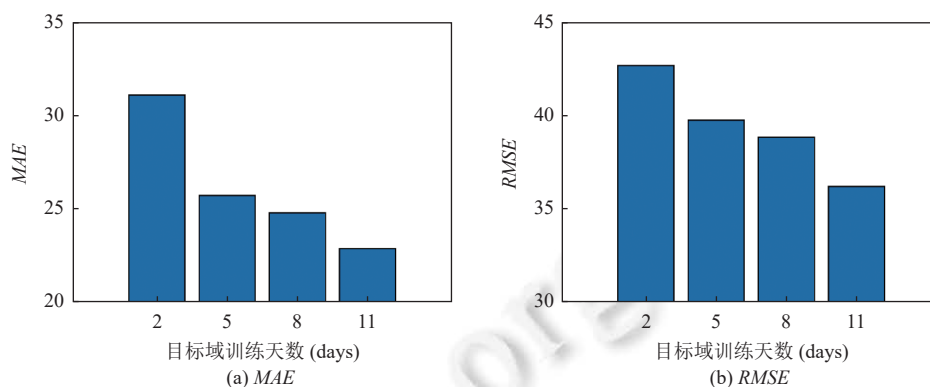


图 5 不同目标域训练集大小的 MAE, RMSE

5 总结

针对数据稀缺、数据质量不高的交通路网的交通流预测问题, 本文提出了一种基于时空图卷积网络的迁移学习模型 TL-STGCN. TL-STGCN 从拥有大量优质数据的源交通路网中学习可迁移知识, 以辅助数据稀缺的目标交通路网完成交通流预测任务. 模型中, 结合时间注意力机制和时空图卷积模块构建了一个时空卷积网络, 从源域和目标域中提取交通流数据的时空特征表示. 将时空卷积网络作为特征提取模块, 与域判别器、流量预测模块结合, 通过对抗性域适应的方法学习两个交通路网域间有差异性的特征表示和域不变的特征表示, 完成目标域的预测任务. 本文在 PEMS04、PEMS08、PEMS-BAY、METR-LA 这 4 个真实的数据集上进行了实验. 实验结果表明, 对于数据稀缺的交通路网的交通流预测问题, TL-STGCN 相较于其他模型在性能上有较大的提升. 在未来的工作中考虑外界因素 (天气、节假日等) 对交通流的影响, 将引入更多额外的因素作为特征; 且考虑到不同交通路网间数据可能不共享的问题, 如何结合联邦学习进行训练也是将来的研究重点.

References:

- [1] Yuan HT, Li GL. A survey of traffic prediction: From spatio-temporal data to intelligent transportation. *Data Science and Engineering*, 2021, 6(1): 63–85. [doi: [10.1007/s41019-020-00151-z](https://doi.org/10.1007/s41019-020-00151-z)]
- [2] Saleem TJ, Chishti MA. Deep learning for the Internet of Things: Potential benefits and use-cases. *Digital Communications and Networks*, 2021, 7(4): 526–542. [doi: [10.1016/j.dcan.2020.12.002](https://doi.org/10.1016/j.dcan.2020.12.002)]
- [3] Cheng XY, Zhang RQ, Zhou J, Xu W. DeepTransport: Learning spatial-temporal dependency for traffic condition forecasting. In: *Proc. of the 2018 Int'l Joint Conf. on Neural Networks (IJCNN)*. Rio de Janeiro: IEEE, 2018. 1–8. [doi: [10.1109/IJCNN.2018.8489600](https://doi.org/10.1109/IJCNN.2018.8489600)]
- [4] Zhuang WQ, Cao YB. Short-term traffic flow prediction based on CNN-BiLSTM with multicomponent information. *Applied Sciences*, 2022, 12(17): 8714. [doi: [10.3390/app12178714](https://doi.org/10.3390/app12178714)]
- [5] Guo K, Hu YL, Qian Z, Hao L, Zhang K, Sun YF, Gao JB, Yin BC. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Trans. on Intelligent Transportation Systems*, 2021, 22(2): 1138–1149. [doi: [10.1109/TITS.2019.2963722](https://doi.org/10.1109/TITS.2019.2963722)]
- [6] Cui ZY, Henrickson K, Ke RM, Wang YH. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Trans. on Intelligent Transportation Systems*, 2020, 21(11): 4883–4894. [doi: [10.1109/TITS.2019.2950416](https://doi.org/10.1109/TITS.2019.2950416)]
- [7] Feng N, Guo SN, Song C, Zhu QC, Wan HY. Multi-component spatial-temporal graph convolution networks for traffic flow forecasting.

- Ruan Jian Xue Bao/Journal of Software, 2019, 30(3): 759–769 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5697.htm> [doi: 10.13328/j.cnki.jos.005697]
- [8] Yu B, Yin HT, Zhu ZX. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence. Stockholm: AAAI Press, 2018. 2634–2640.
 - [9] Wang SZ, Miao H, Li JY, Cao JN. Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks. IEEE Trans. on Intelligent Transportation Systems, 2022, 23(5): 4695–4705. [doi: 10.1109/TITS.2021.3055207]
 - [10] Li JY, Guo FC, Sivakumar A, Dong YJ, Krishnan R. Transferability improvement in short-term traffic prediction using stacked LSTM network. Transportation Research Part C: Emerging Technologies, 2021, 124: 102977. [doi: 10.1016/j.trc.2021.102977]
 - [11] Wan XB, Liu H, Xu H, Zhang XC. Network traffic prediction based on LSTM and transfer learning. IEEE Access, 2022, 10: 86181–86190. [doi: 10.1109/ACCESS.2022.3199372]
 - [12] Wang LY, Geng X, Ma XJ, Liu F, Yang Q. Cross-city transfer learning for deep spatio-temporal prediction. arXiv:1802.00386, 2018.
 - [13] Tang YH, Qu A, Chow AHF, Lam WHK, Wong SC, Ma W. Domain adversarial spatial-temporal network: A transferable framework for short-term traffic forecasting across cities. In: Proc. of the 31st ACM Int'l Conf. on Information & Knowledge Management. Atlanta: ACM, 2022. 1905–1915. [doi: 10.1145/3511808.3557294]
 - [14] Mou LT, Zhao PF, Xie HT, Chen YY. T-LSTM: A long short-term memory neural network enhanced by temporal information for traffic flow prediction. IEEE Access, 2019, 7: 98053–98060. [doi: 10.1109/ACCESS.2019.2929692]
 - [15] Poonia P, Jain VK. Short-term traffic flow prediction: Using LSTM. In: Proc. of the 2020 Int'l Conf. on Emerging Trends in Communication, Control and Computing (ICONC3). Lakshmanarh: IEEE, 2020. 1–4. [doi: 10.1109/ICONC345789.2020.9117329]
 - [16] Zhang D, Kabuka MR. Combining weather condition data to predict traffic flow: A GRU-based deep learning approach. IET Intelligent Transport Systems, 2018, 12(7): 578–585. [doi: 10.1049/iet-its.2017.0313]
 - [17] Ma XL, Dai Z, He ZB, Ma JH, Wang Y, Wang YP. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. Sensors, 2017, 17(4): 818. [doi: 10.3390/s17040818]
 - [18] Ye JX, Zhao JJ, Ye KJ, Xu CZ. How to build a graph-based deep learning architecture in traffic domain: A survey. IEEE Trans. on Intelligent Transportation Systems, 2022, 23(5): 3904–3924. [doi: 10.1109/TITS.2020.3043250]
 - [19] Wu B, Liang X, Zhang SS, Xu R. Advances and applications in graph neural network. Chinese Journal of Computers, 2022, 45(1): 35–68 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2022.00035]
 - [20] Zhao L, Song YJ, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li HF. T-GCN: A temporal graph convolutional network for traffic prediction. IEEE Trans. on Intelligent Transportation Systems, 2020, 21(9): 3848–3858. [doi: 10.1109/TITS.2019.2935152]
 - [21] Zhang ZC, Li M, Lin X, Wang YH, He F. Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. Transportation Research Part C: Emerging Technologies, 2019, 105: 297–322. [doi: 10.1016/j.trc.2019.05.039]
 - [22] Guo SN, Lin YF, Feng N, Song C, Wan HY. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence. Honolulu: AAAI, 2019. 922–929. [doi: 10.1609/aaai.v33i01.3301922]
 - [23] Zhao WZ, Yuan G, Zhang YM, Qiao SJ, Wang SZ, Zhang L. Multi-view fused spatial-temporal dynamic GCN for urban traffic flow prediction. Ruan Jian Xue Bao/Journal of Software, 2024, 35(4): 1751–1773 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/7018.htm> [doi: 10.13328/j.cnki.jos.007018]
 - [24] Zhuang FZ, Qi ZY, Duan KY, Xi DB, Zhu YC, Zhu HS, Xiong H, He Q. A comprehensive survey on transfer learning. Proc. of the IEEE, 2021, 109(1): 43–76. [doi: 10.1109/JPROC.2020.3004555]
 - [25] Ren Y, Chen X, Wan S, Xie KQ, Bian KG. Passenger flow prediction in traffic system based on deep neural networks and transfer learning method. In: Proc. of the 4th Int'l Conf. on Intelligent Transportation Engineering (ICITE). Singapore: IEEE, 2019. 115–120. [doi: 10.1109/ICITE.2019.8880220]
 - [26] Shen J, Qu YR, Zhang WN, Yu Y. Wasserstein distance guided representation learning for domain adaptation. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 4058–4065. [doi: 10.1609/aaai.v32i1.11784]
 - [27] Hussain B, Afzal MK, Ahmad S, Mostafa AM. Intelligent traffic flow prediction using optimized GRU model. IEEE Access, 2021, 9: 100736–100746. [doi: 10.1109/ACCESS.2021.3097141]
 - [28] Wu F, Zhang TY, de Souza AH Jr, Fifty C, Yu T, Weinberger KQ. Simplifying graph convolutional networks. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: ICML, 2019. 6861–6871.
 - [29] Mo JQ, Gong ZG. Cross-city multi-granular adaptive transfer learning for traffic flow prediction. IEEE Trans. on Knowledge and Data Engineering, 2023, 35(11): 11246–11258. [doi: 10.1109/TKDE.2022.3232185]

附中文参考文献:

[7] 冯宁, 郭晟楠, 宋超, 朱琪超, 万怀宇. 面向交通流量预测的多组件时空图卷积网络. 软件学报, 2019, 30(3): 759–769. <http://www.jos.org.cn/1000-9825/5697.htm> [doi: 10.13328/j.cnki.jos.005697]

[19] 吴博, 梁循, 张树森, 徐睿. 图神经网络前沿进展与应用. 计算机学报, 2022, 45(1): 35–68. [doi: 10.11897/SP.J.1016.2022.00035]

[23] 赵文竹, 袁冠, 张艳梅, 乔少杰, 王森章, 张雷. 多视角融合的时空动态 GCN 城市交通流量预测. 软件学报, 2024, 35(4): 1751–1773. <http://www.jos.org.cn/1000-9825/7018.htm> [doi: 10.13328/j.cnki.jos.007018]



李云(1974—), 男, 博士, 教授, 博士生导师, 主要研究领域为云计算, 边缘智能, 无线网络资源管理.



夏士超(1991—), 男, 博士, 讲师, CCF 专业会员, 主要研究领域为云计算, 边缘智能, 无线网络资源管理.



高雅(1999—), 女, 硕士生, 主要研究领域为时空数据挖掘, 深度学习.



吴广富(1980—), 男, 博士, 高级工程师, 主要研究领域为移动通信, 图像智能处理.



姚枝秀(1994—), 女, 博士生, 主要研究领域为移动边缘计算, 无线网络资源管理.