

图神经网络驱动的交通预测技术：探索与挑战

周毅¹, 胡姝婷¹, 李伟¹, 承楠², 路宁³, 沈学民⁴

(1. 河南大学, 河南 郑州 450046; 2. 西安电子科技大学, 陕西 西安 710071;
3. 加拿大女王大学, 加拿大 金斯顿 K7L 3N6; 4. 滑铁卢大学, 加拿大 滑铁卢 N2L 3G1)

摘 要: 随着物联网及人工智能技术的快速发展, 对交通数据进行精准的分析 and 预测成为智慧交通的首要环节。近年来, 交通预测方法逐渐从经典的模型驱动转变为数据驱动, 然而, 如何通过大数据有效分析路网的时空特性是预测过程中面临的关键难题之一。时空大数据分析是交通预测的利器, 将交通路网建模为图网络, 将深度学习方法在图网络上进行扩展, 通过图神经网络建立时空预测模型, 采用图卷积的方式有效地获取路网传感器节点之间的时空相关性, 可以显著提高交通预测模型的精度。针对图神经网络驱动的交通预测技术进行了探索, 基于深度时空特性分析提炼了两大类交通预测模型, 并通过实例进行分析和验证, 探讨了图神经网络在交通预测领域的技术优势和主要挑战, 挖掘了图神经网络预测机制的潜在研究方向。

关键词: 交通预测; 图神经网络; 时空相关性; 同步卷积; 图注意力网络

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2021.00235

Graph neural network driven traffic prediction technology: review and challenge

ZHOU Yi¹, HU Shuting¹, LI Wei¹, CHENG Nan², LU Ning³, SHEN Xuemin (Sherman)⁴

1. Henan University, Zhengzhou 450046, China

2. Xidian University, Xi'an 710071, China

3. Queen's University, Kingston K7L 3N6, Canada

4. University of Waterloo, Waterloo N2L 3G1, Canada

Abstract: With the rapid development of Internet of things and artificial intelligence technology, accurate analysis and prediction of traffic data have become the primary target of intelligent transportations. In recent years, the method of traffic forecasting has gradually changed from the classical model-driven type to the data-driven type. However, how to effectively analyze the spatial-temporal characteristics of road networks through big data is one of the key issues in the traffic prediction process. Spatiotemporal big data analysis is a powerful tool for the traffic prediction. The traffic network can be modeled as a graph network, while the deep learning method can be extended on the graph network. Utilizing graph neural networks, we can build the spatiotemporal prediction model, and obtain the spatial-temporal correlation between the sensor nodes in road networks effectively by using graph convolution, which can significantly improve the accuracy of traffic prediction models. The traffic forecasting technology driven by graph neural networks was explored, and two kinds of traffic prediction models based on the analysis of deep spatial-temporal characteristics were extracted. The actual cases were analyzed and evaluated to discuss the technical advantages and key challenges of graph neural networks in the traffic prediction. The potential issues of graph neural network driven prediction mechanisms were also excavated.

Key words: traffic prediction, graph neural networks, spatial-temporal correlation, synchronous convolution, graph attention networks

收稿日期: 2021-04-06; 修回日期: 2021-07-24

通信作者: 承楠, nancheng@xidian.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62176088, No.61701170, No.62071356); 河南省科技攻关项目 (No.202102310198, No.212102210412)

Foundation Items: The National Natural Science Foundation of China (No.62176088, No.61701170, No.62071356), The Henan Science and Technology Development Program (No.202102310198, No.212102210412)

1 引言

近年来,城市车辆数量的急增引发了交通事故、交通拥堵等一系列安全问题,提高通行效率是城市交通发展的重要方向。在物联网时代的大背景下,城市交通逐渐向智能化、网联化^[1]、绿色化等方向发展。智能交通系统(ITS, intelligent transportation system)充分利用人工智能、边缘计算^[2]、物联网、传感网、移动通信^[3]等前沿技术,可实现车辆对道路交通态势的智慧感知及人、车、路等要素的高效协同^[4]。而交通预测是实现 ITS 的首要环节,也是助力交通控制、路径优化、车辆调度等各种应用的必要条件。交通事故、高峰拥堵等事件的频繁发生使交通状况随时间动态变化,导致了时间序列的非平稳性;另外,路网上布置的传感器之间存在着复杂的空间依赖性,即使两个传感器在欧几里得空间中的距离较近,也可能呈现完全不同的交通状态。因此,要获得准确的交通预测结果,必须同时解决时间相关性和空间相关性这两大交通预测挑战。

交通预测方法分类及优缺点见表 1。典型的交通预测方法主要有两大类,即模型驱动方法和数据驱动方法。模型驱动方法主要研究流量、速度和密度之间的瞬时和稳态关系,其主要依赖先验知识进行系统建模^[5],如交通速度模型、排队论模型等。但实际道路中的交通态势易受交通事故、天气状况等多种因素的影响,很难用固定的模型模拟实际的交通状况。因此,随着交通物联网及大数据技术的快速发展,研究人员逐渐将注意力转向数据驱动方法。数据驱动方法主要根据数理统计分析法进行预测^[6],早期有历史平均分析预测法,它将历史数据的平均值作为最终预测值,这种方法虽然计算简单但预测精度较低。随后,又出现了统计分析和机器学习两大类数据驱动预测模型。

经典的统计分析模型对历史数据进行分析,然后利用回归函数进行预测,主要有线性回归预测法、自回归综合移动平均(ARIMA, auto-regressive integrated moving average)法^[7]、卡尔曼滤波分析预测法^[8]等。统计分析法虽然算法和模型简易,实现方便,但其主要建立在时间序列数据平稳的假设前提下。由于交通数据具有极强的动态性和高度非线性,统计分析模型不能有效地分析未来交通状况的发展态势。机器学习模型通过提取与交通相关的

特征并对模型进行训练,可以用于非线性数据的模型构建,同时降低预测误差。常用的机器学习模型有 K 近邻、支持向量回归、人工神经网络等。Dell'Acqua 等^[9]设计了一种时间感知多元近邻回归的预测方法。Luo 等^[10]将离散傅里叶变换和支持向量回归相结合,提出了一种混合预测方法。传统的机器学习模型依赖于人工提取的交通特征,虽然可以对非线性关系进行建模,但在捕捉时间特征时,其模型架构浅易、参数有限且计算效率较低致使其仅限于交通状况简单或交通数据规模较小的场景。近年来,前馈神经网络、深度信念网络、递归神经网络(RNN, recurrent neural network)等深度学习模型被广泛用于各种复杂关系的预测任务中,RNN 作为一种特殊设计的神经网络,能够学习交通数据的长期时间相关性。Duan 等^[11]将 RNN 的变体长短期记忆网络(LSTM, long short-term memory)用于出行时间的预测。Wang 等^[12]结合 LSTM 和高斯过程回归对网络流量负荷进行预测。然而这些模型只考虑了交通大数据的时间特性,对路网结构中存在的空间特性考虑不足。

随着深度学习在图网络领域的发展,从时空特征的分析角度研究交通预测势在必行。卷积神经网络(CNN, convolutional neural network)和 RNN 是提取时空相关性的典型方法。Han 等^[13]提出了一种大规模路网中短期预测的深度学习框架,通过基于 CNN 的深度聚类方法提取道路的形状特征,并在此基础上建立了基于 RNN 的深度预测模型。但 CNN 只能处理欧几里得域中的图像数据,路网符合一个非欧几里得空间的数据结构,这必然限制了 CNN 的能力。图神经网络(GNN, graph neural network)^[14-17]是一种在图域上运行的基于深度学习的扩展方法,它能够从底层数据中提取出网络中复杂的关联性,且在自然语言处理、推荐系统等任务上表现良好。随后 GNN 逐渐被用于交通领域,GNN 分为图卷积网络(GCN, graph convolutional network)、图注意力网络(GAT, graph attention network)、图自动编码器、图生成网络和图时空网络等五大类^[15]。Kipf 等^[18]采用了图卷积网络对图数据进行半监督学习,但该方法仍然不能同时捕获路网的时空特征及动态相关性。Li 等^[19]将道路结构整合为一张有向的路网拓扑图,提出了一种扩散卷积递归神经网络(DCRNN, diffusion convolutional recurrent neural network)模型,模型使用扩散卷积网络和门控递归

单元 (GRU, gated recurrent unit) 分别捕获路网中的空间相关性和时间相关性。DCRNN 和时空图卷积网络 (STGCN, spatio-temporal graph convolutional network) [20] 作为最早的基于图的交通预测模型, 通过结合 GCN 和 RNN 模型捕获路网的时空相关性。随后, 越来越多的研究倾向于用图神经网络建模复杂的、非线性的、具有时空相关性的路网结构。Guo 等 [21] 提出了一种优化的图卷积递归神经网络, 在训练阶段通过数据驱动的方式学习一个优化图, 利用交通数据揭示路段之间潜在的关系。由于图卷积赋予节点的权重矩阵是固定的, Cirstea 等 [22] 在 DCRNN 的基础上引入注意力机制获得节点的自适应权重矩阵, 进一步提高了预测精度, 充分表明基于图神经网络的交通预测模型的有效性。

尽管上述基于图神经网络的预测模型通过获取路网的时空相关性达到了较优的预测效果, 但它们忽略了交通数据极强的周期规律性。例如, 每个工作日的交通状况、每周整体的交通状况都具有一定的相似性。采用神经网络 (RNN、多层感知器等) 对这种周期规律性进行特征提取, 然后与时空特征进行融合作为时空预测模型的输入将有利于提高预测精度。文献 [23] 分析了预测目标近期、日周期、周周期的 3 种时间维特性, 构建了一个多组件时空图卷积网络架构, 3 个组件的输入分别经过多个时空卷积块和一个全连接层, 经融合输出信息得到最终预测结果, 其中每个时空卷积块都包含空间和时间两个维度的图卷积。为了提高预测精度, Guo 等 [24] 在多组件图卷积网络的基础上又加入了注意力机制, 构成一种基于注意力机制的时空图卷积网络 (ASTGCN, attention based spatial-temporal graph convolutional network), 由时空注意力模块和时空卷积模块共同构成时空块, 进一步降低了预测误差。Zhou 等 [25] 同时考虑路网的时空变量及周期规

律性, 通过融合时空变量和外部特征向量 (日周期、周周期) 构造输入向量, 输入到由注意力机制和门控递归单元为基本架构的编码器模型, 捕获时空相关性; 为了发现历史时间序列的本质隐藏状态, 在解码过程中也加入了注意力机制, 结果表明注意力机制能够有效地提高预测精度。Shi 等 [26] 设计了一种获取动态空间关联性的注意力机制, 可以有效捕获路网的空间、时间及长期周期相关性。然而, 交通事故等突发事件也会导致周期规律性存在偏差。Zhao 等 [27] 考虑了非周期性拥堵下的预测, 提出了一种优化的 GRU 算法, 并分别在工作日、周末、雨天、事故 4 种场景下验证了模型的有效性, 但其缺乏对路网时空相关性的考虑。

基于图神经网络的交通预测模型从时空相关性的角度对路网特征进行分析。基于图神经网络的预测模型见表 2, 可以看出, 在 GNN 的基础上, 研究者又引入了随机扩散、注意力、膨胀因果卷积 (DCC, dilated causal convolution) 等方法提高预测精度。虽然上述方法已经展现了优越的性能, 但仍存在一些亟待解决的问题。许多工作侧重于多步预测, 常用的方法有全连接层 (FC, fully connected layer)、序列到序列 (Seq2Seq, sequence to sequence) 模型 [37] 等。FC 作为输出层, 是获得所需输出形状最简单的方法 [23,29]。而大多数基于图的交通预测方法都依赖于 Seq2Seq 模型架构 [19,25,32], 它主要包含两个组件: 一个用于读取输入序列的编码器和一个根据编码器的输出产生最终输出序列的解码器, 且通常选用 RNN 作为编码器和解码器的内部结构。然而, 当前的 Seq2Seq 模型存在一个重要缺陷, 在训练过程中, 模型的损失由真实样本即传感器监测到的真实数据监督, 解码器根据真实样本中当前时刻的真实值计算下一时刻的输出 [38]。但是在测试的过程中, 模型将当前时刻得到的预测值作为下一时刻

表 1 交通预测方法分类及优缺点

方法	优点	缺点
模型驱动	主要研究流量、速度和密度之间的瞬时和稳态关系, 主要依赖先验知识进行系统建模	模型固定, 难以准确地模拟变化多端的真实交通状况
数据驱动		
统计分析模型	算法和模型简易, 实现方便	模型建立在时间序列数据平稳的假设前提下
机器学习模型	可以提取与交通相关的特征, 用于非线性数据的模型构建	依赖于人工提取的交通特征, 模型架构浅易、参数有限且计算效率较低
深度学习模型	能够让计算机自动学习路网特征, 减少对人工提取特征的依赖性	只考虑路网空间或时间特征
图神经网络+深度学习	从时间和空间两个角度提取路网特征	学习时间相关性效率不高, 长期预测精度有待进一步提高

表 2 基于图神经网络的预测模型

文献	空间模型	时间模型	数据集
[5]	GCN	GRU	SZ-taxi; Los-loop
[28]	GCN	LSTM	出租车的 GPS 数据 (上海 TIC)
[23]	GCN	标准二维卷积	PeMSD4; PeMSD8
[20]	GCN	门控 CNN	BJER4; PeMSD7
[29]	GCN	DCC	PeMSD4; PeMSD7
[30]	GCN	门控 TCN	METR-LA; PEMS-BAY
[31]	考虑高阶邻域的 GCN	LSTM	LOOP; INRIX
[32]	GCN	GRU+注意力	A-map
[19]	扩散卷积网络	GRU	METR-LA; PEMS-BAY
[21]	基于优化图矩阵的 GCN	GRU	D.C.; Philadelphia; PeMSD4
[22]	GAT	GRU	METR-LA
[24]	GCN+注意力	标准二维卷积+注意力	PeMSD4; PeMSD8
[26]	注意力	LSTM+注意力	PeMSD4; PeMSD8
[33]	GAT	LSTM	PeMSD7
[34]	注意力	注意力	Xiamen; PeMS
[35]	时空同步卷积	时空同步卷积	PEMS03; PEMS04; PEMS07; PEMS08
[36]	注意力	注意力	METR-LA; PEMS-BAY

的输入。训练时解码器的输入来自真实数据分布，而测试时解码器的输入来自模型分布，这会导致模型在训练和测试过程中产生的下一个决策存在差异，这个过程产生的误差便会迅速累积，这种差异又称为曝光误差 (exposure bias) [39]。

为了避免曝光误差的影响，Daumé 等 [40] 首次提出在训练过程中减少对真实样本的依赖，令模型在训练的过程中使用自身的预测结果进行下一步预测。Li 等 [19] 使用调度采样 (scheduled sampling) 的方法，通过采样随机决定用概率为 ε 的真实数据，或概率为 $1 - \varepsilon$ 的预测值作为模型的输入。虽然调度采样可以简单地缓解曝光误差问题，但它在地面真实值或模型输出之间随机选择，大大增加了误差反向传播的复杂度。Yu 等 [41] 采用生成对抗网络避免曝光误差问题的发生。生成对抗网络由生成器和判别器组成，生成器依据输入的真实样本尝试生成与之相似的数据，然后经过判别器识别生成数据与真实样本是否接近，鉴别完成后判别器会反馈给生成器一个奖励值，生成器根据奖励值的大小优化模型参数以生成更接近真实样本的数据。生成对抗网络类似于调度采样，在模型训练时依赖模型自身的输出避免对真实值的依赖，从而缓解曝光误差问题。然而，生成对抗网络没有损失函数的监督，在训练

过程中很难判断是否正在取得进展。随后，Keneshloo 等 [42] 将强化学习用于 Seq2Seq 模型。Zhou 等 [25] 用强化学习中的蒙特卡罗策略梯度算法挖掘出相关性最大的信息，用自身的预测值训练模型，而不仅仅依赖数据分布，这在很大程度上缓解了 Seq2Seq 模型中的曝光误差问题。当然，强化学习还包含 Actor-Critic、深度确定性策略梯度 (DDPG, deep deterministic policy gradient) 等优于蒙特卡罗策略梯度的算法，如何用它们缓解曝光误差问题值得进一步研究。

上述是本文总结的交通预测的研究现状，当然在本文之前也有许多类似的工作。例如，Lana 等 [43] 总结了交通预测领域的最新技术成果，并对尚未解决的主要技术挑战进行了分析。Nagy 等 [44] 概述了现有的数据来源，详细介绍了如何使用预测方法促进交通流动性。Zhang 等 [45] 主要对短期交通流预测中的 ARIMA 模型、RNN 模型和稀疏自动编码器这 3 类方法做了总结。文献 [46] 对时空流量预测方法在解决城市交通问题上的发展现状进行了总结。Ye 等 [47] 研究了多种基于图的深度学习架构在各种交通问题中的应用，总结了常见的交通问题及其相应的图深度学习解决方案。而本文主要探索了基于图神经网络的交通预测方法及其挑战，图神经网络驱

动的交通预测研究框架如图1所示。本文总结了交通预测的发展背景及研究现状,在此基础上分析了通过图神经网络进行预测比经典方法更能充分利用路网空间相关性的优势。同时,本文综述了用图神经网络捕获路网时间相关性、空间相关性及时空相关性的几种典型方法,并用经典的DCRNN时空模型进行了实际交通数据的预测分析与验证。另外,本文还从间接捕获和直接捕获两个角度提炼了图神经网络驱动下的交通预测模型。最后,深入挖掘了交通预测技术中潜在的研究方向。

2 时空交通预测模型

图是一种由节点和边构成的复杂数据结构,在现实世界中处处可见,如社交网络、知识图、化学分子结构等^[48]。图可以表示来自不同学科的各种不同类型、存在相互关联的数据。交通路网包含错综复杂的连接关系,因此,路网也可以用一个有向图 $G=(V, E, A)$ 描述。其中, V 表示路网中 N 个节点的集合,节点可以是路网上部署的传感器或道路交叉口等(本文以传感器为例); E 表示边的集合,若节点 v_i 与节点 v_j 相邻,则它们之间存在边 e_{ij} ; $A \in \mathbf{R}^{N \times N}$ 表示路网拓扑图的邻接矩阵。用 A_{ij} 表示邻接矩阵 A 中第 i 行、第 j 列的元素,若节点 v_i 与节点 v_j 相邻,则 $A_{ij}=1$,否则 $A_{ij}=0$ 。将路网上观测到的交通信息(即交通流、交通速度等状态信息)作为节点的属性特征,用 $\mathbf{X} \in \mathbf{R}^{N \times P}$ 表示图 G 中节点

的特征矩阵, P 表示节点属性特征的数量, N 表示节点的数量。给定一个路网,首先提取道路中所有的传感器作为节点,然后根据传感器之间的连接创建节点与节点之间的边,从而得到整个路网的拓扑图。交通预测本质是学习一个映射函数 $f(\cdot)$,通过给定的道路结构和 T 个时间步长的历史交通状态,映射出未来 T' 个时间步长的交通状态,即

$$[\hat{y}_1, \dots, \hat{y}_{T'}] = f([\mathbf{X}_1, \dots, \mathbf{X}_T], G) \quad (1)$$

其中, $\mathbf{X}_1, \dots, \mathbf{X}_T$ 表示历史交通数据的时间序列, G 表示路网拓扑图, $\hat{y}_{T+1}, \dots, \hat{y}_{T+T'}$ 表示预测出的未来 T' 个时间步的交通态势。为了方便表达,本文在陈述过程中省略 T ,用 $\hat{y}_1, \dots, \hat{y}_{T'}$ 代替 $\hat{y}_{T+1}, \dots, \hat{y}_{T+T'}$ 。所以,交通预测问题可以描述为基于路网拓扑结构和历史时间序列数据预测未来的交通状态。而如何获取路网的空间特征和时间特征也相应成为了交通预测中亟待解决的难题。

基于图神经网络的时空交通预测模型如图2所示。整个模型建立在Seq2Seq的基础上,将传感器作为节点,相邻传感器之间的距离关系作为边,根据不同传感器所处的地理位置和它们之间的距离构建路网拓扑图。路网的空间相关性主要受路网中边的特征影响,表现为相邻节点之间的影响。而时间相关性主要受节点自身特征的影响,表现为同一传感器节点随着时间的推移所采集的交通数据的变化情况。在编码器中,将历史的时间序列数据 $\mathbf{X}_1, \dots, \mathbf{X}_T$ 输入GNN,通过邻域聚合的方式分别在

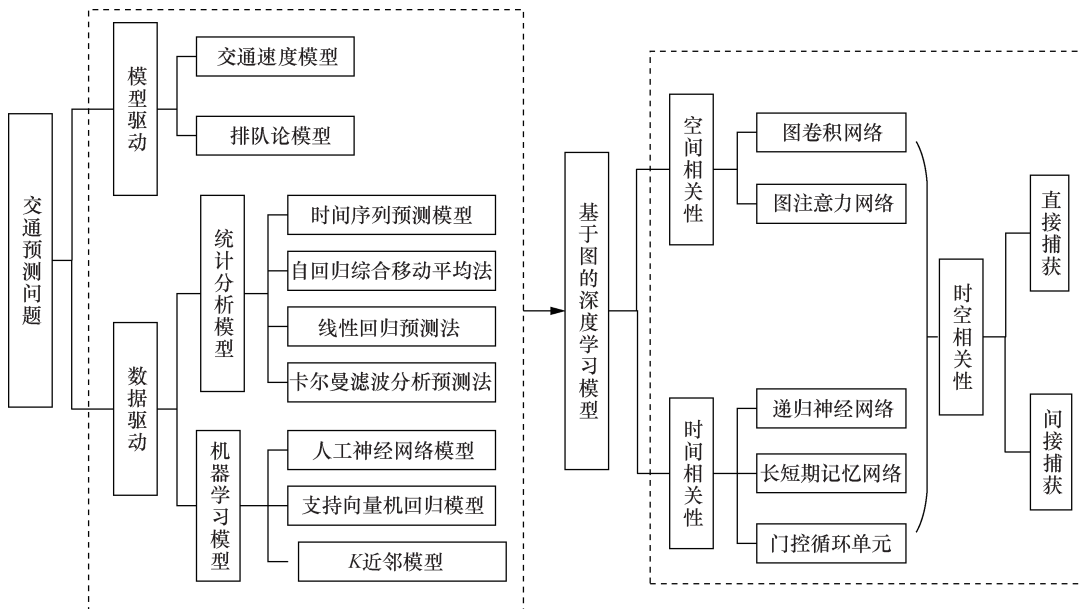


图1 图神经网络驱动的交通预测研究框架

路网拓扑结构中的每个传感器节点上进行传播,并递归聚合邻域节点的状态信息来更新每个传感器节点的隐藏状态,从而捕获路网的空间相关性。再将带有空间相关性的时间序列输入 RNN 来捕获路网的时空相关性。用编码器的最后一个隐藏状态 h_T 初始化解码器,再根据编码器编码的上下文向量 c_T 及 T 时刻的交通状况 X_T 预测 $T+1$ 时刻的交通状况 \hat{y}_1 。

本节阐明了图神经网络驱动下的时空交通预测模型,主要通过结合 GNN 和 RNN 捕获路网的时空特征。下面将分析和探索如何使用 GNN 和 RNN 模型捕获路网的时间相关性和空间相关性。

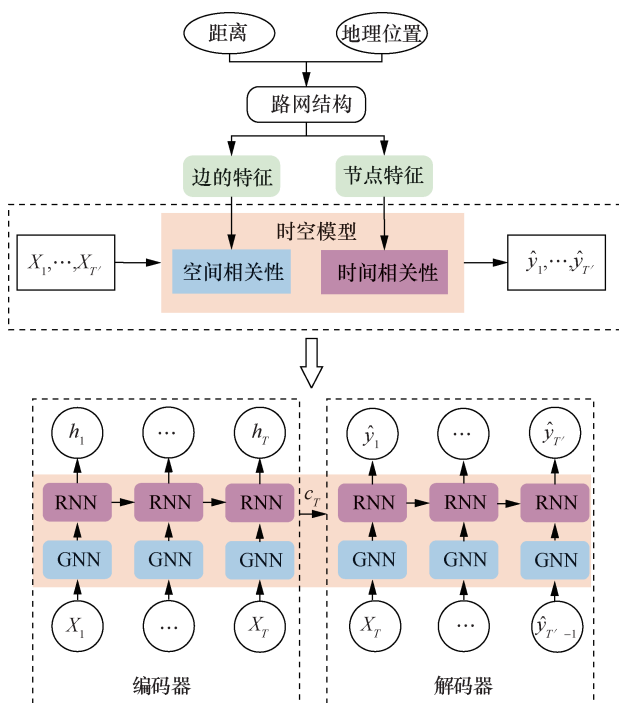


图2 基于图神经网络的时空交通预测模型

3 时间相关性

路网的时间相关性表现为某一传感器在 t 时刻的交通状态将对其在 $t+1$ 时刻的状态造成影响, $t+2$ 时刻的状态也会间接受到影响。也即, 每个传感器节点当前时刻的交通数据受之前交通数据的影响, 同时也影响着当前时刻之后的交通数据, 局部路网时空示意图如图3所示。例如, 某条道路 18:00 发生了拥堵, 则此拥堵将有可能持续到 18:30。不仅如此, 路网的时间相关性还包含日周期、周周期等特性, 如每个工作日的上下班高峰期的交通状况, 本周的交通状况和上周的交通状况等也存在一定的相关性。因此, 准确获取历史交通时间序列数据

中所包含的时间相关特性是交通预测中的一大挑战。

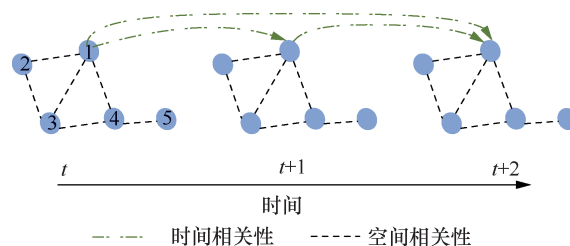


图3 局部路网时空示意图

为了获取路网的时间相关性, 通常将交通预测建模为一个时间序列回归问题, 即用历史时间序列的数据进行统计分析。经典的历史平均预测法、自回归移动平均法、支持向量回归等模型都建立在交通状况平稳的假设下, 不能处理路网上的紧急突发状况。本节重点分析3种用来建模路网时间相关性模型(RNN、LSTM、GRU)的优缺点以及它们之间的关联。

3.1 循环神经网络

RNN^[49]由输入层、隐藏层、输出层组成。RNN可以使用自身的内存单元处理任意输入的时间序列, 因此, 相比于前馈神经网络等深度学习方法, 更适合在动态的交通系统中捕捉交通状态的长时间依赖性^[19,21,28]。RNN结构如图4所示, RNN模型属于一种自循环机制, 它独特的记忆单元可以储存历史时刻交通信息的隐藏状态, 并将其用于下一时刻交通状态的预测。例如, 在 t 时刻, RNN 通过对上一时刻的输出(记忆单元储存的历史隐藏信息 h_{t-1})和当前时刻的输入 x_t 进行加权计算预测出下一时刻的交通状态 \hat{y}_t 。

RNN 在处理时间序列建模的问题上表现较好, 但是由于长时间依赖问题, 即在 t 时刻, 预测未来的交通速度不能仅仅依赖 $t-1$ 时刻的速度, 还要考虑交通状况的日周期、周周期等特性, 输入的时间序列可能很长。然而 RNN 隐藏层的存储空间有限, 当输入的交通数据序列过长时, 后输入的数据将覆盖前面输入的数据, 造成交通数据模糊或缺失。而且在计算传感器节点的输出特征时, RNN 使用的是矩阵乘法, 每个时间步上的梯度会根据矩阵乘法的值发生变化, 当矩阵乘法值过大或过小时, 容易发生梯度爆炸或梯度消失等问题。为了解决此问题, Hochreiter 等^[50]提出了 LSTM, 它不仅能够学习并建模时间序列的长期相关性^[51], 还可以自动确定最优的预测时滞。

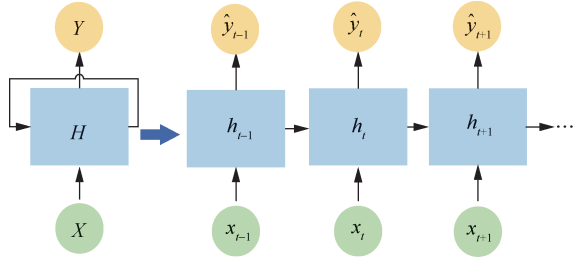


图4 RNN结构

3.2 长短期记忆网络

LSTM^[50,52-54]在RNN的基础上加入了门结构，它的隐藏层由3个门结构（遗忘门、输入门、输出门）和一个记忆单元构成。在RNN中，当前时刻预测的输出 Y 根据输入 X 和隐藏状态 H 计算得到，而LSTM通过门结构去除、增加记忆单元所储存的交通状态的信息，选择性地保留对当前交通状态影响较大的交通信息，与此同时丢弃那些对当前交通状态几乎不产生影响的交通信息，从而避免输入序列过长造成的数据缺失。

LSTM结构如图5所示。其中，遗忘门 f_t 决定在 t 时刻要丢弃哪些交通信息；输入门 i_t 用来确定在 t 时刻需要将多少新输入的交通信息传递给下一步，以便更新旧记忆单元存储的信息；输出门 o_t 用于输出 t 时刻的交通状态。LSTM隐藏层的计算过程如下

$$\begin{cases} f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \\ i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \\ o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \\ \tilde{C}_t = \tanh(W_C[h_{t-1}, X_t] + b_C) \\ C_t = f_t C_{t-1} + i_t \tilde{C}_t \\ h_t = o_t \tanh(C_t) \end{cases} \quad (2)$$

其中， X_t 表示 t 时刻的交通状态， C_{t-1} 表示 $t-1$ 时刻记忆单元所存储的交通信息， \tilde{C}_t 表示 t 时刻输入记忆单元的交通信息， h_{t-1} 为 $t-1$ 时刻的隐藏状态， W_f 、 W_i 、 W_o 、 W_C 及 b_f 、 b_i 、 b_o 、 b_C 表示训练过程中的权重和偏差， $\sigma(\cdot)$ 、 \tanh 均为激活函数。在得到编码器的隐藏状态 h_t 后，通常会将 h_t 输入到一个单层神经网络或全连接层来得到最终的预测结果。

3.3 门控循环单元

LSTM虽然可以缓解梯度消失和梯度爆炸，但其参数较多，计算复杂。GRU和LSTM都是采用门控机制控制记忆单元对交通信息的长期记忆，但

GRU只包含了更新和重置两个门，结构相对简单。因此越来越多的研究趋向于用GRU提取时间相关性，Fu等^[55]用LSTM和GRU两种方法分别进行短期交通流预测，结果表明在相同的实验配置下，用GRU进行预测得到的误差小于LSTM。GRU结构较为简单，参数较少，不仅可以加快训练速度，而且更容易收敛。

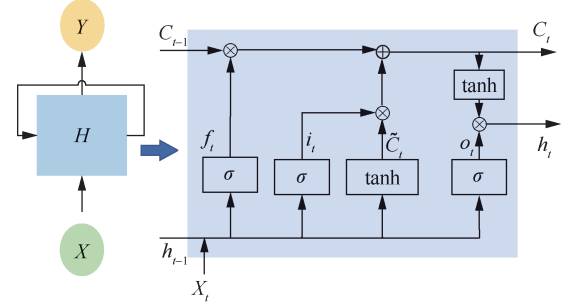


图5 LSTM结构

GRU结构如图6所示，GRU的隐藏层由更新门、重置门和记忆单元构成。其中，更新门 z_t 用来控制在 t 时刻使用 $t-1$ 时刻隐藏状态的程度，重置门 r_t 用来控制在 t 时刻忽略 $t-1$ 时刻的交通信息的程度。在输出的过程中，GRU直接将隐藏状态传递给下一个单元，比LSTM更为简便，不需要再通过输出门进行控制。GRU的传播规律如下

$$\begin{cases} z_t = \sigma(W_z[h_{t-1}, X_t] + b_z) \\ r_t = \sigma(W_r[h_{t-1}, X_t] + b_r) \\ C_t = \tanh(W_C[(r_t h_{t-1}), X_t] + b_C) \\ h_t = z_t h_{t-1} + (1 - z_t) C_t \end{cases} \quad (3)$$

其中， h_{t-1} 表示 $t-1$ 时刻的隐藏状态， X_t 表示 t 时刻输入的交通信息， W_z 、 W_r 、 W_C 和 b_z 、 b_r 、 b_C 表示训练过程中的权重和偏差， $\sigma(\cdot)$ 、 \tanh 均为激活函数。

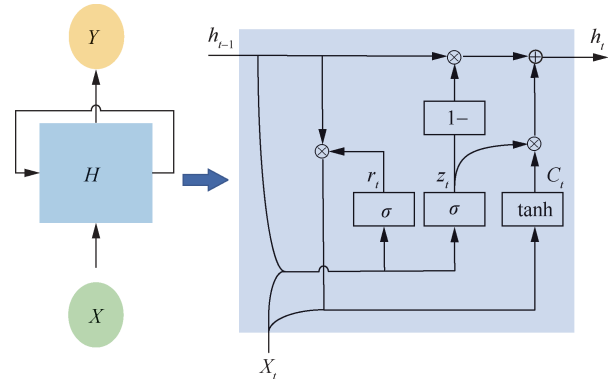


图6 GRU结构

综上,在现有的基于时空相关性的交通预测模型架构中,RNN 被广泛使用于捕获路网的时间相关性。它具有长期记忆功能,每一层 RNN 的输入都会考虑上一层 RNN 的输出,从而捕获路网交通状态的时序关系。然而经典的 RNN 模型存在梯度消失、爆炸等问题,随后出现了 LSTM 和 GRU 等模型,它们通过加入门控机制缓解了 RNN 的不足。但 LSTM 参数较多,结构复杂,训练时间较长,而 GRU 的模型较为简单,参数也相对较少,在训练模型时更加节约时间,因此越来越多的研究趋向于选取 GRU 模型来捕获路网的时间相关性。

4 空间相关性

从图 3 可以看出,路网中不仅存在时间相关性,相邻传感器在空间维度上也存在一定的影响。若节点 1 处发生拥堵,则此拥堵将会在一定时间内传播至与节点 1 相邻的节点 2、节点 3、节点 4。在双向公路网中位于不同方向的两个传感器,即使它们在欧几里得空间内相邻,但在某一特定时刻,相对车道之间交通状况的不同使它们检测到的交通数据具有较大差异。很大程度上,上游交通状况更容易受到下游交通状况的影响,路网的空间相关性与交通预测结果是息息相关的。因此成功捕获路网中的空间相关性对交通预测结果的准确性至关重要。本节将介绍图卷积网络和图注意力网络两种建模路网空间相关性的方法。

4.1 图卷积网络

在诸多研究中,基于交通预测的工作都使用卷积运算提取路网中复杂的空间依赖性。传统的 CNN^[56]适用于欧几里得空间,它将路网建模为网格,网格中的每个顶点都被当作一个节点,通过一个共享参数的卷积核对中心节点及相邻节点特征加权求和从而获取路网的空间特征。然而路网有其独特且复杂的拓扑结构,每个传感器节点可能存在不同数目的相邻节点,GCN 能够将传统网格数据的卷积运算推广到图数据上,可类比于一种在图上进行操作的卷积神经网络^[18]。在图卷积的过程中,卷积核的大小可以随着中心传感器相邻节点的个数而发生变换,只需要通过对相邻节点的特征进行加权计算来更新传感器节点自身的特征,因此采用 GCN 可以更加充分地利用路网的空间特征进行预测。

GCN 架构如图 7 所示,GCN 在每层卷积中分别对每个传感器节点都进行一次卷积操作,从而更

新节点的隐藏特征,然后叠加多层图卷积获取路网的空间相关性。根据输入路网拓扑图 G 得到传感器节点的邻接矩阵 A ,根据传感器观测到的交通信息构建路网的特征矩阵 $X \in \mathbf{R}^{N \times P}$, P 为节点的特征个数。GCN 的本质是找到适用于图的可学习的卷积核。它通过聚合传感器 v_i 自身的速度或流量特征 X_i 、相邻传感器 v_j 的特征 X_j 以及节点与相邻节点之间边的特征更新每个传感器的隐藏特征。每个传感器节点的特征更新规则为

$$X^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X^{(l)} W^{(l)} \right) \quad (4)$$

其中, l 是层数, σ 为激活函数, D 为对角度矩阵, $\tilde{A} = A + I_N$ 表示含自连接的邻接矩阵, I_N 表示单位矩阵, W 是一个可训练的权值矩阵^[18]。

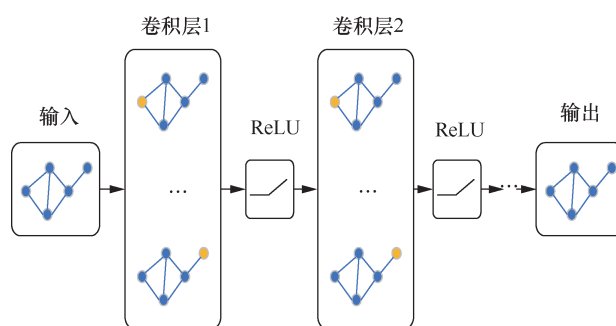


图 7 GCN 架构

GCN 有两种类型:一种是空域卷积,另一种是频域卷积。空域卷积的核心在于聚集邻居节点的信息,它将卷积操作直接定义在每个节点的连接关系上,让节点先在空间域中连接达到层级结构,然后再进行卷积,相当于直接在图的节点上进行卷积。而频域卷积^[57]以分析图的频谱为理论基础,它用傅里叶变换把空域转换到频域,然后进行卷积,主要通过图的拉普拉斯矩阵代替频域上的拉普拉斯算子,再依据频域上的卷积转换为图上的卷积。Yu 等^[20]将路网建模为无向图,并用频谱卷积进行预测,虽然取得了一定的效果,但频谱卷积方法通常要求图结构是无向的,无法得到节点的上下游信息。Li 等^[19]将路网建模为有向加权图,用扩散过程模拟交通流,通过 k 步双向扩散卷积运算对空间相关性进行建模,表明了双向扩散卷积捕获路网交通流随机性的有效性。

4.2 图注意力网络

图注意力网络^[58-59]是在 GCN 的基础上进行改

进而实现的。在捕获空间相关性时, GCN 对传感器节点同阶邻域中不同邻居节点分配的权重是相同的。它无法为邻居中不同的传感器节点分配不同的权重, 这限制了其对路网空间相关性的捕获能力。GAT 和 GCN 的关键区别在如何收集相邻传感器节点的特征表示。与图卷积中固定的标准化操作不同的是, GAT 用注意力机制对邻近节点特征进行加权计算, 迭代学习每个节点的隐藏特征。

用 GAT 捕获路网的空间相关性时, 首先将传感器检测到的历史时间序列数据转换为特征向量集 $\mathbf{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}$, $\vec{h}_i \in \mathbf{R}^P$, P 为每个传感器节点属性特征的数量。对所有传感器节点训练一个权重矩阵 $\mathbf{W} \in \mathbf{R}^{P \times P}$ 表示输入的 P 个传感器属性特征和输出的 P' 个传感器属性特征的关系。然后对每个传感器节点使用自注意力机制, 通过式(5)计算每个传感器节点的注意力系数

$$e_{ij} = a(\mathbf{W}\mathbf{h}_i, \mathbf{W}\mathbf{h}_j) \quad (5)$$

其中, $a(\cdot)$ 表示注意力机制。为了便于比较不同传感器节点的权重系数, 利用 softmax 函数对所有相邻传感器节点的注意力分值进行归一化

$$\alpha_{ij} = \text{soft max}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \quad (6)$$

通过归一化后, 传感器节点的原始注意力分值被转换成所有元素权重之和为 1 的概率分布, 使相关性强(即权重大)的传感器节点更加突出。经过归一化得到每个传感器节点的注意力分值 α_{ij} 后, 对每个传感器节点的邻居节点特征进行线性组合, 再经过激活函数 σ 得到最终的输出特征

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}\mathbf{h}_j \right) \quad (7)$$

其中, σ 为激活函数, 使输出具有非线性特征。对每个节点采用注意力机制, 可以获取不同传感器节点对预测结果的影响程度, 更有利于提取路网的空间相关性。

文献[25]在图卷积的过程中加入注意力机制来学习传感器节点的权值矩阵, 根据节点的权值矩阵和不同时间戳的输入获得传感器节点的自适应邻接矩阵, 从而捕获不同时间序列下路网结构的时空相关性。Zhang 等^[33]用注意力机制代替了上述图卷积网络中的卷积运算, 采用多头注意力机制^[60]使模型

通过多个独立的注意块共同学习路网的空间相关性, 并在真实数据集上进行验证, 实验结果表明基于图的注意力机制在对时空相关性的学习上具有很大的潜力。Zheng 等^[34]使用具有门控融合的空间和时间注意力机制模拟复杂的时空相关性, 通过一个图多注意力网络预测路网中不同位置的超前时间步长的交通状况。Bello 等^[61]开发了一种新的二维相对自注意力机制, 通过自注意力机制扩充卷积算子, 并通过大量的实验证明当卷积与自注意力机制两者相结合时效果最好。

本节总结了捕获路网空间相关性的常用方法。在空间相关性的捕获过程中, GCN 可以有效地聚合邻居节点的特征信息。但在传感器节点特征的更新过程中, 图卷积分配给每个相邻传感器节点的权重是相同的, 而实际道路中相邻节点对传感器节点的影响会随着时间的推移和空间位置的不同而发生变化。图注意力网络采用注意力机制在节点状态的更新过程中给每个相邻传感器节点分配不同的权重, 使传感器节点的隐藏状态更依赖于对其影响较大的相邻节点, 使得模型可以捕获更加精确的空间特征信息, 从而提高预测精度。

5 时空相关性

在上面两节中, 本文依次介绍了路网的时间相关性和空间相关性, 对时间相关性来说, 当前时间步的每个传感器节点可以直接影响下一个时间步的自身; 对于空间相关性来说, 在同一个时间步内, 每个传感器节点可以直接影响其相邻节点。而在实际路网中, 信息的传播同时发生在时间和空间两个维度上。路网的时空相关性如图 8 所示, 时空图中的传感器节点 4 能够影响它的相邻节点(节点 1、节点 3、节点 5)在下一时间步的交通状态。

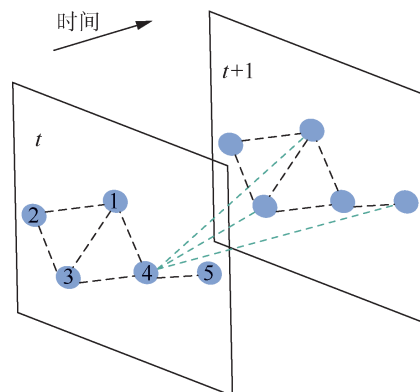


图 8 路网的时空相关性

时空网络是一种典型的数据结构,能够描述交通网络、社交网络、引文网络、化学分子结构、推荐系统等相关应用中的大量数据,对时空网络数据的准确预测可以显著提高这些应用的服务质量。在捕获路网的时空相关性时,使用 T 个历史时间序列的交通数据作为输入,通过图卷积获取路网的空间相关性,然后将含有空间相关性信息的时间序列输入 RNN,通过 RNN 之间的信息传递获得包含时空特征的 T 个隐藏状态。通过卷积得到的空间位置信息只能展现某个时间点的信息,不能反映交通状态随时间变化的趋势,再经过 RNN,可以通过记忆单元所储存的历史时刻隐藏状态辅助预测下一时刻的交通状态。所以,使用时空模型可以同时捕获路网中复杂的空间和时间相关性。

捕获时空相关性的方法有两种,一种是间接捕获,一种是直接捕获。Guo 等^[28]结合基于频谱卷积的 GCN 和 LSTM 提出了一种序列到序列的交通速度预测模型(GC-LSTM),用图卷积操作代替 LSTM 中的矩阵乘法。为避免输入的时间序列数据过长而引起数据丢失,还加入了注意力机制使模型在解码时更加关注输入序列中影响力较大的特定部分,从而使整个 GC-LSTM 模型可以同时预测整个路网的多步交通速度。文献[5]提出一种时间图卷积网络(T-GCN)模型,将历史时间序列输入 GCN 得到路网空间特征,再将获得的具有空间特征的时间序列输入 GRU,从而捕捉时空特征,最后经过一个全连接层得到预测结果。DCRNN 采用双向随机扩散模拟交通流,通过扩散卷积捕获路网的空间相关性,然后通过扩散卷积代替 GRU 中的矩阵乘法,从而获得时空相关性。除此之外,ASTGCN^[24]等方法也是采用间接捕获的方式。它们虽然都从时间相关性和空间相关性两方面捕获了路网的特性,但它们都是使用两个单独的组件分步进行捕获的,先得到路网的空间相关性,然后将空间表示输入 RNN 等时间模块间接获取时空相关性,并没有达到时空相关性的同时捕获。在现有文献中,大多数研究都采用间接捕获的方式,模型架构如图 2 所示。其中,GNN(以 GCN 为例)与 RNN(以 GRU 为例)之间的连接关系如图 9 所示。在 t 时刻,输入每个传感器的交通特征 X_t ,经过 GCN 聚合其邻居节点的特征得到传感器节点的空间特征 \tilde{X}_t ,将其与 $t-1$ 时刻的隐藏特征同时输入 GRU 得到 t 时刻的预测结果 \hat{y}_t 和隐藏特征 h_t 。

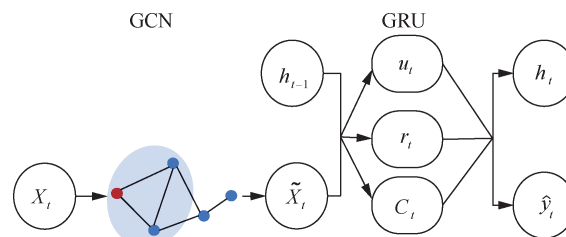


图 9 GCN 与 GRU 之间的连接关系（间接捕获）

另外一种方法是直接捕获,通过一种时空卷积组件实现同时提取路网时空相关性的效果,不需要再通过不同类型的神经网络模型分别获取时间和空间相关性,这种方式有效地揭示了时空网络数据的生成方式。Song 等^[35]提出了一种时空同步图卷积网络模型,通过设计一种时空同步建模机制,把时间和空间的邻接关系统一到局部时空图的邻接矩阵中,以达到同时捕获路网局部时空相关性的效果。将相邻时间步长的单个空间图连接成一个图来构造局部时空图,然后通过一个时空同步图卷积模块捕获构造局部时空图中复杂的局部时空相关性,最后在不同的时间段叠加多个单独的时空同步图卷积模块,通过多个时空图卷积模块的堆叠聚集长期的时空相关性和异质性,以达到更精确的预测结果。

直接捕获方法的结构如图 10 所示,首先输入路网的局部时空网络图,通过输入层将输入特征转换为高维度的空间特征,然后通过叠加多个时空同步卷积模块获取路网中的局部时空相关性,最后经过一个多模块全连接层将最终的输出映射至输出空间,从而得到预测结果。

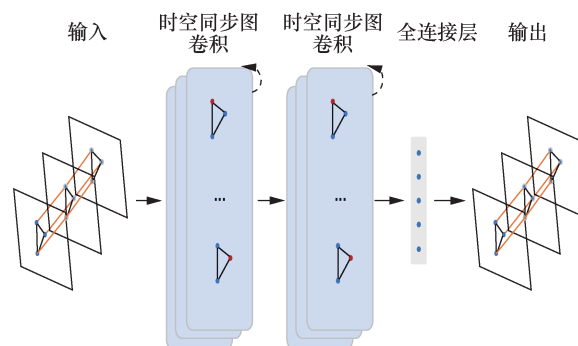


图 10 直接捕获方法的结构

本节综述了捕获时空相关性的两种方法,一种是间接捕获,它通过两个单独的神经网络分别对路网的时间和空间相关性进行建模,然后将带有空间相关性的表示输入时间建模模块,从而间接得到路网的时空相关性。另外一种方法是直接捕获,它将相邻的时间步长构造造成一个局域时空图,通过叠加

时空同步图卷积直接获取路网的局部时空相关性，然后叠加多层时空同步图卷积获取整个路网的长期时空相关性。相比于间接捕获，直接捕获考虑了时空数据的异质性。间接捕获通过叠加神经网络层捕获长时间序列的全局相关性，当时间序列过长或路网较广时，间接捕获可能会发生局部信息丢失的现象。而直接捕获把时间和空间的邻接关系统一到局部时空图中，通过对局部时空图进行处理，同步捕获路网的局部关联性，更有利于揭示路网复杂的时空特性。在现有文献中，大多数研究都采用间接捕获的方法，对直接捕获方法的研究相对较少，有望成为未来的研究方向。

无论是直接捕获还是间接捕获，它们都建立在图神经网络的基础上。为了更好地将节点特征融入模型，研究人员引入了注意力机制、随机扩散等方法，但这些模型在预测时大多只考虑路网的节点特征和时间相关性，忽略了道路周围兴趣点（POI, point of interest）、天气、特殊事件、周期性等外部因素对预测性能的影响。尽管有些研究考虑到了一两种外部因素的影响^[29]，但要想得到更为精确的预测结果，POI、天气、特殊事件、周期性等外部因素的影响都是不可忽略的，尤其是在较为复杂的城市场景下。然而现有的研究缺乏对多种外部因素影响下预测性能的分析。融合多外部因素的时空预测模型如图 11 所示，可以在上述直接捕获、间接捕获模型的基础上，通过 RNN 或多层感知器（MLP, multi-layer perceptron）对外部因素特征进行处理，然后与时空相关性模型得到的结果进行融合，得到基于多种外部因素影响下的预测结果。考虑多种因素的情况下，模型的计算复杂度也会随之增加，如何在多因素影响和计算复杂度之间寻求平衡也是未来研究的一大难题。

6 实例分析

DCRNN 作为最早一批将图神经网络用于交通预测的经典方法之一，为后来的研究奠定了基础。本文在真实的数据集 PEMS-BAY^[62]上用 DCRNN 方法对基于图神经网络的预测模型进行验证分析。PEMS-BAY 数据集以 5 min 为一个时间步采集交通信息，包含了旧金山湾区内 325 个传感器节点，记录了从 2017 年 1 月 1 日到 2017 年 5 月 31 日的交通数据信息，包括交通速度数据、传感器节点的地理位置等信息。

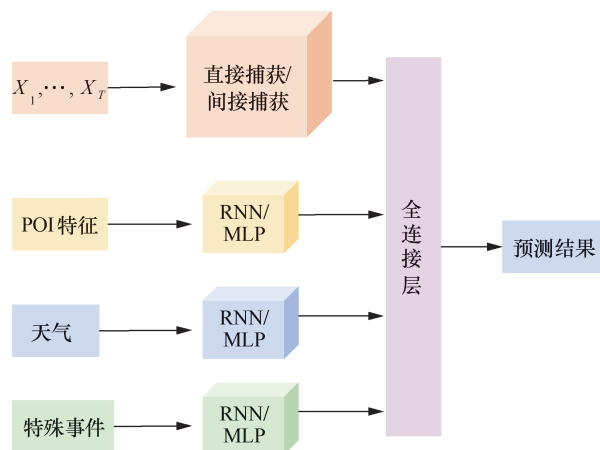


图 11 融合多外部因素的时空预测模型

DCRNN 通过间接捕获的方式，整体采用 Seq2Seq 模型架构，在编码器和解码器中都通过扩散卷积和 GRU 分别建立空间相关和时间相关模型，用扩散卷积代替 GRU 中的矩阵乘法，从而间接得到路网的时空相关性。首先，用一个扩散过程模拟交通流的随机特性，通过扩散卷积更新每个传感器节点的空间隐藏特征，随后将包含空间特征的历史序列输入 GRU 得到历史交通数据对未来交通的影响特性。在训练模型之前，首先将 PEMS-BAY 数据集分为训练集（70%）、验证集（10%）、测试集（20%）3 个部分，然后通过调度采样选取概率为 ε 的真实样本或概率为 $1-\varepsilon$ 的预测值对模型进行训练。将历史时间序列数据输入编码器，编码器会将整个时间序列数据编码为一个固定长度的向量，然后再由解码器从向量中读取下一个时间步的输出，即预测结果。在 PEMS-BAY 数据集中选取一个传感器节点，对预测结果进行可视化，DCRNN 模型的预测结果如图 12 所示。

可以看出，DCRNN 能够预测出与实际道路中高峰期的开始和结束吻合的趋势，这表明了扩散卷积能够很好地捕获路网的空间依赖性。由于路网易受异常事件的影响，图卷积定义的是一个相对平滑的滤波器，它通过不断移动捕获空间特征，这一过程可能使预测结果发生微小的变化，致使 DCRNN 在峰值处的预测能力较差。

将 DCRNN 方法与历史平均（HA, historical average）模型、向量自回归（VAR, vector autoregression）模型、ARIMA、全连接 LSTM（FC-LSTM）、Graph WaveNet^[30]方法进行对比，其中 Graph WaveNet 结合 GCN 和 DCC 获取路网的时空相关性。采用平均绝对误差（MAE, mean absolute error）、

平均绝对百分比误差 (MAPE, mean absolute percentage error) 和均方根误差 (RMSE, root mean squared error) 3 个度量标准对模型的预测结果进行评估。不同预测方法的性能比较见表 3, 展现了对未来 15 min、30 min、60 min 路网信息的预测结果。

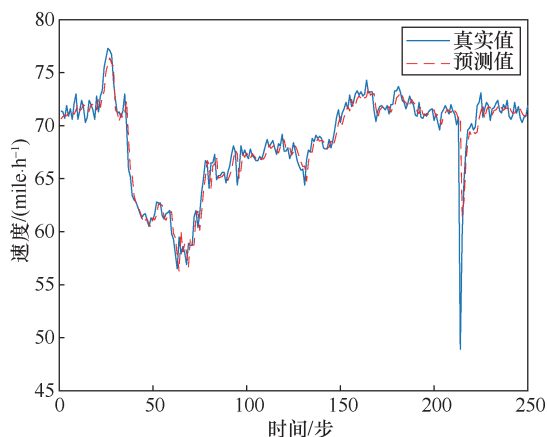


图 12 DCRNN 模型的预测结果

$$\begin{aligned} \text{RMSE}(y, \hat{y}) &= \sqrt{\frac{1}{N} \sum_{i \in N} (y_i - \hat{y}_i)^2} \\ \text{MAPE}(y, \hat{y}) &= \frac{1}{N} \sum_{i \in N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \\ \text{MAE}(y, \hat{y}) &= \frac{1}{N} \sum_{i \in N} |y_i - \hat{y}_i| \end{aligned} \quad (8)$$

其中, N 表示观测样本数, y 表示真实值, \hat{y} 表示预测值。

从表 3 可以看出, 与 HA 等传统方法和 FC-LSTM 等深度学习方法相比, DCRNN 在所有预测范围内的所有误差度量指标都达到最小, 这表明了 DCRNN 时空预测模型的有效性。Graph WaveNet 采用 DCC 代替 GRU 获取路网的时间相关性, 并达到比 DCRNN 更低的误差, 表明 DCC 可以更高效地处理具有长时间序列的时空图数据。然而随着预测时间的增加, 误差逐渐增大, 这是由于在预测的过程中, 没有真实样本的监督, 误差不断累积而造成的。因此, 提高时空预测模型长期预测的精度是实现智能交通系统的一大难题, 有待进一步研究。DCRNN 虽然带我们走进了用图神经网络进行交通预测的大门, 并展示了用图神经网络捕获路网空间相关性对提高预测精度的有效性, 但如何缓解 Seq2Seq 模型中的曝光误差问题、如何提高长期预测的精度等, 仍是交通预测中急需解决的难题。

表 3 不同预测方法的性能比较

时间	模型	MAE	RMSE	MAPE
15 min	HA	2.88	5.59	6.76%
	VAR	1.74	3.09	3.59%
	ARIMA	1.62	3.30	3.50%
	FC-LSTM	2.05	4.19	4.80%
	DCRNN	1.31	2.76	2.73%
	Graph WaveNet	1.30	2.74	2.73%
30 min	HA	2.88	5.59	6.76%
	VAR	2.33	4.15	5.02%
	ARIMA	2.33	4.76	5.40%
	FC-LSTM	2.20	4.55	5.20%
	DCRNN	1.65	3.77	3.72%
	Graph WaveNet	1.63	3.70	3.67%
60 min	HA	2.88	5.59	6.76%
	VAR	2.92	5.11	6.46%
	ARIMA	3.38	6.50	8.30%
	FC-LSTM	2.37	4.96	5.70%
	DCRNN	1.97	4.62	4.71%
	Graph WaveNet	1.95	4.52	4.63%

7 挑战与未来方向

本文对图神经网络驱动下的交通预测方法进行了分析研究, 综述了路网的时间和空间相关性的建模方法。本节将讨论未来交通预测可能的发展方向。

1) 与强化学习相结合

本文介绍了基于 Seq2Seq 架构的预测模型, 它可以较好地捕获路网的时空相关性。但在 Seq2Seq 架构中仍存在曝光误差及模型在训练和测试时不一致等问题。随着强化学习的发展, 研究人员也逐渐用它来解决交通领域的一些难题^[63]。对于上述曝光误差等问题, 可以在解码器中加入 Actor-Critic 网络使模型在训练的过程中根据自身的预测值对下一时间步的交通状况进行预测。Actor-Critic 网络^[64]在策略梯度的基础上加入了价值函数, Actor 使用策略函数生成下一步动作并和环境交互, 而 Critic 使用价值函数评估当前的策略并指导其下一阶段的动作, 即监督每一步的预测结果。然而 Actor-Critic 仍存在难以收敛的问题, 可以使用强化学习中的 A3C、DDPG 等方法, 其中 A3C 采用了多线程的方式使模型易于收敛, DDPG 采用双 Actor

和双 Critic 网络改善了收敛性。这些方法在 Seq2Seq 模型中仍需进行深入研究。

2) 大型路网预测

现有模型方法大多在包含传感器节点较少的数据集上进行验证分析。然而对于一些大型路网(如 PeMS 数据集包含了整个加州地区所有道路上数千个传感器节点),其卷积计算的复杂度令人望而却步,一台计算机的内存也难以承受数千个节点的训练数据,这导致基于 GCN 的模型难以完成整个 PeMS 路网空间特征的提取^[65]。在大型路网中,每个节点的状态都与大量邻接节点的隐藏状态相关,这不仅导致卷积操作计算量的增加,还对计算机的内存有着极高的要求。针对上述问题,可以通过图分割技术将大型路网分割为若干个子图。文献[65]在 DCRNN 的基础上加以改进,使其能够用于大型公路网(PeMS)的预测,采用图形分区包 Metis 将整个路网拓扑图分割为若干个大小大致相同的子图,并同时用一个权值 DCRNN 对所有子图进行训练。图分割方法虽然可以将 DCRNN 扩展到大型路网,但不准确的图划分可能会导致大量空间相关节点被分割到不同的子图中,从而导致预测精度大打折扣。所以还可以考虑通过叠加多层图注意力网络对大型路网进行预测,注意力机制可以使每个传感器节点只依据与自身相关性较大的相邻节点更新自身的隐藏特征。虽然大型路网中的传感器节点存在大量相邻节点,但是可以通过叠加多层图注意力网络的方式使模型提取与每个节点最相关的邻居节点信息,从而减轻计算的复杂度。

3) 长期预测的精度

交通预测模型在短期交通流预测方面表现良好,从表3可以看出,随着预测时间的增加,误差不断变大。随着路网规模的不断扩大,路网中的交叉口及传感器数量随之增加,路网拓扑图也变得更加复杂,历史的交通数据量也更加庞大。由于 RNN 模型的记忆单元有限,当输入数据过多时,数据容易被覆盖或丢失,这会导致预测误差随着预测时间的增加不断累积。在这种情况下,如何有效地提高长期预测的精确度是一项非常具有挑战性的任务,也是交通预测中必须要克服的一个难题。LSTM、GRU 等 RNN 模型是一条从旧记忆单元到新记忆单元的序列路径,由于增加了门结构分支,路径更加复杂,过长的序列路径可能造成数据丢失,从而产生误差并随着时间的推移不断积累。为避免过长的

序列路径引起数据丢失,采用多个注意力模块构成分层注意力结构,将所有过去的向量编码汇总到一个上下文向量中,类似于分层神经网络和时域卷积网络(TCN, temporal convolutional network)^[66-67]。多层注意力可以回顾更多过去的记忆,从而影响未来的决策,缓解误差的积累。TCN 的体系结构比规范的回归网络(LSTM、GRU)更加准确、简单、清晰,不仅如此,在训练时,TCN 所需内存比循环神经网络更低,更适合用于处理长时间序列任务。因此,可以通过 TCN 记忆更长时间的历史交通信息,缓解序列路径过长引起的数据丢失问题,从而提高长期预测的精确度。

4) 非周期拥堵下的预测

在捕获时间相关性方面,研究人员考虑在预测过程中加入日周期、周周期等周期性因素,但降雨、交通事故、大型会议、演出等多种不确定因素对交通预测也存在着很大的影响,而高速公路延误大多是由非周期性拥堵事件引起的,这使得交通状况的预测更加困难。所以提高非周期性拥堵情况下预测模型的能力是一个需要解决的重要问题。可以采用 RNN 或 MLP 处理非周期性的外部因素,也可以考虑通过元学习方法对节点的外部属性特征进行处理。以降雨为例,可以将降雨前后的交通速度数据筛选出来,构成一个新的数据集,通过上述方法构建时空预测模型,用筛选出来的数据集对模型进行训练得到降雨时路网的时空特性,从而进一步实现降雨时的交通预测。在预测非周期性外在因素引起的交通拥堵时,还可以在预测的过程中加入神经网络模型对外在因素(如交通事故,大型会议等)特征进行处理,最后将外在因素特征与时空特征进行融合,预测非周期因素影响下的交通状态。

5) 交通数据缺失

现有的交通预测研究大多基于环形线圈检测器、超声波传感器等设备采集到的数据,这些数据常常会因为天气、传感器失灵、交通管制等原因出现数据缺失、无效、不确定等情况。这不仅会增加数据预处理的难度,还会降低分析结果的准确性和预测模型的效率。在 5G 大数据背景下,可以在预测的过程中,将激光雷达、毫米波雷达等路侧单元监测的交通数据信息作为辅助数据源。先对不同来源的数据进行特征提取,得到相应的特征矩阵,采用全连接层对其进行融合作为模型的输入,经过上述时空预测模型为交通管理者提供更为精确的预

测结果。除此之外,还可以通过生成对抗网络(GAN, generative adversarial network)^[68]根据以往数据生成缺失的数据,生成器根据历史数据仿造数据,判别器用于鉴别生成器仿造的数据是否有效。随后生成器根据判别器的判定结果调参,直至判别器不能判定生成器生成的结果与真实值之间的真假时,再对判别器进行调优,重复上述步骤,通过不断对生成器和判别器进行优化,使模型生成的数据尽可能地达到最优。然后用训练好的模型补全数据集中缺失的数据,从而提高预测精度。在路网模式相同的情况下,还可以考虑通过迁移学习将模型从数据充足的路网转移到数据缺失的路网。所以,在面对交通数据缺失的问题时,不仅可以通过融合多源异构数据进行弥补,还可以考虑 GAN、迁移学习等方法。

6) 实际交通场景应用

现有的交通预测方法大多基于数据集对模型进行训练及测试,而模型在真实场景中的应用是极度匮乏的。然而实现准确的交通预测是为了给交通管理者和出行者提供方便,从而缓解交通拥堵,实现智能交通。在实际应用中,首先收集选定场景周围的交通数据,用上述方法进行交通预测,然后将预测得到的交通数据用于规划出行道路、设计红绿灯的自动调控等。将交通状况分为不拥堵、轻微拥堵、非常拥堵等情况,并为每种情况设定一种交通灯调控方案。让路口的交通灯可以依据提前得知的交通预测信息判断未来的交通状况,并实时采取对应的交通灯调控方案,从而帮助交通管理者避免拥堵的发生。随着 6G 网络^[69]的快速发展,我们认为未来的研究应更加专注于如何将交通预测用于实际场景中,这将有助于交通管理者对交通进行管控,为人们的出行带来方便。准确的交通预测是交通信息服务系统的重中之重,有利于缓解交通拥堵,减少交通事故的发生,从而实现人、车、路协同的智能交通系统。

8 结束语

本文针对如何充分利用路网时间相关性和空间相关性进行预测等问题,综述了将图神经网络中图卷积网络、图注意力网络等关键技术与递归神经网络相结合进行交通预测的时空预测模型。基于上述模型,本文总结了两种捕获路网时空相关性的解决方案:直接捕获、间接捕获。最后,提供了基于

图神经网络的交通预测中与强化学习结合、大型路网预测、长期预测的精度、非周期拥堵下的预测、实际交通场景应用等可能的发展方向。总之,图神经网络驱动下的交通预测不仅可以获取路网的时空特征,还有望解决预测过程中大型路网、长期预测等技术难题,为智能交通系统的实现奠定了坚实的基础。

参考文献:

- [1] LU N, CHENG N, ZHANG N, et al. Connected vehicles: solutions and challenges[J]. IEEE Internet of Things Journal, 2014, 1(4): 289-299.
- [2] QI Y L, TIAN L, ZHOU Y Q, et al. Mobile edge computing-assisted admission control in vehicular networks: the convergence of communication and computation[J]. IEEE Vehicular Technology Magazine, 2019, 14(1): 37-44.
- [3] ZHOU Y Q, TIAN L, LIU L, et al. Fog computing enabled future mobile communication networks: a convergence of communication and computing[J]. IEEE Communications Magazine, 2019, 57(5): 20-27.
- [4] 张士兵, 王婷婷, 张晓格, 等. 智能交通车载网的现状及其发展策略[J]. 通信技术, 2017, 50(7): 1345-1350.
ZHANG S B, WANG T T, ZHANG X G, et al. Development strategy and status of intelligent transportation vehicular networks[J]. Communications Technology, 2017, 50(7): 1345-1350.
- [5] ZHAO L, SONG Y J, ZHANG C, et al. T-GCN: a temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(9): 3848-3858.
- [6] 余辰, 张丽娟, 金海. 大数据驱动的智能交通系统研究进展与趋势[J]. 物联网学报, 2018, 2(1): 56-63.
YU C, ZHANG L J, JIN H. Research progress and trend of big data-driven intelligent transportation system[J]. Chinese Journal on Internet of Things, 2018, 2(1): 56-63.
- [7] 殷礼胜, 唐圣期, 李胜, 等. 基于整合移动平均自回归和遗传粒子群优化小波神经网络组合模型的交通流预测[J]. 电子与信息学报, 2019, 41(9): 2273-2279.
YIN L S, TANG S Q, LISHENG, et al. Traffic flow prediction based on hybrid model of auto-regressive integrated moving average and genetic particle swarm optimization wavelet neural network[J]. Journal of Electronics & Information Technology, 2019, 41(9): 2273-2279.
- [8] GUO J H, HUANG W, WILLIAMS B M. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification[J]. Transportation Research Part C: Emerging Technologies, 2014, 43: 50-64.
- [9] DELL'ACQUA P, BELLOTTI F, BERTA R, et al. Time-aware multi-variate nearest neighbor regression methods for traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(6): 3393-3402.
- [10] LUO X L, LI D Y, ZHANG S R. Traffic flow prediction during the holidays based on DFT and SVR[J]. Journal of Sensors, 2019, 2019: 1-10.
- [11] DUAN Y J, L V Y, WANG F Y. Travel time prediction with LSTM neural network[C]//Proceedings of 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE Press, 2016: 1053-1058.
- [12] WANG W, ZHOU C H, HE H L, et al. Cellular traffic load prediction with LSTM and Gaussian process regression[C]//Proceedings of ICC 2020 - 2020 IEEE International Conference on Communications (ICC).

- Piscataway: IEEE Press, 2020: 1-6.
- [13] HAN L Y, ZHENG K, ZHAO L, et al. Short-term traffic prediction based on DeepCluster in large-scale road networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(12): 12301-12313.
 - [14] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61-80.
 - [15] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
 - [16] ZHOU J, CUI G Q, ZHANG Z Y, et al. Graph neural networks: a review of methods and applications[J]. ArXiv:1812.08434, 2018.
 - [17] ZHANG J W. Graph neural networks for small graph and giant network representation learning: an overview[J]. ArXiv:1908.00187, 2019.
 - [18] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of International Conference on Learning Representations (ICLR). [S.l.:s.n.], 2016.
 - [19] LI Y G, YU R, SHAHABI C, et al. Graph convolutional recurrent neural network: data-driven traffic forecasting[EB]. Proceedings of International Conference on Learning Representations (ICLR). [S.l.:s.n.], 2017.
 - [20] YU B, YIN H T, ZHU Z X. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2018.
 - [21] GUO K, HU Y L, QIAN Z, et al. Optimized graph convolution recurrent neural network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(2): 1138-1149.
 - [22] CIRSTEAN R G, GUO C J, YANG B. Graph attention recurrent neural networks for correlated time series forecasting[EB]. 2021.
 - [23] 冯宁, 郭晟楠, 宋超, 等. 面向交通流量预测的多组件时空图卷积网络[J]. 软件学报, 2019, 30(3): 759-769.
FENG N, GUO S N, SONG C, et al. Multi-component spatial-temporal graph convolution networks for traffic flow forecasting[J]. Journal of Software, 2019, 30(3): 759-769.
 - [24] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 922-929.
 - [25] ZHOU F, YANG Q, ZHANG K P, et al. Reinforced spatiotemporal attentive graph neural networks for traffic forecasting[J]. IEEE Internet of Things Journal, 2020, 7(7): 6414-6428.
 - [26] SHI X M, QI H, SHEN Y M, et al. A spatial-temporal attention approach for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2020.
 - [27] ZHAO J D, GAO Y, YANG Z Z, et al. Truck traffic speed prediction under non-recurrent congestion: based on optimized deep learning algorithms and GPS data[J]. IEEE Access, 2019, 7: 9116-9127.
 - [28] GUO J L, SONG C Y, WANG H. A multi-step traffic speed forecasting model based on graph convolutional LSTM[C]//Proceedings of 2019 Chinese Automation Congress (CAC). Piscataway: IEEE Press, 2019: 2466-2471.
 - [29] GE L, LI H, LIU J L, et al. Temporal graph convolutional networks for traffic speed prediction considering external factors[C]//Proceedings of 2019 20th IEEE International Conference on Mobile Data Management (MDM). Piscataway: IEEE Press, 2019: 234-242.
 - [30] WU Z H, PAN S R, LONG G D, et al. Graph WaveNet for deep spatial-temporal graph modeling[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 1907-1913.
 - [31] CUI Z Y, HENRICKSON K, KE R M, et al. Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(11): 4883-4894.
 - [32] ZHANG Z C, LI M, LIN X, et al. Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies[J]. Transportation Research Part C: Emerging Technologies, 2019, 105: 297-322.
 - [33] ZHANG C H, YU J J Q, LIU Y. Spatial-temporal graph attention networks: a deep learning approach for traffic forecasting[J]. IEEE Access, 2019, 7: 166246-166256.
 - [34] ZHENG C P, FAN X L, WANG C, et al. GMAN: a graph multi-attention network for traffic prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.:s.n.], 2020, 34(1): 1234-1241.
 - [35] SONG C, LIN Y F, GUO S N, et al. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.:s.n.], 2020, 34(1): 914-921.
 - [36] Park C, Lee C, Bahng H, et al. STGRAT: a spatio-temporal graph attention network for traffic forecasting[J]. ArXiv: 1911.13181, 2019.
 - [37] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. [S.l.:s.n.], 2014: 3104-3112.
 - [38] BENGIO S, VINYALS O, JAITLY N, et al. Scheduled sampling for sequence prediction with recurrent neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. [S.l.:s.n.], 2015: 1171-1179.
 - [39] RANZATO M, CHOPRA S, AULI M, et al. Sequence level training with recurrent neural networks[J]. Arxiv:1511.06732, 2015.
 - [40] DAUMÉ H, LANGFORD J, MARCU D. Search-based structured prediction[J]. Machine Learning, 2009, 75(3): 297-325.
 - [41] YU L T, ZHANG W N, WANG J, et al. SeqGAN: sequence generative adversarial nets with policy gradient[C]//Proceedings of 31st AAAI Conference on Artificial Intelligence. [S.l.:s.n.], 2016.
 - [42] KENESHLOO Y, SHI T, RAMAKRISHNAN N, et al. Deep reinforcement learning for sequence-to-sequence models[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7): 2469-2489.
 - [43] LANA I, DEL SER J, VELEZ M, et al. Road traffic forecasting: recent advances and new challenges[J]. IEEE Intelligent Transportation Systems Magazine, 2018, 10(2): 93-109.
 - [44] NAGY A M, SIMON V. Survey on traffic prediction in smart cities[J]. Pervasive and Mobile Computing, 2018, 50: 148-163.
 - [45] ZHANG Y J. Short-term traffic flow prediction methods: a survey[J]. Journal of Physics: Conference Series, 2020, 1486: 052018.
 - [46] XIE P, LI T R, LIU J, et al. Urban flow prediction from spatiotemporal data using machine learning: a survey[J]. Information Fusion, 2020, 59: 1-12.
 - [47] YE J, ZHAO J, YE K, et al. How to build a graph-based deep learning architecture in traffic domain: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2020.
 - [48] ZHANG Z, CUI P, ZHU W. Deep learning on graphs: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2020.
 - [49] JAIN A, ZAMIR A R, SAVARESE S, et al. Structural-RNN: deep learning on spatio-temporal graphs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 5308-5317.
 - [50] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
 - [51] 王祥雪, 许伦辉. 基于深度学习的短时交通流预测研究[J]. 交通运输系统工程与信息, 2018, 18(1): 81-88.
WANG X X, XU L H. Short-term traffic flow prediction based on deep learning[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(1): 81-88.

- [52] MA X L, TAO Z M, WANG Y H, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. Transportation Research Part C: Emerging Technologies, 2015, 54: 187-197.
- [53] TIAN Y X, PAN L. Predicting short-term traffic flow by long short-term memory recurrent neural network[C]//Proceedings of 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). Piscataway: IEEE Press, 2015: 153-158.
- [54] YU R, LI Y G, SHAHABI C, et al. Deep learning: A generic approach for extreme condition traffic forecasting[C]//Proceedings of the 2017 SIAM International Conference on Data Mining. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2017: 777-785.
- [55] FU R, ZHANG Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction[C]//Proceedings of 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). Piscataway: IEEE Press, 2016: 324-328.
- [56] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time-series[J]. The Handbook of Brain Theory and Neural Networks, 1995, 3361(10): 1-14.
- [57] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in Neural Information Processing Systems, 2016, 29: 3844-3852.
- [58] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. Arxiv: 1710.10903, 2017.
- [59] BUSBRIDGE D, SHERBURN D, CAVALLO P, et al. Relational graph attention networks[J]. Arxiv:1904.05811, 2019.
- [60] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [61] BELLO I, ZOPH B, LE Q, et al. Attention augmented convolutional networks[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 3285-3294.
- [62] CHEN C, VARAIYA P. Freeway performance measurement system (PeMS)[EB]. 2003.
- [63] WU W, CHEN N, ZHOU C H, et al. Dynamic RAN slicing for service-oriented vehicular networks via constrained learning[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(7): 2076-2089.
- [64] BAHADANAU D, BRAKEL P, XU K, et al. An actor-critic algorithm for sequence prediction[J]. arXiv: 1607.07086, 2016.
- [65] MALLICK T, BALAPRAKASH P, RASK E, et al. Graph- partitioning-based diffusion convolutional recurrent neural network for large-scale traffic forecasting[J]. Transportation Research Record: Journal of the Transportation Research Board, 2020, 2674(9): 473-488.
- [66] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. Arxiv:1803.01271, 2018.
- [67] LEA C, FLYNN M D, VIDAL R, et al. Temporal convolutional networks for action segmentation and detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 1003-1012.
- [68] 王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络GAN的研究进展与展望[J]. 自动化学报, 2017, 43(3): 321-332.
- WANG K F, GOU C, DUAN Y J, et al. Generative adversarial networks: the state of the art and beyond[J]. Acta Automatica Sinica, 2017, 43(3): 321-332.
- [69] ZHOU Y Q, LIU L, WANG L, et al. Service-aware 6G: an intelligent and open network based on the convergence of communication, computing and caching[J]. Digital Communications and Networks, 2020, 6(3): 253-260.

[作者简介]



周毅(1981-), 男, 博士, 河南大学教授、博士生导师, 河南省车联网协同技术国际联合实验室主任, 主要研究方向为车联网与智能交通、空地协同组网、平行增强学习、协作机器人等。



胡姝婷(1997-), 女, 河南大学硕士生, 主要研究方向为图神经网络、智能交通等。



李伟(1979-), 女, 河南大学副教授、硕士生导师, 主要研究方向为车联网优化控制、协作通信等。



承楠(1987-), 男, 博士, 西安电子科技大学教授、博士生导师, 主要研究方向为车联网与先进交通系统、人工智能、空天地一体化网络等。



路宁(1984-), 男, 博士, 加拿大女王大学助理教授, 主要研究方向为车联网与智能交通、深度强化学习、移动边缘计算等。



沈学民(1958-), 男, 博士, 中国工程院外籍院士, 加拿大工程院、工程研究院院士以及皇家学会会员, 加拿大滑铁卢大学教授, IEEE Fellow, 主要研究方向为空天地一体化网络、车联网、网络安全、人工智能及智能电网等。