



基于图卷积网络的交通流预测方法综述

摘要

近年来,基于深度学习的交通流预测方法一直是交通预测领域的研究热点.与传统卷积神经网络不同,适合处理非欧几里得数据的图卷积网络在空间特征建模方面表现出了强大的能力,而反映路网空间特征的拓扑图、距离图、流量相似图等正是典型的非欧几里得数据.因此,基于图卷积网络及其变体的交通流预测方法成为交通预测领域的一个研究热点,并取得了许多有吸引力的研究成果.本文对近年来基于图卷积网络的交通流预测模型进行了分类和总结.首先,从图卷积网络的基本定义出发,结合空域图卷积和谱域图卷积的定义详述了图卷积的基本原理.其次,根据预测模型的网络结构特点,将基于图卷积网络的交通流预测模型分为“组合型”和“改进型”两大类,并对其中最具有代表性的模型结构进行了详细分析和讨论.此外,对交通流预测领域中常用于模型性能对比的典型数据集进行了综述,并以其中一个真实数据集为例开展仿真测试,展示了4个基于图卷积网络交通流预测模型的预测性能.最后,基于当前的研究现状和发展趋势,对基于图卷积网络的交通流预测方法研究领域中未来的研究热点和难点进行了开放性的讨论和展望.

关键词

深度学习; 交通拥堵; 图卷积网络; 交通流预测

中图分类号 TP181

文献标志码 A

收稿日期 2023-09-05

资助项目 浙江省自然科学基金(LTGS23F030002); 嘉兴市应用性基础研究项目(2023AY11034); 浙江省“尖兵”“领雁”研发攻关计划(2023C01174); 国家自然科学基金(61603154); 工业控制技术国家重点实验室开放课题(ICT2022B52)

作者简介

叶宝林,男,博士,副教授,主要从事深度学习、强化学习及其在智能交通系统中的理论和应用研究.yebaolin@zjxu.edu.cn

0 引言

随着我国城市化进程的不断深入,交通拥堵问题日益凸显.交通拥堵不仅会大幅增加日常的通勤时间,还会产生额外的能源消耗和碳排放.因此,如何缓解交通拥堵是很多大中城市亟待解决的问题.交通流预测可以根据实时监测的交通数据,预测未来一段时间内的交通状态信息.预测得到的交通状态信息可以为交通管理部门的交通管控、调度和决策提供依据^[1-2],提高公共交通资源的利用效率.目前,交通流预测已被广泛应用于城市路网交通流预测和交通态势感知等,是提升路网通行效率的关键技术之一.

从本质上来说,交通流预测是一个典型的时间序列预测问题.与一般的时间序列预测任务不同的是,一个节点的交通流量会受到其附近节点流量的影响,因此,交通流预测时还需要考虑不同节点之间的空间关系.此外,交通流预测还面临着传播时延性、多步预测误差累计、突发异常事件干扰等诸多挑战.在早期的研究中,研究人员大多采用基于数理统计的方法构建交通流预测模型.譬如历史平均值(Historical Average, HA)^[3-4]、自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)^[5-7]、向量自回归模型(Vector Autoregressive Model, VAR)^[8]等.这些方法通常要求数据满足某些假设,来使交通流预测模型发挥出较好的性能.但由于交通流量往往具有复杂多变的特点,这些预测模型在实际交通预测任务中的表现并不太理想.近年来,得益于人工智能技术的迅猛发展,基于机器学习的交通流预测方法的研究成为了一个热门的研究方向.虽然与基于数理统计的方法相比,基于K-近邻算法(K-Nearest Neighbor, KNN)^[9]和支持向量机(Support Vector Machine, SVM)^[10]等早期机器学习的方法可以建模更复杂的数据,但需要非常细致的特征工程,在处理大规模数据时表现不佳.得益于深度学习在语音识别和图像处理等诸多领域取得的重要进展,越来越多的研究人员开始将深度学习方法应用于交通流预测任务中.卷积神经网络(Convolutional Neural Network, CNN)^[11]、长短期记忆网络(Long Short-Term Memory, LSTM)^[12]及其变体门控循环单元(Gated Recurrent Unit, GRU)^[12]是

1 浙江理工大学 信息科学与工程学院,杭州,310018

2 嘉兴大学 信息科学与工程学院,嘉兴,314001

3 浙江大学 智能控制与系统研究所,杭州,310027

交通流预测领域中最常用、最具代表性的深度学习方法,这些方法将交通流预测的准确性提升到了一个全新的高度.值得注意的是,LSTM和GRU主要用于时间序列预测,不能有效捕捉路网拓扑的空间关联关系.在完全失去空间特征信息的情况下,这些网络对路网拓扑的表达能力将会受到严重限制.为了充分利用路网拓扑的空间关联关系,CNN被引入交通流预测领域,并与LSTM和GRU等时间序列模型结合起来共同提取交通流数据中隐含的时空特征信息^[13-14].

然而,CNN最初是用于处理图像、视频等具有规则网格结构的欧几里得数据,它使用一个具有固定形状的卷积核进行卷积.但对于不规则的“图”来说,图中相邻节点间的关联关系均是随时间发生变化的,很难使用一个固定形状的卷积核进行卷积.因此,传统的卷积神经网络无法有效地处理图结构的数据.事实上,实际的道路网结构更接近于不规则的“图(graph)”而非规则的“图片(image)”,使用CNN来建模路网的空关联关系无法很好地贴合实际路网.在近几年的研究中,图卷积网络(Graph Convolutional Network, GCN)将传统的卷积操作推广到图结构的数据中,解决了CNN无法处理非欧几里得数据的问题.由于GCN能够更好地建模复杂路网的空间结构,越来越多的研究人员开始关注基于图卷积网络GCN的交通流预测模型^[15],并取得了一系列振奋人心的研究成果.因此,本文系统地梳理了近年来国内外有关基于图卷积网络的交通流预测模型,对该领域的研究现状和未来发展趋势进行了深入探讨和展望.考虑到现有最新研究成果均是在最基础的图卷积网络上拓展而来,本文将基于图卷积网络的交通流预测模型分为“组合型”和“改进型”两大类,并重点分析和讨论了其中一些典型模型和代表性方法的特色、优势和局限性,以期为该领域后续进一步的研究提供一些可供借鉴和参考的思路.

本文其余章节的组织如下:第1节具体介绍了空域图卷积和频域图卷积;第2节对基于图卷积网络GCN的交通流预测模型进行了分类和总结;第3节介绍了交通流预测领域常用的公开数据集;第4节结合具体的实例展示了基于图卷积网络GCN的交通流预测方法的预测性能;第5节,对基于图卷积网络GCN的交通流预测方法研究领域未来的研究方向进行了开放性的讨论;第6节对全文进行了总结和展望.

1 图卷积网络

传统的卷积操作通常只适用于对规则欧几里得空间数据进行卷积,而图卷积将卷积操作推广到了一般的图结构中.在目前的研究中,图卷积主要有两种计算方法:空域图卷积方法^[16]和基于图傅里叶变换的频域图卷积方法^[17].前者将图中的顶点按照某种规则重新排列成特定的网格形式,使其可以直接在空间域上进行卷积运算,即空域图卷积.后者则是利用卷积定理将卷积转换至谱域进行计算,即频域图卷积.

1.1 空域图卷积

空域图卷积的通用框架主要有两种:一种是消息传递神经网络(Message Passing Neural Network, MPNN)^[18],它的出发点是节点之间的信息聚合,其核心是聚合函数;另一种是混合卷积网络(Mixture model Networks, MoNet)^[19],它立足于图的平移不变性,通过合适的映射函数将拓扑上每个节点的局部结构映射为相同大小的向量以便进一步学习共享卷积核.消息传递神经网络(MPNN)对空域图卷积聚合函数的一般架构进行了定义,利用聚合函数对各个节点的自身信息和周围信息进行聚合和更新.空域卷积过程被分解为消息传递与状态更新两个过程,可分别用消息传递函数 $M_t(\cdot)$ 与更新函数 $U_t(\cdot)$ 来表示.消息传递过程中,图中节点隐含状态 h_x^t 通过传递的消息 m_x^{t+1} 进行更新,具体更新公式如下:

$$m_x^{t+1} = \sum_{y \in N(x)} M_t(h_x^t, h_y^t, e_{xy}), \quad (1)$$

$$h_x^{t+1} = U_t(h_x^t, m_x^{t+1}). \quad (2)$$

式中: h_x^t 和 h_y^t 分别表示节点 x 和节点 y 在第 t 步输出的隐藏状态; e_{xy} 表示连接节点 x 和节点 y 的边上的特征; $N(x)$ 表示节点 x 的邻居节点构成的集合.虽然基于MPNN框架可以直接在空域进行卷积,但面对庞大的图数据时,基于MPNN框架的空域卷积方法需要占用庞大的计算资源.另外,平移不变性的缺失还会给图卷积网络的定义带来困难.

考虑图上的任意节点 x 和它的邻域节点 $y \in N(x)$, MoNet利用图上定义的伪坐标系,将节点 y 映射为伪坐标系下的 d 维向量 $\mathbf{u}(x, y)$,以便学习共享卷积核函数. MoNet^[19]中定义了一个如式(3)所示的加权函数:

$$\mathbf{w}_\Theta(\mathbf{u}) = (w_1(\mathbf{u}), \dots, w_i(\mathbf{u}), \dots, w_n(\mathbf{u})). \quad (3)$$

式中: Θ 表示可学习的参数; $w_i(\cdot)$ 表示第 i 个权重函数; n 表示总的权重函数个数.

利用构建的核函数可为图中各个节点获得相同维度的向量表示:

$$D_i(x)f = \sum_{y \in N(x)} w_i(\mathbf{u}(x, y))f(y), i = 1, \dots, n. \quad (4)$$

式中: $N(x)$ 表示节点 x 的邻居节点构成的集合, $f(y)$ 表示 y 在函数 f 上的取值. 基于式(4) 计算的 $D_i(x)f$ 可得到在非欧几里得域上的卷积空间推广:

$$(f *_c g)(x) = \sum_{i=1}^n g_i D_i(x)f. \quad (5)$$

式中: g 表示图卷积核; $*_c$ 表示图卷积算子.

1.2 频域图卷积

1.2.1 图信号的频域处理

卷积定理是傅里叶变换的一个重要性质, 它可以将时域上的卷积运算变换到频域上的乘积运算, 具体计算公式如下:

$$F(f *_c g) = F(f) \cdot F(g). \quad (6)$$

式中: f 和 g 分别表示两个原始的输入信号; $F(f)$ 和 $F(g)$ 分别表示输入信号 f 和 g 对应的傅里叶变换; $*$ 表示卷积运算. 对式(6) 两端作傅里叶逆变换, 可以得到:

$$f *_c g = F^{-1}(F(f) \cdot F(g)). \quad (7)$$

利用式(6) 可以将图数据的卷积运算转换成频域中乘积运算, 然后再利用式(7) 将其转换至空间域, 以实现完整的图卷积过程. 图拉普拉斯矩阵是实现图傅里叶变换的基础. 对于一个图信号 \mathbf{x} , 其图傅里叶变换可表示为

$$\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x}. \quad (8)$$

式中: \mathbf{x} 为空间域的原始信号; $\hat{\mathbf{x}}$ 为转换至频谱域后的信号; \mathbf{U}^T 表示特征向量矩阵的转置. 类似地, 图信号 \mathbf{x} 的傅里叶反变换可表示为

$$\mathbf{x} = \mathbf{U} \hat{\mathbf{x}}. \quad (9)$$

进一步, 基于卷积定理实现的图卷积操作过程可表示为

$$\mathbf{x} *_c \mathbf{y} = \mathbf{U}(\mathbf{U}^T \mathbf{x}) \odot (\mathbf{U}^T \mathbf{y}). \quad (10)$$

式中: \mathbf{x}, \mathbf{y} 表示图信号; \mathbf{U} 表示特征向量矩阵; $*_c$ 表示图卷积算子; \odot 表示哈达玛积. 若用对角阵 \mathbf{g}_θ 替换 $\mathbf{U}^T \mathbf{y}$, 式(10) 的哈达玛积便可转化为矩阵乘法. 同时, 式(10) 可以表述为卷积核 \mathbf{g}_θ 对信号 \mathbf{x} 进行滤波, 图卷积便可进一步表示为如下形式:

$$\mathbf{U} \mathbf{g}_\theta \mathbf{U}^T \mathbf{x}. \quad (11)$$

1.2.2 切比雪夫多项式近似

在谱图分析中, 图结构的数据一般用其相应的拉普拉斯矩阵表示. 通过分析拉普拉斯矩阵及其特征值, 可以得到对应图结构的性质. 图的拉普拉斯矩

阵被定义为 $\mathbf{L} = \mathbf{D} - \mathbf{A}$, 其规范化形式是 $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \in \mathbf{R}^{N \times N}$, 其中, \mathbf{A} 表示邻接矩阵, \mathbf{I}_n 表示单位矩阵, \mathbf{D} 表示度矩阵, $\mathbf{D} \in \mathbf{R}^{N \times N}$ 为对角矩阵, D_{ii} 表示节点的度, $D_{ij} = \sum_j D_{ij}$. 拉普拉斯矩阵 \mathbf{L} 可分解为 $\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$, 其中, $\mathbf{\Lambda} = \text{diag}([\lambda_0, \dots, \lambda_{N-1}]) \in \mathbf{R}^{N \times N}$ 是一个对角矩阵, \mathbf{U} 是傅里叶基. 以交通流量为例, 假设图上的信号为 $\mathbf{x} \in \mathbf{R}^{t \times N \times N}$, 由式(8) 知信号的图傅里叶变换为 $\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x}$, 其对应的傅里叶反变换是 $\mathbf{x} = \mathbf{U} \hat{\mathbf{x}}$. 作为一种卷积运算, 图卷积利用在傅里叶域内对角化的线性算子来代替经典的卷积算子^[20]. 在此基础上, 卷积核 \mathbf{g}_θ 对图 G 上的信号 \mathbf{x} 滤波表示为

$$\mathbf{g}_\theta *_c \mathbf{x} = \mathbf{g}_\theta(\mathbf{L}) \mathbf{x} = \mathbf{g}_\theta(\mathbf{U} \mathbf{\Lambda} \mathbf{U}^T) \mathbf{x} = \mathbf{U} \mathbf{g}_\theta(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{x}. \quad (12)$$

其中: $*_c$ 表示图卷积算子; $\mathbf{g}_\theta(\mathbf{\Lambda})$ 为一个对角阵. 由卷积定理可知, 图信号的卷积运算等于图信号傅里叶变换到频谱域的信号的乘积, 因此, $\mathbf{g}_\theta *_c \mathbf{x}$ 可以理解为将 \mathbf{g}_θ 和 \mathbf{x} 分别进行傅里叶变换到频谱域, 将其变换后的结果相乘, 然后通过傅里叶反变换得到卷积运算的最终结果. 由于 \mathbf{U} 相乘的复杂度为 $O(N^2)$, 当图的维度较大时, 直接对拉普拉斯矩阵进行特征值分解复杂度较大. 因此, 采用切比雪夫多项式^[21] 进行近似达到高效运算的目的:

$$\mathbf{g}_\theta *_c \mathbf{x} = \mathbf{g}_\theta(\mathbf{L}) \mathbf{x} = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}}) \mathbf{x}. \quad (13)$$

式中: 参数 $\theta \in \mathbf{R}^K$ 表示多项式系数向量; $\tilde{\mathbf{L}} = \frac{2}{\lambda_{\max}} \mathbf{L} - \mathbf{I}_n$, λ_{\max} 是拉普拉斯矩阵的最大特征值. 切比雪夫多项式的递归定义为: $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, 其中, $T_0(x) = 1, T_1(x) = x$. 利用切比雪夫多项式近似展开该求解公式, 对应于利用卷积核 \mathbf{g}_θ 提取图中以每个节点为中心的 0 到 $K-1$ 阶邻域节点的信息. 通过多项式近似递归地计算 K 阶邻域卷积, 等式(13) 的时间复杂度从 $O(N^2)$ 降为 $O(K|E|)$. 最后, 图卷积模块使用 $\text{ReLU}(\cdot)$ 作为激活函数, 形式上可以表示为 $\text{ReLU}(\mathbf{g}_\theta *_c \mathbf{x})$.

用图拉普拉斯的一阶近似叠加多个局域图卷积层可以定义分层线性公式^[22]. 由于需要对数据进行标准化和归一化操作, 可以进一步假设 $\lambda_{\max} = 2$. 因此, 式(13) 可简化为

$$\mathbf{g}_\theta *_c \mathbf{x} \approx \theta_0 \mathbf{x} + \theta_1 \left(\frac{2}{\lambda_{\max}} \mathbf{L} - \mathbf{I}_n \right) \mathbf{x} \approx \theta_0 \mathbf{x} - \theta_1 (\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}) \mathbf{x}. \quad (14)$$

式中: θ_0, θ_1 是核的共享参数. 为了进一步限制参数的数量以解决过拟合问题, θ_0 和 θ_1 用单一参数 θ 代替, 令 $\theta = \theta_0 = -\theta_1$, 图卷积操作可进一步简化为

$$g_{\theta} *_{\text{c}} x \approx \theta (I_n + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) x. \quad (15)$$

为了避免深度图卷积网络中梯度爆炸、消失问题并稳定数值性能, 一般需要在网络中进行归一化操作, 即用 $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ 代替 $I_n + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. 其中, $\tilde{A} = A + I_n, \tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. 最后, 图卷积运算可表示为

$$g_{\theta} *_{\text{c}} x \approx \theta (\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}) x. \quad (16)$$

2 基于图卷积的交通流预测模型

交通流数据本质上是包含了时空信息的时间序列数据. 如何改进预测模型, 使其能够从交通流数据中更高效、更准确地提取时空特征, 是交通流预测的重点研究内容. 采用图结构定义的道路交通网络, 其节点的特征信息可以视为图上的信号^[20]. 为了充分利用交通路网的空间拓扑特征信息, 可以在每个时间步上采用图卷积网络来直接对路网空间特征信息进行处理, 从而更好地利用路网交通流数据在空间维度上的特征相关性, 来提升交通流预测模型的性能. 近年来, 在基于图卷积网络(GCN)交通流预测模型的基础上, 已经衍生出了许多基于图卷积网络改进的交通流预测模型. 这些改进工作大致可以分成两种: 一种致力于构建更加有效的图结构, 以更好地表示不同节点之间的空间依赖关系; 另一种则侧重于研究如何更有效地将 GCN 与其他神经网络相结合, 以充分挖掘交通流数据中的时空关联关系, 提升交通流预测模型的性能. 根据这两种改进思路, 本文将基于 GCN 的交通流预测模型分为“改进型”和“组合型”两大类. 其中: 改进型可细分为动态图卷积网络、多图卷积网络、自适应图卷积网络 3 种; 组合型可细分为时空图卷积网络、残差网络+图卷积网络、自编码器+图卷积网络、注意力机制+图卷积网络、生成对抗网络+图卷积网络、Transformer+图卷积网络、模式分解+图卷积网络 7 种. 需要注意的是, 不同的分类之间可能存在一定程度的交叉. 下面将根据上述分类对具有代表性的模型的预测结果进行详细阐述.

2.1 改进型

相比于传统的卷积神经网络(CNN), 图卷积网络(GCN)能够更好地建模路网中不同节点之间的空间关联关系, 是近年来交通流预测领域中最常用的

神经网络之一. 但图卷积网络中常用的预定义的静态图存在一些明显的缺陷, 导致其建模的空间关联关系可能和实际情况存在一定偏差. 首先, 预定义的静态图假设不同节点之间的空间关联关系是预先确定且不随时间变化的, 然而, 实际路网中的空间关联关系是随时间变化的, 静态的图结构难以反映这种时变的时空关联关系. 此外, 预定义的静态图主要是基于不同节点之间的距离构建图结构, 并未考虑其他深层的空间关联关系. 这种单一的图结构所承载的信息较少, 难以准确地描述复杂路网中隐含的丰富拓扑结构信息. 另外, 由于设备故障等原因, 真实数据集中经常存在部分数据缺失的问题. 如果仅利用这些数据构建图结构, 可能会丢失部分空间信息. 针对上述缺陷, 研究人员主要从动态图卷积网络、多图卷积网络和自适应图卷积网络 3 个方向进行改进.

2.1.1 动态图卷积网络

在早期基于图卷积网络的交通流预测方法中, 描述路网中不同节点之间空间关联关系的结构图在训练过程中始终保持不变. 但事实上, 实际路网中不同节点之间空间关系是随时间变化的, 静态的空间结构图不能很好地描述这种变化的空间拓扑结构. 为了解决该问题, Diao 等^[23]提出一种基于动态时空图卷积(DGCNN)模型. 在 DGCNN 模型中, 实时交通数据被分解为一个依赖长期时空交通关系的、稳定的全局节点嵌入, 以及一个捕获交通变化的局部节点嵌入. 他们使用动态拉普拉斯矩阵估计器来感知交通流量的变化, 并据此对路网空间关系的变化进行动态建模. Zheng 等^[24]提出了一个动态时空相邻图卷积网络(DSTAGCN). DSTAGCN 是利用基于模糊神经网络的数据驱动模块来学习交通图的邻接矩阵, 可以更有效地挖掘交通数据之间潜在的和不确定的关联关系. 该模型也是第一个利用模糊神经网络来学习图卷积方法中邻接矩阵的模型.

动态图卷积是基于注意力机制实现的, 然而传统注意力机制计算的注意力得分可能存在较大误差. 因此, Hu 等^[25]提出一种基于多注意力机制时空图卷积网络(MASTGCN)的交通流预测模型, 用以实现对时空特征的动态调整. 此外, 随着 Transformer 在交通流预测领域的广泛应用, 基于传统注意力机制的动态图也面临着巨大的挑战. 相比于传统的注意力机制, Transformer 中的自注意力机制计算的注意力得分更加准确. 因此, 一些研究者开始使用自注意

力机制来代替传统的注意力机制构建动态图卷积网络.例如, Ma 等^[26]提出一种基于多头自注意力机制时空信息图卷积网络 (MSASTIGCN) 的交通流预测方法.在该方法中,将传统的注意力机制替换为多头自注意力机制,使模型能够更准确地捕获路网空间关系的动态变化,在一定程度上提高了模型的预测性能.

2.1.2 多图卷积网络

在大多数基于单图的图卷积网络中,相邻节点之间的空间相关性是根据它们之间的距离定义的.但事实上,节点之间的距离并不能完全地代表它们之间的空间关联关系.例如,距离较近的两个节点之间可能并无直接连接关系,距离较远的两个节点之间的流量也可能存在某种相似性.单一的图结构难以充分地描述这些隐含的空间关联关系.因此,一些研究人员使用多图融合的思想,从不同的角度来更加全面地描述节点之间的空间关联关系.例如, Lv 等^[27]提出一种基于时间多图卷积网络 (T-MGCN) 的交通流预测模型,并将研究重点放在建模更多可能的遥远节点之间的语义相关性上.为了更加有效地建模全局空间关联关系, Lv 等^[27]不仅构建了描述空间关联关系的拓扑图,也构建了交通模式图和功能图.另外,为了提升模型的预测性能,该文还设计了一个门控循环单元 (GRU) 捕获时间相关性.为了提高 GCN 捕获节点间空间关联关系的能力, Yang 等^[28]设计了一种可分解图卷积结构 (FactorGCN).在 GCN 中,图被广泛用于表示实体之间的关联关系.多数情况下,这种异构的关系是交织在一起的.可分解图卷积结构 FactorGCN 则明确地解开了编码在图中的这种关系.该文先将一个原始的图分解为多个子图,每个子图表示节点之间潜在的、分解的空间关联关系.然后,在每个分解的潜在空间中分别聚合节点的特征以产生分离的特征,从而进一步提高下游预测任务的性能. Yin 等^[29]提出一种基于多时间-多图神经网络 (MTMGNN) 的地铁客流预测模型.该模型采用了 6 种图来表示不同模式下地铁站不同站点间的空间连接.另外,针对复杂但有规律的地铁客流波动特征,该文同时利用近期客流数据和长期客流数据,并从长期客流数据中挖掘与人们出行习惯密切相关的惯性特征. Lee 等^[30]从路网中节点间的距离、方向、位置关系 3 个维度分别构建空间关联关系图,并在此基础上提出了一种基于多图卷积网络的交通流预测模型.仿真测试结果表明,该模型在解决

复杂的城市路网长期交通流预测问题上具有一定优势.针对网约车需求预测问题, Geng 等^[31]将不同区域间的非欧几里得相关性编码为多个图,然后使用多图卷积对这些相关性进行建模,并在此基础上提出了一种基于时空多图卷积网络 (ST-MGCN) 的网约车需求预测模型.该模型可以更加充分地挖掘相邻区域和非相邻区域的空间相关性,提高了网约车需求预测的准确性.针对交通流预测问题, Khaled 等^[32]提出一种带有多图卷积网络的生成对抗网络 (TFGAN) 模型.该文从空间相似性、相关性、空间距离 3 个不同的视角创建邻接矩阵描述路网的空间关联关系,并利用多个 GCN、GRU、自注意力机制来学习路网中节点历史流量序列的局部相关性.

2.1.3 自适应图卷积网络

现有的 GCN 一般利用拓扑关系,或是各种相似度量来定义物理或虚拟图,并以此来衡量不同节点之间的空间相关性.这种图一般由相关领域的专家基于一些先验知识构建,图中的节点之间有着固定的边缘连接,无法很好地贴合实际情况.因此,一些研究人员尝试利用数据驱动学习图结构的方式来实现图的自动更新.例如, Wu 等^[33]提出的 Graph WaveNet,通过构造一个自适应邻接矩阵,保留了隐藏的空间依赖性.同时能够从数据中发掘不可见的图结构,而无需任何先验知识的指导. WaveNet 的卷积层可表示为

$$\mathbf{Z} = \sum_{k=0}^K \mathbf{P}_f^k \mathbf{X} \mathbf{W}_{k1} + \mathbf{P}_b^k \mathbf{X} \mathbf{W}_{k2} + \tilde{\mathbf{A}}_{\text{apt}}^k \mathbf{X} \mathbf{W}_{k3}. \quad (17)$$

式中: $\tilde{\mathbf{A}}_{\text{apt}}^k$ 为自适应邻接矩阵; $\mathbf{P}_f^k, \mathbf{P}_b^k$ 分别为正向和反向转换矩阵.

Bai 等^[34]设计了一个 DAGG 模块,可以从数据中自适应地推断隐藏的空间依赖关系. DAGG 模块首先为所有节点随机初始化一个可学习的节点嵌入 $\mathbf{E}_A \in \mathbf{R}^{N \times d_{\text{emb}}}$, 其中,每一行 \mathbf{E}_A 表示单个节点的嵌入, d_{emb} 表示节点嵌入的维数.然后,类似于求节点相似度来定义图,通过将 \mathbf{E}_A 和 \mathbf{E}_A^T 相乘来推断每对节点之间的空间依赖性.结果表明,使用自适应图的模型比依赖单一静态图结构的模型具有更加优秀的性能.然而,一般的自适应图仅能在训练阶段发挥作用,无法在测试阶段利用测试数据自适应地更新图结构.由于交通数据经常受到时间序列中的异常变化的影响,上述缺点可能会导致某些情况下预测的性能明显下降.为了解决上述问题, Shin 等^[35]提出一种渐进式图卷积模型 (PGCN).该模型通过学习图节

点之间的趋势相似性,实现了构造渐进式邻接矩阵对路网拓扑进行渐进式学习.为了捕获动态空间相关性,Wang 等^[36]提出一种时变图卷积模型(TVGCN).他们设计了一个时变空间卷积模块,可以在不需要任何先验知识的情况下对复杂的路网拓扑结构进行提取.TVGCN 中构建了静态、动态两种自适应图.一个图以数据驱动的方式学习来捕获交通节点之间的静态空间相关性,而另一个图用于自适应地建模不同时间的动态空间相关性.在图卷积网络(GCN)的框架下,他们将这两个图分层组合在一起同时捕获交通图的静态和动态空间相关性.谷振宇等^[37]提出的 STG-CRNN 在预定义的距离图、交通模式图的基础上结合自适应图构建与现实路网更加贴合的图结构,提升模型对空间相关的建模能力.

2.2 组合型

交通流数据中包含了丰富的时间和空间信息,如何更有效地提取、利用这些时空信息是提高交通流预测精度的关键所在.在主流的深度学习方法中,由于 GCN 对非欧几里得数据有较强的建模的能力,因此,常被用于捕获路网中不同节点之间的空间相关性.然而,GCN 对时间序列数据不敏感,难以捕获某个节点的流量随时间的变化关系.因此,GCN 常会与 CNN、循环神经网络(RNN)、Transformer 等结合使用,以共同捕获时空相关性.此外,一些研究人员还将 GCN 与注意力机制(Attention Mechanism)、生成对抗网络(GAN)等结合以提高其捕获空间相关性的能力.总的来说,有关组合型 GCN 的研究主要从时空图卷积网络、残差网络+图卷积网络、自编码器+图卷积网络、注意力机制+图卷积网络、生成对抗网络+图卷积网络、Transformer+图卷积网络、模式分解+图卷积网络 7 个方向展开.需要注意的是,2.1 节中对图结构的改进同样可以应用至本小节的模型中,对于此类改进,本文将不再展开分析.

2.2.1 时空图卷积网络

如图 1 所示,由于交通数据存在明显的时序特征,许多时序预测模型和图卷积相结合被应用于交通预测问题.时空图卷积网络主要可分为两类:基于 RNN 的时空图卷积网络和基于 CNN 的时空图卷积网络.

1) 基于 RNN 的时空图卷积网络

RNN 是一个经典的时序预测模型,曾被广泛应用于各类时序预测任务.但 RNN 在计算过程中存在大量的连乘操作,这会导致严重的梯度消失和梯

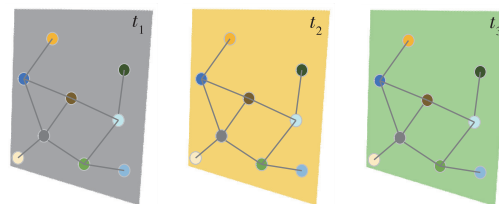


图 1 图结构化的交通数据

Fig. 1 Graphically structured traffic data

度爆炸问题.不同于 RNN, LSTM 缓解了梯度消失和梯度爆炸问题.另外, LSTM 中的门控机制可以有选择性地关键信息保存至记忆细胞中,大幅提高了建模长期时间依赖性的能力.因此, LSTM 常与 GCN 共同组成时空图卷积网络.例如,为了建模路网中交通流的时空相关性, Cui 等^[38]提出一种基于 TGC (Traffic Graph Convolution) 和 LSTM 的路网交通流预测模型.该模型的建模过程可描述为

$$f_t = \sigma_g(W_f \cdot GC_t^{(K)} + U_f \cdot h_{t-1} + b_f), \quad (18)$$

$$i_t = \sigma_g(W_i \cdot GC_t^{(K)} + U_i \cdot h_{t-1} + b_i), \quad (19)$$

$$o_t = \sigma_g(W_o \cdot GC_t^{(K)} + U_o \cdot h_{t-1} + b_o), \quad (20)$$

$$\tilde{C}_t = \tanh(W_c \cdot GC_t^{(K)} + U_c \cdot h_{t-1} + b_c), \quad (21)$$

$$C_t = f_t \odot C_{t-1}^* + i_t \odot \tilde{C}_t, \quad (22)$$

$$h_t = o_t \odot \tanh(C_t). \quad (23)$$

式中: W, U, b 均为 LSTM 网络中的可学习参数; σ 是 sigmoid () 激活函数; $C_{t-1}^* = W_N \odot (\tilde{A}^K \odot F_{FR}) \cdot C_{t-1}$; W_N 为学习相邻细胞状态贡献的权重矩阵; F_{FR} 为自由流可达矩阵,它表示相邻节点对和附近节点对之间影响的可达性; \tilde{A}^K 是由式子 $\tilde{A}_{ij}^K = \min((A + I)_{ij}^K, 1)$ 定义的 k 跳邻域矩阵,其中, A 是邻接矩阵, I 是单位矩阵.

Ali 等^[39]提出一种用于城市人流预测的动态深度时空神经网络(GCN-DHSTNet)模型.该模型采用 LSTM 分别提取每小时的流量特征、每天的流量特征和每周的流量特征,利用 GCN 提取空间关联关系来完成预测任务.为了建模交通流数据中的正向和反向时间特征信息,张阳等^[40]提出一种基于 GCN-BiLSTM 的短时交通流预测模型,其中, Bi-LSTM 网络可以同时从正向和反向提取交通流量数据中的时间特征信息, GCN 网络则用于挖掘交通流量数据中的空间特征信息. Zhang 等^[41]在 LSTM 中引入残差连接并与图卷积组合,提出一种基于残差图卷积长短时记忆网络(RGC-LSTM)的短时流量预测模型.

GRU 是 RNN 的另一种变体.与 LSTM 相比, GRU 将 3 个控制门减少至 2 个,在保持预测精度基本相同

的前提下,降低了模型的计算复杂度.因此,很多研究人员用 GRU 代替 LSTM 提取序列数据中的时间特征.例如,Zhao 等^[42]设计了一种基于时间图卷积网络(Temporal Graph Convolutional Network, T-GCN)的交通流预测方法,该方法采用 GRU 与 GCN 联合提取时空特征.Chen 等^[43]在其提出的时空图神经网络(Spatial Temporal Graph Neural Network, STGNN)中设计了一种具有位置级注意机制的图神经网络层,以有效地聚集来自相邻道路的空间关联信息.另外,该模型使用 GRU 和 Transformer 分别捕获局部和全局时间依赖性.Zhao 等^[44]提出一种基于 GCN 和双向门控递归单元(BiGRU)的交通流预测模型实现多步交通流预测.BiGRU 包含一个正向传播 GRU 和一个反向传播 GRU,可以对交通时间序列数据同时进行前向和后向建模,挖掘数据中隐藏的上下文信息,并捕获交通时间序列的长期时间相关性.

2) 基于 CNN 的时空图卷积网络

RNN 及其变体虽然能有效地捕获时间序列中的时间特征,但由于在 RNN 中下一时刻的输出完全依赖于上一时刻的输入,导致难以进行并行化训练.为了能够进行并行化训练,通常采用 CNN 来提取时间特征.譬如,陈拓等^[45]构建了一种基于时间卷积神经网络(Temporal Convolutional Network, TCN)和 GCN 的交通流预测模型,共同提取时空特征用于出租车需求预测.Yu 等^[46]提出一种新的时空图卷积网络(STGCN),由于同时包含有图卷积和门控时间卷积,STGCN 模型可以在提取空间特征的同时,有效地捕获时间特征.由于 STGCN 模型完全由 CNN 组成,因此其训练速度明显优于基于 RNN 的时空图卷积模型.另外,为了捕获路网中动态变化的空间关联

性,Diao 等^[23]使用动态拉普拉斯矩阵估计器对原有的时空图卷积网络 STGCN 模型进行了改进,提出一种动态图卷积神经网络(DGCNN)模型.如图 2 所示,在改进的 DGCNN 模型中,拉普拉斯矩阵可以有效感知空间关联性的动态变化.为了更精准地捕获数据中的时空依赖关系,Wu 等^[33]将图卷积与扩展因果卷积相结合,提出一种新的图神经网络体系结构 GWN(Graph WaveNet).图神经网络模型 GWN 具有一个堆叠的扩展 1D 卷积组件,其感受野随着层数的增加呈指数增长,因此能够处理长时间序列.该模型中的扩张因果卷积结构如图 3 所示.

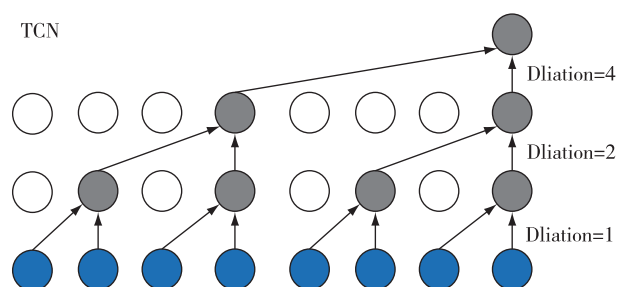


图 3 扩张因果卷积^[33]

Fig. 3 Dilated casual convolution^[33]

扩张因果卷积操作的具体定义如下:

$$\mathbf{x} * \mathbf{f}(t) = \sum_{s=0}^{K-1} \mathbf{f}(s) \mathbf{x}(t - d \times s). \quad (24)$$

式中: \mathbf{f} 是扩张卷积核; d 是扩张系数;*表示卷积运算.由于在模型中引入了自适应邻接矩阵,该模型无需任何先验知识的指导也能够从数据中自动学习隐藏的图结构.该模型中时间特征处理单元采用了门控时间卷积,具体计算过程如下:

$$\mathbf{h} = g(\boldsymbol{\theta}_1 * \mathbf{X} + \mathbf{b}) \odot \sigma(\boldsymbol{\theta}_2 * \mathbf{X} + \mathbf{c}). \quad (25)$$

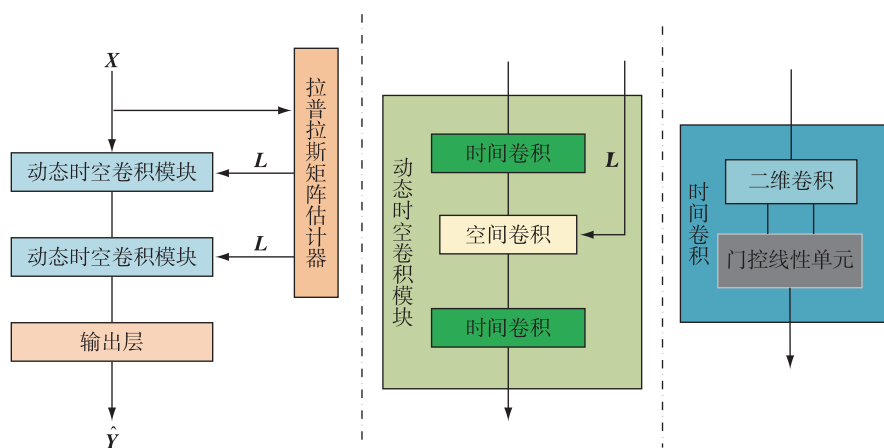


图 2 动态时空图卷积网络框架示意^[23]

Fig. 2 Framework of dynamic spatio-temporal graph convolution network^[23]

式中: $\theta_1, \theta_2, \mathbf{b}, \mathbf{c}$ 为模型参数; $g(\cdot)$ 和 $\sigma(\cdot)$ 分别是 $\tanh(\cdot)$ 和 $\text{sigmoid}(\cdot)$ 激活函数; \odot 表示哈达玛积. 为了解决 GWN (Graph WaveNet) 只能在训练集上学习隐藏图结构的缺陷, Shin 等^[35] 提出一种基于 GWN 的渐进图卷积网络 (PGCN) 模型. 在 PGCN 模型中, 他们同样采用了扩张因果卷积来捕获数据序列中的时间特征. 不同于 GWN 模型, 因为构造了一个能适应数据的渐进图, PGCN 模型可以更好地学习到随时间变化的空间相关性. 在 PGCN 模型中, 具体计算过程定义为

$$\mathbf{z}_t = \mathbf{X}_t *_{\mathbf{G}} f_w = \sum_{k=0}^{K-1} (\mathbf{P}^k \mathbf{X}_t \mathbf{W}_{k,1} + \mathbf{P}^{T^k} \mathbf{X}_t \mathbf{W}_{k,2} + \mathbf{A}_p^t \mathbf{X}_t \mathbf{W}_{k,3}). \quad (26)$$

式中: $*_{\mathbf{G}} f_w$ 表示采用 f_w 作为卷积算子的图卷积操作; $\mathbf{W}_{k,1}, \mathbf{W}_{k,2}$ 和 $\mathbf{W}_{k,3}$ 是可学习的参数; \mathbf{P} 和 \mathbf{P}^T 分别用于反映正向和反向扩散过程; \mathbf{A}_p^t 是第 t 个时间步处的渐进邻接矩阵, 它包含从节点信号相似性中学习到成对权值.

2.2.2 残差网络+图卷积网络

在深度神经网络中, 网络的性能并非始终与网络深度呈正相关关系. 换句话说, 如果只是简单地增加网络的深度, 当网络深度超过某个阈值后, 预测模型在训练集和测试集上的性能均有一定下降. 这明显不同于过拟合时预测模型在训练集上性能有所提升, 但在测试集上的性能下降. 直到 2016 年, He 等^[47] 在图像识别领域提出了残差网络以解决上述问题. 残差网络可以使上一个残差块的信息传递到下一个残差块, 提高信息的传递效率, 并且避免了随着网络深度的增加出现的梯度消失问题和退化问题. Zhang 等^[41] 将残差引入 LSTM 提出残差图卷积长短时记忆 (RGC-LSTM) 模型用于交通预测. 该模型具有参数少、计算量小、收敛速度快的优势. Zhao 等^[44]

则在图卷积层中加入残差网络 (ResGCN), 以解决深度卷积网络的退化问题, 其结构示意图如图 4 所示, 在 ResGCN 中还使用双向门控递归单元 (BiGRU) 对交通时间序列数据进行前向和后向建模.

2.2.3 自编码器+图卷积网络

由于自编码器在预测任务上具有更加稳定的表现且泛化效果好^[48], 因此, 研究人员开始将自编码器与图卷积网络结合起来解决交通流预测问题. 譬如, Zhu 等^[49] 提出一种带有注意力机制的多图卷积 Seq2Seq 模型 (AMGC-Seq2Seq) 用于共享汽车流量预测. 在 AMGC-Seq2Seq 模型中的编码器部分, 他们利用 LSTM 网络提取时间特征, 再用多图卷积网络 (M-GCN) 提取空间关联关系. 然后, 再使用一个 LSTM 网络对提取的时空特征进行融合, 以实现同时对空间和时间关系的同时编码. 在解码器中, 一个单独的 LSTM 被用于解码由编码器传递来的上下文向量, 以获得多步预测输出. 其结构示意图如图 5 所示. 另外, 为了建模交通流中的时间关联关系和空间关联关系, Li 等^[50] 将交通流量建模为在有向图上进行的扩散过程, 提出一种基于扩散卷积循环神经网络 (DCRNN) 的交通流预测方法. DCRNN 在利用扩张卷积和 GRU 提取时空依赖的同时, 加入自编码器结构来提升模型性能. 在模型的训练阶段, 历史交通序列被输入到编码器中, 并得到隐藏状态. 解码器利用隐藏状态和真实观测值生成预测结果. 在测试阶段, 首先使用训练阶段获得的最终状态初始化解码器. 与训练阶段不同的是, 解码器的输入由真实的观测值替换为预测值来实现多步预测. 另外, 为了解决训练和测试两个阶段中数据分布之间存在差异性的问题, DCRNN 模型利用计划采样的方法对输入数据进行采样处理. Guo 等^[51] 提出一种用于速度预测的嵌入式图卷积长短时记忆网络 (EGC-LSTM), 将 GCN

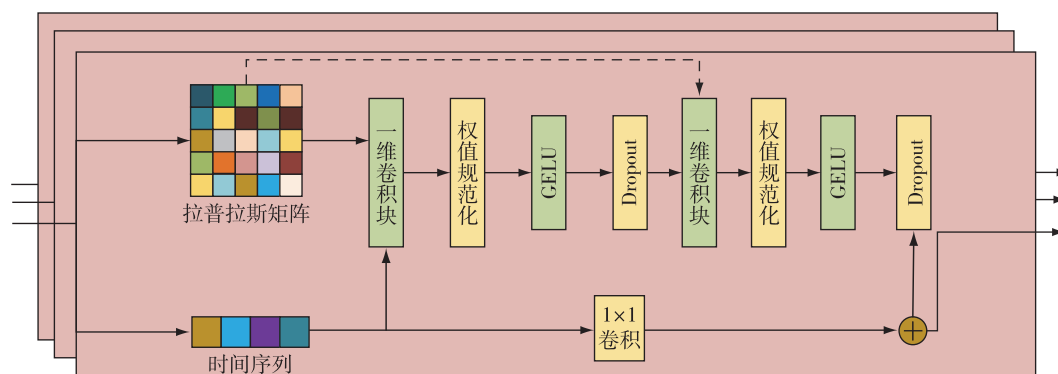


图 4 ResGCN 网络结构^[44]

Fig. 4 ResGCN network structure^[44]

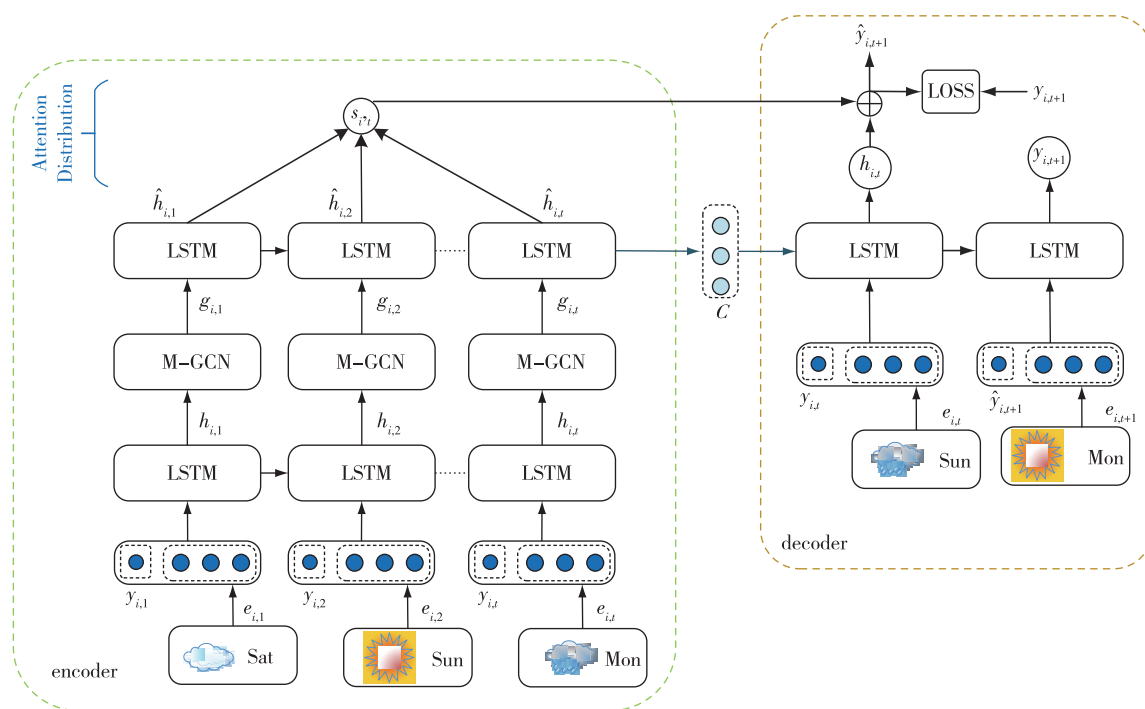


图 5 AMGC-Seq2Seq 结构^[49]

Fig. 5 AMGC-Seq2Seq structure^[49]

和 LSTM 进行融合提取时空特征,并引入注意机制计算特征权重来提高模型在时间维度上的可解释性.为了进一步提高预测准确性,他们在模型中引入了自编码器结构,并将辅助特征进行 one-hot 编码后输入到解码器部分.特征权重可以反映特征的重要性,间接实现了特征选择;将特征权重传递到 GCN-LSTM 中,提高了模型的预测性能.

2.2.4 注意力机制+图卷积网络

注意力机制最初被用于解决自然语言处理领域的相关问题^[52],随后被推广到许多其他的相关任务中.在交通流预测模型中,注意力机制也常与各类神经网络配合使用.例如,葛宇然等^[53]提出一种基于时空域联合学习的动态分层网络(JST-DHNet)模型,并采用注意力机制将通过两种不同的时空学习模块学到的整体时空特征和局部时空特征进行了融合.Zhu等^[49]在其提出的带有注意力机制的多图卷积Seq2Seq(AMGC-Seq2Seq)模型中,利用注意力机制对时空图卷积网络输出的隐藏向量计算注意力分数,并将注意力用于解码器以捕获历史信息中的重点内容.Liu等^[54]在其提出的基于多头注意力机制的时空图卷积网络中,将自注意力定义为

$$A(Q, K_s, V_s) = \text{softmax}\left(\frac{W_i Q_i (W_s K_s)^T}{\sqrt{d_s}}\right) V_s, \quad (27)$$

式中: $Q_i = [q_1, q_2, \dots, q_{T_i}]$ 为查询向量; $K_s = [k_1, k_2, \dots, k_{T_s}]$ 是键向量; $V_s = [v_1, v_2, \dots, v_{T_s}]$ 是值向量; W_i, W_s 为可学习参数; d_s 为查询向量和键向量的维度.为了捕捉丢失的局部时间信息, Liu等^[54]提出了如式(28)所示的局部自注意力块:

$$A_L = \text{softmax}\left(\frac{W_i Q_i (W_s K_s)^T * W_m}{\sqrt{d_s}}\right) V_s. \quad (28)$$

式中: $W_m \in \mathbf{R}^{T \times T}$ 是控制注意力范围的掩码矩阵.多头注意力机制可以并行计算,具有更高的学习和推理效率.他们采用多头注意力机制将所有子空间的结果聚集起来,该过程可表示为

$$A_{\text{Multi-LA}} = \text{concat}(A_{L,1}, A_{L,2}, \dots, A_{L,u}) W_0. \quad (29)$$

式中: W_0 是学习的线性变换矩阵; u 是注意力头的数量.模型利用多头注意力来考虑序列中每个位置在整个序列中的重要性,从而在长期时间依赖条件下提取周期性时间特征.

通过在多步耦合图卷积网络中引入注意力机制, Huang等^[55]提出一种具有时间关注的多步耦合图卷积神经网络(MCGCN)来捕获路网中不同层级的时空关联关系.他们设计了一个多步耦合图卷积模块(MCGC),通过耦合学习关系矩阵来建模交通网络的空间特征,以获取交通路网的不同层次信息.然后,将 MCGC 提取的交通网络信息反馈到多步耦

合图门控递归单元(MCGRU)模块中,实现时空特征的融合.最后,利用多步耦合图注意机制(MCGCAtt)提取历史交通流量中包含的时间信息,从而进一步提升模型的预测性能.为了利用时空注意机制有效捕获交通数据中的动态时空相关性,Guo 等^[56]提出一种基于时空注意力的图卷积网络(ASTGCN)模型用于解决交通流预测问题.该模型中,注意力矩阵的学习过程可以定义为

$$\mathbf{A} = \mathbf{V} \times \sigma((\mathbf{x}_h^{r-1} \mathbf{W}_1) \mathbf{W}_2 (\mathbf{W}_3 \mathbf{x}_h^{r-1})^T + \mathbf{b}_s), \quad (30)$$

$$\mathbf{A}'_{i,j} = \frac{\exp(A_{i,j})}{\sum_{j=1}^N \exp(A_{i,j})}. \quad (31)$$

式中: \mathbf{x}_h^{r-1} 是第 r 层时空块的输入; $\mathbf{V}, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{b}_s$ 是可学习的参数; σ 是 sigmoid() 激活函数.在进行图卷积和时间卷积操作的过程中,注意力矩阵 \mathbf{A}' 将分别动态调整节点间和时刻间的重要程度以捕捉动态时空相关性.

2.2.5 生成对抗网络+图卷积网络

受博弈论中的零和博弈启发,Goodfellow 等^[57]将生成问题视为生成器和判别器这两个网络的对抗和博弈,并提出一种生成对抗网络(Generative Adversarial Network, GAN).生成对抗网络(GAN)由一个生成器和一个判别器组成,生成器通过在潜在空间中随机采样生成网络输入,并通过不断训练使生成器的输出逼近训练集中的真实样本.判别器的输入包括真实样本和生成器的输出,其目的是将生成器的输出从真实样本中区分出来.生成对抗网络擅长无监督学习,且在多种生成模型中能够生成最逼真的结果,是交通流预测研究常用的神经网络之一.例如,代亮等^[58]提出一种基于梯度惩罚的 Wasserstein 生成对抗网络(WGAN-GP)模型用于路网交通流预测.该模型通过构建时间判别器和空间判别器分别从时间和空间两个特征维度对生成数据进行判别,从而引导生成器生成接近真实数据分布的预测数据.Jin 等^[59]提出一种用于短时交通速度预测的 PL-WGAN 模型,其特点是采用一种如图 6 所示的平行学习框架,构建了一个融合时空注意力机制、GCN 和 GRU 的混合模型捕获交叉口和路网的时空特征信息.该方法将 Wasserstein 生成对抗网(WGAN)应用于数据驱动交通建模,并通过在杭州的一个大规模区域网络中的实实验证了其有效性和可扩展性.为了精确预测城市轨道交通网络的短期客流,Zhang 等^[60]提出一种基于深度学习的时空图

生成对抗网络(STG-GAN)预测模型.该模型构建了一个由门控时间卷积网络(TCN)和图卷积网络(GCN)构成的生成网络,以捕获交通路网结构的时空依赖性并生成预测数据.同时,也构建了一个包含空间判别器和时间判别器的判别网络,以增强预测模型建模空间和时间约束的能力.

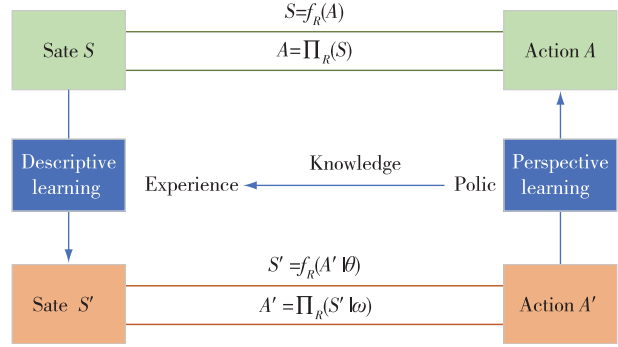


图 6 平行学习框架^[59]

Fig. 6 Parallel learning framework^[59]

2.2.6 Transformer+图卷积网络

2017 年,Google 的机器翻译团队提出一种与循环神经网络(RNN)完全不同的 Transformer 模型,用于自然语言处理(NLP)任务.由于具有很强的建模上下文信息的能力,Transformer 模型在 NLP 任务中得到了广泛的应用,并逐步扩展至其他时序预测相关的任务之中.Transformer 模型的核心是多头注意力机制.在交通流领域中,传统的注意力机制主要用来计算不同特征间的权重,对路网的空间结构图(如拓扑图、距离图)进行动态调整来实现动态图卷积,以及对输入流量数据进行动态加权来得到研究中更为广泛的门控机制.而多头注意力机制可以使用多组注意力同时学习得到不同特征的最优注意力权重.相较于传统的 RNN,Transformer 模型通过考虑不同历史时间步长的不同尺度的依赖关系,提升了模型对远距离时间特征建模的能力.例如,Xu 等^[61]通过联合利用动态的有向空间依赖关系和长程时间依赖关系,提出一种新的时空 Transformer(STTN)模型来提高长期交通流量预测的准确性.他们设计了一个包含空间 Transformer 和时间 Transformer 的时空模块.其中,空间 Transformer 利用空间多头注意力机制建模不同模式(连通性、相似性等)的动态空间依赖性,时间 Transformer 则被用于建模多个时间步长下的长程时间依赖性.仿真实验结果表明,与现有其他模型相比,STTN 模型在处理长程时空依赖关系时

更高效且可扩展。随着研究的深入, 研究人员发现 Transformer 模型对近距离时间相关性不够敏感。因此, 为综合利用 RNN 擅长建模近距离时间相关性和 Transformer 模型擅长建模长程时间相关性的优势, 许多研究人员将 RNN 与 Transformer 模型或是将 TCN 与 Transformer 模型进行融合, 以同时捕获数据中的近距离和远距离时间相关性。例如, Wang 等^[62]提出一种新的具有位置级注意机制 (position-wise attention mechanism) 的图神经网络, 能动态地聚合来自相邻道路的历史交通流信息。为了更好地提取时间特征, 他们同时使用 RNN 和 Transformer 模型来捕获局部和全局的时间相关性。为了进一步提高 Transformer 模型在处理交通流预测任务时的有效性, 一些研究者对 Transformer 的结构进行了改进。譬如, 针对 Transformer 模型适合处理的数据与典型时间序列数据之间存在较大差异, Seq2Seq 和 Transformer 模型的结构易导致累计误差、Transformer 模型不能兼容邻接矩阵等问题, Yan 等^[63]提出一种 Traffic Transformer 模型, 他们设计了一种特殊的编码和特征嵌入来解决 Transformer 与交通流数据结构不兼容的问题, 并将 Transformer 模型的原始编码器和解码器结构改进为全局编码器和全局-局部解码器两个组成部分, 将多个全局编码器和全局-局部解码器块堆叠起来, 形成一个具有层次特征的深度预测网络。针对交通流预测问题, Guo 等^[64]提出一种基于时空注意力机制的图神经网络 (ASTGNN) 模型。为了缓解 Transformer 模型中自注意力机制对局部时间特征不敏感的问题, 该文设计了一种基于 CNN 实现的自注意力模块提升局部时间特征感知能力, 并构建了一个带有自注意力的动态图卷积模块捕获动态空间相关性。刘起东等^[65]在其基于时间感知 Transformer 的交通流预测方法中提出一种可以根据不同时间特征确定不同的注意力计算机制, 从而更加精准地捕获不同时刻下流量的时空相关性。在基于 Transformer 的预测模型中, 大多需要将多个相同的模块进行堆叠, 以捕获深层次的特征。这个特性使得基于 Transformer 的预测模型往往具有网络参数多、网络收敛速度慢的缺点。针对该问题, Bachlechner 等^[66]提出一种 Rezero-Transformer 模型。该文对 Transformer 的归一化和残差连接部分进行了改进, 改进前和改进后分别由式 (32) 和式 (33) 表示:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + F(\text{norm}(\mathbf{x}_i)), \quad (32)$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \cdot F(\mathbf{x}_i). \quad (33)$$

式中: \mathbf{x}_i 是模型输入; α_i 表示一个学到的残差参数; $\text{norm}()$ 表示归一化操作; $F()$ 表示某一个层的变换。

2.2.7 模式分解+图卷积网络

经验模态分解 (Empirical Mode Decomposition, EMD) 是 Huang 等^[67]提出的一种信号处理方法, 它可以自适应地将原始时间序列分解为几种不同的内在模态函数 (IMFs) 和残差。得到的 IMFs 和残差序列具有特定的物理意义, 代表原始时间序列的不同时间尺度特征。为了降低预测数据的维度以便更好地对时空相关性进行建模, 主成分分析 (Principal Component Analysis, PCA) 方法通过对数据进行归一化等操作将 m 维原始数据降为 k 维, 然后对归一化得到的数据计算协方差矩阵、特征值和特征向量, 将特征值降序排列, 并选择最大的 k 个特征值及其对应的特征向量。最后, 使用这些特征向量构造新的特征空间, 并在新的特征空间上完成流量预测。借鉴 EMD 和 PCA 等方法通过对时间序列进行分解得到不同时间尺度特征的思路, Chen 等^[68]提出一种基于 EMD 的多元时间图卷积神经网络预测模型。该文先采用 EMD 方法对原始时间序列进行分解, 将得到不同尺度下的时间特征作为节点的初始特征生成图模型。然后使用多头注意机制来学习节点之间的隐藏依赖关系, 并使用 GCN 网络提取节点空间特征。最后使用一个 TCN 网络为节点嵌入建立时间关系, 以执行多变量时间序列预测。Li 等^[69]设计了一个基于主成分分析 (PCA)、图卷积网络 (GCN) 和长短期记忆网络 (LSTM) 的主时空图卷积网络 (PST-GCN) 模型。具体来说, PCA 首先对原始输入数据进行降维, 然后构建 GCN 网络学习城市路网的拓扑结构, 以获得路网中不同路段和节点间的空间相关性。另外, 他们在 PST-GCN 模型中构建了一个 LSTM 网络用来捕获路网中的时间相关性。

3 公开数据集

自行采集宽范围跨时空交通数据集的难度大、成本高。本节具体介绍了 5 个交通流预测领域常用于模型性能对比分析的公开数据集。

3.1 PeMS

PeMS 是一个大型的交通数据库 (<http://pems.dot.ca.gov>), 数据由美国加州运输公司及其合作机构提供。该数据集主要包含加州高速公路上各类车流量相关数据。这些数据可以让决策者对高速公路通

行效率进行统一、全面地评估,基于对高速公路网络当前状态的了解做出运营决策,分析拥堵瓶颈以确定潜在的补救措施,并做出更好的整体决策.该数据集由超过 44 681 个检测器每 30 s 报告一次数据,30 s 的数据完成编译后,就会被聚合成 5 min 的增量.该数据集包含有众多交通效率评价指标,如流量、速度、延迟、车辆行驶里程(VMT)、车辆行驶小时(VHT)、行驶时间和年平均日交通流量(AADT).值得注意的是,PeMS 有多个衍生子数据集,譬如 PeMS-03、PeMS-BAY 等.

3.2 METR-LA

METR-LA 数据集记录了美国洛杉矶高速公路上 207 个监测点的位置信息(包括检测器间的距离和拓扑关系),以及各监测点在 2012 年 3 月 1 日—6 月 30 日共计 4 个月时间内记录的 34 272 条交通速度数据,且数据的采集时间间隔为 5 min.该数据集可通过其官网(<https://www.metro.net>)下载.

3.3 滴滴盖亚开发数据集

滴滴盖亚开发数据集(<https://gaia.didichuxing.com/>)数据来自于滴滴“盖亚数据开放计划”,统计了网约车在成都、西安、海口等城市行驶中的所有传感器数据.该数据集包括的数据有:车辆平均速度、OD 信息、轨迹信息、驾驶场景、POI 检索数据等.

3.4 Urban 数据集

Urban 数据集(<http://topis.seoul.go.kr/>)记录了韩国首尔江南区(Urban1)和麻浦区(Urban2)两个区域的真实车流数据.Urban1 和 Urban2 是首尔交通流量最大的两个地区,且均具有高度复杂的城市交通网络.数据集的采样周期为 2018 年 4 月 1 日至 2018 年 4 月 30 日.数据集主要是基于 GPS 采集的 7 万多辆出租车的轨迹数据,采样时间间隔为 5 min.采集的交通流数据经过数据预处理得到的是各链路的平均速度.

3.5 NYC Taxi

NYC Taxi 数据集包括了 2009—2020 年纽约黄色和绿色出租车行程记录数据,主要记录了接送日期/时间、接送地点、行程距离、分项票价、费率类型、付款类型和司机报告的乘客数量等信息.该数据集是一个学术界常用的出租车数据集.下载网址:<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>.

4 仿真案例

应用基于图卷积的短时交通流量预测模型解决交通流预测问题的主要步骤包括:数据预处理、模型训练及验证、模型测试等.本节简述了上述 4 个步骤的常用方法或具体过程,并以基于 LSTM 网络的交通流预测方法和 DCRNN 网络的交通流预测方法为参考基准,通过仿真对比分析展示了 4 种基于图卷积神经网络的交通流预测方法在 Urban1 和 Urban2 两个真实交通流数据集上的预测误差.值得注意的是,参与对比的基于 MGCN 的交通流预测方法是笔者设计的.

4.1 数据预处理

深度学习模型是通过数据进行驱动的,数据集的好坏直接影响着模型的预测性能.从路网采集的原始数据中通常存在数据缺失和异常等问题,数据的缺失或异常将直接影响模型的预测精度.因此,在进行模型训练前,往往需要对所采集的原始数据进行预处理,以提高数据的完整性和真实性.对于缺失的数据往往使用数据补全的方法进行处理,对于异常数据则常用数据剔除或高峰去噪等方法进行处理.一般路网采集的原始数据的分布较为分散,并不满足正态分布.因此,需对数据进行标准化和归一化处理,方便模型对数据进行高效学习.在对数据进行标准化和归一化后,将不同的数据集按 6:2:2 或按 7:1:2 的比例划分为训练集、验证集和测试集.

4.2 模型训练、验证及测试

首先,利用训练集来优化模型参数.更具体地说,将训练集输入到构建的预测模型中,利用模型输出与真实值之间的误差定义模型损失,并通过梯度下降法对模型参数进行优化.验证集则主要用于对已经训练好的模型进行评价,筛选出在验证集上泛化性更好的模型参数.在交通流量预测研究领域,评价模型泛化能力的常用指标包括:平均绝对误差(Mean Absolute Error, MAE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)和均方根误差(Root Mean Squared Error, RMSE).在验证集上,利用上述 3 个评价指标筛选出泛化性最佳的模型参数,并在测试集上进行预测.

4.3 结果分析与讨论

表 1 和表 2 分别展示了基于 LSTM、DCRNN,以及 4 种图卷积神经网络的交通流预测方法在 Urban1 和 Urban2 两个真实交通流数据集上的应用,预测步

表 1 不同预测方法在 Urban1 数据集上的预测误差

Table 1 Prediction error of different prediction methods on dataset Urban1

预测方法	MAE			MAPE/%			RMSE		
	15 min	30 min	60 min	15 min	30 min	60 min	15 min	30 min	60 min
LSTM ^[70]	3.28	3.47	3.83	13.52	13.74	15.38	4.82	5.11	5.61
DCRNN ^[50]	2.92	3.37	4.15	12.02	13.98	17.24	4.73	5.32	6.11
STGCN ^[46]	2.78	3.17	3.80	11.60	13.50	16.13	4.21	4.74	5.60
Graph WaveNet ^[33]	2.77	3.18	3.69	11.83	13.94	16.61	4.28	4.77	5.39
STTN ^[61]	2.89	3.21	3.57	12.41	14.07	16.01	4.48	4.88	5.32
MGCN	2.93	3.25	4.37	11.20	12.77	16.25	4.30	4.85	6.27

表 2 不同预测方法在 Urban2 数据集上的预测误差

Table 2 Prediction error of different prediction methods on dataset Urban2

预测方法	MAE			MAPE/%			RMSE		
	15 min	30 min	60 min	15 min	30 min	60 min	15 min	30 min	60 min
LSTM ^[70]	3.08	3.45	4.22	12.04	13.47	15.75	4.48	4.87	5.70
DCRNN ^[50]	2.81	3.15	3.62	12.04	13.64	15.75	4.31	4.69	5.24
STGCN ^[46]	2.80	3.19	3.67	12.13	13.90	16.43	4.19	4.64	5.24
Graph WaveNet ^[33]	2.72	3.08	3.52	11.86	13.61	15.89	4.13	4.53	5.04
STTN ^[61]	2.73	2.96	3.21	12.10	13.16	14.38	4.06	4.31	4.59
MGCN	2.75	3.14	3.73	11.32	12.65	14.58	4.09	4.54	5.11

长分别为 15、30 和 60 min 时的预测误差。不难发现, 与两种基准方法相比, 采用图卷积网络建模交通流量中空间关联关系的方法具有更低的平均预测误差。另外, 图 7—8 进一步展示了预测步长为 5 min 时, 不同预测方法分别在 1、2、3、6、9、12 不同预测时长下的预测误差。可以看出, 在数据集 Urban1 和 Urban2 上, 参与对比分析的交通流预测方法均能较好地拟合真实的交通流量。基于 LSTM 的交通流预测方法, 仅将交通流数据当作简单的时间序列进行处理, 忽略了路网中不同节点间的复杂空间相关性。由于不同节点间的空间相关性对于长期预测是至关重要的, 因此, 虽然在预测步长较小时 LSTM 与其他预测方法之间的预测性能差距较小, 但随着预测步长的增加, 它们之间的差距逐步扩大。在上述几种基于时空图卷积建模的典型交通流预测方法中, 从整体预测性能指标上看, 基于 STGCN 的方法稍逊于基于 Graph WaveNet 的方法。这可能是由于 STGCN 方法采用预定义方式对空间相关性进行描述, 但在其预定义的图结构中包含了一些不合理的空间信息描述, 导致所定义的图未能较好地契合道路实际空间结构。Graph WaveNet 采用以数据驱动学习的方式来学习与真实路网更加契合的图结构, 解决了预定义图存在的局限性。值得注意的是, 基于多图卷积网络 MGCN 的交通流预测方法, 也采用预定义的图结构

且可能会引入一些不合理的空间相关性描述, 但由于使用了多个图卷积网络 GCN 从不同的维度提取路网交通流中的空间特征, 并将所提取的多个空间特征进行了融合, 这在一定程度上缓解了预定义的图结构中可能存在不合理描述的问题。因此, 从 3 个预测误差评价指标上看, 基于多图卷积网络 MGCN 整体上也优于基于时空图卷积网络 STGCN。由于路网的时空特征是复杂多变的, 采用预定义或者数据驱动学习方式生成的静态图结构难以对路网中不同节点间的空间相关性进行有效建模。譬如, 静态图无法有效刻画路网中随时间实时动态变化的空间特征。因此, 基于静态图的交通流预测模型存在一定的局限性。为了解决该问题, 基于多头注意力机制的时空 Transformer (STTN) 模型利用时间多头注意力和空间多头注意力对路网的动态时空相关性进行建模。因此, 在 6 种基于典型深度神经网络的交通流预测方法中, 基于 STTN 模型的交通流预测方法具有最优越的预测性能。

5 未来发展方向

近年来, 基于图卷积网络的交通流预测方法成为了交通流预测领域中最热门的研究方向之一。虽然图卷积网络广泛应用于交通流预测领域并取得了一些有吸引力的研究结果, 但基于图卷积网络的交通流预测方法研究中还存在一些开放性的问题有待

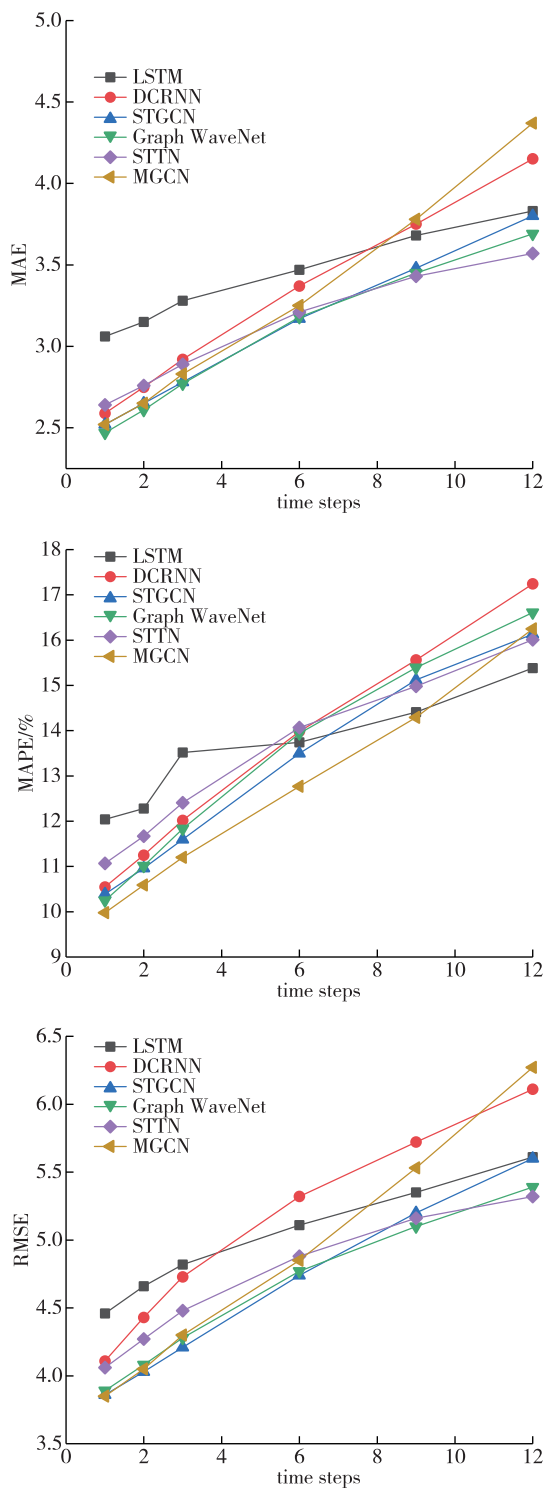


图7 Urban1数据集上典型模型结果对比

Fig. 7 Comparison of typical model results on dataset Urban1

进一步解决.因此,本节对图卷积网络及其在交通流领域未来的研究和发展方向进行了开放性的讨论,以期为该领域的研究人员提供一些可供参考和借鉴的研究视角.

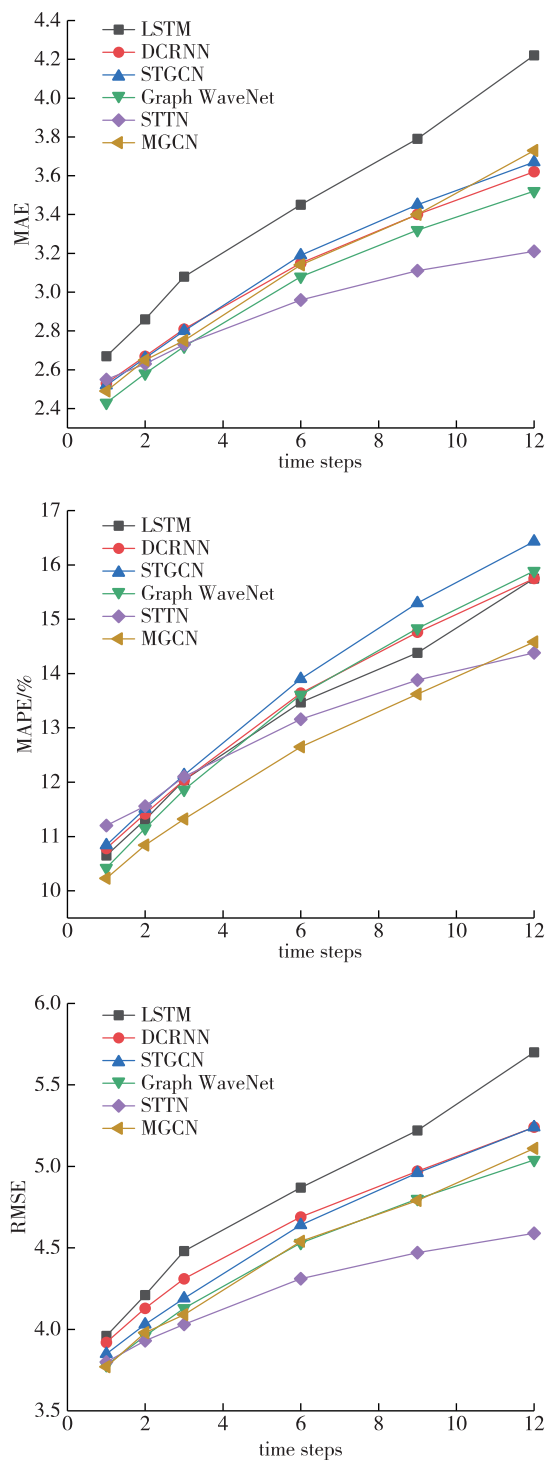


图8 Urban2数据集上典型模型结果对比

Fig. 8 Comparison of typical model results on dataset Urban2

5.1 提取图中更深层次的隐含信息

在一些基于图卷积网络 GCN 的交通流预测模型中,研究人员仅利用了不同节点之间的距离关系来构建邻接矩阵.显然,这种构建邻接矩阵的方式不能够充分挖掘交通图中隐含的信息.与仅使用拓扑

图、距离图的图卷积网络相比,多图和超图都是更加有效的新型图卷积结构,可以对交通图中更深层次的隐含信息进行更加精细化的建模.例如,Lv等^[27]在他们提出的基于时间多图卷积网络 T-MGCN 的交通流预测模型中分析了道路之间存在的两种语义相关性(即历史交通模式相关性和局部区域功能相似性),并使用多个图对异构的空间和语义相关性进行编码,然后,使用一个多图卷积网络来建模和融合不同图所提取的空间特征信息并用于交通流预测.该研究表明,传统的图表示存在一定的局限性,无法表示多重关联,而超图的引入则有利于表示和学习拓扑结构中的更深层次的相互关联关系.考虑到路网中连接节点的边,以及节点之间存在着复杂的相关性且相互作用,Chen等^[43]提出了一种基于多范围注意力双分量图卷积网络的交通流预测方法.该方法构建了一个节点图和一个边缘图分别挖掘节点与节点之间的空间关联关系,以及边与边之间的空间关联关系,然后,使用一个带有多范围注意力机制的双分量图卷积网络显式地建模节点和边的相互作用.这些研究都表明了交通图中具有很多深层的隐含信息,如何更有效地利用这些信息是提高基于图卷积网络 GCN 交通流预测方法预测性能的关键所在.

5.2 多源数据融合

事实上,城市中不同的交通工具是同时运行的,它们在共同的时空场景下相互影响、相互补充,动态地构成了完整的交通系统.因此,来自不同交通工具的流量数据表面上是异构的,但内部具有隐含的相关性.此外,天气、交通事故,以及很多外部因素也会对交通模式产生显著影响.通过融合多源数据并进行协同分析,可以找到这些数据与交通流量之间的相关性,并借此来提高模型的预测性能.例如,Chen等^[71]利用多源城市数据(如 POI、路网、事件、天气等)构建多源属性图,同时考虑静态因素和动态因素(如空间距离、语义距离、道路特征、道路状况和全球背景)来预测短时交通流.由于典型的单数据驱动模型并不普遍适用于异构流量数据,Wang等^[72]提出一种用于多源流量预测问题的多任务超图卷积网络(MT-HGCN).该框架由一个主任务和一个相关任务组成,且两个任务都基于超图卷积网络(HGCN)提取空间关联关系.然后,构建特征融合模块拼接任务,该模块可以对相关性进行建模并共享潜在特征,以提高主任务的性能.Zhang等^[73]在进行多区域上

的客运需求短期预测时,提出一种基于端到端多任务学习时间卷积神经网络(MTL-TCNN)的交通流预测方法.该方法设计了一个 ST-DTW 算法来量化两个预测任务之间的时空相关性,提高了出租车需求预测的精度.因此,发掘更多可靠的数据来源,并建立这些数据与交通流量数据之间的相关性,也是提高模型预测性能的主要方法之一.

5.3 小样本预测

在处理大规模数据时,深度学习方法可以展现出非常优秀的性能,但在实际应用时,可能会存在数据采集不足、不同城市的数据质量参差不齐等问题.另外,交通预测的辅助特征如天气、异常状况等采集的频率远低于交通数据的采集频率,这可能会使我们无法获得足够的训练数据.而在训练数据较少时,基于深度学习方法的预测模型往往会产生明显的性能下降,模型的泛化能力也难以满足要求.因此,针对小样本学习进行优化,是提高预测模型实际性能的重要途径.彭云聪等^[74]将小样本学习分为表征学习、数据扩充、学习策略 3 种,为小样本交通流预测提供了多种解决思路.通过表征学习,可以将缺失或者稀疏的交通数据变换到特征域学习.由于特征域具有更低的特征维度以及相关的语义信息,大大降低了预测模型的学习难度.迁移学习也是实现小样本学习的重要方法之一,其基本思想是相似或者相近的任务之间具有共通性,其本质是基于源域数据和目标域数据之间共同的特征、关系以及模型参数共享等方式进行迁移学习,并将模型用于目标类数据分类.就交通流预测而言,迁移学习旨在将数据丰富的城市或者区域迁移到数据稀疏的城市或者区域.例如,Wang等^[75]通过对有历史数据的交叉口进行迁移学习完成对没有历史数据的交叉口的交通流量预测.但遗憾的是,目前的迁移学习模型均存在泛化能力不足的问题.因此,通过加强模型的小样本学习能力来提高预测模型的实际性能,也是未来重要的研究方向之一.

5.4 异常状况交通流预测

在目前有关交通流预测的研究中,大多是针对日常情形下的交通流量建立预测模型,这些研究大多致力于通过更有效地提取时空特征来优化模型,对其他影响因素考虑得相对较少.但事实上,极端天气、交通事故等异常情况往往会对交通模式产生重大的影响.当路网中出现此类异常情况时,通用的流量预测模型将难以提供准确的预测结果.目前,已经

有研究人员致力于解决此类问题.例如,为了分析交通事故等异常事件对路网交通状况的影响, Liu 等^[54]采用位置编码的方式来检测复杂交通情况下异常事件的发生,以探索异常事件对区域交通拥堵的影响. An 等^[76]提出一种基于模糊卷积神经网络(F-CNN)的交通流预测方法.该方法采用模糊方法来表示交通事故特征,先利用模糊推理机制从真实的交通流数据中生成交通事故信息并融入交通流特征信息中,再利用 CNN 提取数据中的特征信息,处理数据中的不确定性问题.但到目前为止,此类模型的泛化能力仍然相对较弱,其性能仍有很大的提升空间.因此,异常状况下的交通预测也是有待进一步开展深入研究的方向之一.

5.5 时空同步建模

在大多数交通流预测模型中,时间与空间特征都是分别由不同的神经网络模块提取,再将不同神经网络模块的输出进行融合,以达到同时提取时间和空间特征的目的.然而,堆叠过多的神经网络可能会导致网络结构复杂、模型参数过多等问题,并降低了模型的可解释性.对时空特征进行同步建模,在提升模型预测准确度的同时,还可以简化网络结构、提升计算效率、增强模型可解释性.有关时空同步建模的相关研究目前还相对较少.例如, Song 等^[77]在其提出的时空同步图卷积网络(STSGCN)中,设计了一种时空同步建模机制,使得该网络能够同步捕获复杂的时间和空间相关性. Li 等^[78]在其提出的时空图融合卷积神经网络 STFGNN 模型中,将空间图 A_{SG} 、时间图 A_{TC} 、时间连接图 A_{TC} 组合为一个融合图 A_{STFG} . 融合图 A_{STFG} 同时包含相邻时间步长的异质时空相关性.同时,该文还设计了一种新的时空融合图模块来同步捕获时空依赖性用于交通流预测.结果表明,时空同步建模使得预测模型的性能获得了进一步提高.因此,时空同步建模也可以作为未来的研究内容之一.

5.6 长时交通流预测

长时交通流预测是指预测周期为半小时至数小时、一天,甚至更长的交通流预测.与短时交通流预测相比,长时交通流预测可以帮助管理者更好地做出决策、采取措施、统筹安排,对提高路网交通管理和服务水平有着重要的作用.在长时交通流预测领域, Peng 等^[79]提出一种基于强化学习动态图卷积网络的长时交通流量预测模型.他们利用路网的历史流量数据,设计了交通流量传输图和交通流量概率

图来表示不同节点之间的流量转移关系.针对实际应用中可能存在的数据缺陷,他们还将动态图生成模块应用于交通流量的长时预测任务来减少交通流量的预测误差.然而,相比于短时交通流预测,长时交通流预测往往需要处理更复杂的时空依赖关系,且面临着更多不确定性因素的干扰.另外,多步预测模型在进行长时预测时,会出现严重的误差累积问题,导致预测准确性随着预测时长的增加而快速下降.因此,如何继续在现有研究工作基础上提高长时交通流预测模型的准确性,是未来的重要研究内容之一.

5.7 大规模图网络

随着城市化进程的不断深入,城市道路交通网络的规模愈发庞大.由于卷积是基于邻居节点进行的,对于某个节点进行卷积操作时用到的邻域节点的个数也会随着路网规模的增加呈指数级增长.因此,在建模大规模图网络时,目前的图卷积网络模型也面临着严峻的挑战.针对该问题, Hamilton 等^[80]在其提出的一个节点特征信息通用归纳框架 GraphSAGE 模型中使用分批量的训练方法,结合邻居节点采样的方法,将每次计算所需的节点数目控制在一定范围之内.虽然 GraphSAGE 不能解决图卷积网络 GCN 难以有效解决大规模路网交通流预测的问题,但也为图卷积网络 GCN 在大规模数据集上应用提供了一种可参考的思路.因此,如何设计更加高效的模型结构以提高在大规模路网上利用图卷积建模复杂空间相互关联关系的效率,是未来交通流预测领域值得研究的方向之一.

6 总结

交通流预测不仅是保障智能交通系统高效运行的基础,也是提高路网利用率、缓解交通拥堵的重要方法.而图卷积网络则是目前交通流预测模型中使用最广泛的神经网络之一.本文对近几年基于图卷积网络的交通流预测模型进行了全面的回顾与总结.简要回顾了交通流预测的发展历史,介绍了图卷积网络诞生的背景、优势及基本原理.另外,收集了近年来交通流预测领域中基于图卷积网络的热点模型,包括时空图卷积网络和图卷积 Transformer 等,并对其进行了分类和详细讨论.这些模型在交通流预测方面取得了一系列有价值的研究成果,为提升路网通行效率、解决交通拥堵问题提供了一些新的思路和方法.此外,还收集了交通流领域常用的一些公

开数据集,并以其中一个真实数据集为例验证了基于图卷积网络交通流预测模型的有效性以及相比于基准模型的优势.最后,本文对基于图卷积网络的交通流预测方法未来可能的研究方向和挑战进行了开放式的讨论.图卷积网络在交通流预测中的应用前景广阔,但仍存在一些挑战和难点,如模型的可解释性和对大规模数据的处理能力等.未来的研究可以聚焦于这些问题,进一步探索更加高效和精确的交通流预测方法.

参考文献

References

- [1] Wang Y B, Yu X H, Zhang S Y, et al. Freeway traffic control in presence of capacity drop [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22 (3) : 1497-1516
- [2] 陆化普, 孙智源, 屈闻聪. 大数据及其在城市智能交通系统中的应用综述 [J]. 交通运输系统工程与信息, 2015, 15(5): 45-52
LU Huapu, SUN Zhiyuan, QU Wencong. Big data and its applications in urban intelligent transportation system [J]. Journal of Transportation Systems Engineering and Information Technology, 2015, 15(5): 45-52
- [3] Pan B, Demiryurek U, Shahabi C. Utilizing real-world transportation data for accurate traffic prediction [C] // 2012 IEEE 12th International Conference on Data Mining. December 10-13, 2012, Brussels, Belgium. IEEE, 2013: 595-604
- [4] Sun Y J, Zhang G H, Yin H H. Passenger flow prediction of subway transfer stations based on nonparametric regression model [J]. Discrete Dynamics in Nature and Society, 2014, 2014: 1-8
- [5] Williams B M, Durvasula P K, Brown D E. Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models [J]. Transportation Research Record: Journal of the Transportation Research Board, 1998, 1644: 132-141
- [6] Lee S, Fambro D B. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting [J]. Transportation Research Record: Journal of the Transportation Research Board, 1999, 1678: 179-188
- [7] Williams B M. Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling [J]. Transportation Research Record: Journal of the Transportation Research Board, 2001, 1776: 194-200
- [8] Zivot E, Wang J. Vector autoregressive models for multivariate time series [M] // Zivot E, Wang J H. Modeling Financial Time Series with S-PLUS. 2nd ed. New York, NY: Springer New York, 2007: 385-429
- [9] Van Lint J W C, Van Hinsbergen C. Short-term traffic and travel time prediction models [J]. Artificial Intelligence Applications to Critical Transportation Issues, 2012, 22 (1) : 22-41
- [10] Jeong Y S, Byon Y J, Castro-Neto M M, et al. Supervised weighting-online learning algorithm for short-term traffic flow prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(4): 1700-1707
- [11] Ma X L, Dai Z A, He Z B, et al. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction [J]. Sensors, 2017, 17(4): 818
- [12] Fu R, Zhang Z, Li L. Using LSTM and GRU neural network methods for traffic flow prediction [C] // 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). November 11-13, 2016, Wuhan, China. IEEE, 2017: 324-328
- [13] Yao H X, Wu F, Ke J T, et al. Deep multi-view spatial-temporal network for taxi demand prediction [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 2588-2595
- [14] Liu Y P, Zheng H F, Feng X X, et al. Short-term traffic flow prediction with Conv-LSTM [C] // 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP). October 11-13, 2017, Nanjing, China. IEEE, 2017: 1-6
- [15] Nagy A M, Simon V. Survey on traffic prediction in smart cities [J]. Pervasive and Mobile Computing, 2018, 50: 148-163
- [16] Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs [C] // International Conference on Machine Learning. PMLR, 2016: 2014-2023
- [17] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs [J]. arXiv e-print, 2014, arXiv: 1312. 6203
- [18] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry [C] // Proceedings of the 34th International Conference on Machine Learning. August 6-11, 2017, Sydney, NSW, Australia. New York: ACM, 2017: 1263-1272
- [19] Monti F, Boscaini D, Masci J, et al. Geometric deep learning on graphs and manifolds using mixture model CNNs [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 5425-5434
- [20] Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data [J]. arXiv e-print, 2015, arXiv: 1506. 05163
- [21] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [J]. arXiv e-print, 2016, arXiv: 1609. 02907
- [22] Liu Q H, Zhang R, Wang Y J, et al. Daily prediction of the Arctic sea ice concentration using reanalysis data based on a convolutional LSTM network [J]. Journal of Marine Science and Engineering, 2021, 9(3): 330
- [23] Diao Z L, Wang X, Zhang D F, et al. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 890-897
- [24] Zheng Q, Zhang Y Y. DSTAGCN: dynamic spatial-temporal adjacent graph convolutional network for traffic forecasting [J]. IEEE Transactions on Big Data, 2023, 9

- (1):241-253
- [25] Hu J, Chen L Y. Multi-attention based spatial-temporal graph convolution networks for traffic flow forecasting [C]//2021 International Joint Conference on Neural Networks (IJCNN). July 18–22, 2021, Shenzhen, China. IEEE, 2021: 1-7
- [26] Ma Y T, Lan S Y, Wang W W, et al. Modeling of spatial-temporal dependency in traffic flow data for traffic forecasting [J]. SSRN Electronic Journal, 2022. DOI: 10.2139/ssrn.4142192
- [27] Lv M Q, Hong Z X, Chen L, et al. Temporal multi-graph convolutional network for traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3337-3348
- [28] Yang Y D, Feng Z L, Song M L, et al. Factorizable graph convolutional networks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. December 6–12, 2020, Vancouver, BC, Canada. New York: ACM, 2020: 20286-20296
- [29] Yin D, Jiang R H, Deng J W, et al. MTMGNN: multi-time multi-graph neural network for metro passenger flow prediction[J]. GeoInformatica, 2023, 27(1): 77-105
- [30] Lee K, Rhee W. DDP-GCN: multi-graph convolutional network for spatiotemporal traffic forecasting[J]. Transportation Research Part C: Emerging Technologies, 2022, 134: 103466
- [31] Geng X, Li Y G, Wang L Y, et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 3656-3663
- [32] Khaled A, Elsir A M T, Shen Y M. TFGAN: traffic forecasting using generative adversarial network with multi-graph convolutional network [J]. Knowledge-Based Systems, 2022, 249: 108990
- [33] Wu Z H, Pan S R, Long G D, et al. Graph WaveNet for deep spatial-temporal graph modeling [J]. arXiv e-print, 2019, arXiv: 1906. 00121
- [34] Bai L, Yao L N, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17804-17815
- [35] Shin Y, Yoon Y. PGCN: progressive graph convolutional networks for spatial-temporal traffic forecasting [J]. arXiv e-print, 2022, arXiv: 2202. 08982
- [36] Wang Y H, Fang S, Zhang C X, et al. TVGCN: time-variant graph convolutional network for traffic forecasting [J]. Neurocomputing, 2022, 471: 118-129
- [37] 谷振宇, 陈聪, 郑家佳, 等. 考虑时空相似性的动态图卷积神经网络交通流预测[J]. 控制与决策, 2023, 38(12): 3399-3408
- GU Zhenyu, CHEN Cong, ZHENG Jiajia, et al. Traffic flow prediction based on STG-CRNN[J]. Control and Decision, 2023, 38(12): 3399-3408
- [38] 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
- [39] Ali A, Zhu Y M, Zakarya M. Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction [J]. Neural Networks, 2022, 145: 233-247
- [40] 张阳, 胡月, 陈德旺, 等. 基于 GCN-BiLSTM 的短时交通流预测模型[J]. 武汉理工大学学报(交通科学与工程版), 2023, 47(5): 802-806
- ZHANG Yang, HU Yue, CHEN Dewang, et al. Short-term traffic flow prediction model based on GCN-BiLSTM [J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2023, 47(5): 802-806
- [41] Zhang Y, Cheng T, Ren Y B, et al. A novel residual graph convolution deep learning model for short-term network-based traffic forecasting[J]. International Journal of Geographical Information Science, 2020, 34(5): 969-995
- [42] 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
- [43] Chen W Q, Chen L, Xie Y, et al. Multi-range attentive bi-component graph convolutional network for traffic forecasting [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 3529-3536
- [44] Zhao W, Zhang S Q, Zhou B, et al. Multi-spatio-temporal fusion graph recurrent network for traffic forecasting [J]. arXiv e-print, 2022, arXiv: 2205. 01480
- [45] 陈柘, 刘嘉华, 赵斌, 等. 基于 GCN 和 TCN 的多因素城市路网出租车需求预测[J]. 控制与决策, 2023, 38(4): 1031-1038
- CHEN Zhe, LIU Jiahua, ZHAO Bin, et al. Multi-factor taxi demand forecasting for urban road network based on GCN and TCN [J]. Control and Decision, 2023, 38(4): 1031-1038
- [46] Yu B, Yin H T, Zhu Z X. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[J]. arXiv e-print, 2017, arXiv: 1709. 04875
- [47] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27–30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778
- [48] Dai A M, Le Q V. Semi-supervised sequence learning [J]. Advances in Neural Information Processing Systems, 2015: 3079-3087
- [49] Zhu H, Luo Y, Liu Q, et al. Multistep flow prediction on car-sharing systems: a multi-graph convolutional neural network with attention mechanism [J]. International Journal of Software Engineering and Knowledge Engineering, 2019, 29(11/12): 1727-1740
- [50] Li Y G, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [J]. arXiv e-print, 2017, arXiv: 1707. 01926
- [51] Guo J L, Song C Y, Zhang H, et al. Multi-step traffic speed prediction model with auxiliary features on urban road networks and its understanding [J]. IET Intelligent Transport Systems, 2020, 14(14): 1997-2009
- [52] Kipf T N, Welling M. Variational graph auto-encoders

- [J].arXiv e-print, 2016, arXiv: 1611. 07308
- [53] 葛宇然, 付强. 基于时空联合学习的城市交通流短时预测模型[J]. 计算机工程, 2023, 49(1): 270-278
GE Yuran, FU Qiang. Short-time prediction model for urban traffic flow based on joint spatio-temporal learning [J]. Computer Engineering, 2023, 49(1): 270-278
- [54] Liu Z C, Zhang R, Wang C, et al. Spatial-temporal consequence learning with accident encoding for traffic flow prediction [J]. IEEE Transactions on Network Science and Engineering, 2022, 9(3): 1765-1775
- [55] Huang X H, Ye Y M, Yang X F, et al. Multistep coupled graph convolution with temporal-attention for traffic flow prediction [J]. IEEE Access, 2022, 10: 48179-48192
- [56] 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(1): 922-929
- [57] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63(11): 139-144
- [58] 代亮, 梅洋, 钱超, 等. 基于生成对抗网络的大规模路网交通流预测算法 [J]. 控制与决策, 2021, 36(12): 2937-2945
DAI Liang, MEI Yang, QIAN Chao, et al. Traffic flow forecasting algorithm for large-scale road network based on GAN [J]. Control and Decision, 2021, 36(12): 2937-2945
- [59] Jin J C, Rong D D, Zhang T, et al. A GAN-based short-term link traffic prediction approach for urban road networks under a parallel learning framework [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 16185-16196
- [60] Zhang J L, Li H A, Zhang S X, et al. A spatiotemporal graph generative adversarial networks for short-term passenger flow prediction in urban rail transit systems [J]. International Journal of General Systems, 2023, 52(6): 694-721
- [61] Xu M X, Dai W R, Liu C M, et al. Spatial-temporal transformer networks for traffic flow forecasting [J]. arXiv e-print, 2020, arXiv: 2001. 02908
- [62] Wang X Y, Ma Y, Wang Y Q, et al. Traffic flow prediction via spatial temporal graph neural network [C]//Proceedings of the Web Conference 2020. April 20 -24, 2020, Taipei, China. New York: ACM, 2020: 1082-1092
- [63] Yan H Y, Ma X L, Pu Z Y. Learning dynamic and hierarchical traffic spatiotemporal features with transformer [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 22386-22399
- [64] Guo S N, Lin Y F, Wan H Y, et al. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(11): 5415-5428
- [65] 刘起东, 刘超越, 邱紫鑫, 等. 基于时间感知Transformer的交通流预测方法 [J]. 计算机科学, 2023, 50(11): 88-96
LIU Qidong, LIU Chaoyue, QIU Zixin, et al. Time-aware transformer for traffic flow forecasting [J]. Computer Science, 2023, 50(11): 88-96
- [66] Bachlechner T, Majumder B P, Mao H, et al. ReZero is all you need: fast convergence at large depth [C]//Uncertainty in Artificial Intelligence. PMLR, 2021: 1352-1361
- [67] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995
- [68] Chen Y W, Ding F Q, Zhai L B. Multi-scale temporal features extraction based graph convolutional network with attention for multivariate time series prediction [J]. Expert Systems with Applications, 2022, 200: 117011
- [69] Li N, Jia S C, Li Q. Traffic message channel prediction based on graph convolutional network [J]. IEEE Access, 2021, 9: 135423-135431
- [70] Zhao Z, Chen W H, Wu X M, et al. LSTM network: a deep learning approach for short-term traffic forecast [J]. IET Intelligent Transport Systems, 2017, 11(2): 68-75
- [71] Chen L, Shao W, Lv M Q, et al. AARGNN: an attentive attributed recurrent graph neural network for traffic flow prediction considering multiple dynamic factors [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 17201-17211
- [72] Wang J C, Zhang Y, Wang L X, et al. Multitask hypergraph convolutional networks: a heterogeneous traffic prediction framework [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 18557-18567
- [73] Zhang K P, Liu Z J, Zheng L. Short-term prediction of passenger demand in multi-zone level: temporal convolutional neural network with multi-task learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(4): 1480-1490
- [74] 彭云聪, 秦小林, 张力戈, 等. 面向图像分类的小样本学习算法综述 [J]. 计算机科学, 2022, 49(5): 1-9
PENG Yuncong, QIN Xiaolin, ZHANG Lige, et al. Survey on few-shot learning algorithms for image classification [J]. Computer Science, 2022, 49(5): 1-9
- [75] Wang H Q, Zhang R Q, Cheng X, et al. Hierarchical traffic flow prediction based on spatial-temporal graph convolutional network [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 16137-16147
- [76] An J Y, Fu L, Hu M, et al. A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information [J]. IEEE Access, 2019, 7: 20708-20722
- [77] 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 [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 914-921
- [78] Li M Z, Zhu Z X. Spatial-temporal fusion graph neural networks for traffic flow forecasting [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(5): 4189-4196
- [79] Peng H, Du B W, Liu M S, et al. Dynamic graph convolu-

tional network for long-term traffic flow prediction with reinforcement learning [J]. Information Sciences, 2021, 578:401-416

[80] Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs [C] // Proceedings

of the 31st International Conference on Neural Information Processing Systems. December 4-9, 2017, Long Beach, California, USA. New York: ACM, 2017: 1025-1035

A survey of traffic flow prediction based on graph convolutional networks

YE Baolin^{1,2} DAI Bena^{1,2} ZHANG Mingjian^{1,2} GAO Huimin² WU Weimin³

1 School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China

2 College of Information Science and Engineering, Jiaying University, Jiaying 314001, China

3 Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, China

Abstract In recent years, deep learning has been a hot research topic in traffic flow prediction. Graph convolutional networks outperform traditional convolutional neural networks in spatial feature modeling, in view of their powerful capabilities in processing non-Euclidean data such as topological map, distance map and flow similarity map. Therefore, graph convolutional network and its variants have become a research hotspot in traffic flow prediction, and many attractive research results have been obtained. This article classifies and summarizes traffic flow prediction models based on graph convolutional networks in recent years. First, the graph convolution is elaborated by combining the definitions of spatial convolution and spectral convolution. Second, in view of the network structure of the prediction model, the graph convolutional network based traffic flow prediction models are divided into two major categories of combined type and improved type, each of which are analyzed and discussed in detail with representative model structures. In addition, typical datasets commonly used in traffic flow prediction for model performance comparison are reviewed, and a simulation test is conducted using one real dataset to demonstrate the prediction performance of four traffic flow prediction models based on graph convolutional networks. Finally, the future research hotspots and challenges in traffic flow prediction based on graph convolutional networks are prospected.

Key words deep learning; traffic congestion; graph convolutional network; traffic flow prediction