

面向交通流量预测的多组件时空图卷积网络^{*}

冯 宁^{1,2}, 郭晟楠^{1,2}, 宋 超^{1,2}, 朱琪超^{1,2}, 万怀宇^{1,2}

¹(北京交通大学 计算机与信息技术学院, 北京 100044)

²(交通数据分析与挖掘北京市重点实验室(北京交通大学), 北京 100044)

通讯作者: 万怀宇, E-mail: hywan@bjtu.edu.cn



摘 要: 流量预测一直是交通领域研究者和实践者关注的热点问题. 流量数据具有高度的非线性和复杂性, 对其进行精准预测具有很大的挑战. 现有的预测方法大多不能很好地捕获数据的时空相关性. 提出一种新颖的基于深度学习的多组件时空图卷积网络(MCSTGCN), 以解决交通流量预测问题. MCSTGCN 通过 3 个组件分别建模流量数据的近期、日周期、周周期特性. 每个组件同时利用空间维图卷积和时间维卷积有效捕获交通数据的时空相关性. 在美国加利福尼亚州高速公路流量公开数据集上进行了实验, 结果表明, MCSTGCN 模型的预测效果优于现有的预测方法.

关键词: 交通流量预测; 时空相关性; 图卷积网络; 多组件融合

中图法分类号: TP311

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

英文引用格式: Feng N, Guo SN, Song C, Zhu QC, Wan HY. Multi-component spatial-temporal graph convolution networks for traffic flow forecasting. Ruan Jian Xue Bao/Journal of Software, 2019, 30(3): 759–769 (in Chinese). <http://www.jos.org.cn/1000-9825/5697.htm>

Multi-component Spatial-temporal Graph Convolution Networks for Traffic Flow Forecasting

FENG Ning^{1,2}, GUO Sheng-Nan^{1,2}, SONG Chao^{1,2}, ZHU Qi-Chao^{1,2}, WAN Huai-Yu^{1,2}

¹(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

²(Beijing Key Laboratory of Traffic Data Analysis and Mining (Beijing Jiaotong University), Beijing 100044, China)

Abstract: Forecasting the traffic flows is a hot issue for researchers and practitioners in the transportation field. It is very challenging to forecast the traffic flows due to the high nonlinearity and complexity of the data, and most of the existing methods cannot effectively capture the spatial-temporal correlations of traffic flow data. In this paper, we propose a novel deep learning based model, multi-component spatial-temporal graph convolution networks (MCSTGCN), to solve the problem of traffic flow forecasting. MCSTGCN employs three components to respectively model the recent, daily and weekly characteristics of traffic flow data. Each component uses graph convolutions in the spatial dimension and convolutions in the temporal dimension to effectively capture the spatial-temporal correlations of traffic data. Experiments on a public California freeway dataset show that the prediction performance of the MCSTGCN model is better than other existing prevalent methods.

Key words: traffic flow forecasting; spatial-temporal correlation; graph convolutional network; multi-component fusion

智能交通系统(ITS)^[1]正在大力发展, 交通数据预测问题是其中的重要组成部分. 车流量是反映公路运行状态的主要参数之一, 如果能够提前准确地对车流量进行预测, 事先对车辆进行疏导, 可以提高路网的运行能力和

* 基金项目: 国家自然科学基金(61603028)

Foundation item: National Natural Science Foundation of China (61603028)

本文由智能数据管理与分析技术专刊特约编辑樊文飞教授、王国仁教授、王朝坤副教授推荐.

收稿时间: 2018-07-20; 修改时间: 2018-09-20; 采用时间: 2018-11-01

效率,减少交通事故,降低人们的出行时间和成本,降低环境污染。

交通流量预测是典型的时空数据预测问题,不同类别的交通数据内嵌于连续空间,并且随时间动态变化,因此,有效提取时空相关性对解决这类问题至关重要。图 1 所示为流量数据(也可以是车速、车道占用率等其他交通数据)的时空相关性示意图,时间维包含 3 个时间片,空间维的 6 个节点(A~F)表示公路的网状结构。在空间维上,节点的交通状况之间会相互影响(绿色虚线);时间维上,某节点历史不同时刻流量会对该节点未来不同时刻流量产生影响(蓝色虚线);同时,节点历史不同时刻的流量值也会对其关联节点未来不同时刻的流量产生影响(红色虚线)。可见,交通流量在时空维度都存在很强的相关性。而如何从这些复杂、非线性的时空数据中挖掘出隐含的时空模式,并对这些模式进行分析从而提取出有价值的信息,是时空数据挖掘任务的一项极大挑战。

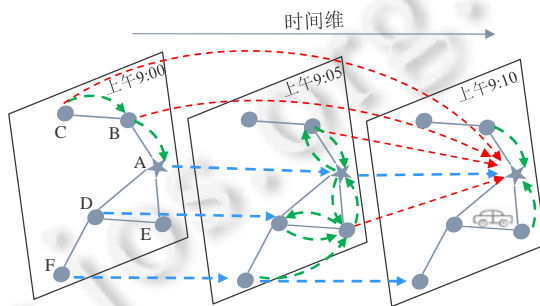


Fig.1 Spatio-Temporal correlation of traffic flow data

图 1 交通流量数据的时空相关性

随着交通行业的发展,道路上部署了许多摄像头、传感器等信息采集设备,这些设备拥有独特的地理空间位置,不断采集各类交通数据,如流量、车速、车道占用率等,积累了大量丰富的带有地理信息的交通时间序列数据,为流量预测提供了很好的数据基础。目前,已有很多学者在这类问题上进行了大量尝试。早期的时间序列预测模型只能用于数据相对稳定、呈线性变化的预测问题,很难适应实际需求。随后发展起来的传统机器学习方法虽然能对更复杂的数据进行建模,但仍然很难同时考虑交通数据的时空相关性,而且这类方法的预测效果很大程度上依赖于特征工程,这往往需要依赖领域专家的经验 and 大量的实验。近年来,很多人将深度学习的方法应用于时空数据预测问题,如利用卷积神经网络(CNN)可以有效提取网格数据的空间特征,但对于图结构的时空数据的特征描述和时空相关性分析,目前仍然还在探索之中。

为了应对上述挑战,我们提出了一种基于深度学习的交通流量预测方法——多组件时空图卷积网络(MCSTGCN)模型,使用基于时空卷积网络的深度学习模型来处理交通路网这类图结构数据,能够同时有效捕获数据的时空特性。该模型分别设置近期、日周期和周周期等 3 个组件来描述交通数据的时空特征,最后将 3 个组件的输出进行融合得到最终的预测结果。本文的主要贡献概括如下:

- 同时考虑并建模了交通流量数据固有的 3 种时间维特性,即近期依赖、日周期性和周周期性,并利用 3 个学习组件分别建模这 3 种特性;
- 提出了 MCSTGCN 模型,能够直接从基于图结构的交通数据中同时有效捕获空间和时间维特征;
- 在真实的高速公路流量数据集上进行实验,验证本文提出模型的预测效果优于其他现有的预测方法。

1 相关工作

交通量预测问题历经多年的不断研究和实践,取得了大量的成果。传统的统计学方法有历史均值法、自回归积分滑动平均法(ARIMA)^[2]等。历史均值法是以历史一段时间数据的均值为依据进行预测,ARIMA 将预测对象随时间推移而形成的数据序列视为一个随机序列,用一定的数学模型近似描述这个序列。这类方法对复杂非线性的交通数据处理能力不足,因为它们都要求数据满足一些前提假设,但实际中的数据往往很难满足这些假设。 K 近邻、支持向量机等机器学习方法能对更复杂的数据进行建模,采用 K 近邻方法的思想是识别相似的交

通状况用于预测^[3],而采用支持向量机方法的思想是通过核函数将低维非线性流量数据映射到高维空间后再进行线性分类^[4].这类方法仍然无法同时有效考虑交通数据的时空相关性,且需要大量的特征工程.随着深度学习在语音识别、图像处理等领域取得的成果,越来越多的人将深度学习应用于时空数据预测问题.Zhang 等人^[5]设计了基于残差卷积单元的 ST-ResNet 来对城市人流量进行预测,这种方法虽然提取了流量数据的时空特征,但将输入限制为标准的 2 维或 3 维网格数据,因而不能用于图结构的高速路网上的交通预测问题.

传统的卷积神经网络可以有效提取数据的局部特征,但只能作用于标准的网格数据.而图卷积可以直接在图结构的数据上实现卷积操作.目前,主流的图卷积方法包括空间方法和谱图方法.空间方法直接把卷积核应用到图上的节点及其邻域,这种方法的核心在于如何选择节点的邻域.Niepert 等人^[6]为中心节点启发式线性选择邻域,将其映射为向量再进行卷积,在社交网络任务中取得了不错的效果;Li 等人^[7]在人体动作识别任务中引入图卷积,提出多种划分策略将节点的邻域划分为不同子集,通过控制子集的个数,保证不同节点可以共享卷积核权重.谱图方法通过图拉普拉斯矩阵将网格数据上的卷积操作推广到图结构数据上.Bruna 等人^[8]提出了一个通用的图卷积框架,将拉普拉斯矩阵的特征向量变换到谱域,再通过样条插值的方法近似求解;随后,Defferrard 等人^[9]优化了该方法,将样条插值替换成切比雪夫多项式的 K 阶截断近似求解,并证明了卷积核的范围严格限制在中心节点的 K 阶邻居内,同时降低了模型复杂度;Seo 等人^[10]提出了图卷积循环网络(GCRN),但在特定设置下很难确定循环网络和图卷积的最佳组合;随后,Li 等人^[11]成功地使用门控循环单元(GRU)和图卷积进行长期交通预测;Yu 等人^[12]提出了一个带有门控机制的图卷积网络,并应用于交通量预测问题.但这些模型都没有考虑交通数据在时间维度上的周期性和趋势性等多种固有特性.

交通流量预测问题的核心在于如何有效捕获数据的时空维特征及相关性.图卷积可以直接对图结构数据进行特征提取,自动挖掘交通数据的空间模式.沿时间轴做卷积操作,可以挖掘交通数据的时间模式.因此,本文采用基于时空图卷积网络的深度学习模型,将二者结合起来同时捕获交通数据的时空特性,有效解决交通流量预测问题.

2 问题定义

2.1 空间路网

我们将空间路网定义为无向图 $G=(V,E,A)$,如图 2(a)所示,其中, V 为节点集, $|V|=N$ 为节点个数; E 为边集,表示节点间的连通性; $A \in \mathbb{R}^{N \times N}$ 为图 G 的邻接矩阵.设在空间路网 G 上的每个节点都会检测 F 个采样频率一致的时间序列数据,即每个节点在每个时间戳都会产生一个长度为 F 的特征向量,如图 2(b)(数据归一化后的示意图)实线所示.

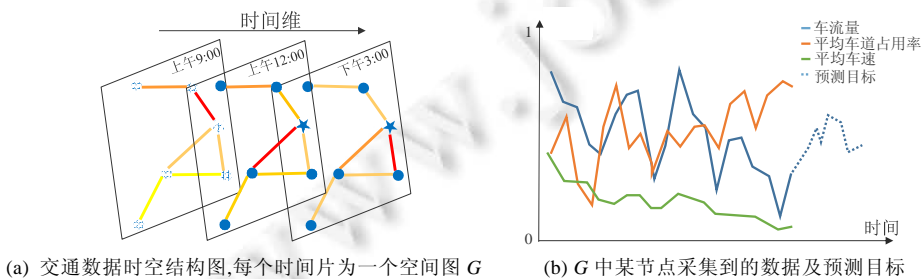


Fig.2

图 2

2.2 流量预测

设路网 G 中各节点的第 $f \in (1, \dots, F)$ 个时间序列为流量序列,未来某一段时间范围内的流量为我们的预测目标.用 $x_t^{c,i} \in \mathbb{R}$ 表示第 i 个节点的第 c 维时间序列在 t 时刻的值, $\mathbf{x}_t^c \in \mathbb{R}^N$ 表示所有节点的第 c 维时间序列在 t 时

刻的值, $\mathbf{X}^c = (\mathbf{x}_1^c, \mathbf{x}_2^c, \dots, \mathbf{x}_\tau^c) \in \mathbb{R}^{N \times \tau}$ 表示所有节点的第 c 维时间序列在 τ 时段内的值, $\mathbf{X}_t = (\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^F)^T \in \mathbb{R}^{F \times N}$ 表示所有节点的所有时间序列在 t 时刻的值, $\mathbf{X} = ((\mathbf{X}^1)^T, (\mathbf{X}^2)^T, \dots, (\mathbf{X}^F)^T)^T = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_\tau) \in \mathbb{R}^{F \times N \times \tau}$ 表示所有节点的所有时间序列在 τ 时段内的值. 设 $y_t^i = x_t^{f,i} \in \mathbb{R}$ 是第 i 个节点在 t 时刻的流量值, 已知网络上所有节点的历史 τ 时段内的所有时间序列值 \mathbf{X} , 我们的预测目标是未来所有节点在窗口长度为 T_p 的流量序列 $\hat{\mathbf{Y}} = (\hat{\mathbf{y}}^1, \hat{\mathbf{y}}^2, \dots, \hat{\mathbf{y}}^N)^T \in \mathbb{R}^{N \times T_p}$, 其中, $\hat{\mathbf{y}}^i = (\hat{y}_{\tau+1}^i, \hat{y}_{\tau+2}^i, \dots, \hat{y}_{\tau+T_p}^i) \in \mathbb{R}^{T_p}$ 表示节点 i 的预测目标, 如图 2(b) 中的蓝色虚线所示. 表 1 列出了本文的符号约定.

Table 1 Mathematical notation

表 1 符号约定

符号	描述
$\mathbf{x}, \mathbf{x}, \mathbf{X}, \mathbf{X}$	输入数据
τ, T	时间序列长度
q	每日采样频率
$\mathbf{L}, \mathbf{D}, \mathbf{A}$	图的矩阵表示 (拉普拉斯矩阵, 度矩阵, 邻接矩阵)
\ast_G, \ast	卷积操作 (图卷积, 卷积)
θ, Θ, Ψ	卷积核参数

3 基于深度学习的多组件时空图卷积网络

图 3 展示了本文提出的多组件时空图卷积 (MCSTGCN) 网络的总体架构, 它由 3 个主要部分组成, 分别用来建模预测目标的近期、日周期及周周期信息 (GCN: 空间维图卷积; Conv: 时间维卷积; FC: 全连接). 例如, 对某地下午 2 点的车流量进行预测, 那么与预测时段直接相邻的历史 1 小时内的交通数据、对应的 1 天前及 1 周前下午 2 点的交通数据都会对该预测问题提供有用信息, 而当天上午 9 点的交通状况往往与预测目标关系不大. 为了得到充足的时间维信息, 同时减少处理不相关历史信息带来的开销, 我们设计了 3 个组件分别用于描述交通数据的 3 种时间维特性: 近期特性、日周期特性及周周期特性.

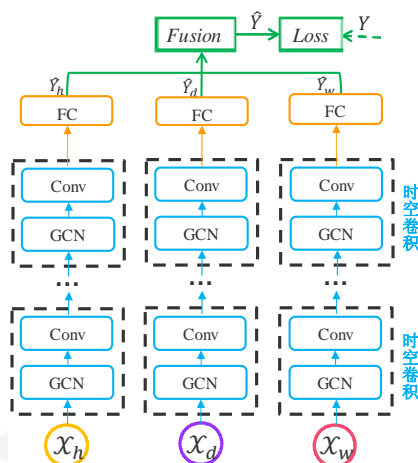


Fig.3 Architecture of MCSTGCN

图 3 MCSTGCN 模型架构图

假设采样频率为每天 q 次, 即时间序列上每天包含 q 个数据点. 设当前时刻 t_0 , 待预测的时间窗口长度为 T_p , 我们沿时间轴分别截取长度为 T_h, T_d 和 T_w 的 3 个时间序列片段作为模型中近期、日周期和周周期 3 个组件的输入, 如图 4 所示 (设预测时段长度 T_p 为 1 小时, T_h, T_d 和 T_w 均取 T_p 的 2 倍), 其中, T_h, T_d 和 T_w 均为 T_p 的整数倍. 3 个时间序列片段分别如下.

- (1) 近期片段: $\mathcal{X}_h = (X_{t_0-T_h+1}, X_{t_0-T_h+2}, \dots, X_{t_0}) \in \mathbb{R}^{F \times N \times T_h}$, 即与预测时段直接相邻的一段历史时间序列片段, 如图 4 中绿色部分所示. 直观地, 道路上车辆的汇聚和发散是逐渐形成的, 因此, 一个节点的上一时刻的流量必然会对其下一时刻流量产生很大的影响;

- (2) 日周期片段:

$$\mathcal{X}_d = (X_{t_0-(T_d/T_p) \times q+1}, \dots, X_{t_0-(T_d/T_p) \times q+T_p}, X_{t_0-(T_d/T_p-1) \times q+1}, \dots, X_{t_0-(T_d/T_p-1) \times q+T_p}, \dots, X_{t_0-q+1}, \dots, X_{t_0-q+T_p}) \in \mathbb{R}^{F \times N \times T_d},$$

由预测时段之前若干天中与预测目标时段相同的序列片段组成, 如图 4 中红色部分所示. 由于受早晚高峰周期、人们的每天作息规律等因素影响, 流量数据在每天的同一时刻往往存在很强的相似性. 日周期组件的目的就是要建模交通数据中以天为单位的周期模式;

- (3) 周周期片段:

$$\mathcal{X}_w = (X_{t_0-7 \times (T_w/T_p) \times q+1}, \dots, X_{t_0-7 \times (T_w/T_p) \times q+T_p}, X_{t_0-7 \times (T_w/T_p-1) \times q+1}, \dots, X_{t_0-7 \times (T_w/T_p-1) \times q+T_p}, \dots, X_{t_0-7 \times q+1}, \dots, X_{t_0-7 \times q+T_p}) \in \mathbb{R}^{F \times N \times T_w},$$

由预测时段之前的若干周中与预测目标星期属性相同且时段相同的序列片段组成, 如图 4 中蓝色部分所示. 同样地, 交通流量数据存在明显的周周期模式, 例如, 星期一的交通模式和历史上星期一的交通模式具有一定的相似性, 而与周末的交通模式可能存在差异. 周周期组件的目的就是要建模交通数据中以周为周期的变化规律.

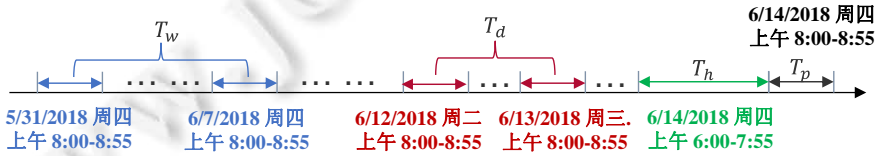


Fig.4 An example of constructing the input of time series segments

图 4 输入时间序列片段构建示例

3 个组件共享相同的网络结构, 由时空卷积模块(如图 5 所示)和全连接层组成, 其中, 时空卷积模块包括空间维度的图卷积和时间维度的标准 2 维卷积两部分. 模型最后将 3 个组件的输出结果基于参数矩阵进行融合, 得到最终的预测结果. 这种结构能够有效捕获交通数据的时间维和空间维特征及其时空相关性.

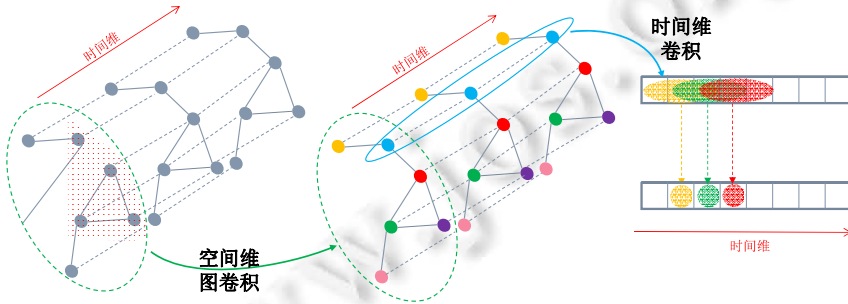


Fig.5 Architecture of spatio-temporal convolutions of MCSTGCN

图 5 MCSTGCN 模型的时空卷积结构

3.1 空间维图卷积

我们首先只考虑某一时间片上的空间图 G , 以了解空间特征的建模过程. 在我们的模型中, 采用谱图方法^[9]将卷积操作推广到图结构数据, 将数据视为图上的信号, 然后直接在图上对图信号进行处理, 来捕获空间中有意义的模式和特征.

谱图方法主要通过将图转化为代数的形式来分析图结构, 在本文中, 我们主要关心图结构中节点间的连通

关系及相互影响,在谱图论中,可以将一个图用其对应的拉普拉斯矩阵来表示.通过分析拉普拉斯矩阵及其特征值就可以得到图结构的性质.图的拉普拉斯矩阵定义为 $L=D-A$,其规范化形式为 $L=I_N-D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \in \mathbb{R}^{N \times N}$,其中, A 为邻接矩阵, I_N 为单位阵,度矩阵 $D \in \mathbb{R}^{N \times N}$ 是节点度数组成的对角矩阵, $D_{ii} = \sum_j A_{ij}$. 对拉普拉斯矩阵进行特征值分解 $L=U\Lambda U^T$,其中, $\Lambda=\text{diag}([\lambda_0, \dots, \lambda_{N-1}]) \in \mathbb{R}^{N \times N}$ 是 L 的特征值组成的对角矩阵, U 是傅里叶基^[12]. 以 t 时刻流量数据为例,图信号为 $\mathbf{x}=\mathbf{x}_t^f \in \mathbb{R}^N$,对图信号进行傅里叶变换可表示为 $\hat{\mathbf{x}}=U^T\mathbf{x}$. 根据拉普拉斯矩阵的性质可知 U 是正交矩阵,因此得到逆傅里叶变换 $\mathbf{x}=U\hat{\mathbf{x}}$. 图卷积是利用定义在傅里叶域中对角化的线性算子来等价代替经典卷积算子^[13]实现的卷积操作,用卷积核 g_θ 对图 G 进行卷积操作^[9]:

$$g_\theta \times_G \mathbf{x} = g_\theta(L)\mathbf{x} = g_\theta(U\Lambda U^T)\mathbf{x} = U g_\theta(\Lambda) U^T \mathbf{x}.$$

由于对图信号进行卷积操作再做傅里叶变换等于对这些信号进行傅里叶变换后的乘积^[14],上式可以理解对 g_θ 和 \mathbf{x} 分别做傅里叶变换到谱域,然后对二者的变换结果进行乘法操作,再做傅里叶逆变换得到卷积操作的结果.

将图变换到谱域实现图上的卷积操作即为图卷积,但当图的规模较大时,直接对拉普拉斯矩阵进行特征值分解代价昂贵,因此,本文采用切比雪夫多项式近似展开求解:

$$g_\theta \times_G \mathbf{x} = g_\theta(L)\mathbf{x} \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})\mathbf{x}.$$

$\theta_k \in \mathbb{R}^K$ 为切比雪夫多项式系数, $\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N$, λ_{\max} 表示拉普拉斯矩阵的最大特征值.切比雪夫多项式的递归定义: $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$,其中, $T_0(x) = 1, T_1(x) = x$. 用切比雪夫多项式近似展开求解,相当于对图中的每个节点采用卷积核提取了以该节点为中心的周围 $0 \sim K-1$ 阶邻居的信息^[9]. 图卷积模块使用线性修正单元(ReLU)为激活函数,即 $\text{ReLU}(g_\theta \times_G \mathbf{x})$.

以 $K=2$ 为例,对每个节点提取其 $0 \sim 1$ 阶邻居信息,对拉普拉斯矩阵特征值进行缩放使 $\lambda_{\max}=2$,上述图卷积操作表示为

$$g_\theta \times_G \mathbf{x} \approx \theta_0 T_0(\tilde{L})\mathbf{x} + \theta_1 T_1(\tilde{L})\mathbf{x} = \theta_0 \mathbf{x} + \theta_1 (L - I_N)\mathbf{x} = \theta_0 \mathbf{x} + \theta_1 (-D^{-\frac{1}{2}} A D^{-\frac{1}{2}})\mathbf{x}.$$

令 $\theta_0 = -\theta_1 = \theta \in \mathbb{R}$ 以减少参数,则 $g_\theta \times_G \mathbf{x} \approx (I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}})\mathbf{x}\theta$, 图上所有节点共享卷积核权重 θ . 同时,为避免数值不稳定、梯度爆炸或梯度消失^[15],我们令 $\tilde{A} = A + I_N, \tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, 得到: $g_\theta \times_G \mathbf{x} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \mathbf{x}\theta$. 拓展到多维数据,第 r 层卷积的输入数据是 $\mathcal{X}_h^{(r-1)} = (\mathbf{X}_h^{1,(r-1)}, \mathbf{X}_h^{2,(r-1)}, \dots, \mathbf{X}_h^{C_{r-1},(r-1)}) \in \mathbb{R}^{C_{r-1} \times N \times T_{r-1}}$, 其中: $r \in \{1, \dots, l\}$ (l 是时空卷积层数); C_{r-1} 是第 r 层网络的输入数据的通道数,当 $r=1$ 时, $C_0=F$; T_{r-1} 是输入数据时间维长度,当 $r=1$ 时,近期组件 $T_0=T_h$ (日周期组件 $T_0=T_d$ 、周周期组件 $T_0=T_w$). 对 $\mathcal{X}_h^{(r-1)}$ 进行图卷积操作,表示为 $g_\theta \times_G \mathcal{X}_h^{(r-1)} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \mathcal{X}_h^{(r-1)} \Theta$, 其中, $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_{C_r}) \in \mathbb{R}^{1 \times C_{r-1} \times C_r}$ 是卷积核参数.

以 $C_{r-1}=1$ 为例,图 6 展示了图卷积过程中对空间节点的 $0 \sim 1$ 阶邻居信息进行提取的过程.图 6(a)为空间路网结构,图 6(b)为根据图 6(a)计算得到邻接矩阵 A 和度矩阵 $D, \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ 及输入数据 $\mathbf{X}^{1,(r-1)}$ 的矩阵表示如图 6(c)所示,由此计算出用卷积核 g_θ 对数据 $\mathbf{X}^{1,(r-1)}$ 进行图卷积操作的结果 $g_\theta \times_G \mathbf{X}^{1,(r-1)}$ 如图 6(d)所示,其中,节点 F 在 $t=2$ 时刻的值由原来的 $x_{6,2}^1$ 变为 $\Theta_1(0.35x_{4,2}^1 + 0.50x_{6,2}^1)$,即,输入数据被其 $0 \sim 1$ 阶邻居的信息更新.其他特征及卷积核同理,对整个输入数据 $\mathcal{X}_h^{(r-1)}$ 进行一次图卷积操作,得到 $g_\theta \times_G \mathcal{X}_h^{(r-1)} \in \mathbb{R}^{C_r \times N \times T_{r-1}}$, 每个节点被该节点的 $0 \sim K-1$ 阶邻居的信息更新.

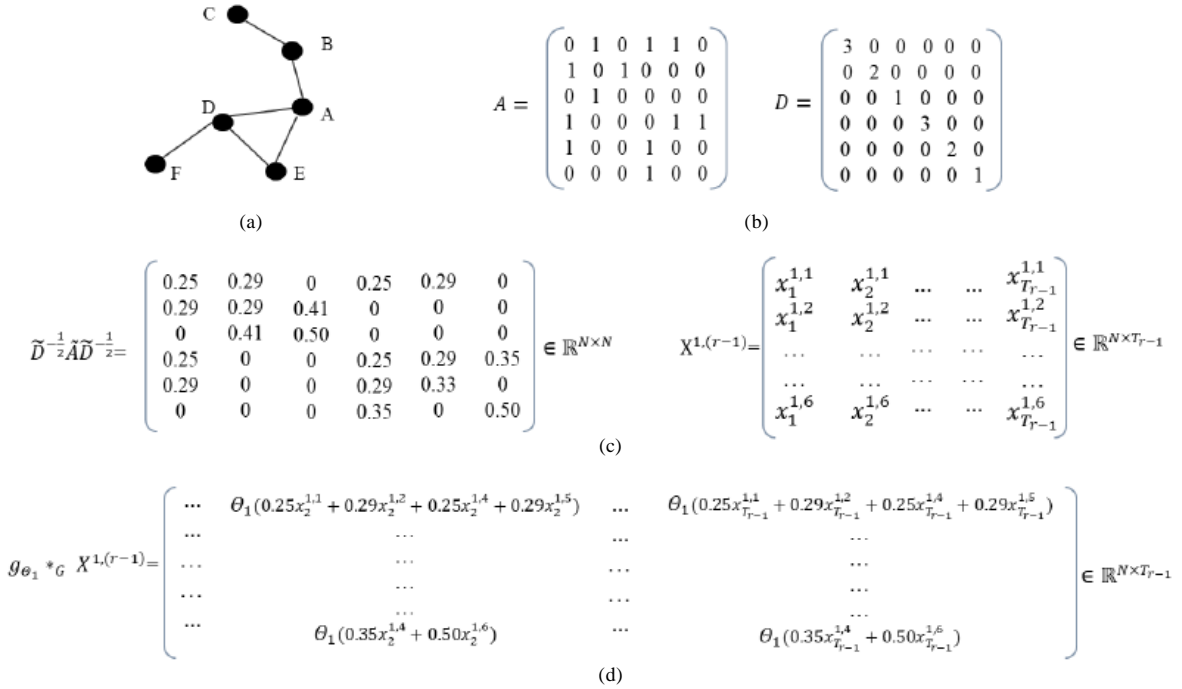


Fig.6 An example of graph convolutions and its matrix representation

图 6 图卷积示例及其矩阵表示

3.2 时间维卷积

通过空间维图卷积操作对输入数据的空间特征进行建模之后,再用标准 2 维卷积捕获时间维特征.我们用线性修正单元激活函数,以提取近期特征为例,得到 $\mathcal{X}_h^{(r)} = \text{ReLU}(\Psi \times (\text{ReLU}(g_{\theta} \times_G \mathcal{X}_h^{(r-1)}))) \in \mathbb{R}^{C_r \times N \times T_r}$ (Ψ 为时间维卷积核的参数),具体卷积过程如图 5 所示.

经过一层时间维卷积之后,节点的信息被该节点相邻时间片信息更新,而节点及其相邻时间片信息在经过图卷积操作后已包含其相邻节点同时刻的信息.因此,通过一层时空卷积操作之后,就会捕获到数据的时间维和空间维特征以及时空相关性.我们使用多层时空卷积,以提取时空维上更远的信息,再通过全连接操作使时空卷积的结果与预测目标维数一致,全连接模块同样使用线性修正单元作为激活函数.

3.3 多组件融合

在本节中,我们将讨论如何融合近期、日周期和周周期组件的输出.以对整个路网周五上午 8 点的流量进行预测为例,有些地方早晚高峰周期规律明显,其日周期、周周期组件的预测结果更为重要;而有些地方不存在明显的交通周期规律,日周期、周周期组件对预测目标起到的帮助较小.可见,不同节点受不同组件的影响程度是不一样的,在融合不同组件的输出结果时,其影响程度也应该不同.因此,应该从历史数据中进行学习.融合后的最终预测结果表示为

$$\hat{\mathbf{Y}} = \mathbf{W}_h \odot \hat{\mathbf{Y}}_h + \mathbf{W}_d \odot \hat{\mathbf{Y}}_d + \mathbf{W}_w \odot \hat{\mathbf{Y}}_w,$$

其中, \odot 是矩阵对应元素相乘的哈达马积; $\mathbf{W}_h, \mathbf{W}_d, \mathbf{W}_w$ 是学习参数,反映了近期、日周期、周周期这 3 种时间维特性对预测目标的影响程度.

4 实验与结果分析

为了检验本文提出模型的性能,我们在两个真实世界的数据集上进行了对比实验.本节将对数据集和实验

设置进行说明,并对实验结果进行详细的对比分析.

4.1 数据集介绍

我们使用美国加利福尼亚州的真实高速公路数据集 PeMSD4 和 PeMSD8 来验证我们的模型. PeMS 是由 Caltrans Performance Measurement System 实时采集的高速公路交通数据,该系统拥有超过 39 000 个传感器站,部署在加利福尼亚州高速公路系统的主要大都市区^[16].数据集是基于 30s/次的频率采样得到的原始数据汇总成的以 5m 为时间间隔的样本.数据集包含带时间戳的车流量、平均车速、平均车道占用率这 3 个维度的特征及采集这些信息的检测器的地理位置信息.

- PeMSD4: San Francisco Bay 区域的数据,共包含 29 条路上的 3 848 个检测器,我们选取时间范围为 2018 年 1 月~2 月的数据进行实验,其中,前 50 天作为训练集,后 9 天作为测试集;
- PeMSD8: San Bernardino 区域的数据,共包含 8 条路上的 1 979 个检测器,我们选取时间范围为 2016 年 7 月~8 月的数据进行实验,其中,前 50 天作为训练集,后 12 天作为测试集.

4.2 数据预处理

我们对道路上的检测器进行筛选,去掉那些距离过近的检测器,保证检测器节点间距离大于 3.5 英里.样本以 5m 为时间间隔,因此,道路上每个节点每天包含 288 个数据点.用线性插值法填充缺失值.另外,对数据进行 0 均值标准化(zero-mean)操作 $x' = x - \text{mean}(x)$,即让数据的平均值为 0.

4.3 实验参数设置

我们基于 Pytorch 框架实现了 MCSTGCN 模型,图卷积使用 32 个相同大小的卷积核.时间维卷积同样使用 32 个相同大小的卷积核,这些卷积核沿空间轴维度为 1,沿时间轴维度为 3,通过控制步长调整时间维长度.整个模型训练时间大约为 20s/轮.3 个组件的输入数据长度会对实验结果产生影响,我们通过实验选定 $T_h=3$, $T_d=1$ 和 $T_w=1$ 为最佳的组合.均方误差是反映估计量与被估计量之间差异程度的一种度量,因此,本文采用该度量指标作为损失函数.

4.4 基准方法

我们将本文提出的模型与以下 7 种已有的时间序列预测方法进行比较.

- HA: 历史均值法,其预测值为近期历史流量状况的平均值.在本文中,我们使用最近 12 个时间片的平均值来预测下一个时间片的值;
- ARIMA^[2]: 自回归积分滑动平均法,是时间序列分析中的一种经典方法;
- LSTM^[17]: 长短期记忆网络,一种特殊的 RNN 模型. LSTM 单元由细胞、输入门、输出门和遗忘门组成;
- GRU^[18]: 门控循环单元网络^[18],一种特殊的 RNN;
- STGCN^[7]: 一种基于空间方法定义的时空图卷积模型,采用了多种划分策略将中心节点的邻居划分到不同子集,实现卷积核的参数共享;
- GCGRU^[11]: 一种将门控循环单元网络与图卷积网络结合的方法;
- GLU-STGCN^[12]: 一种带有门控机制的图卷积网络,在交通数据预测问题上取得了很好的效果.

本文采用均方误差(RMSE)和平均绝对误差(MAE)作为评价指标,具体计算公式下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2},$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|.$$

4.5 实验结果及分析

我们将 MCSTGCN 模型在数据集 PeMSD4 和 PeMSD8 上与前述 7 种基准方法进行了比较,表 2 展示了对未来 1h 内的流量进行预测的结果.

Table 2 Performance comparison of different approaches on the PeMSD4 and PeMSD8 dataset

表 2 不同方法在数据集 PeMSD4 和 PeMSD8 上的性能比较

Model	PeMSD4		PeMSD8	
	RMSE	MAE	RMSE	MAE
HA	54.14	36.76	44.03	29.52
ARIMA	68.13	32.11	43.30	24.04
LSTM	45.82	29.45	36.96	23.18
GRU	45.11	28.65	35.95	22.20
STGCN	38.29	25.15	27.87	18.88
GCGRU	37.25	25.00	30.24	20.69
GLU-STGCN	38.41	27.28	30.78	20.99
MCSTGCN(ours)	35.64	22.73	26.47	17.47

从表 2 可以看出,我们的模型在两种评价指标中均达到最佳性能.我们还可以观察到,传统的时间序列预测方法由于建模能力有限,预测结果并不理想.基于深度学习的方法获得了比传统方法更好的预测结果,而包含时空图卷积机制的模型由于考虑了时空依赖性,预测结果要优于一般的深度学习模型.

此外,我们还比较了各种方法随着预测时长的增长性能变化情况.我们以 5m 为间隔,让预测时长从 5m 增长到 1h,实验结果如图 7 所示.

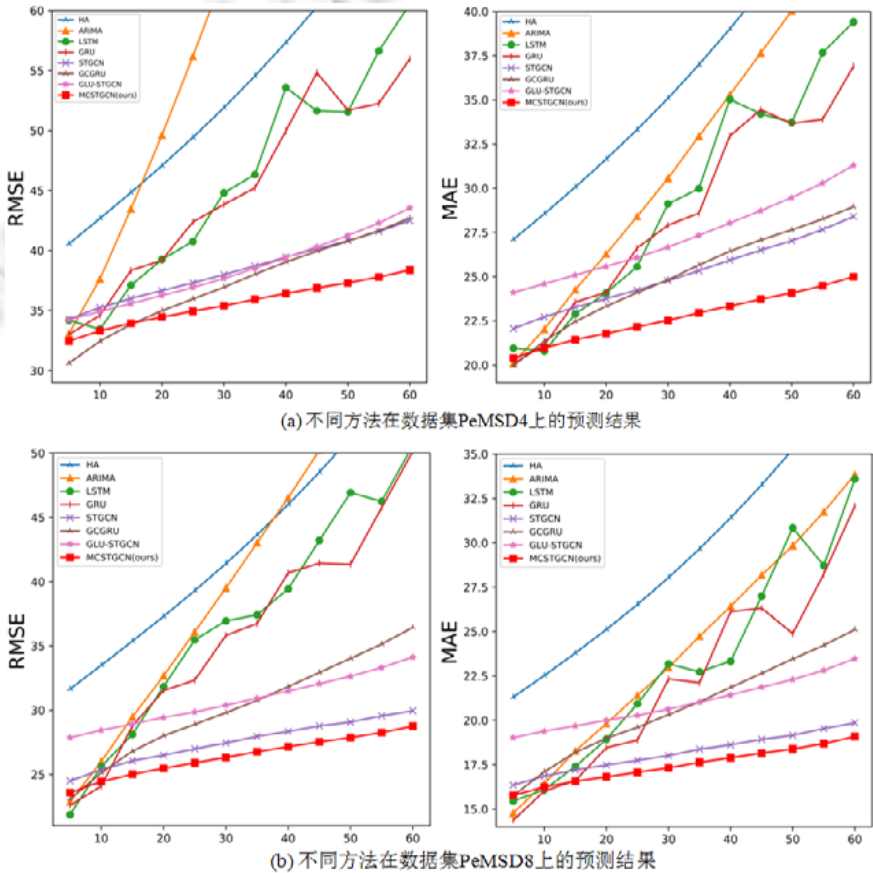


Fig.7 Performance changes of different methods along with the forecasting duration

图 7 不同方法随着预测时长的性能变化情况

从图 7 可以看出,从整体上,随着预测时长的逐步增大,预测难度越来越大,误差整体呈上升趋势.我们的 MCSTGCN 模型在短期预测中就取得了较优的预测结果,这表明了多组件与时空图卷积结合的策略能充分

挖掘数据的时空模式;而随着预测时长的增大,我们的模型预测误差比其他方法增长更加缓慢,这是由于该模型显式建模了时间维度上的多种周期特性,因而在中长期预测中显示更加明显的优势。

最后,为了观察每个组件的输入数据的长度变化对预测结果的影响,我们在 PeMSD8 数据集上将各组件的输入数据长度分别设置为 $T_h \in \{1, 2, 3\}$, $T_d \in \{1, 2, 3\}$, $T_w \in \{1, 2, 3\}$ 进行了实验对比,其结果见表 3。

Table 3 MAE with different component input lengths on dataset PeMSD8

表 3 不同组件输入数据长度在数据集 PeMSD8 上的 MAE

	$T_d=1, T_w=1$	$T_d=1, T_w=2$	$T_d=1, T_w=3$	$T_d=2, T_w=1$	$T_d=2, T_w=2$	$T_d=2, T_w=3$	$T_d=3, T_w=1$	$T_d=3, T_w=2$	$T_d=3, T_w=3$
$T_h=1$	19.26	19.27	19.25	18.10	19.25	19.26	19.25	19.25	19.28
$T_h=2$	18.36	18.47	18.71	18.18	18.01	18.71	18.70	18.70	18.70
$T_h=3$	17.48	17.79	17.47	17.58	17.67	17.47	17.75	17.83	17.48

可以看出, T_h 的变化对预测结果影响较大,而 T_d 和 T_w 影响相对较小,我们选定 $T_h=3, T_d=1$ 和 $T_w=1$ 为最佳的组合。

5 总 结

本文提出一种新颖的多组件时空图卷积网络 MCSTGCN,该模型结合图卷积和标准卷积构造时空卷积块来同时捕获交通数据的时空特性。在真实的高速公路流量数据集上的实验表明,本文提出的模型的预测效果优于其他已有的交通数据预测方法,验证了该模型在捕获时空特征及时空相关性方面具有优势。事实上,除了高速公路流量预测任务外,我们提出的模型也适用于处理其他基于图结构表示的时空交通数据。未来,我们将进一步通过注意力机制等策略来优化模型的网络结构,进一步提升模型的预测能力。

References:

- [1] Zhang JP, Wang FY, Wang KF, Lin WH, Xu X, Chen Ch. Data-driven intelligent transportation systems: A survey. *IEEE Trans. on Intelligent Transportation Systems*, 2011, 12(4):1624–1639.
- [2] Williams BM, Hoel LA. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, 2003, 129(6):664–672.
- [3] Van Lint H, Van Hinsbergen C. Short-term traffic and travel time prediction models. *Transportation Research E-circular*, 2012, 22(1):22–41.
- [4] Jeong YS, Byon YJ, Castro-Neto MM, Easa SM. Supervised weighting-online learning algorithm for short-term traffic flow prediction. *IEEE Trans. on Intelligent Transportation Systems*, 2013, 14(4):1700–1707.
- [5] Zhang J, Zheng Y, Qi D, Li RY, Yi XW, Li TR. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence*, 2018, 259:147–166.
- [6] Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs. In: *Proc. of the ICML*. 2016. 2014–2023.
- [7] Li C, Cui Z, Zheng W, Xu C, Yang J. Spatio-temporal graph convolution for skeleton based action recognition. In: *Proc. of the AAAI*. 2018.
- [8] Bruna J, Zaremba W, Szlam A, Lecun Y. Spectral networks and locally connected networks on graphs. In: *Proc. of the ICLR*. 2014.
- [9] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: *Proc. of the NIPS*. 2016. 3844–3852.
- [10] Seo Y, Defferrard M, Vandergheynst P, et al. Structured sequence modeling with graph convolutional recurrent networks. *arXiv preprint arXiv:1612.07659*, 2016.
- [11] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926v1*, 2017.
- [12] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: *Proc. of the IJCAI*. 2018.
- [13] Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*, 2015.

- [14] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: Proc. of the CVPR. 2017. 29–38.
- [15] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: Proc. of the ICLR. 2017.
- [16] Chen C, Petty K, Skabardonis A, Varaiya P, Jia ZF. Freeway performance measurement system: Mining loop detector data. Journal of the Transportation Research Board, 2001,1748:96–102.
- [17] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997,9(8):1735–1780.
- [18] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv: Neural and Evolutionary Computing, 2014.



冯宁(1993—),女,天津人,硕士,CCF 学生会员,主要研究领域为时空数据挖掘,图数据挖掘.



郭晟楠(1992—),女,博士,CCF 学生会员,主要研究领域为时空数据挖掘,深度学习.



宋超(1995—),男,硕士,主要研究领域为时空数据挖掘,图数据挖掘.



朱琪超(1994—),女,硕士,主要研究领域为时空数据挖掘.



万怀宇(1981—),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为社交网络挖掘,交通数据挖掘.