

DOI: 10.3785/j.issn.1008-973X.2020.06.011

基于图卷积神经网络的城市交通态势预测算法

闫旭, 范晓亮, 郑传潘, 臧彧, 王程, 程明, 陈龙彪

(厦门大学信息学院, 福建厦门 361000)

摘要: 为了实时准确地预测城市交通流量, 提高城市交通态势感知和预测准确度, 提出一种改进的时空图卷积神经网络算法: 基于自由流动可达矩阵的时空图卷积神经网络 (FAST-GCN). 利用图卷积神经网络有效表达城市复杂路网的结构特性, 引入自由流动可达矩阵来挖掘复杂路网的时空依赖性, 从而提高交通态势预测准确度; 对交通流速及站点地理位置数据进行数据预处理; 在现有的时空图卷积神经网络算法的基础上, 增加基于自由流动可达矩阵的图卷积模块, 以有效挖掘城市交通路网的独特空间特征; 通过一个全连接的输出层输出交通流预测结果; 在真实世界数据集 PeMS 上对算法效果进行验证. 结果表明, 采用提出的 FAST-GCN 算法能够有效获取交通路网独特的物理特性, 从而捕获交通数据的时空依赖性, 优于时空图卷积 (STGCN) 等基线算法, 其在 45 min 的预测准确率最好可提高 5.656%; 相比基线模型, 所提算法能够适应大规模路网的交通流预测, 且具有可扩展性.

关键词: 交通流预测; 深度学习; 图卷积神经网络; 时空依赖性; 自由流动可达矩阵

中图分类号: TP 399 **文献标志码:** A **文章编号:** 1008-973X(2020)06-1147-09

Urban traffic flow prediction algorithm based on graph convolutional neural networks

YAN Xu, FAN Xiao-liang, ZHENG Chuan-pan, ZANG Yu, WANG Cheng,
CHENG Ming, CHEN Long-biao

(School of Informatics, Xiamen University, Xiamen 361000, China)

Abstract: An improved spatio-temporal graph convolutional networks traffic prediction algorithm, named free-flow reachable matrix-based spatio-temporal graph convolutional networks (FAST-GCN), was proposed, in order to predict real-time traffic flows accurately and improve the sensing and prediction of citywide traffic situation. The characteristics of urban complex road network structure were expressed effectively by the graph convolutional neural network, and the spatio-temporal dependency in complex road networks was explored by introducing free-flow reachable matrices. Thus the accuracy of traffic situation prediction was improved. First, preprocess traffic speeds and sensors location data. Second, with the existing spatio-temporal graph convolutional networks, the graph convolution module based on free flow reachable matrix was integrated to effectively capture the unique spatial characteristics of the urban traffic road networks. Finally, the prediction results were generated through a fully connected output layer. The proposed model was evaluated on a real-world traffic dataset PeMS. The experimental results show that this model could capture physical characteristics of road network and spatio-temporal dependency, and outperform the baselines such as spatio-temporal graph convolutional networks (STGCN), and the prediction accuracy in 45 minutes was improved by up to 5.656%. In addition, compared with baselines, the proposed model can adapt to traffic flow prediction in large-scale road networks and has superior scalability.

Key words: traffic prediction; deep learning; graph convolution neural network; spatio-temporal dependency; free-flow reachable matrix

收稿日期: 2020-01-03. **网址:** www.zjujournals.com/eng/article/2020/1008-973X/202006011.shtml

基金项目: 国家自然科学基金资助项目 (61872306, U1605254, 61971363, 61802325); 厦门市科技局资助项目 (3502Z20193017).

作者简介: 闫旭 (1997—), 女, 硕士生, 从事智能交通系统研究. orcid.org/0000-0001-6072-069X. E-mail: yanxu97@stu.xmu.edu.cn

通信联系人: 范晓亮, 男, 高级工程师. orcid.org/0000-0002-6958-665. E-mail: fanxiaoliang@xmu.edu.cn

为了解决交通拥堵与交通事故、城市敏感区的人群聚集风险以及发生自然灾害后的人群出行等一系列交通问题,世界各国研究者纷纷对智能交通系统(intelligent transportation system, ITS)展开了研究. ITS的核心是交通控制与诱导系统,而实现交通控制与诱导系统的核心工作是实时准确地进行交通流预测,即有效地融合历史交通规律和实时交通数据来预测城市交通系统的运行态势^[1].

精准地感知和预测城市交通态势是ITS领域的一项重要内容^[2-3],相关研究工作对缓解交通拥堵、减少交通事故及预防人群异常聚集事件等均具有重要应用价值. 现今的交通态势感知和预测研究主要面临以下挑战. 1)交通数据的复杂时空依赖性导致交通态势预测面临实时性困难. 随着时间变化,2个不同地点的空间相关性亦发生变化. 例如,从工作日到周末:在工作日,家庭住所与公司的相关性较强;而在周末,家庭住所与公司的相关性较弱. 此外,对于相同的2个地点,不同时段的时间相关性并不是线性的,即与当前时段最相关的很可能是距离该时段很远的时段. 例如,工作日早高峰时段与其前后平峰时段的相关性较弱,而往往与工作日晚高峰时段的时间相关性更强. 2)交通流量感知设备采集的数据存在缺失、错误等问题. 当前道路状况检测的检测设备(如环形感应线圈、雷达、摄像头等)^[4-6]已经可以获得较高准确度的交通状况数据,但仍存在使用寿命有限及设备故障等问题. 此外,对于检测设备自身的信息(如传感器的邻接信息等)部分,相关数据也难以完全统计.

为了有效应对上述挑战,本文提出一种改进的时空图卷积深度神经网络算法,基于自由流动可达矩阵的时空图卷积深度神经网络(free-flow reachable matrix-based spatio-temporal graph convolutional networks, FAST-GCN). 一方面利用图卷积神经网络有效表达城市复杂路网结构的特性,另一方面引入自由流动可达矩阵来挖掘复杂路网的时空依赖性,从而有效提高交通态势预测准确度. 首先,对交通流速及站点地理位置数据进行数据预处理,清理交通流速数据并根据地理位置数据计算自由流动可达矩阵. 其次,在现有的时空图卷积深度神经网络算法的基础上,增加基于自由流动可达矩阵的图卷积模块,以有效挖掘城市交通路网的独特空间特征. 最后,通过一个全连接的输出层输出交通流预测结果.

1 相关工作

1.1 早期交通流预测算法

交通流预测是ITS最具挑战性的问题之一,其目标为根据给定的一系列历史观测数据对接下来一段时间的交通状况进行预测.

交通流预测方法主要分为2类,分别是动态建模方法和数据挖掘方法. 动态建模方法主要使用数学工具与物理方法,模拟交通系统的动态变化,从而生成预测交通流的公式^[7]. 数据挖掘方法主要是在交通状况的历史数据中挖掘数据规律,从而对未来的数据走势进行预判. 数据挖掘方法受到更多的关注,逐渐发展成为交通流预测的主流方法.

统计学方法与机器学习方法是挖掘交通数据的2种主要方法. 在时间序列分析中,传统的统计学方法占有主导地位,自回归积分滑动平均模型(autoregressive integrated moving average model, AR-IMA)^[8]在短期预测方面表现出明显的优势^[9],然而在中长期预测方面,其准确率大幅度降低. 尽管机器学习算法,如:支持向量机回归(support vector regression, SVR)^[10]等,优于传统的统计学方法,但其仅在时间维度挖掘序列数据变化趋势,很难解决复杂的交通流预测问题. 为了挖掘时间序列以外的交通特征,研究人员将注意力逐渐转移到深度学习方法.

将深度信念网络(deep belief nets, DBN)应用至交通流预测^[11],能够有效挖掘交通数据的高维特征,从而一定程度地降低预测误差,但其难以从输入数据中提取具体的时空特征. 循环神经网络(recurrent neural network, RNN)、长短期记忆网络(long short-term memory, LSTM)^[12]以及门控循环单元(gated recurrent unit, GRU)^[13]等基于RNN的模型能够为相邻时刻的交通数据建立联系,并通过门控等方式保存记忆,以学得交通流序列的长期依赖关系,因而在交通流预测问题上表现出强大的优势与潜力^[14]. 然而这些模型只能从噪声中捕捉复杂的时间特征,并不能够模拟真实世界中的交通路网空间结构. 为了进一步捕捉空间特征, Ma等^[15-16]使用卷积神经网络(CNN)^[17]学习交通路网的空结构. 然而,传统的CNN更适合用于捕捉欧拉数据的空间关系,交通路网属于图结构的非欧拉数据,因此使用CNN学习空间特征并不是学习交通路网结构的最优方法. 可

见,早期的交通流预测算法存在局限性,早期算法大多结合基于RNN的模型与基于CNN的模型,将二者分别用于捕捉交通数据的时空特征,然而,基于CNN的方法并不完全适用于捕获图结构的路网空间特征.于捕获图结构的路网空间特征.

可见,早期的交通流预测算法存在局限性,早期算法大多结合基于RNN的模型与基于CNN的模型,将两者分别用于捕捉交通数据的时空特征,然而,基于CNN的方法并不完全适用于捕获图结构的路网空间特征.

1.2 基于图神经网络的交通流预测算法

近年来,由于图在众多领域(如社交网络、知识图谱等)强大的表现力,图神经网络(graph neural network, GNN)^[18]受到了越来越多的关注;GNN令人信服的性能和较高的可解释性,使其成为一种应用广泛的图结构分析方法^[19].将图卷积运算扩展到图结构数据中可捕获交通路网信息,Li等^[20-27]利用图卷积神经网络(graph convolutional networks, GCN)对交通流预测问题进行建模.

利用GCN进行图卷积的方法主要有2类.第一类方法为图谱卷积,即利用图谱理论,通过设计基于图拉普拉斯矩阵的图谱滤波器完成卷积.Li等^[20]融合图谱卷积、Seq2seq体系与预定抽样等技术捕捉交通流的时空依赖性,获得了较高的预测准确率,但时间复杂度较高.Yu等^[21]将2个捕捉时间特征的门控卷积层与1个捕捉空间特征的图谱图卷积层结合,组装成时空卷积块进行交通流预测,打造了准确率更高的模型并降低了时间复杂度.虽然两者均将卷积应用于图结构数据以捕获图的空间特征并取得了良好的效果,但图谱卷积不能完全适应交通网络的物理特性.第二类图卷积方法是直接在图上定义卷积,从邻居结点中获取信息.Cui等^[22]使用自由流动速度计算自由流动可达矩阵,结合关系矩阵、邻接矩阵与自流矩阵,构造图卷积核捕获交通路网的物理特性.然而,这种方法要求数据集提供传感器的邻接信息,对数据集信息的要求过多,缺乏普适性.

以上研究在有效捕获交通数据的时空依赖性以及充分利用城市复杂路网的特性这两方面均存在不足.因此,本文提出一种改进的时空图卷积深度神经网络算法,一方面利用图卷积神经网络能够有效表达城市复杂路网结构的特性,另一方面引入自由流动可达矩阵来挖掘复杂路网的时空依赖性,从而有效提高交通态势预测准确度.

2 理论基础

2.1 交通流预测问题定义

交通流预测是典型的时间序列预测问题,即,给定 p 个时间间隔的交通状况观测值,预测接下来 q 个时间间隔最可能的交通状况(例如:交通流速或交通流量):

$$\mathbf{v}_{t+1}, \mathbf{v}_{t+2}, \dots, \mathbf{v}_{t+q} = \arg \max_{\mathbf{v}_{t+1}, \mathbf{v}_{t+2}, \dots, \mathbf{v}_{t+q}} \left[\log P(\mathbf{v}_{t+1}, \mathbf{v}_{t+2}, \dots, \mathbf{v}_{t+q} | \mathbf{v}_{t-p+1}, \dots, \mathbf{v}_{t-1}, \mathbf{v}_t) \right]. \quad (1)$$

式中: $\mathbf{v}_t \in \mathbf{R}^N$ 为 t 时刻 N 个路段的交通流速观测值向量,对应每一个路段在每个时间戳存在的一个交通流速数值; P 为概率函数.

由于交通流预测为结构化时间序列预测,将图结构应用于交通网络,即在图上定义交通路网.本文定义交通网络为 $G(\mathbf{V}, \mathbf{E}, \mathbf{W})$.其中, $\mathbf{V} \in \mathbf{R}^N$ 为顶点,即交通路网中的传感器站点; $\mathbf{E} \in \mathbf{R}^{N \times N}$ 为边,即交通路网中传感器站点之间的链路; $\mathbf{W} \in \mathbf{R}^{N \times N}$ 为权重矩阵,即交通路网中每2个传感器站点之间的相关系数.

2.2 基于自由流动矩阵的图卷积

由于非欧拉的图结构数据不能直接进行常规的网格卷积,研究者对图域内卷积的方法进行了文献探索^[20-21].目前图域中卷积的主流方法主要有2种:一种是图谱卷积,利用图谱傅里叶变换在谱域中进行操作,在频谱域中应用卷积的频谱框架;另一种是非图谱卷积,扩展卷积的空间定义,即将图的顶点排列为网格形式进行网格卷积.前者将图数据应用至图谱中模拟图结构,但方法较为固定;后者提取邻居信息的方式更加灵活,更有利于捕捉交通路网独特的物理结构,因此本研究选用非图谱卷积方法模拟路网结构.

2.2.1 自由流动矩阵 为了定义图卷积,首先在图上定义邻接矩阵 $\mathbf{A} \in \mathbf{R}^{N \times N}$ 表示图中结点的邻接关系,若结点 i 与节点 j 间存在链路,则 $A_{ij}=1$,否则 $A_{ij}=0$ (当 $i=j$ 时, $A_{ij}=0$).基于此邻接矩阵,定义一个链路计算方程 $d(v_i, v_j)$,计算从结点 i 到结点 j 经过的最少链路数.然后定义一个距离矩阵 \mathbf{D} , D_{ij} 表示结点 i 到结点 j 所须经过的最短路网距离(当 $i=j$ 时, $D_{ij}=0$).

为了展示更清晰的邻居关系,为每个结点 i 定义 k 阶邻居 $\mathbf{B}_i^k = \{\mathbf{B}_i^k \in \mathbf{R}^N, v_j \in \mathbf{V} | d(v_i, v_j) \leq k\}$,表示结点 i 的邻居结点为其经过 k 步可到达的结点 j . k 阶邻居矩阵 \mathbf{A}^k 可通过 \mathbf{A} 的阶乘来计算,定义为

$$\tilde{A}^k = Z \left(\prod_{i=1}^k A + I \right) = Z(A^k + I). \quad (2)$$

式中: $Z(\cdot)$ 为将矩阵 $(A^k + I) \in \{0, 1\}^N$ 归一化的函数, 将单位矩阵 I 加入至 k 阶邻接矩阵 A^k 中可实现结点的自我可达. 因此, k 阶图卷积 g^k 定义如下:

$$g^k = (W_{g^k} \odot \tilde{A}^k) X_t. \quad (3)$$

式中: \odot 为矩阵对应元素相乘操作, W_{g^k} 为 k 阶邻接矩阵的权重矩阵, 表示道路网络各个结点间的关系, $X_t \in \mathbf{R}^{N \times M}$ 为 t 时刻结点的特征向量, 本文仅使用交通流速这一个特征, $M=1$, $X_t \in \mathbf{R}^{N \times 1}$ 为 t 时刻图中 N 个结点的交通流速向量.

考虑真实交通路网中车辆的物理特性, 某路段对相邻路段的影响与当前交通状况、路段特征、驾驶员性格和车辆特征均相关, 结点 i 不能绕过中间节点到达其非邻接结点 j , 因此本文使用可达矩阵 $F^m \in \mathbf{R}^{N \times N}$ 近似表示 k 阶邻接矩阵, 定义如下:

$$F_{ij}^m = \begin{cases} 1, & S_{ij}^F m \Delta t - D_{ij} \geq 0; \\ 0, & \text{其他.} \end{cases} \quad (4)$$

式中: S_{ij}^F 为车辆自由流动速度, 即从路段 i 行驶至路段 j 在不发生交通拥堵与其他意外情况 (如封路、台风、沙尘暴等) 时的平均速度, m 为单位时间间隔的个数, Δt 为单位时间间隔. 此时, 图卷积 g^m 被定义为

$$g^m = (W_{g^m} \odot F^m) X_t. \quad (5)$$

其中, m 可以控制 F^m 的稀疏程度, 用来确定结点间的邻近性. m 越大, 则对应 k 阶邻近矩阵中的 k 越大, 意味着每个路段的相邻路段越多. g^m 的计算时间复杂度为 $O(n^2)$.

2.2.2 图谱卷积 图谱卷积 (spectral graph convolution, SGC)^[28] 为基于图的拉普拉斯算子的傅里叶域中的卷积运算, 定义如下:

$$\vartheta *_G X = \vartheta(L)X = \vartheta(U\Lambda U^T)X = U\vartheta(\Lambda)U^T X. \quad (6)$$

式中: ϑ 为卷积核, $*_G$ 为图谱卷积运算符, $X \in \mathbf{R}^N$ 为输入图信号, $\vartheta(\Lambda)$ 为滤波器, 是对角阵, 归一化的图拉普拉斯算子为

$$L = I_N - D^{-1/2} W D^{-1/2} = U\Lambda U^T \in \mathbf{D}^{N \times N}. \quad (7)$$

式中: I_N 为单位矩阵; W 为图的关系情况; $D_{ii} = \sum_j W_{ij}$, 为图的入度矩阵, 是对角阵; $U \in \mathbf{R}^{N \times N}$ 为 L 的特征向量; $\Lambda \in \mathbf{R}^{N \times N}$ 为 L 的特征矩阵, 是对角阵.

Defferrard 等^[29] 提出了局部图谱卷积 (localized spectral graph convolution, LSGC), 其使用多项式

滤波器, 定义如下:

$$\vartheta(\Lambda) = \sum_0^{K-1} \theta_k T_k(\Lambda). \quad (8)$$

式中: $T_k(\Lambda)$ 为 k 阶多项式, θ_k 为其对应的参数系数, LSGC 的优势在于训练参数少, 仅有 K 个参数; 且其无须进行特征向量的计算.

此处, Yu 等^[21] 提出的 STGCN 算法所使用的图卷积方法为 LSGC, 所用多项式 $T_k(\Lambda)$ 为切比雪夫多项式.

2.2.3 与图谱卷积的区别 基于自由流动可达矩阵的非图谱卷积 FAST-GC 与 SGC 和 LSGC 的区别与联系如表 1 所示. 与 LSGC 相比, 在参数数量与时间复杂度 2 个方面, FAST-GC 的参数更多, 但由于图卷积的时间复杂度均来源于矩阵乘法, 三者的时间复杂度相同. 在是否能够提取空间特征这一方面, SGC 由于没有限制滤波器的空间, 不能提取局部特征; LSGC 通过 k 阶多项式, 将卷积限制在一个以 k 为半径的球空间内, 能够提取 k 阶局部特征; FAST-GC 结合了自由流动可达矩阵 F_m , 提取 m 阶局部特征, 同时可以表示交通路网的物理特性. 在提取路网物理特性这一维度, FAST-GC 优于 LSGC. 综上, 提出的基于自由流动矩阵的非图谱卷积在交通流预测方面优于 SGC 与 LSGC.

表 1 FAST-GC、SGC 与 LSGC 的性质比较

Tab.1 Character comparison between FAST-GC, SGC and LSGC

模型	定义公式	参数数量	时间复杂度
FAST-GC	$W_{g^m} \odot F^m$	N^2	$O(n^2)$
SGC	$U\vartheta(\Lambda)U^T$	N	$O(n^2)$
LSGC	$\sum_0^{K-1} \theta_k T_k(\Lambda)$	K	$O(n^2)$

3 FAST-GCN 算法

3.1 算法框架

本文提出的基于自由流动可达矩阵的时空图卷积深度神经网络 (FAST-GCN) 算法的总体框架如图 1 所示. 图 1(b) 中, C 为通道数, l 为时空卷积块的层数; 图 1(c) 中, $\sigma(\cdot)$ 为 Sigmoid 函数; 图 1(d) 中, $\text{ReLU}(\cdot)$ 为线性整流函数. 算法框架由 1 个时空卷积块和 1 个全连接的输出层连接组成, 其中时空卷积块由 2 个时间门控卷积模块和 1 个空间图卷积模块按“时间-空间-时间”结构连

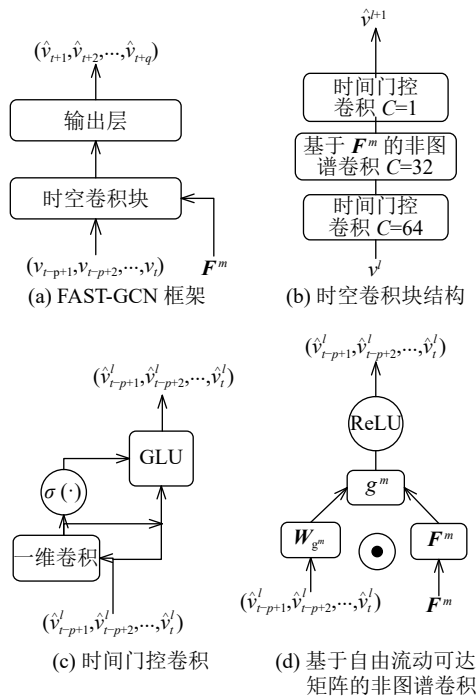


图 1 基于自由流动可达矩阵的时空图卷积深度神经网络 (FAST-GCN) 算法总体框架

Fig.1 Overall framework of free-flow reachable matrix-based spatio-temporal graph convolutional networks (FAST-GCN) algorithm

接组成. 模型训练所运用的损失函数如下:

$$L(W_\theta, \hat{v}) = \sum_t \sum_i \|\hat{v}_{t+1}^i - v_{t+1}^i\|^2 \quad (9)$$

式中: W_θ 为模型中所有可学习参数, \hat{v} 为交通流速预测值向量, \hat{v}_{t+1}^i 为 t 时间戳站点 i 的交通流速预测值, v_{t+1}^i 为 t 时间戳站点 i 的交通流速实际值.

3.2 空间特征提取模块

空间特征提取的传统方法主要为 CNN, 即将交通网络划分为一个个均等的网格, 从而对每个网格进行卷积. 然而, 这样的交通路网划分方式只考虑到了局部路网连接性, 忽略了路网结构的全局连接性. 图谱卷积在提取交通数据空间特征的过程中, 将路网作为一个整体, 能够成功学得交通路网的全局连接性, 然而此种图卷积不能有效挖掘交通路网独特的物理特性. 考虑提取全局连接的交通路网独特的物理特性, 本研究参考 Cui 等^[22] 提出的交通图卷积方法, 使用基于自由流动可达矩阵的非图谱卷积方法对路网进行建模, 该方法的详细介绍见第 2.2.1 节, 网络结构见图 1(d). 基于自由流动可达矩阵的非图谱卷积定义如下:

$$W *_{G} X = g^m = (W_{g^m} \odot F^m) X \quad (10)$$

式中: W 为图卷积神经网络的卷积核, $*_{\Gamma}$ 为图卷

积操作, X 为相应输入数据.

3.3 时间特征提取模块

虽然基于 RNN 的模型 (如: LSTM) 在时间序列预测方面表现良好, 但是均存在模型训练时间开销过大以及对数据的动态变化反应迟缓等问题. CNN 在训练模型时具有训练快速、结构简单以及能快速捕捉数据变化等优势, 因此本研究参考 Yu 等^[21] 提出的方法, 使用基于门控线性单元 (gated linear unit, GLU) 的门控卷积神经网络捕捉交通流的时间动态变化特征, 网络结构见图 1(c). 时间门控卷积的定义如下:

$$\Gamma *_{\Gamma} X = W_a [*_a(X) + X] \odot \sigma(W_b *_{\Gamma} X) \quad (11)$$

式中: Γ 为门控卷积神经网络的卷积核, $*_{\Gamma}$ 为门控卷积操作, $*_a$ 和 $*_b$ 为卷积操作, W_a 和 W_b 为卷积核, $W_a [*_a(X) + X]$ 和 $\sigma(W_b *_{\Gamma} X)$ 为 GLU 的门控输入.

3.4 预测模块

为了融合第 3.2 节与 3.3 节分别捕捉的空间特征与时间特征, 从而捕获交通数据的时空依赖性, 本研究参考 Yu 等^[21] 提出的方法, 采用瓶颈策略产生 1 个“时间-空间-时间”结构, 如图 1(b) 所示, 将 1 个图卷积神经网络嵌入 2 个门控神经网络中, 使两者组成 1 个时空卷积块, 作用于图结构的时间序列数据. 此外, 为了防止过拟合问题的出现, 本研究还对时空卷积块的输出结果进行归一化处理.

整个时空卷积块的计算公式如下:

$$v^{j+1} = \Gamma^1 *_{\Gamma} \text{ReLU} [W *_{G} (\Gamma^0 *_{\Gamma} v^j)] \quad (12)$$

式中: Γ^0 和 Γ^1 分别为时空卷积块中第 1、2 个时间门控卷积神经网络的卷积核.

为了获得预测结果, 将时空卷积模块与输出层连接, 如图 1(a) 所示. 其中, 输出层由 1 个门控卷积神经网络、1 个卷积神经网络和 1 个全连接层连接组成.

3.5 算法伪代码

算法一: 基于自由流动可达矩阵的时空图卷积算法

输入: $v_{1,2,\dots,t}^j$ 交通流速数据; F^m 自由流动可达矩阵;

输出: FAST-GCN 训练模型.

1. 根据第 4.3.4 节设置模型训练参数
2. 输入数据标准化
3. 将交通流速数据按 7:2:1 划分成训练集、测试集和验证集
4. F^m 读入并设置为图卷积核
5. **For** 每个 epoch:

6. $v_{t-p+1,t-p+2,\dots,t}^{l+1} = \text{st_conv}(v_{t-p+1,t-p+2,\dots,t}^l, F^m)$
7. $v_{t+1,t+2,\dots,t+q}^{l+1} = \text{Output_ful}(v_{t-p+1,t-p+2,\dots,t}^{l+1})$
8. 计算损失函数, 更新模型参数
9. 保存模型
10. **Function** $\text{st_conv}(v_{t-p+1,t-p+2,\dots,t}^l, F^m)$:
 11. 时间门控卷积层: 激活函数=GLU
 12. 空间图卷积层: 激活函数=ReLU
 13. 时间门控卷积层: 激活函数=ReLU
14. **Function** $\text{output_ful}(v_{t-p+1,t-p+2,\dots,t}^{l+1})$:
 15. 时间门控卷积层: 激活函数=GLU
 16. 时间门控卷积层: 激活函数=Sigmoid.

4 实验设计与验证

4.1 数据集描述

PeMS 项目^[6]在加利福尼亚州主要大城市的州际公路上部署了超过 39 000 个传感器站点, 由 Caltrans 绩效测量系统 (PeMS) 实时收集数据. 本文所用数据为 PeMS 项目从 2012 年 5 月至同年 6 月在加利福尼亚州第 7 区 (D7 区) 以 5 min 为时间间隔获取的站点流速数据集, 共有 80 619 552 条数据. 同时, 该数据集还包括每个站点的地理位置信息, 共 4 589 条数据. 因此, 数据可分为两部分: 交通流速数据和站点地理位置数据. 本文在 D7 区随机选择一个包括 181 个站点的小型数据集和一个包括 1 045 个站点的大型数据集, 分别命名为 PeMSD7(S) 和 PeMSD7(L).

4.2 数据预处理

交通流速数据的预处理步骤如下: 1) 提取交通流速数据; 2) 清理交通流速数据, 即将数据字段均为空值且存在数据异常的站点剔除; 3) 使用线性插值方法填充缺失数据; 4) 使用 z-score 方法对交通流速数据进行标准化处理并筛选站点, 生成 PeMSD7(S) 与 PeMSD7(L) 两个数据集, 并将每个数据集的工作日数据与周末数据分开作为数据输入.

站点地理位置数据的预处理步骤如下: 1) 提取站点地理位置数据. 2) 利用站点地理位置数据中的经纬度, 计算 2 个数据集各自的近似路网距离矩阵 D . 3) 参照式 (4), 计算 2 个数据集各自的自由流动可达矩阵 F^m , 其中 S_{ij}^F 使用路段的平均限速表示, Δt 根据数据情况自行定义.

4.3 实验设置

4.3.1 实验环境 本实验使用 Python 3.6.7 基于

TensorFlow 1.9.0 实现, 实验的编译和测试均在 Linux 服务器上进行, 服务器配置信息如下: CPU 为 Intel(R) Xeon(R) CPU @ 2.30 GHz, GPU 为 Tesla T4, 内存为 15 079 MB, 操作系统为 Ubuntu 18.04.2.

4.3.2 评价指标 本研究引入 3 个评价指标: 平均绝对误差 (mean absolute error, MAE)、平均绝对百分误差 (mean absolute percentage error, MAPE) 和均方根误差 (rooted mean square error, RMSE).

4.3.3 基线模型 基线模型包括以下 5 种: 1) 自回归积分滑动平均模型 (ARIMA)^[8]; 2) 支持向量机回归 (SVR)^[10]; 3) 卷积神经网络 (CNN)^[15]; 4) 长短期记忆网络 (LSTM)^[12]; 5) 时空图卷积网络 (STGCN)^[21].

4.3.4 参数设置 本文算法在进行多次调参的对比实验后, 最终确定的实验参数设置如下: 所有 CNN 卷积核数量均为 1, 时空卷积块中的第 1 个时间门控卷积网络输入通道数量为 1, 输出通道数量为 32, 激励函数为 GLU; 空间图卷积神经网络的输入通道数量为 32, 输出通道数量为 32, 激励函数为 ReLU; 第 2 个时间门控卷积网络的输入通道数量为 32, 输出通道数量为 64, 激励函数为 ReLU. 使用 RMSprop 最小化 50 轮的均方误差来训练模型, 批量大小为 50, 初始学习率为 0.001.

模型输入的历史时间窗格为 60 min, 即有 12 个交通流速数据观测值 ($p=12$), 预测的未来时间窗格为 15、30、45 min, 即有 3、6、9 个交通流速数据观测值 ($q=3, 6, 9$).

4.4 实验结果和分析

4.4.1 模型确定 为了寻找能够得到最优预测结果的自由流动可达矩阵 F^m , 使用 5 个不同的 m 值, 即 $m=1, 2, 3, 4, 5$, 计算出 5 个 PeMSD7(S) 的 F^m 作为数据输入, 并使用该数据集的周末数据进行训练与测试, 以选择能够得到算法预测准确率最高的 F^m . 如图 2 所示为在不同 m 值下预测未来 15、30 和 45 min 内交通状况的 MAE. 如图 3 所示为在不同 m 值下训练模型的训练损失情况, 为了更好地观察结果, 图中每隔 15 个点取一个点. 图 3 中, L 为训练损失, S 为训练步数.

观察图 2 与 3 发现, 当 $m=1$ 时, 模型测试结果的 MAE 最小, 模型训练过程中的收敛速度较快, 与其他 m 值对应的结果相差不大, 训练损失最低; 随着 m 值的增大, 模型测试结果的 MAE 指标也随之增大, 模型训练过程中的收敛速度减缓, 训练损失也增加. 出现此结果的原因可能如下: 本研

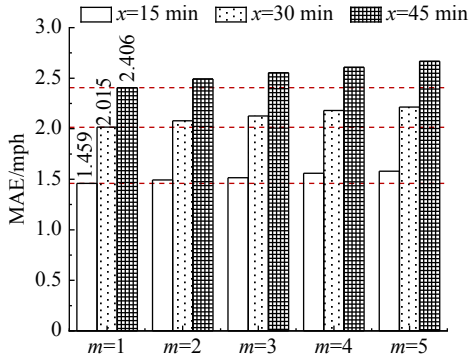


图 2 基于 PEMS7 (S) 周末数据的 FAST-GC 中不同阶数下的训练模型的平均绝对误差比较

Fig.2 MAE comparison of FAST-GC with different orders based on data on weekends of PEMS7 (S)

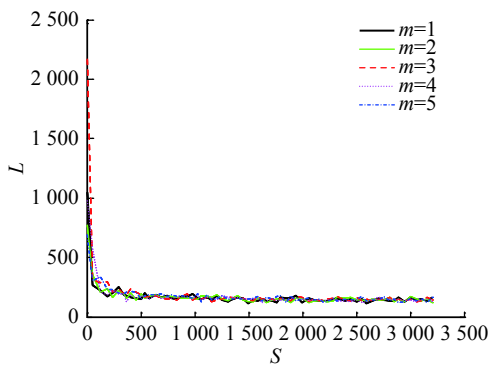


图 3 基于 PEMS7 (S) 周末数据的 FAST-GC 中不同阶数下的训练效率比较

Fig.3 Training efficiency comparison of FAST-GC with different orders based on data on weekends of PEMS7 (S)

研究所用的 PeMSD7(S)数据集涵盖地域范围较小,此数据集内的传感器站点距离较近,可达性较高;当 m 过大时,所有站点彼此均可达,此时的 F^m 矩阵无法有效描述路网特征,因此随着 m 值增大,模型预测准确率下降.相较于其他 m 值,当 $m=1$ 时生成的自由流动可达矩阵 F^m 能更有效地表示站

点间的可达关系并描述路网特征,因此本研究选择 $m=1$ 时生成的 F^m 作为数据输入,进行接下来实验结果的对比分析.在使用不同数据集进行模型训练与测试时,须选择与自身数据相匹配的自由流动可达矩阵 F^m .

4.4.2 实验结果和分析讨论

(1) FAST-GCN 的预测精确度分析.

观察表 2、3 的实验结果(加粗部分)可知,FAST-GCN 算法的准确率在所有评价指标上远优于经典算法 ARIMA、SVR、CNN 和 LSTM;FAST-GCN 算法的预测准确率较 STGCN 有了明显的提升,其在 15、30 和 45 min 的预测准确率比 STGCN 分别提高了 2.524%、4.413% 和 5.656%. 分别观察模型在短时预测与中、长时预测的实验结果,发现在短时交通流预测(15 min)上,FAST-GCN 的预测效果均优于 STGCN;在中、长时交通流预测(即 30 和 45 min)方面,FAST-GCN 的部分指标略低于 STGCN,主要原因是中、长时交通流预测存在显著的扰动因素(如:交通事故)及误差累积等问题,因此,中、长时交通流预测结果的参考价值有限.

(2) FAST-GCN 的可扩展性分析.

将算法在数据集 PeMSD7(S)与 PeMSD7(L)的实验结果作对比,即分别观察并对比表 2~3,发现 SVR 与 LSTM 算法在 PeMSD7(L)上的效果远差于其在 PeMSD7(S)上的效果.这意味着随着数据集规模增大,传统算法的数据挖掘与数值预测能力变差,不能适应大规模的路网交通流预测,不具有可扩展性.本文算法在 PeMSD7(S)和 PeMSD7(L)上的效果相差不大,甚至在 PeMSD7(L)上的效果更优,因此,本文 FAST-GCN 算法能够适应大规模的路网交通流预测,在路网规模上具有可扩展性.

表 2 基于 PEMS7 (S) 和 PEMS7 (L) 的工作日数据采用不同方法训练模型的交通流预测准确度结果

Tab.2 Traffic prediction performance comparison of different approaches for model training based on weekdays data of PeMSD7(S) and PeMSD7(L)

算法	MAE(15/30/45 min)		MAPE(15/30/45 min)		RMSE(15/30/45 min)	
	PeMSD7(S)	PeMSD7(L)	PeMSD7(S)	PeMSD7(L)	PeMSD7(S)	PeMSD7(L)
ARIMA	3.635/4.069/4.462	3.398/3.793/4.147	9.486/10.438/11.302	8.703/9.553/10.338	8.594/9.158/9.704	8.133/8.632/9.131
SVR	4.026/4.628/5.090	3.830/4.433/4.864	12.373/13.992/15.193	11.873/13.272/14.264	8.605/9.388/10.007	8.344/9.142/9.709
CNN	3.256/3.721/3.876	3.292/3.417/3.436	7.995/9.350/10.084	8.182/8.652/8.814	5.618/6.524/6.858	5.928/6.254/6.294
LSTM	3.091/3.240/3.383	3.202/3.238/3.289	7.510/7.925/8.315	8.037/8.132/8.258	5.742/6.124/6.473	6.088/6.171/6.283
STGCN	1.878/2.564/3.052	1.742/2.434/2.953	4.359/6.233/7.560	4.095/5.842/7.043	3.839/5.440/6.454	3.669/5.293/6.410
FFR-STGCN	1.842/2.574/3.094	1.745/2.387/2.850	4.306/6.279/7.731	4.098/5.828/6.999	3.780/5.391/6.460	3.631/5.130/6.092

表 3 基于 PEMS7 (S) 和 PEMS7 (L) 的周末数据采用不同方法训练模型的交通流预测准确度结果

Tab.3 Traffic prediction performance comparison of different approaches for model training based on weekends data of PeMS7(S) and PeMS7(L)

算法	MAE(15/30/45 min)		MAPE(15/30/45 min)		RMSE(15/30/45 min)	
	PeMS7(L)	PeMS7(S)	PeMS7(L)	PeMS7(L)	PeMS7(S)	PeMS7(L)
ARIMA	2.511/2.778/3.019	2.124/2.350/2.546	5.773/6.285/6.761	4.824/5.259/5.646	6.498/6.861/7.212	5.768/6.075/6.361
SVR	4.157/4.562/4.825	3.536/3.890/4.135	11.289/12.053/12.542	8.832/9.442/9.862	8.984/9.395/9.662	7.829/8.239/8.516
CNN	3.502/3.863/3.976	2.863/2.930/3.093	8.040/9.133/9.663	6.652/6.807/7.295	6.506/7.391/7.694	5.727/5.887/6.216
LSTM	3.254/3.359/3.457	2.743/2.753/2.768	7.522/7.794/8.036	6.373/6.417/6.478	6.460/6.725/6.958	5.854/5.900/5.960
STGCN	1.530/2.122/2.527	1.322/1.759/2.057	3.185/4.577/5.486	2.896/4.077/4.787	3.249/4.691/5.569	3.006/4.271/5.011
FFR-STGCN	1.486/2.045/2.428	1.310/1.741/2.029	3.108/4.469/5.339	2.855/4.119/4.919	3.167/4.484/5.254	2.992/4.214/4.891

4.4.3 训练效率分析 比较 STGCN 与 FAST-GCN 的训练效率. 观察记录模型训练过程中每轮的训练时间, STGCN 每轮的训练时间为 8.526 s, FAST-GCN 为 5.554 s, FAST-GCN 比 STGCN 降低了 34.858%. 总体上, FAST-GCN 训练效率更高. 因此, FAST-GCN 在训练效率上优于 STGCN.

5 结 语

本研究针对城市交通态势预测问题, 提出了一种改进的时空图卷积深度神经网络 (FAST-GCN) 算法. 该算法在现有的时空图卷积算法中, 引入了自由流动可达矩阵, 有效挖掘出交通路网独特的物理特性与复杂路网的时空依赖性. 本研究在 PeMS 数据集上开展了大规模的实验和验证工作, 结果表明, 相比于其他经典算法, FAST-GCN 在预测准确度、路网规模的可扩展性以及训练效率方面均具有一定优势, 所提模型预测准确率在 15、30 和 45 min 的预测上, 最多比 STGCN 提高了 2.524%、4.413% 和 5.656%, 并且在 PeMS7(L) (1045 个站点) 上的预测准确率高于在 PeMS7(S) (181 个站点) 上的预测准确率. 此外, 由于模块的缩减, 每轮的训练时间也缩减了 34.858%.

未来改进工作主要如下: 一是继续探索如何改善深度学习算法的可解释性, 并尝试发现模型具有良好预测准确度的因果关系; 二是将本文模型扩展到其他时空数据场景, 如: 水资源消耗、空气质量预测等.

参考文献 (References):

[1] 刘静, 关伟. 交通流预测方法综述 [J]. *公路交通科技*, 2004,

21(3): 82–85.

LIU Jing, GUAN Wei. Overview of traffic flow prediction methods [J]. *Journal of Highway and Transportation Research and Development*, 2004, 21(3): 82–85.

[2] 李德仁, 姚远, 邵振峰. 智慧城市中的大数据 [J]. *武汉大学学报: 信息科学版*, 2014, 58(6): 631–640.

LI De-ren, YAO Yuan, SHAO Zhen-feng. Big data in smart cities [J]. *Journal of Wuhan University: Information Science Edition*, 2014, 58(6): 631–640.

[3] 陆化普, 李瑞敏. 城市智能交通系统的发展现状与趋势 [J]. *工程研究: 跨学科视野中的工程*, 2014, 6(1): 6–19.

LU Hua-pu, LI Rui-min. Development status and trends of urban intelligent transportation systems [J]. *Engineering Research: Engineering in Interdisciplinary Perspectives*, 2014, 6(1): 6–19.

[4] HAJIMOLAHOSEINI H, AMIRFATTAHI R, SOLTANIAN-ZADEH H. Robust vehicle tracking algorithm for nighttime videos captured by fixed cameras in highly reflective environments [J]. *IET Computer Vision*, 2014, 8(6): 535–544.

[5] SEN R, MAURYA A, RAMAN B, et al. Kyun queue: a sensor network system to monitor road traffic queues [C] // *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. Toronto: ACM, 2012: 127–140.

[6] CHEN C, PETTY K, SKABARDONIS A, et al. Freeway performance measurement system: mining loop detector data [J]. *Transportation Research Record*, 2001, 1748(1): 96–102.

[7] VLAHOGIANNI E I. Computational intelligence and optimization for transportation big data: challenges and opportunities [M] // *Engineering and Applied Sciences Optimization*. Cham: Springer, 2015: 107–128.

[8] MAKRIDAKIS S, HIBON M. ARMA models and the Box–Jenkins methodology [J]. *Journal of Forecasting*, 1997, 16(3): 147–163.

[9] HAMED M M, AL-MASAEID H R, SAID Z M B. Short-term prediction of traffic volume in urban arterials [J]. *Journal of Transportation Engineering*, 1995, 121(3): 249–254.

- [10] SMOLA A J, SCHOLKOPF B. A tutorial on support vector regression [J]. **Statistics and Computing**, 2004, 14(3): 199–222.
- [11] HUANG W, SONG G, HONG H, et al. Deep architecture for traffic flow prediction: Deep belief networks with multitask learning [J]. **IEEE Transactions on Intelligent Transportation Systems**, 2014, 15(5): 2191–2201.
- [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. **Neural Computation**, 1997, 9(8): 1735–1780.
- [13] CHO K, VAN MERRIENBOER B, GULVEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. **arXiv Preprint**, arXiv: 1406.1078, 2014.
- [14] CUI Z, KE R, WANG Y. Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction [J]. **arXiv Preprint**, arXiv: 1801.02143, 2018.
- [15] ZHANG J, ZHENG Y, QI D. Deep spatio-temporal residual networks for citywide crowd flows prediction [C] // **Thirty-First AAAI Conference on Artificial Intelligence**. San Francisco: AAAI, 2017.
- [16] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. **Neural computation**, 1989, 1(4): 541–551.
- [17] MA X L, DAI Z, 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.
- [18] 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.
- [19] ZHOU J, CUI G, ZHANG Z, et al. Graph neural networks: A review of methods and applications [J]. **arXiv Preprint**, arXiv: 1812.08434, 2018.
- [20] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [J]. **arXiv Preprint**, arXiv: 1707.01926, 2017.
- [21] YU, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting [J]. **arXiv Preprint**. arXiv: 1709.04875, 2017.
- [22] CUI Z, HENRICKSON K, KE R, 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**, 2019.
- [23] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting [C] // **Proceedings of the AAAI Conference on Artificial Intelligence**. Hawaii: AAAI, 2019, 33: 922–929.
- [24] ZHANG N, GUAN X, CAO J, et al. A hybrid traffic speed forecasting approach integrating wavelet transform and motif-based graph convolutional recurrent neural network [J]. **arXiv Preprint**. arXiv: 1904.06656, 2019.
- [25] SUN J, ZHANG J, LI Q, et al. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks [J]. **arXiv Preprint**. arXiv: 1903.07789, 2019.
- [26] CHEN C, LI K, TEO S G, et al. Gated residual recurrent graph neural networks for traffic prediction [C] // **Proceedings of the AAAI Conference on Artificial Intelligence**. Hawaii: AAAI, 2019, 33: 485–492.
- [27] ZHENG C, FAN X, WANG C, et al. GMAN: a graph multi-attention network for traffic prediction [J]. **arXiv Preprint**, arXiv: 1911.08415, 2019.
- [28] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains [J]. **IEEE Signal Processing Magazine**, 2013, 30(3): 83–98.
- [29] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering [C] // **Advances in Neural Information Processing Systems**. Barcelona: NIPS, 2016: 3844–3852.