

基于 GRU 的密集连接时空图注意力网络的城市交通预测^①

郭海锋^{②*} 许宏伟^{**} 周子盛^{**}

(* 浙江工业大学网络空间安全研究院 杭州 310014)

(** 浙江工业大学信息工程学院 杭州 310023)

摘要 城市道路拓扑结构的复杂性、交通流量的实时变化以及多元的外部环境等因素给交通预测带来了极大的困难。现有方法对交通路网的时空特征挖掘性不足,缺乏对外部因素的考虑,为此本文提出了一种基于门控循环单元(GRU)的时空图注意力密集连接网络,通过门控循环单元来捕获路网数据的动态规律,并以图注意力密集连接网络来提取路网复杂的空间结构特征,建立城市交通网络对时空的依赖关系。针对外部客观因素,采用独热编码的方式对城市各路段发生的交通事件进行数据建模,增强交通网络的信息属性。以杭州申花路及周围共 309 个路段为例,对所提出模型的预测能力和可行性进行验证。实验结果表明,模型预测精度最高达到了 81.64%,与传统数学模型和主流的神经网络模型对比,预测精度较 ARIMA 提高了 35.42%,较图注意力网络(GAT)和 GRU 神经网络分别提高了 17.45% 和 3.02%。实验证明该方法可以适应复杂的交通流进行长期的交通预测任务,同时也能增强交通管理能力,减少交通拥堵成本。

关键词 交通预测;时空特征;神经网络;门控循环单元(GRU);密集连接;图注意力网络(GAT)

作为交通管理系统的重要组成部分,交通预测在交通规划、交通管理和控制中发挥着重要作用。交通预测是指对城市道路网络数据进行分析,进而预测未来的交通状况,包括速度、流量以及通行时间等。掌握未来交通状态可以大幅缓解交通拥堵带来的不便,减轻城市环境压力^[1],同时也能提高整个交通网络的可靠性。路网结构的复杂、交通流数据的时变性以及环境因素和路网状态的不确定性给交通预测带来了极大的困难。因此,充分捕捉城市交通路网的时间和空间特征是完成交通预测任务的关键。

近年来,随着智能交通系统和机器学习的不断发展,交通预测成为城市交通领域主要的研究课题之一。现有方法主要分为模型驱动方法和数据驱动

方法 2 类。模型驱动方法是通过全面的先验知识进行系统建模以解释交通流量、速度和密度之间的瞬态和稳态关系,其代表方法有排队论模型^[2]、交通速度模型^[3]和微观基本图模型^[4]等。实际交通数据受到多方面因素的影响,这些模型依赖人的经验和先验知识的假设,无法准确描述复杂的现实环境中交通数据的变化。数据驱动模型方法是基于数据的规律性来分析变化趋势,并应用到交通状态的预测和评估中^[5-6],主要可以分为统计学模型、机器学习模型以及深度神经网络模型。统计学模型有以历史平均数据作为预测结果的人类联想记忆(human association memory, HAM)模型^[7],其计算简单,但是预测精度低;统计学模型还包括差分整合移动平均自回归模型(autoregressive integrated moving aver-

① 国家自然科学基金(52072343),浙江省自然科学基金(LY20E080023)和国家重点研发计划(2019YFE0126100)资助项目。

② 男,1977 年生,博士,副教授;研究方向:智能交通系统和人工智能;联系人,E-mail: guohf@zjut.edu.cn。

(收稿日期:2023-01-26)

age, ARIMA) 模型、线性回归^[8]和卡尔曼滤波等模型。ARIMA 时间序列模型是将观测到的时间序列拟合为一个参数模型来预测未来的数据;线性回归模型建立了一个基于历史交通数据的回归函数来预测交通流量^[9];卡尔曼滤波模型则是根据测量不同时间下的交通数据,考虑各时间下的联合分布来对未来的交通数据进行预测。随后出现了基于机器学习的交通预测模型,最初是将 k-最近邻算法(k-nearest neighbors, KNN)应用于交通流量,然后使用贝叶斯推理方法^[10]和支持向量机(support vector machines, SVM)^[11]进行研究。这些模型可以模拟更加复杂的交通流特征,但对非线性特征的捕捉能力有限。近几年,深度学习在捕捉非线性和复杂模式上的优势引起了广泛关注,有研究者将深度学习引入到交通预测领域,如深度信念网络(deep belief network, DBN)^[12-13]、堆栈自编码神经网络(stacked auto encoder, SAE)^[14]。这些方法可以独立处理交通流信息,但是无法建立数据在时间维度上的依赖关系。为此,本文研究了基于时间序列的循环神经网络(recurrent neural network, RNN),该网络可以利用历史时间序列数据来预测未来的时序。但 RNN 在反向传播时会出现梯度消失或梯度爆炸的情况,对此有学者提出长短时记忆网络(long short-term memory, LSTM)^[15-16]和门控循环单元(gate recurrent unit, GRU)^[17-18]用来解决上述问题。文献[19]利用长短时记忆网络来对交通速度进行预测,克服了 RNN 模型梯度消失的问题。GRU 相对于 LSTM 仅有 2 个门,参数量较少,容易达到收敛状态,在一些细节较多的交通预测任务中,GRU 预测效果更佳^[20]。

交通流数据除了具有时间依赖性,还会受到路网拓扑结构的影响,相邻路段的交通状态会相互牵制,如上游路况会对中下游路况产生短期连锁反应。通过引入卷积神经网络(convolutional neural networks, CNN)提取空间特征信息,并与 LSTM^[21-22]相结合来捕捉路网时间和空间特征,提高预测精度。但 CNN 比较适用于欧几里得空间设计,城市路网空间结构具有随机性和复杂性,CNN 在处理这种拓扑结构时具有较大的局限性。文献[23]提出用图卷积网络(graph convolutional network, GCN)来处理道

路空间结构^[24],可以更好地建模路段对交通路网的空间依赖性^[25-26]。GCN 在汇聚邻近路段信息时,无法区分出相邻路段对目标路段不同的影响力。在 GCN 基础上,文献[27]引入图注意力网络(graph attention network, GAT)来对相邻路段分配不同的权重系数,解决了该问题。还有研究利用局部空间图来进行时空注意层堆叠^[28],但 GCN 和 GAT 网络本质属于低通滤波器,经过多层汇聚后出现过平滑现象^[29],节点特征呈相似性;因此图神经网络存在的瓶颈^[30]会导致对空间特征提取不足,实现深度图卷积^[31]是解决该问题的关键。城市交通还会受到各种外界因素影响,例如天气、交通事件、周围环境等。除了提取路网的时间和空间信息外,对外部因素建模可以增加交通信息属性,提高交通预测精度。文献[32]集成了一个基于 LSTM^[33]的编码器,对外部信息进行编码,并将多模态数据建模为序列输入,利用外部信息进行交通预测任务。文献[34]通过对动态和静态因素分别进行编码,并引入到时空模型中增加输入信息的特征,该方法被证明相较于单一的时间序列,多源数据信息可明显提高交通预测性能。此外,特殊事件也会造成交通流的剧烈变化,有学者捕捉某些特殊事件的隐藏变量^[35]来学习时间序列中的动态模式,完善交通预测任务。

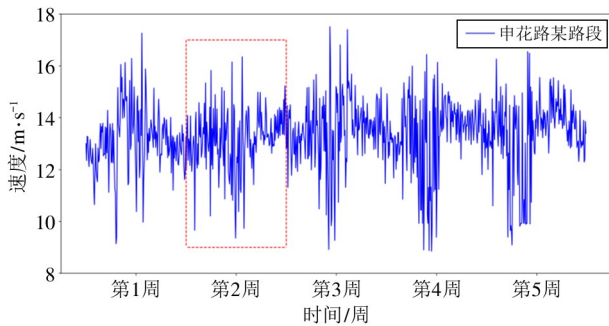
综上所述,考虑多维的路网特征是交通预测任务未来的方向。本文提出了一种新的时空神经网络,该方法不仅可以从交通数据中捕获复杂的时空特征,还可以通过对路网外部因素建模,用于城市道路的交通预测任务,提高预测精度。

1 数据分析和问题描述

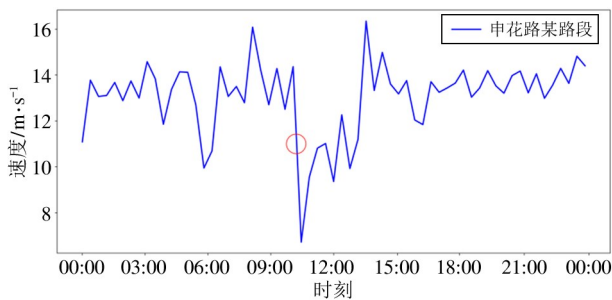
本文将城市道路的实时平均速度作为实验研究对象,选取杭州申花路及周围共 309 条路段进行 2 个月的数据记录。收集的数据分为 2 个部分:第 1 部分是记录各路段的速度时序,时间间隔为 15 min,记作 $X = (X_1, X_2, \dots, X_T)$, $X \in R^{T \times N}$,其中 T 代表记录的时间步数, N 代表路网中所包含的路段数量;第 2 部分是记录整个交通路网所发生的交通事件,为了与速度时序相匹配,记录时间窗口同样为

15 min, 对其进行独热编码, 即当前时间步发生交通事件的路段记为“1”, 其他为“0”, 建模后的交通事件矩阵 $\mathbf{K} = (K_1, K_2, \dots, K_T)$, $\mathbf{K} \in R^{T \times N}$ 。通过对收集的路网数据进行分析, 发现城市交通网络在时间和空间上都具有较强的相关性。

时间特性: 选取某一条路段连续 5 周的交通速度进行可视化, 如图 1(a) 所示, 区域框代表第 2 周的路段速度变化情况, 其整体趋势与其他几周较为吻合, 城市路段的交通状态受到时间约束, 呈现周期性变化。将该路段其中一天的交通速度可视化, 如图 1(b) 所示, 一天内的交通状态会随时间发生变化, 交通流受到之前或更长历史时刻的交通状况的影响, 当前时刻的交通状况对未来交通流也会产生影响。数据表明交通流随时间动态变化, 主要体现在周期性和趋势上。



(a) 路段速度在连续 5 周内的周期性变化



(b) 路段速度在一天内的变化趋势

图 1 交通流的时间特性

空间特性: 选取其中 3 个相邻的路段, 对一段时间内的交通速度可视化, 根据道路方向将 3 个路段分别标记为上游路段、中游路段以及下游路段。如图 2 所示, 上游路段速度在第 30 min 有明显上升趋势, 与其相邻的中游路段速度在第 45 min 同样出现了上升趋势, 下游路段在之后时间步的速度也得到

了相应的提升。这说明由于路网在短期内流动传输的强相关性, 3 个路段的交通状态变化趋势比较接近, 上游路段通过传递效应影响下游路段, 下游路段通过反馈也会影响上游路段, 反映了城市道路交通状态具有空间依赖性。

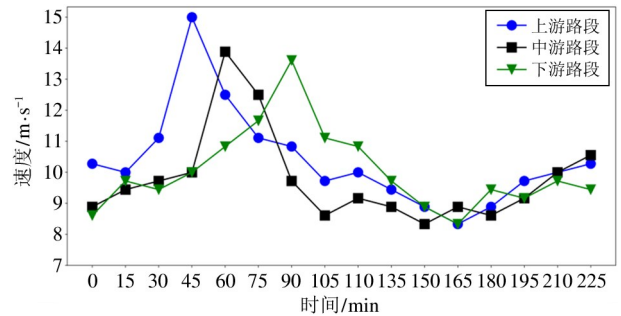


图 2 交通流的空间特性

交通预测任务不仅依赖于历史交通信息和空间关系, 交通事件对交通流的影响也较大。图 3 描述了 2 个路段发生交通事件后, 路段的速度随时间的变化情况。路段 1 和路段 2 分别在 135 min 和第 180 min 发生了交通事件, 在此期间路段速度发生了骤降, 同时由于交通流通性, 也会波及到周围相邻路段。交通事件的发生极大地阻碍了路段车辆的通行, 使得交通流数据产生尖锐的变化, 产生诸多未知因素, 对未来交通状态影响较大, 造成了交通预测的困难。因此, 对该外部因素进行合理的建模可以增加对路网提取特征的维度, 增强模型对外部信息的感知, 从而提高交通预测的精度。

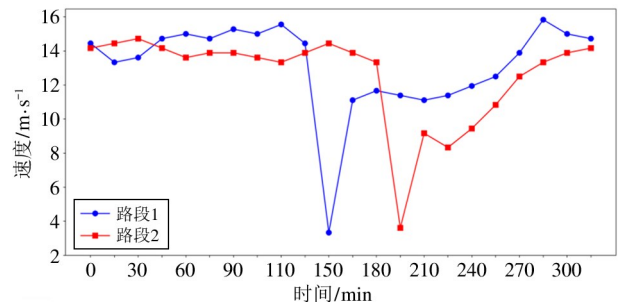


图 3 交通事件对交通流的影响

综上所述, 城市交通道路网络受到时间、空间以及交通事件等外在因素的影响。本文提出了一种时空图注意力神经网络模型来对路网进行特征提取, 并将外部因素建模为路段中的属性信息, 提高模型

对外部信息的感知能力。交通预测的主要任务是利用前 l 项观测到的速度时序 $X = (X_{t-l}, X_{t-l+1}, \dots, X_t)$ 来预测未来 T 个时间步的速度 $X = (X_{t+1}, X_{t+2}, \dots, X_T)$, 其中 X_t 是 t 时刻各路段的实际速度, X_{t+1} 是下一时间步 $t + 1$ 时刻的预测速度。

2 模型搭建

2.1 空间特征建模

获取路网复杂的空间依赖关系是交通预测的关键,卷积神经网络(CNN)利用卷积核来提取局部特征,该方法适用于图像和规则性的网格结构。城市道路网络呈现无规则拓扑结构,路段排列随机且复杂,CNN 无法准确地捕捉到路段与路段之间的潜在联系。本文引入图注意力网络 GAT 来处理路网空间结构,将注意力机制和 GCN 结合。如图 4(a)所示,GAT 在汇聚节点信息时,对每个邻居节点赋予不同的权重系数,在传播过程中引入自注意力机制,每个节点的隐藏状态通过注意其邻居节点来计算。该网络既可以准确地捕捉到路网的空间特征,又能判断出相邻路段对目标路段不同的影响力,更好地学习到局部特征的依赖关系。对节点 (i, j) 的注意力系数计算方法为

$$e_{ij} = \alpha^T [W \vec{h}_i \parallel W \vec{h}_j] \quad (1)$$

其中, e_{ij} 是节点 j 对节点 i 的注意权重系数; h_i 和 h_j 是当前输入层的节点 i 和节点 j 的特征表示; $W \in R^{F \times F'}$ 是线性变化矩阵,属于共享参数; F 表示输入特征的维度; F' 是输出特征的维度; \parallel 为向量的拼接。最后对目标节点所有的注意权重系数进行归一化操作:

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \quad (2)$$

其中 N_i 是节点 i 的所有邻接节点,利用得到的权重系数进行节点信息汇聚时可以捕捉重要节点的隐藏特征。GAT 也可以实现多头(multi-heads)注意力机制,如图 4(b)所示,将每一层单头信息传递的向量进行汇聚,使用多组注意力权重处理时间序列可以学习不同的语义信息,更加有效地进行特征提取,具体过程如式(3)所示。

$$\vec{h}_i = \sigma \left(\frac{1}{k} \sum_{k=1}^k \sum_{j \in N_i} \alpha_{ij}^k W^k \vec{h}_j \right) \quad (3)$$

其中, K 为多头注意力的个数, α_{ij}^k 为第 k 个注意力机制 (α^k) 计算的归一化。

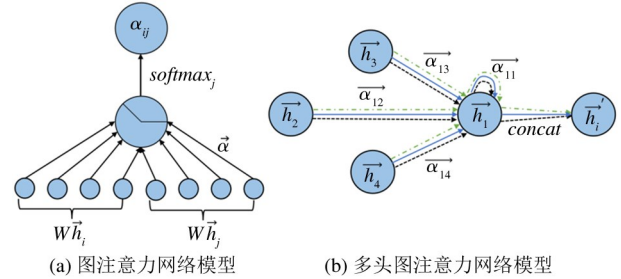


图 4 图注意力神经网络模型框架

GAT 网络属于低通滤波器,在多层叠加后会出现平滑现象,节点呈现相似性,表征能力下降;另外在邻接节点权重较小的情况下,对次邻接节点汇聚时会导致信息损失。为解决上述问题,通过引入密集连接网络^[36](dense network)对 GAT 进行深度建模,组成密集连接块(Dense-GAT)。如图 5 所示,每层多头图注意力网络与前面所有层的输入输出进行向量拼接,利用特征重用的方法来提高空间提取能力,提高模型对路网空间结构的感知度。密集连接网络借用了残差网络^[37](ResNet)的思想,在加强空间特征传递的同时,也减轻了模型在反向传播过程中出现的梯度消失。传播方式如式(4)所示。

$$x_i = H_i([x_0, x_1, \dots, x_{i-1}]) \quad (4)$$

其中, x_i 表示第 i 层 GAT 网络的输出, H_i 表示非线性转换函数。

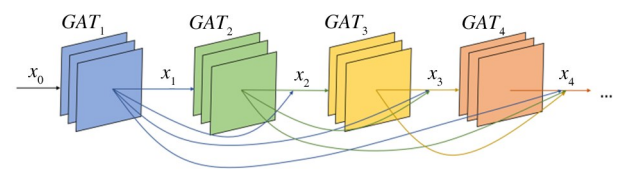


图 5 密集连接的多头图注意力网络

2.2 时间特征建模

获取时间依赖关系是交通预测中的另一个关键问题,目前处理时间序列的神经网络模型主要是递归神经网络(RNN)。由于 RNN 在训练过程中容易出现梯度消失和梯度爆炸等问题,对于预测交通流这种长序列数据存在较大的局限性。LSTM 和 GRU

模型是递归神经网络的一种变体,可以有效解决长序列带来的梯度消失,同时利用门控机制来保存序列的长期信息。与 LSTM 相比,GRU 只使用了 2 个门控开关,理论参数量是 LSTM 的 3/4,训练时更容易收敛,对于精细化的交通预测任务效果更好^[20]。

本文选用 GRU 模型来获取交通数据的时间依赖关系,其内部结构如图 6 所示,其中 h_{t-1} 表示 $t-1$ 时刻的隐藏状态,其中也包括历史序列信息; e_t 为 t 时刻的交通信息; r_t 是重置门,控制前一时刻隐藏信息的忽略程度,决定前一状态写入候选集 \tilde{h}_t 的信息量; z_t 是更新门,控制前一时刻状态信息进入当前状态的程,决定如何将新的输入信息和记忆信息相结合,更新门的值越大表示前一时刻信息量被写入得越多; \tilde{h}_t 为 t 时刻存储的记忆内容,使用 \tanh 激活函数保证 \tilde{h}_t 输出向量元素均在 $-1 \sim 1$ 之间; h_t 是当前 t 时刻的输出状态,由重置门和更新门计算后得到。通过门控开关使得 GRU 在获取当前时刻交通信息的同时,仍保留了历史交通信息的变化趋势,具有捕获时间依赖性的能力。GRU 前向传播公式为

$$z_t = \sigma(\mathbf{W}_z \cdot [h_{t-1}, x_t]) \quad (5)$$

$$r_t = \sigma(\mathbf{W}_r \cdot [h_{t-1}, x_t]) \quad (6)$$

$$\tilde{h}_t = \tanh(\mathbf{W} \cdot [r_t \odot h_{t-1}, x_t]) \quad (7)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (8)$$

其中, $[\]$ 表示 2 个向量相连, \cdot 表示矩阵乘法, \odot 表示哈达玛积。

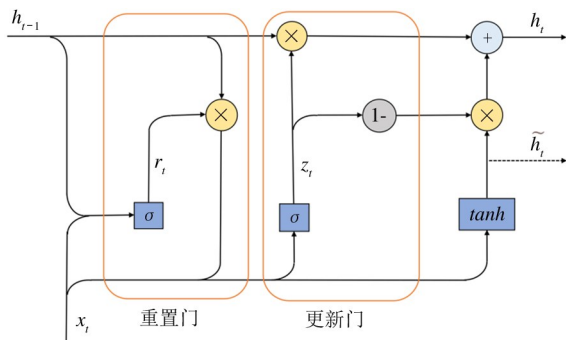


图 6 门控循环单元内部结构

2.3 基于 GRU 的密集连接图注意力神经网络

为了充分捕获路网的时空特征,本文提出了一

种新的基于 GRU 的密集连接图注意力神经网络 (dense graph attention network of GRU, DG-GRU), 模型如图 7 所示。首先利用 2.2 节的 GRU 模型构建循环神经网络来捕捉路网的时间特征, 建立时间序列的长期依赖关系, 其次利用 GAT 网络组成密集块 (Dense-GAT) 来获取路网空间特征, 重复多次利用先前层的输入输出信息, 层与层之间使用向量拼接 (concat), 深度挖掘节点特征。

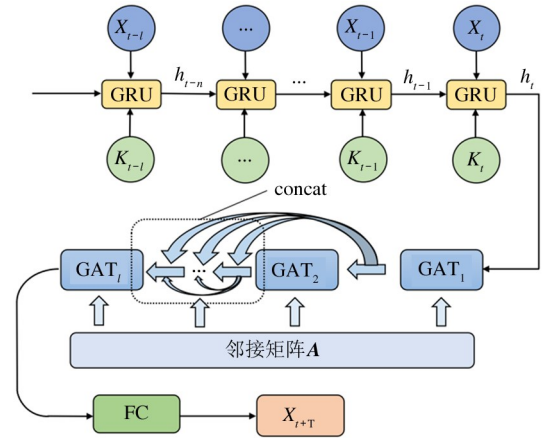


图 7 基于 GRU 的密集连接图注意力网络

模型输入数据除了原始时序信息外, 每个时间步还加入了交通事件矩阵 \mathbf{K} 作为先验知识, 两者在时间维度上进行连接, 帮助模型学习路网潜在的干扰因素。前向传播过程如下, 将速度时序 $X = (X_{t-1}, X_{t-1+1}, \dots, X_t)$ 和交通事件矩阵 $\mathbf{K} = [K_{t-1}, K_{t-1+1}, \dots, K_t]$ 进行向量拼接, 输入维度为 $R^{2l \times N}$, 在 t 时刻门控循环单元输出包含时间维度的隐藏层 h_t 后, 经过 Dense-GAT 网络提取空间特征, 最后通过全连接层预测出未来 T 个时间步的交通速度。如式(9)所示。

$$[X_{t+1}, X_{t+2}, \dots, X_{t+T}] = f(G, X | \mathbf{K}) \quad (9)$$

其中, G 为道路网络, 用无向图 $G = (V, E)$ 来描述路网的拓扑结构, 将路段作为节点; V 代表一组道路节点, $V = (v_1, v_2, \dots, v_N)$; N 是节点数量, E 是节点与节点连接的边, 用邻接矩阵 \mathbf{A} 来描述节点之间的关系, 其中 $\mathbf{A} \in R^{N \times N}$ 。对于相邻节点记作“1”, 其他记为“0”, 通过该方式把整个城市路网的拓扑结构映射到无向图中, 对其进行空间建模, 每个节点 v 都包含路段的交通流时序数据和空间状态信息。

3 案例应用与结果分析

3.1 实验环境和数据

实验环境为 Pytorch, 采用 GPU 加速的训练方法, 选取了杭州市申花路周围路段的交通速度数据集来评估 DG-GRU 模型的预测性能。数据集共由 3 个部分组成: (1) 309×309 的邻接矩阵 \mathbf{A} , 用来描述路段之间的空间关系; (2) 记录各个路段的速度序列 $X = (X_1, X_2, \dots, X_T), X \in R^{T \times N}$, 记录间隔为 15 min; (3) 描述各路段交通事件发生状况的矩阵 $\mathbf{K} = (K_1, K_2, \dots, K_T), \mathbf{K} \in R^{T \times N}$ 。同时将所有数据进行归一化操作以提高模型的训练速度, 其中 80% 的数据作为训练集, 20% 的数据作为测试集。实验预测了未来 15 min、30 min 和 45 min 3 个时间段的交通速度。

3.2 评价指标

为了比较直观地评估 DG-GRU 的预测性能, 实验采用 5 个主流的评价指标来描述真实交通速度 Y_t 和预测速度 \hat{Y}_t 之间的差距。包括以下 5 个指标:

(1) 均方根误差 (RMSE)

$$RMSE = \sqrt{(Y_t - \hat{Y}_t)^2} \quad (10)$$

(2) 平均绝对误差 (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_t - \hat{Y}_t| \quad (11)$$

(3) 精确度 (Accuracy)

$$Accuracy = 1 - \frac{\|Y - \hat{Y}_t\|_F}{\|Y\|_F} \quad (12)$$

(4) 可决系数 (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{i=1}^n (Y_t - \bar{Y}_t)^2} \quad (13)$$

(5) 解释方差分数 (Var)

$$Var = 1 - \frac{Var\{Y - \hat{Y}_t\}}{Var\{Y\}} \quad (14)$$

其中, RMSE 和 MAE 代表 Y_t 和 \hat{Y}_t 之间的真实误差, 数值越小, 代表模型预测能力越好; Accuracy 数值越大, 代表预测精度越高; R^2 和 Var 是用来衡量预测值代表真实值的能力, 数值越大, 说明模型预测能力越高。

3.3 模型参数设计

DG-GRU 模型超参数主要包括学习率、批量大

小、训练次数、GAT 多头注意力头数、Dense-GAT 层数以及 GRU 单元隐藏层的数量。为加快模型收敛速度, 将学习率设为 0.001, 批量大小设为 32, 训练次数为 3 000 次。模型选用 Xavier 初始化权重等参数, 使用 Adam 优化器更新参数。模型训练过程中, 目的是减少预测值和真实交通速度之间的误差。本文模型选用的损失函数为

$$loss = ||Y_t - \hat{Y}_t|| + \lambda L_{reg} \quad (15)$$

其中, Y_t 表示实际的交通速度, \hat{Y}_t 表示模型预测的速度。式(15)前一项用于减少两者之间误差, 后面一项是 L_2 正则化, λ 为可训练参数。通过观察损失函数的变化趋势来判断模型预测性能, 损失函数数值越小, 模型预测性能越好。

3.4 实验结果与模型性能分析

在实验过程中, 超参数对模型预测性能具有较大影响, 其中包括 GAT 多头注意力头数、Dense-GAT 层数以及 GRU 单元隐藏层的数量。为获得最佳预测效果, 通过对三者进行交叉验证来进一步分析实验结果。当设定 Dense-GAT 层数 P 后, 改变多头注意力头数和隐藏层数量来观察模型预测精度变化趋势; 选择模型参数的最优值, GAT 多头注意力的头数分别设为 [2, 3, 4, 5, 6], 隐藏单元数量的候选集设为 [8, 16, 32, 64, 100]。如图 8 所示, 当 GAT 多头注意力的头数和隐藏单元数量不断增加时, DG-GRU 模型预测性能先提高后降低。分析发现模型对 GRU 隐藏单元数量较为敏感, 在与注意力头数交叉实验的过程中, 当隐藏层数量设为 64、多头注意力头数设为 3 时, RMSE 值最低, 即模型预测效果最好。不同的 GAT 密集连接层数对预测精度也会产生影响。交叉实验结果显示, 当 Dense-GAT 层数设为 3 时 DG-GRU 模型达到最佳预测性能, 而不断增加层数会导致参数量变多, 计算量增加导致过拟合现象, 效果变差。

为验证本文的思路, 选取 5 个基线模型来进行实验对比, 包括 ARIMA、GCN、GAT、Dense-GAT、GRU, 其中 Dense-GAT 模型设置层数为 3 层时达到最佳性能。表 1、2 和 3 分别显示了各模型在未来 15 min、30 min 和 45 min 预测的性能指标, 其中 * 代表数值太小, 忽略不计。从表格中可以明显看出, 对

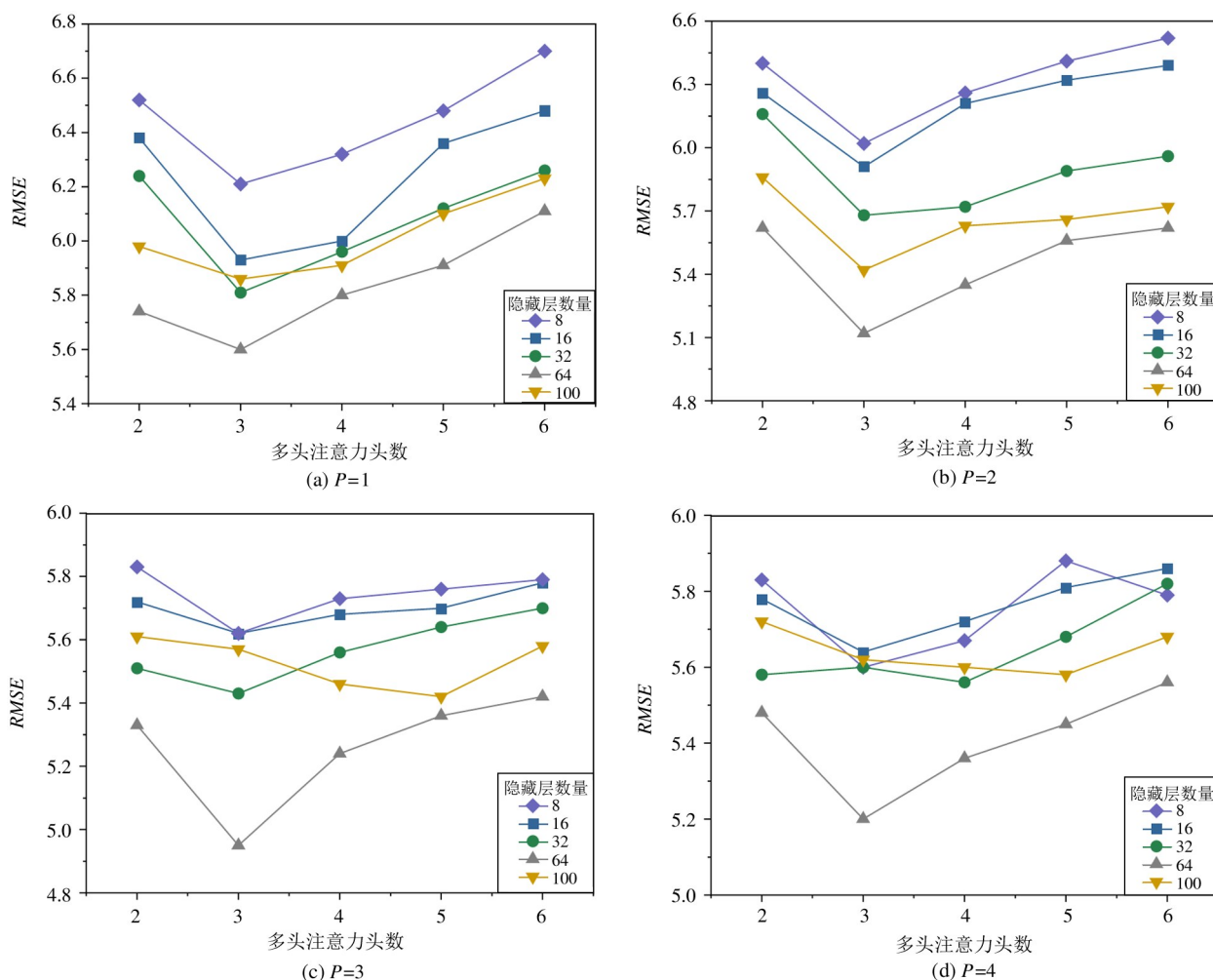


图 8 多头注意力和隐藏层的交叉验证

表 1 DG-GRU 和基线模型对未来 15 min 交通状况的预测结果

时间	性能指标	ARIMA	GCN	GAT	Dense-GAT	GRU	DG-GRU
15 min	RMSE	10.910 0	9.560 0	9.410 0	6.900 0	5.770 0	4.950 0
	MAE	8.620 0	7.640 0	7.520 0	5.310 0	4.130 0	3.700 0
	Accuracy	0.462 2	0.641 9	0.648 0	0.742 6	0.786 2	0.816 4
	R2	*	0.424 0	0.451 0	0.509 1	0.648 3	0.729 2
	Var	*	0.430 0	0.453 0	0.533 1	0.668 0	0.734 5

表 2 DG-GRU 和基线模型对未来 30 min 交通状况的预测结果

时间	性能指标	ARIMA	GCN	GAT	Dense-GAT	GRU	DG-GRU
30 min	RMSE	11.090 0	9.560 0	9.430 0	7.000 0	5.970 0	5.720 0
	MAE	8.750 0	7.650 0	7.600 0	5.390 0	4.460 0	4.370 0
	Accuracy	0.453 1	0.641 7	0.646 1	0.739 2	0.779 3	0.798 2
	R2	*	0.393 0	0.443 0	0.477 0	0.620 1	0.644 9
	Var	*	0.398 0	0.446 0	0.482 6	0.635 1	0.653 6

表3 DG-GRU和基线模型对未来45min交通状况的预测结果

时间	性能指标	ARIMA	GCN	GAT	Dense-GAT	GRU	DG-GRU
45 min	RMSE	11.210 0	9.570 0	9.480 0	7.410 0	7.330 0	6.020 0
	MAE	8.960 0	7.650 0	7.620 0	5.920 0	5.330 0	4.580 0
	Accuracy	0.451 2	0.641 5	0.642 2	0.721 9	0.728 5	0.787 4
	R2	*	0.381 0	0.441 0	0.506 5	0.462 6	0.602 0
	Var	*	0.385 0	0.443 0	0.601 3	0.492 5	0.620 2

于3个不同的预测视野,DG-GRU在所有评估指标中都优于其他模型,证明DG-GRU在交通预测任务中的有效性。

对实验结果分析发现以下几点。

(1) 空间特征提取能力。与只考虑时间依赖的自回归模型ARIMA以及GRU网络进行对比,对于15min的交通预测任务,DG-GRU预测均方根误差分别减少了54.6%和14.2%,预测精度分别提高了43.4%和14.7%,说明该模型具有从交通数据中提取潜在空间特征的能力,建立对路网的空间依赖性,比单一化利用路网时间信息的模型更加有效。而ARIMA是通过计算每个节点的误差和平均来预测的,很难长期处理波动较大的时序,导致预测精度远低于GRU。将Dense-GAT和GAT模型对比可以发现,在3个预测范围中,Dense-GAT模型性能均优于GAT,证明密集连接网络可以加深图卷积模型的空间挖掘,增加节点表征能力。

(2) 时间特征提取能力。为了验证DG-GRU模型是否有能力从交通数据中描绘时间特征,将其

与空间建模的GCN、GAT、Dense-GAT模型进行对比,在15min预测范围下,DG-GRU预测的RMSE分别减少了48%、47.4%和28.3%,预测精度分别提高了30.1%、20.6%和9.3%,说明该模型可以很好地捕捉时序间的非线性特征并建立路网的时间依赖性。

(3) 长期预测能力。对DG-GRU预测未来3个时间段的RMSE和Accuracy进行分析,从图9(a)可以看出,在不同的预测范围下其RMSE和精度的变化幅度小,对外界敏感度低,不易发生震荡;图9(b)显示,在不同的预测范围下,DG-GRU模型的RMSE误差均为最低,预测效果优于其他基线模型。证明该模型既可以用于短期预测任务,也可以进行长期的预测。

(4) 交通事件矩阵消融实验。为了验证加入交通事件矩阵的有效性,本文对其进行了消融实验,实验结果如图10所示,其中DG-GRU-K表示加入交通事件矩阵的密集连接网络。结果表明,在加入矩阵K后,DG-GRU模型的RMSE下降了3.7%,预测精

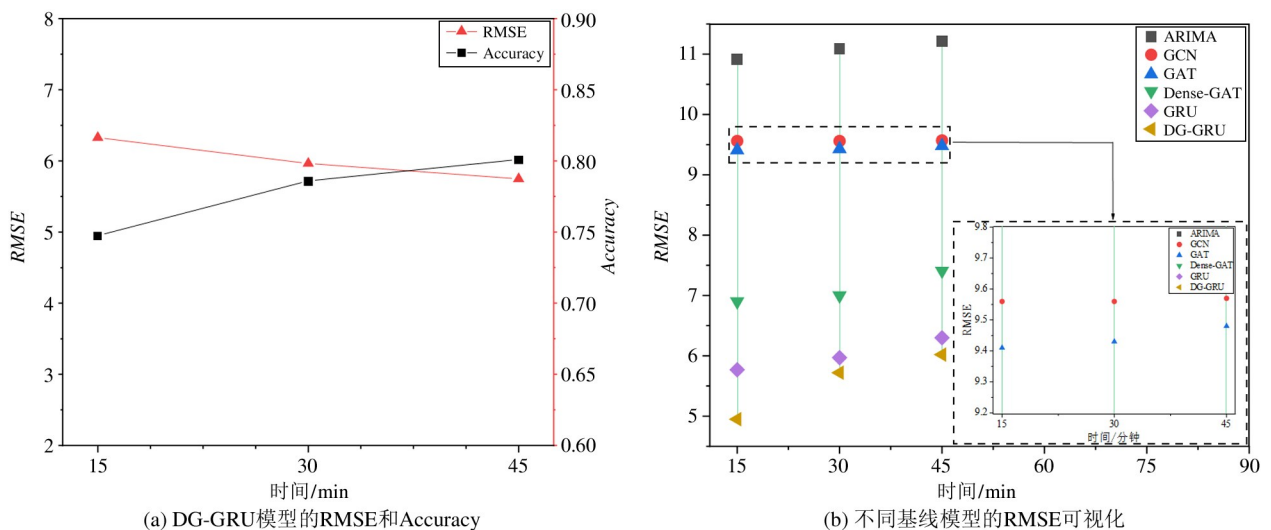


图9 DG-GRU与基线模型预测性能对比分析

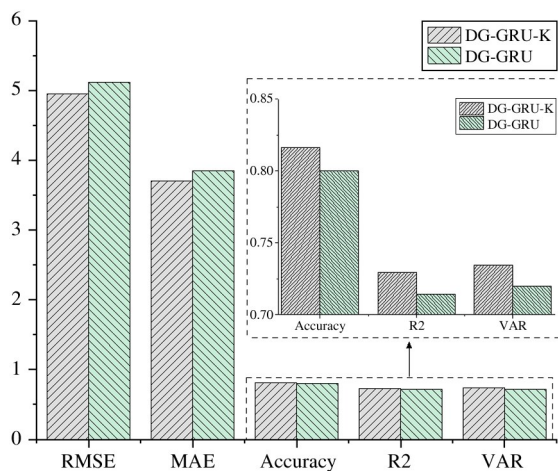


图 10 交通事件矩阵的消融实验

度提升了 2.5%, 证明对外部环境因素建模可以有效增加路段节点的特征多样性, 增强信息属性, 提高模型预测精度。

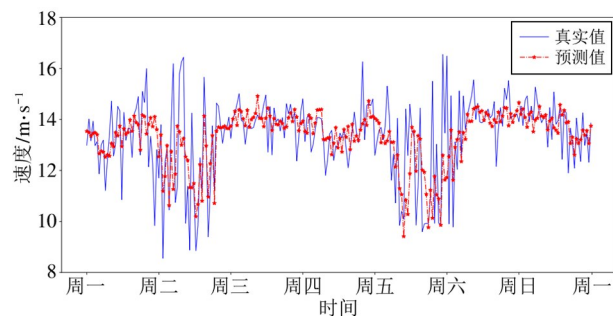
此外, 针对基线模型, 本文分析了在训练过程中各模型参数量、训练时间(s/epoch)以及达到收敛状态的迭代次数, 如表 4 所示。结果表明, GRU 模型相对于空间建模的 GCN、GAT、Dense-GAT 网络参数量更多, 带来的效果也更好, 说明路网数据本身是时间序列, 对时间维度特征具有较大的依赖。本文模型在时空维度上建模, 参数量最多, 带来的训练复杂度相对较大, 但是通过较少的迭代次数就能达到收敛状态, 获得最佳预测效果, 同时在一定程度上可以避免模型产生过拟合。

表 4 模型计算复杂度分析

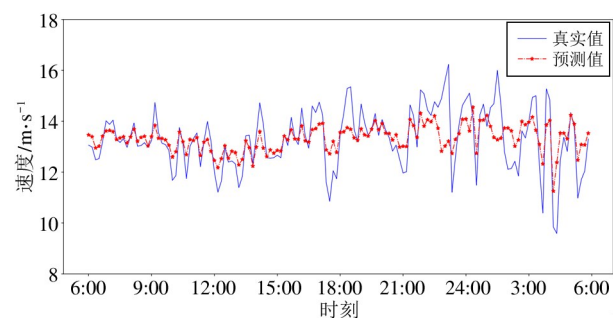
	GCN	GAT	Dense-GAT	GRU	DG-GRU
参数量	139	222	2 370	12 672	18 544
训练时间 (s/epoch)	3.4	3.8	6.7	18.2	22.3
迭代次数	523	516	421	216	156
精确度	0.641 9	0.648 0	0.742 6	0.786 2	0.816 4

为了更直观地检验本文提出模型的预测性能, 从交通路网中随机选取一条路段, 对连续一周的预测结果进行可视化。如图 11(a)、12(a)和 13(a)所示, 其分别代表了未来 15 min、30 min 和 45 min 交通速度预测的可视化结果。为了观察局部预测细节, 本文将其中一天的预测结果进行可视化, 结果如

图 11(b)、12(b)和 13(b)所示。连续一周的预测结果表明, DG-GRU 模型预测的结果在整体趋势上具有较好的拟合度, 模型能捕捉到真实速度的尖锐特征, 不会出现过度平滑的表现。

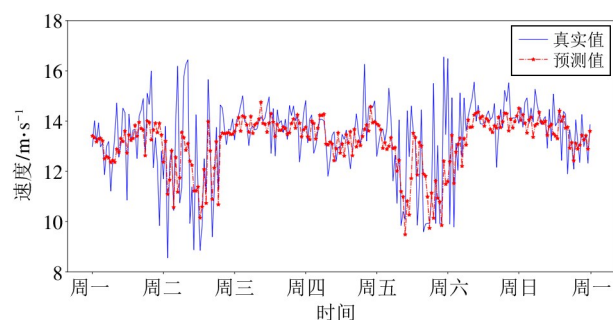


(a) 连续一周预测可视化结果

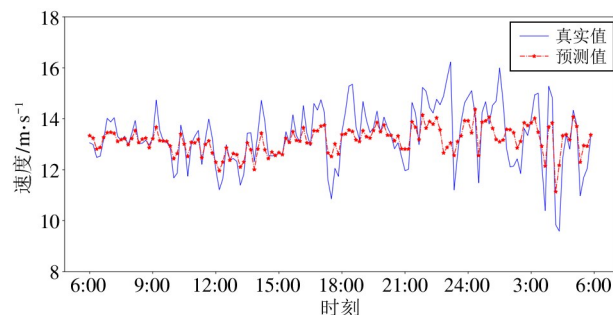


(b) 连续一天预测可视化结果

图 11 未来 15 min 模型预测可视化结果



(a) 连续一周预测可视化结果



(b) 连续一天预测可视化结果

图 12 未来 30 min 模型预测可视化结果

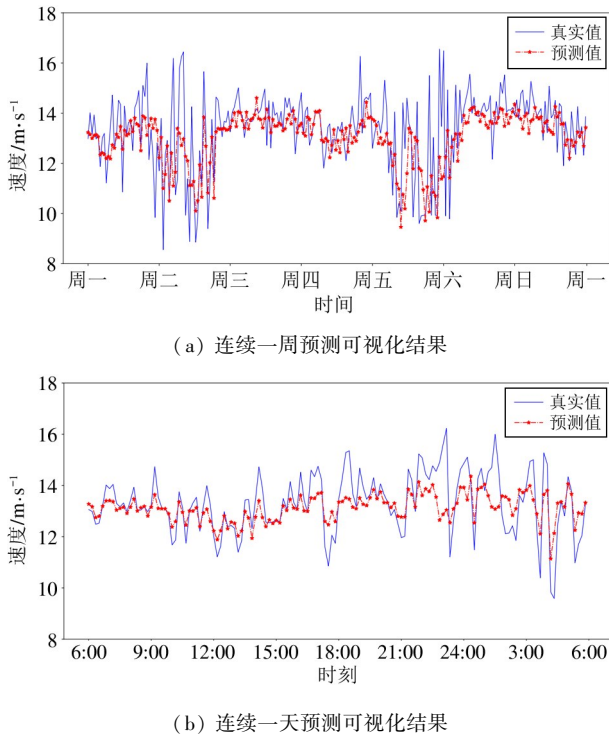


图 13 未来 45 min 模型预测可视化结果

将一天的可视化结果提取出来,可以更加明显地看出 DG-GRU 模型的预测效果,其短期预测曲线更接近真实曲线,说明该模型在长期预测中丢失部分信息的同时,能较好地捕捉短期依赖关系;对于捕捉变速趋势的转折点,DG-GRU 模型在高、低峰处有偏差,原因可能是交通状态的突然变化不仅受到本研究中使用交通事件的影响,还受到其他外在因素的综合影响。另外,该模型不会因为进行长期预测产生严重的震荡而脱离预测目标,说明 DG-GRU 具有长期预测能力,对外界的敏感度较低。

综上,DG-GRU 模型在复杂的交通流数据建模上具有优势,尤其在捕捉时空特征上优于其他基线模型。通过学习路网数据潜在的非线性关系进而较好地拟合其未来的变化趋势,使模型可以达到更好的预测性能。

4 结论

(1) 通过对城市交通网络数据的分析,构建了基于 GRU 的密集连接的图注意力网络。一方面,利用 GRU 来捕捉节点属性的动态变化,获得时间依赖

性;另一方面,利用 Dense-GAT 来捕捉路网的拓扑结构,获得空间依赖性。同时加入交通事件作为先验知识,使模型能够捕捉时序数据的尖锐变化,实现了城市交通道路速度预测的任务。

(2) 本文搭建的模型主要基于时空特征方面的考虑,同时加入交通事件的先验知识是对考虑外部客观因素的初步尝试。从实验结果可以发现,相较于考虑单一化特征的 GCN 模型和 GRU 模型,DG-GRU 模型在捕捉交通路网时空特征和加入客观先验知识后,其预测性能得到了较大的提升,说明对路网数据和外界信息的充分挖掘是交通预测任务的重点方向。后续研究将着重考虑更多的外部因素,加强交通预测任务对象特征的整合,进一步提高模型预测的准确性和灵活性。

(3) 实验结果表明,DG-GRU 模型能够很好地捕捉路网的时空信息,在预测精度上明显优于其他基线模型。对比传统模型 ARIMA,DG-GRU 模型预测性能在 3 个预测范围内分别提高了 43.3%、43.2% 和 42.7%。对比主流神经网络模型 GCN 和 GRU,DG-GRU 模型预测性能在 3 个预测范围内分别提高了 21.4%、19.6%、18.5% 和 3.7%、2.4%、7.5%。DG-GRU 预测模型对外界敏感度较低,具有长期预测能力。

对城市交通网络的精准预测有助于改善路段拥堵状况,提高道路的通行能力,同时对城市路段规划和区域划分起着重要作用。此外,DG-GRU 模型不仅限于交通预测,还可用于其他时空预测任务。

参考文献

- [1] ABDOLLAHI M, KHALEGHI T, YANG K. An integrated feature learning approach using deep learning for travel time prediction [J]. Expert Systems with Applications, 2020,139:1-11.
- [2] XU X Y, LIU J, LI H Y, et al. Analysis of subway station capacity with the use of queueing theory [J]. Transportation Research Part C: Emerging Technologies, 2014, 38:28-43.
- [3] QI W, LI L, HU J, et al. Traffic velocity distributions for different spacings [J]. Journal of Tsinghua University, 2011,51(3):309-312.
- [4] 许菲菲,何兆成,沙志仁. 交通管理措施对路网宏观基本图的影响分析 [J]. 交通运输系统工程与信息,

- 2013,13(2):185-190.
- [5] SHAN Z, ZHAO D, XIA Y. Urban road traffic speed estimation for missing probe vehicle data based on multiple linear regression model [C] // The 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). Busan, Korea; IEEE, 2013:118-123.
- [6] SHEN G J, WANG X H, KONG X J. Short-term traffic volume intelligent hybrid forecasting model and its application [J]. Systems Engineering-Theory and Practice, 2011,31(3):561-568.
- [7] SMITH B L, DEMETSKY M J. Traffic flow forecasting: comparison of modeling approaches [J]. Journal of Transportation Engineering, 1997,123(4):261-266.
- [8] SUN H, ZHANG C, RAN B. Interval prediction for traffic time series using local linear predictor [C] // The 7th International IEEE Conference on IEEE. Washington, USA; IEEE, 2004:410-415.
- [9] ZHANG X L, GUO G, H E, HUA-PU L U. Short-term traffic flow forecasting based on k -nearest neighbors non-parametric regression [J]. Journal of Systems Engineering, 2009,24(2):178-183.
- [10] SUN S, ZHANG C, YU G. A Bayesian network approach to traffic flow forecasting [J]. IEEE Transactions on Intelligent Transportation Systems, 2006,7(1):124-132.
- [11] HU W, YAN L, LIU K, et al. A short-term traffic flow forecasting method based on the hybrid PSO-SVR [J]. Neural Processing Letters, 2016,43(1):155-172.
- [12] 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.
- [13] JIA Y, WU J, DU Y. Traffic speed prediction using deep learning method [C] // IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). Rio de Janeiro, Brazil; IEEE, 2016:1217-1222.
- [14] LV Y, DUAN Y, KANG W, et al. Traffic flow prediction with big data: a deep learning approach [J]. IEEE Transactions on Intelligent Transportation Systems, 2015,16(2):865-873.
- [15] TIAN Y, LI P. Predicting short-term traffic flow by long short-term memory recurrent neural network [C] // 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). Guadalajara, Mexico; IEEE, 2015:153-158.
- [16] 季学武, 费聪, 何祥坤, 等. 基于 LSTM 网络的驾驶意图识别及车辆轨迹预测 [J]. 中国公路学报, 2019,32(6):34-42.
- [17] RUI F, ZUO Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction [C] // The 31st Youth Academic Annual Conference of Chinese Association of Automation. Wuhan, China; IEEE, 2016:324-328.
- [18] ZHANG D, KABUKA M R. Combining weather condition data to predict traffic flow: a GRU based deep learning approach [C] // IEEE International Conference on Dependable, Autonomic and Secure Computing; IEEE International Conference on Big Data Intelligence and Computing; IEEE International Conference on Pervasive Intelligence and Computing; IEEE Cyber Science and Technology Congress. Orlando, USA; IEEE, 2017:1216-1219.
- [19] MA X, TAO Z, WANG Y, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data [J]. Transportation Research Part C: Emerging Technologies, 2015,54:187-197.
- [20] RUI F, ZUO Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction [C] // The 31st Youth Academic Annual Conference of Chinese Association of Automation. Wuhan, China; IEEE, 2016:324-328.
- [21] LIU Y, ZHENG H, FENG X, et al. Short-term traffic flow prediction with Conv-LSTM [C] // The 9th International Conference on Wireless Communications and Signal Processing. Nanjing, China; IEEE, 2017:1-6.
- [22] WU Y, TAN H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework [EB/OL]. (2016-12-03) [2022-12-26]. <https://arxiv.org/pdf/1612.01022.pdf>.
- [23] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks [EB/OL]. (2016-09-09) [2022-12-26]. <https://arxiv.org/pdf/1609.02907.pdf>.
- [24] ZHOU J, CUI G, HU S, et al. Graph neural networks: a review of methods and applications [J]. AI Open, 2020,1:57-81.
- [25] ZHAO L, SONG Y, ZHANG C, et al. T-GCN: a temporal graph convolutional network for traffic prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2019,21(9):3848-3858.
- [26] LI Z, XIONG G, CHEN Y, et al. A hybrid deep learning approach with GCN and LSTM for traffic flow prediction [C] // 2019 IEEE Intelligent Transportation Systems Conference (ITSC). Auckland, New Zealand; IEEE, 2019:1929-1933.
- [27] LIN C Y, SU H T, TUNG S L, et al. Multivariate and propagation graph attention network for spatial-temporal prediction with outdoor cellular traffic [C] // The 30th ACM International Conference on Information and Knowledge Management. Birmingham, USA; Association for Computing Machinery, 2021:3248-3252.
- [28] AOSONG F, LEANDROS T. Adaptive graph spatial-temporal transformer network for traffic forecasting [C] // The 31st ACM International Conference on Information and Knowledge Management. New York, USA; IEEE, 2022:

- 3933-3937.
- [29] YAN Y, HASHEMI M, SWERSKY K, et al. Two sides of the same coin: heterophily and oversmoothing in graph convolutional neural networks[EB/OL]. (2021-02-12) [2022-12-26]. <https://arxiv.org/pdf/2102.06462.pdf>.
- [30] ALON U, YAHAV E. On the bottleneck of graph neural networks and its practical implications[EB/OL]. (2020-06-09) [2022-12-26]. <https://arxiv.org/pdf/2006.05205.pdf>.
- [31] LI G, M MÜLLER, THABET A, et al. Deep GCNs: can GCNs go as deep as CNNs? [C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea:IEEE, 2020:9266-9275.
- [32] LIAO B, ZHANG J, WU C, et al. Deep sequence learning with auxiliary information for traffic prediction[C]//The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. London, UK: Association for Computing Machinery, 2018:537-546.
- [33] VAN LINT J W C, HOOGENDOOM S P, VAN ZUYLEN H J. Freeway travel time prediction with state-space neural networks: modeling state-space dynamics with recurrent neural networks[J]. Transportation Research Record Journal of the Transportation Research Board, 2002, 1811(1):347-369.
- [34] ZHU J, WANG Q, TAO C, et al. AST-GCN: attribute-augmented spatiotemporal graph convolutional network for traffic forecasting[J]. IEEE Access, 2021, 9: 35973 - 35983.
- [35] WANG Z, JIANG R, XUE H, et al. Event-aware multi-modal mobility nowcasting [EB/OL]. (2021-12-14) [2022-12-26]. <https://arxiv.org/pdf/2112.08443.pdf>.
- [36] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA:IEEE, 2017:4700-4708.
- [37] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA:IEEE, 2016:70-778.

Urban traffic prediction based on densely connected spatial-temporal graph attention network of GRU

GUO Haifeng*, XU Hongwei**, ZHOU Zisheng**

(* Institute of Cybersecurity, Zhejiang University of Technology, Hangzhou 310014)

(** College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

Abstract

Due to the complexity topology of urban traffic network, the real-time change of traffic flow and external environmental factors, there are huge difficulties in traffic prediction. In view of the inadequacy of existing methods in mining the spatio-temporal features of road network and the insufficient consideration of external factors, a spatial-temporal network of dense graph attention network based on gated recurrent unit (GRU) (DG-GRU) is proposed. The function of gated recurrent unit is used to capture the dynamic changes of road network data. Densely connected graph attention network (GAT) is used to extract the complex spatial structure characteristics of the road network. They can establish the dependence of urban traffic network data on time and space. Considering the influence of external factors, the one-hot encoding is used to model the traffic events that occur in urban sections to enhance the information attributes of transportation network. Taking Shenhua Road and its surrounding sections in Hangzhou as an example to verify the predictive ability and feasibility of the network. The experimental results illustrate that the prediction accuracy of the method is up to 81.64%. Compared with traditional mathematical model and mainstream neural network model, the prediction accuracy of DG-GRU is 35.42% higher than that of ARIMA. Compared with graph attention network (GAT) and GRU neural networks, its prediction accuracy is improved by 17.45% and 3.02%, respectively. Experimental results show that the model in this paper can adapt to complex traffic flow and carry out long-term traffic forecasting tasks. Meanwhile, it can enhance traffic management ability and reduce the costs traffic congestion.

Key words: traffic prediction, spatio-temporal features, neural network, gated recurrent unit (GRU), densely connected, graph attention network (GAT)