

基于图注意力机制的交通流预测模型

周安众, 谢丁峰

(湖南工业职业技术学院信息工程学院, 湖南 长沙 410208)

✉ sprite4@163.com; coolboyxie@163.com



摘要:针对现有交通流预测模型在预测精度上的不足,提出一种基于注意力机制的图模型。首先,利用多头注意力机制在交通图中编码高阶邻域结构,提取交通网络中的高阶空间特征。然后,嵌入长距离时间结构注意力机制提取长期性的历史周期信息。模型采用注意力机制替代传统的局部卷积核结构,可以有效提取长距离时空依赖关系。在 METR-LA(洛杉矶路网)、PeMS-BAY(加州湾区路网)、PeMS-S(加州小型路网)三个真实的交通数据集上进行实验证明,模型在预测未来 60 min 的交通流精度上较传统深度学习方法, RMSE(均方根误差)平均降低 3.1%、3.9%和 1.8%,表明所提模型的长时间预测能力优势明显。

关键词:注意力机制;图模型;时空依赖;交通流预测

中图分类号:TP391 **文献标志码:**A

A Traffic Flow Prediction Model Based on Graph Attention Mechanism

ZHOU Anzhong, XIE Dingfeng

(Department of Information Engineering, Hunan Industry Polytechnic, Changsha 410208, China)

✉ sprite4@163.com; coolboyxie@163.com

Abstract: This paper proposes a graph model based on attention mechanism to address the accuracy issues of existing traffic flow prediction models. Firstly, the multi-head attention mechanism is used to encode high-order neighborhood structures in the traffic map and extract high-order spatial features in the traffic network. Then, a long-distance time structure attention mechanism is embedded to extract long-term historical cycle information. The model uses attention mechanism to replace traditional local convolutional kernel structure, which can effectively extract long-distance spatio-temporal dependencies. Experiments are carried out on three real traffic data sets, METR-LA (Road Network of Los Angeles), PeMS-BAY (Bay Area Road Network of California), and PeMS-S (Small Road Network of California). Experimental results show that the RMSE (Root Mean Square Error) of the traffic flow prediction accuracy of the proposed model in the next 60 minutes is 3.1%, 3.9%, and 1.8% lower on average than that of traditional deep learning method, which indicates that the proposed model has obvious advantages in long-term prediction ability.

Key words: attention mechanism; graph model; spatio-temporal dependencies; traffic flow prediction

0 引言(Introduction)

近年来,研究人员为了消除交通流预测中的时空因素的影响,采用了各种深度学习模型^[1]。其中,图卷积神经网络(Graph Convolutional Network, GCN)用于建模交通网络这类非欧结构数据已被证明是有效的^[2]。在时间建模方面,递归神经网络(Recurrent Neural Network, RNN)是短时交通流预测

(30 min 以内)中的常用方法^[3]。然而,对更长时间(45 min 以上)的交通流预测效果仍不理想^[4],具体原因如下。

(1)路网中节点信息不仅随相邻节点变化,还间接受到更远距离节点(高阶节点)的影响。传统深度学习模型以局部特征为中心,需要叠加多层结构提取全局特征,训练时容易出现梯度弥散问题,不能有效提取长距离空间依赖关系。

(2)交通流量发生周期性的变化,如日高峰时间或其他周期事件,有必要引入长期的历史时间依赖关系捕捉周期特征。长短期记忆网络和门控循环单元常用来提取长距离时间依赖关系,但在迭代训练时可能会导致误差累积^[5-7]。

综上所述,本文提出一种结构化图注意力网络模型(Structural Graph Attention Network, SGAN)。一方面,SGAN采用注意力机制替代了传统深度学习架构中的局部卷积核,使其在长距离约束情况下能高效提取时空特征。另一方面,SGAN在计算注意力系数时利用节点的结构信息,不仅考虑了节点之间的特征相关性,还考虑了节点的连接方式。

1 问题描述(Problem statement)

对于交通流预测任务,建立图 $G=(V, \epsilon, A)$,其中 V 是图 G 中的节点集, ϵ 是边集,反映节点之间的连通性。 $A \in \mathbb{R}^{N \times N}$ 表示 G 的邻接矩阵。同时,建立矩阵 $X^{(t)} \in \mathbb{R}^{N \times d}$ 作为 t 时刻的输入特征矩阵,其中 N 表示节点数, d 表示特征的维度。给定图 G 中 T 个时间步的输入特征序列($X=[x^{(t-T+1)}, \dots, x^{(t)}]$),通过学习映射函数 f 预测未来 T' 个时间步的交通流($Y=[x^{(t+1)}, \dots, x^{(t+T')}]$)。图 G 提供了图的空间信息, T 则提供了图中交通流的时间信息,学习函数 f 将输入的 T 个历史时间步数据映射到未来的 T' 个时间步上。

2 模型架构(Model architecture)

2.1 整体架构

本文提出的SGAN整体架构如图1所示,由多个时空注意力模块组成,每个模块包含一个空间依赖层(Spatial Dependencies Layer, SDL)和一个时间依赖层(Temporal Dependencies Layer, TDL),并加入了残差结构^[8]。历史周期数据经过特征转换后作为时空注意力模块的输入,并通过多个SDL和TDL生成最终输出。SDL主要关注节点在图上的结构依赖关系,而TDL则主要关注节点在时间步上的顺序依赖关系。

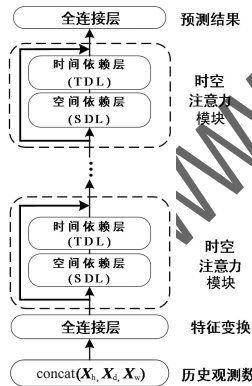


图1 SGAN模型的整体架构图

Fig. 1 The overall architecture of the SGAN model

一般来说,输入的 T 个历史时间步与待预测的 T' 个时间步相邻,为最近一个小时内的历史观测数据。此外,每日和每周的长期历史观测数据也能帮助预测未来交通流的周期变化^[9],日数据设为待预测时间前一天的相同时间步,周数据设为待预测时间前一周的相同时间步,则最近的小时数据记为 X_h ,日数据记为 X_d ,周数据记为 X_w ,将三个数据进行连接作为SGAN的最终输入数据:

$$X = \text{concat}(X_h, X_d, X_w) \quad (1)$$

数据经过连接后包含交通流不同的周期模式,有利于模型

提取长期影响下的周期变化特征,然后通过一个全连接层对其特征进行变换,得到一个高维向量,输入到时空注意力模块进行特征提取^[10]。

2.2 空间依赖层

交通网络图中不同位置的节点因不同的连接方式导致其重要性存在差异,SDL中引入图注意力机制自适应地计算每个节点的注意力系数,即节点的重要度^[11]。对于节点 v_i ,对应特征为 x_i ,则其输出特征 z_i 通过公式(2)计算:

$$z_i = \sigma \left(\sum_{j=1}^n \alpha_{ij} (x_j W) \right) \quad (2)$$

本质上,公式(2)是对 v_i 的邻居节点的聚合操作^[12]。 x_j 是邻居节点特征, $W \in \mathbb{R}^{d_x \times d_z}$ 是待学习的参数矩阵, σ 是激活函数, α_{ij} 是 v_i 相对于 v_j 的注意力系数,由以下公式计算:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})} \quad (3)$$

其中, n 是 v_i 的邻居节点数量, e_{ij} 利用公式(2)中的共享参数 W 对特征进行变换,再由单层前馈神经网络函数 a 得到:

$$e_{ij} = a(Wx_i, Wx_j) \quad (4)$$

公式(2)通过聚合 v_i 的邻居节点特征得到 z_i ,没有考虑每个节点的连接方式。如图2所示,假设以 A 为中心节点,接收来自其邻居节点的信息,图3中 D 的不同边数将导致传递到节点 A 的信息强度不同。因此,需要考虑中心节点与其相邻节点之间的拓扑关系,并对拓扑结构进行编码,从而得到图的空间依赖关系。

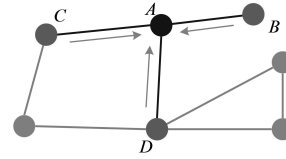


图2 A节点接收邻居节点传递的信息

Fig. 2 Node A receiving information transmitted by neighboring nodes

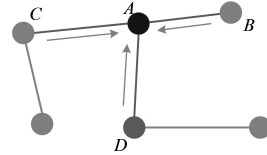


图3 邻居节点的结构不同导致信息传递强度不同

Fig. 3 Different information intensities caused by different structures of neighboring nodes

基于以上问题,在注意力机制中引入结构信息。首先将单位矩阵 I 添加到邻接矩阵 A 中,为每个节点创建一个自循环:

$$\hat{A} = A + I \quad (5)$$

定义转移矩阵 $\hat{P}$,用于表达节点的连接方式:

$$\hat{P} = \hat{D}^{-1} \hat{A} \quad (6)$$

$\hat{D}$ 为对角矩阵,对角元素 $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$,则与 v_i 连接的边数为 $\hat{D}_{ii}$,元素 $\hat{P}_{ij}$ 表示信息从节点 v_i 转移到 v_j 的概率。类似于PageRank算法衡量网页重要性的思想^[13], $\hat{P}_{ij}$ 包含一定的结构信息,新的注意力系数计算公式如下:

$$\hat{e}_{ij} = \hat{P}_{ij} e_{ij} \quad (7)$$

$$\hat{\alpha}_{ij} = \frac{\exp(\hat{e}_{ij})}{\sum_{k=1}^n \hat{e}_{ik}} \quad (8)$$

$\hat{\alpha}_{ij}$ 表示 v_i 的邻居节点 v_j 的相对重要性。如图 4 所示,为了得到更高阶邻居结构信息,需进一步扩展注意力机制以提取更长距离的节点依赖关系。

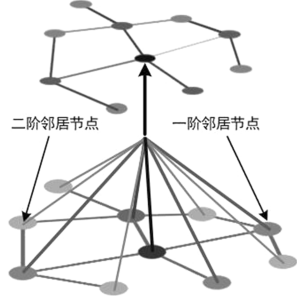


图 4 SDL 提取高阶邻居节点的结构信息

Fig. 4 SDL extracting structural information of high-order neighbor

A 为一阶邻域矩阵,通过计算 A 的 k 次幂 A^k 表示 k 阶邻接矩阵。节点 v_i 的第 k 阶邻居节点为 $N_i^k = \{v_j \in V \mid d(v_i, v_j) = k\}$ 。则 $\hat{P}_i^k$ 提供了节点 v_i 与 k 阶邻居之间的结构性度量:

$$\hat{P}_i^k = \hat{D}_i^{-1} A^k \quad (9)$$

随着 k 的增加,注意力机制将应用于更高阶的节点,即更长距离的节点。结合公式(2),定义 v_i 与第 k 阶邻居的注意力机制如下:

$$z_i^k = \sum_{j \in N_i^k} \hat{\alpha}_{ij}^k x_j \mathbf{w}^k \quad (10)$$

将提取到的不同阶特征连接起来:

$$\tilde{z}_i = [z_i^1, z_i^2, \dots, z_i^k] \quad (11)$$

2.3 空间依赖层

TDL 提取时间维度上的长距离依赖关系,如图 5 所示,节点的输出特征由该节点与历史时间步上的特征融合而成,为了对不同时间步进行区分,需加入关于时间步的位置信息,使模型能区分时间的序列结构。初始化 one-hot 编码向量 $S_i^t \in \mathbb{R}^{d_z}$ 为节点 v_i 在时间步 t 的结构向量, S_i^t 值为 1 的元素指示了 v_i 在时间 t 的位置, TDL 的输入特性为 $\mathbf{h}_i^t = [\tilde{z}_i^t, S_i^t]$, $\tilde{z}_i^t$ 是 SDL 的输出。

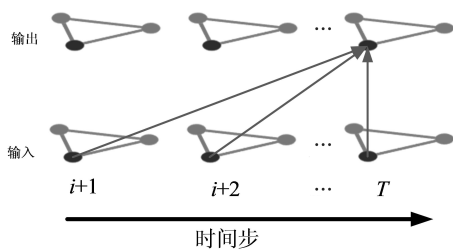


图 5 TDL 提取节点不同时间步之间的依赖关系

Fig. 5 TDL extracting dependencies between nodes at different time steps

为了计算时间注意力系数,每个节点的特征先由参数矩阵 $\mathbf{W}$ 进行变换,分别用 $\mathbf{W}^Q$ 、 $\mathbf{W}^K$ 和 $\mathbf{W}^V \in \mathbb{R}^{d_z \times d_h}$ 表示,这些参数在每一层中分别进行训练,时间注意力系数 β 由以下公式计算:

$$\mathbf{b}_i^{tr} = \frac{(\mathbf{h}_i^t \mathbf{W}^Q)(\mathbf{h}_i^t \mathbf{W}^K)}{\sqrt{d_h}} \quad (12)$$

$$\beta_i^{tr} = \frac{\exp(\mathbf{b}_i^{tr})}{\sum_{k \in \tau} \mathbf{b}_i^k} \quad (13)$$

时间 t 的输出特征 $\mathbf{h}_i^t$ 由 $\mathbf{W}^V$ 变换后的加权求和得到:

$$\mathbf{h}_i^t = \sum_{\tau \in T} \beta_i^{\tau} \mathbf{h}_i^{\tau} \mathbf{W}^V \quad (14)$$

为了提高特征的表达能力,采用多头注意力^[14],由 n 个并行的注意力机制共同输出,并连接为最终的输出:

$$\mathbf{H} = [\mathbf{h}^{(1)}, \dots, \mathbf{h}^{(n)}] \quad (15)$$

模型将上述多个时空模块进行堆叠,形成多层结构,并采用具有 ReLU 激活函数的全连接层作为模型的最终输出:

$$\mathbf{Y} = \text{ReLU}(\phi \mathbf{H} + \mathbf{b}) \in \mathbb{R}^{T'} \quad (16)$$

公式(16)不仅再次提高了特征表示的能力,还将最终的输出变换为真实预测值的维度,其中 T' 是用于预测的时间步长数。

3 实验(Experiment)

3.1 数据集描述

本文在以下三个真实交通数据集上验证 SGAN 的性能。

(1)METR-LA:该数据集包含从美国洛杉矶高速公路上的探测器收集的交通信息,选择了其中 207 个传感器从 2012 年 3 月 1 日到 2012 年 6 月 30 日采集的数据。

(2)PeMS-BAY:该数据集由美国加利福尼亚州运输机构(CalTrans)收集,包含 325 个传感器从 2017 年 1 月到 2017 年 5 月采集的数据。

(3)PeMS-S:该数据集也是由美国加利福尼亚州运输机构(CalTrans)收集,由 228 个传感器从 2012 年 5 月至 2012 年 6 月采集的数据。

以上所有数据集的时间步长的间隔为 5 min,应用 Z-Score 标准化,70%的数据用于训练,20%的数据用于测试,其余 10%的数据用于验证。

3.2 实验设置

本文实验均由谷歌 TensorFlow 深度学习平台实现,运行在 NVIDIA Titan RTX GPU 平台上,使用 RMSprop 优化器,迭代训练 50 次。三个周期数据的时间步长为 $T_h = 12$, $T_d = 12$, $T_w = 12$,历史时间步长总和 $T = 36$ 。预测时间步长 $T' = 12$,通过全连接神经网络将输入的原始周期数据转换为 16 维的特征向量,作为第一个时空注意力模块的输入。整个架构堆叠 2 个时空注意力模块,每个模块采用 8 头注意力机制,高阶参数 k 为 3。

3.3 基准模型与评价指标

本文采用三个度量标准用于评估 SGAN 的性能,包括平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分比误差(MAPE),将 SGAN 与以下代表性的基准模型进行比较。

(1)ARIMA:差分自回归移动平均模型,一种时间序列分析方法^[15]。

(2)FC-LSTM:解决了 LSTM 只考虑时序,没有考虑空间相关性的问题,使用了带有全连接层的 LSTM 模型提取时空依赖关系^[16]。

(3)STGCN:时空图卷积模型,使用 CNN 卷积的方式提取时空依赖关系^[17]。

(4)DCRNN:扩散卷积循环神经网络,将交通流建模为扩散过程,采用扩散卷积提取空间依赖关系,采用 RNN 提取时间

依赖关系^[18]。

(5)Graph WaveNet:一种图神经网络架构,采用扩散图卷积提取空间依赖关系,采用带空洞的卷积核提取时间依赖关系^[19]。

(6)ASTGCN:时空注意力模型,结合注意力机制和卷积核同时捕获时空依赖关系^[9]。

(7)SLCNN:结构学习卷积模型,将传统的 CNN 扩展到图域并学习图的结构信息用于交通流预测^[20]。

(8)NS-SGAN:在 SGAN 的基础上,去除了 SDL 和 TDL 中的转移矩阵和结构化向量,评估 SGAN 中结构化设计的有效性。

3.4 实验结果

表 1 和表 2 分别显示了在 METR-LA 和 PeMS-BAY 数据集上,SGAN 与其他基准模型在 15 min、30 min 和 60 min 时的预测结果比较,从中可以看出,基于图结构的 STGCN、GraphWaveNet 以及本文的 SGAN 大部分数据皆优于 FC-LSTM 等非图结构模型,说明交通网络的结构信息对交通预测至关重要。与其他时空模型相比,采用了注意力机制的 SGAN 大部分数据优于基于卷积的 STGCN 和基于递归的 DCRNN 模型,而 SGAN 采用注意力机制提取长距离依赖关系,在 60 min 的预测中优势更明显,表明注意力机制在长时预测方面更有效。同时,SGAN 的结果明显优于去除结构化信息的 NS-SGAN,证明了注意力机制中的结构信息是重要的。

表 1 SGAN 和基准模型在 METR-LA 上的性能比较

Tab.1 Performance comparison of SGAN and benchmark models on METR-LA

模型	METR-LA(15 min/30 min/60 min)		
	MAE	RMSE	MAPE/%
ARIMA	3.99/5.15/6.90	8.21/10.45/13.23	9.60/12.70/17.40
FC-LSTM	3.44/3.77/4.37	6.30/7.23/8.69	9.60/10.90/13.29
DCRNN	2.77/3.15/3.60	5.38/6.45/7.60	7.34/8.80/10.50
STGCN	2.88/3.47/4.59	5.74/7.24/9.40	7.63/9.57/12.70
Graph WaveNet	2.69/3.07/3.53	5.15/6.26/7.37	6.90/8.37/10.01
SLCNN	2.53/2.88/3.30	5.18/6.15/7.20	6.70/8.00/9.70
NS-SGAN	2.77/3.16/3.58	5.10/5.85/7.07	6.84/8.38/10.03
SGAN	2.73/3.11/3.51	4.92/5.80/6.48	6.73/8.31/9.95

表 2 SGAN 和基准模型在 PeMS-BAY 上的性能比较

Tab.2 Performance comparison of SGAN and benchmark models on PeMS-BAY

模型	PeMS-BAY(15 min/30 min/60 min)		
	MAE	RMSE	MAPE/%
ARIMA	1.62/2.33/3.38	3.30/4.76/6.50	3.50/5.40/8.30
FC-LSTM	2.05/2.20/2.37	4.19/4.55/4.96	4.80/5.20/5.70
DCRNN	1.38/1.74/2.07	2.95/3.97/4.74	2.90/3.90/4.90
STGCN	1.36/1.81/2.49	2.96/4.27/5.69	2.90/4.17/5.79
Graph WaveNet	1.30/1.63/1.95	2.74/3.70/4.52	2.73/3.67/4.63
SLCNN	1.44/1.72/2.03	2.90/3.81/4.53	3.00/3.90/4.80
NS-SGAN	1.57/1.83/2.11	2.99/3.80/4.69	3.07/3.83/4.76
SGAN	1.48/1.71/1.93	2.81/3.65/4.35	2.92/3.61/4.36

表 3 比较了 SGAN 与基准模型在 PeMS-S 数据集上针对 15 min、30 min、45 min 和 60 min 交通流的预测结果。采用图结构和注意力机制的模型,其表现优于包括 ARIMA 和 FC-LSTM 在内的时间序列模型,与基于 GCN 和 RNN 结构的

STGCN、DCRNN 模型相比,采用注意力机制的 SGAN 效果更好。由于 ASTGCN 只是部分采用注意力机制,仍然结合了卷积核进行特征提取,因此其无法有效提取长距离依赖关系,需要多个独立模块分别提取不同的周期特征,增加了模型的复杂性。相比 Graph WaveNet,SGAN 在 45 min 和 60 min 的预测上表现更好,证明了注意机制对更长时间交通流预测的能力;Graph WaveNet 由于采用了小感受野的卷积核,因此在提取短距离依赖方面的表现不错,在长距离依赖方面不如 SGAN 有效,特别是 SGAN 采用完全的注意力机制,可以直接提取长距离时空依赖性,不需要通过叠加太深的层增加感受野,从而降低了模型训练的难度。

表 3 在 PeMS-S 数据集上的实验结果对比

Tab.3 Comparison of experimental results on the PeMS-S dataset

模型	PeMS-S(15 min/30 min/45 min/60 min)		
	MAE	RMSE	MAPE/%
ARIMA	5.61/5.89/6.40/7.04	9.09/9.15/9.45/9.89	12.89/13.91/15.33/18.15
FC-LSTM	3.55/3.97/4.15/4.50	6.15/6.92/7.57/8.15	8.62/9.47/10.29/10.83
ASTGCN	2.73/3.40/3.82/4.09	4.95/6.21/6.95/7.41	7.29/8.98/10.04/10.74
DCRNN	2.27/3.16/3.85/4.42	4.27/5.95/7.15/8.33	5.26/7.68/9.01/10.83
STGCN	2.25/3.02/3.56/4.05	4.05/5.70/6.77/7.68	5.22/7.33/8.77/10.03
Graph WaveNet	2.16/2.82/3.18/3.48	4.10/5.45/6.21/6.60	4.96/6.97/8.10/8.69
NS-SGAN	2.28/2.91/3.22/3.49	4.17/5.41/6.12/6.55	5.38/7.22/8.11/8.69
SGAN	2.23/2.85/3.17/3.45	4.08/5.35/6.04/6.48	5.31/7.15/8.07/8.65

本文还分析了高阶情况下模型的性能,图 6 显示了在 PeMS-BAY 数据集上 MAE 取不同 k 值的结果比较。可以看出, k 值越大,数据集上的 MAE 结果越小。当模型考虑更高阶的邻居时, k 必须设置得足够大,以覆盖每个节点足够远的邻居。然而,高阶信息意味着更多的参数,当参数数量增加时,训练时过拟合问题会导致效果不明显。此外,对于较大的 k 值,经过多层特征提取后,图上节点的信息传递会过于平滑,注意力系数的权重会趋向于平均分配,对节点失去区分性。

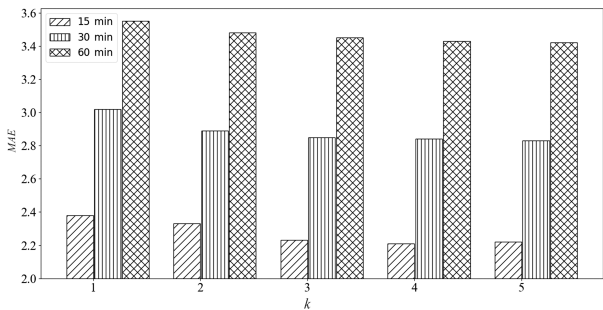


图 6 PeMS-BAY 数据集上取不同 k 值的 MAE 结果

Fig. 6 MAE results of different values of k on PeMS-BAY

为了观察注意力机制的影响,从 SDL 提取出学习到的注意力系数矩阵并可视化。从 PeMS-S 数据集中随机选择的 16 个传感器(节点),其连接的结构如图 7 所示。图 8 是当 $k=1$ 时的注意力系数热力图,点 (x,y) 处的像素值表示节点 x 对节点 y 的注意力系数,像素值颜色越深,表示对应的注意力系数权重越高。由于邻接矩阵 A 的对称性,因此系数矩阵也具有对称的形式。当 k 大于 1 时,观察到许多较深色的非对角块,因为阶数增加时,在感受野中有更多的节点,相当于可以提取到更长距离的节点依赖关系。以节点 11 为例,图 8 只考虑一

阶邻居节点,如节点 2、7、8 及其自身。当 k 值增大时,可以从图 9 和图 10 中看到,节点 11 与更远处的节点产生了明显的注意力系数,即考虑了长距离的空间依赖性。

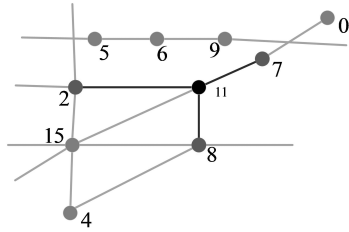


图 7 随机选取节点形成的结构图

Fig. 7 Structure diagram formed by randomly selecting nodes

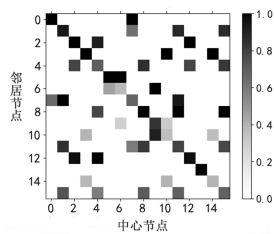


图 8 k 为 1 时注意力系数热力图

Fig. 8 Heat map of attention coefficient when k is 1

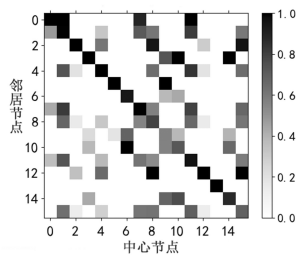


图 9 k 为 2 时注意力系数热力图

Fig. 9 Heat map of attention coefficient when k is 2

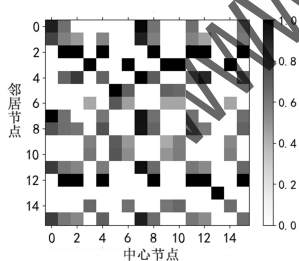


图 10 k 为 3 时注意力系数热力图

Fig. 10 Heat map of attention coefficient when k is 3

4 结论 (Conclusion)

本文提出了一种基于图注意力机制的交通流预测模型,该模型摒弃了传统的卷积、递归单元,采用注意力机制能有效地提取交通网络图中的长距离时空依赖关系。同时,在注意力机制中引入了空间结构和时间结构信息,使模型不仅考虑了节点之间的特征相关性,还考虑了节点的连接方式,在长距离时空上能有效区分不同的节点。在真实数据集上的实验表明,该模型的预测结果明显优于现有模型。

参考文献 (References)

- [1] BUI K H N, NGUYEN N D, YI H. Dynamic spatial transformer wavenet network for traffic forecasting[J]. Vietnam Journal of Computer Science, 2023, 10(1): 25-38.
- [2] 甘萍, 林基明, 农丽萍, 等. 一种用于交通预测的自适应时空图神经网络[J]. 桂林电子科技大学学报, 2023, 43(1): 7-13.
- [3] BOGAERTS T, MASEGOSA A D, ANGARITA-ZAPATA J S, et al. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data [J]. Transportation Research Part C: Emerging Technologies, 2020, 112: 62-77.
- [4] HOU Z S, LI X Y. Repeatability and similarity of freeway traffic flow and long-term prediction under big data[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(6): 1786-1796.
- [5] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [6] BELHADI A, DJENOURI Y, DJENOURI D, et al. A recurrent neural network for urban long-term traffic flow forecasting [J]. Applied Intelligence, 2020, 50(10): 3252-3265.
- [7] ZHANG Z C, LI M, LIN X, et al. Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies [J]. Transportation Research Part C: Emerging Technologies, 2019, 105 (8): 297-322.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Computer Society and Computer Vision Foundation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [9] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting [C]//Association for the Advancement of Artificial Intelligence (AAAI). Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI press, 2019: 922-929.
- [10] ZHENG C P, FAN X L, WANG C, et al. GMAN: a graph multi-attention network for traffic prediction [C]//Association for the Advancement of Artificial Intelligence (AAAI). Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI press, 2020: 1234-1241.
- [11] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [EB/OL]. (2018-02-04) [2023-05-22]. <https://arxiv.org/abs/1710.10903>.
- [12] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for quantum chemistry [C]//International Conference on Machine Learning. Proceedings of Machine Learning Research. New York: ACM, 2017: 1263-1272.
- [13] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the Web; SIDL-WP-1999-0120 [R]. California: Stanford InfoLab, 1999.

(下转第 62 页)