



计算机科学

COMPUTER SCIENCE

基于多邻接图与多头注意力机制的短期交通流量预测

尹恒, 张凡, 李天瑞

引用本文

尹恒, 张凡, 李天瑞. 基于多邻接图与多头注意力机制的短期交通流量预测[J]. 计算机科学, 2023, 50(4): 40-46.

YIN Heng, ZHANG Fan, LI Tianrui. Short-time Traffic Flow Forecasting Based on Multi-adjacent Graph and Multi-head Attention Mechanism [J]. Computer Science, 2023, 50(4): 40-46.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于合成图像和Xception改进模型的安卓恶意家族分类方法](#)

Android Malware Family Classification Method Based on Synthetic Image and Xception Improved Model

计算机科学, 2023, 50(4): 351-358. <https://doi.org/10.11896/jsjcx.220300200>

[基于交互注意力和图卷积网络的方面级情感分析](#)

Aspect-level Sentiment Classification Based on Interactive Attention and Graph Convolutional Network

计算机科学, 2023, 50(4): 196-203. <https://doi.org/10.11896/jsjcx.220100105>

[基于用户长短期偏好的序列推荐模型](#)

Sequential Recommendation Model Based on User's Long and Short Term Preference

计算机科学, 2023, 50(4): 47-55. <https://doi.org/10.11896/jsjcx.220100264>

[传播树结构结点及路径双注意力谣言检测模型](#)

Dual-attention Network Model on Propagation Tree Structures for Rumor Detection

计算机科学, 2023, 50(4): 22-31. <https://doi.org/10.11896/jsjcx.220200037>

[基于BERT和多特征融合嵌入的中文拼写检查](#)

Chinese Spelling Check Based on BERT and Multi-feature Fusion Embedding

计算机科学, 2023, 50(3): 282-290. <https://doi.org/10.11896/jsjcx.220100104>

基于多邻接图与多头注意力机制的短期交通流量预测

尹恒¹ 张凡^{1,2} 李天瑞^{1,2,3}

1 西南交通大学计算机与人工智能学院 成都 611756

2 四川省制造业产业链协同与信息化支撑技术重点实验室 成都 611756

3 综合交通大数据应用技术国家工程实验室 成都 611756

(yinheng714995@163.com)

摘要 交通流预测在智慧城市系统中占有重要地位,是许多交通方向应用的基石。该任务的难点在于如何有效地建模交通流的时空依赖。现有方法大都使用图卷积网络(Graph Convolution Networks,GCN)建模空间关系,使用卷积神经网络网络(Convolution Neural Network,CNN)或者循环神经网络(Recurrent Neural Network,RNN)建模时间关系,但在建模空间关系时往往只利用邻接矩阵建模了局部关系而忽略了全局空间信息。而在整个路网中存在一些道路,其周围的路网结构相似,这些道路的路网中承载的作用是相似的,这些相似道路的特征也可以作为流量预测的依据。因此,提出一种基于多邻接图与多头注意力机制的时空网络模型 MA-STGCN,包括:1)利用 node2vec 算法计算路网中道路的向量表示,通过阈值计算出相似矩阵用于图卷积操作,抽取全局空间信息;2)利用多通道自注意力机制深入挖掘模型的时空特征。在公开数据集 PEMS04 与 PEMS08 上进行的实验验证了该模型的有效性,其准确率与主流模型相比均有提高。

关键词: 交通流预测;空洞卷积;时空网络;注意力机制;节点嵌入

中图分类号 TP391

Short-time Traffic Flow Forecasting Based on Multi-adjacent Graph and Multi-head Attention Mechanism

YIN Heng¹, ZHANG Fan^{1,2} and LI Tianrui^{1,2,3}

1 School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

2 Manufacturing Industry Chains Collaboration and Information Support Technology Key Laboratory of Sichuan Province, Chengdu 611756, China

3 National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Chengdu 611756, China

Abstract Traffic flow forecasting is the cornerstone of many applications in transportation which has a great importance in smart city system. The difficulty of this task is how to effectively model the temporal and spatial dependence. Existing methods usually use GNN to model temporal correlation and CNN or RNN to model temporal correlation. When modeling the spatial correlation, only the adjacency matrix is applied to model local relationships while ignoring global spatial information. However, there are some roads in the entire road network whose surrounding structures are similar, and these roads carry similar functions in the road network. Therefore, the characteristics of these similar roads can also be used as the basis for traffic prediction. This paper proposes a traffic flow forecasting model based on multi-adjacent matrix and multi-head attention mechanism. It includes: 1) the node2vec algorithm is applied to calculate the vector representation of the road in road network, and the similarity matrix is calculated through the threshold for graph convolution operation to extract global spatial information; 2) the multi-channel self-attention mechanism is used to mine the spatial and temporal features of the model. Experiments on public datasets PEMS04 and PEMS08 demonstrate the proposed model's effectiveness. Its accuracy is improved compared with the mainstream models.

Keywords Traffic forecasting, Dilated convolution, Spatial-Temporal network, Attention mechanism, Node embedding

1 引言

随着城市化进程的不断推进与经济的高速发展,交通

拥堵的负面影响也越来越大,会造成严重的经济损失。通过建立有效的交通流量预测模型,可以提前预知路况,通过交通管控缓解道路拥堵,同时为行程规划、地区功能划分等上层

到稿日期:2022-02-16 返修日期:2022-09-05

基金项目:国家自然科学基金(61773324);四川省重点研发项目(2020YFG0035)

This work was supported by the National Natural Science Foundation of China(61773324) and Sichuan Key R&D Project, China (2020YFG0035).

通信作者:张凡(fan.zhang@swjtu.edu.cn)

任务提供基础,更好地服务于智慧城市的建设。

城市路网复杂且车流量较大,同时车流数据具有高度的时空相关性。首先,在空间上来说,道路相互连接,车流只能通过相连接的道路进行流动,因此一个路段在某一时刻的流量势必会在后面的时刻对周围路段的流量造成影响;其次,道路流量是时间序列数据,具有明显的时间相关性,同一路段的连续几个时间片的数据会对后面的流量数据造成影响。

近来有许多工作通过图神经网络建模空间信息并取得了良好的效果,但是这些工作单纯地使用原始邻接矩阵进行图卷积,只利用到了局部空间信息,而忽略了全局信息。在整个路网中,存在一些局部结构相似的路段,它们在路网中所承担的功能也是相似的,一段时间内这些路段的交通流量趋势应当是相似的,因此可以引入其他相似路段的信息来辅助预测道路未来的流量。

据此,本文提出一种基于多邻接矩阵进行时空卷积的神经网络预测模型 MA-STGCN。

首先,通过图嵌入算法构造包含路网空间信息的路段编码,并计算出自适应的邻接矩阵,进一步辅助原始邻接矩阵参与图卷积计算,从而抽取到更加丰富的空间特征;然后,将卷积神经网络与多头时间注意力机制相结合来抽取时间维度上的特征;最后,通过多层堆叠的时空层抽取结果并在每层之间增加残差连接,以充分捕获时空特征,得到更加准确的预测结果。

本文的贡献如下:1)提出了基于多邻接图和多头注意力的神经网络模型 MA-STGCN,用以解决交通流量预测问题;2)模型引入了道路空间相似矩阵,能学习到全局层面的道路信息,同时引入多头注意力机制来更好地学习到特征表示;3)在 PEMS04 和 PEMS08 数据集中进行的大量实验表明,MA-STGCN 的效果优于其他基线模型。

2 相关工作

2.1 交通流预测

早期的交通流预测算法主要基于 SVR^[1], ARIMA^[2] 和 Random Forest Regression^[3] 等机器学习模型,这些模型拥有较强的特征表达能力,在小规模的数据集上能取得不错的效果,但是在大规模数据上的表现则较差。伴随着深度学习^[4] 的崛起,研究者开始使用深度学习方法如 LSTM^[5] 和 GRU^[3] 来预测交通流,并在其基础上做出优化^[6],但是这些方法仅仅考虑了单条道路上的交通流量变化,而没有考虑路段之间的连接信息。为了解决这个问题, CNN 开始被用于空间关系建模, ConvLSTM^[7] 和 ST-ResNet^[8] 被相继提出,其使用 CNN 抽取空间特征的同时使用 LSTM 抽取时间特征。然而,这种方法只能用于网格化的数据,而现实中的交通流数据都基于路网,是一种更为复杂的图结构, CNN 无法处理。

GCN^[9] 被用于处理图结构数据,与路网的图结构高度契合,后续的研究大都使用 GCN 抽取空间特征。GCRN^[10] 使用 GCN 和 RNN 来解决序列建模问题;DCRNN^[11] 使用双向随机游走对空间依赖进行建模,再使用编码器-解码器^[12] 架构对时间序列进行建模;DCGRU-RF^[13] 在 DCRNN 的基础上使用 GRU,结合 Random Forest 进行建模;STGCN^[14] 则

使用卷积网络和图卷积网络组成时空卷积块,通过叠加抽取特征;ASTGCN^[15] 在时空卷积的基础上增加了时间注意力机制和空间注意力机制,同时考虑了邻近、日、周的时间序列作为特征;Wu 等^[16] 提出的 GraphWaveNet,使用空洞卷积^[17] 捕获时间特征,同时引入门控机制来控制信息的流入流出,在空间特征的部分引入自适应的邻接矩阵用于学习隐含的连接关系,在一定程度上弥补了路网结构对节点间关系表征的不足;Li 等^[18] 在 GraphWaveNet 的基础上引入了三支门控和自回归分量;STSGCN^[19] 引入了局部时空图概念,同时计算时空特征,通过大量堆叠特征抽取层抽取特征进行计算。

2.2 图嵌入

许多复杂问题涉及的数据为图结构。为了计算的便利,需要将图中的每个节点使用低维向量进行表示。图嵌入,便是研究如何使嵌入的向量能更好地保留图中信息的问题。

Perozzi 等^[20] 提出了 DeepWalk,使用随机游走采样获得节点序列,将其当作文本序列,通过 word2vec^[21] 模型进行训练,得到每个节点的嵌入;Tang 等^[22] 提出了 Line 模型,通过定义一阶相似度和二阶相似度作为优化目标计算节点编码,同时引入了负采样和边采样等优化方法来减少计算量。DeepWalk 和 Line 分别是基于 DFS 邻域和 BFS 邻域的图嵌入算法,前者注重节点的同质性,后者则考虑节点的结构相似性。node2vec^[23] 综合考虑同质性和结构相似性,提出一种有偏随机游走算法,通过变量控制随机游走的访问倾向,能学习到更好的节点向量表示。

3 问题定义

定义路网为图 $G=(V,E,A)$, V 代表图的节点集合,即路网的道路集合, $|V|=N$ 代表共有 N 条路段; E 代表图中的边,即道路的连接关系; A 为邻接矩阵。

某一时刻 t 的路网状态被定义为 $X_G^{(t)} \in \mathbb{R}^{N \times C}$, 表示所有 N 条道路在时刻 t 时的 C 个特征。

交通流量预测的问题可以被定义为:给定历史数据 $(X_G^{(t-T+1)}, X_G^{(t-T+2)}, \dots, X_G^{(t)})$, 预测 $(X_G^{(t+1)}, X_G^{(t+2)}, \dots, X_G^{(t+T)})$, 其中 T 表示历史数据序列长度, T' 则表示预测序列长度。

4 MA-STGCN 模型

针对交通流预测问题,本文提出了 MA-STGCN 模型,其结构如图 1 所示。模型主要分为 3 个部分:输入层、时空卷积层和输出层。

(1)输入层:将归一化的数据输入模型,将时间编码和空间编码通过广播机制添加到输入中,然后通过线性变换层提升特征维度;根据原始邻接矩阵,通过图嵌入算法计算路段编码,利用阈值筛选出相似邻接矩阵;再通过自适应路段编码计算出关联邻接矩阵。

(2)时空卷积层:将数据连续通过 3 个时空卷积块,每层都通过残差连接到输出结果中。每个时空卷积块由一个膨胀卷积块、一个图卷积块和一个多头时间注意力块组成。

(3)输出层:前面得到的特征通过连续的线性变换得到输出结果。

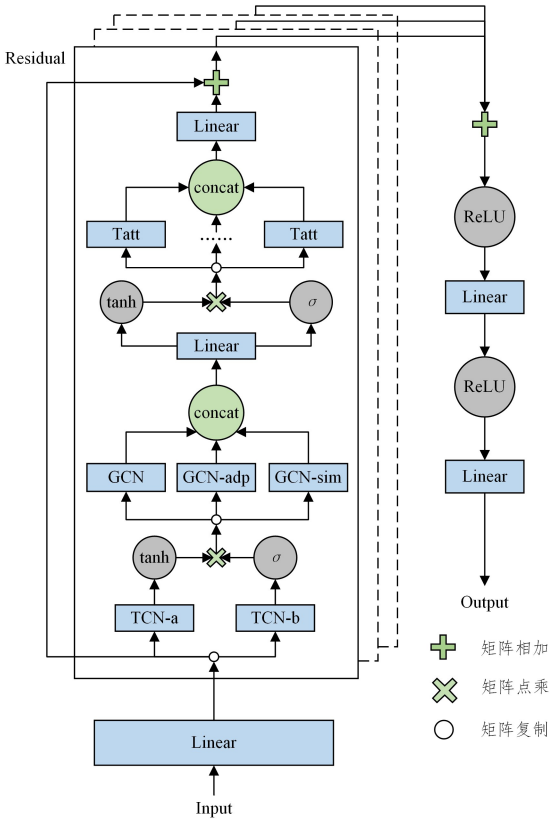


图1 MA-STGCN 模型

Fig.1 MA-STGCN model

4.1 相似邻接矩阵与自适应矩阵的构建

本模型除了原始的邻接矩阵 $A \in \mathbb{R}^{N \times N}$ 外,还需要构建自适应邻接矩阵 $A_{\text{adj}} \in \mathbb{R}^{N \times N}$,以及相似邻接矩阵 $A_{\text{sim}} \in \mathbb{R}^{N \times N}$ 。

构造两个可学习的变量 $emb_1 \in \mathbb{R}^{N \times C_{\text{emb}}}$, $emb_2 \in \mathbb{R}^{C_{\text{emb}} \times N}$,其中 C_{emb} 代表节点编码的维度,维度越高则节点可以包含的信息越多,同时学习也更加困难。模型训练完毕后, emb_1 与 emb_2 都可以一定程度上学习到节点表示。自适应邻接矩阵可由两者计算得到,如式(1)所示:

$$A_{\text{adj}} = \text{SoftMax}(\text{ReLU}((emb_2)^T emb_1)) \quad (1)$$

显式的邻接矩阵由于数据的不完整性,不一定能反映真实的空间依赖,通过自适应邻接矩阵可以在模型训练中学习到这些隐式依赖。

本文采用 node2vec 算法计算路段嵌入。首先通过有偏的随机游走进行节点采样:

$$P(C_i = x | C_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z}, & \text{if } (v, x) \in E \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$\pi_{vx} = \alpha_{pq}(t, x) \cdot W_{vx} \quad (3)$$

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p}, & \text{if } d_{tx} = 0 \\ 1, & \text{if } d_{tx} = 1 \\ \frac{1}{q}, & \text{if } d_{tx} = 2 \end{cases} \quad (4)$$

其中, t, v, x 分别代表前一个节点、当前节点以及下一个节点。通过式(2)获取从当前节点转移到下一个节点的概率,采样得到节点序列。其中, Z 为归一化常数, W_{vx} 为边 (v, x) 的权重, p 和 q 为控制游走策略的模型参数, d_{tx} 代表下一个

节点 x 与前一个节点 t 之间的关系, 0 表示 t 就是 x , 1 表示 t 与 x 相连, 2 表示 t 与 x 不相连。

在采样得到的节点序列上,求得使得序列出现的概率最大化的函数 f :

$$\max_f \sum_{u \in V} [-\log Z_u + \sum_{n_i \in N_i(u)} f(n_i) \cdot f(u)] \quad (5)$$

为了简化计算,实验中使用的是上述采样中得到的邻居节点集合。最终得到所有节点的向量表示 $E_{\text{node}} \in \mathbb{R}^{N \times C_{\text{node}}}$ 。根据自行设定的阈值 th_{sim} 计算出其相似矩阵 $A_{\text{sim}} \in \mathbb{R}^{N \times N}$ 。

$$A_{\text{sim}}(i, j) = \frac{E_{\text{node}}(i) \cdot E_{\text{node}}(j)}{\|E_{\text{node}}(i)\| \|E_{\text{node}}(j)\|} \quad (6)$$

$$A_{\text{sim}}(i, j) = \begin{cases} 0, & A_{\text{sim}}(i, j) < th_{\text{sim}} \\ 1, & A_{\text{sim}}(i, j) \geq th_{\text{sim}} \end{cases} \quad (7)$$

4.2 输入层

每条道路都有自己的空间特点,这一特征是不变的。为此,我们使用位置编码来表示每条道路的空间特征,这可以增强模型建模空间关联的能力;同理,使用时间编码来增强时间序列特征,整体加强模型的时空关联性。

创建一个可学习的空间嵌入变量 $S_{\text{emb}} \in \mathbb{R}^{N \times C}$ 与时间嵌入变量 $T_{\text{emb}} \in \mathbb{R}^{T \times C}$ 。模型输入数据 $X \in \mathbb{R}^{N \times C \times T}$,通过广播机制将这两个嵌入与输入相加,得到新的输入表示:

$$X^{\text{new}} = X + S_{\text{emb}} + T_{\text{emb}} \in \mathbb{R}^{N \times C \times T} \quad (8)$$

S_{emb} 与 T_{emb} 两个编码随着模型的训练会慢慢包含对应的空间和时间信息,帮助模型更好地捕获时空特征。

随后,通过线性变换将输入 X^{new} 映射到更高的维度 D_1 :

$$X^{\text{new}} = X^{\text{new}} W_m + b_m \quad (9)$$

4.3 时空卷积块

4.3.1 残差连接层

在每个时空卷积块的最开始,会通过一个全连接层将输入映射到维度 D_{res} ,连接到最终的输出:

$$X^{\text{res}} = X^{\text{new}} W_{\text{res}} + b_{\text{res}} \quad (10)$$

4.3.2 时间卷积层

通过一维 CNN 进行卷积,采用步长为 kt 的卷积核进行时间维度上的卷积。如果单纯地使用普通一维卷积来捕捉时间维度特征,会导致需要较多的卷积层数才能捕获到全局特征。为此,使用膨胀卷积来获取更大的感受域,使模型能够在较少的层数内充分学习到全局的信息。相较于基于 RNN 的模型,基于膨胀卷积的模型可以更好地捕获长序列的信息,避免梯度爆炸。同时,由于 CNN 特性可以进行并行运算,其运算速度更快。

膨胀卷积的示意图如图 2 所示。显然,使用膨胀卷积可以在很少的层数内获取到全局的信息。

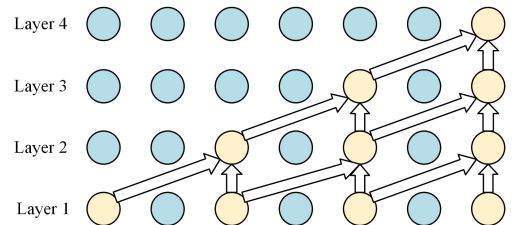


图2 膨胀卷积

Fig.2 Dilated convolution

$$X_1^{\text{temp}} = \text{Conv}^{\text{dil}}(X^{\text{new}}) \quad (11)$$

$$X_2^{\text{temp}} = \text{Conv}^{\text{dil}}(X^{\text{new}}) \quad (12)$$

$$X^{\text{temp}} = g(X_1^{\text{temp}}) \sigma(X_2^{\text{temp}}) \quad (13)$$

其中, $\text{Conv}^{\text{dil}}(\cdot)$ 表示一维膨胀卷积操作,在卷积过后输入的特征维度将会变换至 D_{dil} 。 $g(\cdot)$ 代表 \tanh 激活函数, $\sigma(\cdot)$ 则代表 sigmoid 激活函数,两者组合形成门控机制。门控机制可以有效地控制信息的流入和流出,保证模型不会由于层数过多出现梯度消失而无法收敛的问题。

通过时间卷积层学习到了时间特征后,再利用图卷积层来学习空间特征。

4.3.3 多图卷积层

本模型的图卷积层涉及 3 个邻接图: \mathbf{A} , \mathbf{A}_{adp} , \mathbf{A}_{sim} 。参考 DCRNN^[11] 的做法,将邻接矩阵转换为对应的概率转移矩阵:

$$\mathbf{P}_{\text{adj}} = \mathbf{A} / \text{rowsum}(\mathbf{A}) \quad (14)$$

$$\mathbf{P}_{\text{adp}} = \mathbf{A}_{\text{adp}} / \text{rowsum}(\mathbf{A}_{\text{adp}}) \quad (15)$$

$$\mathbf{P}_{\text{sim}} = \mathbf{A}_{\text{sim}} / \text{rowsum}(\mathbf{A}_{\text{sim}}) \quad (16)$$

该概率转移矩阵 $\mathbf{P}(i, j)$ 表示从 i 节点转移到 j 节点的概率;对概率转移矩阵做阶乘,即可得到第 k 阶的概率转移矩阵 $\mathbf{P}^k(i, j)$,其表示从 i 节点经过 k 步转移到 j 节点的概率,通过设定更高的阶数 k ,可以涵盖多个步长的空间转移信息。

通过设定阶数为 $order$,得出图卷积模块公式如下:

$$X_{\text{adj}}^{\text{sp}} = \sum_{k=1}^{order} \mathbf{P}_{\text{adj}}^k X^{\text{temp}} \mathbf{W}_{k, \text{adj}} \quad (17)$$

$$X_{\text{adp}}^{\text{sp}} = \sum_{k=1}^{order} \mathbf{P}_{\text{adp}}^k X^{\text{temp}} \mathbf{W}_{k, \text{adp}} \quad (18)$$

$$X_{\text{sim}}^{\text{sp}} = \sum_{k=1}^{order} \mathbf{P}_{\text{sim}}^k X^{\text{temp}} \mathbf{W}_{k, \text{sim}} \quad (19)$$

通过式(17)–式(19)分别计算出 3 个邻接矩阵对应的卷积结果后,将其拼接在一起。

$$X^{\text{sp}} = [X_{\text{adj}}^{\text{sp}}, X_{\text{adp}}^{\text{sp}}, X_{\text{sim}}^{\text{sp}}] \quad (20)$$

拼接结果 X^{sp} 通过线性变换,得到两个结果 X_1^{sp} 与 X_2^{sp} ,通过式(23)输出图卷积层结果。

$$X_1^{\text{sp}} = (X^{\text{sp}})^T \mathbf{W}_{\text{gcn},1} + b_{\text{gcn},1} \quad (21)$$

$$X_2^{\text{sp}} = (X^{\text{sp}})^T \mathbf{W}_{\text{gcn},2} + b_{\text{gcn},2} \quad (22)$$

$$X^{\text{sp}} = g(X_1^{\text{sp}}) \sigma(X_2^{\text{sp}}) \quad (23)$$

式(23)中的 g 代表 \tanh 激活函数, σ 代表 sigmoid 激活函数。

4.3.4 多头时间注意力层

在时空卷积块中增加一个时间注意力机制,提升时间维度上的特征抽取能力。通过式(24)–式(26)计算出注意力值 Att 。

$$S_1 = (X^{\text{sp}})^T \mathbf{W}_{\text{att},1} \quad (24)$$

$$S_2 = X^{\text{sp}} \mathbf{W}_{\text{att},3} \quad (25)$$

$$Att = \text{SoftMax}(\sigma((S_1 \mathbf{W}_{\text{att},2} (S_2)^T + b_{\text{att}}))) \quad (26)$$

通过式(27)对注意力进行归一化并计算出结果 X^{att} 。

$$Att'_{i,j} = \frac{\exp(Att_{i,j})}{\sum_{j=1}^T \exp(Att_{i,j})} \quad (27)$$

$$X^{\text{att}} = Att'_{i,j} X^{\text{sp}} \quad (28)$$

受到 Vaswani^[24] 的启发,设置多个时间注意力头,计算出多个 X^{att} ,拼接后得到 $X_{\text{cat}}^{\text{att}}$,通过线性变换后与最开始的残差求和即可获得时空卷积块的最终输出。

$$X_{\text{out}} = ((X_{\text{cat}}^{\text{att}})^T \mathbf{W}_{\text{out}} + b_{\text{out}}) + \text{trunc}(X^{\text{res}}) \quad (29)$$

其中, $\text{trunc}(\cdot)$ 代表将在时间维度上将 X^{res} 截断到 X_{out} 的长度,默认从尾部开始截断(数据在模型中由于时间卷积操作,在时间维度上变短)。

4.4 输出层

通过多个堆叠的时空卷积块后,将输出结果通过线性变换,输出最终的结果。

$$Y^{\text{pred}} = \text{ReLU}((X^{\text{att}})^T \mathbf{W}_{\text{out1}} + b_{\text{out1}}) \quad (30)$$

$$Y^{\text{pred}} = \text{ReLU}(Y^{\text{pred}})^T \mathbf{W}_{\text{out2}} + b_{\text{out2}} \quad (31)$$

其中, $Y^{\text{pred}} \in \mathbb{R}^{N \times T}$ 。

此处一次性直接预测出所有结果,而非使用 RNN 模型,避免了循环模型所导致的误差放大问题。

5 实验

在两个高速公路数据集 PEMS04 以及 PEMS08^[25] 上对所提模型进行了实验。两个数据集信息如表 1 所列。

表 1 数据集信息
Table 1 Dataset description

Datasets	Number of sensors	Time range
PEMS04	307	2018-01-01–2018-02-28
PEMS08	170	2016-01-07–2016-08-31

5.1 数据集划分

流量数据以 5 min 为单位进行统计,使用前一个小时的流量数据预测下一个小时的流量数据,即使用前 12 个数据点预测后 12 个数据点。数据集按照 6:2:2 的比例划分为训练集、验证集和测试集,每部分数据均做归一化处理。

5.2 评价指标

本文选用 3 个评价指标:平均绝对误差(Mean Absolute Error, MAE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)以及均方根误差(Root Mean Square Error, RMSE)。三者的定义如下:

$$MAE = \frac{1}{m} \left(\sum_{i=1}^m |y_i - \hat{y}_i| \right) \quad (32)$$

$$MAPE = \frac{1}{m} \left(\sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \quad (33)$$

$$RMSE = \sqrt{\frac{1}{m} \left(\sum_{i=1}^m (y_i - \hat{y}_i)^2 \right)} \quad (34)$$

其中, m 为样本数量, y_i 与 \hat{y}_i 分别表示预测值和实际值。另外, $MAPE$ 在计算 $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$ 时,为了避免分母为 0 导致数值溢出,在 $y_i = 0$ 时将 $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$ 计算为 0。

5.3 损失函数

使用 MAE 作为损失函数,预测结果为 Y^{pred} ,真实结果为 Y^{true} ,损失函数如下:

$$\text{loss} = \frac{\sum_{i=1}^N \sum_{j=1}^T |Y^{\text{pred}}(i, j) - Y^{\text{true}}(i, j)|}{T' \times N} \quad (35)$$

5.4 基准方法

选用 7 个交通流量预测模型作为基准模型来与所提模型进行比较。

SVR^[1]: 向量支持器回归, 使用线性核进行回归。

LSTM^[5]: 长短期记忆网络, 通过门控机制来解决梯度消失和长期记忆问题。

DCRNN^[11]: 扩散卷积循环神经网络, 使用扩散图卷积和 seq2seq 来编码时空特征信息。

STGCN^[14]: 时空图卷积网络, 使用 ChebNet^[26] 编码空间信息。

ASTGCN^[15]: 使用时空注意力机制来帮助学习时空特征, 同时引入了本周、本日以及当前一个小时的 3 种特征来更充分地学习到动态的时间信息。为了对比的公平, 实验中不使用其多时间特征部分的模型。

Graph WaveNet^[16]: 使用膨胀卷积来捕获更大范围的时间特征, 并引入自适应邻接矩阵来补充原始邻接矩阵中可能缺失的连接。

STSGCN^[19]: 通过构建局部时空图来同步地捕获时间特征与空间特征。

5.5 实验设置

在所提模型中, GCN 的 *order* 被设置为 2, 时间卷积核的大小为 2, 膨胀卷积参数为 2; 特征维度分别为 $C_{in} = 1$, $D_1 = 32$, $D_{res} = 512$, $D_{et} = 256$ 。用于计算自适应邻接矩阵的节点嵌入的维度 C_{emb} 为 10。

该模型使用 Pytorch1.7.0 实现, 在内存为 16 GB 的服务器上进行实验, 显卡使用 Titan V, 其显存为 12GB。Batchsize 设置为 64, 使用 Adam 优化算法, 学习率为 0.001。同时, 为了增强模型的泛化能力, 使用 dropout^[27] 机制在训练过程中随机抛弃神经元, 避免模型过拟合, dropout 的超参数设置为 0.3。

5.6 实验结果与分析

5.6.1 Batchsize 超参实验

Batchsize 是深度学习中一个十分重要的超参数, Batchsize 太小会导致模型无法收敛, 太大则会导致模型难以学习, 为此需要选择合适的 Batchsize 参数。参考以往的工作^[14, 19], 拟采用 Batchsize 为 64, 为了寻求更好的模型效果, 增加 16, 32, 128, 共计 4 个候选 Batchsize 在数据集 PEMS08 上进行实验, 结果如图 3 所示。

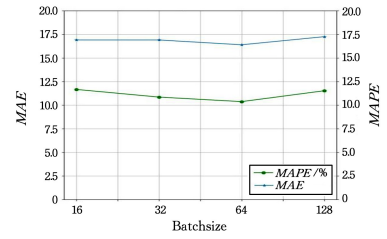


图 3 Batchsize 实验结果

Fig. 3 Results of different number of Batchsize

如图 3 所示, 模型效果一开始随着 Batchsize 的上升而提升, 随后在 Batchsize 继续上升时效果变差, 因此模型在 Batchsize 取 64 时能取得最好的效果。

5.6.2 路段相似度阈值实验

本文基于路段相似性构建相似矩阵, 其中涉及超参数阈值 th_{sim} 的选择, 其范围为 $[0, 1]$ 。当阈值 th_{sim} 太小时, 所得到的相似矩阵为全 1 矩阵, 相当于一个全连接图, 不仅不能帮助模型, 还会对模型造成负面影响; 反之, 当阈值太大时, 则可能导致相似矩阵全为 0, 学习不到任何信息。由于没有先前的工作可供参考, 在 PEMS08 上, 以 0.05 为间隔在 $[0, 1]$ 范围内进行实验, 实验结果如图 4 所示。

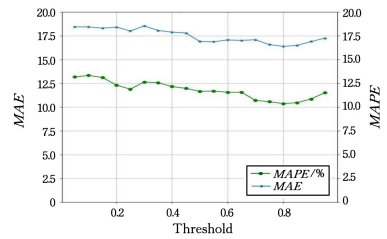


图 4 th_{sim} 实验结果

Fig. 4 Results of different number of th_{sim}

如图 4 所示, MAE 与 MAPE 指标一开始随着 th_{sim} 的增大而呈现下降趋势; 在 0.8 处达到最低值, 随后开始上升, 为此选择 th_{sim} 为 0.8。

5.6.3 对比实验

将本文提出的模型 MA-STGCN 与 7 个基线模型进行对比实验, 结果如表 2 所列。

表 2 实验结果

Table 2 Experiment results

Baseline methods		SVR	LSTM	DCRNN	STGCN	ASTGCN	Graph WaveNet	STSGCN	MA-STGCN
Datasets	Metric								
PEMS04	MAE	28.70	27.14	24.70	22.70	22.93	25.45	21.19	20.40
	MAPE/%	19.20	18.20	14.59	14.59	16.56	17.29	13.9	14.10
	RMSE	44.56	41.59	35.55	35.55	35.22	39.70	33.65	31.83
PEMS08	MAE	23.25	22.2	18.02	18.02	18.61	19.13	17.13	16.39
	MAPE/%	14.64	14.20	11.40	13.08	13.08	12.68	10.96	10.37
	RMSE	36.16	36.16	27.83	28.16	28.16	31.05	26.80	25.50

本文提出的模型在 PEMS04 和 PEMS08 两个数据集上都取得了较好的效果, 除了在 PEMS04 数据集上其 MAPE 指标略差于 STSGCN 模型外, 其他指标均优于其他所有参与比较的模型, 说明本文模型在此任务上效果更好; 同时, 深度学习模型的效果比传统机器学习模型如 SVR 的效果更好, 说明深度学习在此任务中抽取有用特征进行建模的能力更强。

5.6.4 消融实验

为了证明本文所提出方法的有效性, 对所提模型 MA-STGCN 进行消融实验, 测试其相似邻接矩阵、多头时间注意力模块的作用。为此, 在原模型基础上分别去掉相似邻接矩阵部分和多头时间注意力部分, 得到对照模型 MA-STGCN-WS 与 MA-STGCN-WATT。在 PEMS08 数据集上进行实验

比较,实验结果如表3所列。

表3 消融实验结果

Table 3 Ablation experiment results

	MAE	MAPE/%	RMSE
MA-STGCN-WS	16.57	10.67	25.79
MA-STGCN-WATT	17.16	11.45	26.92
MA-STGCN	16.39	10.37	25.50

可以看出,多头注意力机制和相似邻接矩阵均能提升模型的效果。上述实验结果证明了本文所提方法的有效性。

5.6.5 模型性能实验

将LSTM,STGCN,ASTGCN,STSGCN以及本文提出的MA-STGCN在数据集PEMS08上进行性能实验,对比各个模型的训练时间、推理时间以及显存占用,结果如表4所列。

表4 模型性能实验

Table 4 Computation time and GPU memory occupation

Model	Train/ (s/epoch)	Inference/ (s/epoch)	Memory/ MB
LSTM	4.80	0.78	1765
STGCN	9.30	1.67	4587
ASTGCN	11.10	2.41	3553
MA-STGCN	28.64	5.33	3815
STSGCN	104.47	21.48	5718

从表4可以看出,相较于LSTM,STGCN和ASTGCN等结构较为简单的模型,MA-STGCN虽然速度相对较慢,但是预测准确率远优于这些模型,且在5min预测一次的应用场景下其计算速度基本没有受到影响;而与预测结果相近的STSGCN相比,MA-STGCN的训练推理时间远远减少。该实验结果足以证明本文模型的综合性能更加优异。

结束语 本文提出一种基于多邻接图与多头注意力机制的时空网络模型MA-STGCN,包括:利用node2vec算法计算路网中道路的向量表示,通过阈值计算出相似矩阵用于图卷积操作,抽取全局空间信息;利用多通道自注意力机制深入挖掘模型的时空特征。在PEMS04与PEMS08公开数据集上进行的实验验证了该模型的有效性,其MAP,MAPE和RMSE等指标与对比模型相比均有提升。

模型对道路相似度的定义仅仅使用了路网的空间信息,后续工作可以考虑引入道路在时间序列上的相似度构成的相似矩阵,进一步利用全局信息。

另一方面,本文原定引入时空注意力机制,但是在实验过程中引入空间注意力机制会导致模型效果变差,反复实验均无法解决该问题,后续将探索如何改进空间注意力使得模型效果获得提升;除此之外,MA-STGCN的性能可以做进一步的优化提升,后续将研究如何提升其计算性能,例如优化图卷积计算方式,将其转换为在线学习模型,通过增量学习来避免重新使用大量数据训练模型等。

参考文献

[1] AGARAP A F M. A neural network architecture combining gated recurrent unit(GRU)and support vector machine(SVM)for intrusion detection in network traffic data[C]//Proceedings of

the 2018 10th International Conference on Machine Learning and Computing. 2018;26-30.

- [2] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results[J]. Journal of Transportation Engineering, 2003, 129(6): 664-672.
- [3] LESHEM G, RITOV Y. Traffic flow prediction using adaboost algorithm with random forests as a weak learner[C]// Proceedings of World Academy of Science, Engineering and Technology. 2007;193-198.
- [4] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [5] CUI Z, KE R, PU Z, et al. Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction[J]. arXiv:1801.02143, 2018.
- [6] ZHANG X H, FENG A M. Short-Time traffic flow prediction based on and LSTM network[J]. Application Research of Computers, 2021, 41(1): 225-230.
- [7] SHI X J, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [C] // Advances in Neural Information Processing Systems. 2015: 802-810.
- [8] ZHANG J, ZHENG Y, QI D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]// Proceedings of Thirty-first AAAI Conference on Artificial Intelligence. 2017: 1566-1661.
- [9] ESTRACH J B, ZAREMBA W, SZLAM A, et al. Spectral networks and deep locally connected networks on graphs[C]// Proceedings of 2nd International Conference on Learning Representations. 2014: 1-14.
- [10] SEO Y, DEFFERRARD M, VANDERGHEYNST P, et al. Structured sequence modeling with graph convolutional recurrent networks[C]// Proceedings of International Conference on Neural Information Processing. Cham: Springer, 2018: 362-373.
- [11] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[C]// Proceedings of International Conference on Learning Representation. 2018: 1-16.
- [12] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]// Advances in Neural Information Processing Systems. 2014: 3104-3112.
- [13] XIONG T, QI Y, ZHANG W B. Short-Time traffic flow prediction in road network based on DCGRU-RF model[J]. Computer Science, 2020, 47(5): 84-89.
- [14] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[J]. arXiv:1709.04875, 2017.
- [15] 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. 2019: 922-929.
- [16] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling[J]. arXiv:1906.00121, 2019.

- [17] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. arXiv:1511.07122, 2015.
- [18] LI H, WANG F, XIE S Y, et al. A traffic prediction model with dual autoregressive components based on improved graph wave network[J]. Computer Science, 2021, 48(S2): 159-165, 183.
- [19] SONG C, LIN Y, GUO S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 914-921.
- [20] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: Online learning of social representations [C] // Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014: 701-710.
- [21] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C] // Advances in Neural Information Processing Systems, 2013: 3111-3119.
- [22] TANG J, QU M, WANG M, et al. Line: Large-scale information network embedding [C] // Proceedings of the 24th International Conference on World Wide Web, 2015: 1067-1077.
- [23] GROVER A, LESKOVEC J. node2vec: Scalable feature learning for networks [C] // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 855-864.
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // Advances in Neural Information Processing Systems, 2017: 5998-6008.
- [25] WEISS M, BONNEL P, HUMMEL R, et al. Analyzing on-road emissions of light-duty vehicles with Portable Emission Measurement Systems (PEMS) [R]. JRC Scientific and Technical Reports, EUR, 2011.
- [26] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering [J]. Advances in Neural Information Processing Systems, 2016, 29: 3844-3852.
- [27] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.



YIN Heng, born in 1997, postgraduate. His main research interests include traffic flow prediction and urban computing.



ZHANG Fan, born in 1989, Ph.D, research assistant. His main research interests include fault diagnosis, pattern recognition and data fusion.

(责任编辑:柯颖)