

一种基于时空动态图注意力网络的共享出行需求预测方法

骈纬国¹, 吴映波², 陈 蒙², 蔡俊鹏²

(1. 重庆大学汽车协同创新中心, 重庆 400044; 2. 重庆大学大数据与软件学院, 重庆 400044)

摘 要: 基于图卷积神经网络的共享出行需求预测一般采用非时间特定性的静态空间图结构提取非欧氏空间相关性特征, 这种方式所构建的城市结构图是一种在不同时间间隔的静态空间图结构, 而不能动态提取不同时间间隔的空间相关性特征. 针对这一问题, 本文提出了一种基于时空动态图注意力网络(Spatial-Temporal Dynamic Graph Attention Networks, STDGAT)的共享出行需求预测方法. 该方法基于区域间通勤关系动态构建时间特定性城市空间图结构, 以实现动态空间相关性建模, 并采用图注意力网络和长短期记忆网络自适应提取动态空间相关性特征和时间依赖性特征. 使用一个全连接层作为输出预测层将联合时空特征映射为真实的需求值以完成预测. 实验结果表明, 该方法在RMSE, MAPE和MAE等3个评价指标上均优于实验基准比较方法.

关键词: 需求预测; 深度学习; 图注意力网络; 时空数据预测; 共享出行

中图分类号: TP399; U495

文献标识码: A

文章编号: 0372-2112(2022)02-0432-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210179

A Spatial-Temporal Dynamic Graph Attention Network Based Method for Sharing Travel Demand Prediction

PIAN Wei-guo¹, WU Ying-bo², CHEN Meng², CAI Jun-peng²

(1. Automotive Collaborative Innovation Center, Chongqing University, Chongqing 400044, China;

2. School of Big Data & Software Engineering, Chongqing University, Chongqing 400044, China)

Abstract: Sharing travel demand prediction methods based on graph convolutional network generally adopt non-time-specific static graph structure to extract non-Euclidean spatial features. The urban structure graph in this way is a static spatial graph structure at different time intervals, which cannot extract spatial feature at different time intervals dynamically. To tackle this problem, this paper proposes a spatial-temporal dynamic graph attention network(STDGAT) for sharing travel demand prediction. STDGAT constructs a commuting-based time-specific dynamic graph structure to achieve the dynamic spatial correlation modeling. After that, STDGAT applies graph attention network and the long short-term memory to adaptively extract the dynamic spatial feature and temporal feature respectively. A fully connected layer is used as the prediction layer to map the joint spatial-temporal feature to real demand values. Experiment results demonstrate the significant improvement of our method on three evaluation metrics RMSE, MAPE, and MAE over state-of-the-art baselines.

Key words: demand prediction; deep learning; graph attention network; spatial-temporal prediction; sharing travel

1 引言

随着移动互联网和共享经济在全球的迅速发展, 共享出行在出行服务中开始扮演越来越重要的角色. 滴滴出行、优步、Grab和长安出行等在线共享出行平

台, 为人们的出行提供了一种更加便利的方式. 准确的共享出行需求预测可有效协助平台分配车辆、提高车辆利用率、减少用户等待时间、缓解交通堵塞等^[1].

共享出行需求预测是一类基于历史数据的空间和

收稿日期: 2021-01-27; 修回日期: 2021-03-10; 责任编辑: 王天慧

基金项目: 国家重点研发计划(No.2019YFB1706101); 重庆市技术创新与应用发展专项重点项目(No.cstc2019jscx-mbxdX0047); 中央高校基本科研业务费(No.2020CDCGRJ050)

时间特征来预测未来需求的时空数据挖掘任务. 在最新的相关研究工作中, 学者们往往使用图卷积网络 (Graph Convolutional Network, GCN) 来提取城市中的非欧氏空间相关性特征^[1-3]. 但传统的 GCN 模型在提取空间相关性特征时存在着为不同的邻居节点 (区域) 分配相同的重要性系数或权重^[4]的问题. 此外, 现有的基于 GCN 的时空预测方法在构建城市结构图时, 往往是基于区域地理近邻关系来建立图中节点 (区域) 之间边的关系. 这种方式所构建的城市结构图是一种在不同的时间间隔中保持一致的非时间特定性的静态空间图结构, 不利于针对不同的时间间隔来动态地提取空间相关性特征.

针对上述问题, 本文提出一种基于时空动态图注意力网络 (Spatial-Temporal Dynamic Graph Attention Networks, STDGAT) 的共享出行需求预测方法. 在 STDGAT 中, 图注意力网络 (Graph Attention Network, GAT)^[4]被用来提取区域间的非欧氏空间相关性特征. 基于注意力机制^[5], GAT 可为城市结构图中节点的不同的邻居节点自适应分配不同的重要性系数, 以实现对不同邻居区域的自适应重要性分配. 此外, 该方法使用基于通勤关系的时间特定性动态空间图结构, 以建立在不同时间间隔具有时间特定性的动态空间关联性. 具体而言, 如图 1 所示, 若在时间间隔 t 中存在从区域 A 到区域 B 的用车通勤, 但不存在从区域 B 到区域 A 的用车通勤, 则在时间间隔 t 中的城市空间图中存在从节点 A 到节点 B 的有向边, 但不存在从节点 B 到节点 A 的有向边, 即对于节点 A 来说, 节点 B 是其邻居节点, 但对于节点 B 来说, 节点 A 不是其邻居节点. 如果在时间间隔 t 中, 区域 A 和区域 B 之间不存在用车通勤 (双向均不存在), 则在时间间隔 t 中节点 A 和节点 B 之间不存在任何边连接关系, 即节点 A 和节点 B 互相不为邻居节点. 通

过这种方式, 城市空间图可由静态的图结构变为具有时间特定性的动态图结构, 以此来建立具有时间特定性的动态空间关联性.

通过在一个大规模的共享出行数据集上对本文提出的基于 STDGAT 的共享出行需求预测方法进行实验验证. 实验结果表明, 该方法在 RMSE, MAPE 和 MAE 3 个评价指标上均优于相关基准比较方法.

2 相关工作

传统的时空预测方法主要依赖时间序列中的统计信息来回归得到最终的预测结果. 自回归移动平均模型 (Autoregressive Integrated Moving Average, ARIMA) 是其中的一个代表性方法, 并在一些传统的交通预测任务中得到了广泛的应用^[6,7]. 胡文斌等人^[8]提出了用于城市交通导航的多路口导航量搜索方法. 为了提高模型的预测准确率, 一些额外的信息也被研究人员加入模型中来辅助模型的预测, 例如天气状况和节假日信息等^[9-12].

近年来, 研究人员开始广泛应用深度学习方法来解决相关问题. Zhang 等人^[9]提出了一个基于深度卷积神经网络的方法来预测城市人流量, 在此之后, 他们进一步将残差连接^[13]加入他们的模型中, 并提出了用于城市人流量预测的 ST-ResNet^[10]. Wang 等人^[14]提出了用于出租车供需预测的多层全连接神经网络. Yu 等人^[15]和 Zhao 等人^[16]分别提出了基于长短期记忆 (Long Short-Term Memory, LSTM) 和双向长短期记忆 (Bi-directional Long Short-Term Memory, BiLSTM) 的时空预测方法. Yao 等人^[17]使用卷积神经网络 (Convolutional Neural Network, CNN) 和循环神经网络 (Recurrent Neural Network, RNN) 来提取联合时空特征, 并在此基础上使用了图嵌入 (Graph Embedding) 方法来获取远距离区域之间的语义相关性特征. Qiu 等人^[18]提出了一个用于出租车起始点-终点需求预测的情景化时空网络.

为了捕获不规则的非欧几里得空间相关性, Li 等人^[3]提出了用于交通预测的基于扩散卷积的循环神经网络; Yu 等人^[2]提出了基于 GCN 的 STGCN 模型; 在这之后, Geng 等人^[1]构建了基于邻居关系、功能相似性和交通连通性的多图结构模型, 以此来捕获更加多样性的非欧氏空间相关性.

3 基于时空动态图注意力网络的共享出行需求预测方法

3.1 共享出行需求预测问题定义

共享出行需求预测任务是一种基于历史时空数据的时间序列预测任务, 具体而言, 是根据过去多个连续时间间隔中的需求序列数据来预测未来一个时间间隔



图1 城市的不同区域

中全局需求量的任务. 其中给定的历史需求序列是从时间间隔 $t - L + 1$ 到当前时间间隔 t .

本文首先根据真实的地理坐标(经度和纬度)对整个城市进行分割, 因此每个区域的真实形状是规则的矩形, 并被抽象成一个点来进行建模. 因此, 预测任务可以表示为

$$X^{t+1} = F(\{X^{t-L+1}, \dots, X^{t-1}, X^t\}) \quad (1)$$

其中, X^t 表示在 t 时刻所有区域的需求量的集合, $F(\cdot)$ 表示预测模型.

3.2 模型整体结构

图2展示了STDGAT的整体框架结构. STDGAT由空间模块、时间模块和输出预测层3个部分组成.

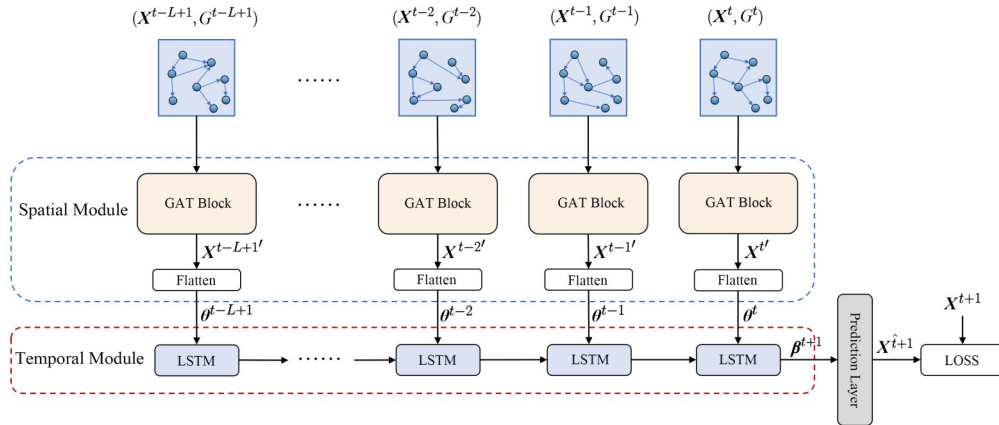


图2 模型整体框架结构

3.2.1 空间模块

空间模块(Spatial Module)用于在每个时间间隔中, 提取全局的空间特征. 空间模块包含一个由多个图注意力层(Graph Attention Layer, GAT)层组成的GAT Block. 其内部结构如图3所示.

GAT层通过学习节点间的注意力因子来更新每个节点的隐藏特征. 通过使用GAT层来提取空间相关性特征, 模型为城市结构图中节点的不同邻居节点分配了不同的重要性系数, 从而实现了自适应的空间相关性特征提取. 具体的操作过程如下所述.

被GAT层操作的图被定义为 $G = (V, E)$, 其中, V 和 E 分别表示图 G 中的节点和边的集合. 在节点集合 V 中 v_i 表示节点集合 V 中的第 i 个节点. 基于上述定义, 节点 v_i 在第 l 层的特征向量表示为 $\mathbf{h}_i^l \in \mathbb{R}^{d(l)}$, 其中, $d(l)$ 表示节点 v_i 在第 l 层的特征向量的长度. 为了在不同的时间间隔中提取具有时间特定性的动态空间相关性特征, STDGAT基于区域间通勤关系构建了具有时间特定性的动态空间图结构, 并在空间模块中使用GAT层来自适应地提取具有时间特定性的动态空间关联性特征. 因此, 上述概念被进一步定义为 $v_i^t, G^t = (V, E^t)$ 和 $\mathbf{h}_i^t \in \mathbb{R}^{d(l)}$, 分别表示在时间间隔 t 中的节点、图和特征向量.

如图4所示, 节点 v_i^t 和 v_j^t 之间的注意力因子的计算过程被表示为

$$e_{ij}^t = a(\mathbf{W}\mathbf{h}_i^t, \mathbf{W}\mathbf{h}_j^t) \quad (2)$$

其中, $\mathbf{W} \in \mathbb{R}^{d(l+1) \times d(l)}$ 是第 l 层的可训练参数, $a(\cdot, \cdot)$ 是

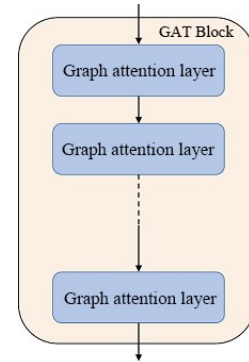


图3 GAT Block内部结构

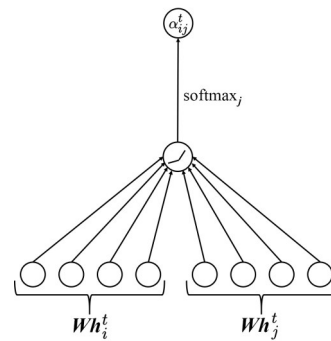


图4 图注意力层

用于计算节点 v_i^t 和节点 v_j^t 之间关联性的函数. 需要注意的是, $v_j^t \in N_{v_i^t}$, 其中 $N_{v_i^t}$ 表示节点 v_i^t 在时间间隔 t 中的邻居节点的集合. 关于函数 $a(\cdot, \cdot)$ 的选择, 学者们

通常使用一个可训练的前馈神经网络^[4]. 因此,在 t 时刻,节点 v_i^t 与其邻居节点 v_j^t 之间的注意力因子可以被表示为

$$e_{ij}^t = \text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i^t \parallel \mathbf{W}\mathbf{h}_j^t]) \quad (3)$$

其中, $(\cdot)^T$ 和 $\parallel$ 分别表示矩阵的转置操作和拼接操作. 在进行完上述操作之后,softmax 函数被用来对注意力因子进行归一化操作.

$$\alpha_{ij}^t = \text{softmax}(e_{ij}^t) = \frac{\exp(e_{ij}^t)}{\sum_{v_k \in N_{v_i}^t} \exp(e_{ik}^t)} \quad (4)$$

最后,通过使用上述归一化注意力因子来加权求和得到节点 v_i^t 新的特征向量,具体过程为

$$\mathbf{h}_i'^t = \sigma \left(\sum_{v_j \in N_{v_i}^t} \alpha_{ij}^t \mathbf{W}\mathbf{h}_j^t \right) \quad (5)$$

其中, $\mathbf{h}_i'^t$ 表示经过注意力操作后,节点 v_i^t 新的特征向量. 需要注意的是,图注意力层中的参数 $\mathbf{W}$ 和 $\mathbf{a}$ 在不同时间间隔中是共享的,因此,GAT层的参数量不会随着输入序列的长度变化而变化.

在GAT Block中,每个GAT层的操作被表示为

$$\mathbf{X}_{l+1}^t = f_l(\mathbf{X}_l^t) \quad (6)$$

其中, $\mathbf{X}_l^t \in \mathbb{R}^{N \times d(l)}$ 表示在时间间隔 t 中第 l 个GAT层的输入, $f_l(\cdot)$ 表示第 l 个图注意力层的操作, N 和 $d(l)$ 分别表示城市的区域总数和第 l 层中每个节点的特征向量长度. 因此,时间间隔 t 中的需求 $\mathbf{X}^t$ 在经过GAT Block之后变为

$$\mathbf{X}'' = \text{GATBlock}(\mathbf{X}^t) \quad (7)$$

其中, $\mathbf{X}'' \in \mathbb{R}^{N \times d}$ 表示在时间间隔 t 中从GAT Block输出的空间特征, d 表示在经过了GAT Block之后每个节点的特征向量长度. 在经过了图注意力操作之后,在时间间隔 t 中从GAT Block输出的特征矩阵 $\mathbf{X}''$ 被展开为一个特征向量 $\boldsymbol{\theta}^t \in \mathbb{R}^{Nd}$. 最后,长度为 L 的需求序列经过空间模块后,输出的空间特征序列 $\mathbf{S}_{t+1} \in \mathbb{R}^{L \times Nd}$ 表示为

$$\mathbf{S}_{t+1} = [\boldsymbol{\theta}^n | n = t, t-1, t-2, \dots, t-L+1] \quad (8)$$

3.2.2 时间模块

在时间依赖性建模中,循环神经网络(Recurrent Neural Networks, RNN)已被证明可以取得良好的效果^[19],长短期记忆(LSTM)的引入克服了传统的RNN难以学习长期依赖性关系的缺点^[20]. 此外,已有的关于时空预测的文献表明,LSTM在处理这种时序数据时具有优异的表现^[1,17,18]. 因此,本文在STDGAT的时间模块(Temporal Module)中,使用了LSTM来对需求序列的时间依赖性进行建模.

LSTM引入了一个记忆单元 c_t 来累积序列中之前时刻的信息. 具体而言,在 t 时刻,给定输入 $\mathbf{x}_t$,LSTM使用一个输入门 i_t 和一个遗忘门 f_t 来更新记忆单元 c_t ,并且使用一个输出门 o_t 来控制隐藏状态 $\mathbf{h}_t$. 其表达式为

$$\begin{aligned} i_t &= \sigma(\mathbf{W}_{ii}\mathbf{x}_t + \mathbf{b}_{ii} + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_{hi}) \\ f_t &= \sigma(\mathbf{W}_{if}\mathbf{x}_t + \mathbf{b}_{if} + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_{hf}) \\ g_t &= \tanh(\mathbf{W}_{ig}\mathbf{x}_t + \mathbf{b}_{ig} + \mathbf{W}_{hg}\mathbf{h}_{t-1} + \mathbf{b}_{hg}) \\ o_t &= \sigma(\mathbf{W}_{io}\mathbf{x}_t + \mathbf{b}_{io} + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_{ho}) \\ c_t &= f_t \circ c_{t-1} + i_t \circ g_t \\ \mathbf{h}_t &= o_t \circ \tanh(c_t) \end{aligned} \quad (9)$$

其中, $\circ$ 表示Hadamard乘积, σ 表示sigmoid激活函数, $\mathbf{W}_{pq}$ 和 $\mathbf{b}_{pq}$ ($p \in (i, h), q \in (i, f, g, o)$)为LSTM的可训练参数, c_t 和 $\mathbf{h}_t$ 分别表示在 t 时刻记忆单元状态和隐藏状态. 关于更多LSTM的细节,请参考文献[20~22].

如图2所示,LSTM将空间模块输出的空间特征序列 $\mathbf{S}_{t+1}$ 作为输入,然后输出时空联合特征向量 $\boldsymbol{\beta}^{t+1} \in \mathbb{R}^k$.

3.2.3 输出预测层

输出预测层(Prediction Layer)旨在将联合特征向量 $\boldsymbol{\beta}^{t+1}$ 映射为需求向量,以此作为最终的预测结果. 本文使用了一个具有 N 个神经元的全连接层作为最后的输出预测层(N 代表城市中的区域总数). 输出预测层的公式可表示为

$$\hat{\mathbf{X}}^{t+1} = f(\mathbf{W}_{FC}\boldsymbol{\beta}^{t+1} + \mathbf{b}_{FC}) \quad (10)$$

其中, $\hat{\mathbf{X}}^{t+1}$ 表示最终的预测结果, $\mathbf{W}_{FC}$ 和 $\mathbf{b}_{FC}$ 分别表示输出预测层的权重和偏置, $f(\cdot)$ 为激活函数.

4 实验与分析

4.1 实验数据集与参数设置

本文实验使用来自滴滴出行的共享出行数据集(<https://gaia.didichuxing.com>). 该数据集包含海口市从2017年5月1日到2017年10月31日的共享出行订单数据,数据集包含有121个区域,每个区域的大小约为 $1 \text{ km} \times 1 \text{ km}$. 在本文实验中,时间间隔设置为1 h,数据集的总订单数为12 185 427,总时间间隔数为4 416.

并将2017年5月1日至2017年9月30日的数据作为训练集,剩余数据(2017年10月1日至2017年10月31日)作为测试集. 在训练集中,80%的数据用来训练模型,剩余20%用作验证集.

在STDGAT中,输入序列长度 L 被设置为5. 在空间模块中,GAT Block包含3个GAT层,每一层均包含32个隐藏单元. 在时间模块中,LSTM的隐藏层数量为1,并包含512个神经元. GAT层中的激活函数为LeakyReLU,在输出预测层中的激活函数为ReLU. 本文通过Adam^[23]优化器来对网络进行优化. 在训练过程中,学习率和权重衰减值分别为 $1\text{e-}3$ 和 $5\text{e-}5$. 模型的代码基于深度学习库Pytorch^[24]编写,并在两块NVIDIA 1080Ti GPU上进行训练,最大训练轮数为200.

4.2 损失函数

在训练阶段,STDGAT使用均方误差(Mean Square Error, MSE)作为损失函数,并通过使其最小化的方式来训练模型. MSE的公式可表示为

$$\mathcal{L}(\Theta) = \frac{1}{z} \sum_i (y_i - \hat{y}_i)^2 \quad (11)$$

其中, Θ 表示模型的所有可训练参数, y_i 和 $\hat{y}_i$ 分别表示真实值和预测值, z 表示训练样本的总数.

4.3 评价指标

本文采用了3种适用于共享出行需求预测任务的评价指标,分别为均方根误差(Rooted Mean Square Error, RMSE)、平均百分比误差(Mean Average Percentage Error, MAPE)和平均绝对误差(Mean Absolute Error, MAE),计算式为

$$\text{RMSE} = \sqrt{\frac{1}{z} \sum_{i=1}^z (y_i - \hat{y}_i)^2} \quad (12)$$

$$\text{MAPE} = \sqrt{\frac{1}{z} \sum_{i=1}^z \frac{|y_i - \hat{y}_i|}{y_i}} \quad (13)$$

$$\text{MAE} = \frac{1}{z} \sum_{i=1}^z \frac{|y_i - \hat{y}_i|}{y_i} \quad (14)$$

其中, y_i 和 $\hat{y}_i$ 分别表示真实值和预测值, z 表示测试样本的总数.

4.4 与基准模型的对比

将STDGAT分别与历史平均(Historical Average, HA)、自回归移动平均(Autoregressive Integrated Moving Average, ARIMA)、Lasso回归(Lasso regression, Lasso)、岭回归(Ridge regression, Ridge)、XGBoost^[25]、多层感知机(Multiple Layer Perception, MLP)、DMVST-Net^[17]、DCRNN^[3]、STGCN^[2]、ST-MGCN^[1]基准预测方法进行实验对比分析. 表1展示了基准模型和STDGAT的实验结果. 从表中可以看出,STDGAT在3个评价指标上均达到了最佳效果. HA和ARIMA的预测误差最大,这是由于这2个模型只能依靠历史真实的数据值来进行预测,并没有考虑任何相关特征的提取. 由于考虑了序列中更多的上下文信息,线性回归模型(Lasso和Ridge)的表现要优于HA和ARIMA. 但是,线性回归模型同样无法提取更多的特征来进行预测,因此线性回归方法的预测误差依然较大. XGBoost和MLP进一步对输入序列进行了变换,并提取了序列中的隐藏特征,因此,它们的表现相比于上述4种方法,得到了进一步的提升. 但是,XGBoost和MLP依然没能在空间维度或时间维度上对数据进行建模.

4个深度学习方法(DMVST-Net, DCRNN, STGCN, ST-MGCN)进一步考虑了时空依赖,因此它们达到了较为理想的效果. 与这3种方法相比,由于考虑了不同邻

居区域间的自适应空间相关性和具有时间特定性的动态空间图结构,STDGAT的表现更加优异. 具体而言,DMVST-Net在提取全局空间特征时,仅仅将一个固定的静态图嵌入到一个向量中,这种方式对于提取自适应和动态空间相关性具有局限性. DCRNN, STGCN和ST-MGCN对区域间的非欧氏空间相关性进行了建模. 但是,由于这两种方法在对非欧氏空间相关性进行建模时,使用的是传统的GCN模型,因此这两种方法只能使每个区域为其不同的邻居区域分配相同的重要性系数. 此外,DCRNN, STGCN与ST-MGCN并没有考虑具有时间特定性的动态空间图结构,因此,它们的效果相比于STDGAT要略显逊色.

表1 与基准模型的对比

方法	RMSE	MAPE	MAE
HA	24.162 9	0.471 0	13.583 5
ARIMA	22.544 0	0.376 9	12.476 2
Lasso	19.585 1	0.308 0	11.308 2
Ridge	19.596 8	0.306 9	11.317 8
XGBoost	11.314 3	0.257 2	6.574 0
MLP	10.475 3	0.236 7	5.680 6
DMVST-Net	8.869 5	0.191 4	4.536 5
DCRNN	8.783 6	0.189 3	4.438 9
STGCN	8.615 3	0.190 9	4.557 9
ST-MGCN	8.519 0	0.182 7	4.402 2
STDGAT	7.881 1	0.174 4	4.088 1

4.5 在不同时间的实验结果

为探讨在不同时间段上预测效果的差异,本文分别在不同时间段中进行了实验.

图5展示了在一周7天中,STDGAT和所有基准模型的表现. 图6展示了在一周7天中,STDGAT与4个深度学习基准模型的表现. 如图中所示,STDGAT在各个时间段上都取得了最佳的表现,这证明了STDGAT具有很好的鲁棒性. 表2展示了STDGAT和基准模型分别在

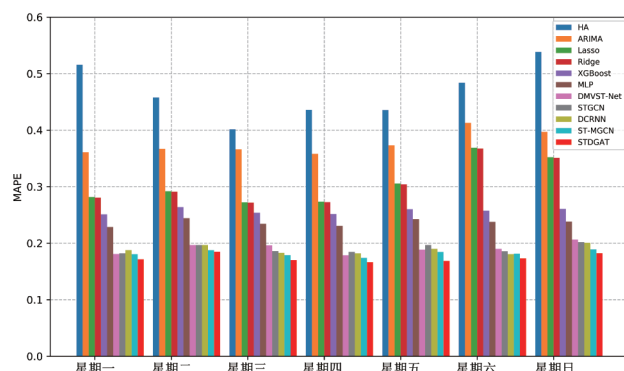


图5 模型在不同天的表现

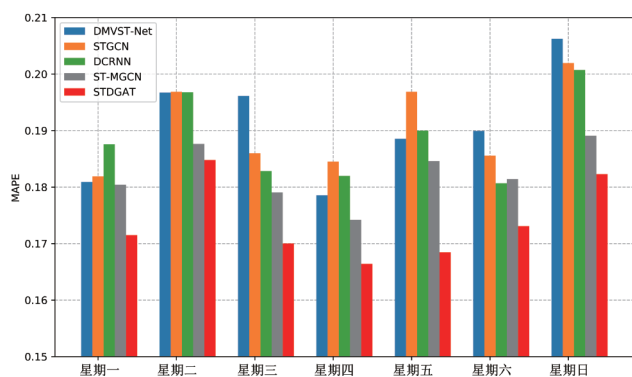


图6 深度学习模型在不同天的表现

工作日和周末的实验结果。如表中所示,STDGAT无论是在工作日还是周末都具有最佳的效果。

但实验结果中所有方法在工作日的表现均优于在周末的表现。其中的原因在 Yao 等人^[17]的工作中得到了相关解释,即相比于工作日的需求模式,周末的需求模式要更加不规则。由于规则的需求模式更加有利于模型去学习,因此在工作日中,模型的表现要更加优于在周末的表现。

表2 工作日与周末的实验结果(MAPE)

方法	工作日	周末
HA	0.452 7	0.513 2
ARIMA	0.365 0	0.404 4
Lasso	0.285 4	0.360 0
Ridge	0.284 4	0.358 6
XGBoost	0.256 3	0.259 3
MLP	0.236 2	0.238 1
DMVST-Net	0.188 2	0.198 7
DCRNN	0.188 3	0.191 4
STGCN	0.189 4	0.194 4
ST-MGCN	0.181 5	0.185 5
STDGAT	0.172 8	0.178 0

4.6 STDGAT的变体模型

STDGAT有2个不同的特征提取模块:空间模块和时间模块。为了探索这2个模块对预测结果的影响,本文构建如下变体模型。(1) Spatial module + Prediction layer. 该网络由STDGAT的空间模块和输出预测层组成。该网络仅提取空间特征来进行预测,以此来探讨在缺乏时间特征提取时模型的性能。(2) Temporal module + Prediction layer. 该网络由STDGAT的时间模块和输出预测层组成。该网络仅提取时间特征来进行预测,以此来探讨在缺乏空间特征提取时模型的性能。

表3展示了上述变体模型与完整版STDGAT的实验结果对比。在缺少了空间模块或时间模块时,STDGAT的效果会得到大幅度下降。由此可见,空间模块和时间模块在STDGAT中均具有重要的作用,两者缺一不可。

表3 不同模块的表现对比

方法	RMSE	MAPE	MAE
Spatial module + Prediction layer	9.627 3	0.214 3	5.213 1
Temporal module + Prediction layer	9.315 1	0.194 6	4.596 4
STDGAT	7.881 1	0.174 4	4.088 1

实验也表明了STDGAT在共享出行需求预测任务中达到了优异的效果,也验证了不同特征提取模块的合理性和有效性。

针对基于通勤关系的时间特定性空间图结构,本文构建了STDGAT-fixed变体模型进行实验对比,以此来验证其有效性。

STDGAT-fixed是STDGAT的变体,在提取空间特征时,该模型使用了一种在不同时间间隔中保持固定的静态空间图结构,这种图结构是基于真实的区域地理邻接关系构建的,以此来代替STDGAT中基于通勤关系的时间特定性空间图结构。

STDGAT-fixed与STDGAT的实验结果对比如表4所示。其中,STDGAT的3个评价指标上的表现均优于STDGAT-fixed,这意味着本文所提出的基于通勤关系的时间特定性空间图结构在共享出行需求预测任务中,具有比传统的基于区域真实地理邻接关系的图结构更加优异的表现。这也证明了在图学习中,相比于静态的空间图结构,动态空间图结构是一种更加可行有效的图表示学习方式。

表4 STDGAT与变体模型STDGAT-fixed的对比

方法	RMSE	MAPE	MAE
STDGAT-fixed	8.233 4	0.178 8	4.191 7
STDGAT	7.881 1	0.174 4	4.088 1

4.7 序列长度与GAT层数对实验结果的影响

实验分别探讨了输入到模型中的序列长度与GAT层数对预测结果的影响。

图7展示了输入序列长度对预测结果的影响。如图中所示,当输入序列长度为5时,模型达到最佳性能。当输入序列长度小于5时,预测误差随着序列长度的增加而减小,这表示随着序列长度的增加,模型可以学习到更长的时间依赖性信息,这些信息有助于对结果的预测。但是当输入序列长度大于5时,模型的效果会有略微下降,并伴随着一些波动。造成这种现象的一个潜在原因是,随着输入序列长度的增加,输入到模型中的信息量呈线性增长,因此模型需要将更多的信息与梯度计算过程联系起来,并会造成一定程度上的信息丢失,因此训练过程会变得更加困难。

图8展示了空间模块中GAT Block中的GAT层数对实验结果的影响。从图中可以看出,随着GAT层数由0增加到5, RMSE值逐渐降低。这意味着,随着GAT

层数的增加,模型的效果变得更好.造成这种现象的原因可解释为,随着层数的加深,原始的特征将进一步和其邻居中的特征进行聚合,这将使较深的层具有更大的感受野.由于更大的感受野可以捕获更多的空间关联性,因此,随着层数的加深,模型可以学习更多的空间信息来提高其性能.

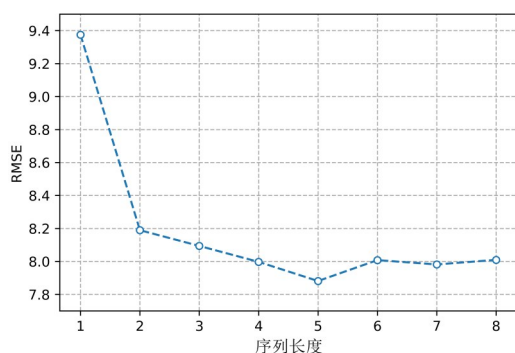


图7 不同输入序列长度的实验结果(RMSE)

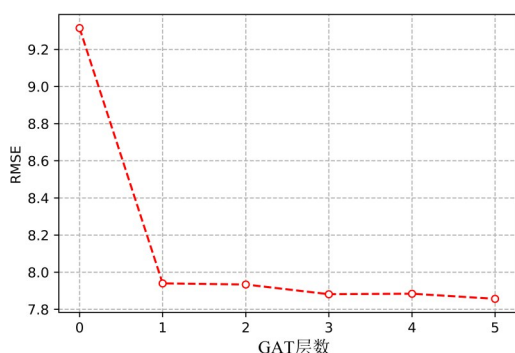


图8 不同GAT层数的实验结果(RMSE)

5 结论

本文提出了一种用于共享出行需求预测的时空动态图注意力网络.针对共享出行需求预测在空间特征提取时存在的非自适应邻居区域重要性分配和静态空间相关性建模的问题,本文基于图注意力网络和时间特定性动态空间图结构有效地解决了这2个问题,并通过大量的实验验证了模型的有效性和合理性,为未来的相关工作提供了参考.

参考文献

- [1] GENG X, LI Y, WANG L, et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii, USA: AAAI, 2019, 33: 3656-3663.
- [2] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[C]//Proceedings of the 27th International Joint

- Conference on Artificial Intelligence. Stockholm, Sweden: Morgan Kaufmann, 2018: 3634-3640.
- [3] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting [C]//International Conference on Learning Representations. Vancouver, Canada: Vancouver Convention Center, 2018: 1-16.
- [4] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//International Conference on Learning Representations. Vancouver, Canada. 2018: 1-12.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Annual Conference on Neural Information Processing Systems(NeurIPS). Long Beach, USA: MIT Press, 2017: 5998-6008.
- [6] MOREIRA-MATIAS L, GAMA J, FERREIRAM, et al. Predicting taxi-passenger demand using streaming data[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(3): 1393-1402.
- [7] SHEKHAR S, WILLIAMS B M. Adaptive seasonal time series models for forecasting short-term traffic flow[J]. Transportation Research Record, 2007, 2024(1): 116-125.
- [8] 胡文斌, 聂聪, 邱振宇, 等. 一种城市交通路网实时动态多路口路径导航量子搜索方法[J]. 电子学报, 2018, 46(1): 104-109.
- HU W B, NIE C, QIU Z Y, et al. A route guidance method based on quantum searching for real-time dynamic multi-intersections in urban traffic networks[J]. Acta Electronica Sinica, 2018, 46(1): 104-109. (in Chinese)
- [9] ZHANG J, ZHENG Y, QI D, et al. DNN-based prediction model for spatio-temporal data[C]//Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Burlingame, USA: ACM, 2016: 1-4.
- [10] ZHANG J, ZHENG Y, QI D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI, 2017: 1655-1661.
- [11] PAN B, DEMIRYUREK U, SHAHABI C. Utilizing real-world transportation data for accurate traffic prediction [C]//2012 IEEE 12th International Conference on Data Mining. Brussels, Belgium: IEEE, 2012: 595-604.
- [12] WU F, WANG H, LI Z. Interpreting traffic dynamics using ubiquitous urban data[C]//Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Burlingame, USA: ACM, 2016: 1-4.

- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, USA: IEEE, 2016: 770-778.
- [14] WANG D, CAO W, LI J, et al. DeepSD: Supply-demand prediction for online car-hailing services using deep neural network[C]//2017 IEEE 33rd International Conference on Data Engineering. San Diego, USA: IEEE, 2017: 243-254.
- [15] YU R, LI Y, SHAHABI C, et al. Deep learning: A generic approach for extreme condition traffic forecasting [C]//Proceedings of the 2017 SIAM international Conference on Data Mining. Houston, USA: SIAM, 2017: 777-785.
- [16] ZHAO J, XU J, ZHOU R, et al. On prediction of user destination by sub-trajectory understanding: A deep learning based approach[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Lingotto, Italy: ACM, 2018: 1413-1422.
- [17] YAO H, WU F, KE J, et al. Deep multi-view spatial-temporal network for taxi demand prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI, 2018: 1-9.
- [18] QIU Z, LIU L, LI G, et al. Taxi origin-destination demand prediction with contextualized spatial-temporal network[C]//2019 IEEE International Conference on Multimedia and Expo. Shanghai, China: IEEE, 2019: 760-765.
- [19] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Annual Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014: 3104-3112.
- [20] HOCHREITER S, BENGIO Y, FRASCONI P, et al. Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-term Dependencies[M]. Piscataway, NJ, USA: Wiley-IEEE Press, 2001.
- [21] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [22] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Annual Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015: 802-810.
- [23] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]//International Conference on Learning Representations. San Diego, USA: JMLR, 2015: 1-15.
- [24] PASZKE A, GROSS S, MASSA F, et al. PyTorch: An

imperative style, high-performance deep learning library [C]//Advances in Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2019: 8026-8037.

- [25] CHEN T, GUESTRIN C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining. San Francisco, USA: ACM, 2016: 785-794.

作者简介



骈纬国 男, 1996年出生于河北省. 现为重庆大学在读硕士研究生. 主要研究方向为时空数据挖掘.
E-mail: pwg@cqu.edu.cn



吴映波(通讯作者) 男, 1978年出生于湖北省. 现为重庆大学大数据与软件学院教授. 主要研究方向为分布式智能计算.
E-mail: wyb@cqu.edu.cn



陈蒙 男, 1989年出生于重庆市. 现为重庆大学大数据与软件学院助理研究员. 主要研究方向为云计算、智能网卡卸载.



蔡俊鹏 男, 1994年出生于福建省. 现为重庆大学大数据与软件学院在读博士生. 主要研究方向为服务计算.