

文章编号: 2095-2163(2022)11-0253-06

中图分类号: TP183,U491.1

文献标志码: A

时空注意力卷积网络的交通流量预测研究

李志鹃

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 准确实时的交通预测在交通管理和规划中起重要作用,而交通数据是典型的时空数据,目前大部分的预测方法没有完全考虑交通数据的动态时空相关性、周期性和稳定预测性能等要素。针对这些要素,本文提出了一种时空注意力卷积网络(spatial-temporal attention graph convolutional network, STAGCN),该网络采用了Encoder-Decoder框架,其中 k 阶切比雪夫(ChebNet)图卷积网络用于提取交通网络的空间相关性,同时将历史时间步长和与预测时间段间隔相同的周交通数据堆叠作为输入,更好地捕获时间和周期性。在解码器中加入时间注意机制,自适应地学习编码器输出中更重要的部分。此外,设计了残差网络稳定预测性能。最后,在2个真实的交通数据集上进行大量实验,结果表明STAGCN优于其他现有的方法。

关键词: 交通预测; 图卷积网络; 时空相关性; 注意机制; 编码器-解码器

Research on traffic flow prediction of spatial-temporal attention graph convolutional networks

LI Zhijuan

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] Real-time traffic prediction plays an important role in traffic management and planning. Traffic data is typical spatial-temporal data. However, most forecasting methods do not fully consider the dynamic spatial-temporal correlation, periodicity and stable forecasting performance of traffic data. In this paper, an spatial-temporal attention graph convolutional network(STAGCN) is proposed, which adopts the Encoder-Decoder framework. The k -order Chebyshev graph convolution network(ChebNet) is used to extract the spatial correlation of the traffic network. The weekly traffic data stack with the same historical time step as the predicted time interval is used as input to better capture periodicity. The attention mechanism is added to the decoder to adaptively learn the more important part of the encoder output. Furthermore, the residual network is designed to stabilize prediction performance. A large number of experiments on two real traffic datasets demonstrate that STAGCN is better than other existing baselines in traffic flow prediction.

[Key words] traffic prediction; graph convolutional networks; spatial-temporal correlation; attention mechanism; encoder-decoder

0 引言

交通预测是基于历史交通数据对未来时刻的交通状况(如交通速度、交通流量和交通占用率等)进行预测。准确实时的交通预测有助于交通规划和管理,进而缓解交通拥堵^[1]。

近年来,国内外的研究者都致力于交通预测的研究且取得不错的成果。研究大致可以分为传统方法和深度学习方法。传统方法主要代表为历史平均(historical average, HA)、差分自回归移动平均模型(auto-regressive integrated moving average model, ARIMA)^[2]、向量自回归(vector auto-regressive, VAR)^[3],这种方法不能捕获数据复杂的非线性关系,特征提取能力有限。2009年,Castro等人^[4]提出

支持向量回归(Support Vector Regression, SVR)方法用于交通流量预测,该方法分析了交通流的非线性特征,但挖掘时空关系有限。随着深度学习的发展,交通预测方法得到进一步改进。2014年,Huang等人^[5]提出深度信念网络(Deep Belief Nets, DBN)。2015年,Lv等人^[6]提出深度学习模型、自动堆叠编码器(Stacked Autoencoder, SAE)用于预测交通流量,以上2种方法均采用栅格模型,只能应用于标准网格数据且忽略了数据采集点本身的空间关系。

为更好地建模交通网络的空间关系,近期研究将图神经网络(graph convolutional neural networks, GCN)应用于交通预测方法^[7-9]。2018年,Yu等人^[10]提出了一种深度学习框架、时空图卷积网络

作者简介: 李志鹃(1997-),女,硕士研究生,主要研究方向:深度学习、智能交通预测。

通讯作者: 李志鹃 Email:LZJ296949658@163.com

收稿日期: 2022-03-15

(STGCN),利用GCN捕获空间相关性,利用卷积提取时间特征。2019年,Guo等人^[11]提出(Attention based spatial-temporal graph convolutional network, ASTGCN)模型,从注意机制的角度出发,该网络使用2个注意力层分别捕获空间依赖性和时间相关性。2020年,Song等人^[12]的研究提出一种可以同步提取局部时空关系的时空模型(spatial-temporal synchronous graph convolutional network, STSGCN)。

受上述研究工作的启发,为了更好地捕获交通网络的时空相关性,进一步提高预测性能,本文提出一种时空注意力卷积网络(spatial-temporal attention graph convolutional network, STAGCN)的交通预测模型,该方法能够捕获交通数据复杂的动态时空特征。本文主要贡献如下:

(1)设计了一种Encoder-Decoder框架,具体来说,编码器和解码器均采用门控循环单元(gate recurrent unit, GRU),采用 k 阶ChebNet图卷积^[13]对历史交通数据的空间依赖关系建模,在解码器中加入时间注意机制,自适应地学习编码器输出中更重要的部分,动态学习时空相关性。

(2)提出将历史时间步长和与预测时间段间隔相同的周交通数据堆叠作为输入,更好地捕获时间和周期性。

(3)利用残差网络稳定预测性能,并在2个真实交通数据集上进行消融实验,验证了残差、 k 阶ChebNet图卷积和Encoder-Decoder框架的有效性。

1 整体设计

1.1 问题描述

在本文中,交通数据包含 N 个道路节点(传感

器)、 T 个时间步、 C 个特征。文中将交通道路网络的拓扑结构定义为图 $G=(V,E,A)$,其中 V 是一组节点(如交通传感器); E 是表示各节点之间的连接性的边; $A \in R^{N \times N}$ 是邻接矩阵,表示节点之间的依赖程度(按路网距离测量)。 t 时刻节点 i 的交通流量数据用 $x_{t,i} \in R^C$ 表示,则 $X_t=(x_{t,1},x_{t,2},\dots,x_{t,N}) \in R^{N \times C}$ 表示 t 时刻所有节点的特征,如交通速度、交通流量和交通占用率等。本文研究主题为:

利用 N 个观测点的 p 个历史时间步长的数据预测未来时间步长为 q 的所有节点的交通流量。即输入为 $\mathbf{X}=(X_1,X_2,\dots,X_p)^T \in R^{N \times C \times p}$,预测输出表示为 $\hat{\mathbf{Y}}=(\hat{y}_1,\hat{y}_2,\dots,\hat{y}_q)^T \in R^{q \times N}$ 。即:

$$(\hat{y}_1,\hat{y}_2,\dots,\hat{y}_q)=f(X_1,X_2,\dots,X_p;G) \quad (1)$$

1.2 网络结构

本研究提出的基于时空注意力卷积网络的交通流量预测模型(STAGCN)框架如图1所示。是由GRU构成Encoder-Decoder框架,在编码器部分,采用 k 阶的ChebNet图卷积对历史数据的空间依赖关系进行建模,将 T_w 周与预测时段相同时间间隔的交通数据 X_w 及输入数据 X 进行堆叠作为模型的输入,同时可以捕获历史数据的周期性。解码器部分引入注意机制动态学习编码器不同时刻的隐藏层状态。此外,加入残差网络模块提升网络的性能。由ChebNet图卷积、注意机制的编-解码器和残差网络共同组成时空注意模块,堆叠 N 个上述时空注意模块,最后通过卷积预测层预测未来道路的交通流量。

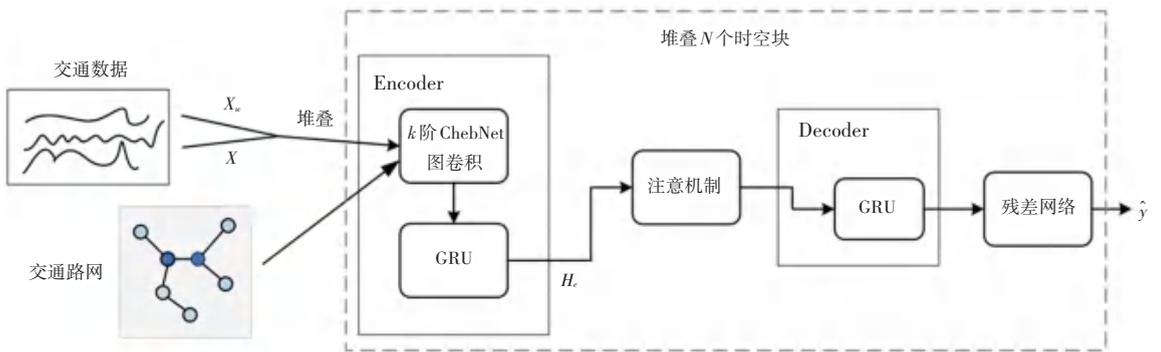


图1 STAGCN总体框架图
Fig. 1 The architecture of STAGCN

1.2.1 空间相关性

道路上不同地点的交通流量是动态变化的,节

点上的流量不仅受之前自身交通状况影响,而且也与相邻节点的交通量息息相关。图2是交通节点空

间相关性示意图。

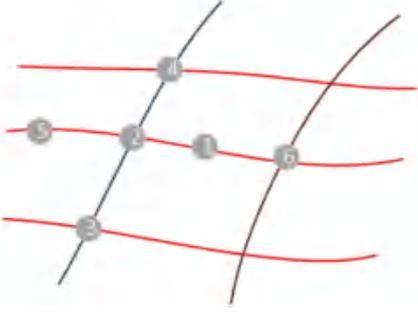


图 2 交通节点空间相关性示意图

Fig. 2 The diagram of spatial correlation of traffic nodes

以节点 1 为例,其流量受到节点 2,3,4 和节点 5 的影响,其中 2 号节点的影响程度最大,因其在空间网络上更接近 1 号节点。本文为更好地捕获节点间的空间相关性,应用 ChebNet 图卷积捕获交通数据的空间依赖关系。与图卷积神经网络相比, ChebNet 图卷积利用 k 阶多项式拟合卷积核的方法,在降低计算复杂度的同时增大了感受野。 k 阶 ChebNet 多项式函数表示如下:

$$\text{chebnet}(x) = \text{ReLU}(g_\theta * x) \quad (2)$$

$$g_\theta * x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x \quad (3)$$

$$\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N \quad (4)$$

$$L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \quad (5)$$

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x) \quad (6)$$

$$T_0(x) = 1, T_1(x) = x \quad (7)$$

其中, θ_k 是可训练参数; L 是邻接矩阵 A 和度矩阵 D 构成的对称规范化拉普拉斯矩阵; λ_{\max} 表示 L 的最大特征值; $T_k(x)$ 是切比雪夫多项式的定义; 因切比雪夫多项式自变量范围为 $[-1, +1]$, 故使用 \tilde{L} 。

本文通过调整 k 阶切比雪夫多项式的阶数,调整感受野,捕获 k 阶邻居节点的空间相关性,确定空间特征的影响权重。

1.2.2 时间相关性

交通数据在时间维度上也存在动态相关性,未来 $t+1$ 时刻的交通流量可能受 t 时刻的交通状况影响较大,若 $t-\alpha$ 时刻发生交通事故或社会事件,则受之前 $t-\alpha$ 时刻的影响可能更大。此外,不同道路上,相同时间间隔的历史交通流量对未来时刻流量的影响也是不同的。为此,本文采用注意机制来动态捕获不同地点、不同时间间隔的依赖性。具体而言,利用 Attention 机制的 Encoder-Decoder 框架,

在解码器部分动态调整编码器不同地点、不同时刻的隐藏层状态对预测结果的影响权重。Attention 机制计算如下:

$$\alpha_i = \tanh(W_e H_e + W_d H_{(t-1),d}) \quad (8)$$

$$\beta_i = \frac{\exp(\alpha_i)}{\sum_{k=1}^T \exp(\alpha_k)} \quad (9)$$

$$w_i = \beta_i H_e \quad (10)$$

$$x_d = \text{Decoder}(x_t, w_i, H_{(t-1),d}) \quad (11)$$

其中, \tanh 是激活函数; W_e, W_d 是可训练参数; $H_e = [H_{0,e}, H_{1,e}, \dots, H_{T,e}]$ 是各个时刻隐藏层的状态; T 是编码器隐藏层状态数; $H_{(t-1),d}$ 是解码器在 $t-1$ 时刻隐藏层的状态; β_i 是注意系数向量; w_i 表示加权注意权重; x_t 表示 t 时刻解码器的输入; x_d 表示解码器的输出。

1.2.3 周期性

据生活规律来看,工作日的早高峰和晚高峰会有一定的相似性,而与周末的交通流模式不同。为更好地捕获这一周期性,本文将交通数据中过去几周与预测时段相同时间间隔的 T_w/p 周 (T_w 为时间步 p 的整数倍) 数据进行单独划分,如图 3 所示。

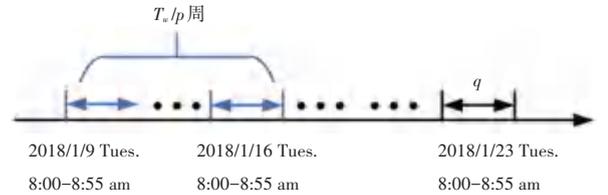


图 3 T_w/p 周划分示意图

Fig. 3 The schematic diagram of T_w/p

将 T_w/p 周与预测时段相同时间间隔的交通数据 x_w 作为编码器输入的一部分,提取交通数据的周期特征。

接下来,模型中还加入了稳定性能的残差网络层,而后通过卷积预测层输出:

$$x_r = \text{ReLU}(x_d + \text{ReLU}(\Phi * (x_d))) \quad (12)$$

$$\hat{y} = \text{Conv2d}(x_r) \quad (13)$$

其中, ReLU 是激活函数; x_d 表示经过解码器后的输出; “*”, “ Φ ” 分别是卷积操作和卷积核参数; x_r 是残差网络层的输出。

具体来说,本文中的残差网络同时使用了先降维减少计算的卷积层和还原输入维度的卷积层,这样可以在减少计算量的同时保证精度。研究推得,模型的预测结果为 \hat{y} 。

2 实验及分析

2.1 数据集

本文在 PeMSD4, PeMSD8 两个真实高速公路交通数据集上进行了大量的实验,数据由加州高速路网 Caltrans Performance Measurement System (PeMS)^[14]收集。路网中收集实时数据的独立传感器超过 39 000 个,且每 5 min 汇总一次交通数据,即 1 h 有 12 组数据。具体信息见表 1。

表 1 PeMS 数据集描述

Tab. 1 PeMS dataset description

数据集	节点数	数据长度	时间跨度
PeMSD4	307	16 992	2018/1/1~2018/2/28
PeMSD8	170	17 856	2016/7/1~2016/8/31

采用与文献[11]中相同的数据预处理步骤,初步预处理后,数据集中若有缺失值,则进行线性插值填充并通过 Z-Score^[15]方法进行标准化。本文将数据集按照 6:2:2 的比例分为训练集、验证集和测试集进行实验。

2.2 实验设置

本实验在 Windows10 操作系统下 (CPU: Inter (R) Core (TM) i9-10900X @ 3.70 GHz GPU: NVIDIA GeForce RTX 3090), 基于 Pytorch 深度学习框架,使用了 CUDA10.1+cuDNN 进行 GPU 加速。

在研发的模型中,利用 $p = 12$ 个历史时间步长来预测未来 $q = 12$ 个时间步长(即未来 1 h)的交通流量。本文中选择 L_2 loss 作为损失函数,初始学习率为 0.001,采用 Adam 优化器进行训练。模型中所有的卷积层都使用 32 个卷积核,批处理大小为 32,门控循环单元 (GRU) 的单元数为 32,层数为 2,周序列长度 T_w 取 12。此外,据文献[11]的研究启发,测试了切比雪夫多项式阶数 $k \in \{1, 2, 3\}$ 和时空注意模块 $N \in \{1, 2, 3\}$ 对预测结果的影响,随着 k 和 N 的增大,性能略有提高,综合计算效率和预测性能提升的效果,选择 $k = 2, N = 2$ 的参数配置。为防止训练过拟合,本文使用了 EarlyStopping 方法, $patience$ 设为 20。

2.3 评价指标

本文采用 3 个常用于评估流量预测性能的评测指标:平均绝对误差 (MAE)、均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE)。其定义分别如下:

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (15)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (16)$$

其中, n 是样本数; \hat{y}_i 表示预测值; y_i 表示真实值。

2.4 基线

将本文的模型与 8 种模型方法进行对比,对这 8 种模型可概述如下。

(1) HA (historical average): 历史平均模型,使用平均值作为预测结果。

(2) ARIMA (Williams and Hoel, 2003)^[2]: 自回归移动平均模型。

(3) VAR (vector auto-regressive)^[3]: 捕获交通序列间成对相关性的向量自动回归模型。

(4) LSTM (Hochreiter and Schmidhuber, 1997)^[16]: 长短时记忆网络,一种特殊的循环神经网络。

(5) DCRNN (Li et al., 2018)^[17]: 利用扩散卷积捕获时空特性的循环神经网络。

(6) STGCN (Yu et al., 2018)^[10]: 一种基于空间方法的时空图卷积模型。

(7) ASTGCN (Guo et al., 2019)^[11]: 使用注意力层和卷积网络提取时空特性。

(8) STSGCN (Song et al., 2020)^[12]: 一种用于结合不同时间、节点间信息的时空同步机制。

2.5 实验结果对比与分析

本文提出的 STAGCN 模型与基线模型在 PeMSD4、PeMSD8 数据集上的预测性能对比见表 2。

表 2 不同模型在 PeMS 数据集上的平均性能

Tab. 2 Average performance comparison of different algorithms on the PeMS dataset

模型	PeMSD4			PeMSD8		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
HA	36.76	54.14	21.83	29.52	44.03	16.59
ARIMA	35.19	58.05	20.68	24.02	43.23	15.89
VAR	33.63	51.62	19.73	19.19	29.81	13.10
LSTM	29.32	45.72	18.59	23.10	36.89	13.15
DCRNN	21.22	37.23	14.17	16.82	26.36	10.92
STGCN	21.16	35.69	13.83	17.50	27.09	11.29
ASTGCN	21.48	33.70	14.88	18.70	28.41	12.37
STSGCN	21.52	34.14	14.46	17.88	27.36	11.71
STAGCN	20.29	32.57	13.73	15.15	24.49	10.57

从表 2 中可以看出, STAGCN 模型在 PeMSD4、PeMSD8 两个数据集上的预测性能都达到了最优, 证明本文的模型是有效的、且对数据的适应性强。传统的预测方法 HA、ARIMA 仅考虑了时间相关性、而忽略了交通数据的空间特征, 因此预测效果不好。神经网络方法 LSTM 在 PeMSD4 数据集上的预测误差小于 VAR 等传统方法, 但在 PeMSD8 数据集上预测误差大于 VAR, 其模型适应性不佳。DCRNN 等基于图结构的预测方法效果较好, DCRNN 利用循环神经网络捕获时间依赖性, 在 PeMSD8 数据集上的效果比 STGCN 方法更好, 但 PeMSD4 数据集较差, 说明模型对交通数据的适应性不佳。ASTGCN 与 STSGCN 方法较之前的方法又有所改进, 但是捕获长时空依赖性的能力有限, 预测效果还有待改进。

2.6 消融分析

为进一步看出 STAGCN 中各模块的影响, 下面进行消融对比实验。基于本文研发的 STAGCN 模型, 设计了 3 个变种, 详见如下。

- (1) STAGCN-C: 移除了 ChebNet 图卷积模块。
- (2) STAGCN-E: 移除了 Encoder-Decoder 框架。
- (3) STAGCN-R: 移除了残差网络。

图 4、图 5 分别是在 PeMSD4、PeMSD8 数据集进行的变种模型实验的预测误差条形图。

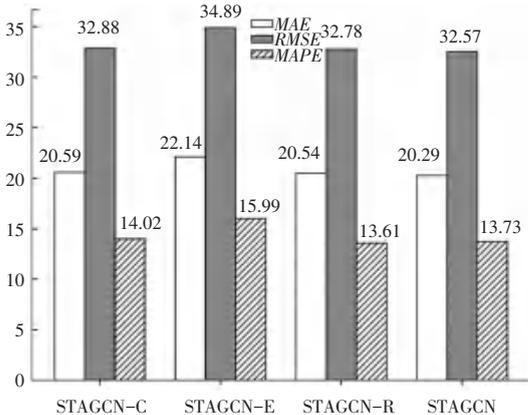


图 4 各个变种在 PeMSD4 数据集上的性能比较

Fig. 4 Performance comparison of variant models on PeMSD4 dataset

从图 4、图 5 中可以看到, 本文研究提出的模型中各模块都能发挥出作用, 提高预测性能。其中, Encoder-Decoder 框架对模型的影响最大, 对提高预测精度的能力更强, 这说明解码器中添加注意机制能捕获不同时刻的重要性, 自适应地调整权重。 k 阶 ChebNet 图卷积影响位居其后, 残差网络又次之, 这是因为考虑到计算效率和预测效果提高的程度, 时

空模块和网络深度并未过大, 残差网络的作用不明显。综上, 本文提出的 STAGCN 模型中各个模块都有效果, 能够帮助模型更好地进行预测任务。

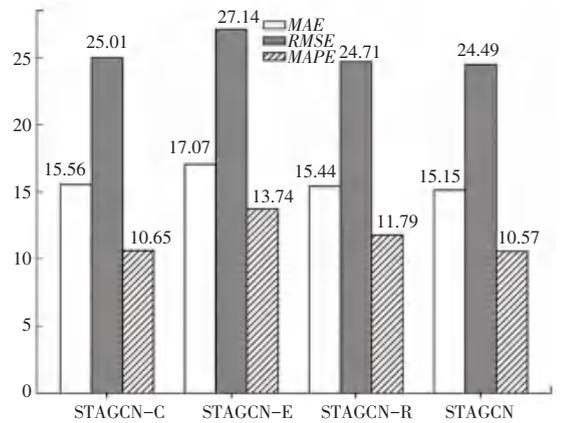


图 5 各个变种在 PeMSD8 数据集上的性能比较

Fig. 5 Performance comparison of variant models on PeMSD8 dataset

3 结束语

本文提出了一种时空注意力卷积模型 STAGCN 应用于交通流量预测。STAGCN 是一个应用 Encoder-Decoder 框架, 其中采用 k 阶 ChebNet 图卷积网络提取交通流的空间特征、同时将过去的周期交通数据作为编码器输入的一部分, 捕获交通数据的周期性。在解码器部分, 利用注意机制学习编码器不同隐藏状态的权重, 基于此还设计了残差网络模块稳定模型的性能, 避免因增加网络深度引起性能降低的情况发生。在 2 个真实的交通数据集上进行大量的实验, 结果表明, 本文提出的 STAGCN 模型表现要优于其他的模型, 可以有效地捕获交通流的动态时空特性。

本文的模型在训练效率上略显不足, 在未来的工作中, 会将其考虑进去, 同时增加现实不可控因素 (如天气和社会事件等), 提高预测任务的效率和精度。此外, 可以进一步地将研发的 STAGCN 模型应用在别的时空数据预测中。

参考文献

- [1] ZHANG Junping, WANG Feiyue, WANG Kunfeng, et al. Data-driven intelligent transportation systems: A survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4): 1624-1639.
- [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.

(下转第封三)