

基于自注意力机制 TCN-BiGRU 的交通流预测^{*}

郝椿琳 张 剑

(上海工程技术大学航空运输学院 上海 201620)

摘要: 为更精准地预测道路交通流,本文提出了基于自注意力机制 TCN-BiGRU 的交通流预测模型。该预测模型首先利用时间卷积网络(TCN)的卷积特性跨时间步的提取交通流数据中的时间相关性;其次,利用 BiGRU 双向捕捉交通流的时间相关特性,经过更新门和复位门后更全面的提取交通流的时间特性;考虑到双向门控循环单元在双向计算过程中存在有并行性较低和部分特征无法捕捉的情况,引入自注意力机制能够让模型能够注意到全局中不同输入之间的相关性,让模型能够不受序列长度限制的特征捕捉的难题,最大限度的保留特征进而提高模型的鲁棒性,最终得到交通流的预测值。为验证模型的适用性,本文选取真实的交通数据进行多组预测对比实验,在单一路段将预测结果与基准模型和多路段的经典模型以及消融进行对比,结果表明基于自注意力机制 TCN-BiGRU 对于多特征的单一路段或多路段的预测结果表现为:单一路段的 MAE, MAPE/%, R^2 平均值分别为 15.91, 10.89, 0.976;多路段的 MAE, MAPE/%, R^2 分别为 19.62, 13.53, 0.982, 具有较好的预测效果,所建立的组合预测模型在预测精度上表现出更好的水平,为交通流的预测提供了良好的参考价值。

关键词: 交通流;时间卷积网络;双向 GRU;自注意力机制

中图分类号: U491.14 文献标识码: A 国家标准学科分类代码: 580.2030

Traffic flow prediction based on self-attention mechanism TCN-BiGRU

Hao Chunlin Zhang Jian

(School of Air Transport, Shanghai University of Engineering and Technology, Shanghai 201620, China)

Abstract: In order to enhance the accuracy of road traffic flow prediction, this paper proposes a traffic flow prediction model based on the self-attention mechanism TCN-BiGRU. Firstly, the predictive model utilizes the convolutional property of time convolutional network (TCN) to extract temporal correlations within traffic flow data across different time steps. Secondly, bidirectional BiGRU is employed to comprehensively capture time-related characteristics of traffic flow by updating and resetting gates. Recognizing that bidirectional gated recurrent unit has limited parallelism and may not capture certain features during bidirectional calculation, the introduction of self-attention mechanism allows the model to focus on global correlation between different inputs, overcoming limitations posed by sequence length in feature capture and maximizing feature retention for improved robustness. Finally, predicted values for traffic flow are obtained. To validate the applicability of the model, real traffic data is selected for multiple prediction and comparison experiments against benchmark models and ablation experimental models across various road segments. The results demonstrate that single or multiple section predictions using multiple features based on self-attention mechanism TCN-BiGRU yield mean MAE values of 15.91 and 19.62 respectively; MAPE/% values of 10.89 and 13.53 respectively; as well as R^2 values of 0.976 and 0.982 respectively-indicating strong predictive performance.

Keywords: traffic flow; temporal convolutional network; bi-directional gated recurrent unit; self-attentive mechanism

0 引言

随着智能交通的不断发展,道路总量不变的情况下所产生的交通拥堵成为当下迫切需要解决的问题,准确实时

的交通流预测可以很好的为交通管理者提供决策意见。智能交通系统(intelligent transportation system, ITS)在深度学习的大环境下迅速发展^[1],通过交通流的预测形成对交通的诱导控制和信号管理,可以有效地提升人、车、路、环境

之间的舒适度^[2]。因此如何确保交通流预测的精度成为了当下交通领域的研究重点方向之一。

传统的基于数理统计的模型的短时客流预测有历史平均模型(HA)、最小二乘法、ARIMA、逻辑回归、卡尔曼滤波模型、K 近邻模型等^[3-5]，这一类的研究方法主要是单输入单输出。近年来，国内外基于机器学习的组合预测方法在交通领域也得到广泛应用，Lin 等^[6]提出了将 ARIMA 模型和 GARCH 模型结合起来，实现交通流量的预测；Wang 等^[7]利用了混沌理论和 SVR 提出了一种基于多个交通流量基本参数的短期交通流量预测方法；Tang 等^[8]通过对含有噪声的原始交通流数据进行处理，结合去除噪声方案和支持向量机提出一种预测方法提高了预测精度。

随着深度学习在交通流预测方向的不断发展，组合预测的深度学习模型逐渐弥补了单个模型预测精度不高的缺陷，对于交通流的预测有较好的促进作用。殷礼胜等^[9]提出一种基于改进 VMD-GAT-GRU 模型的交通流预测模型，对高速公路交通流预测且精度较高；Zhang 等^[10]提出一种基于 GA-TCN 的短时交通流预测模型，利用遗传算法优化时间卷积神经网络的超参数，以确定最佳的适应度预测模型；Ashish 等^[11]提出注意力机制，其优点在于它能够全局捕捉联系，得到的模型质量较好；Ye 等^[12]为了更好关注影响网约车需求预测的关键特征，提出了一种基于注意力机制的 TCN 预测模型；吴莹莹等^[13]通过引入注意力机制并借助一维卷积单元提取数据的高维特征，利用 GRU 单元学习数据中的时序特征实现对超长序列的学习；目前对于交通流的预测大多基于对交通流的单一特征进行预测，较少考虑到其他特征，比如考虑到车辆速度对于交通流的影响、交通流的周期性以及车辆间距交通流大小对预测的影响考虑比较少。考虑到交通流量与各个特征(如：速度、不同间距之间的车流量、周期性、日期类型以及交叉口等)之间的相互影响，同时考虑到在预测中对历史数据的需求，将自注意力机制结合深度学习的组合预测模型对交通流预测是一个值得研究的话题。

在本文的交通流预测问题，着重分析关于深度学习与自注意力机制结果模式下交通流在不同特征影响下的交通

流预测，提出了基于自注意力机制 TCN-BiGRU 的交通流预测模型。本文的难点和创新点如下：首先考虑到交通流的随机性和不确定性，为了避免梯度消失和梯度爆炸，利用时间卷积网络(TCN)的卷积特性去跨时间步的提取交通流数据中的时间相关性；其次，将交通流数据重新拼接后经过双向门控循环单元(BiGRU)的更新门(update gate)和复位门(reset gate)后更全面的捕捉交通流的时间特性；考虑到双向门控循环单元在双向计算过程中存在有并行性较低和部分特征无法捕捉的情况，引入自注意力机制让模型能够注意到全局中不同输入之间的相关性，让模型能够不受序列长度限制的特征捕捉问题，最大限度的保留特征进而提高模型的鲁棒性。选取真实的交通数据进行预测对比测试，在单一路段将预测结果与基准模型和多路段的消融实验模型进行对比，最终得到预测值。

1 交通流问题描述与基础模型介绍

1.1 交通流问题描述

交通流预测的问题是指对交通流量、速度、交通状态以及交通拥堵情况等的相关信息集合进而实现对交通流的预测，交通流的是具有不确定性、非线性和复杂性的特点，因而对于交通流从预测不仅仅局限于交通的流量、速度、状态、天气、拥堵等情况。另外根据预测情况的步长不同，可分为单步预测和多步预测。单步预测是每次只预测未来一个值，单步预测准确率较高，因为它利用了最近的数据和已训练的长短期依赖性，但无法直接预测多步，否则可能导致时间悖论，而多步预测是每次预测未来多个值。短期预测使用单步预测，而长期预测可以使用多步预测或单步迭代训练预测。交通流预测根据预测时间的长短可分为短期(5~30 min)、中期(30~60 min)和长期(超过 1 h)预测。

1.2 长短期记忆网络模型结构

长短期记忆模型^[14-15](LSTM)是输入层、循环隐藏层和输出层组成。每个块由 3 个乘法门组成，即一个输入门、一个输出门和一个遗忘门，乘法门使神经单元能够在很长一段时间内保存和检索输入的交通流信息，从而解决并消除梯度消失问题。具体结构如图 1 所示。

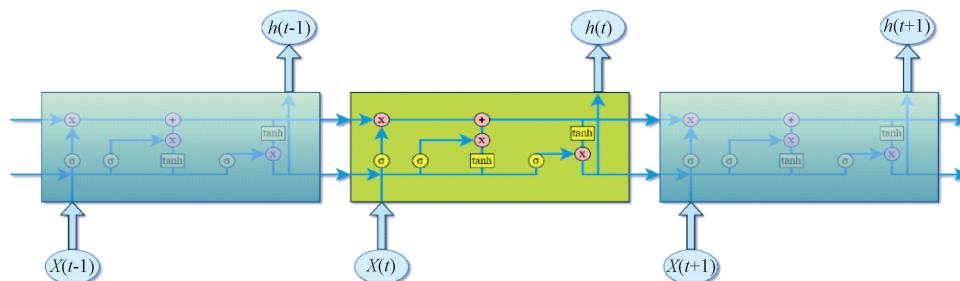


图 1 LSTM 的内部框架图

同时每个神经单元将单元状态和隐藏状态传输到紧邻的下一个单元，而输入门 i 控制存储单元是否更新，遗忘

门 f 通过告诉记忆单元 C_t 是否设置为 0，即 LSTM 网络将不再需要的交通流通过与遗忘门的过滤器相乘而被删除。

下列等式为 LSTM 结构。

$$\begin{cases} i = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{s}_{t-1} + \mathbf{V}_i \mathbf{C}_{t-1} + \mathbf{b}_i) \\ f = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{s}_{t-1} + \mathbf{V}_f \mathbf{C}_{t-1} + \mathbf{b}_f) \\ o = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{s}_{t-1} + \mathbf{V}_o \mathbf{C}_{t-1} + \mathbf{b}_o) \\ \mathbf{C}_t = f \cdot \mathbf{C}_{t-1} + i \cdot \mathbf{g}(\mathbf{W}_c \mathbf{x}_t + \mathbf{W}_s \mathbf{s}_{t-1} + \mathbf{b}_c) \\ \mathbf{s}_t = o \cdot h(\mathbf{C}_t) \end{cases} \quad (1)$$

其中, i, f, o 分别代表输入、遗忘和输出门的激活向量, \mathbf{C}_t 是细胞状态向量, $h(\mathbf{C}_t)$ 是输出向量。 $\mathbf{W}, \mathbf{U}, \mathbf{V}$ 是权重矩阵。

1.3 时间卷积网络模型结构

时间卷积网络^[16] (TCN)思想是一个因果卷积和膨胀卷积的过程。时间卷积网络(TCN)在本文中的主要作用是对交通流数据的时间序列特征提取,在训练的过程中对内存的要求不高,可以一维的滑动卷积核去接受交通流时间序列的任意长度,意味着时间卷积网络在接受有效的时间序列时可以实现滤波器的跨层共享,反向传播的多少取决于设置的网络深度。TCN 的结构框架如图 2 所示。

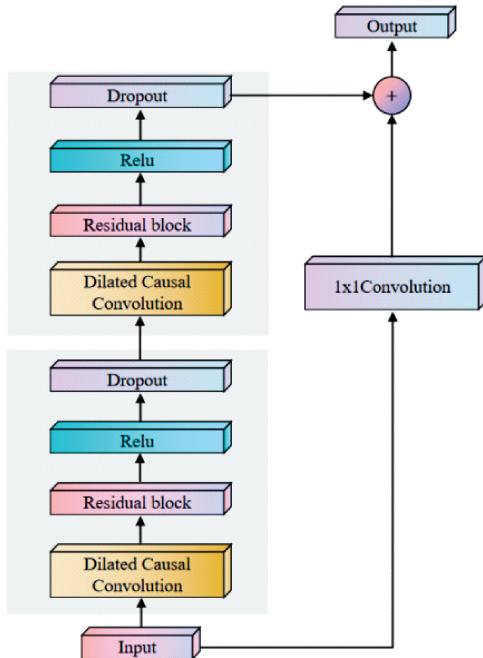


图 2 TCN 的结构框架图

1.4 双向门控循环单元模型结构

门控循环单元^[17] (GRU)与长短期记忆模型(LSTM)不同,GRU 不包含单独的内存块,这使得它在训练时更加高效和简单,GRU 将遗忘门和输入门合并为重置门。但是单项的门控循环单元只能处理交通流的正向时序特征,因为交通流之间是具有前后联系的,所以考虑到这一特性,引入双向门控循环单元(BiGRU),BiGRU 的前向隐含层神经元结构与后向隐含层神经元结构相同,如图 3 所示, BiGRU 可以把时间序列的正反两个方向的时序特征提取到,正向的 GRU 模型接受正向输入,反向的 GRU 学习反

向的输入。BiGRU 的公式表达如下:

$$\overrightarrow{\mathbf{h}}_t = GRU(x_t, \overrightarrow{\mathbf{h}}_{t-1}) \quad (2)$$

$$\overleftarrow{\mathbf{h}}_t = GRU(x_t, \overleftarrow{\mathbf{h}}_{t-1}) \quad (3)$$

$$y_t = \overrightarrow{\mathbf{h}}_t \oplus \overleftarrow{\mathbf{h}}_t \quad (4)$$

式中: $\overrightarrow{\mathbf{h}}_t$ 是 BiGRU 的前向输出, $\overleftarrow{\mathbf{h}}_t$ 是 BiGRU 的反向输出, y_t 为整个 BiGRU 的模型输出。

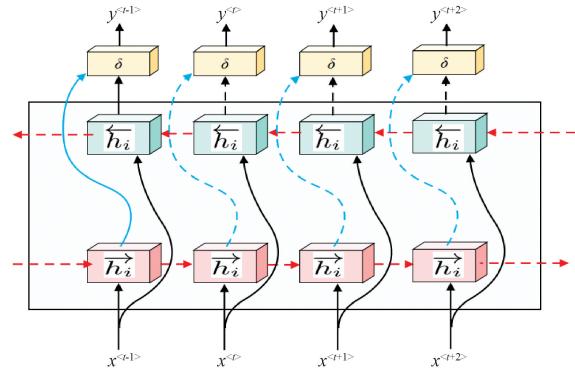


图 3 BiGRU 的网络结构

1.5 自注意力机制原理

由于路段与路段之间存在着空间相关的隐藏关系,且这种内在的隐藏关系没有被充分利用,可能会导致模型的训练效果较差,因此引入自注意力机制^[18] (self-attention),自注意力机制是一个映射函数通过对全局信息掌握的过程,进而提取交通流信息的各个特征之间的关键联系,注意力机制对模型进行权重的重新调整并对输入信息进行学习并训练之后更新参数^[19],使模型能注意到输入的交通流序列之间相关性,提高预测精度。其中注意力的公式如式(5)所示。

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (5)$$

其中, \mathbf{Q} 为输出单元的输出向量, $\sqrt{d_k}$ 是调节因子。 $softmax$ 值为 0 或者 1, 其意义是归一化。在本文中 \mathbf{Q} 指的是 BiGRU 输出单元的序列向量, K 为路段之间的空间特征关系。

2 基于自注意机制 TCN-BiGRU 的交通流预测模型

2.1 自注意机制 TCN-BiGRU 框架

为了提升交通流的预测效果,本文提出了基于自注意力机制 TCN-BiGRU 的交通流预测 (self-attention-TCN-BiGRU)。通过改进后的模型可以更完整的保留好交通流的信息,以便实现对交通流的更高精度的预测。基于自注意力机制 TCN-BiGRU 的交通流预测模型包括交通流时间序列的输入层(input)、时间卷积网络(TCN)层、双向门控循环单元(BiGRU)层、自注意机制(self-attention)层、输出层(output)。模型预测框架图如图 4 所示。

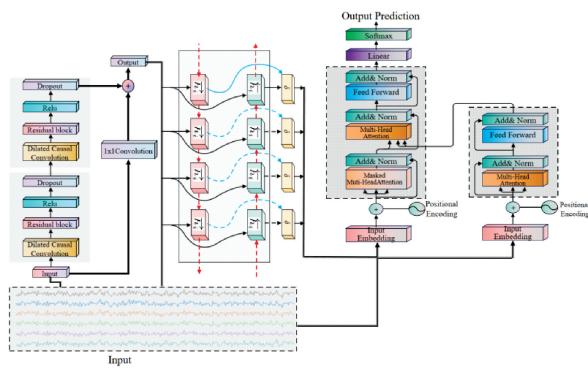


图 4 基于自注意力机制 TCN-BiGRU 的交通流预测模型

基于自注意力机制 TCN-BiGRU 的交通流预测模型的实现思路如下：

步骤 1) 将构建好的交通流数据输入进行读取和划分，将这些输入数据转换为适合模型处理的张量形式后通过读取，初步把数据集划分为训练集、验证集、测试集，然后再将数据标准化处理。

步骤 2) TemporalConvNet 将读取到的交通流数据在每个 TemporalBlock 包括一个一维卷积层和残差连接后，一维卷积层将窗口大小为 k 的滑动窗口应用于输入序列，通过卷积操作提取特征。然后，残差连接将卷积结果与输入序列相加，以便更好地捕捉序列中的重要特征。

步骤 3) 然后输入到双向门控循环单元(BiGRU)模型中，在 BiGRU 模型中，TCN 层特征处理后的交通流时间序列数据输入到 BiGRU 层，分别在 BiGRU 模型中的前向 \vec{h}_t 和反向 \overleftarrow{h}_t 计算。前向 \vec{h}_t 和反向 \overleftarrow{h}_t 计算之后输出 y_t 。BiGRU 通过在时间序列上前向和后向传播信息来捕捉序列中的上下文和长期依赖关系。

步骤 4) 特征再处理(增强)后，利用自注意力机制用于捕捉序列中不同空间位置的重要性的功能，将双向门控循环单元(BiGRU)模型输出的 y_t 输入到自注意力机制层(self-attention)，寻找到交通流序列与空间之间的关系之后，再做数据的标准化。综合特征表示的计算过程是将 BiGRU 的输出 h 和自注意力机制的输出 z 进行融合得到综合特征表示 f ，即 $f = [h; z]$ 。

步骤 5) 预测输出。自注意力机制 TCN-BiGRU 的交通流预测的预测结果输出，得到模型的最终预测值。

2.2 模型的参数设定以及优化器选择

模型的操作平台是在基于 pytorch 进行的，操作过程中的每个 epoch，都是先进行训练，每个 epoch 训练好后，再进行验证，这样操作的目的保证了训练过程中数据缺失或者损失的情况。同时为了保准训练的可靠性和过早的结束训练，本文设置 epoch 的迭代次数有为 1 000 次。

激活函数是网络模型中的一个重要组成部分，在本文模型中，选择的激活函数是修正线性单元(the rectified linear unit)，相较于 Sigmoid 和 tanh，ReLU 函数可以在本文模型中

优化梯度下降和收敛速度的问题。其表达式如下。

$$ReLU(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (6)$$

优化器选择的是自适应时刻估计方法 (adaptive moment estimation)，即 Adam，相较于梯度下降法，Adam 能把每个参数的自适应学习率计算出来，另外还具有环境感知的特点。Adam 的参数更新公式如下：

$$\begin{cases} \hat{\mathbf{M}}_{(t)} = \frac{\mathbf{M}_t}{1 - \beta_1^t} \\ \hat{\mathbf{V}}_t = \frac{\mathbf{V}_t}{1 - \beta_2^t} \\ \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \frac{r}{\sqrt{\hat{\mathbf{V}}_t} + \epsilon} \hat{\mathbf{M}}_{(t)} \end{cases} \quad (7)$$

其中， \mathbf{M}_t 为一阶动量项， \mathbf{V}_t 为二阶动量项， $\hat{\mathbf{M}}_{(t)}$ 为 \mathbf{M}_t 的修正值， $\hat{\mathbf{V}}_t$ 为 \mathbf{V}_t 的修正值。

3 数据的来源及处理和实验环境

3.1 实验和数据的选取

实验的框架基于 Anaconda3 和 pytorch 的深度学习框架，在 Pycharm 环境下导入深度学习包对交通流和速度的训练模型搭建和预测，通过配置 NVIDIA GPU 用 GeForce RTX2060 进行训练和预测。具体的实验环境如表 1 所示。

表 1 Self-attention-TCN-BiGRU 实验环境

实验环境	
CPU: 12th Gen Intel(R) Core(TM) i5-12400F	
硬件环境	GPU: NVIDIA GeForce RTX 2060
	内存: 32 G DDR4
	Windows 10 professional
软件环境	Python3. 10
	torch 1. 13
	cuda11. 6

考虑到交通流之间存在着相互关联性，比如在本文中所进行的实验数据是来自于 MIDAS 系统 2022 年 1~3 月的不同路段交通流数据，此次选取的交通流的数据包含车辆速度、不同间距之间的车流量、不同路段车流量、周期性、日期类型以及交叉口的交通流量，采样的特点是每 15 min 为一个采样间隔的交通流数据，这样做可以将交通流的特征信息更好的保留。在选取的道路中，该道路的分支与其他其他城市连接，有很强的交互作用，因此本文选择 MIDAS 系统内的 6 个路段的交通流数据，其中路段属性具有多样性，如环形交叉口，所选路段如图 5 所示。

3.2 交通流参数定义

对道路交通流数据的预测主要是对交通流的宏观参数预测。道路的宏观交通流数据参数包含：流量、密度、速

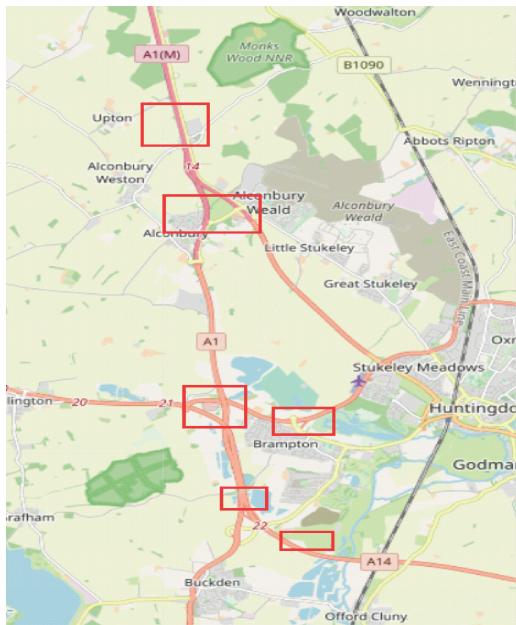


图 5 所选择的研究区域位置

度、流速、饱和度以及延误等。根据道路上的具体情况,本文选择了交通流量、车流量速度两个参数研究,同时考虑了交叉口交通流、节假日特征、不同观察间距交通流对其进行研究。参数的公式如下:

$$q = \frac{n}{s} \quad (8)$$

$$V = \frac{1}{n} \sum_{i=1}^n \bar{v}_i \quad (9)$$

式中: q 为在一段时间内通过选择路段的交通流平均数量; n 表示一段时间内的车辆数目; V 表示所选择路段的区间平均速度; \bar{v}_i 表示第 i 辆车的平均速度。

3.3 数据预处理

对于原始交通流数据其实存在着一些异常缺失的情况,因此对 MIDAS 系统的原始实验数据进行预处理,预处理的过程是首先对重复的数据进行删除,通过对交通流数据的可视化分析出异常值进而剔除异常值。此次的道路交通流数据的采样间隔是 15 min,选取 2022 年 1~3 月 6 条路段的交通流数据构建交通流数据集,该数据集包含不同间距段的交通流数据、总交通流数据和平均速度,部分交通流数据集及速度可视化如图 6 所示。

为了更好的验证本文的自注意力机制 TCN-BiGRU 交通流预测模型 (self-attention-TCN-BiGRU),将预处理后的原始数据划分训练集、验证集、测试集,三者的比例为 0.7:0.15:0.15,训练集用于对模型的训练,验证集用于对模型训练后的验证,测试集用于对模型预测结果前的测试。

4 实验结果与分析

4.1 评价标准

对于预测效果的分析与对比,所选取的预测评价指标

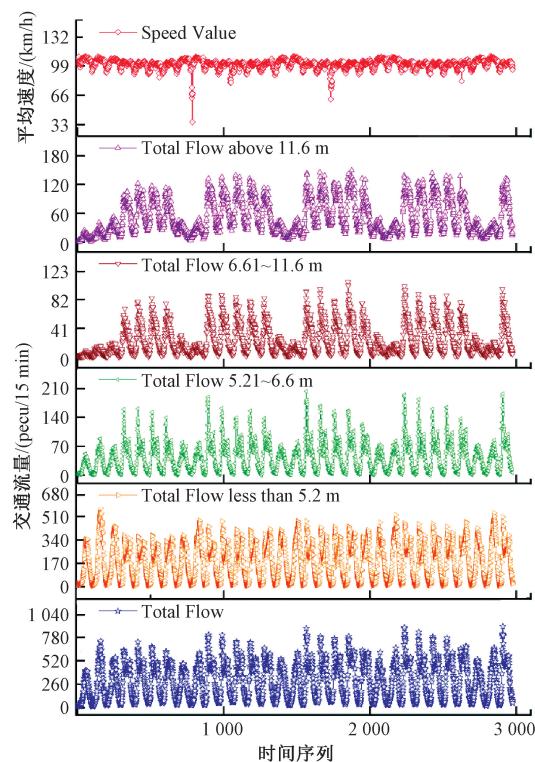


图 6 部分交通流量及速度

可以起到对模型的验证和对预测结果的有效性进行检验。因此本文选取的评价指标为以下 3 种:平均绝对误差 (MAE),平均绝对百分比误差 (MAPE) 以及皮尔逊相关系数 (R^2),这 3 种统计指标都能很好的分析出预测模型性能。平均绝对误差 (MAE) 是实际值与预测值之间误差的绝对平均值,MAE 的值越小表示选取的性能越好;平均绝对百分比误差 (MAPE) 是用来预测值与实际值之间的偏离程度,是为了更好的反映出预测的结果与实际值之间预测精度的比例平均值,当 MAPE 的值越小表示所选取的预测模型性能越好;皮尔逊相关系数 (R^2) 用来衡量预测值与实际值之间的线性强度的相关关系, R^2 的取值范围是在 0~1 之间,当 R^2 的值越接近 1,表示所选取的预测模型的预测能力越好。评价指标公式如下:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (10)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (11)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (12)$$

式中: y_i 表示 i 时刻真实的交通流量数据, \hat{y}_i 表示 i 时刻预测的交通流量数据,其中 N 表示用于预测的样本数量。

4.2 实验结果和分析

实验对所选取道路的 6 个路段交通流进行预测,其中每个路段有 6 个不同的输入特征,输入特征包含速度、4 个

不同间距的交通流、总交通流。对每个路段的 3 个月交通流进行预测,预测对实验模型的比对,在本文 Sa-TCN-BiGRU 模型中,所表现出来的预测效果最佳,具体的实验结果如表 2~7 所示。

表 2 路段 1 各预测模型实验结果

模型	MAE	MAPE/%	R ²
TCN	36.29	15.69	0.955
TCN-LSTM	30.08	13.24	0.967
TCN-GRU	28.19	9.88	0.970
TCN-BILSTM	27.18	10.59	0.971
TCN-BiGRU	27.91	9.61	0.971
Self-attention TCN-BiGRU	22.59	8.88	0.981

表 3 路段 2 各预测模型实验结果

模型	MAE	MAPE/%	R ²
TCN	7.54	28.66	0.825
TCN-LSTM	7.16	35.05	0.847
TCN-GRU	6.68	41.01	0.887
TCN-BILSTM	5.78	28.51	0.905
TCN-BiGRU	6.06	30.40	0.901
Self-attention TCN-BiGRU	4.05	20.02	0.956

表 4 路段 3 各预测模型实验结果

模型	MAE	MAPE/%	R ²
TCN	32.54	15.04	0.962
TCN-LSTM	31.12	12.88	0.963
TCN-GRU	31.56	12.51	0.962
TCN-BILSTM	28.96	10.01	0.966
TCN-BiGRU	29.41	10.32	0.967
Self-attention TCN-BiGRU	25.51	9.63	0.975

表 5 路段 4 各预测模型实验结果

模型	MAE	MAPE/%	R ²
TCN	26.15	16.22	0.928
TCN-LSTM	22.99	15.44	0.941
TCN-GRU	22.62	18.97	0.950
TCN-BILSTM	20.87	15.47	0.954
TCN-BiGRU	23.94	16.08	0.942
Self-attention TCN-BiGRU	19.62	13.53	0.960

表 6 路段 5 各预测模型实验结果

模型	MAE	MAPE/%	R ²
TCN	17.72	20.89	0.936
TCN-LSTM	13.83	12.62	0.955
TCN-GRU	13.44	11.19	0.952
TCN-BILSTM	14.37	14.44	0.953
TCN-BiGRU	13.03	11.67	0.963
Self-attention TCN-BiGRU	9.16	8.56	0.981

表 7 路段 6 各预测模型实验结果

模型	MAE	MAPE/%	R ²
TCN	34.96	14.29	0.944
TCN-LSTM	28.25	12.98	0.961
TCN-GRU	28.64	12.06	0.961
TCN-BILSTM	26.64	10.78	0.964
TCN-BiGRU	27.93	11.75	0.964
Self-attention TCN-BiGRU	20.50	9.08	0.979

在结合 6 个交通流特征对 6 个路段的交通流进行 6 个模型进行预测时,从表 2~7 可以看出,本文所提出的组合模型均表现最佳的预测效果,且较好的捕捉到了交通流的不同特征,因此本文提出的 Self-attention TCN-BiGRU 组合模型具有较好的预测性能。另外本文为了呈现出 6 种模型的可读性以及表现出模型的预测效果的有效性,将 TCN 模型,TCN-LSTM 模型,TCN-GRU 模型,TCN-BiLSTM 模型,TCN-BiGRU 模型,Self-attention TCN-BiGRU 模型这 6 种模型进行预测对比,特此选取路段 2 作 6 个模型的可视化分析,同时对本文所提出的模型作局部特征做了放大处理,如图 7(a)~(f)所示。

在图 7(a)~(f)所展示路段 2 的 6 个模型真实值与预测值可视化视图中,通过输入值和预测值的对比,虽然说图 7(a)~(e)这 5 个模型的总体预测趋势都能捕捉到交通流的一些特征,但是这 5 个模型对较多的突出局部特征捕捉的效果表现其实不佳。如图 7(f)所示,本文所提出的基于自注意力机制 TCN-BiGRU 组合模型预测交通流,对局部特征放大之后发现所提出的模型能较好的捕捉全局特征和局部特征,且能预测反映出较为真实的交通流变化趋势。同时从表 2~7 的数据结果证明本文所选取的 Self-attention TCN-BiGRU 预测模型对交通流预测具有较好的预测效果,提升了预测精度,因此本文所选取的模型预测更符合真实值。

为进一步的验证本文所选取模型的预测可靠性,前面的工作是对单一路段的预测,即多特征单一路段,考虑到交通流具有传递性,且路段之间是具有网络结构,因此将多路段多特征的交通流数据输入到本文选取的模型中,同时也通过对比该领域中比较经典的一些模型进行性能比较,各实验结果如表 8 所示。

相比较于单一的经典模型和其他的组合模型,本文提出的基于 Self-attention TCN-BiGRU 的交通流组合预测模型如图 8 的可视化,通过对局部的放大发现,本文选取的模型预测效果最佳,预测结果能较好的拟合真实值的变化趋势。从表 8 中的预测结果评标指标可以分析出,由于短时交通流的非平稳性特点,因此采用如 TCN、LSTM、GRU、BiGRU 等单一较为经典模型的指标评价都不是很好。与单一的模型相比,采用组合模型的预测方法,模型的预测精度如表 8 所示得到了一定的提升,但预测结果与实际结果的拟合偏离程度如图 8 所示表现得略高。本文

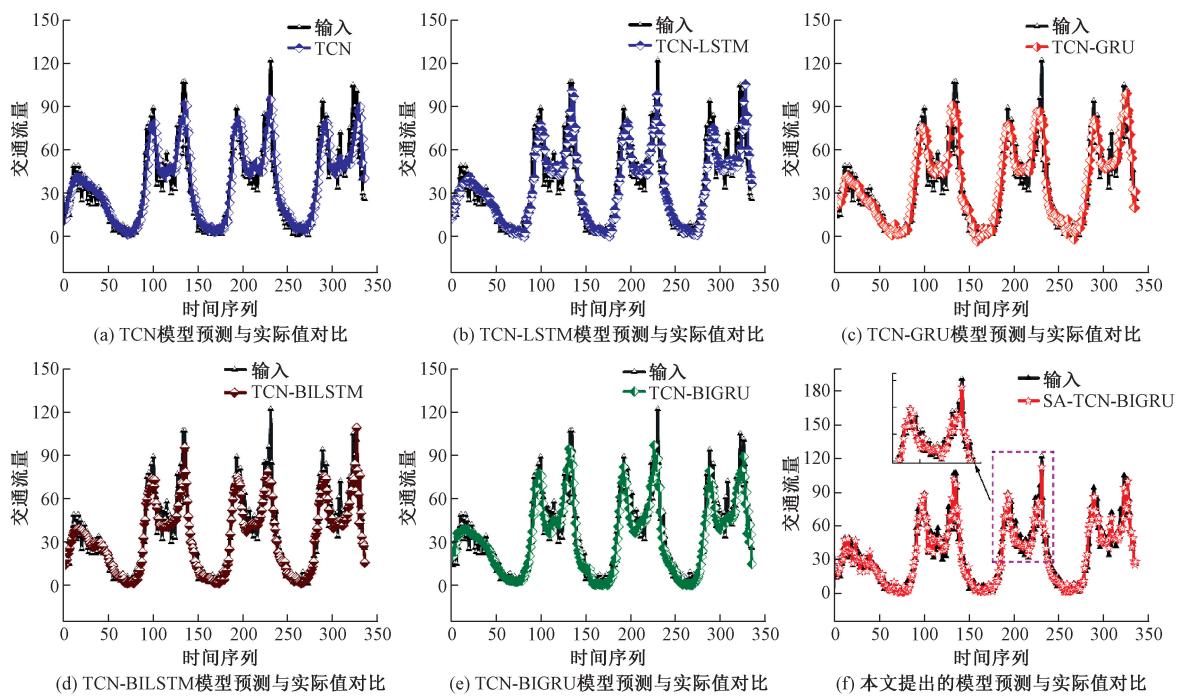


图 7 路段 2 各预测模型预测结果与实际值比较

表 8 多路段多特征各预测模型实验结果

模型	MAE	MAPE/%	R^2
TCN	26.84	20.74	0.931
LSTM	27.81	17.1	0.957
GRU	27.21	16.9	0.944
BiGRU	20.42	15.8	0.961
TCN-LSTM	21.80	17.31	0.955
TCN-GRU	23.26	15.29	0.948
TCN-BiLSTM	18.11	12.23	0.968
TCN-BiGRU	18.58	12.82	0.967
Self-attention TCN-BiGRU	13.63	9.21	0.982

提出的基于自注意力机制 TCN-BiGRU 的交通流预测不管是对于多特征的单一路段还是多路段都有效的提高预测精度,其预测结果和拟合效果都明显优于其他的单一和组合模型,预测的误差也是最小的,MAE 为 13.63,MAPE/% 为 9.21, R^2 为 0.982,证明了组合预测的优点。综上所述,本文提出的基于 Self-attention TCN-BiGRU 的组合预测模型能够提高预测精度,具有较好的预测效果。

5 结 论

考虑到交通流的随机特性,本文提出了基于自注意力机制 TCN-BiGRU 的交通流预测模型。首先利用时间卷积网络(TCN)的卷积特性去跨时间步的提取收集交通流特征,避免梯度消失和梯度爆炸;其次,将交通流数据重新拼接后经过双向门控循环单元(BiGRU)的更新门(update gate)和复位门(reset gate)进行正向和反向更好的捕捉交通流的时间相关特性,然后将双向的 GRU 的输出拼接起来;最后考虑到双向门控循环单元的输入有很多大小不一样的向量,且考虑到其中还有一定的联系,且空间特征无法捕捉的情况,因此引入自注意力机制让模型能够注意到时间和空间位置不同输入之间的相关性,得到交通流的最后预测值。通过仿真结果表明,与其他 5 种预测模型相比,Self-attention TCN-BiGRU 预测模型对于多特征的单一路段还是多路段都可以有效的提高预测精度,具有较好的预测效果,并且在多路段的 MAE,MAPE/%, R^2 分别为 19.62,13.53,0.982 取得了不错结果。本文所提出的模型对于不同特征不同路段的交通流预测均能表现出较好的预测精度,可以作为道路交通管理的参考。

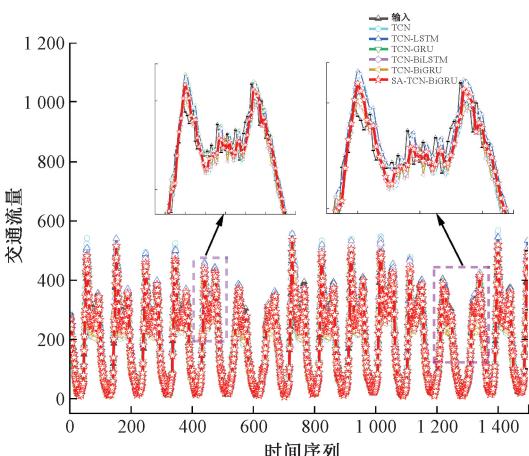


图 8 多路段多特征各预测模型可视化

参考文献

- [1] BOUKERCHE A, WANG J H. Machine learning-based traffic prediction models for intelligent transportation systems [J]. Computer Networks, 2020, 181:107530.
- [2] KAFFASH S, NGUYEN A T, ZHU J. Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis [J]. International Journal of Production Economics, 2021, 231:107868.
- [3] BRIAN L, SMITH, MICHAEL J, et al. Traffic flow forecasting: Comparison of modeling approaches [J]. Journal of Transportation Engineering, 1997, 123: 261-266.
- [4] MOHAMMAD M, HAMED, HASHEM R, et al. Short-term prediction of traffic volume in urban arterials [J]. Journal of Transportation Engineering, 1995, 121:3(249).
- [5] VOORT D V M, DOUGHERTY M, WATSON S. Combining kohonen maps with arima time series models to forecast traffic flow [J]. Transportation Research Part C-Emerging Technologies, 1996, 4(5): 307-318.
- [6] LIN X F, HUANG Y Z. Short-term high-speed traffic flow prediction based on ARIMA-GARCH-M Model [J]. Wireless Personal Communications, 2021, 117 (4): 3421-3430.
- [7] WANG X Q, LI W J, YIN C Y, et al. A multiple-parameter approach for short-term traffic flow prediction [J]. Modern Physics Letters B, 2021, 35(14):2150245.
- [8] TANG J, CHEN X, HU Z, et al. Traffic flow prediction based on combination of support vector machine and data denoising schemes [J]. Physica A: Statistical Mechanics and its Applications, 2019, 534: 120642.
- [9] 殷礼胜,吴洋洋. 基于改进 VMD-GAT-GRU 的交通流量组合预测模型 [J]. 电子测量与仪器学报, 2022, 36(7):62-72.
- [10] ZHANG R J, SUN F, SONG Z W, et al. Short-term traffic flow forecasting model based on GA-TCN [J]. Journal of Advanced Transportation, 2021, 1338607:13.
- [11] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30:5998-6008.
- [12] YE X F, HAO Y, YE Q M, et al. Demand forecasting of online car-hailing by exhaustively capturing the temporal dependency with TCN and Attention approaches [J]. Iet Intelligent Transport Systems, 2023, DOI: 10.1049/itr2.12387.
- [13] 吴莹莹,赵丽宁,袁志鑫,等. 基于注意力机制的 CNN-GRU 船舶交通流预测模型 [J]. 大连海事大学学报, 2023, 49(1):75-84.
- [14] SEPP H, JURGEN S. Long short-term memory [J]. Neural Computation, 1997, 9:1735-1780.
- [15] MA X L, TAO Z M, WANG Y H, 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.
- [16] BAI S J, KOLTER J Z, VLADLEN K, et al. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [J]. Computing Research Repository, 2018, DOI: 10.48550/arXiv.1803.01271.
- [17] ZHANG D, KABUKA M R. Combining weather condition data to predict traffic flow: a GRU-based deep learning approach [J]. IET Intelligent Transport Systems, 2018, 12(7): 578-585.
- [18] WANG S, LIB, KHABSA M, et al. Linformer: self-attention with linear complexity [J]. ArXiv Preprint, 2020, ArXiv:2006:04768.
- [19] ABNAR S, ZUIDEMA W. Quantifying attention flow in Transformers [J]. 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), 2020. <https://doi.org/10.48550/arXiv.2005.00928>.

作者简介

郝椿淋,硕士研究生,主要研究方向为智能交通,时空大数据。

E-mail: haodehen0510@163.com

张剑(通信作者),博士,主要研究方向为智能交通,信息融合与模式识别。

E-mail: 08120003@sues.edu.cn