

基于注意力机制的CNN-BiGRU的短时交通流预测

杨 瑞, 冯慧芳

西北师范大学, 数学与统计学院, 甘肃 兰州

收稿日期: 2021年11月24日; 录用日期: 2021年12月21日; 发布日期: 2021年12月29日

摘 要

短时交通流预测是实现道路交通管理和指导的基础,对道路通行能力、交通安全等具有重要的实际意义。提出一种基于注意力机制的卷积神经网络和双向门控循环单元神经网络相结合的深度学习交通流预测模型。采用卷积神经网络(CNN)提取交通流的空间特征。利用双向门控循环单元(BiGRU)从前向和后向传播中都获取信息,充分提取交通流的时间相关特征。引入注意力机制,将交通流最重要的特征得到了最大限度的保留,以此来提升CNN和BiGRU网络的特征提取效果,达到提升模型的预测精度的目的。在真实数据集上的对比实验和消融实验结果都表明:基于注意力机制的CNN-BiGRU的短时交通流预测模型能够有效捕获交通流的动态时空特征,具有良好的预测性能。

关键词

智能交通, 交通流预测, 卷积神经网络, 双向门控循环单元, 注意力机制

Short-Term Traffic Flow Prediction of CNN-BiGRU Based on the Attention Mechanism

Rui Yang, Huifang Feng

College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

Received: Nov. 24th, 2021; accepted: Dec. 21st, 2021; published: Dec. 29th, 2021

Abstract

Short-term traffic flow prediction is the basis of road traffic management and guidance, and has

important practical significance for road capacity and traffic safety, and so on. A short-term traffic flow prediction based on CNN-BiGRU-Attention (CNN and BiGRU network based on attention mechanism) model is proposed. The Convolutional Neural Networks (CNN) have been employed to extract the spatial features of the traffic flow. The Bidirectional Gated Recurrent Unit (BiGRU) is exploited to obtain information from both forward and backward propagation, and then fully capture temporal dependencies of the traffic flow. The attention mechanism is employed to maximize the retention of the most important features of the traffic flow, so as to improve the feature extraction effect of CNN and BiGRU networks, and achieve the purpose of improving the prediction accuracy of the model. The results of comparison experiments and ablation experiments on the real data set show that the short-term traffic flow prediction model of CNN-BiGRU based on the attention mechanism can effectively capture the dynamic spatio-temporal characteristics of traffic flow, and has good prediction performance.

Keywords

ITS, Traffic Flow Prediction, CNN, BiGRU, Attention Mechanism

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

计算机技术、通信技术、传感技术等现代信息技术为智能交通系统(Intelligent System, ITS)的蓬勃发展提供了重要的技术支撑。智能交通系统在交通领域中的广泛应用,推动了道路交通管理技术的革新,不仅提高了交通效率,而且也使得行人的出行质量和车辆安全都得到了很大的改善与保障。实时、准确地识别与评估城市交通状态,不仅能够为交通拥堵控制决策提供有效的支持,也为居民出行提供指导[1][2]。

短时交通流预测是 ITS 系统的核心模块之一,相关研究及应用受到学术界和业界的广泛关注[3][4][5]。但由于交通流具有非线性、时间与空间的依赖性及随机性等特征,使得短时交通流预测的研究方面仍然还未能达到让人满意的成果,在预测精度和稳定性上还有待突破。根据近年来的研究成果,可将短时交通流预测方法大致归纳为四类:

1) 基于统计的预测方法: Zhang 等[6]提出了差分整合移动平均自回归模型和自回归条件异方差模型的短时交通流量预测模型。李文勇等[7]基于 Box-Cox 指数变换方法,研究了交通流量均值与方差间的关系,充分考虑交通流的长相关特征,建立了改进的差分整合移动平均自回归模型的交通流预测模型。周晓等[8]以美国明尼苏达州双子城快速路的实际数据为样本,建立了基于卡尔曼滤波的道路平均速度预测模型,仿真实验验证了该模型具有良好的预测性能。统计分析法虽然算法和模型简易,实现方便,但其主要建立在时间序列数据平稳的假设前提下。基于统计的预测模型复杂度低,能够刻画交通流的非线性特征,但该类模型是建立在时间序列数据平稳的假设前提下,而且模型仅能刻画交通流时间性关系,忽略道路之间的空间相关性,没有考虑其他外界因素,当交通流状态不确定性较强时,其预测结果可能会与实际状况严重不符。

2) 传统的机器学习预测方法: Zhu [9]和 Ryu [10]等考虑了交通流的时空特征,分别应用贝叶斯网络模型和 K-近邻模型进行短期预测。Tang 等[11]等将基于支持向量机(Support Vector Machine, SVM)和去噪算法相结合,建立了交通流多步预测模型,并用明尼苏达州的三个检测器收集的流量数据验

证了模型有效性。Xie 等[12]分析了影响交通的主要因素,对基于传统机器学习的交通流预测模型进行综述性分析,并基于多源交通数据进行了仿真实验。传统的机器学习方法侧重于短期交通流预测,能够达到较好的预测精度。然而,传统的机器学习模型架构简单、参数有限,适合比较简单的交通环境的流量预测,不能挖掘大数据中深层的、隐含的时空相关性,故对复杂交通环境中的交通流预测能力有限。

3) 基于深度学习的预测方法:深度学习是一种处理大数据非常有效的方法,通过监督学习或者无监督学习,自动学习复杂数据中的模式并提取数据特征。深度学习模型以其强大的分层特征学习能力在各种机器学习任务中取得了巨大的成功,也被广泛应用于交通流预测中。杜圣东[13]等采用序列到序列时空注意力深度学习模型对交通流进行预测。卢生巧[14]等用卷积神经网络和门控循环单元(GRU)神经网络做短时交通流预测,但是忽略了后续序列信息的影响,当时间序列变长时,存在重要信息丢失现象。陆文琦等[15]建立了基于双向长短期记忆神经网络 BiLSTM (Bi-directional Long Short-Term Memory)和注意力机制混合深度学习预测模型,并以北京市真实的交通流数据进行算法性能验证,实验结果表明提出的模型具有较好的单步与多步预测性能。王博文[16]等提出一种基于编码器-解码器(ED)框架的长短期记忆网络(LSTM, Long Short-Term Memory)模型,在 PEMS 数据集上做多步预测,采用 Encoder-Decoder 学习框架,解决了因时间序列较长及多步预测所引起的模型预测误差迅速累积的问题。采用 LSTM 作为编码器对多因素交通流序列的非线性特征进行拟合,充分挖掘了交通流序列中多个变量之间的非线性关系。潘卫鹏[17]等提出一种基于图注意力卷积网络(GACN)的短时交通流预测方法,在数据集 PEMS 上对该模型效果进行验证,结果表明:可以有效地捕获交通数据的动态时空依赖性。Ye [18]等分别对基于深度学习和基于图卷积神经网络的预测模型进行了文献综述,给出了一般预测框架,讨论了深度学习预测模型要解决的核心问题及存在的挑战。Seng 等[19]提出一种基于多图卷积神经网络和门控循环单元的不规则区域交通流量预测模型。并利用公共自行车系统数据集、出租车数据集和无桩共享自行车数据集等三个真实数据集验证了所提模型预测性能。深度学习模型虽然在预测性能上都取得了一定的提升,但是其超参数调整也比较困难,如何对隐藏层神经元数和网络层数进行选择,是值得进一步研究的方向。

4) 基于强化学习的预测方法: Peng 等[20]应用图卷积神经网络学习交通网络的空间特征,结合长短期记忆神经网络学习交通流的时间特征,提出了基于强化学习的长期交通流预测模型,并以纽约自行车数据集验证了模型有效性。目前,强化学习方法在城市交通中的应用更侧重于应用深度强化学习解决交通控制问题,研究城市交通系统中客流量控制策略[21]。

交通网络是一个复杂动态系统,一条道路某时刻的交通状态不仅和该条道路前多个时刻的交通状态有关,而且与其它多条道路的交通状态有关。因此,交通流量预测模型既反映了各相关道路交通流之间的关联关系,又考虑一条道路在不同时间点的交通流之间的相互影响。在短时交通流的预测中,历史数据中隐藏了交通流的一些潜在规律。传统机器学习方法不能深入挖掘历史数据的时序性特点和数据之间所隐藏的潜在规律。深度学习网络能充分挖掘数据之间的关系,获取历史流量数据的周期性规律,提取出较为重要的信息,提高预测精度。针对以上所存在的挑战,本文作如下改进:1) 在流量数据不确定性的情况下,双向门控循环单元神经网络可以有效的对动态时间序列数据进行建模;2) 由于当输入时间序列过长时,双向门控循环单元神经网络会出现信息丢失的现象、从而很难挖掘数据结构之间的信息。为此,通过引入 Attention 机制加强了关键信息的影响,避免了时间序列过长而出现的丢失信息的问题。因此,针对交通流的非线性、时间动态性、空间相关性等特征,利用深度学习模型的优势,本文提出一种基于注意力机制的卷积神经网络和双向门控循环单元神经网络相结合的集成交通流量预测模型。

2. 基于注意力机制的 CNN-BiGRU 的短时交通流预测模型的构建

2.1. 卷积神经网络模型 CNN

卷积神经网络(Convolutional Neural Networks, CNN)是一种用于深度学习的网络架构, 它直接从数据中学习, 无需手动提取特征, 其结构如图 1 所示[22]。CNN 结构包括数据输入层、输出层和多个隐藏层, 隐藏层由卷积层, 池化层及全连接层构成。

数据输入层: 数据输入层是整个网络的输入, 一般情况需要对原始数据进行预处理。

卷积层: 卷积层包括了卷积计算层和激活层, 卷积计算层是将卷积核矩阵与对应位置的数据矩阵进行矩阵运算, 设输入为 $x \in R^{M \times N}$, 卷积核为 $W \in R^{U \times V}$, 卷积定义为 $Z = W * X$, 其中 $*$ 为卷积运算符, 其运算公式为:

$$z_{ij} = \sum_{u=1}^U \sum_{v=1}^V W_{uv} X_{i-u+1, j-v+1} \quad (1)$$

激活层的功能是通过激活函数实现, 该层是应用激活函数将卷积计算层的输出结果进行非线性变换: $Y = f(z)$, 其中 $f(\cdot)$ 为非线性激活函数, 常用的激活函数有 Sigmoid、tanh、ReLU、ReLU 变形等。

池化层: 池化层的主要功能是进行压缩数据, 减少神经网络的参数和计算量, 同时防止过拟合。常用的池化方式为最大池化(Max Pooling), 该方法是将池化区域内的最大值作为该区域的值, 突出数据最重要的特征。其他池化方法还包括平均池化、重叠池化和空间金字塔池化、随机池化等。

全连接层: 全连接层一般是卷积神经网络的最后一层, 该层的神经元与上层的所有神经元全连接。全连接层的作用是整合网络学习到的分布式特征表示, 并将它们映射到样本标签空间。

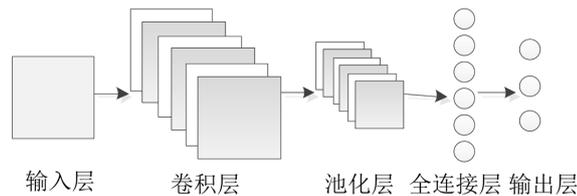


Figure 1. Structure of CNN
图 1. CNN 结构

2.2. 双向 GRU 模型结构

门控循环单元 GRU (Gated Recurrent Unit, GRU)是一类基于 LSTM 的基础上对结构进行了优化改进的循环神经网络[23]。改进的模型预测精度有保证, 训练参数减少, 时间复杂度较低, 能够解决训练中梯度消失与梯度爆炸问题。GRU 结构如图 2 所示, 图中 x_t 为输入, 表示更新门, r_t 表示重置门, h_t 为隐藏层的输出。GRU 门控更新公式为:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (2)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (3)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (4)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (5)$$

其中 W_z 、 W_r 、 W 为可训练的参数矩阵, b_z 、 b_r 、 b_h 为偏置。

单向 GRU 状态总是从前往后输出单向传播, 只能提取当前状态前向信息, 不能利用后续的序列信息。双向的 GRU (Bi-direction GRU, BiGRU) 让网络从前向和后向传播中都获取信息, 充分提取数据相关特征。BiGRU 结构如图 3 所示, 其输出同时由这两个 GRU 的状态共同决定:

$$h_i = \bar{w} \bar{h}_i + \tilde{w} \tilde{h}_i \tag{6}$$

其中 \bar{h}_i 、 \tilde{h}_i 分别为正向传播和方向传播的输出, \bar{w} 和 \tilde{w} 分别为正向和方向传播的权重。这样, 通过前后传播的 BiGRU 对数据特征信息进行叠加, 使得最重要的特征得到了最大限度的保留。

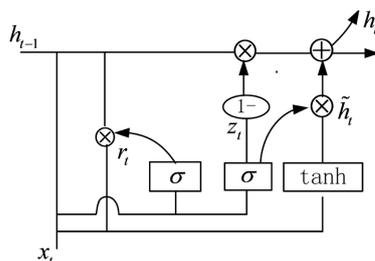


Figure 2. Structure of GRU

图 2. GRU 结构

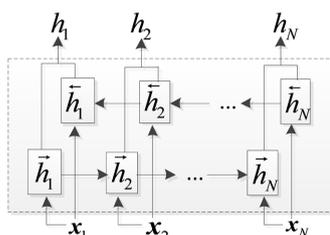


Figure 3. Structure of BiGRU

图 3. BiGRU 结构

2.3. 注意力机制

Attention 机制是一种模拟了人脑注意力的资源分配机制, 在特定时刻, 人脑会将注意力集中在需要重点关注的区域, 减少甚至忽略对其他区域的关注, 以得到更多需要关注的信息, 屏蔽其他无用信息的干扰。其核心思想是改变对信息的注意力, 忽略无关信息并放大所需信息。注意力机制是模拟人类的注意力机制来提高深度学习的性能。Attention 结构如图 4 所示[22]。

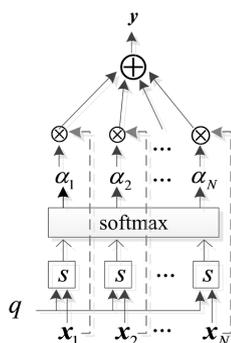


Figure 4. Structure of the attention mechanism

图 4. 注意力机制结构

假设 Attention 层的输入向量为 $X = [x_1, x_2, \dots, x_N]$, 给定一个与任务有关的查询向量 q , 计算输入信息的注意力分布 α_n :

$$\alpha_n = \text{softmax}(s(x_n, q)) \quad (7)$$

其中 $s(x_n, q)$ 为注意力打分函数, 注意力打分函数可采用加法模型、点积模型、双线性模型等。Attention 层输出 y 是注意力分布(即权重系数)对输入信息进行加权求和:

$$y = \sum_{n=1}^N \alpha_n x_n \quad (8)$$

2.4. 基于注意力机制的 CNN-BiGRU 的短时交通流预测模型

本文利用卷积神经网络、双向的 GRU 和注意力机制的优势, 融合三种模型, 提出一种基于注意力机制的 CNN-BiGRU 的短时交通流预测模型。模型框架如图 5 所示[21]。由多层 CNN、BiGRU 和注意力机制组成。具体预测步骤为:

- 1) **输入层**: 将原始历史数据进行预处理后输入到 CNN 层;
- 2) **CNN 层**: CNN 层由两个 CNN 模块和一个全连接层组成, 每个 CNN 模块包括 1 个卷积层和 1 个池化层。根据经验和交通流量实验数据的结构, 将卷积层设计为一维卷积, 并选 ReLU 函数作为卷积层的激活函数, 池化层的池化方法选取最大池化。经过两层卷积层和池化层的处理, 得到特征向量经过全连接层结构并由 Sigmoid 函数激活得到 CNN 层最终的输出向量;
- 3) **BiGRU 层**: CNN 层的输出特征向量输入到 BiGRU 网络中, 通过前后传递的 BiGRU 网络对特征信息进行叠加, 充分挖掘交通流数据的时间相关性;
- 4) **Attention 层**: 将经过 BiGRU 网络训练获得的特征向量作为 Attention 层的输入, 计算注意力系数, 并利用注意力系数对特征向量进行有区别的信息聚合, 将最重要的特征得到了最大程度的保留;
- 5) **输出层**: 将从 Attention 层获取的特征向量输入到全连接层, 通过 Sigmoid 激活函数得到交通流预测值。



Figure 5. The framework of prediction model
图 5. 预测模型框架

3. 实验结果与分析

3.1. 模型性能指标和基线模型

1) 评价指标

本文选取平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)、平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)和拟合优度 R^2 为评价标准, 各指标的定义分别为:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \times 100\% \quad (9)$$

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

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2} \quad (11)$$

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

其中 y_i 为实际交通流, \hat{y}_i 为预测交通流量, \bar{y} 为实际交通流的均值, n 为样本数量。预测模型的 MAPE、MAE、RMSE 等参数的值越小, 说明模型预测性能越好。拟合优度统计量 R^2 越接近 1 表示模型对样本数据的拟合程度越好, 预测性能也越好; 反之, R^2 越小, 模型的预测性能越差。

2) 基线模型设置

本文采用历史平均模型(History average model, HA)、BP 神经网络(back propagation, BP)、混合神经网络 CNN-GRU (Convolutional Neural Networks-Gated Recurrent Unit, CNN-GRU)和 CNN-GRU-Attention 等基线模型进行性能比较。

a) HA 模型: HA 模型是一种经典的时间序列预测方法, 是将一个时段内的历史交通流量的平均值作为未来的预测值。

b) BP 模型: BP 神经网络是一种应用广泛的基于反向传播算法的前馈神经网络。实验中隐藏层神经元为 100。

c) CNN-GRU 模型: 该模型是基于 CNN 和 GRU 混合的神经网络预测模型。模型参数设置 num_epochs 为 50、batch_size 为 256、学习率为 0.005, 隐含层神经元为 10。

d) CNN-GRU-Attention 模型: 是增加了注意力机制的 CNN-GRU 模型。模型参数设置 num_epochs 为 50、batch_size 为 128、学习率为 0.005, 隐含层神经元为 10。

3.2. 数据集和实验参数设置

本研究采用美国加利福尼亚州 PEMS 公开数据集, 被很多学者用于研究, 如文献[16] [17], 本文选自 2017 年 1 月 1 日~7 月 15 日, 33 条高速公路的流量, 数据采用间隔为 5 Minutes, 故每条道路每天有 288 个数据点, 个别缺失数据使用线性插值法进行填充。原始数据需要进行归一化处理:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

其中 x 为原始流量数据, x' 是经过归一化后的数据, x_{\max} 和 x_{\min} 分别是样本数据的最大值和最小值。

选取数据集 70% 的数据作为训练集, 30% 的数据作为测试集。两个 CNN 的卷积核大小均为 3×3 , 卷积核数量分别为 6 和 16, 步长为 1, padding 设置为 SAME, 激活层函数为 ReLU, 池化层核均为 3×3 , 步长为 2。前向 BiGRU 和后向 BiGRU 模型的神经元均为 120, num_epochs 为 100、batch_size 为 256、学习率为 0.003。CNN 训练参数设置: num_epochs 为 100、batch_size 为 32、学习率为 0.003 本文实验环境采用 11th Gen Intel i5-1135G7 处理器, NVIDIA-GeForce MX450 显卡。算法模型采用 Python3.6 作为编程语言, 软件架构基于 Pytorch 框架, 绘图工具采用 matplotlib 绘图模块。设置均方误差(MSE)作为损失函数, 选择 Adam 算法作为参数优化器。

3.3. 实验结果分析

1) 对比实验

图 6 为编号为 3 的监测点在 2017 年 7 月 15 日 00:00 到 23:55 的交通流量真实值和预测值。由图中可知, 除了 HA 模型, 其他神经网络模型的预测与真实值都比较接近, 但本文提出的基于注意力机制的

CNN-BiGRU 的短时交通流预测能够较好地捕获交通流量的整体趋势, 预测值与真实的交通流量值非常接近。夜间时段的交通流明显低于白天时段的交通流。

表 1 为本文所提模型与 HA、BP、CNN-GRU、CNN-GRU-Attention 等模型的预测性能比较。由表 1 可知本文提出的基于注意力机制的 CNN-BiGRU 模型的预测在三个评价指标中均具有最好性能。HA 模型实现简单, 但缺乏对时空时间序列空间特征挖掘, 预测精度不高。BP 和 CNN-GRU 模型对于短期预测表现良好, 但缺乏对交通流复杂特征的深层次挖掘, 相比之下, 本文建立的深度学习预测模型, 在 CNN-BiGRU 混合模型基础上引入 Attention 机制, 模型不仅能捕获交通流量的时空特征, 而且加强了关键信息的提取, 进一步提升了模型的预测性能。

Table 1. The results of comparative experiment

表 1. 对比实验结果

模型	MAPE	MAE	RMSE	R^2
HA	28.23%	57.81	81.12	0.71
BP	10.78%	24.95	32.90	0.95
CNN-GRU	10.78%	26.54	34.93	0.95
CNN-GRU-Attention	10.21%	25.28	34.48	0.95
CNN-BiGRU-Attention	7.71%	19.55	26.66	0.97

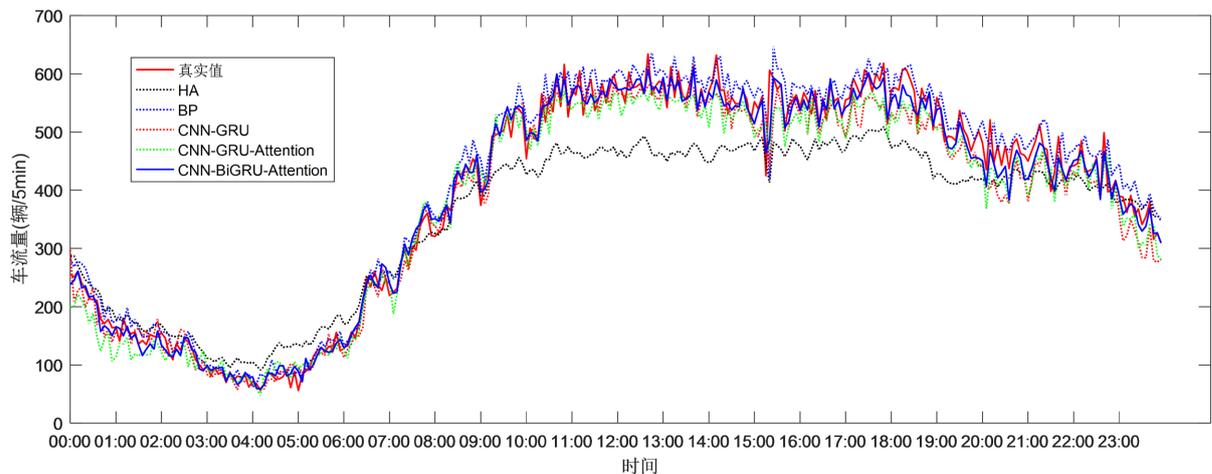


Figure 6. Comparison results of predicted value and actual value

图 6. 真实值与预测值对比

表 1 为本文所提模型与 HA、BP、CNN-GRU、CNN-GRU-Attention 等模型的预测性能比较。由表 1 可知本文提出的基于注意力机制的 CNN-BiGRU 模型的预测在三个评价指标中均具有最好性能。HA 模型实现简单, 但缺乏对时空时间序列空间特征挖掘, 预测精度不高。BP 和 CNN-GRU 模型对于短期预测表现良好, 但缺乏对交通流复杂特征的深层次挖掘, 相比之下, 本文建立的深度学习预测模型, 在 CNN-BiGRU 混合模型基础上引入 Attention 机制, 模型不仅能捕获交通流量的时空特征, 而且加强了关键信息的提取, 进一步提升了模型的预测性能。

2) 消融实验

通过消融实验检验基于注意力机制的 CNN-BiGRU 模型中的每个组件对预测效果的影响, 实验结果如表 2 所示。

Table 2. The results of ablation experiment

表 2. 消融实验结果

模型	MAPE	MAE	RMSE	R^2
CNN	15.37%	31.72	40.78	0.93
BiGRU	11.61%	31.08	37.78	0.94
BiGRU-Attention	10.41%	22.11	29.67	0.96
CNN-BiGRU	10.81%	27.23	35.72	0.94
CNN-BiGRU-Attention	7.71%	19.55	26.66	0.97

由表 2 可知, 单纯 CNN 预测模型性能最差, 是因为该模型没有考虑交通流的时间相关性, 相比之下, 能够捕获时间相关性的 BiGRU 模型预测效果较好。BiGRU 和 CNN-BiGRU 模型的预测性能差别不大, 增加了注意力机制的 BiGRU 模型预测性能明显提高, 这说明注意力机制能够增加深度学习模型的学习能力, 有助于提升模型的预测性能。对比实验和消融实验的结果都表明本文建立的增加了注意力机制的 CNN-BiGRU 模型预测性能明显优于其他模型。

4. 结论

本文利用卷积神经网络、双向的 GRU 网络和注意力机制的优势, 融合三种模型, 提出一种基于注意力机制的 CNN-BiGRU 的短时交通流预测模型。首先, CNN 提取交通流的空间特征, 然后利用 BiGRU 从前向和后向传播中充分提取交通流的时间相关特征, 并用 Attention 机制加强了关键信息的影响, 将交通流最重要的特征得到了最大限度的保留, 以此来提升模型的预测精度。最后, 以真实数据集为样本数据做了对比实验和消融实验。对比实验结果都表明, 相较于 HA、BP、CNN-GRU 和 CNN-GRU-Attention 四种基线模型, 基于注意力机制的 CNN-BiGRU 模型具有更好的预测效果。与变体 CNN、BiGRU、BiGRU-Attention、CNN-BiGRU 等的消融实验结果也表明, 基于注意力机制的 CNN-BiGRU 模型预测效果是最好的。在后续研究中, 我们将考虑天气、交通事故等外部因素, 进一步提升模型的预测准确性。

基金项目

国家自然科学基金/National Natural Science Foundation of China (71761031)。

参考文献

- [1] Liu, Z.M. and Yu, L. (2020) Stay or Leave? The Role of Air Pollution in Urban Migration Choices. *Ecological Economics*, **177**, Article ID: 106780. <https://doi.org/10.1016/j.ecolecon.2020.106780>
- [2] Rudolph, F. and Matrai, T. (2018) Congestion from a Multimodal Perspective. *Periodica Polytechnica Transportation Engineering*, **46**, 215-221. <https://doi.org/10.3311/PPtr.12048>
- [3] Fusco, G., Colombroni, C. and Isaenko, N. (2016) Short-Term Speed Predictions Exploiting Big Data on Large Urban Road Networks. *Transportation Research Part C Emerging Technologies*, **73**, 183-201. <https://doi.org/10.1016/j.trc.2016.10.019>
- [4] Oh, S., Byon, Y.J., Jang, K. and Yeo, H. (2018) Short-Term Travel Time Prediction on Highway: A Review on Model

- Based Approach. *KSCE Journal of Civil Engineering*, **22**, 298-310. <https://doi.org/10.1007/s12205-017-0535-8>
- [5] Zhang, K., Xie, C., Wang, Y., Miguel Ángel, S., Nguyen, T.M.T., Zhao, Q., *et al.* (2021) Hybrid Short-Term Traffic Forecasting Architecture and Mechanisms for Reservation-Based Cooperative ITS. *Journal of Systems Architecture*, **117**, Article ID: 102101. <https://doi.org/10.1016/j.sysarc.2021.102101>
- [6] Zhang, Y., Zhang, Y.L. and Haghani, A. (2014) A Hybrid Short-Term Traffic Flow Forecasting Method Based on Spectral Analysis and Statistical Volatility Model. *Transportation Research Part C Emerging Technologies*, **43**, 65-78. <https://doi.org/10.1016/j.trc.2013.11.011>
- [7] 李文勇, 李俊卓, 王涛. 基于 Box-Cox 指数变换改进的 ARIMA 模型交通流预测方法[J]. 武汉理工大学学报(交通科学与工程版), 2020, 44(6): 974-977.
- [8] 周晓, 唐宇舟, 刘强. 基于卡尔曼滤波的道路平均速度预测模型研究[J]. 浙江工业大学学报, 2020, 48(4): 392-396+404.
- [9] Zhu, Z., Peng, B., Xiong, C.F. and Zhang, L. (2016) Short-Term Traffic Flow Prediction with Linear Conditional Gaussian Bayesian Network. *Journal of Advanced Transportation*, **50**, 1111-1123. <https://doi.org/10.1002/atr.1392>
- [10] Ryu, U., Wang, J., Kim, T., Kwak, S. and Juhyok, U. (2018) Construction of Traffic State Vector Using Mutual Information for Short-Term Traffic Flow Prediction. *Transportation Research Part C: Emerging Technologies*, **96**, 55-71. <https://doi.org/10.1016/j.trc.2018.09.015>
- [11] Tang, J., Chen, X., Hu, Z., Zong, F., Han, C. and Li, L. (2019) Traffic Flow Prediction Based on Combination of Support Vector Machine and Data Denoising Schemes. *Physical A: Statistica Mechanica and Its Applications*, **534**, Article ID: 120642. <https://doi.org/10.1016/j.physa.2019.03.007>
- [12] Xie, P., Li, T.R., Liu, J., Du, S., Yang, X. and Zhang, J. (2020) Urban Flow Prediction from Spatiotemporal Data Using Machine Learning: A Survey. *Information Fusion*, **59**, 1-12. <https://doi.org/10.1016/j.inffus.2020.01.002>
- [13] 杜圣东, 李天瑞, 杨燕, 等. 一种基于序列到序列时空注意力学习的交通流预测模型[J]. 计算机研究与发展, 2020, 57(8): 1715-1728.
- [14] 卢生巧, 黄中祥, 等. 基于深度学习的短时交通流预测模型[J]. 交通科学与工程, 2020, 36(3): 74-80.
- [15] 陆文琦, 芮一康, 冉斌, 等. 智能网联环境下基于混合深度学习的交通流预测模型[J]. 交通运输系统工程与信息, 2020, 20(3): 47-53.
- [16] 王博文, 王景升, 等. 基于长短时记忆网络的 Encoder-Decoder 多步交通流预测模型[J]. 重庆大学学报, 2021, 44(11): 71-80.
- [17] 潘卫鹏, 郭唐仪, 等. 基于图注意力卷积网络的短时交通流预测[C]//世界交通运输工程技术论坛(WTC2021)论文集(上). 北京: 人民交通出版社, 2021: 2495-2500.
- [18] Ye, J.X., Zhao, J., Ye, K.J. and Xu, C.Z. (2020) How to Build a Graph-Based Deep Learning Architecture in Traffic Domain: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, **99**, 1-21. <https://doi.org/10.1109/TITS.2020.3043250>
- [19] Seng, D., Lv, F., Ling, Z., Shi, X. and Fang, Q. (2021) Forecasting Traffic Flows in Irregular Regions with Multi-Graph Convolutional Network and Gated Recurrent Unit. *Frontiers of Information Technology & Electronic Engineering*, **22**, 1179-1193. <https://doi.org/10.1631/FITEE.2000243>
- [20] Peng, H., Du, B., Liu, M., Liu, M., Ji, S., Wang, S., *et al.* (2021) Dynamic Graph Convolutional Network for Long-Term Traffic Flow Prediction with Reinforcement Learning. *Information Sciences*, **578**, 401-416. <https://doi.org/10.1016/j.ins.2021.07.007>
- [21] Jiang, Z.B., Fan, W., Liu, W., Zhu, B. and Gu, J. (2018) Reinforcement Learning Approach for Coordinated Passenger Inflow Control of Urban Rail Transit in Peak Hours. *Transportation Research Part C Emerging Technologies*, **88**, 1-16. <https://doi.org/10.1016/j.trc.2018.01.008>
- [22] 邱锡鹏. 神经网络与深度学习[M]. 北京: 机械工业出版社, 2021: 193-196.
- [23] Cho, K., Merriënboer, B.V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., *et al.* (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. arXiv:1406.1078. <https://arxiv.org/abs/1406.1078>