

刘思林, 廖祝华, 符琦, 等. 多影响因素下的交通流速度预测[J]. 湖南科技大学学报(自然科学版), 2024, 39(3): 109-115. doi:10.13582/j.cnki.1672-9102.2024.03.014

LIU S L, LIAO Z H, FU Q, et al. Prediction of Traffic Flow Speed Under Multiple Influencing Factors [J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2024, 39(3): 109-115. doi:10.13582/j.cnki.1672-9102.2024.03.014

多影响因素下的交通流速度预测

刘思林^{1,2}, 廖祝华^{1,2*}, 符琦³, 刘毅志^{1,2}, 赵肄江^{1,2}

(1. 湖南科技大学 计算机科学与工程学院, 湖南 湘潭 411201;

2. 湖南科技大学 服务计算与软件服务新技术湖南省重点实验室, 湖南 湘潭 411201; 3. 长沙工业学院, 湖南 长沙 410200)

摘要: 及时、准确的交通流预测对于车辆导航规划和智能交通调度具有重要作用。道路交通不仅具有时空相关性, 同时多种环境因素还会对交通状况造成重要影响。为提高道路交通流速度预测的准确率, 首先对试验选择地区的降雨程度、空气污染程度数据进行分级处理, 时间周期划分工作日与非工作日; 然后结合双向长短期记忆网络和特征工程技术, 建立基于多影响因素的交通流速度预测模型(MF-TPM), 并在公开的地区交通速度时序数据集与天气数据集上建模分析; 最后基于大规模真实交通数据进行试验, 结果表明: MF-TPM的预测精度比常用的长短期记忆网络、卷积神经网络和双向长短期记忆网络模型分别高2.20%、4.94%和0.63%。在不同的降雨程度和空气污染等级下, MF-TPM同样具有最佳的预测表现。

关键词: 智能交通; 速度预测; 影响因素; 深度学习; BiLSTM

中图分类号: TP311 文献标志码: A 文章编号: 1672-9102(2024)03-0109-07

Prediction of Traffic Flow Speed Under Multiple Influencing Factors

LIU Silin^{1,2}, LIAO Zhuhua^{1,2*}, FU Qi³, LIU Yizhi^{1,2}, ZHAO Yijiang^{1,2}

(1. School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China;

2. Hunan Provincial Key Laboratory for Service Computing and Novel Software Technology,

Hunan University of Science and Technology, Xiangtan 411201, China; 3. Changsha Institute of Technology, Changsha 410200, China)

Abstract: Timely and accurate traffic flow prediction plays an important role in navigation planning and intelligent traffic dispatch. Road traffic is not only temporal and spatial correlation, but also a variety of environmental factors will have an important impact on traffic conditions. In order to improve the accuracy of road traffic flow speed prediction, firstly, the data of rainfall degrees and air pollution levels are graded, and the time period is divided into working days and non-working days. Then combined with Bidirectional Long-Term and Short-Term Memory network (BiLSTM) and feature engineering technology, we need to establish a Multi-Factor-based Traffic Flow Speed Prediction Model (MF-TPM), and model and analyze the public regional traffic speed time series datasets and weather datasets. Finally, experiments based on large-scale real traffic data show that the prediction accuracy of MF-TPM is 2.20%, 4.94% and 0.63% higher than the commonly used Long-Short-Term Memory network (LSTM), Convolutional Neural Network (CNN) and BiLSTM network models, respectively. MF-TPM also has the best prediction performance under different rainfall levels and air pollution levels.

Keywords: intelligent transportation; speed prediction; influencing factors; deep learning; BiLSTM

收稿日期: 2022-03-10

基金项目: 湖南省自然科学基金资助项目(2021JJ30276); 湖南省教育厅科学研究重点项目资助(19A172); 国家自然科学基金资助项目(41871320)

* 通信作者, E-mail: zhiliao@hnust.edu.cn

准确、实时的交通速度预测能够根据当前交通网络变化规律,帮助交通部门及时了解某路段未来某时间段车流的平均速度,对城市道路进行交通规划以及交通诱导.因此,准确的交通速度预测^[1]的研究具有重要的现实意义和应用前景,一直是智能交通领域^[2]的研究热点.

近年来,交通流预测得到了国内外学者的广泛研究,许多理论和方法^[3-4]被应用于交通预测研究.交通流预测也是基于动态获取的交通流时间序列数据来预测未来的交通流状态.在现代交通系统中,每天都会产生大量的交通数据,导致短期流量预测(STTP)的快速发展,而深度学习方法在STTP中得到了应用^[5].在具有复杂时空关系的交通网络中,深度神经网络通常表现良好,因为它们能够自动提取最重要的特征和模式.文献[6]将图卷积网络(GCN)嵌入LSTM,命名为GC-LSTM,用于动态网络链接预测.它针对不同类型数据集都能取得比传统方法更好的性能.利用BiLSTM^[7]模型构建短时交通流预测模型,可捕捉向前和向后时段各相关特征因子对预测模型的影响^[8-9].文献[10]利用ConvLSTM模块对相邻区域的短时交通流数据进行处理,提取时空特征,然后使用双向LSTM对预测点的历史交通数据进行处理,提取交通流数据的周期特征.为准确挖掘城市路网中的时空特性,文献[11]针对数据连续缺失情况下交通流预测精度下降甚至失效的问题,提出一种时空卷积神经网络预测模型,以实现数据缺失情况下的交通流预测;文献[12]通过建立泊松分布模型和马尔可夫模型,分析区域内交通事故易发点在晴天、阴天、雨天发生交通事故概率和在路线中的概率,然后构建风险系数指标,并针对模型求得的路线安全系数采用最大决策法选出最优安全路线;文献[13]研究主干道街道上的天气状况和交通运行状况之间的关系,并发现随着降雨降水深度的增加,速度降低,而交通流量指标之间的关联强度因降雨深度而异;文献[14]分析了不同天气模式对城市快速路交通速度的影响,表明降雨天气、雾霾天气导致交通量较正常天气有明显降低,而雪天拥堵指数也比非雪天拥堵指数更高.

交通流受多种因素影响,例如天气状态、空气污染程度、法定节假日活动、出行早高峰^[15]等,且交通流具有时空相关性.现有交通预测研究考虑的变量单一,较少结合多种影响交通流的因素进行建模预测^[16].本文为提高预测准确率,对降雨程度、空气污染程度进行分级,并将时间属性分为工作日与非工作日;然后在实际交通流的规律性分析的基础上,建立基于多影响因素的交通流速度预测模型(MF-TPM),并在不同的环境下验证模型的可用性.

1 基于多影响因素的交通流速度预测模型(MF-TPM)

MF-TPM模型整体结构如图1所示,模型考虑了影响交通流的多种环境因素和时空相关性的特点,对多源数据进行融合处理,提取多种特征变量,并建立特征工程.其中,气象特征包括每10 min记录的实时气温及天气状态;空气质量变量为每10 min记录的环境空气质量指数(AQI);时间属性为该时间点的小时数和工作日属性.序列数据按照时间前后进行排列重构并构建特征矩阵^[17],执行窗口拆分后输入BiLSTM训练,验证后通过训练后的模型进行交通流速度预测.

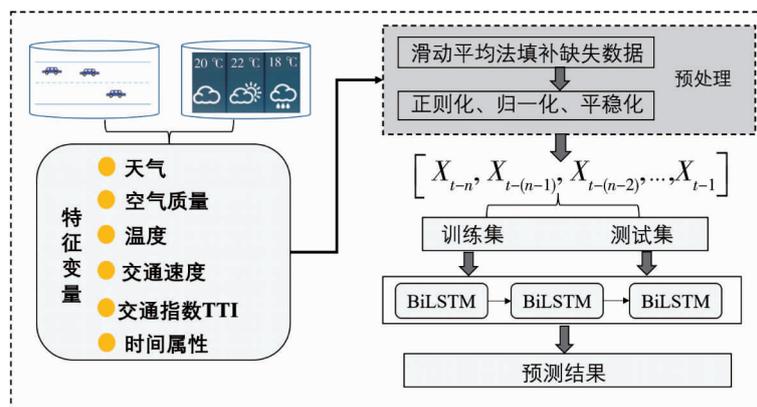


图1 MF-TPM 结构

该 MF-TPM 结合了 BiLSTM 方法,其中 BiLSTM 结构如图 2 所示.

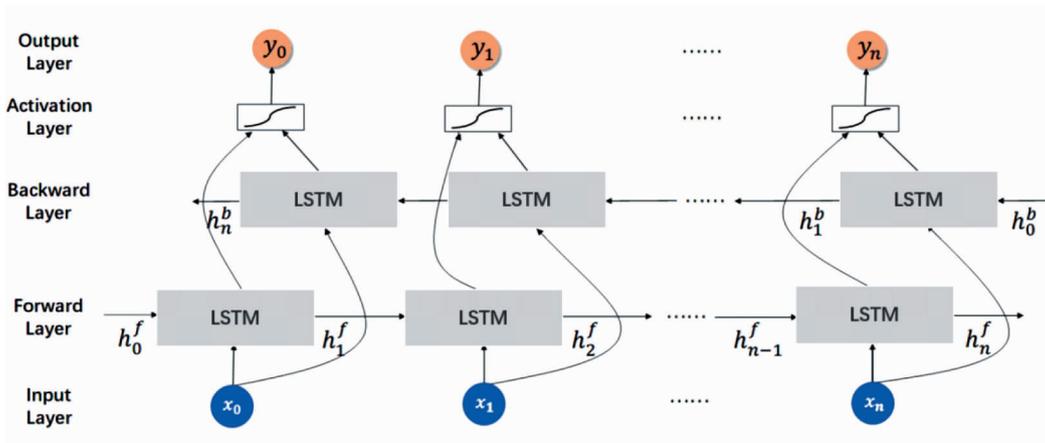


图 2 双向长短期记忆网络

BiLSTM 分为 2 个独立的 LSTM,输入序列分别以正序和逆序输入至 2 个 LSTM 神经网络进行特征提取,将 2 个输出向量(即提取后的特征向量)进行拼接后形成的词向量作为该词的最终特征表达.BiLSTM 的模型设计理念是使 t 时刻所获得特征数据同时拥有过去和将来之间的信息.试验证明:这种神经网络结构模型对文本特征提取效率和性能要优于单个 LSTM 结构模型^[18].

2 数据集获取与处理

2.1 天气和空气质量数据处理

试验所需的真实交通数据选取 2018 年全年成都市道路行驶速度与 TTI(Travel Time Index)时序数据集,为获取同时段的交通外部环境数据,本文通过天气预报网站获取 2018 年全年成都市气象站点记录的历史气象观测数据,包括每 10 min 观测的温度、风速、天气状态、空气质量指数(Air Quality Index, AQI)资料.数据结构如表 1 所示.

表 1 天气及空气质量指数数据样例

路段名	日期	时间	温度/℃	天气	风速/(m/s)	AQI
17	2018-01-01	0:00:00	5	阴	0	115
17	2018-06-26	5:00:00	23	中雨	0	23
17	2018-09-23	17:00:00	25	小雨	0	63

根据《环境空气质量指数(AQI)技术规定(试行)》^[19]规定:空气污染指数划分为 0~50,51~100,101~150,151~200,201~300 和大于 300 这 6 档.这些对应于空气质量的 6 个级别,级别越高,说明污染越严重;数据集中的天气状态分为晴天、阴天、小雨、中雨、大雨等 5 种类型.

2.2 道路拓扑及交通运行时序数据处理

本研究使用 2018 年全年成都市道路行驶速度与 TTI 时序数据集,数据内容包括各城市级、辖区级、道路级滴滴基于浮动车轨迹的城市交通指数^[20]计算服务所得 TTI 和平均行驶速度.

TTI 是业内使用较多的城市拥堵程度的评价指标,是反映实际花费的行程时间与自由流花费行程时间的比值关系,值越大表示交通运行状态越差.

数据集中车辆实际速度估算方法:如果一条路段(link)有 2 个连续时间片,分别为 t_1, t_2 , link 长度为 S ,那么 t_1 到 t_2 这段时间内,link 的平均速度 V 为时间平均速度,计算方式见式(1).

$$V = 2 \frac{S}{t_1 + t_2} \tag{1}$$

在同一条 link 且在一个时间片内,TTI 指数计算的基本方法如式(2).

$$TTI = \text{自由流速度} / \text{实际速度} \tag{2}$$

式中:自由流速度为路段在低交通量,低密度情况下机动车所能达到的行程速度.

试验所用数据集中各字段的具体解释、示例如表 2 所示.

表 2 原始数据解释与示例

字段	释义	数据示例
obj_id	TTI 对象 id	841
batch_time	时间	2018-01-01 00:00:00:00
TTI	交通指数数据	1.186 65
speed	平均速度	47.398 3(km/h)
geom	TTI 对象几何范围	MULTILINESTRING((104.13743 30.60591,104.1383230.60538), ...)

试验选取成都市内 120 条道路作为实验路段,数据集内路段速度记录示例见表 3.

表 3 道路速度与 TTI 数据示例

Obj_id	Batch_time	TTI	Speed
841	2018-01-01 00:00:00	1.186 65	47.398 3
839	2018-01-01 00:00:00	1.025 90	94.673 0
836	2018-01-01 00:00:00	1.226 58	41.218 7

2.3 提取影响交通流特征变量并融合处理

首先,提取时间、速度、TTI、温度、天气状态、AQI 这 6 种特征,并进行融合与处理,得到道路行驶相关的多特征融合数据,结构如表 4 所示.

表 4 多特征融合的道路行驶序列数据样例

时间	TTI	速度	天气	温度/℃	AQI
2018-01-01 00:10:00	1.102 02	76.509 5	阴	5	115
2018-01-01 00:20:00	1.101 31	77.079 2	阴	5	115
2018-01-01 00:30:00	1.094 17	76.251 5	阴	5	115

其次,将序列数据按照时间前后进行排列重构,构建特征矩阵如式(3).

$$\begin{bmatrix} S_{t-n} \\ S_{t-(n-1)} \\ S_{t-(n-2)} \\ \vdots \\ S_t \end{bmatrix} = \begin{bmatrix} V_{t-n} & a_{t-n} & b_{t-n} & \cdots & f_t \\ V_{t-(n-1)} & a_{t-(n-1)} & b_{t-(n-1)} & \cdots & f_{t-(n-1)} \\ V_{t-(n-2)} & a_{t-(n-2)} & b_{t-(n-2)} & \cdots & f_{t-(n-2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ V_t & a_t & b_t & \cdots & f_t \end{bmatrix}. \quad (3)$$

式中: S_t 为 t 时刻的特征值集合; V_t 为 t 时刻该道路的平均速度,即数据集中路段 id 对应的 speed 值; $a_t \sim f_t$ 为 t 时刻输入的特征变量.

数据集选取 70% 的数据作为训练集,其余数据为测试集.预测目标为未来一段时间的平均速度.处理后的数据用于具有 1 个输入层、2 个隐藏层、1 个输出层的 BiLSTM 网络训练并进行交通流预测.Dropout 层(丢失率设置为 0.2)来防止过拟合,最后通过 Dense 全连接层输出.

3 试验结果分析

3.1 考虑多种交通影响因素的交通流平均速度时空性分析

为了分析降雨和工作日与否对道路中车辆平均速度的影响,将数据集划分工作日和非工作日、雨天和非雨天这 4 个数据集,每个数据集大小相同.同时分析不同环路的平均速度均值,主要选择成都市区一环~四环的 120 条道路,并统计分析不同时空特点下各环路的平均速度变化,分析结果如图 3 所示.

从图 3 中可见:交通流速度变化在不同的时间和空间上表现出不同的特性和变化模式,同时受降雨、时间等多种因素影响.而且发现越靠近城市中心道路行驶速度越小,因此越靠近城市中心环路的道路平均

速度越小,同时降雨使得道路的平均速度降低,而环路平均速度在工作日与非工作日下的差别较小.

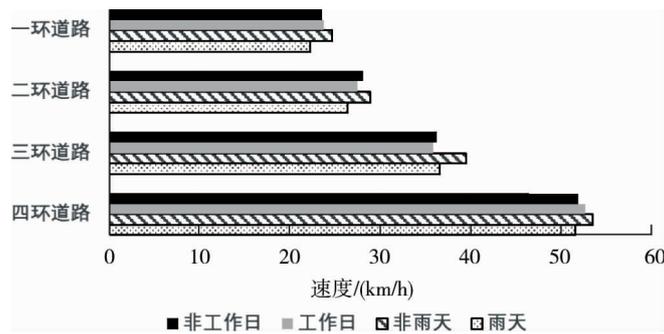


图 3 不同影响因素下的道路交通流平均速度

道路交通流平均速度在多种因素的影响下随时间变化体现出规律性.为分析具体路段的速度变化和规律,我们利用历史交通流速度数据,分析标准周内的道路速度变化.图 4 是一个标准周内的路段交通流平均速度变化,从中可见非工作日的交通流速度变化模式与工作日有明显区别.

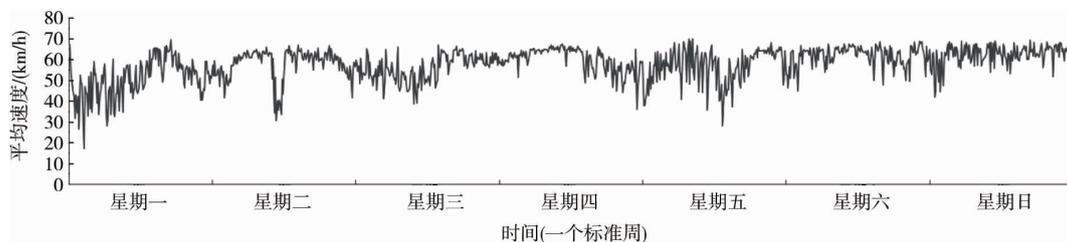


图 4 一个标准周内路段交通流平均速度曲线

图 5 是一个标准周内各天的速度变化对比曲线图.通过分别对各天速度进行统计分析,可知非工作日的速度平均值最大,同时速度离散程度最小,拥有最小的标准差.而周一的速度离散值最大,标准差最大,但速度平均值最小.

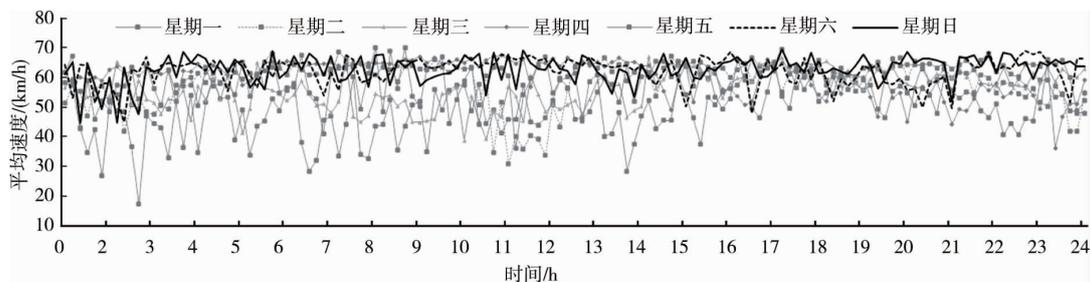


图 5 一个标准周内各天的速度变化对比

3.2 模型预测试验评价与对比

本文引入 3 种误差评价指标来评价模型预测性能,指标计算方式见式(4)~式(6):

均方根误差(RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}. \tag{4}$$

平均绝对误差(MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|. \tag{5}$$

平均绝对百分比误差(MAPE):

$$MAPE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|. \tag{6}$$

式中: y_i 为交通流的真实值; \hat{y}_i 为交通流的预测值; N 为选取样本的数量.

图6是本文预测模型与随机环境下 BiLSTM 预测的路段全天平均速度和实际平均速度值的对比曲线图.与随机环境下的 BiLSTM 模型预测结果相比,融合了环境特征变量的 MF-TPM 模型预测值与真实值拟合程度更高且稳定性更好.

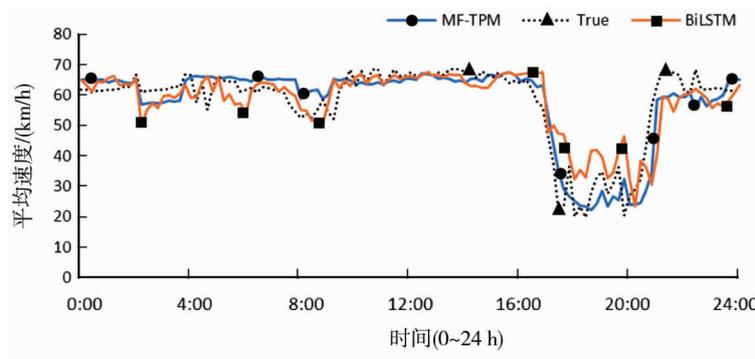


图6 速度预测值与真实值对比

各模型预测性能对比试验结果见表5.从表5中可见:MF-TPM 预测效果要优于其他模型,而 CNN 模型预测误差最大,表现出较差的预测性能.

表5 各模型预测性能

Model	MAPE/%	MAE	RMSE
STM	12.26	3.88	5.84
CNN	15.00	4.62	7.36
BiLSTM	10.69	3.54	5.17
MF-TPM	10.06	3.39	5.01

3.3 降雨程度和空气污染程度对模型预测性能的影响分析

按照天气状态和空气质量等级对测试集进行划分后, MF-TPM 与 BiLSTM, LSTM, CNN 这3种预测模型进行 MAPE 值对比, 试验结果见图7和图8.

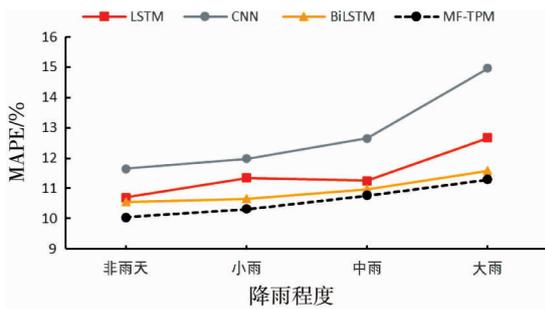


图7 不同降雨程度下各模型 MAPE 值对比

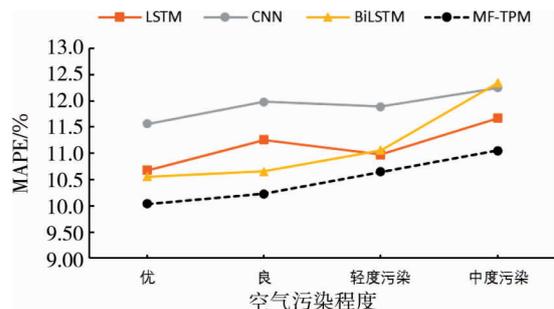


图8 不同空气污染程度下各模型 MAPE 值对比

图7表明各模型在非雨天和小雨天的预测精度变化值较小,随着降雨程度的增强,各个预测模型的预测误差也逐渐增大, CNN 的预测性能表现依旧最差, MF-TPM 误差增长程度最为平缓,准确度最高.从图8的折线图可以看出:不同空气污染状况对各预测模型影响程度不同(实际数据集样本无空气重度污染数据),中度污染以下的空气质量对各模型的预测精度影响较轻,而 MF-TPM 误差最小且表现平稳.

4 结论

1) 提出考虑多种环境因素和时空相关性特点的交通流速度预测模型 MF-TPM,对多源数据进行了融合处理,提取多种影响交通流速度变化的特征变量,并建立特征工程.

2) MF-TPM 模型通过数据集划分,建立工作日、非工作日影响因子,并将降雨程度和空气污染程度进

行分级作为更精细的影响因子,充分考虑了交通流数据的时序特点和环境影响因素。

3)在不同降雨程度和空气污染等级情况下, MF-TPM 模型具有更准确和平稳的预测表现,以及自适应环境的能力。

参考文献:

- [1] QU L C, LYU J, LI W, et al. Features injected recurrent neural networks for short-term traffic speed prediction [J]. *Neurocomputing*, 2021, 451: 290–304.
- [2] TELANG S, CHEL A, NEMADE A, et al. Intelligent transport system for a smart city [M]//TAMANE SC, DEY N, HASSANIEN AE. *Security and Privacy Applications for Smart City Development*. Cham: Springer, 2021: 171–187.
- [3] MISHRA S, SHIBU A, BALAN R, et al. An evolutionary approach for congestion prediction on IoT data streams in smart city environment [C]//*Developments of Artificial Intelligence Technologies in Computation and Robotics*. WORLD SCIENTIFIC, 2020: 84–92.
- [4] NAGY A M, SIMON V. Survey on traffic prediction in smart cities [J]. *Pervasive and Mobile Computing*, 2018, 50: 148–163.
- [5] LEE K, EO M, JUNG E, et al. Short-term traffic prediction with deep neural networks; a survey [J]. *IEEE Access*, 2021, 9: 54739–54756.
- [6] CHEN J, XU X, WU Y, et al. GC-LSTM: graph convolution embedded LSTM for dynamic network link prediction [J]. *Applied Intelligence*, 2018, 52: 7513 – 7528.
- [7] PENG T, ZHANG C, ZHOU J Z, et al. An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting [J]. *Energy*, 2021, 221: 119887.
- [8] CUI Z Y, KE R M, PU Z Y, et al. Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction [EB/OL]. 2018: arXiv: 1801.02143. <http://arxiv.org/abs/1801.02143>
- [9] CUI Z Y, KE R M, PU Z Y, et al. Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values [J]. *Transportation Research Part C: Emerging Technologies*, 2020, 118: 102674.
- [10] LIU Y P, ZHENG H F, FENG X X, et al. Short-term traffic flow prediction with Conv-LSTM [C]//2017 9th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2017: 1–6.
- [11] 张壮壮, 屈立成, 李翔, 等. 基于时空卷积神经网络的数据缺失交通流预测 [J]. *计算机工程与应用*, 2022, 58(7): 259–265.
- [12] 章诗琪, 魏斐斐, 范馨月. 受天气与交通事故影响的安全路线选择模型 [J]. *计算机工程与应用*, 2021, 57(10): 246–251.
- [13] SHBEEB L, HAZIM N. Modelling the Effects of Environmental Factors on Traffic Flow Parameters [J]. *International Journal of Engineering Research and Technology*, 2019, 12(12): 2910–2924.
- [14] 徐月欣. 不同天气条件下的城市快速路交通拥堵评价与短时预测 [D]. 西安: 长安大学, 2017.
- [15] 史文彬. 时间序列的相关性及信息熵分析 [D]. 北京: 北京交通大学, 2016.
- [16] 赵宝新. 基于多源数据融合的城市交通拥堵扩散的研究及其应用 [D]. 深圳: 中国科学院大学(中国科学院深圳先进技术研究院), 2020.
- [17] 许得杰, 毛保华, 戎亚萍. 限速区段列车流特性及运行延误研究 [J]. *计算机工程与应用*, 2015, 51(5): 14–18.
- [18] MA C X, DAI G W, ZHOU J B. Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(6): 5615–5624.
- [19] 潘本锋, 宫正宇, 王帅, 等. 环境空气质量指数在应用中存在的问题及建议 [J]. *中国环境监测*, 2015, 31(1): 64–67.
- [20] 徐锦强, 黄海南, 李林, 等. 基于网络拓扑及交通特性的城市路网弹性评价 [J]. *华东交通大学学报*, 2020, 37(3): 52–59.