

引入通勤特征的车辆行程时间组合预测方法

和雪儿¹, 岳凡凡², 洪源伯³, 黄超铄^{*3}

1. 广州市高速公路有限公司, 广东 广州 510320

2. 广州市交通设计研究院有限公司, 广东 广州 510320

3. 华南理工大学土木与交通学院, 广东 广州 510640

DOI: 10.61369/SSSD.2025080008

摘 要 : 为了缓解通勤高峰期的交通拥堵, 提升交通运行效率和舒适度。本文基于高速公路收费数据集, 设计了表征车辆个体通勤行为特征的指标, 引入通勤特征、天气特征等外部特征, 结合时间等内部特征, 采用4种常用的算法: K-最近邻算法(KNN)、极端梯度提升算法(XGBoost)、多层感知机算法(MLP)、长短时记忆神经网络算法(LSTM)对流量进行预测, 最后采用贝叶斯回归方法对各算法进行融合。选取跨粤港澳大湾区各市、各区县通勤OD交通量较大, 车辆往来频繁的15个路段作为实验对象, 以15min为时间粒度进行统计, 输入步长为8, 预测步长为4, 对车辆的行程时间进行预测。分别对引入通勤特征前后模型效果, 各模型预测误差, 峰值预测效果、模型动态权重几个方面展开分析。结果表明: 引入通勤特征变量能有效提升模型预测准确率, 贝叶斯组合模型在预测精度、预测稳定性及峰值预测效果方面均表现更优, 采用动态调整权重的方式进行融合能克服单一预测方法的局限性, 提升整体模型性能。

关 键 词 : 交通运输规划与管理; 高速公路; 流量预测; 机器学习; 组合预测

Combined Prediction Method of Vehicle Travel Time Incorporating Commuting Characteristics

He Xue'er¹, Yue Fanfan², Hong Yuanbo³, Huang Chaoshuo^{*3}

1.Guangzhou Expressway Co., Ltd., Guangzhou, Guangdong 510320

2.Guangzhou Transportation Design & Research Institute Co., Ltd., Guangzhou, Guangdong 510320

3.School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, Guangdong 510640

Abstract : To alleviate traffic congestion during peak commuting hours and enhance traffic operation efficiency and comfort, this paper designs indicators characterizing the individual commuting behavior of vehicles based on expressway toll datasets. By introducing external characteristics such as commuting characteristics and weather characteristics, and combining internal characteristics like time, four common algorithms are adopted: K-Nearest Neighbor (KNN), eXtreme Gradient Boosting (XGBoost), Multi-Layer Perceptron (MLP), and Long Short-Term Memory neural network (LSTM) to predict traffic flow. Finally, the Bayesian regression method is used to integrate these algorithms. Fifteen road sections with large commuting OD traffic volumes and frequent vehicle movements across cities and districts in the Guangdong-Hong Kong-Macao Greater Bay Area are selected as experimental objects. Statistics are carried out with a time granularity of 15 minutes, an input step size of 8, and a prediction step size of 4 to predict the travel time of vehicles. Analyses are conducted from several aspects, including the model performance before and after introducing commuting characteristics, the prediction errors of each model, the peak prediction effect, and the dynamic weights of the model. The results show that introducing commuting characteristic variables can effectively improve the prediction accuracy of the model. The Bayesian combined model performs better in terms of prediction accuracy, prediction stability, and peak prediction effect. The integration method using dynamically adjusted weights can overcome the limitations of single prediction methods and enhance the overall performance of the model.

Keywords : transportation planning and management; expressway; traffic flow prediction; machine learning; combined prediction

行程时间预测在交通出行和管控等方面扮演重要角色，高效、精准的行程时间预测方法是交通领域学者的研究热点。随着数据采集和检测技术的发展，数据驱动的行程时间预测方法日趋成熟。与传统方法相比，数据驱动方法可根据历史交通数据的统计规律推断变化趋势，具有灵活的优势^[1,2]。其中，参数方法通过建立数学模型来描述行程时间与相关变量之间的关系，以预测未来的行程时间。但需要对模型进行合理的假设和参数选择，这可能会影响预测结果的准确性，导致预测性能不稳定。而非参数方法不需要事先对相关变量的分布做出假设，可以从海量交通数据中挖掘规律，能较好地处理交通流的随机性和非线性特征^[3-5]。

为克服单一模型预测存在的局限性，本研究采用组合预测模型提高模型的整体性能。本研究基于高速公路收费数据集，首先设计表征车辆个体通勤行为特征的9个指标，引入通勤特征，气温、降雨、风速等天气特征，结合时序、节假日、星期、早晚高峰期等时间特征作为输入参数，分别采用KNN、XGBoost、MLP、LSTM对路段行程时间进行预测，再采用贝叶斯回归方法对基模型的预测结果进行动态加权融合，形成一种引入通勤特征的车辆行程时间组合预测模型，并从引入通勤特征前后模型效果，平均误差，峰值预测准确率、模型动态权重等方面进行综合评估^[6-8]。

一、预测模型

（一）数据来源

本研究采用高速公路收费流水数据，该数据集具有数据量大、字段丰富以及数据质量高的优点。高速公路收费流水数据集具有各种类型的出行信息，对数据集进行精简后保留如下字段：通行标识ID、车辆车牌号、入口收费站编号、入口收费站名称、驶入时间、出口收费站编号、出口收费站名称、驶出时间、车型代码、车种代码、计费总里程数。天气数据来源于国家气象信息中心。数据字段包括站点代码、年份、月份、日期、小时、气温、降雨量和风速等，其中3类天气数据均为连续型数据，时间粒度为小时级别。

二、实验案例与分析

实验选取粤港澳大湾区跨各市、各区县通勤OD交通量较大，通勤车辆往来频繁的15个路段作为研究对象，站点间的平均行程时间大概在5min-100min不等，平均里程在5公里-90公里之间。

选取15个OD站点时间范围在2023年9月11日-30日全天00:00-23:59的收费数据和天气数据，以为时间间隔，统计得到车辆平均行程时间作为预测变量。总计有15（OD站点）*20（天）*96（时段）=28800行数据。采用滑动时间窗的数据结构输入，同时采用多个步长同时输出的预测模式，输入步长8，预测步长范围4，实验中，对KNN模型、XGBoost模型、MLP模型和LSTM模型采用相同的输入，贝叶斯回归模型的输入是前4种模型的输出特征结果。在数据集的划分上，将每个站点前70%的数据作为训练集，后30%作为测试集。

为验证引入通勤特征对于预测车辆行程时间的必要性，仅以天气特征和时间特征变量作为模型输入变量，进行车辆行程时间的预测，分别得到其不同预测步长下的平均绝对百分比误差指标，并将其与引入通勤特征后的模型结果进行对比。预测可视化结果如图1所示。

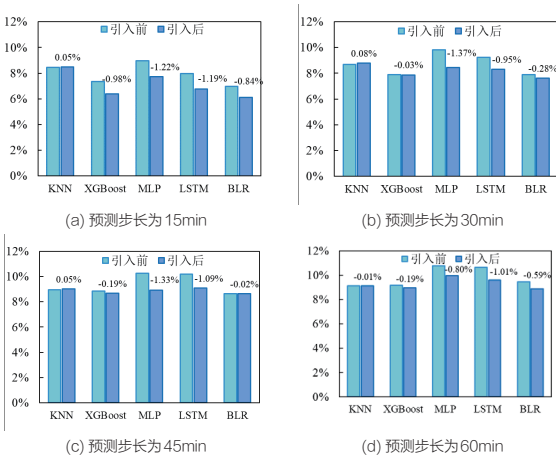


图1 引入通勤特征前后模型预测性能对比图

模型对比结果由MAPE（引入后）/MAPE（引入前）得到，观察图1，分别对比各模型在引入个体因素前后的预测效果，发现除KNN模型外，引入通勤特征后各模型的预测误差值均减小，并且减小幅度各有差异，MLP模型和LSTM在引入通勤特征后模型效果有了较大提升，MAPE值降低1%左右。此外，模型对比结果显示的普遍规律表明，在预测步长较短时，引入通勤特征对模型表现力的提升有更大的贡献。因此，在本文所研究的车辆行程时间预测问题中，引入通勤特征变量具有必要性和合理性，能有效提升预测模型的表现效果^[9]。

为分析各模型在行程时间预测中的预测性能，将KNN模型、XGBoost模型、MLP模型以及LSTM模型得到的预测结果和最后贝叶斯回归融合模型的结果进行对比，不同步长下各模型的误差如表1所示。各模型预测性能指标可视化结果如图2所示。

表1 各模型预测性能对比结果

预测步长		15min	30min	45min	60min
RMSE	KNN	8.44%	8.70%	8.96%	9.13%
	XGBoost	7.35%	7.90%	8.86%	9.14%
	MLP	8.96%	9.83%	10.25%	10.75%
	LSTM	7.96%	9.25%	10.18%	10.62%
	BLR	6.96%	7.90%	8.66%	9.48%

MAE	KNN	8.49%	8.78%	9.01%	9.12%
	XGBoost	6.37%	7.87%	8.67%	8.95%
	MLP	7.74%	8.46%	8.93%	9.95%
	LSTM	6.77%	8.30%	9.09%	9.61%
	BLR	6.11%	7.62%	8.64%	8.88%
MAPE	KNN	0.05%	0.08%	0.05%	-0.01%
	XGBoost	-0.98%	-0.03%	-0.19%	-0.19%
	MLP	-1.22%	-1.37%	-1.33%	-0.80%
	LSTM	-1.19%	-0.95%	-1.09%	-1.01%
	BLR	-0.84%	-0.28%	-0.02%	-0.59%

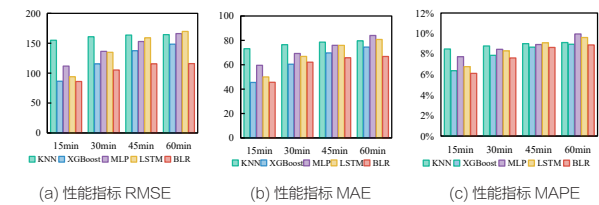


图2 各模型预测结果性能指标

观察表2和图2可知，在 KNN 模型、XGBoost 模型、MLP 模型、LSTM 模型和贝叶斯融合模型中，融合模型的预测误差最小。同时各模型的预测误差都随步长的增加而显著增加，且在不同步长下的表现各有优劣，比如在 15min 的预测步长下，表现

较好的模型为融合模型（6.11%）、XGBoost 模型（6.37%）和 LSTM 模型（6.77%），KNN 模型误差最大，为 8.49%，然而在 60min 的预测步长下，除融合模型外，表现较好的为 XGBoost 模型和 KNN 模型。说明 MLP 和 LSTM 模型受预测步长的影响更大，容易在较短的步长下表现出较好的预测性能。总的来说，对于各项性能指标（RMSE、MAE 和 MAPE），融合模型在各预测步长下都具有最小的误差值，说明与单一的预测模型相比，贝叶斯融合模型的预测精度更高^[10]。

三、结论

本文综合考虑通勤特征、天气特征、时间特征等多个特征变量，构建了基于 KNN、XGBoost、MLP 和 LSTM 的模型，并采用贝叶斯回归动态融合其预测结果的车辆行程时间组合预测模型。同时选取粤港澳大湾区几个典型的 OD 站点，对车辆的行程时间进行预测。实验结果表明：（1）引入通勤特征变量被证明是合理且必要的，它们能够有效地提升预测模型的表现效果，引入后预测模型的准确率提升 0.84%。（2）相较于单一的预测模型，贝叶斯组合模型在预测精度、预测稳定性以及峰值预测效果方面均表现更为优异，预测精度在 93.8% 以上。

参考文献

- [1] 韩超, 宋苏, 王成红. 基于 ARIMA 模型的短时交通流实时自适应预测 [J]. 系统仿真学报, 2004(07): 1530-1532.
- [2] 杨兆升, 朱中. 基于卡尔曼滤波理论的交通流量实时预测模型 [J]. 中国公路学报, 1999(03): 63-67.
- [3] 蔡昌俊, 姚恩建, 王梅英, 等. 基于乘积 ARIMA 模型的城市轨道交通进出站客流量预测 [J]. 北京交通大学学报, 2014, 38(02): 135-140.
- [4] 谭满春, 冯萃斌, 徐建闽. 基于 ARIMA 与神经网络组合模型的交通流预测 [J]. 中国公路学报, 2007(04): 118-121.
- [5] 罗向龙, 焦琴琴, 牛力瑶, 等. 基于深度学习的短时交通流预测 [J]. 计算机应用研究, 2017, 34(01): 91-93.
- [6] 张悦, 张磊, 刘佰龙, 等. 基于时空 Transformer 的多空间尺度交通预测模型 [J]. 计算机工程与科学, 2024, 46(10): 1852-1863.
- [7] 李帅, 杨柳, 赵欣卉. 基于深度学习的城市区域短时交通拥堵预测算法 [J]. 科学技术与工程, 2023, 23(25): 10866-10878.
- [8] 陆百川, 马新露, 郭鹏, 等. 基于注意力机制与时空图卷积网络的短时交通流预测模型 [J]. 交通运输工程学报, 2022, 22(4): 214-225.
- [9] 于德新, 杨苹, 高学英, 等. 基于 CEEMDAN-SSA-BiLSTM 的组合短时交通流预测模型 [J]. 吉林大学学报 (工学版), 2023, 53(5): 1420-1431.
- [10] 黄刚, 柳祖鹏, 王宝杰. 基于多组件时空融合的深度学习交通流预测 [J]. 计算机工程与应用, 2024, 60(1): 89-98.