

DOI:10.20035/j.issn.1671-2668.2022.06.004

基于路警数据与 LightGBM 算法的 高速公路行程时间预测*

景峻¹, 仝瑶^{2,3}, 李鹏¹, 么新鹏¹, 王孜健¹

(1.山东高速集团有限公司, 山东 济南 250098; 2.北京工业大学 城市交通学院, 北京 100124;

3.北京市交通工程重点实验室, 北京 100124)

摘要: 行程时间预测是支撑高速公路交通运行评价、行车诱导、交通智能化管控等应用的关键技术,在高速公路交通大数据日渐丰富的背景下,如何保障高效、准确的行程时间预测结果值得关注。为弥补现有预测数据单一、实时性不佳等问题,文中提出一种基于 LightGBM(Light Gradient Boosting Machine)算法和路警数据融合的行程时间预测模型,基于多源数据,构建交通量、大车占比、天气、日期类型、车型、路段长度、平均行程时间等多维特征集;利用山东济广(济南—广州)高速公路 ETC(电子不停车收费系统)门架系统(出入口收费站和路段 ETC 门架)、视频卡口等数据进行行程时间预测模型训练与验证,采用均方根误差 RMSE、平均绝对误差 MAE、平均绝对百分比误差 MAPE 及运算时间 4 项评价指标对该模型与最邻近(KNN)、随机森林(RF)、支持向量机回归(SVR)3 种常用机器学习算法的预测结果进行对比,结果表明采用该模型,4 个验证路段的 RMSE 为 5.78,分别比 KNN、RF、SVR 模型降低 22.8%、13.5%、21.0%,运算速度分别提高 -60%、98%、96%,可应用于高速公路网实时行程时间预测,并支持面向不同车型的差异化行程时间信息服务。

关键词: 高速公路;行程时间预测;路警融合数据;LightGBM 算法;机器学习

中图分类号:U491.14

文献标志码:A

文章编号:1671-2668(2022)06-0013-07

行程时间是现代交通管理和控制系统的关键输入,也是道路服务水平评价的常用指标,对出行者是最有价值的信息。高速公路行程时间预测是一个十分重要的课题,大量学者对此进行了研究。在预测算法方面,邢雪等采用高速公路实测历史数据构建特征集,建立了一种预测强度修正的 k-means 方法;王翔等基于最邻近算法(KNN),考虑历史交通状态特征,对高速公路短时行程时间进行了预测;刘伟铭等将最小二乘支持向量机(LSSVM)应用于行程时间预测,并利用粒子群算法(PSO)对 LSSVM 的参数进行优化,最终利用 PSO-LSSVM 模型预测行程时间;杭明升等将卡尔曼滤波成功应用于高速公路行程时间动态实时预测;江周等采用多源数据融合的方法,基于卡尔曼滤波算法建立了城市道路网络行程时间预测模型;李萌等考虑恶劣天气情况对高速公路行程时间预测的影响,将降雨量作为影响因素,利用遗传算法优化的径向神经网络(RBF)实现了不同降雨强度下行程时间预测;刘松等基于

高速公路收费站的刷卡数据获取行程时间,利用门限递归神经网络对行程时间进行了预测;林培群等对不同机器学习算法及长短时记忆神经网络模型(LSTM)进行融合,并引入高速公路时空特性,实现了不同时间步长下高速公路行程时间预测;为同时得到预测结果及置信度,文献[10-11]将贝叶斯理论纳入行程时间预测框架。上述方法用于行程时间预测各有所长,其中神经网络具有强大的非线性映射能力,能处理海量数据;机器学习有着更好的解释性,但在处理大量数据时力不从心。LightGBM 算法融合了二者的优点,其运算速度快、内存消耗低、模型精度高、支持并行训练、可处理海量数据,被广泛应用于行程时间、客流等预测。

在数据源方面,现有行程时间预测方法所依赖的数据主要源于传统交通调查、“两客一危”北斗定位及浮动车数据。传统交通调查设备检测精度不高,耐久性较差,布设与维修成本高昂;“两客一危”数据与浮动车数据仅包含旅游客车、危险品运输车

* 基金项目:山东省交通运输厅科技计划项目(2020BZ02-05)

辆或出租车与公交车,预测出的行程时间对于高速公路整体交通状况不具有代表性。此外,受限于视频卡口的覆盖范围及检测精度,视频数据在行程时间预测应用方面的效能一直无法充分展现。2019年中国高速公路“撤站”后,全国共建设完成约2.7万套ETC(电子不停车收费系统)门架,发展ETC用户约2.26亿,日均生成3亿多条天线交易数据和4亿多条车牌识别数据。如此广阔的覆盖范围和庞大的数据体量及高精度检测特性使ETC门架数据在交通流预测、交通事件识别、交通协同管控等方面具有先天优势。当前视频卡口数据主要归属于公安交警,ETC门架数据主要归属于高速公路管理部门。融合路警双方的数据源,在充分挖掘ETC门架数据的基础上,应用视频卡口数据对其进行补充,可有效提高检测器与数据采集密度,支撑更加准确的行程时间预测应用,助力高速公路精细化监管。因此,本文以高速公路ETC门架与交警视频卡口数据为基础,应用LightGBM机器学习算法进行高速公路行程时间预测,以快速获取较高精度的行程时间预测结果,为高速公路智能化实时管理控制与出行信息服务等应用提供数据支撑。

1 数据处理与特征集构建

1.1 数据来源与数据结构

研究数据源为ETC门架交易数据和视频卡口过车数据。ETC门架原始交易数据主要字段见表1,包括通行标识、通行介质、门架ID、车牌号、车牌颜色、交易时间、车型等信息。相邻点位交易信息比对结果表明,全日ETC门架检测车辆检出率可达99%,可识别车型达16种。

表1 ETC门架原始数据主要字段

字段名称	字段含义
pass_id	通行标识
media_type	通行介质
gantry_id	门架ID
vehicle_plate	车牌号/车牌号码_车牌颜色
trans_time	交易时间
vehicle_type	车型

视频卡口原始数据主要字段见表2,包括卡口桩号、车速、号牌号码、采集时间、车牌颜色、车辆颜色等信息。相较于ETC门架数据,卡口数据具有更多的交通检出信息(包括车速、分方向车道的交通量等)。值得注意的是,由于高清视频卡口检测依赖于

能见度,在夜间等能见度不足的情况下存在漏检、错检等问题。

表2 卡口原始数据主要字段

字段名称	字段含义	字段名称	字段含义
stake_mark	卡口桩号	pass_time	采集时间
vehicle_lane	车道/方向	plate_color	车牌颜色
vehicle_speed	车速	vehicle_color	车辆颜色
vehicle_plate	号牌号码		

1.2 数据预处理

由于ETC门架原始数据中仅存储交易时间数据,且ETC交易分为OBU(车载单元)收费和CPC(复合通行卡)收费两种方式,交易延迟不统一,存在实际通行时间不明确的问题,难以与视频卡口车辆通行时间匹配。视频卡口检测存在重复检测、车牌字段缺失等问题。因此,原始数据不能直接用于行程时间预测,需对数据进行预处理,校正时间并剔除异常数据。此外,两种数据源的车型、车牌分类标准存在差异,需对ETC门架及卡口数据进行关联融合处理。数据预处理流程如下:

(1) ETC门架数据时间校准。根据通行介质字段区分两种交易方式,分别针对两种交易方式,根据交易系统平均延时将交易时间字段校正为通行时间。

(2) 卡口异常数据清洗。剔除卡口原始数据中车牌号为空或错误的过车数据,对数据以30min为窗口进行去重。

(3) 路段单元构建。根据各ETC门架及卡口安装桩号及检测范围,对检测设备先后顺序进行排序,将连续两设备及其中间路段作为一路段单元,并计算路段单元长度。

(4) 路警数据融合与车辆匹配。通过车牌号、车牌颜色字段对路段单元起终点设备检测到的过车数据(ETC门架或卡口)进行关联,获取完整通过路段单元(未经立交驶入、驶出高速公路)的车辆及其通过路段起终节点的时间。

(5) 行程时间计算与异常值剔除。计算车辆行程时间,结合各路段单元长度,剔除过长或过短的行程时间数据。

1.3 特征集构建

在训练模型前,从数据集中提取对车辆行程时间可能有影响的特征。特征的选取直接影响预测精度。通过对数据的初步分析,参考文献[3,9,16—19],选取以下8个特征作为模型的输入(见表3):

表3 特征集含义及分类

特征名称	特征含义及分类
link_flow	交通量
large_mix_rate	大车占比
weather	天气(0-晴天,1-雨天,2-雾天)
is_holiday	日期类型(0-工作日,1-周末,2-节假日)
time	时间
vehicle_type	车型
link_distance	路段长度
avg_travel_time	上一预测周期平均行程时间

(1) 路段流量。当前路段历史各时段交通流量。本研究中,由于两检测器间可能存在立交及高速公路出入口,选取路段单元起点检测器流量作为

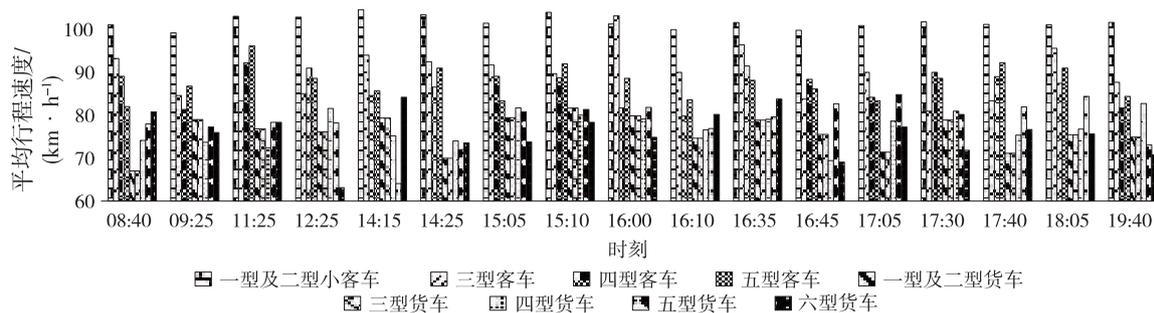


图1 各车型行程时间

(7) 路段长度。当前路段道路长度,为静态值。

(8) 平均行程时间。当前路段历史各时段的平均行程时间。当前时间路段行程时间可能与历史上一时段的行程时间具有相关性。

以 5 min 为一个时间单元进行模型训练及预测,在该时间单元内,交通状态及天气环境不会出现较大变化,且能保证样本车辆完整通过每一路段单元(检测器平均间隔 5 km,符合道路限速条件的最大行程时间为 5 min)。

将数据集中的参数 link_flow、large_mix_rate、weather、is_holiday、time、vehicle_type、link_distance、avg_travel_time 分别设为 $U = \{X_1, X_2, X_3, \dots, X_8\}$,基于数据的历史时间序列对原始数据集进行重构,形成预测的特征集:

$$Q = \left\{ \begin{matrix} X_1^1, X_2^1, X_3^1, \dots, X_8^1 \\ X_1^2, X_2^2, X_3^2, \dots, X_8^2 \\ \vdots \\ X_1^j, X_2^j, X_3^j, \dots, X_8^j \end{matrix} \right\}; j = 1, 2, \dots, N \quad (1)$$

式中: X_i^j 表示第 i 个特征的第 j 个历史时间段的数

路段流量。

(2) 大车混入率。当前路段历史各时段二型以上大车占总体交通量的比例。

(3) 天气。当前时刻天气状况(晴/雨/雾)。不良的天气对驾驶员驾驶行为具有较大影响,可能导致车速不稳定和行程时间增加。天气信息从互联网爬取,每小时更新。

(4) 日期类型。当前日期属工作日或节假日等。

(5) 时间。当前时段起始时间。

(6) 车型。根据 ETC 门架交易数据将车辆分为 14 种类型。以某路段 2021 年 5 月部分时段数据为例,进行分车型的行程时间统计,结果见图 1。从图 1 可见各车型间行程速度差异较大。

据集。

将特征集矩阵 Q 按行展开成一个行向量,并基于整体数据构建训练集 Q_{tre} :

$$Q_{\text{tre}} = \left\{ \begin{matrix} x_{1,1}^1, \dots, x_{1,1}^j, x_{2,1}^1, \dots, x_{2,1}^j, x_{8,1}^1, \dots, x_{8,1}^j \\ x_{1,2}^1, \dots, x_{1,2}^j, x_{2,2}^1, \dots, x_{2,2}^j, x_{8,2}^1, \dots, x_{8,2}^j \\ \dots \\ x_{1,m}^1, \dots, x_{1,m}^j, x_{2,m}^1, \dots, x_{2,m}^j, x_{8,m}^1, \dots, x_{8,m}^j \end{matrix} \right\} \quad (2)$$

式中: $x_{i,k}^j$ 表示第 i 个特征的第 j 个历史时间段的第 k 个数据; m 为每个数据特征的训练数据个数。

2 模型构建

2.1 LightGBM 算法

LightGBM 算法是 XGBoost(极限梯度提升树)的改进算法,与 XGBoost 相比有着更快的计算速度和更高的精度。LightGBM 与 XGBoost 是在 GBDT (Gradient Boosting Decision Tree)基础上发展而来的大规模并行计算提升树,与 GBDT 不同的是,LightGBM 的目标函数在损失函数的基础上增加了正则化项,考虑了树过于复杂而带来的过拟合问题。

LightGBM 的目标函数如下:

$$f_{\text{obj}} = \sum_{i=1}^n \text{Loss}(\hat{y}_i, y_i) + \sum_{t=1}^k \omega(f_t) \quad (3)$$

式中: $\text{Loss}(\cdot)$ 为模型预测的损失函数; $\omega(f_t)$ 为正则化项。

Boosting 算法是通过加法模型与前向分布算法得到的树模型。令 \hat{y}_i^t 为模型对第 i 个输入数据在第 t 个时间步的预测值, \hat{y}_i^{t-1} 为当前模型的计算值(为已知常量), $f_t(x_i)$ 为当前生成的决策树的计算附加值, 则有:

$$\hat{y}_i^t = \hat{y}_i^{t-1} + f_t(x_i) \quad (4)$$

目标函数可表示为:

$$f_{\text{obj}}^t = \sum_{i=1}^n \text{Loss}(\hat{y}_i^{t-1} + f_t(x_i), y_i) + \sum_{i=1}^t \omega(f_i) \quad (5)$$

根据式(5), 只需求出 $f_t(x_i)$, 便可进一步得到目标函数 f_{obj} 。根据二阶泰勒展开公式, 得:

$$f(x + \Delta x) = f(x) + f'(x) \cdot \Delta x + \frac{1}{2} f''(x) \cdot \Delta x^2 \quad (6)$$

将目标函数 f_{obj} 展开为二阶泰勒公式, 得:

$$f_{\text{obj}}^t = \sum_{i=1}^n \text{Loss}(\hat{y}_i^{t-1}, y_i) + \text{Loss}'(\hat{y}_i^{t-1}, y_i) \cdot f_t(x_i) + \frac{1}{2} \text{Loss}''(\hat{y}_i^{t-1}, y_i) \cdot f_t(x_i)^2 + \sum_{i=1}^t \omega(f_i) \quad (7)$$

问题的核心是怎样对特征空间进行划分, 即递归地寻找最优切分变量和切分点。最优切分变量和切分点的计算方式与 GBDT 类似, 基于贪心算法的思想, 对于每个节点, 针对每个特征采用启发式算法计算最优切分点, 再遍历所有特征变量, 找到最佳切分变量。重复上述算法直到满足停止要求。XGBoost 算法对损失函数应用了二阶泰勒展开, 相对于 GBDT 的一阶展开其精度更高; 加入正则化项, 可有效避免过拟合; 可自动对缺失值进行处理, 且可并行计算, 计算速度更快。但 XGBoost 算法在计算最佳切分变量及最优切分点时需遍历每个特征及每个数据, 会消耗大量内存, 计算复杂。LightGBM 算法在 XGBoost 算法之上进行改进, 使用带深度限制的 Leaf-wise 算法防止模型过拟合, 采用直方图加速算法降低内存消耗和计算复杂度。

带深度限制的 Leaf-wise 算法中, 树的生长策

略有两种生成方法: 一是 Level-wise 基于层的生长方法, 对每层每个节点都进行分裂; 二是 Leaf-wise 方法, 对于每层, 只分裂增益较大的节点, 可减少计算量, 同时利用最大深度限制防止过拟合(见图 2)。

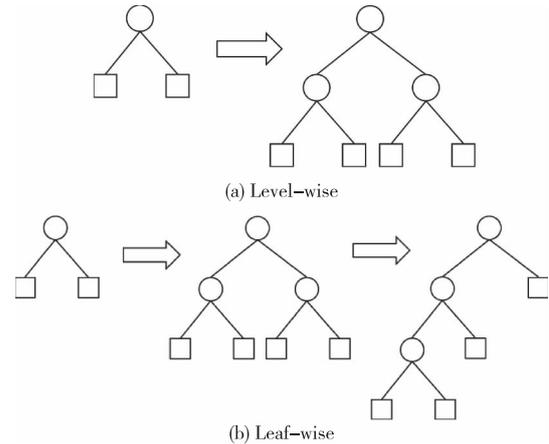


图 2 不同的树生长策略

基于直方图的加速算法将输入层中已排序的参数数据集转换为具有指定数量的数据区间或 Bins 的直方图(见图 3)。转换之后, 每个数据间隔(Bin)拥有相同的索引。该算法在大大提高 LightGBM 模型训练速度的同时只需较少的内存消耗。

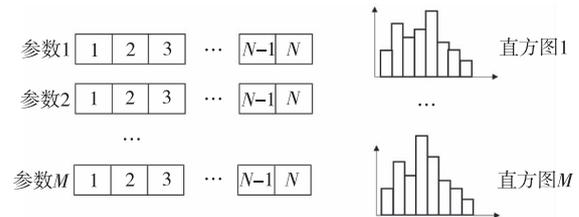


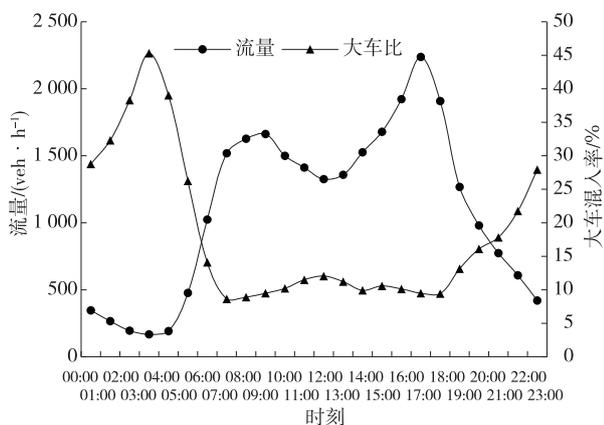
图 3 直方图加速算法示意图

2.2 LightGBM 模型训练

2.2.1 数据准备

研究路段为山东省济南市济广(济南—广州)高速公路唐王收费站至济南零点收费站区间, 共 22 km, 单向四车道, 自东向西途径临港、小许家、东客、华山立交。对济广高速公路 2021 年 5 月车流量及大车混入率按小时进行统计, 结果如下: 济广高速公路车辆出行时间主要集中在 6:00—19:00, 以二型及以下小型车为主, 小型车占比为 90% 左右; 夜间流量较低, 以大型车为主, 大型车占比最高达 45% (见图 4)。

该路段单向共设置 3 处 ETC 门架、2 处视频卡口。选取这 5 个点位 2021 年 5 月共 2 450 387 条 ETC 交易数据及 2 145 281 条视频卡口过车数据进



流量为一个月各时段平均值

图 4 流量与大车混入率关系图

行高速公路行程时间预测。

2.2.2 模型主要参数设置

LightGBM 模型有较多的超参数,不同参数值对模型性能有显著影响,需针对具体应用场景对模型超参数进行优化。采用文献[18]中方法对参数进行优化调整。为防止数据训练过拟合,采取 K 折交叉验证(KCV)的方法将特征数据集分成 m 份,其中 c 份用作训练集, $m - c$ 份数据用作验证集,对模型进行训练。模型的主要参数及相关字段见表 4。

表 4 模型重要参数表及相关字段

参数名称	类型	范围
learning_rate(学习率)	Double	>0
n_estimators(基学习器个数)	Int	>0
max_depth(树的最大深度)	Int	>0
num_leaves(最大叶子数)	Int	1~131 072
max_bin(直方图分块数)	Int	1~255
min_data_in_leaf(叶子节点最小记录数)	Int	≥ 0
feature_fraction(特征分数)	Double	0~1
bagging_fraction(每次迭代所用数据比例)	Double	0~1
bagging_freq(抽样次数)	Int	≥ 0
min_gain_to_split(最小分裂增益)	Double	≥ 0

基学习器的数量对模型预测精度有很大影响,数量过大或过小会导致模型训练结果过拟合或欠拟合。为解决这个问题,以均方根误差 RMSE 为指标,通过不断迭代学习寻找合适的基学习器数量。如图 5 所示,模型预测的 RMSE 随着基学习器的数量增大而减少,最后达到收敛状态。选取 500 个基学习器。

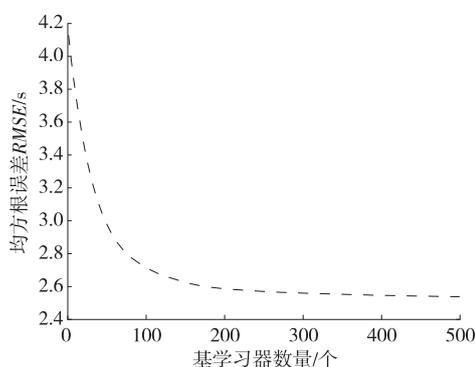


图 5 RMSE 随模型基学习器数量的变化

LightGBM 模型的 learning_rate 对算法的计算速度及预测精度有很大影响。在 LightGBM 模型中,learning_rate 的默认值为 0.1。在其他参数不变的情况下,将 learning_rate 分别设置为 0.01、0.05、0.1,分析不同 learning_rate 下模型误差与迭代次数的关系,结果见图 6。由图 6 可知:对于上文构建的预测模型及数据集,LightGBM 算法中 learning_rate 对模型预测精度几乎没有影响,但 learning_rate 越大,算法收敛越快。

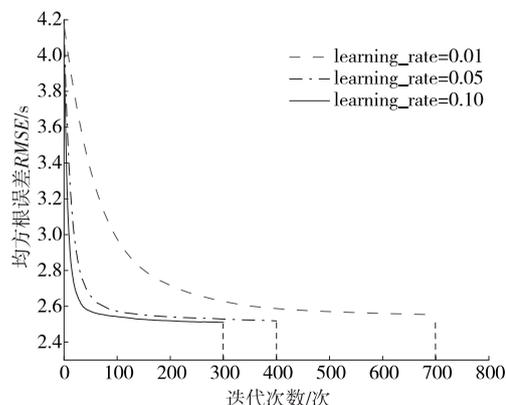


图 6 不同 learning_rate 下模型误差与迭代次数的关系

在 LightGBM 模型中,max_depth 和 num_leaves 控制决策树的复杂性,其取值过大会使模型陷入过拟合,导致预测精度达不到预期。为此,研究误差随不同 max_depth 和 num_leaves 参数值组合的变化,获取超参数的优化值,结果见图 7。由图 7 可知:虚线圈中颜色最深,max_depth 的最佳取值约为 4,对应于 num_leaves 参数的最佳取值约为 10,与利用文献[15]中基于贝叶斯的参数方法所得结果吻合;随着 max_depth 与 num_leaves 取值的增大,误差逐渐增大,表明模型可能出现过拟合。因此,两参数的取值不宜过大,该结论与 LightGBM 模型本身的性质也相符。

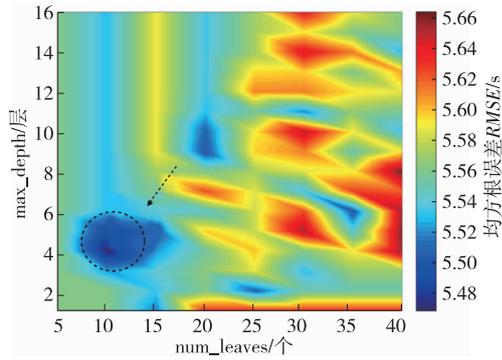


图7 误差随 num_leaves 与 max_depth 不同取值的变化热力图

其他超参数的取值见表5。

表5 超参数的优化取值

参数名	类型	取值
n_estimators	Int	500
max_depth	Int	4
num_leaves	Int	10
max_bin	Int	165
min_data_in_leaf	Int	16
feature_fraction	Double	0.88
bagging_fraction	Double	0.79
bagging_freq	Int	6
min_gain_to_split	Double	0.112

3 模型结果分析

3.1 模型验证指标

选取均方根误差 $RMSE$ 、平均绝对误差 MAE 、平均绝对百分比误差 $MAPE$ 作为预测模型精度验证指标。 $RMSE$ 反映预测结果与实际值之间的偏差,计算公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^{\text{real}} - y_i^{\text{predict}})^2}$$

式中: y_i^{real} 表示第 i 个实际值; y_i^{predict} 表示第 i 个预测值。

MAE 是预测值与实际值绝对误差的平均值,反映误差的真实情况和大小,计算公式如下:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i^{\text{real}} - y_i^{\text{predict}}|$$

$MAPE$ 是预测值与实际值绝对误差和实际值的比值的百分比,反映预测结果的可靠度,计算公式如下:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i^{\text{real}} - y_i^{\text{predict}}|}{y_i^{\text{real}}} \times 100$$

3.2 模型对比验证

为验证 LightGBM 模型在高速公路行程时间预测方面的优越性,选取随机森林(RF)、支持向量机回归(SVR)、最邻近(KNN)3种常用机器学习算法进行对比验证。根据5个ETC门架及视频卡口检测器分布将该高速公路分为4段分别进行预测。各路段不同模型预测值与实际值的对比见图8。

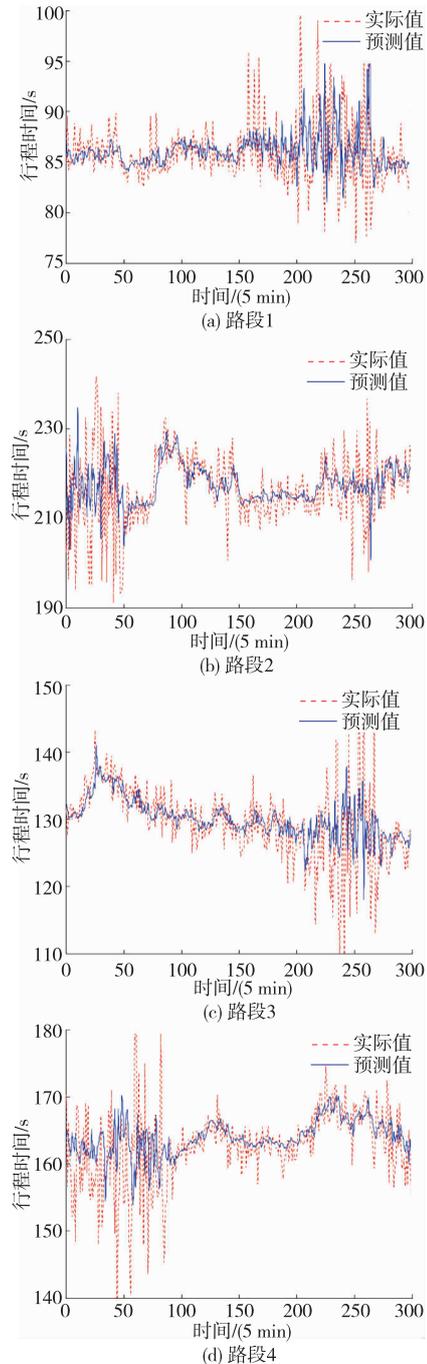


图8 不同模型预测值与实际值对比

从图8可看出:LightGBM 模型能较好地预测数据的趋势和走向。实际值中存在一些异常值(偏

离均值程度很大),可能是由高速公路上突发状况(如车祸、异常天气等)所致。LightGBM 模型对于这些异常情况的预测效果不是很好,但总体预测精度较高。

为进一步分析 LightGBM 模型的预测性能,将该模型的 $RMSE$ 、 MAE 、 $MAPE$ 及耗时与 KNN、RF、SVR 算法进行对比,结果见表 6。

表 6 各路段不同模型预测效果评价

路段	模型	$RMSE$	MAE	$MAPE$	耗时/s
路段 1	LightGBM	3.92	1.50	16.13	0.70
	KNN	4.25	1.57	16.70	0.40
	RF	4.37	1.55	16.42	43.21
	SVR	4.40	1.55	16.40	18.81
路段 2	LightGBM	8.67	2.39	16.21	0.72
	KNN	9.74	2.51	16.93	0.37
	RF	9.08	2.43	16.41	42.88
	SVR	9.10	2.44	16.40	13.93
路段 3	LightGBM	4.92	1.78	15.67	0.69
	KNN	5.44	1.87	16.32	0.40
	RF	5.13	1.79	15.64	41.86
	SVR	8.37	1.79	15.67	15.03
路段 4	LightGBM	5.60	1.90	14.77	0.30
	KNN	10.54	2.41	18.36	0.35
	RF	8.15	2.08	16.14	25.02
	SVR	7.40	2.13	16.35	12.08

从表 6 可看出:1) LightGBM 模型预测值的 $RMSE$ 、 MAE 低于 KNN、RF、SVR 模型,除路段 3 的 $MAPE$ 略高于 RF 模型外,其他路段均低于 KNN、RF、SVR 模型。对于路段 1 和路段 2, KNN、RF、SVR 模型的 3 个指标预测精度大致相同;对于路段 3, SVR 模型预测值的 $RMSE$ 明显大于其他模型;对于路段 4, RF、SVR 模型的预测精度类似, KNN 模型的预测精度低于其他模型。2) 从运算耗时来看, KNN 模型因其简易的计算原理,耗时最短,其次是 LightGBM 模型,两模型的耗时差距较小,都小于 1 s; SVR 模型的耗时为 10~20 s,排名第三;耗时最长的是 RF 模型,路段 1~3 的运算时间在 40 s 以上。总体来说, RF、SVR 模型的运算时间远高于 KNN、LightGBM 模型,不适用于实时预测; KNN 的耗时最短,但与 LightGBM 模型的差距较小,而後者的预测精度和稳定性优于前者。综合来看, LightGBM 模型满足高效、准确的实时预测要求,是性能最好的高速公路行程时间预测模型。

4 结语

针对高速公路行程时间预测中存在的数据来源单一、预测时效性不足等问题,本文提出一种基于路警融合数据的 LightGBM 模型,实现高速公路行程时间的高效预测。利用 ETC 门架、视频卡口多源融合关联数据,采用该模型进行预测,与其他传统机器学习算法相比, LightGBM 算法的综合性能最好,有助于节省计算与存储资源,可应用于高速公路路网级的行程时间实时预测。但该预测模型采用的训练和测试路段线形较平顺,且采用的是交通畅通条件下数据,后期需利用更多不同路段、不同交通状态下数据进行模型验证,以推广到整个高速公路网,提高路网整体监测与管理能力。

参考文献:

- [1] PIRC J, TURK G, ZORA M. Highway travel time estimation using multiple data sources[J]. IET Intelligent Transport Systems, 2016, 10(10): 649-657.
- [2] 邢雪, 于德新, 田秀娟, 等. 基于数据挖掘的高速公路行程时间预测[J]. 华中科技大学学报(自然科学版), 2016, 44(8): 36-40.
- [3] 王翔, 陈小鸿, 杨祥妹. 基于 K 最近邻算法的高速公路短时行程时间预测[J]. 中国公路学报, 2015, 28(1): 102-111.
- [4] 刘伟铭, 雷焕宇, 翟聪, 等. 基于 PSO-LSSVM 的高速公路短时行程时间预测[J]. 公路与汽运, 2017(3): 36-39+48.
- [5] 杭明升, 杨晓光, 彭国雄. 基于卡尔曼滤波的高速道路行程时间动态预测[J]. 同济大学学报(自然科学版), 2002, 30(9): 1068-1072.
- [6] 江周, 张存保, 许志达, 等. 基于多源数据的城市道路网络行程时间预测模型[J]. 交通信息与安全, 2014, 32(3): 27-31.
- [7] 李萌, 谷远利, 王硕, 等. 降雨条件下高速公路短时行程时间预测研究[J]. 交通信息与安全, 2018, 36(4): 90-96.
- [8] 刘松, 彭勇, 邵毅明, 等. 基于门控递归单元神经网络的高速公路行程时间预测[J]. 应用数学和力学, 2019, 40(11): 1289-1298.
- [9] 林培群, 夏雨, 周楚昊. 引入时空特征的高速公路行程时间预测方法[J]. 华南理工大学学报(自然科学版), 2021, 49(8): 1-11.
- [10] MACKAY D J C. Probable networks and plausible predictions: A review of practical Bayesian methods for supervised neural networks[J]. Network: Computation in Neural Systems, 1995, 6(3): 469-505.

(下转第 31 页)

和较小范围内受影响的规律相反。总体上看,在相关色温为 2 500~3 100 K 的较大区间内,白色三类反光膜的逆反射测量结果受相关色温偏离的影响较小;在 2 856 K \pm 50 K 的较小区间内,白色三类反光膜的逆反射测量结果受相关色温偏离的影响较大。

(3) 有色反光膜逆反射测量结果随相关色温变化的趋势反映了光源光谱分布变化趋势。相关色温升高时,波长较小的分量增加,波长较大的分量减少,故黄色和红色反光膜的逆反射测量结果降低,蓝色和绿色反光膜的逆反射测量结果升高。

(4) 相关色温变化对逆反射测量的影响总体较显著。相关色温偏离的半宽度为 50 K 时,各色反光膜的逆反射系数示值误差不大于 4.5%,不确定度分量最大为 1.732 cd/(lx \cdot m²)。因此,对测量结果精度要求较高时,应保证相关色温稳定在 2 856 K 处。

参考文献:

- [1] 王义旭,彭璐,苗娜,等.逆反射测量装置校准中角度控制方法研究[J].中国计量,2018(11):96-98.
- [2] 王露婉,韩晓坤,何华阳,等.道路交通标线动静态测量比对研究[J].公路与汽运,2021(2):43-46.
- [3] 何华阳,苏文英,王蕊.路面平整度对标线逆反射测试的静态影响[J].中国公路学报,2022,35(5):211-220.
- [4] 李奕,周秉直,赵飞虎,等.逆反射亮度系数校准装置研制及测量误差影响因素分析[J].徐州工程学院学报(自然科学版),2016,31(4):76-82.
- [5] 杨勇,张智勇,李旭,等.光源光谱对标志逆反射系数的影响[J].光谱学与光谱分析,2014,34(1):12-15.
- [6] 徐何辰,饶丰,李加,等.汽车前照灯光谱分布差异对回反射器光度色度的影响[J].照明工程学报,2016,27(2):21-24.
- [7] 罗语丹.相对测量法检测交通标志反光膜逆反射系数的可靠性分析[J].公路与汽运,2019(6):34-38.
- [8] 杨学政,滕方勇.标志反光膜逆反射性能评价方法与影响因素分析[J].山东交通科技,2020(2):106-109.
- [9] 韩飞.交通标志反光膜逆反射系数比对及能力验证结果分析[J].福建交通科技,2018(2):155-157.
- [10] 王劲惠.公路交通标志反光膜常态逆反射系数的不确定度评定[J].计量技术,2011(2):65-68.
- [11] 李伟,朱传征.逆反射测量系统计量检定技术[J].公路交通科技,2012,29(2):132-135.
- [12] 谷天明.公路标志标线的逆反射系数测量[J].交通标准化,2003(5):29-31.
- [11] VAN HINSBERGEN C, HEDGY A, VAN LINT J W C, et al. Bayesian neural networks for the prediction of stochastic travel times in urban networks[J]. IET Intelligent Transport Systems, 2011, 5(4): 259-265.
- [12] LIANG W, LUO S, ZHAO G, et al. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms[J]. Mathematics, 2020, 8(5): 765-781.
- [13] 王芳杰,王福建,王雨晨,等.基于 LightGBM 算法的公交行程时间预测[J].交通运输系统工程与信息, 2019, 19(2): 116-121.
- [14] 韩皓,徐圣安,赵蒙.考虑线网结构的 LightGBM 轨道交通短时客流预测模型[J].铁道运输与经济, 2021, 43(10): 109-117.
- [15] 兰良,蒙祖强.基于高速公路联网收费数据的行程时间预测研究[J].信息与电脑(理论版), 2021, 33(7): 5-9.
- [16] 杨成连,文涛,李洋洋.基于 ARMAX 模型的高速公路行程时间预测研究[J].公路交通科技(应用技术版), 2020, 16(4): 301-307.
- [17] 张娟,孙剑.基于 SVM 的城市快速路行程时间预测研究[J].交通运输系统工程与信息, 2011, 11(2): 1009-1014.
- [18] 黄雪峰,周剑峰.高速公路车辆行程时间预测研究[J].公路, 2009(2): 116-120.
- [19] 樊博,马筱栋.基于收费数据的高速公路旅行时间预测[J].公路与汽运, 2020(6): 24-28.
- [20] 李松江,宋军芬,杨华民,等.基于聚类分析的高速公路行程时间预测[J].计算机仿真, 2019, 36(2): 384-389.
- [21] 李苗苗.不良天气下高速公路可变限速控制研究[D].兰州:兰州交通大学, 2021.
- [22] 张腾月,翁小雄.基于收费数据的高速公路行程时间可靠性估计模型[J].广西师范大学学报(自然科学版), 2016, 34(4): 70-77.
- [23] GAN M, PAN S, CHEN Y, et al. Application of the machine learning LightGBM model to the prediction of the water levels of the lower Columbia River[J]. Journal of Marine Science and Engineering, 2021, 9(5): 496.

收稿日期:2022-03-07

收稿日期:2022-04-08