

# 基于收费数据的高速公路旅行时间预测\*

樊博<sup>1,2</sup>, 马筱标<sup>1,2</sup>

(1.重庆交通大学 交通运输学院, 重庆 400074; 2.重庆市交通运输工程重点实验室, 重庆 400074)

**摘要:** 针对高速公路检测器铺设密度较低现状,以高速公路联网收费数据为基础,开展收费站间旅行时间预测研究。首先对从收费数据中提取的旅行时间进行清洗修正和填补处理,以获取真实旅行时间数据集;然后构建旅行时间解释变量,建立支持向量回归模型的旅行时间预测算法;最后以重庆某通道真实数据为基础,分不同场景对模型进行验证,结果表明模型预测精度总体较好,其中一般情况场景和暑期场景下预测结果的平均相对误差分别约为 5.17%、8.30%。

**关键词:** 交通工程;旅行时间预测;高速公路;收费数据;支持向量回归

中图分类号:U492.8

文献标志码:A

文章编号:1671-2668(2020)06-0024-05

高速公路旅行时间是诱导公众出行的重要依据,是评价道路实时运行状态的重要指标,也是辅助道路管理人员进行管控资源配置的决策参考。旅行时间预测对于交通使用者和管理者都具有重要意义和应用价值,可以说,旅行时间预测是高速公路交通管理系统、出行者信息系统中不可或缺的部分。

由于国外检测器铺设密度相对较高,旅行时间预测研究可以在高精度交通流数据下开展,如 Moonam H. 等基于历史真实数据,使用卡尔曼滤波器,针对不同类型道路预测旅行时间。对于利用收费数据进行旅行时间预测的研究,王翔等根据交通事故持续时间和不同时段交通状态的特征,构建基于改进最近邻非参数回归的方法预测旅行时间,算法中各类别历史真实数据集中  $K$  值采用交叉验证的方法标定;赵建东等采用决策级融合策略,对收费数据和微波车检数据进行融合,通过建立模型分配权重,使用 BP 神经网络进行旅行时间预测,同时引入遗传算法对 BP 神经网络进行优化,模型预测结果的相对误差在 10% 以内。该文以高速公路中主线收费站路段为基础进行分析,构建影响旅行时间的解释变量,采用支持向量回归算法,提出一种基于收费数据的高速公路站间旅行时间预测模型。

## 1 数据准备

### 1.1 数据来源与数据结构

数据来自重庆高速公路网 2017—2019 年联网收费数据,删除其中异常数据(包括但不限于字段缺

失、字段取值无效和车辆进出收费站的时间不合逻辑等)后提取每条信息中的车辆 OD 信息。经过数据预处理后一条完整的收费数据所包含的主要字段及含义见表 1。

表 1 收费数据中包含的主要字段

字段	字段含义
VEHPLATE	车辆车牌号
VEHCLASS	车型所属等级
TRUCKFLAG	客、货车标签
ENTIME	车辆进站时间
EXTIME	车辆出站时间
ENSTATION	进站编号
EXSTATION	出站编号

### 1.2 研究对象

选取重庆 G65 包茂(包头—茂名)高速公路渝湘段为研究对象。该项目全长 84.5 km,由巴南主线收费站起,出城单向途经南彭、接龙、石龙、大观、南川、金佛山 8 个收费站后到水江收费站为止。

以 2018 年为例,路网上总流量分布受客车流量分布影响最大,货车流量则基本上维持在相对稳定状态(见图 1)。客车中的一型客车,其交通量占客车流量的绝大部分,具有一定的代表性。同时已有研究表明,高速公路交通系统中可用一型客车的旅行时间来推算其他车型的旅行时间。因此,以一型客车作为主要研究车型。

随机统计工作日、周末和节假日路网流量小时分布,结果见图 2。由图 2 可知:路网夜晚和凌晨时段流量相对稀少,该时段车辆以自由流速度行驶,无

\* 基金项目:重庆市科委基础前沿研究专项(cstc2018jcsx—mszdX0112)

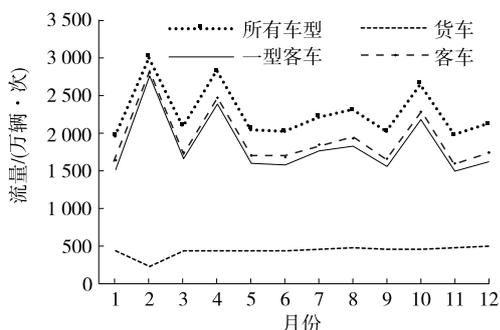


图 1 重庆高速公路网 2018 年全年流量月份分布

预测必要。因此,以 7:00—20:00 作为预测时段。

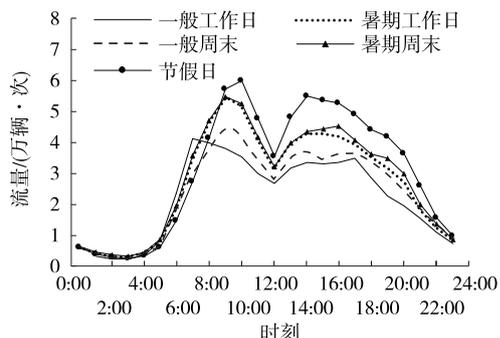


图 2 重庆高速公路网流量小时分布

## 2 旅行时间提取及处理

### 2.1 旅行时间提取

如图 3 所示,高速公路 A、B 收费站之间的路段称为  $k$  站间,站间旅行时间是指在某时间周期内由 A 入口驶入、B 出口驶出高速公路的所有车辆的平均行驶时间,也可称为旅行时间。

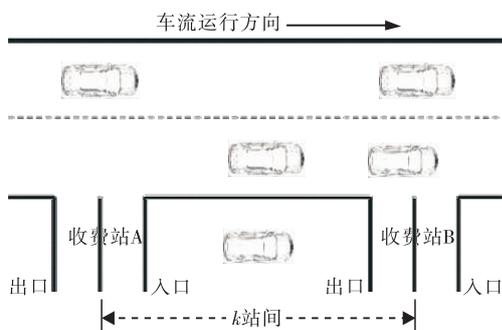


图 3 高速公路收费站间示意图

可根据收费数据中车辆经收费站进出高速公路的时间计算  $k$  站间在周期  $T$  内的旅行时间:

$$t_{k,T} = \frac{1}{N} \sum_{i=1}^N t_k^{(i)} \quad (1)$$

式中: $t_{k,T}$  为  $T$  周期内由 A 入口驶入、B 出口驶出的车辆平均旅行时间,即  $k$  站间在周期  $T$  内的旅行时

间;周期  $T$  为一个时间间隔,一般取 10~15 min,这里取 15 min; $N$  为  $T$  周期内 A 入口驶入、B 出口驶出的车辆总数; $i=1,2,3,\dots,N$ ; $t_k^{(i)}$  为从 A 入口驶入、B 出口驶出的第  $i$  辆车的旅行时间(s),通过式(2)计算得到。

$$t_k^{(i)} = t_{k,ex}^{(i)} - t_{k,en}^{(i)} \quad (2)$$

式中: $t_{k,ex}^{(i)}$ 、 $t_{k,en}^{(i)}$  分别为车辆  $i$  驶出、驶入高速公路的时间,即收费数据中 EXTIME、ENTIME 字段。

由于部分时间相邻收费站间 OD 记录较少,甚至某些相邻站间在某些时段内无 OD 记录,无法提取站间真实交通运行状态的旅行时间。另外,由于某些特殊驾驶行为(如车辆长时间以低于最低限速或超速行驶、车辆进入服务区休息等)产生的异常旅行时间对提取 OD 间准确的旅行时间有所影响,有必要对该类数据进行控制。

### 2.2 数据质量控制

对旅行时间数据的控制主要分为数据清洗和数据修补。数据清洗流程如下:1) 删除旅行时间大于一个自然日的数据;2) 删除平均速度超过道路最大限速 20%的数据;3) 计算各周期内车辆旅行时间的均值  $\mu$  和标准差  $\delta$ ,按照  $2\delta$  原理删除旅行时间不在  $[\mu - 2\delta, \mu + 2\delta]$  范围内的数据;4) 重复步骤 3 直至剔除所有不在有效范围内的数据。

完成数据清洗后,对缺失数据进行修补。利用式(1)计算提取真实旅行时间时, $T$  周期内  $k$  站间旅行时间为  $t_{k,T}$ ,若该周期内无有效  $t_k^{(i)}$ ,则无法计算  $t_{k,T}$ ,需对无旅行时间的周期进行数据修补。通过  $k$  站间前、后站间真实旅行时间  $t_{k-\beta}$  和  $t_{k+\beta}$  值对  $k$  站间旅行时间进行修补,其中  $\beta=1,2,3,\dots$ 。修补过程如下:若  $k$  站间在  $T$  周期下旅行时间  $t_{k,T}$  用  $t_k$  表示, $t_{k-\beta,k}$  为  $k-\beta$  站间收费站入口到  $k$  站间收费站出口的旅行时间, $t_{k,k+\beta}$  为  $k$  站间收费站入口到  $k+\beta$  站间收费站出口的旅行时间,则有式(3)~(5)。式(4)在随着  $\beta$  逐渐增大的计算过程中,一旦  $t_k$  取值不为零,式(3)~(5)停止计算,此时  $t_k$  取值即为  $k$  站间在  $T$  周期下的修补值。

$$t_k = \begin{cases} 0, t'_k \text{ 与 } t''_k \text{ 均为零} \\ 1/2(t'_k + t''_k), t'_k \text{ 与 } t''_k \text{ 均不为零} \\ t'_k + t''_k, t'_k \text{ 与 } t''_k \text{ 中仅有一项为零} \end{cases} \quad (3)$$

$$t'_k = (t_{k-\beta,k} - t_{k-\beta}) \frac{l_k}{l_{k-\beta,k}} \cdot \gamma_{k-\beta,k} \quad (4)$$

$$t''_k = (t_{k,k+\beta} - t_k) \frac{l_k}{l_{k,k+\beta}} \cdot \gamma_{k,k+\beta} \quad (5)$$

式中: $t'_k$ 与 $t''_k$ 分别为 $k-\beta$ 站间、 $k+\beta$ 站间的真实旅行时间,通过式(4)、式(5)计算; $l_k$ 为 $k$ 站间长度; $l_{k-\beta,k}$ 、 $l_{k,k+\beta}$ 分别为 $k-\beta$ 到 $k$ 站间的路段长度、 $k$ 到 $k+\beta$ 站间的路段长度; $\gamma_{k-\beta,k}$ 、 $\gamma_{k,k+\beta}$ 分别表示 $k-\beta$ 、 $k+\beta$ 各自到 $k$ 站间的数据是否充足,有效数据充足取1,否则取零。

### 3 解释变量构建

在构建影响旅行时间的解释变量时,将变量分为前期旅行时间,月份、周天、时段,大小车旅行时间差值,道路环境、天气条件。

#### 3.1 前期旅行时间

选取该预测站间 $k$ 的前期旅行时间为备选变量,为确定变量中前期旅行时间的周期数量,随机选取该项目某2个月、某2个不相邻站间 $p$ 和 $q$ 的前5个周期旅行时间进行相关性分析,相关系数计算结果见表2。

表2 当前周期旅行时间与前期旅行时间的相关系数

路段	前期旅行时间	相关系数
路段 $p$ (6.5 km)	$t_{p,T-1}$	0.910
	$t_{p,T-2}$	0.918
	$t_{p,T-3}$	0.907
	$t_{p,T-4}$	0.827
	$t_{p,T-5}$	0.894
路段 $q$ (25 km)	$t_{q,T-1}$	0.921
	$t_{q,T-2}$	0.936
	$t_{q,T-3}$	0.895
	$t_{q,T-4}$	0.778
	$t_{q,T-5}$	0.801

由表2可知:当前时间下的旅行时间与其前3个周期旅行时间的相关性较高。因此,选择前3个周期旅行时间作为备选变量。

#### 3.2 月份、周天、时段

以月份、周天和时段为变量对主线收费站交通量分布规律进行分析,确定各变量分类。

对于时段,根据图2路网流量分布可将预测时段分为两类:一类为出行高峰时段7:00—11:00和13:00—18:00;另一类为平峰时段11:00—13:00和18:00—20:00。

对于月份,统计2018年各月进入主线收费站的日均流量,结果见图4。按照各月日均流量分布可将全年12个月分为两类:第一类为1、3、4、5、10、11和12月;第二类为2、6、7、8和9月。

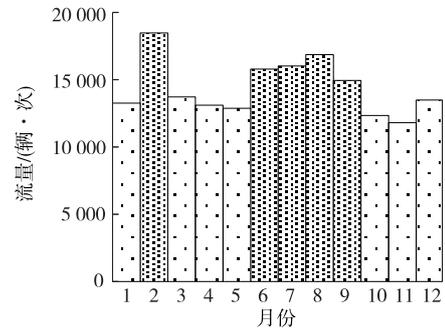


图4 通道主线收费站日均流量月份分布

再统计各周天进入主线收费站的日均流量,并进行系统聚类,结果见图5。根据图5,将周天变量分为三类:周一到周四为一类;周日为一类;周五、周六为一类。

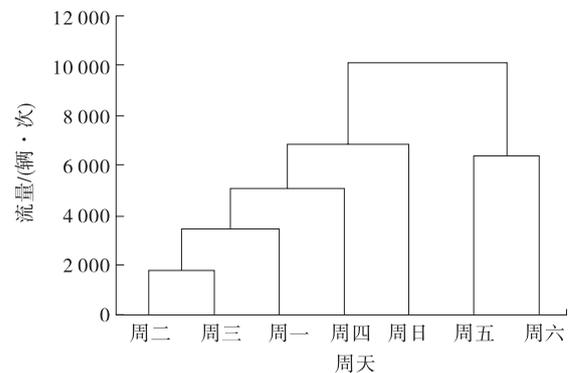


图5 周天聚类谱系图

#### 3.3 大小车旅行时间差值

相比城市交通系统,高速公路系统各周期内货车平均数量较多,交通流量中货运车辆所占比例对交通运行状态有一定影响,而交通运行状态又与车辆旅行时间密切相关。小车(一型客车)旅行时间与货车(所有货车)旅行时间的差值一定程度上能反映站间交通状态,在运行畅通的路段上,大、小车旅行时间基本维持在相对固定的差值;较为拥堵时,大、小车旅行时间差值会降低,意味着交通状态已经或正在发生某些改变,这些改变又会对下一时刻的旅行时间产生影响。因此,引入周期内大小车旅行时间差值作为变量之一。

#### 3.4 道路环境、天气条件

就道路环境而言,根据重庆山区高速公路多桥梁、隧道的特点,选取站间内桥梁、隧道长度作为备选变量。对于天气条件,按晴天、阴雨天和雾天雪天对其进行划分。

综上,影响高速公路旅行时间的解释变量及含义见表3。

表3 解释变量

变量名称	变量含义
$t_{k,T-1}, t_{k,T-2}, t_{k,T-3}$	站间 $k$ 前三周期的旅行时间
month	月份分类(分为2类)
day_of_week	周天分类(分为3类)
period_of_day	时段分类(分为2类)
dif_tt	大小车旅行时间差值
$l_s$	站间桥梁、隧道长度
weather	天气(分为3类)

## 4 基于支持向量回归的旅行时间预测模型

### 4.1 支持向量回归

使用支持向量回归(SVR)建立旅行时间预测模型。先建立样本集  $D = \{(x_u, y_u)\}_{u=1}^U = \{(x_1, y_1), (x_2, y_2), \dots, (x_u, y_u)\}$ , 其中  $x_u \in R^d$  为模型输入, 模型输出为  $y_u \in R$ 。引入拉格朗日乘数将 SVR 模型转换为:

$$\max \sum_{u=1}^U y_u \alpha_u - \epsilon \sum_{u=1}^U |\alpha_u| - \frac{1}{2} \sum_{u=1}^U \sum_{v=1}^U \alpha_u \alpha_v \kappa(x_u, x_v) \quad (6)$$

$$\text{s.t.} \sum_{u=1}^U \alpha_u = 0; -C \leq \alpha_u \leq C; u = 1, 2, \dots, U$$

式中:  $\alpha_u = \beta_u - \beta_u^*$ ,  $\beta_u$  和  $\beta_u^*$  为拉格朗日乘子,  $\alpha_v$  同理;  $\epsilon$  和  $C$  分别为损失值、惩罚因子, 分别取 0.1、1;  $\kappa(x_u, x_v)$  为 SVR 模型核函数, 选择高斯径向基为核函数, 其表达式见式(7)。

$$\kappa(x_u, x_v) = \exp(-\delta \|x_u - x_v\|^2) \quad (7)$$

式中:  $\delta$  为超参数, 设置为 1。

对式(6)进行求解, 得:

$$f(x) = \sum_{u=1}^U \alpha_u \kappa(x_u, x_v) + b \quad (8)$$

式中:  $b$  为偏置误差项, 可借助 KKT 条件求解。

### 4.2 实验验证

#### 4.2.1 性能评价指标

选取均方根误差( $R_{MSE}$ )、平均相对误差( $M_{APE}$ )作为模型的评价指标, 计算公式如下:

$$R_{MSE} = \sqrt{\frac{1}{n} \sum_{T=1}^n (t_T - \hat{t}_T)^2} \quad (9)$$

$$M_{APE} = \frac{1}{n} \sum_{T=1}^n \left| \frac{t_T - \hat{t}_T}{t_T} \right| \cdot 100\% \quad (10)$$

式中:  $n$  为周期数;  $t_T$  为周期  $T$  的旅行时间真实值;  $\hat{t}_T$  为周期  $T$  的旅行时间预测值。

#### 4.2.2 预测结果分析

以某站间某天的旅行时间为例, 通过修正, 旅行时间相比未经处理的原数据波动范围有所减小, 缺少有效数据的周期也得到修补(见图6)。

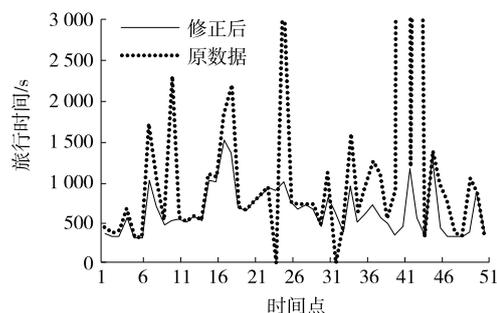
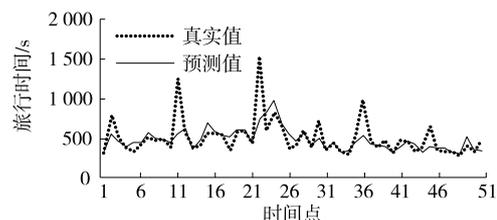
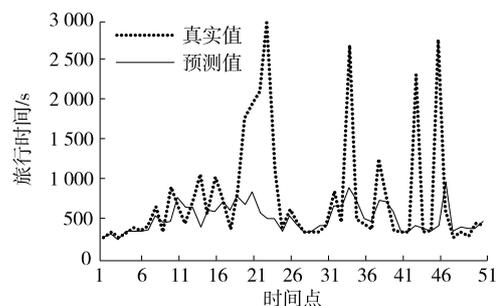


图6 修正旅行时间

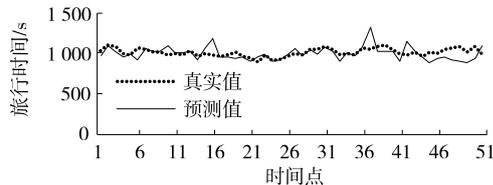
将真实旅行时间按 7 : 3 划分训练集和测试集, 选取  $m$  站间(路段长度为 6 km, 其中桥梁、隧道长度为零)和  $n$  站间(路段长度为 16 km, 其中桥梁、隧道长度为 4.85 km)。图 7 为  $m$  站间在一般情况场景下某周一和某事故场景下的模型预测结果, 图 8 为  $n$  站间在暑期场景下某周日和端午节假日下的模型预测结果。

(a)  $m$ 站间一般情况场景下(周一)(b)  $m$ 站间事故场景下(周五)图7  $m$ 站间分场景预测结果

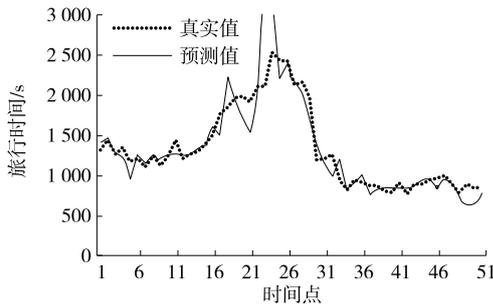
由图7可知:模型在一般情况场景下表现较好;由于构建的解释变量中缺少能准确描述偶然事件下旅行时间的影响因素,模型对偶然事件场景下的预测能力表现不足,但在非事故影响周期内,模型仍有一定的预测能力。

由图8可知:模型在暑期场景下的预测精度总体较高;在节假日场景下也具有一定的预测能力。

综上,模型总体表现能力较优,尤其是在一般场景和暑期场景下,平均相对误差分别为5.17%、



(a)  $n$ 站间暑期场景下(周日)



(b)  $n$ 站间节假日场景下(端午节)

图8  $n$ 站间分场景预测结果

8.30%;相较于一般场景和暑期场景,节假日场景下短时间内或某些周期内的车流量更多,模型预测误差较高。不同场景下模型预测结果的误差见表4。

表4 模型预测结果评价

场景	$R_{MSE}$	$M_{APE}/\%$
一般情况	23.42	5.17
暑期	35.66	8.30
节假日	58.01	14.48

## 5 结论

(1) 构造 SVR 模型预测高速公路旅行时间可行、有效,模型在一般情况场景和暑期场景下的预测精度较高,对节假日也有一定的预测能力。

(2) 要实现模型对偶然事件下的准确预测,仅考虑大小车旅行时间差值还不够,需引进更能描述偶然事件的参数变量。

(3) 下一步可考虑道路线形等相关变量,进一步提高模型的可靠性和适用性;同时使用特征选择方法,去掉解释变量中可能存在的冗余变量,进一步提高模型的计算能力。

## 参考文献:

[1] MOONAM H M, QIN X, ZHANG J. Utilizing data mining techniques to predict expected freeway travel time from experienced travel time [J]. Mathematics and Computers in Simulation, 2019, 155: 154-167.

[2] 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.

[3] 赵建东, 徐菲菲, 张琨, 等. 融合多源数据预测高速公路站间旅行时间[J]. 交通运输系统工程与信息, 2016, 16(1): 52-57.

[4] LU C, DONG J. Estimating freeway travel time and its reliability using radar sensor data[J]. Transportmetrica B (Transport Dynamics), 2018, 6(2): 97-114.

[5] LIU X B, CHIEN S, KIM K. Evaluation of floating car technologies for travel time estimation[J]. Journal of Modern Transportation, 2013, 20(1): 49-56.

[6] 景立竹, 李群善, 许金良, 等. 基于 V/C 比和载重汽车混入率的高速公路基本路段车辆平均行程时间预测模型[J]. 长安大学学报(自然科学版), 2018, 38(5): 106-113.

[7] CACERES H, HWANG H, HE Q. Estimating freeway route travel time distributions with consideration to time-of-day, inclement weather, and traffic incidents [J]. Journal of Advanced Transportation, 2016, 50: 967-987.

[8] JUNG S Y, QIN X, NOYCE D A. Rainfall effect on single-vehicle crash severities using polychotomous response models [J]. Accident Analysis & Prevention, 2010, 42(1): 213-224.

[9] 李长城, 文涛, 刘小明, 等. 基于高速公路收费数据的行程时间可靠性模型研究[J]. 公路交通科技, 2014, 31(12): 110-115.

[10] 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.

[11] 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.

[12] 尹志鹏, 刘伟铭, 庄岩浩. 基于高速公路收费数据的旅行时间可靠性分析[J]. 公路与汽运, 2014(5): 52-57.

[13] 李松松. 基于收费数据挖掘的高速公路旅行时间预测和交通状态判别应用研究[D]. 广州: 华南理工大学, 2017.

[14] 赵建东, 王浩, 刘文辉, 等. 基于收费数据的高速公路站间旅行时间预测[J]. 同济大学学报(自然科学版), 2013, 41(12): 1849-1854.