

基于决策树理论的交通流参数短时预测

薛红军[▲] 陈广交 李鑫民 顾 理

(南通市公路管理处 江苏 南通 226001)

摘 要 针对现有交通流参数短时预测方法的不足,考虑到交通流数据序列的非线性特征,提出一种基于决策树理论的非参数预测方法。采用时间序列滞后项将交通流参数序列转化成非参数模型能处理的数据格式。考虑到交通流参数之间存在长期协整关系,构建流量速度滞后项的组合向量,为预测模型提供基础数据。构建基于分类回归树(CART)的交通流参数短时预测模型。基于实际采集的道路交通流数据,对模型在不同等级道路不同速度区间下的性能进行评估。结果表明,所提出的模型相较于常用的时间序列模型,精度有所提高;速度预测准确性普遍高于流量,速度平均绝对百分比误差基本小于 13%,而流量预测则达到了 30%;采用工作日和周末数据分别建模能够有效提升预测性能;不同速度区间下的预测性能评估显示,模型在各等级道路中速区间的预测结果具有较高的准确性与稳定性。

关键词 城市交通;交通流参数;短时预测;决策树;非参数模型

中图分类号:U491 文献标志码:A doi:10.3963/j.issn 1674-4861.2016.03.009

A Short-term Forecasting of Traffic Flow Parameters Based on Decision Tree Theory

XUE Hongjun[▲] CHEN Guangjiao LI Xinmin GU Li

(Nantong Road Management Department, Nantong 226001, Jiangsu, China)

Abstract: Due to the non-linearity feature of data series of traffic flow, a nonparametric predicting method based on the theory of decision tree is proposed to overcome the deficiencies of current short-term forecasting methods of traffic flow. Using time series lags, the data series of traffic flow are converted into a format which can be recognized by a non-parameter model. Considering the long-term co-integration relationship between traffic flow parameters, combinational vectors of volume and speed lags are established, providing a basis for this forecasting model. Then a decision tree model based on classification and regression tree (CART) is developed for the prediction for the parameters of short-term traffic flow. Based on actual data of traffic flow, the performance of this CART decision tree model is evaluated under different types of road and speed intervals. The results show that the CART decision tree model is superior to the general time series model in terms of the prediction accuracy. In addition, the accuracy of speed prediction is generally higher than that of volume prediction. The mean average percentage error (MAPE) of speed prediction is less than 13% while the MAPE of volume prediction is 30%. Besides, this model can perform better when it is constructed with the data from weekdays or weekends separately. The evaluation of performance under different speed intervals indicates that the CART decision tree model presents high accuracy and stability in medium intervals speed for all types of road.

Key words: urban traffic; traffic flow parameters; short-term forecasting; decision tree; non-parametric model

0 引 言

交通流参数预测是指对具有非周期性、非线性和随机性的交通数据序列根据当前和历史数据

的特征,对未来态势做出合理的推测和判断,对于实现道路交通状态的提前预知、交通管控和诱导方案的制定具有重要意义^[1]。一般而言,交通状态参数预测以 5 min 或 15 min 的短时预测为代

收稿日期:2016-01-05 修回日期:2016-06-11

▲第一作者(通信作者)简介:薛红军(1972—),本科,研究方向:公路交通管理、智慧交通,E-mail:ljw1202@sina.cn

表。至今为止,国内外学者在交通状态参数短时预测方面展开了大量的研究,并提出了不少预测模型和方法,主要分为经验法和数据驱动法两类。早期研究中,由于缺乏有效的数据收集技术和手段,道路交通状态参数的数量和质量都不能很好地得到保证,严重制约了交通流预测技术的发展,因此当时采用的主要技术是以随机漫步法和历史平均法为代表的经验法。随着交通检测技术的不断进步,研究人员开始从获取的大量实测交通数据中探索和挖掘有效的知识模式,并在此基础上提出了一系列基于数据驱动思想的交通流短时预测模型。Xia^[2]采用考虑了季节因素的差分自回归移动模型(ARIMA)实现短时交通流参数预测;Guo^[3]构建了 Kalman 滤波模型预测拥挤状况下的交通流速度;Krantz^[4]以历史交通流量、速度、占有率、行程时间和日期为输入构建神经网络模型预测 5 min 至 15 min 后的交通流量;杨兆升等^[5]认为路段交通流量与之前时段的流量存在关系,并将其作为输入构建支持向量机模型预测下一时刻的流量。在实际应用中,有很大一部分研究人员倾向于以 ARIMA 为代表的时间序列模型。其原因在于,该类模型除了能够满足预测的实时性要求之外,在建模过程中能够嵌入时间序列数据的前后依赖关系,同时考虑量化趋势性、周期性、季节性等影响因子提高了模型预测的精度和可解释性^[6-8]。

上述交通流参数短时预测模型依然存在一些缺点:①经验法主要以交通流的历史趋势为依据进行参数预测,但难以捕捉交通流的实时短期波动;②时间序列模型通常假定数据服从一定的分布(如正态分布、t 分布等),这种假设在现实情形中往往不成立,导致模型在某些情形下的预测结果不是很可靠。此外,部分时间序列模型,如 ARIMA,在本质上是一类线性模型,而交通流时间序列数据往往会呈现显著的非线性特征,这进一步限制了此类模型在实际中的应用。③非参数模型尽管不要求数据满足特定的分布假设,但大部分模型,如神经网络和支持向量机方法,训练时间过长导致模型的实时性无法得到保证,且易出现“过拟合”问题^[9-10]。

针对传统交通流参数短时预测方法存在的问题,考虑到决策树模型能够很好地捕捉交通流数据中的动态变化模式和非线性特征,笔者提出一种基于决策树理论的非参数交通流短时预测方

法,基于向量预测思想,在建模过程中考虑交通流参数之间的关系,采用流量和速度 2 个参数进行组合预测,建立了交通流参数短时预测的多变量向量预测模型。

1 基于决策树理论的交通流参数预测

1.1 分类回归决策树

决策树^[11-14]是一种用于分类和预测的数据挖掘技术,其基于信息熵理论和树状分类规则构建样本属性变量与目标变量之间的映射关系。在决策树理论的众多模型中,以 Quinlan 提出的 C4.5^[15]算法和 Breiman 等人^[16]提出的分类回归树(classification and regression tree, CART)算法最为典型。考虑到 C4.5 算法在构造树的过程中,需要对数据集进行多次的顺序扫描和排序操作,实际应用效率较低,不易满足交通流参数预测实时性要求;CART 算法具有不需要任何先验假设,对噪声的干扰和数据缺失具有较强的鲁棒性,等优点,基于此,笔者选用 CART 算法构建决策树模型。

CART 算法在训练过程中采用二分递归分割机制,将当前训练样本集按照一定的分割标准划分为 2 个样本子集,使得生成的每个非叶子节点都含有 2 个不同的分支,叶子节点则对应样本的分类类别或预测变量值。表 1 给出了 CART 算法的输入输出及构建过程。

1.2 模型构建

交通流参数信息(流量、速度、占有率等数据)以时间序列形式呈现,道路交通状态波动性特征使得相邻时间间隔的交通流数据之间相互依赖,即交通流波动变化过程是一种动态随机过程。因此,从 ITS 检测数据特征角度来看,交通流数据具有典型的时间序列特性。CART 算法属于非参数模型,不能直接用于处理时间序列数据。因此,本文的模型构建过程从以下 3 个方面出发将传统决策树模型应用于交通流参数预测:①首先,提取交通流参数时间序列滞后项作为参数变量;②基于向量预测思想,采用流量和速度参数构建组合预测模型;最后,③确定样本集,生成决策树并进行参数预测。

模型构建前,首先,需要将时间序列数据转换为决策树模型能够处理的不含时间顺序的参数变量。由于时间序列滞后项包含了丰富的历史信息,能够很好地反映时间序列数据的短期波动性

和趋势性,因此,本研究提取滞后项作为模型输入的参数变量。在此基础上,基于向量预测思想,以流量和速度参数作为建模变量,考虑到两者间存

在长期协整关系,将 2 个参数组合成向量,采用多参数组合预测的方式提高模型精度。时间序列滞后项提取及参数组合过程表述如下:

表 1 CART 算法描述
Tab.1 Description of CART

算法: CART(classification and regression tree)

输入: 训练样本集 $T = \{S_i\}, 1 \leq i \leq N$, 样本 S_i 对应的属性集合为 $F = \{F_j\}, 1 \leq j \leq M$

输出: 分类回归决策树模型

模型构建过程:

- (1) 根据训练样本集创建根节点;
- (2) 如果根节点对应的样本集 T 都属于同一个类别或仅剩一个样本,则返回该节点为叶子节点,并为其分配相应类别或计算其预测输出值;
- (3) 对于样本的每一个属性 F_j 执行该属性上的一个划分,并计算该划分所对应的分割指标值(针对分类分割标准采用 Gini 指数^[18],针对预测分割标准采用最小二乘偏差);
- (4) 根据分割指标值,选取最优分割属性,并据此将样本集 T 划分为 T_1 和 T_2 2 个样本子集;
- (5) 对于 T_1 和 T_2 重复进行步骤(1)到步骤(4)的处理,直至生成的决策树不再有新的分支产生;
- (6) 基于后剪枝策略^[17]对决策树进行优化,并输出最终分类回归树模型。

其中, N 为训练样本总数, M 为分类属性个数。

令 $Z_t = [q_t, v_t]$, 其中, q_t, v_t 分别表示 t 时刻的交通流量和速度。预测过程包括寻找输入变量 X 及其与预测变量 Y 的映射关系 $f(\cdot)$, 满足 $Y = f(X)$ 。令 $X_t = \{Z_{t-i} \mid i = 0, 1, \dots, p\}$, X_t 表示 t 时刻的输入变量,由流量和速度的滞后项组合而成, p 表示时间序列滞后阶数,若 p 值过小则可能损失部分有效信息,影响预测精度,若过大又会带来不必要的资源浪费(多次试验数据表明对于 5 min 时间间隔的交通流参数预测取 6 较为合理)。同时,模型中预测变量 Y_t 对应为 $t+1$ 时刻的交通流量和速度的估计值 $\hat{Z}_{t+1} = [\hat{q}_{t+1}, \hat{v}_{t+1}]$ 。对于包含有 N 组训练样本的数据集 Ω ,则可以表示为 $\Omega = \{(X_n, Z_{n+1}) \mid n = 1, 2, 3, \dots, N\}$ 。

预测模型构建分为以下 3 个步骤:

步骤 1. 构建训练样本集。确定训练样本集包含的样本组数 N , 从历史数据库中分别提取 $N+1$ 组连续 5 min 时间间隔的交通流量和速度数据,由此构建包含有 N 组训练样本的数据集 Ω 。

步骤 2. 生成决策树模型:通过第一步构建的训练样本集来生成决策树,过程主要包括将样本递归划分进行建树过程和用验证数据进行剪枝过程。对于递归划分过程,需要确定划分的标准,即确定属性变量的临界值,将样本分为 2 类。每个属性的划分按照能减少的杂质的量来进行排序,而杂质的减少量定义为划分前的杂质减去划分后的每个节点的杂质量划分所占比率之和。文中选用 Gini 指标作为杂质度量方法,假设一个

样本共 C 类,那么节点 A 的 Gini 不纯度可表示为

$$\text{Gini}(A) = 1 - \sum_{i=1}^C p_i^2 \quad (1)$$

式中: p_i 表示属于 i 类的概率。

当分类回归树划分得太细时,会对噪声数据产生过拟合作用,因此我们要通过剪枝来解决。剪枝又分为前剪枝和后剪枝:前剪枝是指在构造树的过程中就知道哪些节点可以剪掉,不对这些节点进行分裂;后剪枝是指构造出完整的决策树之后再再来考查哪些子树可以剪掉。文中采用后剪枝中的代价复杂性剪枝策略,计算分类回归树中的每一个非叶子节点的表面误差率增益值 α ,找到 α 值最小的非叶子节点,进行剪枝。

$$\alpha = \frac{R(a) - R(A_a)}{|N_{A_a}| - 1} \quad (2)$$

$$R(a) = r(a) \times p(a) \quad (3)$$

式中: $|N_{A_a}|$ 为子树中包含的叶子节点个数; $R(a)$ 为如果节点 a 被剪枝的误差代价; $r(a)$ 为节点 a 的误差率; $p(a)$ 是节点 a 上的数据占有所有数据的比例; $R(A_a)$ 为若该节点不被剪枝,子树 A_a 的误差代价,它等于子树 A_a 上所有叶子节点的误差代价之和。

步骤 3. 交通流参数实时预测:对于任意时刻 t ,构建输入变量 $X_t = \{Z_{t-i} \mid i = 0, 1, \dots, p\}$,将其作为预测模型的输入,即可得到 $t+1$ 时刻的交通流参数预测值 $\hat{Z}_{t+1} = [\hat{q}_{t+1}, \hat{v}_{t+1}]$ 。

预测模型框架见下图

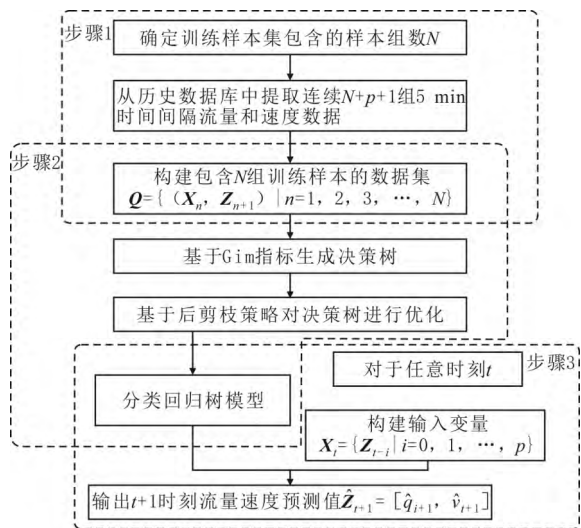


图1 预测模型框架

Fig.1 Framework of proposed method

2 实例验证及性能评估

2.1 数据描述

本文实验数据来源于实际采集的南通市干线

公路和昆山市主次干道道路断面交通流数据,包括 5 min 时间间隔的交通流量和速度数据。为了更加有效地验证和评估所提出的非参数决策树预测模型性能,本文还构建了精度较高的 ARIMA 模型以作对比。针对 2 类模型对于建模数据需求量的不同,本文采用不同的数据集用作模型标定,以最大限度发挥各模型性能。同时,为了确保预测结果具有可比性,2 类模型性能评估的数据集则一致,数据基本信息如表 2 所示。

2.2 评估指标

本文采用统计学中常用的 3 种误差估计指标来评估模型的预测性能,分别为平均绝对误差(MAE)、均方根误差(RMSE)以及平均绝对百分比误差(MAPE)。其中,MAE 和 RMSE 这 2 个指标分别从残差和方差的绝对数值角度来评估预测值与实际观测值之间的误差,MAPE 则是一种反映预测误差相较于实际观测值的相对指标,也是数理统计中方程估计和预测评价中应用最直接和最广泛的指标。3 种误差估计指标的计算公式

表 2 道路断面交通流数据基本信息描述

Tab.2 Traffic flow data at selected traffic-counting stations

道路类型	道路名称	断面编号	建模数据日期	性能评估数据日期
干线公路	G204	101100203	20150809(0828)~20150829	20150830~20150831
	S335	109102401	20150809(0828)~20150829	20150830~20150831
主干道	萧林路	102111	20150622(0711)~20150712	20150713~20150714
	前进路	108221	20150622(0711)~20150712	20150713~20150714
次干道	同丰路	106101	20150622(0711)~20150712	20150713~20150714
	震川路	110217	20150622(0711)~20150712	20150713~20150714

注:20150809(0828)~20150829 表示决策树理论建模数据日期为 20150809~20150829,ARIMA 为 20150828~20150829,下同。

如式(4)~(6)所示。

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - \hat{X}_i| \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{n}} \quad (5)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right| \quad (6)$$

式中: n 为样本个数; X_i 为实际观测值; \hat{X}_i 为预测值。

2.3 总体性能评估

表 3 给出了基于决策树理论的预测模型和 ARIMA 模型在各等级道路共 6 个断面总体预测性能对比。另外,考虑到工作日和周末交通流特性的差异性,本文进一步对工作日和周末交通流分开预测处理,评估结果见表 4 和 5。

从表 3 中可以看出:

1)提出的非参数决策树模型相较于时间序列模型 ARIMA,在流量预测方面具有显著的性能优势。表现为在所有检测断面上,决策树模型 3 个评估指标的值都明显小于 ARIMA 模型。

2)在速度预测方面,决策树模型在干线公路和城市主干道上仍然具有较为明显的性能优势,对于主干道交通流速度预测的 MAPE 值达到了 4%左右,低于 ARIMA 模型。在城市次干道方面,决策树模型在 106101 断面预测 MAPE 值为 12.09%,高于 ARIMA 模型的 11.87%;而决策树模型在 110217 断面预测 MAPE 值为 13.12%,要稍低于 ARIMA 模型的 13.26%。因此,决策树模型在次干道的 2 个断面上预测结果没有明显的优势,预测精度与 ARIMA 模型不相上下,甚至略低于 ARIMA 模型。

3)综合来看,2 类模型在不同种类的道路上

流量的预测性能波动范围要明显大于速度的预测性能波动范围,流量预测准确率的浮动范围从70%到87%,而速度预测的准确率则都接近或高于90%,这说明:①不同种类道路上的交通流所受外界干扰程度存在较大的差异性,就流量而言,

干线公路交通流受外界干扰最大,交通参数表现出明显的波动性,导致预测准确性降低;②相较于速度,交通流所受外界干扰对流量预测的准确性的影响程度要更大。

从表4和表5中可以看出:

表3 总体预测性能比较

Tab.3 Comparison of overall performance

断面编号	预测模型	流量			速度		
		MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
		$/(pcu \cdot 5 min^{-1})$	$/(pcu \cdot 5 min^{-1})$		$/(km \cdot h^{-1})$	$/(km \cdot h^{-1})$	
101100203	决策树	7.24	9.30	29.76	4.38	6.10	6.43
	ARIMA	7.43	9.70	30.09	4.41	6.29	6.48
109102401	决策树	6.35	8.18	30.62	3.96	5.40	6.86
	ARIMA	7.98	10.09	33.90	4.45	5.94	7.68
102111	决策树	4.12	5.95	15.27	1.22	2.36	3.41
	ARIMA	4.37	6.20	18.42	1.64	2.54	3.59
108221	决策树	3.27	4.42	12.91	1.42	1.96	4.54
	ARIMA	3.67	4.83	13.69	1.82	2.31	5.27
106101	决策树	5.05	6.56	20.34	3.31	4.35	12.09
	ARIMA	5.17	6.71	21.80	3.24	4.27	11.87
110217	决策树	8.44	12.02	24.68	4.03	5.15	13.12
	ARIMA	8.55	12.13	25.79	3.99	5.04	13.26

表4 工作日模式下模型预测性能比较

Tab.4 Performance comparison (in weekdays)

断面编号	预测模型	流量			速度		
		MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
		$/(pcu \cdot 5 min^{-1})$	$/(pcu \cdot 5 min^{-1})$		$/(km \cdot h^{-1})$	$/(km \cdot h^{-1})$	
101100203	决策树	7.71	10.00	25.33	4.22	5.51	6.29
	ARIMA	7.44	9.78	26.07	4.30	5.67	6.46
109102401	决策树	7.37	9.48	29.82	3.82	5.27	6.92
	ARIMA	8.86	11.36	31.10	4.43	5.89	7.82
102111	决策树	3.91	5.43	14.82	1.21	2.34	3.52
	ARIMA	4.85	6.76	17.80	1.66	2.60	3.61
108221	决策树	3.59	4.43	11.81	1.15	2.14	5.18
	ARIMA	4.05	5.28	13.26	1.82	2.29	5.43
106101	决策树	5.49	7.21	17.64	3.08	4.10	11.61
	ARIMA	5.60	7.30	18.83	3.15	4.13	11.87
110217	决策树	9.36	13.19	22.84	3.93	5.04	13.29
	ARIMA	9.89	13.79	25.39	4.01	5.11	13.70

表5 周末模式下模型预测性能比较

Tab.5 Performance comparison (in weekdays)

断面编号	预测模型	流量			速度		
		MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
		$/(pcu \cdot 5 min^{-1})$	$/(pcu \cdot 5 min^{-1})$		$/(km \cdot h^{-1})$	$/(km \cdot h^{-1})$	
101100203	决策树	8.04	10.47	24.47	4.33	5.49	6.52
	ARIMA	8.32	10.76	26.15	3.87	5.07	5.91
109102401	决策树	7.58	9.51	19.67	3.30	4.26	5.61
	ARIMA	7.65	9.43	19.45	3.23	4.13	5.44
102111	决策树	3.85	5.61	13.41	3.21	5.16	7.46
	ARIMA	4.15	5.82	14.81	3.65	5.49	8.90
108221	决策树	3.37	4.25	10.23	1.25	1.93	4.76
	ARIMA	3.80	4.75	11.19	1.73	2.28	5.25
106101	决策树	5.38	6.62	15.91	3.18	4.31	13.61
	ARIMA	5.18	6.40	15.82	3.32	4.45	14.32
110217	决策树	8.12	10.74	17.37	3.85	4.90	13.84
	ARIMA	7.71	10.45	17.51	3.97	4.99	14.26

在划分了时间模式之后,2类模型的流量预测准确性均得到了不同程度的提升,而速度预测的性能提升却不是很明显,与不划分时间模式的总体预测性能相当。这说明速度预测的稳定性要高于流量预测的稳定性,同时又一次说明交通流所受外界干扰对流量预测准确性的影响程度更大。另一方面,周末流量预测的精度普遍高于工作日,原因可能是工作日存在上下班高峰期,拥堵状态下的交通流波动较大,预测难度增加。而工作日和周末速度预测的精确度无明显差异,这也是外界干扰对速度预测准确性影响较小导致的。另外,在相同的时间模型下,本文提出的决策树模型相较于 ARIMA 模型在流量和速度预测方面依然具有显著优势。因此,建议在实际交通流短时

预测中,应针对工作日和周末分别构建决策树模型以提高预测的准确度。

为了更加直观地观察本文提出的决策树模型预测性能,选取了主干道 102111 号断面,做出该断面模型预测值与实际值拟合效果图,如图 2 所示。由图 2 中可见,在大多数时段内流量预测值与观测值较为吻合,且 2 条曲线具有相同的变化趋势,说明模型拟合程度较高,预测值能够较好的反映实际交通流量变化,个别时段由于检测误差或噪声点的存在,导致预测误差略大于正常时段。此外,与流量预测结果相比,模型在速度数据集上具有较高的预测准确性,究其原因在于速度的波动性要显著小于流量,一定程度降低的预测的难点。

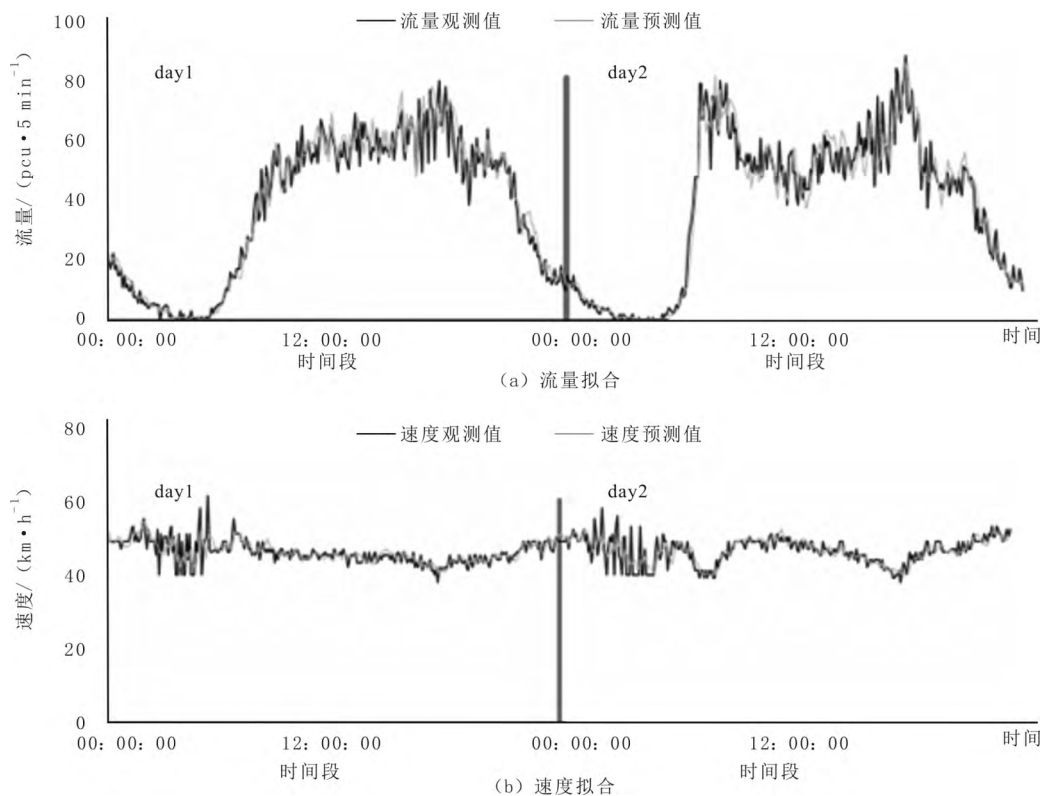


图 2 102111 号断面决策树模型流量速度预测值与观测值拟合效果
Fig.2 Plots of observed and forecasted traffic volumes and speeds for station 102111

2.4 不同交通水平下的模型性能评估

为了量化评估所提出模型在不同交通水平下的预测性能,本文对测试数据集上的速度区间进行了划分,并分别计算各区间对应的误差评估指标。当区间内样本量过少时,样本存在较大的偶然性,为了确保评估结果合理,当区间内样本量少于 10 时,该区间不予评估,以“—”代替。模型在不同速度区间内的评估结果如表 6 所示。

从表 6 中可以看出:模型在不同速度区间的

表现性能不尽相同,同时在相同速度区间不同类型道路的情况下,模型性能依然存在较大的差异。针对干线公路而言,本文提出的模型在速度高于 60 km/h 的情形下的预测性能总是要好于速度低于 60 km/h 的情形。针对主干道和次干道而言,在速度介于 20~40 km/h 的情形下,速度预测具有最高的预测准确性,而针对流量预测,其准确性随着速度区间的不断递增而逐渐降低。分析其原因,一是由于各等级道路通行能力的不同,导致相

同交通状态在不同等级道路下对应不同的速度区间。二是不同交通状态下,预测模型的精度会有较大差异。在非稳定交通状态下,如交通拥挤(低速区间)或离散性大(高速区间),此时的交通波动

较大,给预测带来了困难,导致预测精度的降低。因此,本文构建的模型在不同等级道路相同速度区间上的预测性能也不尽相同。总体而言,模型在中速区间内具有较好的预测性能。

表 6 不同速度区间下模型总体预测性能比较

Tab.6 Performance comparison under different speed interval

断面编号	速度区间/ (km · h ⁻¹)	流量			速度		
		MAE (pcu · 5 min ⁻¹)	RMSE (pcu · 5 min ⁻¹)	MAPE/%	MAE (km · h ⁻¹)	RMSE (km · h ⁻¹)	MAPE/%
101100203	0 ≤ v ≤ 20						
	20 < v ≤ 40						
	40 < v ≤ 60	7.38	9.04	39.01	5.42	6.42	9.93
	v > 60	7.20	9.35	27.67	4.02	5.55	5.63
109102401	0 ≤ v ≤ 20						
	20 < v ≤ 40						
	40 < v ≤ 60	6.35	8.31	25.96	3.50	4.92	6.79
	v > 60	6.33	7.99	36.28	4.48	5.83	6.68
102111	0 ≤ v ≤ 20						
	20 < v ≤ 40	3.78	4.76	13.12	1.03	2.17	3.65
	40 < v ≤ 60	4.03	5.93	14.66	1.01	1.97	3.03
	v > 60	5.02	7.87	30.71	2.37	3.73	5.32
108221	0 ≤ v ≤ 20	2.77	3.52	10.17	1.85	2.18	5.59
	20 < v ≤ 40	2.98	3.75	11.94	1.22	1.69	4.02
	40 < v ≤ 60	5.94	6.18	28.83	3.17	5.54	7.31
	v > 60						
106101	0 ≤ v ≤ 20	7.20	8.82	14.16	3.93	5.08	23.49
	20 < v ≤ 40	5.09	6.62	19.30	2.77	3.50	9.65
	40 < v ≤ 60	3.10	3.80	59.21	8.25	9.04	18.20
	v > 60						
110217	0 ≤ v ≤ 20	12.55	15.55	15.71	6.66	7.08	38.55
	20 < v ≤ 40	9.51	13.24	20.73	3.33	4.28	11.14
	40 < v ≤ 60	4.56	6.29	32.01	5.34	6.56	11.54
	v > 60						

3 结束语

笔者提出了一种决策树理论的交通流参数短时预测方法,用非参数决策树模型中的 CART 算法构建了向量预测模型对转换后的数据样本进行训练和建模,并进行最终的交通流短时预测。首先,针对非参数模型对数据形式的要求,采用时间序列滞后项将交通流序列转化成模型所需形式;接着,考虑到交通流参数之间存在的稳定关系,将流量和速度滞后项组合成向量形式作为模型构建的基础数据;最后,基于转化后的数据,选择 CART 算法生成决策树并实现交通流参数的预测。

选用 3 种道路类型共 6 个检测断面采集的交通流实测数据对所提出的模型进行了效果评估和有效性验证,实验结果表明,笔者所提出的模型相较于时间序列模型具有显著的性能优势,同时模型

具备了大部分传统非参数预测模型所不具备的预测实时性优点。因此,所提出的模型是行之有效的,在实际的交通流短时预测应用中可以作为时间序列模型的一种有效替代或选择。

参考文献

References

- [1] LEUNG H, FAOUZI N E E, KURIAN A. Intelligent Transportation System (ITS)[J]. Information Fusion, 2011, 12(1):2-3.
- [2] XIA J. Dynamic freeway travel time prediction using single loop detector and incident data [D]. Lexington, Kentucky:University of Kentucky, 2006.
- [3] GUO J, XIA J, SMITH B, et al. Kalman filter approach to speed estimation using single loop detector measurements under congested conditions [J]. Journal of Transportation Engineering, 2009, 135

- (12): 927-934.
- [4] KUMAR K, PARIDA M, KATIYAR V K. Short term traffic flow prediction for a non urban highway using artificial neural network[J]. Procedia-Social and Behavioral Sciences, 2013, 104: 755-764.
- [5] 杨兆升, 王媛, 管青. 基于支持向量机方法的短时交通流量预测方法[J]. 吉林大学学报(工学版), 2006, 36(6): 881-884.
YANG Zhaosheng, WANG Yuan, GUAN Qing. Short-term traffic flow prediction method based on SVM[J]. Journal of Jilin University (Engineering and Technology Edition), 2006, 36(6): 881-884. (in Chinese)
- [6] 祁伟, 李晔, 汪作新. 季节性 ARIMA 模型在稀疏交通流下的预测方法[J]. 公路交通科技, 2014, 31(4): 130-135.
QI Wei, LI Ye, WANG Zuoxin. Utilizing seasonal ARIMA model for traffic estimation in sparseness[J]. Journal of Highway and Transportation Research and Development, 2014, 31(4): 130-135. (in Chinese)
- [7] 邱敦国, 杨红雨. 一种基于双周期时间序列的短时交通流预测算法[J]. 四川大学学报(工程科学版), 2013(5): 011.
QIU Dunguo, YANG Hongyu. A short-term traffic flow forecasting algorithm based on double seasonal time series[J]. Journal of Sichuan University (Engineering Science Edition), 2013(5): 011. (in Chinese)
- [8] 聂庆慧, 夏井新, 钱振东. 城市道路交通流短时预测及可靠性分析[J]. 西南交通大学学报, 2013, 48(5): 955-960.
NIE Qinghui, XIA Jingxin, QIAN Zhendong. Short-term traffic flow forecasting and reliability analysis of urban road[J]. Journal of Southwest Jiaotong University, 2013, 48(5): 955-960. (in Chinese)
- [9] 陆化普, 孙智源, 屈闻聪. 大数据及其在城市智能交通系统中的应用综述[J]. 交通运输系统工程与信息, 2015, 15(5): 45-52.
LU Huapu, SUN Zhiyuan, QU Wencong. Big data and its application in urban intelligent transportation system[J]. Journal of Transportation Systems Engineering and Information Technology, 2015, 15(5): 45-52. (in Chinese)
- [10] 朱征宇, 刘琳, 崔明. 一种结合 SVM 与卡尔曼滤波的短时交通预测模型[J]. 计算机科学, 2013, 40(10): 248-251.
ZHU Zhengyu, LIU Lin, CUI Ming. Short-term traffic flow forecasting model combining SVM and kalman filtering[J]. Computer Science, 2013, 40(10): 248-251. (in Chinese)
- [11] CAMERON A C, Tri P K. Regression analysis of count data[M]. Cambridge: Cambridge University Press, 2013.
- [12] Box G E P, Gwilym M J. Time series analysis: forecasting and control [M]. Oakland, Calif: Holden-Day, 1976.
- [13] Landgrebe D. A survey of decision tree classifier methodology [J]. Systems Man & Cybernetics IEEE Transactions on, 2013, 21(3): 660-674.
- [14] Ho T K. The random subspace method for constructing decision forests [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(8): 832-844.
- [15] QUINLAN J R. Induction of decision trees [J]. Machine Learning, 1986, 1(1): 81-106.
- [16] BREIMAN L, FRIEDMAN, J H, OLSHEN, R A, et.al. Classification and regression trees [M]. Wadsworth, Belmont, CA: CRC Press, 1984.
- [17] 张宇. 决策树分类及剪枝算法研究 [D]. 哈尔滨: 哈尔滨理工大学, 2009.
ZHANG Yu. Study on decision tree classification and pruning [D]. Harbin: Harbin University of Science and Technology, 2009. (in Chinese)