

面向区域交通状态预测的时空集成模型

吴家成, 王洪钰, 肖建力

(上海理工大学, 光电信息与计算机工程学院, 上海 200093)

摘要: 在智能交通系统中, 交通状态预测发挥着至关重要的作用。针对现有的交通预测方法集中于中微观层面, 且时间和空间维度单一的问题, 提出了一种面向区域宏观交通状态预测的集成模型。该模型以交通指数为依据, 在时间维度上采用时间序列预测方法获得时间预测结果, 在空间维度上采用支持向量回归方法获得空间预测结果, 并在集成模型中将两者的结果融合。在交通指数云图上的实验结果表明, 与单一维度的时间或空间模型相比, 该模型能显著提高预测精度。

关键词: 区域交通流; 交通状态预测; 时间预测模型; 空间预测模型; 模型集成

中图分类号: TP 181 **文献标志码:** A

Spatial-temporal integrated model for regional traffic state prediction

WU Jiacheng, WANG Hongyu, XIAO Jianli

(School of Optical-Electrical and Computer Engineering, University of Shanghai
for Science and Technology, Shanghai 200093, China)

Abstract: Traffic state prediction plays an important role in intelligent transportation systems. Aiming at the problem that the existing traffic forecasting models focus on the medium and micro level, and their temporal and spatial dimensions are single, an integrated model for macroscopic traffic state forecasting was proposed. Based on the traffic index, the time series prediction model was adopted to obtain the temporal predictive results in the time dimension, and the support vector regression model was adopted to obtain the spatial predictive results in the spatial dimension. The results of two models were fused in an integrated model. Through experiments on the traffic index cloud maps, the results show that significant improvement in the prediction accuracy can be achieved by the proposed integrated model compared with the single temporal or spatial model.

Keywords: regional traffic flow; traffic state prediction; temporal prediction model; spatial prediction model; model integration

在现代智能交通系统中, 交通流分析已成为一个重要的组成部分。在这方面先前已经有了大

量的研究工作^[1-5], 而其中值得关注的是交通指数云图这一种创新的宏观交通流分析方法。交通指

收稿日期: 2020-01-23

基金项目: 国家自然科学基金资助项目 (61603257)

第一作者: 吴家成(1995-), 男, 硕士研究生。研究方向: 大数据技术与智能交通。E-mail: 182570484@st.usst.edu.cn

通信作者: 肖建力(1982-), 男, 副教授。研究方向: 数据挖掘与大数据技术、机器学习与计算机视觉、智能交通与智慧城市。E-mail: audyxiao@sjtu.edu.cn

数云图的灵感来源于气象云图,它关注的是区域的宏观交通状态,而不是中观分析中的道路状态或微观分析中的车辆状态。在交通指数云图中,一张城市路网被划分为若干网格,每个网格都可视为一个小区域,通过获取区域中所有道路的相关信息,再利用交通指数公式就能生成相应区域的交通指数。交通指数可以反映区域的交通状态,为了可视化的方便,交通指数在云图上被映射为不同的颜色,例如,红色表示一个区域是拥堵的,绿色表示一个区域是畅通的。完整的交通指数云图对于交通分析很有帮助,然而在某些情况下,由于计算交通指数的相关数据的丢失,导致了某些区域交通指数的缺失。这些缺失不具有规律性,从而对宏观交通流分析产生了阻碍。因此,本文的目标就是从时间和空间的角度出发,对区域交通指数进行预测分析。

交通流分析中的预测方法大致可分为两类:时间序列分析方法和机器学习方法^[6]。自回归移动平均 (ARIMA) 模型是时间序列分析方法的典型代表,它通常用于规律性的交通流量预测^[7]。这种类型的模型试图建立一种数学算法,以规则的时间间隔解释一系列过去的行为,然后将其应用于预测未来的值^[8-9]。该模型适用于挖掘出交通流在时间上的变化规律。

支持向量机 (SVM) 作为一种著名的机器学习方法,具有良好的自学习和非线性预测能力,在实时交通流预测中具有显著的优势^[10-11]。交通流在短时间内可能出现不规律的变化,而 SVM 模型正适合捕获这种变化。除此之外,即使在训练样本有限的情况下,它仍然可以获得相对更好的预测结果。因此,该模型适用于实时预测系统^[12]。

有关时间序列的研究中,ARIMA 及其季节性模型 SARIMA 已被广泛用于交通流预测中。例如, Xu 等^[13] 使用 ARIMA 和 Kalman 滤波器来预测实时道路交通状态, Luo 等^[14] 提出了一种基于改进的 SARIMA 和遗传算法的交通流预测算法。他们改进了原始的 ARIMA 模型并获得了令人满意的结果,这证明 ARIMA 模型从时间的角度出发的确解决了一些交通预测的问题。但是,作为一种线性模型,其在处理非线性的问题上表现并不好。

SVM 是一种基于结构风险最小化原则的机器学习方法,也属于非参数化模型的代表^[15]。机器学习方法在交通上的应用已经越来越受到学者的

关注,而 SVM 模型正是其中表现良好的一种方法^[16-17]。SVM 的分类模型 SVC 常被用于车辆的分类检测或是道路的交通分析^[18-19],而回归模型 SVR 常被用于交通流的预测分析^[20-21]或是交通事故的防范预警^[22-23]。SVM 模型之所以得到学者的关注,一方面是由于本身的精度较高,另一方面也是因为其处理交通非线性问题的能力优异。以交通流为例,交通流在一个短期时间内可能呈现出许多不规律的变化特性,对这部分预测也具有难度。传统线性预测模型不仅需要大量的历史数据,并且预测的效果可能也不尽如人意。相比之下, SVM 模型所需的训练数据较少,并且在这种短期非线性的预测问题上效果更好。

考虑到本文的区域交通状态预测不同于现有大量研究中的中微观交通流预测,在这种宏观层面的研究较少,并且区域的交通状态是复杂的,总体既在时间上呈现出一定规律性的变化,局部又在空间上表现出一种不规律的分布特点。因此,如何定量分析交通状态并进行状态预测是本文的核心。受前人研究的启发,本文在城市交通状态指数的基础上,同时从时间与空间两个角度进行状态预测,并基于两者的预测结果,尝试构建出一个集成模型,进一步提高预测精度。

1 交通指数及交通指数云图

交通指数 TI 是衡量一个区域交通状态的指标,取决于区域内的路段信息、路网信息以及时间等因素,其计算公式为^[4]

$$TI = \left[\frac{\sum_{r=1}^R P_r \left(\sum_{t=1}^T w_t \left(\sum_{i=1}^N k_i l_i \frac{V_i}{V_{fr}} \right) \right)}{\sum_{r=1}^R P_r \left(\sum_{t=1}^T w_t \left(\sum_{i=1}^N k_i l_i \right) \right)} \right] \times 100 \quad (1)$$

式中: V_i 为路段的实际交通流速度; V_{fr} 为路段的自由流速度; l_i 为路段的里程长度; k_i 为路段上的车道数; N 为路段数; w_t 为时间权重系数,在高峰时段和非高峰时段取值不同; T 为全天的时段数; P_r 为路网的权重,取值取决于路网的类型; R 为路段类型的数目。

单纯的交通指数不能直观地呈现区域的交通状态,因此,需要交通指数云图 (TICM) 这样的可视化工具。TICM 由交通速度云图改进而来^[24],其具体生成步骤如下:

- a. 将整个城市地图划分成若干 $n \times n$ 大小的网格;
- b. 根据交通指数公式计算各网格的指数值, 值越小表示区域越拥挤, 值越大表示区域越畅通;
- c. 为了可视化交通指数值, 使用了一系列颜色将指数值映射到 TICM 上, 例如, 绿色表示区域畅通, 黄色表示区域拥挤, 红色表示区域堵塞;
- d. 选定划分数目 n 和时间段, 就可以获得最终的 TICM。

TICM 中指数值的缺失大部分是由于某段时间内某些路段原始交通数据的缺失导致, 通常在时间上表现为一段 10~20 min 的数据空白, 并且在凌晨至中午的时间段内出现的频率较高; 在空间上表现为缺失值的区域分布零散, 夹杂在正常区域之中。为了从空间角度研究区域特性, 将缺失值的区域连同其周围 8 个区域划分在一起, 这样一个 3×3 的大区域称之为区域块。对于其他的正常区域, 也可以作这样的区域块划分, 目的是为了之后训练模型的需要。根据区域块的中心区域是否有指数值, 可以将区域块划分为包含待预测区域的目标区域块和非目标区域块, 其示例图如图 1 所示。

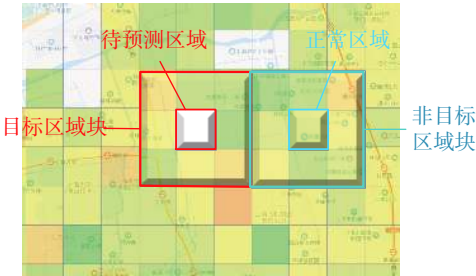


图 1 目标区域块和非目标区域块示例

Fig.1 An example of a target block and a non-target block

2 方法和模型

2.1 方法框架

基于区域块划分和时空结合预测的思想, 构建出区域交通状态指数预测的框架, 如图 2 所示。方法框架主要由 3 个部分组成: 时间预测部分、空间预测部分和最终预测部分。目标区域块在时间和空间预测部分的结果将被输入到集成模型中, 而非目标区域块的时空预测结果将被用于训练和学习集成模型的参数, 两者共同获得最终的预测结果。

在时间预测部分, 使用 ARIMA 模型基于之前

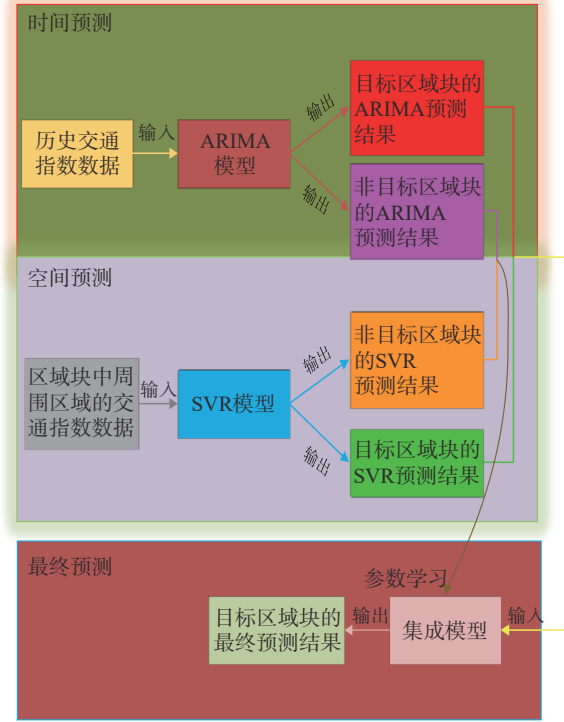


图 2 方法框架图

Fig. 2 Framework of proposed model

的历史数据进行预测。由于不同区域的状态变化情况都不尽相同, 因此, 需要对每个区域建立唯一的 ARIMA 模型。

在空间预测部分, 需要挖掘中心区域与其周围区域空间上的关系。这种空间联系不像时间那样呈现出一定的规律性, 因此, 很难用线性模型加以描述。本文采用 SVR 模型来进行空间预测, 非目标区域块的数据将作为训练集用来进行 SVR 模型的训练和参数学习。

2.2 时间 ARIMA 模型与空间 SVR 模型

ARIMA 模型是描述时间序列的模型, 它由 3 个部分组成, 分别是自回归 (AR) 部分、积分 (I) 部分和移动平均值 (MA) 部分。其中, AR 部分描述了当前值和历史值之间的关系, I 部分使采样时间序列满足平稳性的要求, MA 部分描述了模型中误差的积累^[25]。在本文中, 假设区域在 t 时间段交通指数值的 ARIMA 预测结果为 y_t 。

$$y_t = \mu + \sum_{i=1}^p \lambda_i y_{t-i} + e_t - \sum_{i=1}^q \theta_i e_{t-i} \quad (2)$$

式中: μ 为常数项; λ_i 为自相关系数; y_{t-i} 为前 i 个时间段的交通指数值; p 为 AR 部分的阶数; q 为 MA 部分的阶数; e_t 为误差项; θ_i 为误差项的系数。

对于每一个区域, 都可以通过式 (2) 获得 ARIMA

的预测结果。在每一个区域块中,本文主要关注中心区域的预测结果,并将待预测区域的预测值记为 y_T^A ,正常区域的预测值记为 y_N^A 。

SVR模型的空间预测主要以区域块为单位。假设第 i 个区域块的中心区域的指数值为 y_i ,周围8个邻近区域的指数值可以表示为一个向量 $\mathbf{x}_i = \{x_1, x_2, \dots, x_8\}$,那么,所有中心区域为正常区域的区域块所组成的模型训练集为

$$\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_i, y_i), \dots, (\mathbf{x}_n, y_n)\} \quad (3)$$

其中, n 为训练集的总个数。

SVR模型的决策函数

$$f(\mathbf{x}) = \omega^T \varphi(\mathbf{x}) + b \quad (4)$$

式中: ω 为权重向量; b 为偏差; $\varphi(\mathbf{x})$ 表示将输入映射到更高维度的非线性映射。

模型的任务首先是参数学习,即解决如下所示的最优化问题:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n l_\varepsilon(f(\mathbf{x}_i) - y_i) \quad (5)$$

式中: C 为惩罚因子; l_ε 为 ε -不敏感损失函数。

$$l_\varepsilon(\sigma) = \begin{cases} 0, & |\sigma| \leq \varepsilon \\ |\sigma| - \varepsilon, & \text{其他} \end{cases} \quad (6)$$

式中: ε 为决策平面到分界面的最大不敏感距离; σ 为距离值。

通过引入松弛变量 ξ_i 和 $\hat{\xi}_i$,式(5)可以改写为

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \hat{\xi}_i) \quad (7)$$

$$\text{s.t.} \begin{cases} f(\mathbf{x}_i) - y_i \leq \varepsilon + \xi_i \\ y_i - f(\mathbf{x}_i) \leq \varepsilon + \hat{\xi}_i \\ \xi_i > 0, \hat{\xi}_i > 0, i = 1, 2, 3, \dots, n \end{cases} \quad (8)$$

直接根据式(8)的约束条件对式(7)进行求解是很困难的,通常的做法是引入拉格朗日乘子 α_i 和 $\hat{\alpha}_i$,将问题转化为拉格朗日对偶问题。通过计算对偶问题的偏导数,可以求得 ω 的表达形式为

$$\omega = \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i) \varphi(\mathbf{x}_i) \quad (9)$$

将式(9)代入式(4),并且在处理拉格朗日乘数问题的必要条件(KKT条件)下^[26],可以得到决策函数的最终形式:

$$f(\mathbf{x}) = \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i) K(\mathbf{x}, \mathbf{x}_i) + b \quad (10)$$

式中: $K(\mathbf{x}, \mathbf{x}_i)$ 为核函数; b 为偏差。

在本文中选择了径向基核函数,它的形式为

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2) \quad (11)$$

对于每一个区域块,可以根据式(10)计算出中心区域的空间预测结果,并将待预测区域的预测值记为 y_T^S ,正常区域的预测值记为 y_N^S 。

2.3 集成模型

ARIMA模型的预测结果从时间规律性上描述了区域可能的状态,SVR模型的预测结果从空间非线性的关系中描述了区域可能的状态。但是,后续的实验部分证明,这两种模型各自的预测精度都不是很高,各有其优点和缺陷。因此,如何考虑时空相关性,以合适的方式将两者的结果进行组合与优化,是构建集成模型的关键。

在本文中,假设最终预测值是时间预测和空间预测结果的线性组合,即赋予时间和空间各一个权重,权重和为1,以此来反映时间和空间对最终预测值的贡献程度。假设第 k 个目标区域块中待预测区域的最终预测值为 $y_{T_k}^F$,根据上述假设,可以将它表示为ARIMA模型的预测结果 $y_{T_k}^A$ 和SVR模型的预测结果 $y_{T_k}^S$ 的线性组合,

$$y_{T_k}^F = \eta y_{T_k}^S + (1 - \eta) y_{T_k}^A \quad (12)$$

式中, η 为待求解的参数。

η 是集成模型参数学习的关键。对于不同的时间段, η 的值通常是不一样的,它应该动态反映时间和空间预测值之间的关系。

对于 η 的学习,如果借鉴交通流分析中速度预测的思想,通常的想法是根据一个月左右的历史数据训练模型,来学习模型的参数。但是,交通状态指数是一个特殊的、具有时效性的指标,通常只会保存几天左右的历史记录,因此,不适合这种参数学习的方式。

基于交通云图本身的特点,本文提出了一种利用非目标区域块来学习参数的自适应方法。对于第 i 个非目标区域块的中心区域,假设其最终预测值为 $y_{N_i}^F$,那么,它也可以表示为ARIMA模型的预测结果 $y_{N_i}^A$ 和SVR模型的预测结果 $y_{N_i}^S$ 的线性组合。

$$y_{N_i}^F = \alpha y_{N_i}^S + (1 - \alpha) y_{N_i}^A \quad (13)$$

由于此区域为正常区域,真实值是已知的,因此,可以判断预测结果与真实值之间的偏差。对于参数 α ,可以先假定其值,再通过偏差进行修正,来找到最优值。对于偏差的衡量,可以选择相对误差 Δ ,计算公式为

$$\Delta = \frac{|y_{N_i}^F - y_{N_i}|}{y_{N_i}} \quad (14)$$

式中, y_{N_i} 表示 i 区域的真实值。

为了表示所有非目标区域块预测结果的偏差情况, 计算相对误差的平均值

$$\bar{y}_\alpha = \frac{\sum_{i=1}^n \frac{|y_{N_i}^F - y_{N_i}|}{y_{N_i}}}{n} \quad (15)$$

式中, n 为非目标区域块的总数。

对于取不同值的 α , 都可以计算得到一个相对应的 $\bar{y}_\alpha$ 。 α 是反映时间与空间预测结果所占权重的指标。如果时间与空间中一方的结果明显优于另一方, 那么, $\bar{y}_\alpha$ 一定是随着 α 单调变化的。但是, 后续的实验表明, $\bar{y}_\alpha$ 随 α 的变化是非单调的, 并且可以求得其最优值。这说明存在一个最优的权重 α , 它衡量了时间和空间预测结果与正确预测值之间的偏差程度, 并对其进行修正调节。

由于非目标区域块与目标区域块来自同一个交通指数云图, 它们之间的时空关系近似相同, 因此, 通过非目标区域块的预测结果学习出最优的参数 α_{ool} 后, 可以令式 (12) 中的 η 等于此时的 α_{ool} , 以获得待预测区域的最终预测结果。具体的参数学习过程如算法 1 所示。

算法 1 集成模型的参数学习过程。

输入: 参数 α , ARIMA 和 SVR 模型的预测结果 y_N^A 和 y_N^S 。

输出: 参数 η 。

- 令 α 等于 0.01;
- 通过式 (15) 计算出平均相对误差 $\bar{y}_\alpha$, 并记录此时的 α 和 $\bar{y}_\alpha$;
- 令 $\alpha = \alpha + 0.01$, 如果 $\alpha < 1$, 返回步骤 b, 否则进入步骤 d;
- 画出 $\bar{y}_\alpha$ 随 α 变化的曲线图, 找到最小值的 $\bar{y}_\alpha$ 和对应的 α_{ool} 。

3 实验与结果分析

3.1 实验数据

为了评价模型的性能, 在 TICM 平台上进行了实验。之前提到, 时间上缺失值的产生多出现在凌晨和上午的时间段。同时, 从实验结果验证的角度考虑, 需要待预测区域有真实值进行对比。因此, 选取了缺失值较少的一天的凌晨至中午时

间段的数据, 时间跨度为 0:00—12:00, 数据获取的时间间隔为 10 min, 并进行了数据清洗与预处理的工作。为了更好地区分时段的特性, 将整个时间划分为 0:00—6:00 和 6:00—12:00 这 2 个时间段, 并且假定每个时间段最后 1 h 的数据是缺失的。

在空间方面, 所用的 TICM 以上海市为背景, 尺寸为 48×48 大小, 共 2304 个区域。对 TICM 进行之前所述的区域块划分操作, 并剔除异常和无效数据后, 总共有 211 个区域块。将目标区域块和非目标区域块以 3:7 的比例划分, 共得到 64 个目标区域块和 147 个非目标区域块。

3.2 评价指标

本实验中, 选用预测分析中常用的 3 个评价指标: 平均绝对误差 (MAE)、均方根误差 (RMSE) 和平均绝对百分率误差 (MAPE), 用于评估时间预测模型、空间预测模型和集成模型的性能。评价指标的值越低, 说明模型的性能越好^[12]。

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (17)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (18)$$

式中: $\hat{y}_i$ 为预测值; y_i 为真实值; n 为样本总数。

3.3 实验过程和结果分析

3.3.1 时间预测部分

在时间预测部分, 需要通过 ARIMA 模型预测每个时间段最后 1 h 的缺失值。值得注意的是, 作为一种基于历史值的预测模型, ARIMA 在离真实历史值较近的几个时间点, 预测结果通常是比较好的, 但是, 模型性能会随着时间的增加而下降。通过 ARIMA 预测结果和真实值的对比, 发现 1 h 中前 20 min 的 ARIMA 预测结果可近似认为达到了期望水平, 但是, 之后的结果则越来越不理想。因此, 实验部分着重关注了每个时间段中模型表现最差的最最后 1 个时间间隔, 即 5:50—6:00 和 11:50—12:00。对于 64 个目标区域块中的待预测区域, ARIMA 的预测值和真实值的对比如图 3 所示。

可以看出, ARIMA 模型预测值的变化趋势与真实值基本相同, 但是, 一些地方出现了很大的偏差。这一方面可能是由于某个时间交通流的剧烈变化, 例如, 早高峰, 对模型产生了干扰; 另

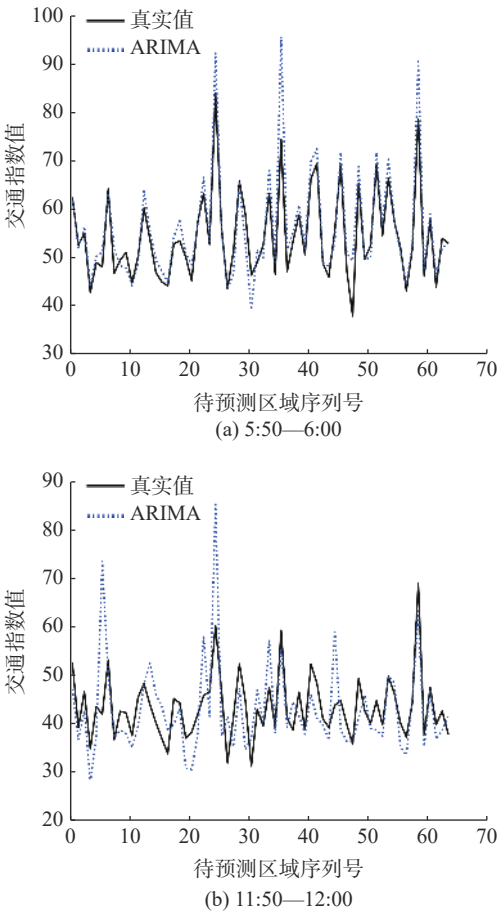


图3 ARIMA 预测值和真实值的对比

Fig.3 Comparison between the predicted value of ARIMA and true value

一方面,也是受限于模型本身的学习和预测能力,导致总体的预测精度不高。

3.3.2 空间预测部分

在空间预测部分,需要考虑区域块中心区域与周围区域的关系,对于SVR模型来说,就是目标与特征之间的关系。最直接的方法是将所有周围区域都作为特征输入,但是,结果证明并不是特征越多越好,特征间可能产生相互干扰,反而降低了模型预测的准确性。

通过借鉴现有的相关研究结果,使用随机森林的方式对模型进行特征选择^[27-28]。对于周围区域组成的一个特征集 $X=\{X1,X2,\cdots,X8\}$,通过随机森林方法可以计算得到各个特征的重要程度,如表1所示。

可以看到,X2,X4,X5,X7这4个特征的重要性明显高于其余4个,因此,可以只将它们作为输入特征。区域块中特征的分布以及特征选择的示意图如图4所示。保留的4个特征刚好为中心区域的4个最近邻,说明了这4个特征已经包含了

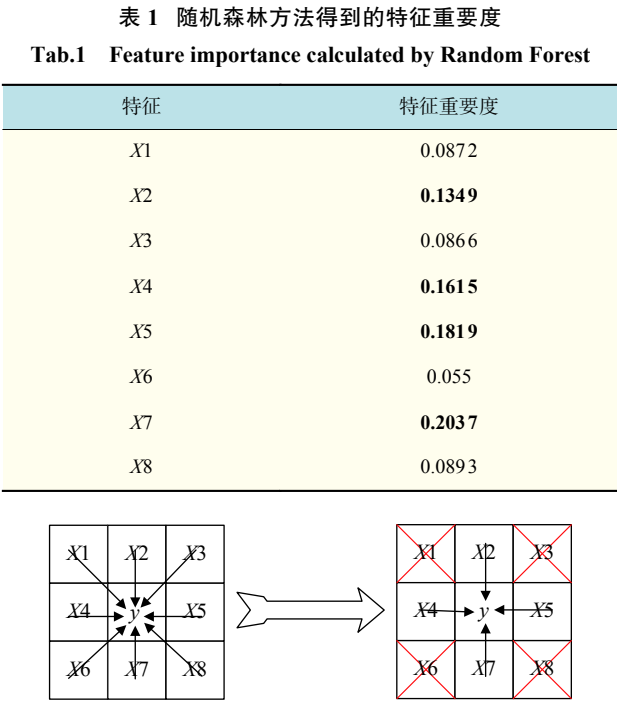


图4 用随机森林进行特征选择

Fig.4 Feature selection by Random Forest

足够的信息,再增加特征只会产生不必要的冗余。

经过特征选择后,使用SVR模型对待预测区域进行预测,得到的预测值和真实值的对比如图5所示。可以看出,SVR模型的预测值变化较为平稳,但对于其中的峰值,SVR显然没有捕获到这种变化。结合之前ARIMA预测的结果来看,这些区域可能产生了指数值突变,并且与其周围区域的联系不是很紧密,因此,用SVR模型很难准确预测。

3.3.3 集成模型预测部分

虽然ARIMA模型和SVR模型各自的预测结果都不是让人满意,但从上面的结果图中可以看到,两者的结果形成了一些互补,尤其是对于几个峰值点,ARIMA模型对变化的幅度产生了过量的预测,而SVR模型则对变化的预测不足。如果可以找到一种方法来衡量模型在不同时刻的表现,并设置合适的权重系数,那么,这样构建的集成模型应该可以有效提升预测准确性。

基于这样的思想,在方法部分详述了集成模型的构建过程。具体地,对于5:50—6:00和11:50—12:00这2个时间段,集成模型参数 α 的确定过程如图6所示。相应地,这2个时间段的集成模型可以表示为

$$y_{T_k}^F = 0.34y_{T_k}^S + 0.66y_{T_k}^A \tag{19}$$

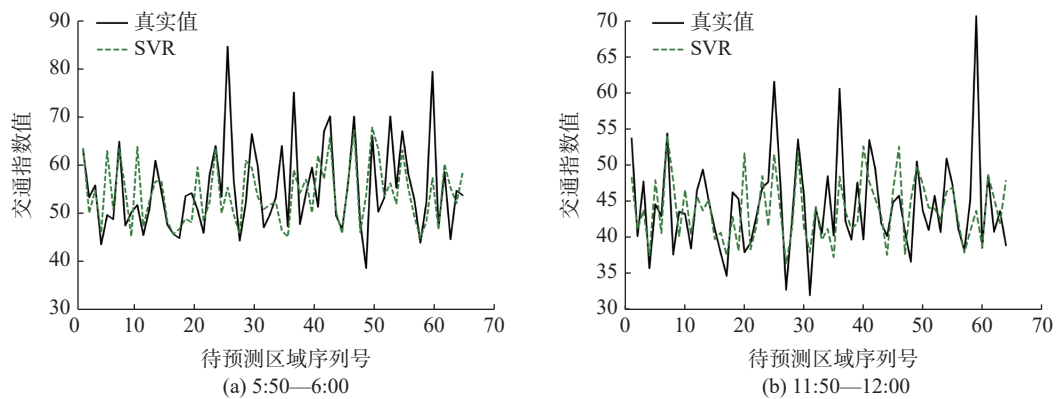


图 5 SVR 预测值和真实值的对比

Fig.5 Comparison between the predicted value of SVR and true value

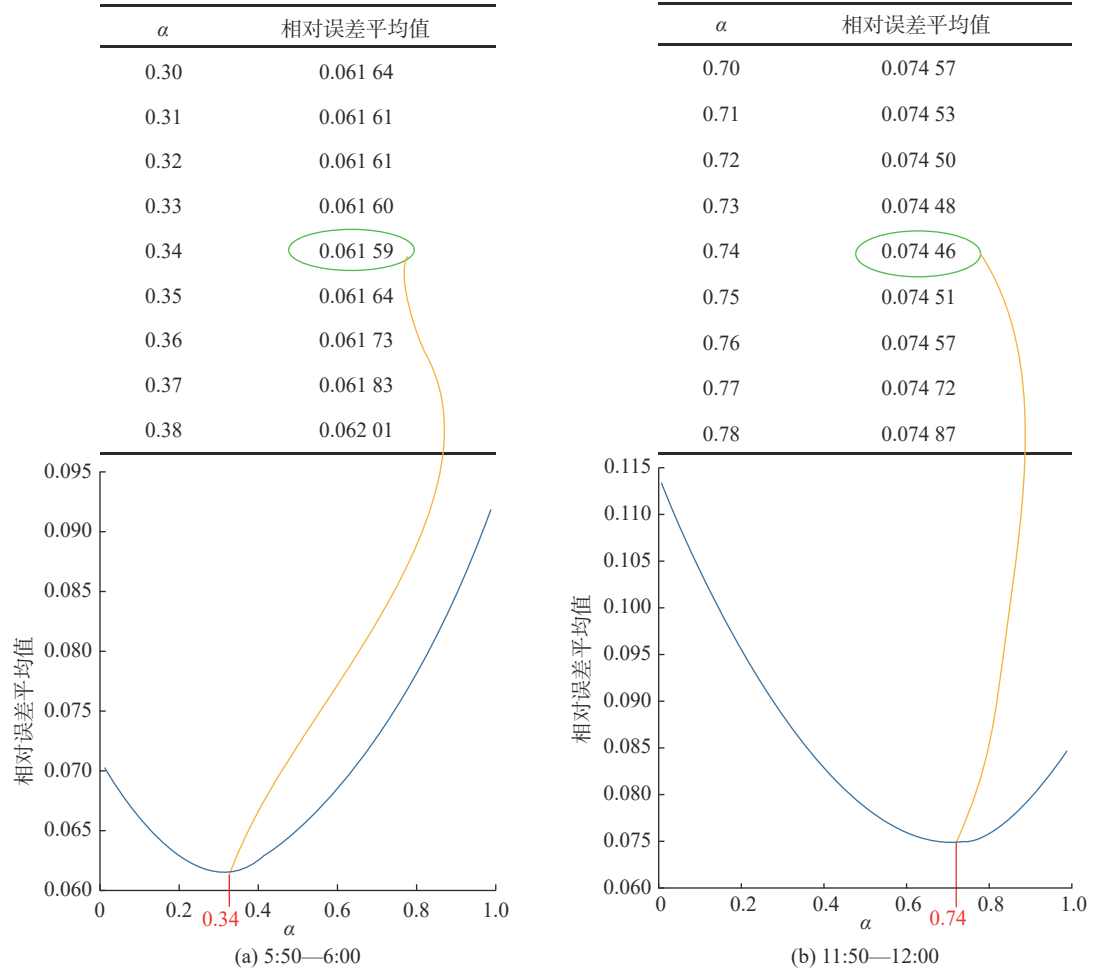


图 6 集成模型参数确定过程

Fig.6 Parameter determination of the integrated model

$$y_{T_k}^F = 0.74y_{T_k}^S + 0.26y_{T_k}^A \quad (20)$$

式 (19) 代表 5:50—6:00 时间段的模型，式 (20) 代表 11:50—12:00 时间段的模型。可以看出在 2 个不同的时间段，模型训练得到的参数相差很大。

图 7 显示了集成模型预测的值，以及其与

ARIMA 和 SVR 以及真实值进行对比的结果。可以看出，集成模型对真实值的拟合效果更好，并且对于一些峰值点也能有更准确的预测效果。评价指标对比结果如表 2 和表 3 所示，从指标值中可以看出，ARIMA 和 SVR 在 2 个不同时间段的表现确实相差较大。在 5:50—6:00 时间段，ARIMA

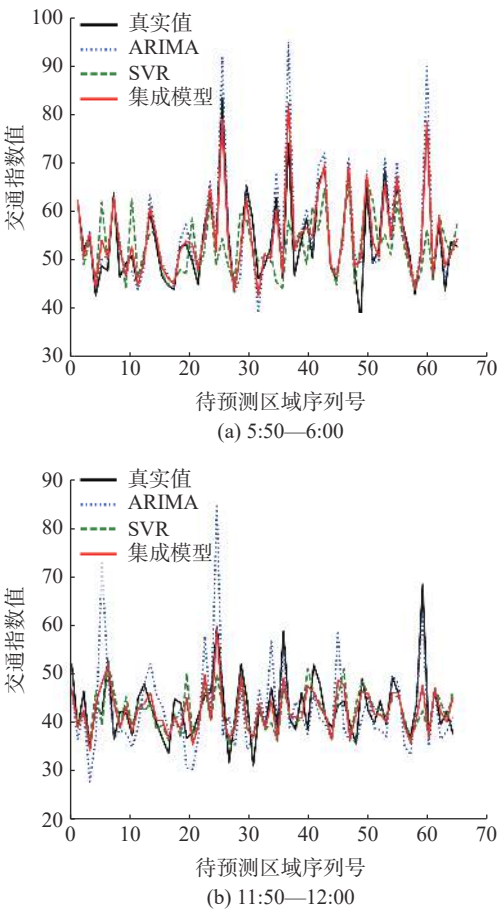


图 7 3 种模型与真实值对比结果

Fig. 7 Comparison between the predicted value of three models and true value

表 2 5:50—6:00 时 3 个模型的评价指标值
Tab.2 Evaluation indicators of three models during the time period from 5:50 to 6:00

模型	MAE	RMSE	MAPE/%
ARIMA	2.8214	4.3314	5.0636
SVR	5.4734	7.8801	9.6963
集成模型	2.2365	3.0954	4.3071

表 3 11:50—12:00 时 3 个模型的评价指标值
Tab.3 Evaluation indicators of three models during the time period from 11:50 to 12:00

模型	MAE	RMSE	MAPE/%
ARIMA	5.4070	7.5084	12.3578
SVR	3.9816	5.8251	8.9344
集成模型	3.5770	4.8208	8.1136

模型的性能明显优于 SVR 模型的性能；而在 11:50—12:00 时间段，SVR 模型的性能明显优于 ARIMA 模型的性能。

上述结果对比说明了交通流的变化是一个未知且复杂的过程，只从一个角度，用一种模型进行分析，很难得到理想的效果，这就体现了构建集成模型的必要性。时间预测模型与空间预测模型各有其优劣，它们各自的预测性能可能受预测时间、早高峰、异常拥堵区域、模型学习过程等诸多因素影响，并不存在一种模型优于另一种的情况。因此，对于集成模型来说，参数设置是否合理对最终结果有重要影响，否则，得到的结果甚至会不如单个时间或空间模型的结果。

需要强调的是，本文提出的集成模型是一种用于指数云图类预测问题的框架，所用的时间预测模型和空间预测模型可根据实际情况进行选择，并不局限于文中提到的模型。实验部分的设计也是为了模拟在一个较坏的情况下各模型间性能的对比。在实际情况中出现大范围指数值缺失的情况较少，因此，模型的性能也应该更好。

3.4 集成模型参数对效果的影响

集成模型性能好坏的关键在于参数的确定，如果集成模型的参数未按照文中提出的自适应方法进行学习，而是用经验方法人为确定，那么，模型的效果会显著下降。表 4 和表 5 呈现了评价指标 MAE 随集成模型参数取值不同对应的结果。表中前 3 列的参数是根据经验选取的，最后 1 列的参数是根据自适应方法学习得到的。可以看

表 4 5:50—6:00 不同模型参数对应的 MAE 值
Tab.4 MAE values corresponding to different model parameters during the time period from 5:50 to 6:00

模型	η			
	0.25	0.5	0.75	0.34
集成模型	2.4630	2.7034	3.9794	2.2365
ARIMA	2.8214	2.8214	2.8214	2.8214
SVR	5.4734	5.4734	5.4734	5.4734

表 5 11:50—12:00 不同模型参数对应的 MAE 值
Tab.5 MAE values corresponding to different model parameters during the time period from 11:50 to 12:00

模型	η			
	0.25	0.5	0.75	0.74
集成模型	4.4881	3.8124	3.5843	3.5770
ARIMA	5.4070	5.4070	5.4070	5.4070
SVR	3.9816	3.9816	3.9816	3.9816

出,根据经验选取的参数值对应的 MAE 都比自适应方法得到的大,说明在不同的时间段,自适应方法得到的参数总能取得相对较好的预测结果。此外,在 5:50—6:00 时间段, $\eta = 0.75$ 时的集成模型效果不如 ARIMA 模型;在 11:50—12:00 时间段, $\eta = 0.25$ 时的集成模型效果不如 SVR 模型。这验证了集成模型的性能受参数影响很大,经验方法得到的参数可能会使得集成模型效果不如一维的时间或空间模型。

4 结 论

以交通指数为基础进行了区域交通状态预测研究,将现有的中微观层面的交通流预测方法运用到了宏观层面。通过从时间与空间 2 个维度进行预测分析,提出了一种适用于云图类预测问题的时空结合集成模型框架。在交通指数云图上进行的实验表明,单独的时间或空间预测模型性能各有其不足,而自适应的集成模型弥补了它们的缺陷,并且得到的结果更为精确。

本文的研究主要以交通指数云图为依据,交通指数云图具有原创性,因此,较难找到这方面的实验对比方法。在后续研究中,将围绕所提出的集成模型框架,将更多的时间和空间预测方法代入,进行充分的实验和对比分析,以进一步阐明集成模型的合理性,并进一步提高预测结果的准确率。

参考文献:

- [1] LO S C, HSU C H. Cellular automata simulation for mixed manual and automated control traffic[J]. Mathematical and Computer Modelling, 2010, 51(7/8): 1000–1007.
- [2] HOOGENDOORN S P, BOVY P H L. State-of-the-art of vehicular traffic flow modelling[J]. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2001, 215(4): 283–303.
- [3] LI Y N, XIAO J L. Traffic peak period detection using traffic index cloud maps[J]. Physica A: Statistical Mechanics and its Applications, 2020, 553: 124277.
- [4] 肖建力. 智能交通中的多核支持向量机与分类器集成方法研究[D]. 上海: 上海交通大学, 2013: 88–94.
- [5] XIAO J L, LI H, WANG X, et al. Traffic peak period detection from an image processing view[J]. Journal of Advanced Transportation, 2018, 2018: 2097932.
- [6] MORI U, MENDIBURU A, ÁLVAREZ M, et al. A review

- of travel time estimation and forecasting for advanced traveller information systems[J]. Transportmetrica A: Transport Science, 2015, 11(2): 119–157.
- [7] WONG K I, HSIEH Y C. Short-term traffic flow forecasting for urban roads using space-time ARIMA[C]// 15th International Conference of Hong Kong Society for Transportation Studies. Hong Kong, China, 2010: 583–584.
- [8] LI L C, HE S L, ZHANG J, et al. Short-term highway traffic flow prediction based on a hybrid strategy considering temporal-spatial information[J]. Journal of Advanced Transportation, 2016, 50(8): 2029–2040.
- [9] BOX G E P, JENKINS G M, REINSEL G C. Time series analysis: forecasting and control[M]. Hoboken, New Jersey: John Wiley & Sons, 2011.
- [10] YU C Y, LAM K C. Applying multiple kernel learning and support vector machine for solving the multicriteria and nonlinearity problems of traffic flow prediction[J]. Journal of Advanced Transportation, 2014, 48(3): 250–271.
- [11] WU C H, HO J M, LEE D T. Travel-time prediction with support vector regression[J]. IEEE Transactions on Intelligent Transportation Systems, 2004, 5(4): 276–281.
- [12] FENG X X, LING X Y, ZHENG H F, et al. Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(6): 2001–2013.
- [13] XU D W, WANG Y D, JIA L M, et al. Real-time road traffic state prediction based on ARIMA and Kalman filter[J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(2): 287–302.
- [14] LUO X L, NIU L Y, ZHANG S R. An algorithm for traffic flow prediction based on improved SARIMA and GA[J]. KSCE Journal of Civil Engineering, 2018, 22(10): 4107–4115.
- [15] CHAPELLE O, VAPNIK V, BOUSQUET O, et al. Choosing multiple parameters for support vector machines[J]. Machine Learning, 2002, 46(1/3): 131–159.
- [16] LI Z, YUAN R, GUAN X. Accurate classification of the internet traffic based on the svm method[C]// 2007 IEEE International Conference on Communications. Glasgow, UK: IEEE, 2007: 1373–1378.
- [17] WANG H Y, WANG Y Z. Vessel traffic flow forecasting with the combined model based on support vector machine[C]// 2015 International Conference on Transportation Information and Safety. Wuhan, China: IEEE, 2015: 695–698.
- [18] ESTE A, GRINGOLI F, SALGARELLI L. Support vector machines for TCP traffic classification[J]. Computer Networks, 2009, 53(14): 2476–2490.

[19] JING N, YANG M, CHENG S Y, et al. An efficient SVM-based method for multi-class network traffic classification[C]//30th IEEE International Performance Computing and Communications Conference. Orlando, FL, USA: IEEE, 2011: 1–8.

[20] HONG W C, DONG Y C, ZHENG F F, et al. Hybrid evolutionary algorithms in a SVR traffic flow forecasting model[J]. Applied Mathematics and Computation, 2011, 217(15): 6733–6747.

[21] LUO X L, LI D Y, ZHANG S R. Traffic flow prediction during the holidays based on DFT and SVR[J]. Journal of Sensors, 2019, 2019: 6461450.

[22] XIAO J L, LIU Y C. Traffic incident detection using multiple-kernel support vector machine[J]. Transportation Research Record, 2012, 2324(1): 44–52.

[23] XIAO J L. SVM and KNN ensemble learning for traffic incident detection[J]. Physica A:Statistical Mechanics and its Applications, 2019, 517: 29–35.

[24] XIAO J L, WANG Z H. Traffic speed cloud maps: a new method for analyzing macroscopic traffic flow[J]. Physica A: Statistical Mechanics and its Applications, 2018, 508: 367–375.

[25] KAVOUSHI-FARD A, KAVOUSHI-FARD F. A new hybrid correction method for short-term load forecasting based on ARIMA, SVR and CSA[J]. Journal of Experimental & Theoretical Artificial Intelligence, 2013, 25(4): 559–574.

[26] JAHN J. Karush–Kuhn–Tucker conditions in set optimization[J]. Journal of Optimization Theory and Applications, 2017, 172(3): 707–725.

[27] ZHANG L Z, ALHARBE N R, LUO G C, et al. A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction[J]. Tsinghua Science and Technology, 2018, 23(4): 479–492.

[28] GENUER R, POGGI J M, TULEAU-MALOT C. Variable selection using random forests[J]. Pattern Recognition Letters, 2010, 31(14): 2225–2236.

（编辑：石 瑛）



[上接第 442 页]

[10] 严佳, 姜虹. 盲探气管插管装置联合呼气末二氧化碳监测用于困难气道患者经鼻气管插管的效果 [J]. 中华麻醉学杂志, 2012, 32(5): 579–581.

[11] APFELBAUM J L, HAGBERG C A, CAPLAN R A, et al. Practice guidelines for management of the difficult airway: an updated report by the American society of anesthesiologists task force on management of the difficult airway[J]. Anesthesiology, 2013, 118(2): 251–270.

[12] RHEE S H, YUN H J, KIM J, et al. Risk factors affecting the difficulty of fiberoptic nasotracheal intubation[J]. Journal of Dental Anesthesia and Pain Medicine, 2020, 20(5): 293–301.

[13] LANGHAN M L, EMERSON B L, NETT S, et al. End-tidal carbon dioxide use for tracheal intubation: analysis from the national emergency airway registry for children (NEAR4KIDS) registry[J]. Pediatric Critical Care Medicine, 2018, 19(2): 98–105.

[14] FIADJOE J E, HIRSCHFELD M, WU S, et al. A randomized multi-institutional crossover comparison of the GlideScope® Cobalt Video laryngoscope to the flexible fiberoptic bronchoscope in a Pierre Robin manikin[J]. Pediatric Anesthesia, 2015, 25(8): 801–806.

[15] LÁSZLÓ C J, SZŰCS Z, NEMESKÉRI Á, et al. Human cadavers preserved using Thiel's method for the teaching of fibreoptically-guided intubation of the trachea: a laboratory investigation[J]. Anaesthesia, 2018, 73(1): 65–70.

[16] PIEPHO T, WERNER C, NOPPENS R R. Evaluation of the novel, single-use, flexible aScope® for tracheal intubation in the simulated difficult airway and first clinical experiences[J]. Anaesthesia, 2010, 65(8): 820–825.

（编辑：董 伟）