

一种结合 SVM 与卡尔曼滤波的短时交通流预测模型

朱征宇 刘琳 崔明

(重庆大学计算机学院 重庆 400044) (软件工程重庆市重点实验室 重庆 400044)

摘要 针对短时交通流量的预测问题,提出了一种结合卡尔曼滤波与支持向量机的预测模型。该模型采用预测误差平方和与相关系数极大化准则智能选取恰当的预测方式,综合利用了支持向量机的稳定性与卡尔曼滤波的实时性,发挥了两种模型各自的优势。实验结果表明,该模型误差指标均低于单项预测模型。特别地,该模型在高峰时段的预测性能最佳,平均相对误差保持在8%以内,是短时交通流预测的一种有效可行的方法。

关键词 交通流,组合预测,支持向量机,卡尔曼滤波

中图分类号 TP391 **文献标识码** A

Short-term Traffic Flow Forecasting Model Combining SVM and Kalman Filtering

ZHU Zheng-yu LIU Lin CUI Ming

(Department of Computer Science, Chongqing University, Chongqing 400044, China)

(Software Engineering Chongqing Key Lab, Chongqing 400044, China)

Abstract Aiming at the issue about short-term traffic flow forecasting, a prediction model combining with Kalman filtering and support vector machine was proposed. The model adopts appropriate forecast method intelligently in each prediction period by the standards of error sum of squares and vector cosine of the angle, utilizes the stability of SVM and the real-time nature of Kalman filter comprehensively, and takes respective advantages of the two models. Experiments show that the model's error indicators are lower than single forecast model. In particular, the model in the peak hours, which average relative error is maintained at less than 8%, is a feasible and effective method of short-term traffic flow forecasting.

Keywords Traffic flow, Combining forecasting, Support vector machine, Kalman filter

1 引言

实时准确的短时交通流预测是城市交通控制与道路诱导系统的研究基础,其主要问题在于,如何有效利用历史和实时的交通数据信息来滚动预测未来一段时间的交通流状况。一般认为预测时段小于15分钟的为短时交通流预测。

目前可用于交通流预测的模型与方法很多,包括回归分析方法、时间序列预测方法、灰色预测理论、Markov预测、神经网络方法以及卡尔曼滤波等^[1,2]。不同的预测方法各有其优点和缺点^[3]。由于每种预测方法利用的数据源不尽相同,不同的数据源都是从不同的角度提供各方面有用的信息,因此各种预测方法之间并不是相互排斥,而是相互联系、相互补充的^[4]。组合预测就是把不同的预测模型组合起来,综合利用各种预测方法所提供的信息,以适当的方式得出组合预测模型和最佳预测结果,从而达到提高预测精度和增加预测可靠性的效果。

组合预测主要有两大类,分别为线性组合预测和非线性组合预测,已有大量的文献对此进行了研究。线性组合预测的思想是通过制定某些准则,如预测误差平方和、离差绝对值

和达到最小等,来确定各个单项预测模型在组合预测中所占的权重^[5]。而非线性组合预测则通常运用神经网络、支持向量机^[6]等机器学习的方法对样本进行训练,得到单项预测模型在组合预测中的权重。实验研究表明,非线性组合预测方法的总体预测效果优于线性组合预测方法^[7]。两类组合预测模型计算或训练出的权重一旦确定便不再更改,而短时交通流的随机性和不确定性较强,致使这两类方法推广能力有限。文献^[8]则提出了一种利用基因表达式编程的方法来自动生成非线性函数的组合预测模型,改变了过去只依靠各个子方法的简单线性相加的传统组合预测建模方法,所预测的数据准确度也有明显提高。但是挖掘出来的组合预测函数往往是比较复杂的超越函数,难以进一步深刻探究预测函数所蕴含的意义。

此外,除短时交通流组合预测中组合方式较重要外,单项预测方法的选取和预测模型的建立也尤为关键。支持向量机方法运用结构风险最小化原则进行学习,有较高的泛化能力,而且在训练样本很少的情况下具有很好的推广能力,在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势^[9,10]。卡尔曼滤波法则是一种基于线性回归分析的成熟的

到稿日期:2012-12-20 返修日期:2013-04-27 本文受科技部国家科技支撑计划重点项目(2011BAH25B041)资助。

朱征宇(1959—),男,博士,教授,博士生导师,CCF高级会员,主要研究方向为Web智能检索、智能交通、数据库技术,E-mail:zhu_zhengyu@cqu.edu.cn;刘琳(1987—),女,硕士,主要研究方向为智能交通;崔明(1986—),男,硕士,主要研究方向为智能交通。

预测方法,先根据现有数据形成对下一个观测值的最佳预测模型,然后用更新方程将最新的观测值加入预测向量中^[11]。支持向量机因为利用大量历史数据进行训练,其模型中蕴含了交通流量的历史规律。而卡尔曼滤波则是以不断地“预测-修正”的线性迭代方式进行计算,根据实时的交通变化规律进行动态预测。

作者根据支持向量机和卡尔曼滤波的以上特点,在分别提高两种单项预测模型精度的基础上,提出了一种短时交通流组合预测模型。

2 短时交通流预测模型

2.1 支持向量机方法简介^[12,13]

对于样本集 $T = \{(x_i, y_i), i=1, 2, \dots, l\}$, 其中 $x_i \in R^N, y_i \in R$ 首先将变量 x 通过非线性映射 $\phi(\cdot)$ 映射至高维特征空间 F , 再构建线性回归函数 $f(x) = w^T \phi(x) + b$, 其中 w 和 b 是如下凸二次规划问题的解:

$$\begin{aligned} \min_{w, b} P &= \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t. } y_i - (w^T x_i + b) &\leq \varepsilon + \xi_i \\ (w^T x_i + b) - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0, i=1, 2, \dots, l \end{aligned} \quad (1)$$

式中, C 为惩罚系数, 用来调节经验风险与表达能力之间的平衡; ε 为不灵敏损失函数; ξ_i, ξ_i^* 为松弛变量。利用拉格朗日乘子法, 原始问题转化为如下对偶问题:

$$\begin{aligned} \min_{a_i, a_i^*} D &= \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l Q_{ij} (a_i - a_i^*) (a_j - a_j^*) + \\ &\quad \varepsilon \sum_{i=1}^l (a_i + a_i^*) - \sum_{i=1}^l y_i (a_i - a_i^*) \\ \text{s. t. } 0 &\leq a_i, a_i^* \leq C q_i, i=1, 2, \dots, l \\ \sum_{i=1}^l (a_i - a_i^*) &= 0 \end{aligned} \quad (2)$$

式中, $Q_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)$, 这里 $K(\cdot)$ 为核函数, 引入核函数可以避免显示使用非线性映射 $\phi(\cdot)$, 从而避免了维数灾难。 a_i, a_i^* 为对偶问题(2)的解, 所求的回归函数可以重写为:

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x_i, x) + b \quad (3)$$

2.2 卡尔曼滤波方法简介^[14,15]

卡尔曼滤波是一种先进的控制方法, 是以 1960 年卡尔曼提出的滤波原理为基础的, 在应用于短时交通流量预测之前, 已成功应用在交通需求预测领域, 预测精度较高。卡尔曼滤波是针对时变随机信号的一种滤波方法, 实际上就是对状态的线性最小方差估计, 其原理可由一组滤波递推公式表示。随机信号及其测量过程的数学模型分别表示为:

$$X_k = A_{k,k-1} X_{k-1} + w_{k-1} \quad (4)$$

$$y_k = C_k X_k + v_k \quad (5)$$

式中, y_k 为观察向量; X_k 为状态变量; C_k 为观察矩阵; $A_{k,k-1}$ 为状态一步转移矩阵; w_{k-1} 为系统噪声向量; v_k 为测量噪声向量。假设系统噪声 w_{k-1} 和测量噪声 v_k 是互不相关的零均值高斯白噪声。其统计特性如下:

$$E[w_k] = 0, E[w_k w_j^T] = Q_k \delta_{kj}, E[v_k] = 0 \quad (6)$$

$$E[v_k v_j^T] = R_k \delta_{kj}, E[w_k v_j^T] = 0$$

状态向量 X_k 的最小均方差估计值 \hat{X}_k 可由下面的递推

方程组得到。

$$\begin{cases} \hat{X}_k = A_{k,k-1} \hat{X}_{k-1} + K_k (y_k - C_k A_{k,k-1} \hat{X}_{k-1}) \\ K_k = P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + R_k)^{-1} \\ P_{k|k-1} = A_{k,k-1} P_{k-1} A_{k,k-1}^T + Q_{k-1} \\ P_k = (I - K_k C_k) P_{k|k-1} \end{cases} \quad (7)$$

卡尔曼滤波法是一种基于线性回归分析的预测方法, 首先根据现有数据形成对下一个预测值的最佳预测模型, 然后用更新方程将最新的预测值加入预测向量中。

2.3 结合 SVM 与卡尔曼滤波的短时交通流预测模型

组合预测模型的精确度是建立在各单一预测模型基础之上的。所以, 为了提高其预测精度, 必须提高单一预测模型的精度。接下来的 3 小节分别介绍文中提出的基于支持向量机交通流预测模型、卡尔曼滤波预测模型以及结合这两种方法的组合预测模型。

2.3.1 基于支持向量机的交通流预测模型

根据交通流量的时间序列变化规律, 路段上的交通流量与前几个时段的交通流量有着必然的联系, 这样就可以利用路段前几个时段的交通流量数据序列来预测未来时段的交通流量^[16]。假设 x_i 为影响交通流量预测的因素, y_i 为交通流量预测值。基于支持向量机的交通流量预测模型的建立即寻求如下表达式成立:

$$f(x) = \sum_{i=1}^k (a_i - a_i^*) K(x, x_i) + b \quad (8)$$

式中, x 为影响交通流量的因素; x_i 为 k 个样本中的第 i 个样本; $K(x, x_i)$ 为核函数。核函数主要有 4 大类, 分别为线性、多项式、RBF(径向基)函数、sigmoid 等。本文采用的是径向基函数, 如下式所示:

$$K(x, y) = \exp \left| -\frac{\|x - y\|^2}{2\sigma^2} \right| \quad (9)$$

设当前时段为 t , $v_i(t)$ 为路段 i 在 t 时段的交通流量, $v_i(t-1)$ 为路段 i 在 $t-1$ 时段的交通流量。经实验验证, 利用前 3 个时段对预测时段的交通流量进行预测效果最佳, 因此本文采用当前时间段和前两个时间段的交通流量对未来时间段 $t+1$ 的交通流量进行预测。 t 时刻样本的输入值为 $v_i(t)$ 、 $v_i(t-1)$ 、 $v_i(t-2)$ 和 $\bar{v}_i(t+1)$, 即向量 x_i ; $v_i(t+1)$ 为样本的输出值, 即 y_i 。模型可表示为:

$$v_i(t+1) = \varphi_1 v_i(t) + \varphi_2 v_i(t-1) + \varphi_3 v_i(t-2) + \bar{v}_i(t+1) \quad (10)$$

式中, $\varphi_1 + \varphi_2 + \varphi_3 = 1$, 且 $\varphi_1 > \varphi_2 > \varphi_3$ 。不同历史时段对预测时段的影响不同, 越靠近预测时段的权重分配就越大。 $\bar{v}_i(t+1)$ 为历史同期(区分工作日与节假日)平均交通流量。根据建立的模型, 采用台湾大学林智仁(Lin Ctlih-Jen)副教授等开发的 LIBSVM 软件包对短时交通流量序列进行预测。

2.3.2 基于卡尔曼滤波的交通流预测模型

在利用卡尔曼滤波对交通流进行预测时, 仍选取路段前 3 个时间段 $k, k-1, k-2$ 的交通流量来预测 $k+1$ 时刻的交通流量值, 建立交通流量预测模型为:

$$r(k+1) = H_0 V(k) + H_1 V(k-1) + H_2 V(k-2) + \omega(k) \quad (11)$$

式中, $r(k+1)$ 为预测的交通流量系数; $V(k), V(k-1), V(k-2)$ 分别为 $k, k-1, k-2$ 时刻的交通流量系数; H_0, H_1, H_2 分

别为参数矩阵; $\omega(k)$ 为预测噪声,假定为零均值的白色噪声,其协方差矩阵为 $R(k)$ 。 $r(k+1)$ 和 $V(k)$ 定义为:

$$r(k+1) = \frac{q_F(k+1)}{q_M(k+1)} \quad (12)$$

$$V(k) = \frac{q(k)}{q_M(k)} \quad (13)$$

式中, $q_F(k+1)$ 为卡尔曼滤波在 kT 时刻对 $(k+1)T$ 时刻的交通流量的预测值; $q(k)$ 为 kT 时刻的实际交通流量值; $q_M(k)$ 代表路段在历史同期同时段 $((k-1)T, kT]$ 内的交通流量中位数,由于检测器出现设备或通讯故障时,会导致某些时段检测器数据的遗失,此时数据库在存储该时段交通流量时以0代替,因此,与文献[17]中采用历史平均值相比,本文采用历史中位数,更能提高模型的精度。

将上述模型改写为状态空间表示形式,为方便应用卡尔曼滤波理论预估状态变量,作以下变换:

$$A(k) = [V(k), V(k-1), V(k-2)] \quad (14)$$

$$X(k) = [H_0(k), H_1(k), H_2(k)]^T \quad (15)$$

$$y(k) = r(k+1) \quad (16)$$

可得

$$X(k) = B(k)X(k-1) + u(k-1) \quad (17)$$

$$y(k) = A(k)X(k) + \omega(k) \quad (18)$$

把式(17)、式(18)看作卡尔曼滤波的系统状态方程和观测方程,则 $y(k)$ 为观测向量; $X(k)$ 为状态向量; $B(k)$ 为状态转移矩阵,且 $B(k) = I$; $A(k)$ 为观测矩阵; $u(k-1)$ 为模型噪声,假定为零均值的白色噪声,其协方差矩阵为 $Q(k-1)$; $\omega(k)$ 为观测噪声,假定为零均值的白色噪声,其协方差矩阵为 $R(k)$ 。

利用卡尔曼滤波理论,可得如下方程组:

$$\begin{cases} \hat{X}(k) = X(k|k-1) + K(k)[y(k) - A(k)X(k|k-1)] \\ X(k|k-1) = B(k)\hat{X}(k-1) \\ K(k) = P(k|k-1)A^T(k) [A(k)P(k|k-1)A^T(k) + R(k)]^{-1} \\ P(k|k-1) = B(k)P(k-1)B^T(k) + Q(k-1) \\ P(k) = [I - K(k)A(k)]P(k|k-1) \end{cases} \quad (19)$$

式中, $K(k)$ 为卡尔曼滤波增益; $P(k)$ 为滤波误差方差阵。

在计算过程中 $\hat{X}(k_0)$ 为:

$$\hat{X}(k_0) = X(k_0|k_0-1) + K(k_0)[y(k_0) - A(k_0)X(k_0|k_0-1)] \quad (20)$$

以上各式中 $R(k)$ 、 $Q(k)$ 、 P_0 因没有先验数据,都设为对角阵, $\hat{X}(k_0|k_0-1)$ 设为零向量。当 $\hat{X}(k)$ 确定后,

$$r(k+1) = A(k)\hat{X}(k) \quad (21)$$

因此,路段在下一个时段的交通流量预测值为:

$$q_F(k+1) = A(k)\hat{X}(k) \times q_M(k+1) \quad (22)$$

2.3.3 组合预测模型

本文采用平均线性组合预测与0-1组合预测相结合的思想。通过待预测时段的前3个时段预测结果来决定采用何种预测方法。灵活针对当前交通流变化规律,采取相应的预测方法。在这里分两种情况:

1)在各单项预测方法对前3个时段有近似相同的预测误

差平方和时,采用简单平均线性组合方法来预测下一时段的流量。经验及实证表明,这样做是合理的^[2]。其表达式为:

$$\hat{y}_{combine} = \frac{1}{2}(\hat{y}_{sum} + \hat{y}_{kalman}) \quad (23)$$

2)在各单项预测方法预测误差平方和有一定差距时,则采用0-1组合预测模型。即根据当前交通流特性,利用相关系数极大化准则实时决策用于预测的单一模型,使得交通特性不同的时段有不同的与之相适应的预测模型,从而提高预测效果。其模型的数学表达式为:

$$\hat{y}_{combine} = \Theta(r_s - r_k)\hat{y}_{sum} + \Theta(r_k - r_s)\hat{y}_{kalman} \quad (24)$$

式中, r_s 和 r_k 分别为SVM与Kalman预测值与实测流量值前3个时段的相关系数值; \hat{y}_{sum} 、 \hat{y}_{kalman} 、 $\hat{y}_{combine}$ 分别为SVM、Kalman和组合模型在 t 时段的预测值; $\Theta(\cdot)$ 函数定义为:当 $x \leq 0$ 时, $\Theta(x) = 0$;当 $x > 0$ 时, $\Theta(x) = 1$ 。为了使预测值逼近实际值,我们希望预测值向量和实测值向量之间是高度线性相关,相关系数愈大,表明预测效果愈佳,所以 r 越大表示相应的预测方法越有效^[5]。相关系数计算公式如下:

$$r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (25)$$

设预测时段为 t , t 时段的交通流量组合预测值为 $y'_{combine}$;支持向量机预测值为 y'_{svm} ;卡尔曼滤波预测值为 y'_{kalman} 。前3个时段流量实测值向量为 $A = [A_{t-1}, A_{t-2}, A_{t-3}]$;支持向量机前3个时段预测值为 $S = [S_{t-1}, S_{t-2}, S_{t-3}]$;卡尔曼滤波前3个时段预测值为 $K = [K_{t-1}, K_{t-2}, K_{t-3}]$ 。

针对时段 t 的交通流量组合预测的具体步骤如下:

Step1 初始化各参数;

Step2 利用支持向量机和卡尔曼滤波对时段 t 的交通流量进行预测,预测值分别存储为 y'_{svm} 和 y'_{kalman} ;

Step3 计算支持向量机与卡尔曼滤波前3个时段的预测值与实测流量值的误差平方和 $SSES(t)$ 与 $SSEK(t)$,分别为: $SSES(t) = \sum_{i=1}^N (S_{t-i} - A_{t-i})^2$, $SSEK(t) = \sum_{i=1}^N (K_{t-i} - A_{t-i})^2$,其中 $N=3$ 。

Step4 判断 $SSES(t)$ 与 $SSEK(t)$ 的绝对差值是否在阈值 TS 内,若是,则时段 t 的预测值为 $y'_{combine} = \frac{1}{2}(y'_{svm} + y'_{kalman})$,否则转Step5;

Step5 分别计算支持向量机与卡尔曼滤波前3个时段的预测值 S 、 K 与实测流量值 A 的相关系数 r_s 和 r_k :

$$r_s = \frac{\sum_{i=1}^N (A_i - \bar{A})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^N (A_i - \bar{A})^2} \sqrt{\sum_{i=1}^N (S_i - \bar{S})^2}}$$

$$r_k = \frac{\sum_{i=1}^N (A_i - \bar{A})(K_i - \bar{K})}{\sqrt{\sum_{i=1}^N (A_i - \bar{A})^2} \sqrt{\sum_{i=1}^N (K_i - \bar{K})^2}}$$

Step6 判断 r_s 与 r_k 的大小,若 $r_s > r_k$,则 $y'_{combine} = y'_{svm}$;否则, $y'_{combine} = y'_{kalman}$ 。

当 t 时刻观测值到来后,即进行下一轮的预测。图1为组合预测算法流程图。

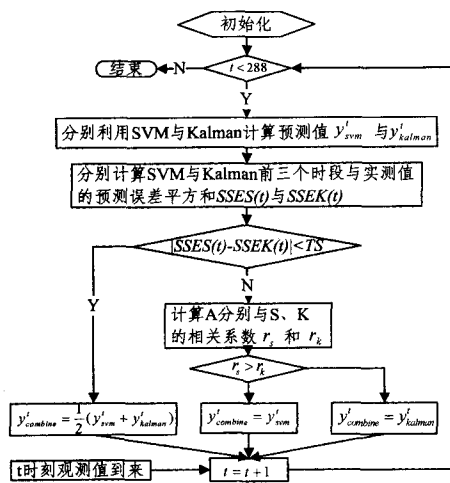


图1 组合预测算法流程图

3 实验结果与分析

3.1 评价指标

预测效果的评价采用文献[18]中定义的修正相对误差及文献[17]中的均等系数。

1. 相对误差 (Relative error, RE): $RE = \frac{|x - \hat{x}|}{x}$, 式中, x

为某时刻实测流量, 而 \hat{x} 为预测流量。当使用该指标评价全天预测效果时, 由于某些时刻的实测流量为 0, 因此引进修正的相对误差。

2. 修正相对误差 (Modified relative error, MRE): $MRE = \frac{|x - \hat{x}|}{\bar{x}}$, 其中, \bar{x} 为当日全天平均流量。该指标主要运用于全天预测效果评价, 而在评价高峰时段预测效果时, 仍使用一般的相对误差。

3. 均等系数: $EC = 1 - \frac{\sqrt{\sum_{i=1}^n (x_i - \hat{x}_i)^2}}{\sqrt{\sum_{i=1}^n (x_i)^2} + \sqrt{\sum_{i=1}^n (\hat{x}_i)^2}}$

EC 表示预测值与实测值的拟合度, 一般 0.90 以上为较好的拟合。

3.2 交通数据获取

本文采用的交通数据与文献[18]一样, 均来自于明尼苏达德卢斯大学 (University of Minnesota Duluth, <http://www.d.umn.edu/tdrl/traffic/>), 网站上提供最近一年的交通数据。可以根据需要, 输出任一检测器在若干时间间隔下记录到的交通流量和占有率两类数据。

为比较预测效果, 本文也选取了编号为 10 的检测器所记录的历史数据, 以 5 分钟为单位时间间隔输出 2012 年 1 月 1 日至 2012 年 10 月 7 日每天该路段交通流量和占有率数据。由于交通流量是公认的反映道路交通状况的最主要指标, 因此本文预测的也是交通流量数据。

3.3 实验分析

实验分别采用 SVM、Kalman 和组合预测模型来预测交通流量, 对 2012 年 10 月 8 日至 10 月 12 日共 5 天的全天和高峰时段的交通流量进行了预测。由于在实际运用中, 人们往往更关注于高峰时段车流量, 因此对高峰时段预测效果单

独评价。通过对 2012 年 1 月到 10 月该路段全年平均车流量的分析, 如图 2 所示, 每日大致 16:00—18:00 为该路段流量的高峰时段(粗线部分)。

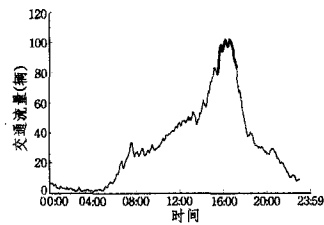


图2 10号检测器2012年1月—10月平均交通流量

以 10 月 12 日为例, 图 3 给出了高峰时段交通流 SVM、Kalman 及组合预测 3 种方法的预测结果, 输出的是 16:00—18:00 每隔 5 分钟的交通流量预测值。可以看出, 两小时内共 25 个时段的预测中, 组合预测曲线与实测值拟合得最好。另外, 卡尔曼滤波在前半段的预测效果普遍优于支持向量机, 因此, 组合预测值在前半段几乎与卡尔曼滤波预测值重合, 体现了组合预测择优而取的特点。而在 18:00 时段, 由于 SVM 与 Kalman 预测值与实测值误差平方和接近, 组合预测选择平均线性组合, 取两者平均值, 使得组合预测值与实测值重合, 预测误差接近于 0。图 4 给出了对应的高峰时段组合预测相对误差结果图。

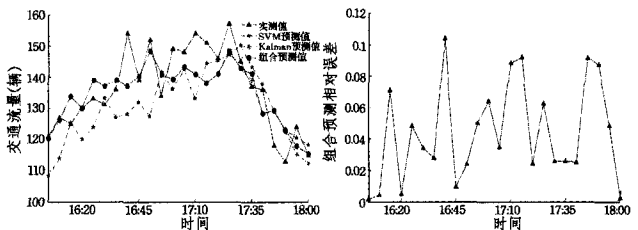


图3 高峰时段实测值与组合预测 图4 高峰时段组合预测相对误差对比示意图

表 1 和表 2 分别记录了 3 种方法下全天预测结果的修正相对误差和高峰时段的相对误差, 并汇总了 5 天的平均预测结果。表 3 记录了 3 种方法对高峰时段预测的均等系数。

表1 SVM、Kalman 及组合预测模型的全天预测修正相对误差汇总

预测模型	10月8日	10月9日	10月10日	10月11日	10月12日	平均修正相对误差
Svm	13.42%	13.93%	14.45%	13.75%	14.84%	14.07%
Kalman	13.37%	13.18%	14.46%	13.08%	14.55%	13.72%
组合预测	12.81%	13.13%	13.49%	12.97%	13.44%	13.16%

表2 SVM、Kalman 及组合预测模型的高峰时段预测的相对误差汇总

预测模型	10月8日	10月9日	10月10日	10月11日	10月12日	平均相对误差
Svm	7.84%	9.68%	7.03%	10.47%	6.95%	8.39%
Kalman	9.44%	9.00%	6.47%	10.59%	5.62%	8.22%
组合预测	8.16%	8.16%	6.51%	9.77%	5.52%	7.62%

表3 SVM、Kalman 及组合预测模型的高峰时段预测的均等系数汇总

预测模型	10月8日	10月9日	10月10日	10月11日	10月12日	均等系数
Svm	0.9494	0.9402	0.9511	0.9453	0.9613	0.9494
Kalman	0.9431	0.9489	0.9571	0.9476	0.9679	0.9529
组合预测	0.9486	0.9467	0.9552	0.9507	0.9683	0.9539

可见, 无论从预测误差, 还是均等系数来看, 组合预测模型的预测效果都是最好的。全天预测中, 组合预测修正相对

(下转第 278 页)

- [4] Priya G G L, Domic S. Edge Strength Extraction using Orthogonal Vectors for Shot Boundary Detection [J]. Procedia Technology, 2012, 6: 247-254
- [5] Kang C-C, Wang W-J, Kang C-H. Image segmentation with complicated background by using seeded region growing [J]. AEU-International Journal of Electronics and Communications, 2012, 66(9): 767-771
- [6] Haris K, Efstratiadis S N, Maglaveras N, et al. Hybrid Image Segmentation Using Watersheds and Fast Region Merging [J]. IEEE Transaction on Image Processing, 1998, 7(12): 1684-1699
- [7] Adams R, Bischof L. Seeded Region Growing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641-647
- [8] 汪阔, 万其明, 张大寿, 等. 光谱、形状特征结合的多精度图像分割算法与应用 [J]. 地球信息科学学报, 2010, 12(2): 261-268
- [9] Ding Jun-di, Ma Ru-ning, Chen Song-can, et al. A Fast Directed Tree Based Neighborhood Clustering Algorithm for Image Segmentation [J]. Lecture Notes in Computer Science, 2006, 4233: 369-378
- [10] Araújo A R F, Costa D C. Local adaptive receptive field self-organizing map for image color segmentation [J]. Image and Vision Computing, 2009, 27(9): 1229-1239
- [11] 黄慧萍, 等. 面向对象图像分析中的尺度问题研究 [D]. 北京: 中国科学院研究生院, 2003

(上接第 251 页)

误差分别比 SVM、Kalman 模型低了 6.91% 和 4.26%; 高峰时段预测中, 组合预测的相对误差分别比 SVM、Kalman 模型低了 10.10% 和 7.87%; 从表 3 中可以看到, 3 种预测方法的均等系数都在 0.90 以上, 且组合预测方法的均等系数均高于其余两种单项预测模型。

此外, 本文还将组合预测方法与文献[18]中提出的基于 K 近邻的预测方法作了比较, 该文中采用的是 10 号检测器 2006 年全年流量数据, 交通规律与本文中 2012 年数据一致, 文中预测了 2007 年 1 月 1 日到 1 月 5 日共 5 天的交通流量。表 4 记录了文献[18]与本文方法的对比结果, 可以看出, 本文提出的组合预测模型在全天预测中, 比基于 K 近邻预测方法误差降低了 11.44%, 在高峰时段的预测中, 误差降低了 17.98%。因此, 本文提出的组合预测方法是令人满意的。

表 4 文献[18]基于 K 近邻方法与本文组合预测方法误差对比

评价指标	文献[18]预测方法	本文预测方法	误差降低率
全天修正相对误差	14.86%	13.16%	11.44%
高峰时段相对误差	9.29%	7.62%	17.98%

结束语 不同的单项预测方法在不同时段的预测效果不一样, 本文提出的组合预测模型旨在智能地选取针对某一时段相应的更准的单项预测方法。其思想简单直观: 通过过去几个时段流量实测值与单项预测方法预测值之间的误差平方和及相关系数两重标准, 来决定下个时段采用何种预测方式, 消除了传统组合预测模型中, 通过改善某种拟合误差进行线性规划得出各单项预测模型所占权重而造成推广性不强的弊端。此外, 支持向量机与卡尔曼滤波两种方法的结合既包含了实时动态变化的趋势, 也潜在包含了历史各时段的流量规律。实验表明, 组合预测模型效果令人满意, 具有普遍性。作者下一步的主要工作就是: 在单步预测的基础上, 将组合预测模型运用到交通流量多步预测上, 为基于预测的导航系统提供服务。

参考文献

- [1] 邵春福, 熊志华, 姚智胜. 道路网短时交通需求预测理论、方法及应用 [M]. 北京: 清华大学出版社, 2011
- [2] 许伦辉, 傅惠. 交通信息智能预测理论与方法 [M]. 北京: 科学出版社, 2009
- [3] 宋驰, 沈国江. 短时交通流预测模型综述 [J]. 自动化博览, 2012(6): 84-87
- [4] 陈华友, 盛昭瀚, 刘春林, 等. 基于向量夹角余弦的组合预测模型的性质研究 [J]. 管理科学学报, 2006, 9(2): 1-8
- [5] 孙李红, 沈继红. 基于相关系数的加权几何平均组合预测模型的性质 [J]. 系统工程理论与实践, 2009, 29(9): 84-91
- [6] 李斌, 郝涛, 史明华, 等. 基于支持向量机的交通流组合预测模型 [J]. 天津工业大学学报, 2008, 27(2): 73-76
- [7] 韩冬梅, 牛文清, 杨荣, 等. 线性与非线性最优组合预测方法的比较研究 [J]. 情报科学, 2007, 25(11): 1672-1678
- [8] 张大斌, 张景广, 彭森, 等. 基因表达式编程在组合预测建模中的应用 [J]. 系统工程理论与实践, 2012, 32(3): 568-573
- [9] Wu Chun-hsin, Wei Chia-chen, Chang Ming-hua, et al. Travel time prediction with support vector regression [J]. IEEE Transaction on Intelligent Transportation systems, 2004, 5(12): 276-281
- [10] Theja P V V K, Vanajakshi L. Short Term Prediction of Traffic Parameters Using Support Vector Machines Technique [C] // Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on. Goa, India, 2010: 70-75
- [11] 龚珊, 尹相勇, 朱爱华, 等. 基于浮动车的路段行程时间卡尔曼滤波预测算法 [C] // 2008 第四届中国智能交通年会论文集. 2008: 1-6
- [12] Chang Ming-wei, Lin Chi-hjen. Leave-one-out Bounds for Support Vector Regression Model Selection [J]. Neural Computation, 2005, 17(5): 1188-1222
- [13] Kim K J. Financial Time Series Forecasting Using Support Vector Machines [J]. Neurocomputing, 2003, 55(3): 307-319
- [14] Kalman R. A new approach to linear filtering and prediction problems [J]. Journal of Basic Engineering, 1960, 82(01): 35-46
- [15] 陆如华, 徐传玉, 张玲, 等. 卡尔曼滤波的初值计算方法及其应用 [J]. 应用气象学报, 1999, 12(3): 63-67
- [16] 杨兆升, 王媛, 管青, 等. 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
- [17] 沈国江, 王啸虎, 孔祥杰, 等. 短时交通流量智能组合预测模型及应用 [J]. 系统工程理论与实践, 2011, 31(3): 561-568
- [18] 张涛, 陈先, 谢美萍, 等. 基于 K 近邻非参数回归的短时交通流预测方法 [J]. 系统工程理论与实践, 2010, 30(2): 376-384