

# 一种改进的短期交通流量预测算法研究

郭 新

(广东交通职业技术学院 交通信息学院, 广东 广州 510800)

**摘 要:**短期交通流量预测已经成为智能交通系统的重要研究领域。为了进行流量动态分配,积极有效地运行交通管理系统,必须要准确估计交通流量。在预测短期流量时,近期流量信息显然对将来短期流量预测具有重要的预示作用,即应该考虑取决于交通流量数据时差的相对重要性。文中提出一种新的短期流量预测模型:基于在线学习的加权支持向量回归模型(OLWSVR)。OLWSVR 模型与多种知名预测模型(包括人工神经网络模型、局部加权回归模型、传统的支持向量回归模型,及在线学习支持向量模型)进行比较。结果表明,文中模型的性能优于其他当前模型的性能。

**关键词:**智能交通系统;交通流;在线学习;支持向量回归模型

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2015)02-0103-05

doi:10.3969/j.issn.1673-629X.2015.02.024

## Research on an Improved Prediction Algorithm for Short-term Traffic Flow

GUO Xin

(Communication and Information College, Guangdong Communication Polytechnic,  
Guangzhou 510800, China)

**Abstract:** Prediction of short-term traffic flow has become one of the major research fields in intelligent transportation systems. Accurately estimating traffic flow are important for operating effectively and proactively traffic management systems in the context of dynamic traffic assignment. For predicting short-term traffic flows, recent traffic information is clearly a more significant indicator of the near-future traffic flow. In other words, the relative significance depending on the time difference between traffic flow data should be considered. It presents a novel prediction model for short-term traffic flow in this paper, called Online Learning Weighted Support-Vector Regression (OLWSVR). The OLWSVR model is compared with several well-known prediction models, including artificial neural network models, locally weighted regression, conventional support-vector regression, and online learning support-vector regression. The results show that the performance of the proposed model is superior to that of existing models.

**Key words:** intelligent transportation systems; traffic flow; online learning; support-vector regression model

## 0 引 言

智能交通系统(Intelligent Transportation System, ITS)在全方位解决交通拥挤与交通运输安全问题上是一种有效方法。而交通诱导和交通控制是 ITS 的重要组成部分,实时准确的交通流量预测是实现智能交通控制和交通诱导的前提和关键<sup>[1-2]</sup>。另外,随着 ITS 的高速发展,其数据存储更新速度更加迅速,每天都有海量数据通过各种交通信息监测网络流进 ITS 数据库。很多有用的信息,包括潜在的交通行为模式、交通管理模式等,都隐藏在巨大的数据流之中;因此,采用

有效的数据挖掘算法发掘 ITS 海量数据中的潜在重要交通规律,对于交通状态预测评估与决策优化具有重要意义。

## 1 相关工作

交通流量是展开各种交通工作不可或缺的基础数据,短期交通流量的预测是进行一系列智能控制的基础和依据,因而在智能交通系统中有着十分广泛的应用。相继有众多的学者提出了一系列方法用于交通流量预测<sup>[3-8]</sup>。如文献[3]针对短时交通流量预测的难

收稿日期:2013-11-29

修回日期:2014-03-09

网络出版时间:2014-12-27

基金项目:国家自然科学基金资助项目(61273126);广东省自然科学基金重点资助项目(10251064101000008)

作者简介:郭 新(1984-),女,博士,研究方向为系统建模、分析与评估。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20141227.1339.001.html>

题,提出一种改进型贝叶斯组合模型。该模型只根据各基本预测模型当前时刻之前几个交通流量的预测表现,通过提出的分配算法实时更新组合模型中各个基本预测模型的权重,从而改善了传统贝叶斯组合模型权重计算迭代步长过长的缺陷,提高了贝叶斯组合模型对各个基本预测模型预测精度的灵敏性。仿真实验结果表明,基于改进型贝叶斯组合模型的预测精度不仅优于单一的预测方法,而且也优于传统的贝叶斯组合模型。文献[4]提出一种混合智能数据挖掘的交通流量预测模型。首先利用自组织神经网络(SOM)的无监督学习方式实现海量数据类型特性的自动标识,降低对专家经验的依赖度;其次采用改进遗传算法(GA)优化模糊神经网络(FNN),对标识数据进行学习,建立交通流量预测模型。通过对ITS的实际数据进行分析,结果表明文中所提出的数据挖掘方法准确有效,预测精度达到95%,比不使用遗传算法优化提高了近8%。文献[5]提出了一种利用差分进化算法优化T-S模糊神经网络预测交通流量的算法。该算法利用差分进化来弥补T-S模糊神经网络连接权值和阈值选择上的随机性缺陷,从而能发挥T-S模糊神经网络泛化的映射能力,而且能使T-S模糊神经网络具有较快的收敛性以及较强的学习能力。仿真结果表明,该算法具有更好的非线性拟合能力和更高的预测准确性,在交通流量预测领域具备可行性和有效性。文献[6]提出一种小波消噪的神经网络短时交通流量预测模型。首先采用小波技术对短时交通流量数据进行消噪处理,然后采用关联维数确定BP神经网络输入变量个数,最后采用BP神经网络建立短时交通流量预测模型。仿真结果表明,与消噪前相比,消噪后模型的预测精度有了显著提高,其预测误差远远小于消噪前,预测结果更具实用价值。文献[7]提出一种蚁群算法(ACO)优化支持向量机(SVM)参数的短时交通流量预测模型(ACO-SVM)。将SVM参数的选取看作参数的组合优化问题求解,采用鲁棒性较强的ACO来搜索最优解。仿真结果表明,ACO-SVM在预测精度、收敛速度、泛化能力等方面均优于参比模型,更适合于短时交通流量的预测。文献[8]中的AOSVR是一种支持向量机的在线更新算法,具有模型在线学习的特点,可应用于交通流量的实时预测,其中模型参数的选择是预测性能的关键因素。利用大连SCOOT系统采集的实时数据,通过训练集求解AOSVR的不敏感损失系数 $\varepsilon$ 和惩罚参数 $C$ ,形成自适应参数选择的AOSVR方法。仿真结果表明,该方法能够满足动态路网交通流量预测的实时性和精确性需求,具有一定的应用价值。

但是,由于交通流量数据的高复杂性和高维数灾,

使得传统预测方法很难起到较好的预测效果,于是发展先进有效的数据挖掘算法及软件技术对于描述和分析海量数据十分迫切。鉴于此,文中在现有工作的基础上,提出了一种基于在线学习的加权支持向量回归模型(OLWSVR)来进行短期交通流量预测,并通过仿真实验对该方法的有效性进行了验证。

## 2 改进的短期交通流量预测算法

文中提出的OLWSVR算法将加权学习方法与基于在线的支持向量机结合起来实现短期流量预测。设训练集 $\{(\mathbf{x}_i, \mathbf{y}_i)\}_i^n$ 有 $n$ 个训练样本,其中 $\mathbf{x}_i \in X \subseteq \mathbb{R}^n$ 为输入向量, $\mathbf{y}_i \in Y \subseteq \mathbb{R}^n$ 是它对应的输出向量,回归函数 $f(x)$ 可写为:

$$f(x) = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

其中, $\mathbf{w} \in \mathbb{R}^n$ 和 $b \in \mathbb{R}$ 为系数。

支持向量机的目的是最小化风险函数:

$$\begin{aligned} \text{Min } \frac{1}{2} \|\mathbf{w}\|^2 + C \left( \sum_{i=1}^n \xi_i^+ + \sum_{i=1}^n \xi_i^- \right) \quad (2) \\ \text{条件: } \begin{cases} y_i - (\mathbf{w}^T \mathbf{x}_i + b) \leq \varepsilon + \xi_i^- \\ (\mathbf{w}^T \mathbf{x}_i + b) - y_i \leq \varepsilon + \xi_i^+ \\ \xi_i^+, \xi_i^- \geq 0, i = 1, 2, \dots, n \end{cases} \end{aligned}$$

其中, $C(>0)$ 表示经验风险和模型平滑度间的折中参数,松弛变量 $\xi_i^+$ 和 $\xi_i^-$ 为正。对 $\varepsilon$ -不敏感损失函数,参数 $\varepsilon$ 预先确定。在支持向量机中,常数 $C$ 表示折中参数,实现训练误差最小化和间隔(margin)最大化。如果 $C$ 对所有训练数据保持恒定,则表示支持向量机模型认为所有训练数据同等重要。然而,在时间序列预测问题中,与目标点非常接近的训练点比其他点更为重要。换句话说,与目标点 $x$ 非常接近的点具有更大的权重,远离目标点的其他点的权重更低。于是,初始支持向量机的表述(2)修正为:

$$\begin{aligned} \text{Min } \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^n C_i (\xi_i^+ + \xi_i^-) \quad (3) \\ \text{条件: } \begin{cases} y_i - (\mathbf{w}^T \mathbf{x}_i + b) \leq \varepsilon + \xi_i^- \\ (\mathbf{w}^T \mathbf{x}_i + b) - y_i \leq \varepsilon + \xi_i^+ \\ \xi_i^+, \xi_i^- \geq 0, i = 1, 2, \dots, n \end{cases} \end{aligned}$$

其中, $C_i = \mathbf{w}_f^T * C$ ,  $\mathbf{w}_f$ 为权重函数。

人们针对其他流量预测问题提出了多种权重函数<sup>[9]</sup>,文中采用高斯内核权重函数。该函数定义为:

$$\mathbf{w}_f^T = \exp \left( - \left( \frac{\sqrt{(\mathbf{x}_i - \mathbf{x})^T (\mathbf{x}_i - \mathbf{x})}}{h} \right)^2 \right) \quad (4)$$

其中, $\mathbf{x}_i$ 为训练样本; $\mathbf{x}$ 为目标样本; $h$ 为本地模型中需要被优化的带宽参数。

邻域中每个点,按其 $\mathbf{x}$ 与 $\mathbf{x}$ 的距离,确定自身权重。

引入拉格朗日乘子 $\alpha_i^+$ 和 $\alpha_i^-$ ,根据Karush-Kuhn-

Tucker (KKT) 最优性条件<sup>[10]</sup>, 优化问题的原始表述转化为如下对偶形式:

$$\text{Min } \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i^- - \alpha_i^+) (\alpha_j^- - \alpha_j^+) \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^n \gamma_i (\alpha_i^- - \alpha_i^+) + \sum_{i=1}^n \varepsilon (\alpha_i^- + \alpha_i^+) \quad (5)$$

$$\text{条件: } \begin{cases} 0 \leq \alpha_i^+, \alpha_i^- \leq C_i, i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i^+ = \sum_{i=1}^n \alpha_i^- \end{cases}$$

式(5)拉格朗日表达式可表述为:

$$L_D = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i^- - \alpha_i^+) (\alpha_j^- - \alpha_j^+) \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^n \gamma_i (\alpha_i^- - \alpha_i^+) + \sum_{i=1}^n \varepsilon (\alpha_i^- + \alpha_i^+) - \sum_{i=1}^n (\delta_i^- \alpha_i^- + \delta_i^+ \alpha_i^+) + \sum_{i=1}^n (u_i^- (\alpha_i^- - C_i) + u_i^+ (\alpha_i^+ - C_i)) + \zeta \sum_{i=1}^n (\alpha_i^- - \alpha_i^+) \quad (6)$$

其中,  $\delta_i^+, \delta_i^-, u_i^+, u_i^-$  和  $\zeta$  为拉格朗日乘子。

因此, 带有内核函数的拉格朗日表达式的 KKT 条件可被重写为:

$$\begin{aligned} \frac{\partial L_D}{\partial \alpha_i^+} &= \sum_{j=1}^n \sum_{j=1}^n K(x_i, x_j) (\alpha_j^+ - \alpha_j^-) + \varepsilon - y_i + \zeta - \delta_i^+ + u_i^+ = 0 \\ \frac{\partial L_D}{\partial \alpha_i^-} &= \sum_{j=1}^n \sum_{j=1}^n K(x_i, x_j) (\alpha_j^+ - \alpha_j^-) + \varepsilon - y_i + \zeta - \delta_i^- + u_i^- = 0 \\ \delta_i^+, \delta_i^- &> 0, \delta_i^+ \alpha_i^+ = 0, \delta_i^- \alpha_i^- = 0 \\ u_i^+, u_i^- &> 0, u_i^+ (\alpha_i^+ - C_i) = 0, u_i^- (\alpha_i^- - C_i) = 0 \end{aligned} \quad (7)$$

其中,  $K(x_i, x_j)$  表示间接计算特征空间两个向量内积的内核函数;  $\zeta$  等于优化后式(1)中的  $b$ ; 根据 KKT 条件,  $\alpha_i^-$  和  $\alpha_i^+$  最多有一个非零, 且两个均非负。因此, 系数差  $\gamma_i$  定义为:

$$\gamma_i = \alpha_i^+ - \alpha_i^- \quad (8)$$

其中,  $\gamma_i$  既确定  $\alpha_i^-$ , 又确定  $\alpha_i^+$ 。

此外, 第  $i$  个样本  $x_i$  的间隔函数 (margin function) 定义为:

$$h(x_i) \equiv f(x_i) - y_i = \sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j) \gamma_j - y_i + b \quad (9)$$

综合式(7)~(9)可得如下 5 种状态:

$$\begin{aligned} h(x_i) &\geq \varepsilon \quad \gamma_i = -C_i \\ h(x_i) &= \varepsilon \quad -C_i < \gamma_i < 0 \\ -\varepsilon &\leq h(x_i) \leq \varepsilon \quad \gamma_i = 0 \\ h(x_i) &= -\varepsilon \quad 0 < \gamma_i < C_i \\ h(x_i) &\leq -\varepsilon \quad \gamma_i = C_i \end{aligned} \quad (10)$$

在 OLWSVR 方法中, 每次增加新样本时回归参数必须逐渐增加或减小。为此, 式(10)中的 5 种状态可以通过支持向量机分类状态等 3 种子集进行辨识, 而训练集  $T$  中的样本刚好可以分类到支持向量机的这些状态中。根据  $h(x_i)$  的符号, 子集  $E$  由两个独立部分组成, 比如  $h(x_i) \geq \varepsilon, \gamma_i = -C_i$  和  $h(x_i) \leq -\varepsilon, \gamma_i = -C_i$ , 而子集  $S$  也由两个独立部分组成, 比如  $h(x_i) = \varepsilon, -C_i < \gamma_i < 0$  和  $h(x_i) = -\varepsilon, 0 < \gamma_i < C_i$ 。

$E$  集合: 误差支持向量。

$$E = \{i \mid |\gamma_i| = C_i\} \quad (11)$$

$S$  集合: 间隙支持向量机。

$$S = \{i \mid 0 < |\gamma_i| < C_i\} \quad (12)$$

$R$  集合: 剩余样本。

$$R = \{i \mid \gamma_i = 0\} \quad (13)$$

算法初始化使用两个样本来生成 OLWSVR 系数, 并使用这些系数以每次 3 个样本的频率对训练集剩余部分进行训练。如果新的样本不是支持向量, 则完全可以安全添加且无需更新参数。然而, 如果加入的样本成为支持向量, 则必须根据文献[10]中的增量算法, 逐渐增加或减小回归参数。训练结束后, 立刻更新在线测试和预测器。

文中在实施 OLWSVR 时, 使用了基于文献[11]嵌入式理论的相空间重建技术。设有时间序列  $\{x(t), t = 1, 2, \dots\}$  及预测源点  $O$  (生成预测的时间), 构建一组训练样本, 即从部分时间序列  $\{x(t), t = 1, 2, \dots, O\}$  中构建一组训练样本  $A_{O,B}$  且  $A_{O,B} = \{x(t), y(t), t = B, \dots, O-1\}$ 。其中  $x(t) = [x(t), x(t+1), \dots, x(t-B+1)]^T, y(t) = x(t+1), B$  为训练集  $A_{O,B}$  的嵌入维度。根据训练集  $A_{O,B}$  训练预测器  $P(A_{O,B}; x)$ ; 然后, 利用  $\hat{x}(O+1) = P(A_{O,B}; x(O))$  预测  $x(O+1)$ 。

为了部署 OLWSVR, 使用了定义为  $\exp(-p |x_i - x_j|^2)$  的 RBF 内核。因此, 文中 OLWSVR 的完整预测模型为:

$$\hat{y}(t) = \mathbf{w}^T \mathbf{x} + b \quad (14)$$

其中,  $\hat{y}(t)$  为  $(t+1)$  时的流量测量预测值。

### 3 仿真实验

文中依托加州高速公路性能评估 7.0 版本系统<sup>[12]</sup> (PeMS) 实时收集 30 s 循环检测器数据。这些数据整合后以 5 min 为周期, 并在互联网上面向研究人员开放。PeMS 中给出了文中所有 VDS (车辆检测站) 的位置, 有 12 个而不是 14 个地点, 因为具有相同里程 (post mileage) 的两个 VDS 在 I5-N 和 I5-S 两种场景下均有出现。在两种场景下采集流量数据集:

场景 1:典型流量状态。此场景下,没有车辆碰撞等可能剧烈改变流量模式的特殊事件发生。对随机选择的 7 个高速公路地点进行数据采集,从早上 5 点到早上 10 点,以 5 min 为间隔,持续 16 天,共获得 107 520(960 个观测值/VDS/day×7 个 VDS 站×16 天)个观测值。只包括周二、周三、周四的数据,因为这些天的流量行为属于正常状态。开始 15 天的数据用于模型训练,第 16 天的数据用于模型测试。

场景 2:非典型性流量状态。在场景 2 中,测试日(第 16 天)要么流量发生特殊情况(假日),要么在被分析的 VDS 附近有意外情况发生(交通事故)。其余 15 天用于模型训练,所有 7 个 VDS 的典型流量模式无假日或事故发生。

为了评估文中模型的预测性能,选择多层感知 NN(MLP-NN)算法、本地加权回归算法(LWR)、SVR、OLSVR 进行比较。为了运行 MLP-NN 模型,需要设置多个参数,包括隐藏层数、隐藏层的神经元数量、学习速率。由于迄今仍然缺乏系统的参数优化方法,通过多重状态下的基本实验来确定参数。MLP-NN 结构包括:输入层的 10 个神经元,具有 4 个神经元的单个隐藏层,及 1 个输出神经元。输入和输出神经元分别为  $\{x(k), k = t - 9, t - 8, \dots, t\}$  和  $x(t + 1)$ , 其中  $t$  为当前时间。隐藏层和输出节点使用正切 s 形曲线函数和线性传递函数作为激活函数。使用梯度下降学习算法来训练网络,更新连接权重和偏差。学习速率设为 0.3。

LWR 算法通过设计空间局部邻域多项式加权拟合来近似表示将来流量与过去和当前流量的关系<sup>[13]</sup>。LWR 预测模型可被表示为  $\hat{y}(t) = x(t)\beta^T + b$ , 其中在  $(t + 1)$  时  $\hat{y}(t) = x(t + 1)$ 。此外,  $\beta^T = [\beta_k, k = t - 9, t - 8, \dots, t]^T, x(t) = [x(k), k = t - 9, t - 8, \dots, t]$ 。使用 LWR 时,参数  $\beta$  可通过式  $\beta = (x^Twx)^{-1}x^Twy$  估计, 其中  $y$  为训练样本预测值向量,  $w$  为权重矩阵, 表示观测值间的连接性。式(4)中的加权函数用于确定 LWR 的权重值。

对在线算法(OLSVR 和 OLWSVR),训练方法如下:前 10 个数据点用于输入,第 11 个数据点作为输出。然后,10 点输入窗口开始“移动”,覆盖第 11 个数据点,进而生成新的 10 点输入窗口,然后将第 12 个数据点作为输入。这一过程持续到最后一个观测值成为目标为止。模型参数调整好后,在每个场景上对模型进行测试。此外,用平均绝对百分误差(MAPE)来定量描述每个模型的预测性能:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$$

(15)

其中,  $\hat{y}_i$  为观测值  $i$  的流量预测值;  $y_i$  为观测值  $i$

的实际流量值;  $n$  为预测值数量。

为进行实验比较,用表 1 中的各个检测器完成每个模型的平均绝对百分误差(MAPE)的计算。

表 1 场景 1 下各模型的总体性能

| 检测站(VDS)        | 平均绝对百分误差/% |     |     |       |        |
|-----------------|------------|-----|-----|-------|--------|
|                 | MLP-NN     | LWR | SVR | OLSVR | OLWSVR |
| I5-N(里程数:12.75) | 5.2        | 5.3 | 4.9 | 5.1   | 4.4    |
| I5-N(里程数:25.47) | 7.7        | 6.3 | 6.1 | 5.4   | 5.0    |
| SR101-N         | 8.0        | 5.5 | 5.2 | 4.8   | 4.1    |
| I5-S            | 9.9        | 9.0 | 8.8 | 9.0   | 8.1    |
| I10-W           | 6.7        | 6.5 | 6.1 | 5.1   | 5.6    |
| I5-S            | 8.3        | 6.3 | 6.5 | 7.1   | 6.1    |
| I880-S          | 5.2        | 5.0 | 4.8 | 4.6   | 4.2    |
| 平均值             | 7.3        | 6.3 | 6.1 | 5.9   | 5.3    |

表 1 表明,在线算法的结果具有可比性,平均 MAPE 值分别为 5.9% 和 5.3%。除这些模型外,传统的支持向量机的整体预测精度最优,平均 MAPE 值为 6.1%。

此外,图 1 给出了 I5-N 各算法的实际值和预测值。

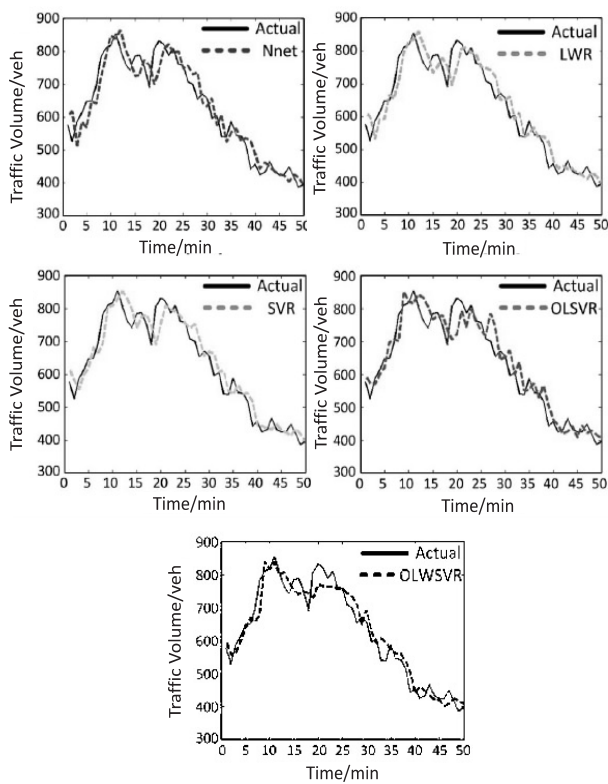


图 1 场景 1 下 I5-N 时的实际值和预测值  
收集了 3 天的测试数据(同样的时间周期),并给出了流量测量均值和预测均值。图中给出了上午 6:20 到 10:00 期间 5 分钟流量提前 1 步预测值,表明相对输出算法,文中算法与实际值更为接近。

表 2 给出了场景 2 下各 VDS 站各模型的 MAPE 值。



表2 场景2下各模型的总体性能

| 检测站<br>(VDS) | 平均绝对百分误差/% |      |      |       |        |
|--------------|------------|------|------|-------|--------|
|              | MLP-NN     | LWR  | SVR  | OLSVR | OLWSVR |
| I5-N         | 8.1        | 7.4  | 7.7  | 7.5   | 7.0    |
| I580-W       | 14.0       | 12.2 | 10.0 | 10.5  | 8.8    |
| I880-S       | 15.0       | 14.5 | 13.2 | 12.4  | 11.0   |
| SR170-S      | 19.2       | 19.2 | 18.5 | 15.0  | 13.5   |
| I5-S         | 10.5       | 10.9 | 10.7 | 10.4  | 9.9    |
| SR57-N       | 13.6       | 13.4 | 12.9 | 12.8  | 12.5   |
| SR91-E       | 22.8       | 22.9 | 22.4 | 23.4  | 20.5   |
| 平均值          | 14.8       | 14.4 | 13.6 | 13.1  | 11.9   |

所有模型中,NN模型的预测性能最低,平均MAPE值为14.8%。此时,OLWSVR模型的总体性能最优,平均MAPE值为11.9%。

此外,图2给出了场景2数据集的一个例子。

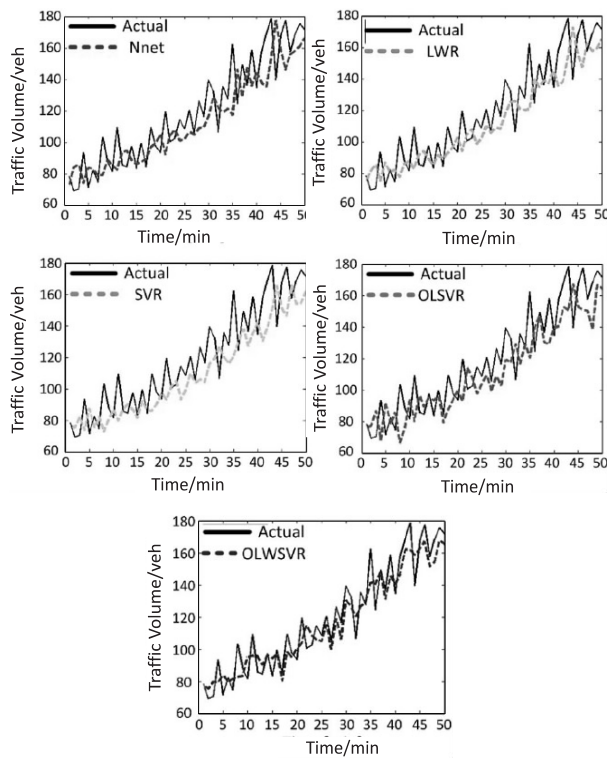


图2 场景2下I5-S时的实际值和预测值

与场景1类似,也收集了3天的测试数据(同样的时间周期),并给出了流量测量均值和预测均值。图中给出了I5-S的实际值和预测值,表明文中OLWSVR对实际值的预测效果良好。

与OLSVR相比,很明显文中模型的参数更多(加权函数 $wf_i$ )。鉴于拟合优度和模型复杂度间所做的折中,有必要研究受到参数数量影响的模型拟合。其中有多种模型评估算法,Akaike信息标准(AIC)和贝叶斯信息标准(BIC)是常用的模型评估和选择方法<sup>[14]</sup>。AIC和BIC分别表述为:

$$AIC = \log(\hat{e}_{MLE}) + \frac{2d}{n}$$

(16)

$$BIC = \log(\hat{e}_{MLE}) + \frac{d\log(n)}{n}$$

(17)

其中, $d$ 为参数数量; $n$ 为观测值数量。此外, $\hat{e}_{MLE}$ 定义为:

$$\hat{e}_{MLE} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

(18)

其中, $y_i$ 为实际流量值; $\hat{y}_i$ 为预测值。

表3和表4给出了各种场景下AIC和BIC的性能。

表3 基于AIC和BIC的模型评估(场景1)

| 监测站 (VDS)          | AIC  |       |        | BIC   |       |        |
|--------------------|------|-------|--------|-------|-------|--------|
|                    | SVR  | OLSVR | OLWSVR | SVR   | OLSVR | OLWSVR |
| I5-N<br>(历程数12.75) | 8.03 | 7.95  | 7.77   | 8.12  | 7.94  | 7.89   |
| I5-N<br>(历程数25.47) | 8.68 | 8.59  | 8.45   | 8.77  | 8.68  | 8.42   |
| SR101-N            | 9.06 | 9.00  | 8.87   | 9.15  | 9.09  | 8.89   |
| I5-S               | 9.36 | 9.29  | 9.07   | 9.45  | 9.38  | 9.11   |
| I10-W              | 9.59 | 9.54  | 9.26   | 9.68  | 9.63  | 9.48   |
| I5-S               | 9.78 | 9.69  | 9.32   | 9.87  | 9.78  | 9.54   |
| I880-S             | 9.94 | 9.84  | 9.60   | 10.03 | 9.93  | 9.72   |
| 均值                 | 9.21 | 9.13  | 8.90   | 9.30  | 9.20  | 9.01   |

表4 基于AIC和BIC的模型评估(场景2)

| 监测站 VDS | AIC   |       |        | BIC   |       |        |
|---------|-------|-------|--------|-------|-------|--------|
|         | SVR   | OLSVR | OLWSVR | SVR   | OLSVR | OLWSVR |
| I5-N    | 9.81  | 9.66  | 9.14   | 9.90  | 9.76  | 9.75   |
| I580-W  | 10.49 | 10.33 | 9.98   | 10.58 | 10.46 | 10.03  |
| I880-S  | 10.90 | 10.75 | 10.21  | 11.00 | 10.86 | 10.54  |
| SR170-S | 11.19 | 11.03 | 10.33  | 1.28  | 11.15 | 10.42  |
| I5-S    | 11.41 | 11.25 | 10.55  | 11.50 | 11.37 | 10.84  |
| SR57-N  | 11.60 | 11.43 | 10.13  | 11.69 | 11.55 | 10.02  |
| SR91-E  | 11.75 | 11.58 | 10.59  | 11.84 | 11.71 | 10.57  |
| 均值      | 11.02 | 10.86 | 10.13  | 11.11 | 10.98 | 10.31  |

从AIC和BIC的评估结果看,文中OLWSVR模型最优。从表中还可以看出,AIC和BIC的模型评估效果相似。

4 结束语

文中提出了一种高速公路短期流量预测OLWSVR模型,并与当前其他几种预测模型进行了性能对比。对于交通流量行为异常的非典型性流量场景,文中算法具有实用价值,原因是最近数据被赋予的权重较高,使得模型处理最近异常变化的速度更快,并在下次预测时加以考虑。实验结果和模型评估结果表明,文中OLWSVR模型无论是处理正常流量场景,还是发生了假日和事故的异常场景,其性能均优于其他模

### 3 结束语

文中基于现有事务提交协议提出了一种适用于无线传感器网络的提交协议,通过原理阐述与实验分析,该协议在一定程度上满足了无线传感器网络的资源和能源有限性的特点,可以保证事务执行结果的一致性,为无线传感器网络的进一步应用打下基础,将来可以进一步研究复杂网络结构下的事务提交协议以适用于大规模传感网的应用,提高服务质量。

#### 参考文献:

- [1] Bernstein P A, Goodman N. Concurrency control in distributed database systems[J]. ACM Computing Surveys, 1981, 13(2): 185-221.
- [2] Reinke C, Hoeller N, Neumann J, et al. Integrating standardized transaction protocols in service oriented wireless sensor networks[C]//Proceedings of the 24th ACM symposium on applied computing. Honolulu: ACM, 2009: 2202-2203.
- [3] Kumar V, Prabhu N, Dunham M H, et al. TCOT—a timeout-based mobile transaction commitment protocol [J]. IEEE Transactions on Computers, 2002, 51(10): 1212-1218.
- [4] Skeen D, Stonebraker M. A formal model of crash recovery in a distributed system[J]. IEEE Transactions on Software Engineering, 1983, SE-9(3): 219-228.
- [5] Nouali N, Doucet A, Drias H. A two-phase commit protocol for mobile wireless environment[C]//Proceedings of the 16th Australasian database conference. Darlinghurst: [s. n.], 2005: 135-143.
- [6] Serrano-Alvarado P, Rouvoy R, Merle P. Self-adaptive com-

ponent-based transaction commit management[C]//Proceedings of the 4th workshop on reflective and adaptive middleware systems. Grenoble: ACM, 2005.

- [7] Rouvoy R, Serrano-Alvarado P, Merle P. Towards context-aware transaction services[C]//Proceedings of the 6th IFIP WG 6.1 international conference on distributed applications and interoperable systems. Bologna: [s. n.], 2006: 272-288.
- [8] Arntsen A B, Mortensen M, Karlsen R, et al. Flexible transaction processing in the argos middleware[C]//Proceedings of the 2008 workshop on software engineering for tailor-made data management. Nantes: [s. n.], 2008: 12-17.
- [9] Helal A, Ku T H, Elmasri R, et al. Adaptive transaction scheduling[C]//Proceedings of the second international conference on information and knowledge management. Washington: [s. n.], 1993: 704-713.
- [10] 何 演, 管有庆. 基于 WS-C/T 协议的 Web 服务业务事务处理研究与实现[J]. 计算机技术与发展, 2011, 21(4): 90-93.
- [11] 管有庆, 程 强. Web 服务组合事务处理研究与实现[J]. 计算机技术与发展, 2013, 23(11): 77-81.
- [12] 胡 侃, 刘云生. 保证数据流融合处理一致性的事务机制[J]. 计算机科学, 2008, 35(1): 112-116.
- [13] 胡 侃, 刘云生. 传感器网络中协作实时数据库事务的提交控制[J]. 计算机学报, 2007, 30(6): 916-923.
- [14] Obermeier S, Böttcher S, Kleine D. A cross-layer atomic commit protocol implementation for transaction processing in mobile ad-hoc networks [J]. Distributed Parallel Databases, 2009, 26(2-3): 319-351.

(上接第 107 页)

型。当前,正对模型进行拓展,在相邻 VDS 中融入空间和时间信息,以进一步提高交通流量的预测精度。

#### 参考文献:

- [1] 江玉八. 短时交通流预测算法研究和应用[D]. 镇江: 江苏大学, 2012.
- [2] 金成均, 常桂然, 程 维, 等. 基于 IPSO 的模糊神经网络优化及交通流量预测[J]. 计算机科学, 2012, 39(10): 190-192.
- [3] 王 建, 邓 卫, 赵金宝. 基于改进型贝叶斯组合模型的短时交通流量预测[J]. 东南大学学报: 自然科学版, 2012, 42(1): 162-167.
- [4] 杨 凡, 严忠贞. 混合神经网络挖掘模型在交通流量预测中的研究[J]. 小型微型计算机系统, 2012, 33(9): 1978-1981.
- [5] 侯 越. DE 优化 T-S 模糊神经网络的交通流量预测[J]. 计算机工程与设计, 2013, 34(9): 3284-3287.
- [6] 于振洋. 小波消噪的神经网络短时交通流量预测模型[J]. 计算机仿真, 2012, 29(9): 360-363.

- [7] 徐 鹏, 姜凤茹. 基于蚁群优化支持向量机的短时交通流量预测[J]. 计算机应用与软件, 2013, 30(3): 250-254.
- [8] 王 凡, 谭国真, 杨际祥, 等. 基于 AOSVR 的交通流预测及参数选择[J]. 小型微型计算机系统, 2010, 31(6): 1245-1248.
- [9] Elattar E E, Goulermas J, Wu Q H. Electric load forecasting based on locally weighted support vector regression[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2010, 40(4): 438-447.
- [10] Ma J, Theiler J, Perkins S. Accurate on-line support vector regression[J]. Neural Computation, 2003, 15(11): 2683-2703.
- [11] Sauer T, Yorke J A, Casdagli M. Embedology[J]. Journal of Statistical Physics, 1991, 65(3): 579-616.
- [12] Freeway Performance Measurement System (PEMS), version 7.0[EB/OL]. 2006. <http://pems.dot.ca.gov/>.
- [13] Sun H, Liu H X, Xiao H, et al. Use of local linear regression model for short-term traffic forecasting[J]. Transportation Research Record, 2003, 1836(1): 143-150.
- [14] Park B B. Hybrid neuro-fuzzy application in short-term freeway traffic volume forecasting [J]. Transportation Research Record, 2002, 1802(1): 190-196.

# 一种改进的短期交通流量预测算法研究

作者：[郭新](#)，[GUO Xin](#)  
作者单位：[广东交通职业技术学院 交通信息学院, 广东 广州, 510800](#)  
刊名：[计算机技术与发展](#)[ISTIC](#)  
英文刊名：[Computer Technology and Development](#)  
年，卷(期)：2015 (2)

引用本文格式：[郭新](#). [GUO Xin](#) [一种改进的短期交通流量预测算法研究](#) [期刊论文]-[计算机技术与发展](#) 2015 (2)