

基于 LS-SVM 的交通流时序数据补齐方法

吴芳, 王晓原, 付宇

WU Fang, WANG Xiao-yuan, FU Yu

山东理工大学 交通与车辆工程学院 智能交通研究所, 山东 淄博 255049

Institute of Intelligent Transportation, School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong 255049, China

E-mail: wangxiaoyuan@sdu.edu.cn

WU Fang, WANG Xiao-yuan, FU Yu. Method for filling time series data of traffic flow based on LS-SVM. Computer Engineering and Applications, 2008, 44(29): 232-235.

Abstract: Real-time and accurate traffic flow data is essential for Intelligent Transportation Systems research. Traffic control and traffic guidance are directly affected by the quality of input data. The detected data is often incomplete and may cause out of order. The model for filling time series data of traffic flow based on LS-SVM is proposed in this paper, missing data can be filled by using traffic flow historical data. The simulation results show that LS-SVM have better generalization ability and strong robustness.

Key words: missing data; filling; Least Squares Support Vector Machines (LS-SVM); times series; traffic flow; Intelligent Transportation Systems (ITS)

摘要: 实时、准确的交通流数据是实现智能运输系统(Intelligent Transportation Systems, 简称 ITS)的关键, 对交通流的控制和诱导有直接影响。由于种种原因, 通过交通检测器获得的数据往往是不完整的, 存在丢失现象, 影响了后续模型的实际应用效果。针对这一问题, 提出一种基于最小二乘支持向量机(Least Squares Support Vector Machines, 简称 LS-SVM)的交通流时间序列数据补齐模型, 利用交通流历史数据对丢失值进行诊断和修补。利用实例仿真验证表明, LS-SVM 具有较好的泛化能力和很强的鲁棒性, 采用基于 LS-SVM 的交通流时间序列模型补齐丢失数据能够取得很好的效果。

关键词: 丢失数据; 补齐; 最小二乘支持向量机; 时间序列; 交通流; 智能运输系统

DOI: 10.3778/j.issn.1002-8331.2008.29.067 文章编号: 1002-8331(2008)29-0232-04 文献标识码: A 中图分类号: U491

1 引言

交通流数据信息是智能运输系统(Intelligent Transportation Systems, 简称 ITS), 尤其是其先进的交通管理子系统(Advanced Traffic Management Systems, 简称 ATMS)和先进的出行者信息子系统(Advanced Traveler Information Systems, 简称 ATIS)研究和开发的基础, 它们的各种子功能都是以信息应用为中心展开的。智能运输系统是一个多维世界, 信息在其中发挥着重要作用。人们的交通选择行为是在信息的支持下完成的, 出行者不仅仅是被动的对象, 更是交通管理的参与者。各先进的交通检测器实时采集动态的交通流数据, 经过种基础信息系统处理后, 可以为 ATMS 提供实时交通流信息, 从而交通管理实现从简单静态到智能动态的转变^[1-7]。

随着检测器技术的不断发展, 可以通过多种方式获得交通流数据。然而, 在实际道路交通系统中, 由于检测设备、传输设备等出现故障, 使得原始数据中各交通流参数经常存在丢失值现象^[8-10], 给交通流数据的分析以及深层次的数据挖掘带来不

利影响。目前, 实现智能运输系统的各子系统的功能大多以完整可用的数据信息集为前提。因此, 在数据预处理阶段, 形成一个合理的预测值, 实现对丢失值的修补, 可以使后续的数据分析获得满意的效果, 对实现 ATMS 和 ATIS 有着非常重要的现实意义^[2, 6, 9-11]。限于篇幅, 仅以交通流流量数据为例对丢失数据进行补齐。对丢失数据的识别可以采用如下判别规则: 根据理论输出数据的总量与实际输出数据的总量进行差值比较, 判断出丢失数据。如果所得的差值为正, 则表明有数据丢失。若为负, 则表明数据冗余^[9-10]。

基于丢失数据的补齐问题与函数逼近和估计问题在本质上是等价的这一事实, 在交通流领域, 可以把丢失数据的补齐问题看作是各交通流参数之间的复杂的非线性函数的逼近问题来处理。目前, 对丢失数据的补齐存在多种方法, 如采用前一天的历史趋势数据或历史趋势数据与前一时段的实测数据的加权估计进行补齐, 但采用上述方法存在地区、时段不同, 权值无法确定, 估计精度降低等问题。Vapnik 等人利用结构风险最

基金项目: 山东省自然科学基金(the Natural Science Foundation of Shandong Province of China under Grant No.Y2006G32); 山东理工大学科研基金重点资助项目(No.2004KJZ02)。

作者简介: 吴芳(1982-), 女, 汉族, 山东理工大学在读硕士研究生, 研究方向为智能运输系统; 王晓原(1970-), 教授, 工学博士, 研究方向为智能运输系统。

收稿日期: 2007-11-28 修回日期: 2008-02-03

小化归纳原则提出了支持向量机算法 (Support Vector Machines, 简称 SVM)^[12-13], 此算法在模式识别、函数估计、金融时间序列预测、医学上的病况检测等领域得到了广泛的应用, 但未在交通领域丢失数据方面实现应用。该算法较好地解决了上述缺陷, 而且在训练样本很少的情况下具有很好的推广能力。本文所采用的最小二乘支持向量机 (Least Squares Support Vector Machines, 简称 LS-SVM) 是 SVM 的一种演变, 即将 SVM 法中的不等式约束改为等式约束, 且将误差平方和 (Sum Squared Error, 简称 SSE) 损失函数作为训练集的经验损失, 这样就把问题转化为一个线性矩阵求解问题。因此, 本文采用 LS-SVM 建立交通流时间序列数据补齐模型, 利用交通流历史数据对丢失值进行了诊断和修补, 取得了很好的效果。

2 SVM 和 LS-SVM 的原理

2.1 SVM 的原理

20 世纪 90 年代, Vapnik 等人提出了一类新机器学习方法——支持向量机, 它建立在统计学习理论基础之上, 具有完备的统计学习理论基础和出色的学习性能, 能较好地解决小样本、非线性、高维数和局部极小点等实际问题^[12-13]。

SVM 方法的基本思想是以结构风险最小化 (Structural Risk Minimization, 简称 SRM) 原则为理论基础, 通过某种特定的非线性映射把样本空间映射到一个高维乃至无穷维的特征空间 (Hilbert 空间), 并在特征空间中寻求最优划分或回归线性超平面, 把此平面作为分类决策面, 从而解决样本空间中的高度非线性分类和回归等问题。SVM 的最终决策函数只由少数的支持向量所确定, 即具有稀疏性。计算的复杂性取决于支持向量的数目, 而不是样本空间的维数, 这在某种意义上避免了“维数灾难”现象的发生^[12-13]。

设给定训练数据集 $\{x_i, y_i\}_{i=1}^N, x_i \in R^n$ 为 n 维系统输入向量, $y_i \in R$ 为系统输出, $f(x)$ 为待估计的未知函数。作非线性映射 $\varphi: R^n \rightarrow H$, 其中 φ 称为特征映射, H 为特征空间, 则采用如下形式的函数对未知函数进行估计^[12-18]:

$$f(x) = \omega^T \varphi(x) + b \quad (1)$$

其中 ω 为 R^n 空间中的权向量, $b \in R$ 为偏置。SVM 算法用原空间的核函数代替高维特征空间的点积运算, 避免了复杂运算, 并利用结构风险最小化为学习规则, 用数学描述为 $\omega^T \omega \leq \text{常数}$ 。标准 SVM 算法以 ε -不敏感损失函数作为结构风险最小化的估计问题, ε -不敏感损失函数的含义是, 当 x 点的观察值 y 与预测值 $f(x)$ 之差不超过事先给定的 ε 时, 则认为在该点的预测值 $f(x)$ 是无损失的, 尽管预测值 $f(x)$ 和观察值 y 可能并不相等。因此建立最优化方程为

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

式中 C 为平衡因子, ξ_i, ξ_i^* 为引入训练集的误差, 它们可表示样本点超出拟合精度的程度, $\varepsilon > 0$ 。引入 Lagrang 函数

$$\begin{aligned} L(\omega, b, \xi_i, \xi_i^*) = & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N (\xi_i + \xi_i^*) - \\ & \sum_{i=1}^N \alpha_i (\xi_i + \varepsilon - y_i + \omega^T \varphi(x_i) + b) - \\ & \sum_{i=1}^N \alpha_i^* (\xi_i^* + \varepsilon + y_i - \omega^T \varphi(x_i) - b) - \\ & \sum_{i=1}^N (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (3)$$

式中 $\alpha_i, \alpha_i^* \geq 0, \eta_i, \eta_i^* \geq 0, i=1, \dots, N$ 。函数的最优值应满足条件

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \eta_i = 0, i=1, \dots, N \\ \frac{\partial L}{\partial \xi_i^*} = 0 \Rightarrow C - \alpha_i^* - \eta_i^* = 0, i=1, \dots, N \end{cases} \quad (4)$$

式(4)中 $\alpha_i - \alpha_i^* \neq 0$ 对应的 x_i 为支持向量, ω 反映了函数的复杂度, 是非线性映射函数 $\varphi(\cdot)$ 的线性组合。引进核函数来代替非线性映射 $\varphi(\cdot)$, 即 $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ 。

将式(4)代入式(3)中, 得到优化问题的对偶形式为

$$\begin{aligned} \max & -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) + \\ & \sum_{i=1}^N (\alpha_i - \alpha_i^*) y_i - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) \\ \text{s.t.} & \begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases} \end{aligned} \quad (5)$$

由式(5)、(6)可求得训练样本集的非线性逼近为 $f(x) =$

$$\sum_{i=1}^N (\alpha_i^* - \alpha_i) K(x_i, x) + b。$$

2.2 LS-SVM 的原理

Suykens 在文献[14]中提出了 LS-SVM 方法。对于 LS-SVM, 选择 ξ_i 的二范数作为损失函数, 设为 e_i , 则此优化问题变为^[12-14]:

$$\min_{\omega, b, e} J(\omega, e) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2, \gamma > 0 \quad (7)$$

$$\text{s.t. } y_i = \omega^T \varphi(x_i) + b + e_i, i=1, 2, \dots, N \quad (8)$$

在这里, γ 为一实数常量, 它决定了 $\frac{1}{2} \omega^T \omega$ 和 $\frac{1}{2} \sum_{i=1}^N e_i^2$ 的相对大小, 能够在训练误差和模型复杂度之间取一个折中, 以便使所求的函数具有较好的泛化能力。LS-SVM 算法定义了与标准 SVM 算法不同的损失函数, 并将其不等式约束改为等式约束, 可在对偶空间上求取 ω 。

引入 Lagrange 函数

$$L(\omega, b, e; \alpha) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i [\omega^T \varphi(x_i) + b + e_i - y_i] \quad (9)$$

其中 $\alpha_i \in R, \alpha_i > 0$ 为 Lagrange 乘子, 于是最优解的条件如下:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^N \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i, i=1, \dots, N \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow y_i = \omega^T \varphi(x_i) + b + e_i, i=1, \dots, N \end{cases} \quad (10)$$

由式(10)消去 ω 和 e_i , 此优化问题转化为求解如下方程:

$$\begin{bmatrix} \mathbf{b} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{B} + \gamma^{-1} \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (11)$$

其中, $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$, $\mathbf{1} = [1, \dots, 1]^T$, \mathbf{B} 为一个方阵, 其第 i 列 j 行的元素为 $B_{ij} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j)$, $i, j = 1, \dots, N$, $K(x_i, x_j)$ 为核函数。

在式(3)的基础上可进一步求出 ω , 从而得到训练数据集的非线性逼近

$$f(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \quad (12)$$

2.3 核函数的选择

SVM 的特点之一在于核函数的引入。低维空间向量集通常难于划分, 解决的方法是将它们映射到高维空间, 但这个办法带来的困难就是计算复杂度的增加, 而核函数正好巧妙地解决了这个问题。在 SVM 理论中, 采用不同的核函数将构成不同的 SVM, 这样也形成了不同的算法。目前研究的核函数主要有以下几种形式^[12-18]:

(1) 线性核函数: $K(x, x_i) = x \cdot x_i$;

(2) 多项式核函数: $K(x, x_i) = [(x \cdot x_i) + 1]^d$, 其中 d 为多项式的阶数;

(3) 高斯径向基核函数(Radial Basis Function, RBF): $K(x, x_i) = \exp\left(-\frac{|x-x_i|^2}{2\sigma^2}\right)$, 其中 σ 为 RBF 函数的宽度参数;

(4) 神经网络核函数(Sigmoid): $K(x, x_i) = \tanh(v(x \cdot x_i) + c)$, 其中 v, c 为常数。

3 交通流丢失数据补齐模型及仿真

所谓采用历史数据对丢失数据进行补齐就是利用过去的观测值来估计丢失值, 实际上是基于这样一个假设, 即丢失值与过去值之间存在某种确定的函数关系。所以补齐的目的就是试图寻找一个函数以确定丢失值和过去值之间的关系, 也就是说采用历史数据对丢失数据的补齐问题与函数逼近问题在本质上是一致的。

本文重点选取交通流量的丢失数据进行补齐, 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。设 $Q(\tau)$ 为所要补齐的 τ 时刻路段 L 上丢失的交通流量数据, $Q(\tau-k)$ 是 $\tau-k$ ($k=1, 2, \dots, m$) 时刻路段 L 上的交通流量向量。本文考虑前 s ($s < m$) 个时段的交通流量(即 $\tau-1, \tau-2, \dots, \tau-s$ 时段的交通流量)对路段 L 的交通流量的影响。所以要在 $Q(\tau)$ 和此路段上前 s 个交通流量(即 $Q(\tau-1), \dots, Q(\tau-s)$)之间寻找一个函数关系或者它的一个逼近。令 $\mathbf{x}(\tau) = [Q(\tau-1), Q(\tau-2), \dots, Q(\tau-s)]^T$, $\mathbf{y} = Q(\tau)$, 则采用历史数

据对丢失数据补齐的模型为:

$$f(x) = \omega^T \varphi(x(\tau)) + b \quad (13)$$

其中 ω, b 是要要求的, 式(13)和式(1)是相同的, 因此可以运用 LS-SVM 对交通流时间序列进行补齐。

3.1 数据来源

本文根据某交通路段存在丢失数据的这一时刻的过去一段时间的交通流量, 补齐这一时刻含有丢失数据的交通流量。为了验证模型的有效性, 所采集的数据是在车流密度不大, 车辆间相互影响微弱, 其他外界干扰因素较少的情况下实测得到的。本文采用的数据为淄博市综合交通规划调查数据, 采集地点为淄博市张店区张周路, 采集时间是 2007 年 5 月 14 日~2007 年 5 月 18 日 07:00A.M.~19:00P.M., 数据采样间隔为 5 min, 将每天的数据定为 1 组, 共为 5 组。

本文所采用的学习样本是 2007 年 5 月 14 日~2007 年 5 月 17 日的前 4 组共 576 个流量数据, 验证样本是 2007 年 5 月 18 日的 1 组共 144 个数据(包含人造丢失数据个数)。原始数据本身并不存在丢失值, 为了验证 LS-SVM 方法补齐丢失数据的有效性, 将原始数据进行改造, 在交通流量数据中随机引入丢失值, 人为地制造一些丢失数据, 然后使用上述模型对丢失数据进行补齐。图 1 所示是 2007 年 5 月 18 日 07:00A.M.~19:00P.M. 人为制造丢失数据后的交通流量数据图。

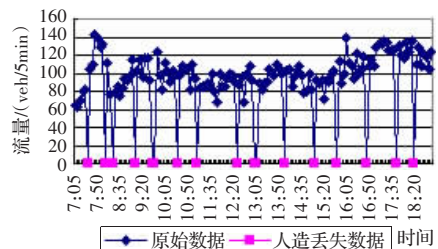


图1 含人造丢失数据的交通流量图

3.2 算法流程

首先需要确定参数 γ 和核宽度 σ^2 , 然后利用上述算法可以很容易的进行丢失数据的补齐, 具体步骤如流程图 2 所示。

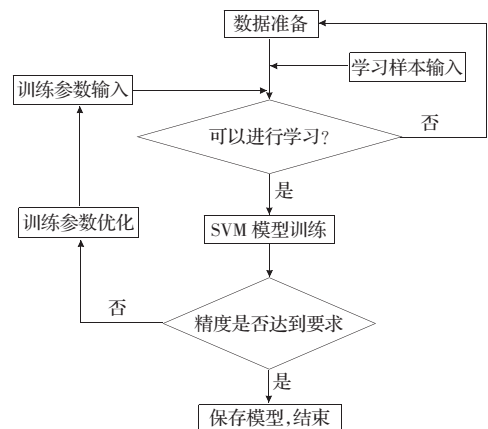


图2 算法流程图

3.3 仿真应用

道路流量在一周内具有统计规律, 本文所采用的数据为星期一到星期五的交通流量, 使得数据统计规律具有较好的稳定性。为了进行丢失数据的补齐, 首先利用前 4 组数据对模型进行训练, 最后 1 组数据用于丢失数据的补齐。丢失数据的补齐

在 Matlab 7.0 环境下完成, 为了降低建模误差, 对输入数据和期望输出数据分别在 $[-1, 1]$ 和 $[0, 1]$ 上做了归一化处理, 建模分析后再将其返还到原始区间。本文核函数采用径向基函数, 如下式所示

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

仿真发现, 参数 γ 和 σ^2 核宽度对补齐效果有较大影响, 需要根据训练数据的具体特性具体确定。通常减小 σ^2 会提高训练精度, 但会降低泛化能力, 增大 γ 也能提高训练精度。对不同参数 γ 和 σ^2 核宽度的训练表明, 当 $\gamma=100, \sigma^2=0.5$ 时经过学习训练的模型进行丢失数据补齐的效果较好(如图 3 所示)。

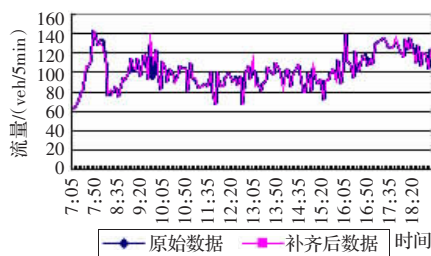


图 3 原始数据与补齐后数据对比图

3.4 仿真结果分析

为了能较好的评价补齐结果, 引入以下 4 个评价指标:

$$\text{均等系数 } EC = 1 - \frac{\sqrt{\sum_t (T_{fill}(t) - T_{real}(t))^2}}{\sqrt{\sum_t T_{fill}(t)^2} + \sqrt{\sum_t T_{real}(t)^2}}$$

$$\text{平均相对误差 } mrerr = \frac{1}{N} \sum_t \frac{T_{fill}(t) - T_{real}(t)}{T_{real}(t)}$$

$$\text{平均绝对相对误差 } marerr = \frac{1}{N} \sum_t \frac{|T_{fill}(t) - T_{real}(t)|}{T_{real}(t)}$$

$$\text{最大绝对相对误差 } mxarerr = \max \left| \frac{T_{fill}(t) - T_{real}(t)}{T_{real}(t)} \right|$$

其中 EC 表示补齐值与真实值之间的拟合, 一般在 0.90 以上表示拟合较好。

补齐结果的误差指标值见表 1 所示。表中的误差指标值 $marerr$ 显示了较低的平均误差 5.47%, 而且表中的误差指标值 EC 达到了 0.9548 的高拟合度。可见, 基于时间序列的 LS-SVM 算法补齐交通流量丢失数据有很好的效果。

表 1 丢失数据补齐所得的各项误差指标

误差指标	EC	$mrerr(\%)$	$marerr(\%)$	$mxarerr(\%)$
结果	0.9548	0.61	5.47	18.72

从图 3 中可以看出, 当相邻两个时间段数据均有丢失时, 算法的补齐精度有所降低, 本算法要求相邻时间段不能同时出现数据丢失情况, 这在实际中很难保证, 仍需进一步的研究, 以提高模型补齐相邻时间段同时出现数据丢失的准确性。

4 结论

(1) 本文利用交通流量的历史数据基于 LS-SVM 的交通流时间序列方法可以实现丢失数据的有效补齐。

(2) 同传统的按照历史趋势数据进行补齐的几种简单方法相比, 本模型在精度、泛化能力、最优性方面显示了较大的优势。

(3) 参数 γ 、核函数的类型以及核参数对补齐结果有很大的影响, 选择什么样的参数及核函数以提高补齐效果是未来的研究向之一。

(4) 本文只是用基于时间序列的 LS-SVM 方法对有限样本进行了丢失数据的补齐, 对于大量数据的多处丢失问题有待进一步研究。

(5) 待参数标定后, 本模型不需要从历史数据库中提取前一天的历史趋势数据, 计算快速、简单, 适用于模型的在线处理。

(6) 式(4)中 $\alpha_i^* - \alpha_i \neq 0$ 对应的 x_i 为支持向量, 即决定了标准 SVM 具有稀疏性, 但 LS-SVM 不存在支持向量, 即不具备 SVM 所拥有的稀疏性。但是对于研究交通流丢失数据的补齐来说, 稀疏性并不像用 SVM 来处理函数估计那么重要, 因此选择 LS-SVM 方法来进行交通流丢失数据的补齐, 取得了很好的效果。

参考文献:

- [1] 丹尼尔 L·鸠洛夫, 马休 J·休伯. 交通流理论[M]. 北京: 人民交通出版社, 1983.
- [2] 杨兆升. 基础交通信息融合技术及其应用[M]. 北京: 中国铁道出版社, 2005.
- [3] 王晓原, 隗志才, 贾洪飞, 等. 交通流突变分析的变点统计方法研究[J]. 中国公路学报, 2002, 15(4): 69-74.
- [4] 张敬磊, 王晓原. 交通事件自动检测算法研究进展[J]. 武汉理工大学学报: 交通科学与工程版, 2005, 19(2): 115-118.
- [5] 王晓原, 刘海红. 基于投影寻踪自回归的短时交通流预测[J]. 系统工程, 2006, 24(3): 20-24.
- [6] 王晓原, 杨新月. 基于三次样条非参数拟合的驾驶员行为仿真模型[J]. 系统仿真学报, 2006, 18(9): 2691-2697.
- [7] 王晓原, 张敬磊, 张开旺, 等. 基于非参数样条拟合的交通流预测方法研究[J]. 计算机工程与应用, 2006, 42(26): 218-220.
- [8] 魏星, 王玲, 万建伟, 等. 基于 SIS 的不完全测量数据处理方法研究[J]. 系统仿真学报, 2006, 18(11): 3309-3312.
- [9] 姜桂艳. 道路交通状态判别技术与应用[M]. 北京: 人民交通出版社, 2004.
- [10] 赵慧, 于雷, 耿彦斌. 国内外实时交通流数据质量控制比较与分析[J]. ITS 通讯, 2005, 7(2): 15-19.
- [11] 韩卫国, 王劲峰, 胡建军. 交通流量数据缺失值的插补方法[J]. 交通与计算机, 2005, 23(1): 39-42.
- [12] 邓乃扬, 田英杰. 数据挖掘中的新方法: 支持向量机[M]. 北京: 科学出版社, 2004.
- [13] 吴今培, 孙德山. 现代数据分析[M]. 北京: 机械工业出版社, 2006.
- [14] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [15] 张朝元, 胡光华, 徐天泽. 基于 LS-SVM 的交通流量时间序列预测[J]. 云南大学学报: 自然科学版, 2004, 26(增刊): 19-22.
- [16] 赵磊, 李国和, 马现峰. 基于支持向量机的缺失数据补齐方法[J]. 计算机工程与应用, 2006, 42(36): 207-208.
- [17] 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.
- [18] Liu Haihong, Wang Xiaoyuan, Tan Derong, et al. Study on Traffic Information Fusion Algorithm Based on Support Vector Machines[C]. Sixth International Conference on Intelligent Systems Design and Applications, 2006: 183-187.