

基于在线支持向量回归算法的短时交通流预测

高学辉¹, 刘艳忠², 王巧芝¹, 贾世胜¹, 孙皓³

(1. 山东科技大学机电系, 山东泰安 271019; 2. 河南省城市交通规划研究中心, 河南郑州 450000;
3. 山东科技大学信息与电气工程学院, 山东青岛 266510)

摘要: 实时准确的短时交通流预测在城市道路交通和高速公路交通中都十分重要, 是交通控制与诱导系统的基础。应用在线支持向量回归算法对交通流进行预测, 并对济南某高架路实测数据进行仿真运算。预测结果表明, 在小样本下, 与 BP 神经网络算法相比, 在线支持向量回归算法明显优于 BP 神经网络算法, 增大样本数, BP 神经网络算法预测精度有所提高, 但仍低于线支持向量回归算法; 在运算时间上, BP 神经网络算法运算时间更短。

关键词: 在线支持向量回归; 核函数; 神经网络; 非参量回归; 短时交通流

中图分类号: U49 文献标志码: A 文章编号: 1672-3767(2011)01-0078-05

Prediction of Short-term Traffic Flow with On-line Support Vector Regression Algorithm

GAO Xuehui¹, LIU Yanzhong², WANG Qiaozhi¹, JIA Shisheng¹, SUN Hao³

(1. Department of Mechanical and Electrical Engineering, Shandong University of Science and Technology, Taian, Shandong 271019, China; 2. Henan Planning and Research Center of Urban Transportation, Zhengzhou, Henan 450000, China;
3. College of Information and Electrical Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266510, China)

Abstract: The prediction of precise real time short-term traffic flow is very important to the transportation not only in urban road nets but also in freeway system and, it is a foundation of traffic control and guide system. The short-term traffic flow (STTF) was predicted by applying the on-line support vector regression (OLSVR) algorithm and simulation was operated with real data measured in one section of Jinan elevated road in this paper. The predicted results showed that on-line support vector regression (OLSVR) algorithm was superior to the BP neural network algorithm in precision under the condition of small samples. The precision of the BP neural network algorithm was raised with the increase of the sample size, but still smaller than that with the OLSVR algorithm. The calculation time with the BP neural network algorithm is shorter than that with the OLSVR algorithm.

Key words: on-line SVR; kernel function; neural network; nonparametric regression; short-term traffic flow (STTF)

交通流控制与诱导是交通管理中的一个关键问题, 其核心之一是实时、准确的短时交通流预测。短时交通流预测一般是指时间间隔不超过 15 min 的交通流预测, 是一个复杂的、非线性的、时变的问题, 做到实时、准确的短时交通预测是十分困难的。

几十年来, 各国学者提出了各种预测方法, 包括历史均值法、Kalman 滤波法、非参数回归法^[1]、神经网络法^[2]、支持向量回归法^[3-7]和小波分析法^[8-10]等。历史均值法和 Kalman 滤波法是传统的预测算法, 对线性系统有较好的预测结果, 对非线性系统其预测精度并不突出。神经网络模型对于非线性交通流预测来说, 可以取得较好的预测结果, 但是其通用性和实时性并不能令人满意, 而且神经网络模型存在局部最小值的问题, 有时候并不能获得最优解。

Smith 等^[1]在 2002 年提出了非参量回归法来预测短时交通流, 非参量回归法的最大优点是不需建立模

收稿日期: 2010-10-09

基金项目: 山东科技大学科学研究“春蕾计划”项目(2008BZC044)。

作者简介: 高学辉(1975—), 男, 山东临朐人, 讲师, 主要从事交通流预测与控制研究。E-mail: xhgao@163.com.

型,仅仅依靠历史数据就可以对交通流进行预测,很好地解决了其他模型通用性不高的缺点。但是,非参量回归法需要的历史数据较多,其预测精度与历史数据密切相关,其核函数的选择也需要进一步的研究,并且实时性不好。

支持向量机不存在局部最小值的问题,在小样本下可以取得较高的预测精度。刘艳忠等^[11]在2008年就提出了在线学习的支持向量回归模型预测短时交通流,Manoel Castro-Neto等^[12]在2009年也提出了在线支持向量回归机预测交通状况。支持向量回归算法可以较好地解决复杂的、非线性的、时变的问题,适用于短时交通流预测问题的研究。

1 在线学习的支持向量增量回归机

支持向量机是 Vapnik 等^[13]在1979年针对分类问题提出的,后来推广到回归模型中形成支持向量回归机。支持向量回归机通过引入核函数巧妙地解决了样本维数的问题,使得算法的复杂度与样本的维数无关,从而避免了“维数灾难”。

非线性情况下, Lagrange 支持向量回归机模型可以描述如下:

$$\begin{aligned} \min \quad & \frac{1}{2} (\|w\|^2 + b^2) + \frac{C}{2} (\xi^{*T} \xi + \xi^T \xi^*) \\ \text{s. t.} \quad & Aw + be - y \leq \epsilon e + \xi \\ & y - Aw - be \leq \epsilon e + \xi^* \\ & w = A^T (\alpha^* - \alpha) \\ & b = e^T (\alpha^* - \alpha) \end{aligned} \quad (1)$$

其中: A 为 $m \times n$ 矩阵; $x_i (i=1, 2, \dots, m)$ 为其行向量; b 是阈值; e 是以 1 为分量的任意维向量; ξ, ξ^* 为允许误差量; C 是惩罚参数,或称正则化参数,体现模型复杂程度和训练误差之间的折中。

迭代的拉格朗日方程为:

$$L = \frac{1}{2} (\alpha^* - \alpha)^T (AA^T + ee^T) (\alpha^* - \alpha) - y^T (\alpha^* - \alpha) + \epsilon e^T (\alpha^* - \alpha) + \frac{1}{2C} (\alpha^{*T} \alpha^* + \alpha^T \alpha) - \delta^{*T} \alpha^* - \delta^T \alpha \quad (2)$$

其中: $\delta, \delta^* \in \mathbf{R}^m$ 是拉格朗日乘法向量。令 $H = \begin{pmatrix} A & e \\ -A & -e \end{pmatrix}, Q = \frac{I}{C} + K(H, H^T)$, 在

$$\begin{cases} \frac{\partial L}{\partial \alpha_i} = \sum_{j=1}^l Q_{ij} (\alpha_j - \alpha_j^*) + \epsilon - y_j + \xi_i - \delta_i + u_i = 0 \\ \frac{\partial L}{\partial \alpha_i^*} = - \sum_{j=1}^l Q_{ij} (\alpha_j - \alpha_j^*) + \epsilon - y_j - \xi_i - \delta_i + u_i = 0 \text{ (KKT) 条件下, 大部分 } \alpha_i^* \text{ 和 } \alpha_i \text{ 是非零的;} \\ \delta_i^* \alpha_i^* = 0 \\ u_i^* (\alpha_i^* - C) = 0 \end{cases}$$

令 $\theta_i = \alpha_i^* - \alpha_i$, 因此, θ_i 决定于 α_i^* 和 α_i 。

对每个样本 (x_i, y_i) , 差值函数 $h(x_i)$ 定义为: $h(x_i) = f(x_i) - y_i = \theta^T Ax_i + e^T - y_i$ 。

$$\begin{cases} -\epsilon \leq h(x_i) \leq \epsilon, \theta_i = 0 \\ h(x_i) < -\epsilon, \theta_i > 0 \\ h(x_i) > \epsilon, \theta_i < 0 \end{cases} \quad (3)$$

因此,最佳回归函数为

$$f(x) = K((x^T \ 1)^T, (Ae)^T (\alpha^* - \alpha)) \quad (4)$$

2 数据处理

传感器和传输线路在采集传输过程中可能会出现故障而导致数据丢失或失真,丢失或失真数据会干扰正常的预测结果,降低预测精度。为了得到更加科学的结论,应该将这些数据加以处理。交通事故也会对正常的交通流量造成干扰,从而影响处理结果。传感器和传输线路故障造成的异常数据相对于交通事故造成

的异常数据来说是相对容易判断的,这些数据要么是缺失,要么是严重脱离前后数据的孤立点,在极个别的情况下(比如数据传输线路附近一段时间突然出现极强干扰源),异常数据跟交通事故造成的异常数据极为相似,这种情况下的异常数据是相对难以判断的。

对于数据的缺失,可以用两个方法处理:一种是时间段法,一种是插值法。时间段法是将采集的时间点延长为时间段,例如传感器应在 9:00 采集一个数据,可以设定采集时间为 8:59-9:01 这一时间段,每隔 30 s 采集一个数据,将这 2 min 内采集的数据取平均值上传,如在这 4 次采集过程中,某一次的数据为零,则舍去,只计算 3 次的平均值上传。将时间点延长为时间段可以有效地预防由于传感器故障或者车辆太多造成的采集出错的情况,但是不能解决传输故障。

插值法是常用的数据处理方法,对于缺失的数据可以通过时间间隔来判断,发现数据缺失后,通过插值法将数据补齐,本文采用的插值公式如下:

$$D_{t_k} = \frac{D_{t_{(k-1)}}}{2!} + \frac{D_{t_{(k-2)}}}{3!} + \frac{D_{t_{(k-3)}}}{4!} \quad (5)$$

插值法可以补足任何故障下产生的数据缺失,但是补足的数据不是真实值,如果数据缺失量大,不宜采用插值法。

交通事故异常数据的处理有多种方法,为保证数据的一般性,本文采用的样本中不包括交通事故数据,因此未做交通事故异常数据处理。

3 实例

图 1 是济南市某高架路段一周的车流量分布图,采样时间从周一 0:00 点开始每隔 10 min 采样 1 次,数据已经经过处理,从图上可以明显看出车流量具有周期性,周一到周五的流量图基本一致,周六和周日的流量图基本一致。

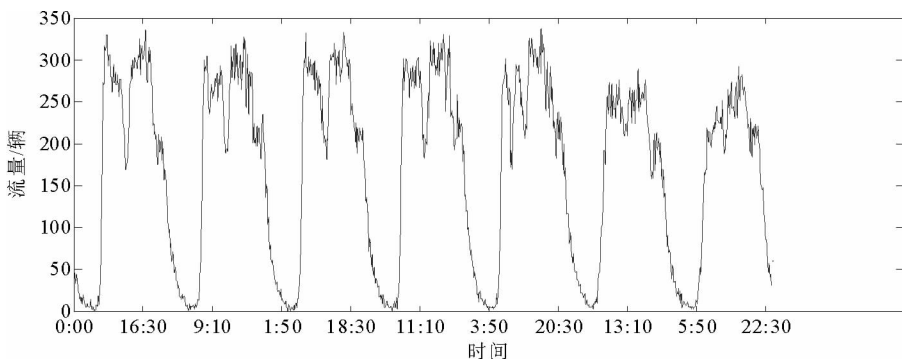


图 1 一周流量分布图

Fig. 1 The distribution of traffic flow in one week

采用支持向量回归算法,核函数采用高斯核函数 $K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$, BP 神经网络算法采用单隐层结构,输入层、隐含层、输出层的节点数分别为 1, 20, 1, 应用 Matlab 进行仿真。

选取周一 8:00 到 20:00 数据作为样本数据建立模型,周二 8:00 到 20:00 数据作为精确值与预测结果对比,如图 2、图 3、表 1 所示。

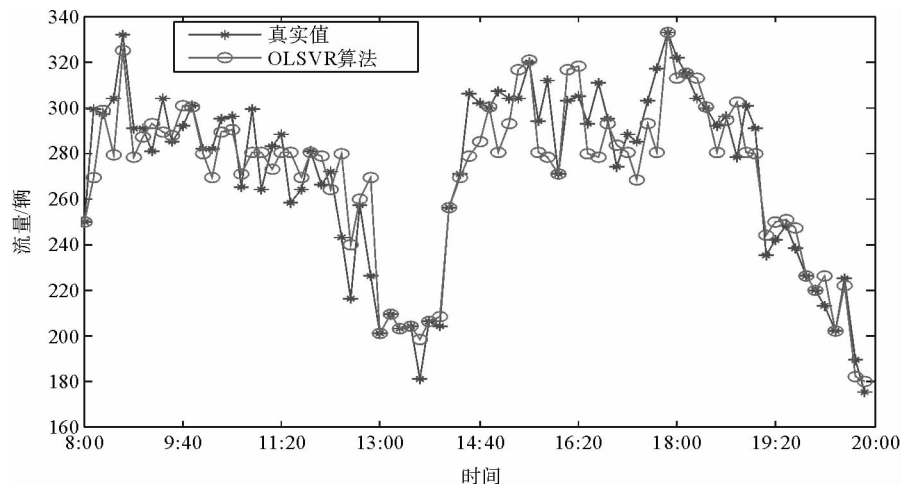


图 2 8:00—20:00 样本数据和 OLSVR 算法预测数据比较图

Fig. 2 The comparison between actual values and predicted ones with OLSVR algorithm at 8:00—20:00

通过图 2、图 3 和表 1 可以看出,在小样本下,在线支持向量回归算法的精确性明显优于 BP 神经网络算法,在图 2 中在线支持向量回归

算法共有 17 个点的预测值与实际值相等,总体走势基本与实际值一致,曲线拟合度为 92.16%;而 BP 神经网络算法仅有 9 个点的预测值与实际值相等,总体走势与实际走势有区别,曲线拟合度也只有 84.68%,但是 BP 神经网络算法运算时间为 3.56 s,小于在线支持向量回归算法的 9.44 s。

表 1 小样本下两种方法性能指标评价表

Tab. 1 The evaluation of performance indices for two algorithms with small samples

方法	BP 神经网络	支持向量回归机
误差/%	14.89	6.67
拟合度	0.846 8	0.921 6
运算时间/s	3.56	9.44

增加样本数,取周一 0:00 到周二 24:00 时间内的数据作为样本数据,周三,周四数据作为实际值与预测值比较。图 4 和图 5 分别是周三 0:00 到 24:00 实际值与预测值的比较图,表 2 是指标评价,从图上可以看出,BP 神经网络的曲线拟合度仍不如在线支持向量回归算法,图的上部大部分表现为平均值,虽然总体误差减小,但是不能反映短时交通流实际情况,在线支持向量回归算法则基本反映了短时交通流实际情况。从表 2 也可以明显看出在线支持向量回归算法曲线拟合度好于 BP 神经网络算法,同样在线支持向量回归算法所用时间远远超过 BP 神经网络算法。

表 2 增加样本下两种方法性能指标评价表

Tab. 2 The evaluation of performance indices of two algorithms with increased samples

方法	BP 神经网络	支持向量回归机
误差/%	6.36	5.17
拟合度	0.923 1	0.965 2
运算时间/s	7.59	19.47

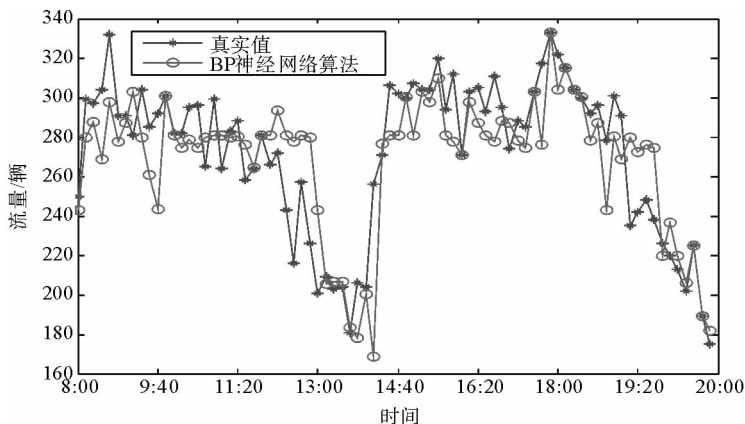


图 3 8:00—20:00 样本数据和神经网络算法预测数据比较图
Fig. 3 The comparison between actual values and predicted ones with neural networks algorithm at 8:00—20:00

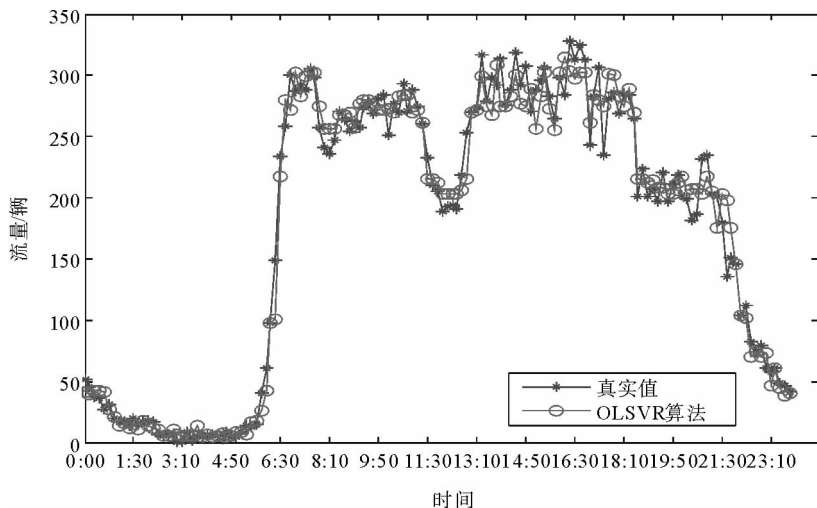


图 4 0:00—24:00 样本数据和 OLSVR 算法预测数据比较图
Fig. 4 The comparison between actual values and predicted ones with OLSVR algorithm at 0:00—24:00

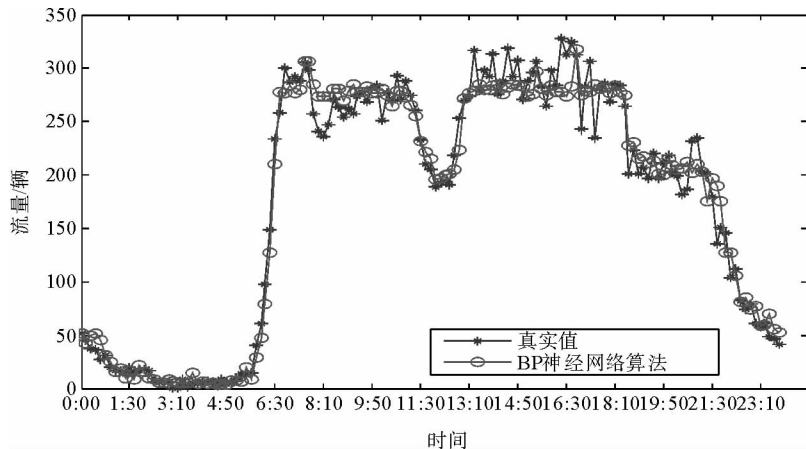


图 5 0:00—24:00 样本数据和神经网络算法预测数据比较图
Fig. 5 The comparison between actual values and predicted ones with neural networks algorithm at 0:00—24:00

4 结论

应用在线支持向量回归算法预测短时交通流数据,与BP神经网络算法相比,在不同样本条件下,不论是预测准确性还是曲线拟合度上,在线支持向量回归算法好于神经网络算法。支持向量在线回归算法与样本数无关,因此可以在较低样本数下得到较好的结果,而BP神经网络算法在小样本下效果不好,随着样本数增多,BP神经网络算法容易陷入局部最小值而使其曲线拟合度不高,在线支持向量回归算法则不存在此问题。

在线支持向量回归算法虽然有以上优点,但是其算法复杂度较高,计算所需时间要比BP神经网络算法所需时间长。因此,对于在线支持向量回归算法,仍需继续改进,使其运算时间进一步缩短,以适应交通流预测的实时性要求。

参考文献:

- [1] SMITH B L, WILLAIMS B M, OSWALD R K. Comparison of parametric and nonparametric models for traffic flow forecasting[J]. *Transportation Research Part C*, 2003, 10(4): 302-321.
- [2] 杜长海, 黄席樾. 基于神经网络和 Markov 链的交通流实时滚动预测[J]. *系统仿真学报*, 2008, 20(9): 2464-2468.
DU Changhai, HUANG Xiyue. Real-time rolling traffic flow forecasting based on neural networks and Markov chains[J]. *Journal of System Simulation Research*, 2008, 20(9): 2464-2468.
- [3] 张新, 潘美芹, 邵福波, 等. 支持向量回归机训练集的并行预处理方法[J]. *山东科技大学学报: 自然科学版*, 2009, 28(5): 85-89.
ZHANG Xin, PAN Meiqin, SHAO Fubo, et al. Parallel pretreatment method for training set of support vector regression machine[J]. *Journal of Shandong University of Science and Technology: Natural Science*, 2009, 28(5): 85-89.
- [4] 胡运红, 董玉林. 支持向量机简化算法中支持向量与违背对的几何意义[J]. *山东科技大学学报: 自然科学版*, 2010, 29(1): 95-99.
HU Yunhong, DONG Yulin. Geometrical meaning on support vectors and violating pairs for the simplified algorithms of SVM[J]. *Journal of Shandong University of Science and Technology: Natural Science*, 2010, 29(1): 95-99.
- [5] CHANG C C, LIN C J. Training V-support vector regression theory and algorithms[J]. *Neural Computation*, 2002, 14(8): 1959-1977.
- [6] MA J, JAMES T, SIMON P. Accurate online support vector regression[J]. *Neural Computation*, 2003, 15(11): 2683-2703.
- [7] 李斌, 郝涛, 史明华. 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.
LI Bin, XI Tao, SHI Minghua. Traffic flow combined forecast model of support vector of machine[J]. *Journal of Tianjin Polytechnic University*, 2008, 27(2): 73-76.
- [8] 高勇, 陈锋. 基于小波分析的短时交通流非参数回归预测[J]. *中国科学技术大学学报*, 2008, 38(12): 1427-1431.
GAO Yong, CHEN Feng. Wavelet analysis-based NPR prediction of short-term traffic flow[J]. *Journal of University of Science and Technology of China*, 2008, 38(12): 1427-1431.
- [9] 高学辉, 孙皓, 刘瑞国, 等. 遗传算法在高速公路路经计算中的应用[J]. *武汉理工大学学报: 交通科学与工程版*, 2007, 31(5): 854-856.
GAO Xuehui, SUN Hao, LIU Ruiguang, et al. Application of genetic algorithm in expressway's path[J]. *Journal of Wuhan University of Technology: Transportation Science & Engineering*, 2007, 31(5): 854-856.
- [10] 高学辉, 王刚, 刘艳忠, 等. 基于移动瓶颈理论的高速公路重载货车影响效应研究[J]. *山东科技大学学报: 自然科学版*, 2010, 29(3): 100-104.
GAO Xuehui, WANG Gang, LIU Yanzhong, et al. Study on the effect of full-load lorry on freeway based on moving bottleneck theory[J]. *Journal of Shandong University of Science and Technology: Natural Science*, 2010, 29(3): 100-104.
- [11] LIU Y Z, SHAO X J, LI X H, et al. The study on a real-time forecasting model for short-term traffic flow based on online incremental LSVR[C]//6th International Conference on Traffic and Transportation Studies, Nanning, August, 2008: 852-861.
- [12] MANOEL C N, JEONG Y S, JEONG M K, et al. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions[J]. *Expert Systems with Applications*, 2009, 36(3): 6164-6173.
- [13] CORTES C, VAPNIK V. Support vector network[J]. *Machine Learning*, 1995, 20: 273-297.