

文章编号:1673-0291(2006)03-0019-04

基于支持向量回归机的 交通状态短时预测方法研究

姚智胜,邵春福,高永亮

(北京交通大学 交通运输学院,北京 100044)

摘要:提出基于支持向量回归机的交通状态短时预测方法.具体的做法是,以交通检测器收集到某时刻前几时段及上下游前几时段的交通流量、占有率、平均速度等交通参数为输入,以对应时段交通流量为输出,选取核函数,对支持向量回归机进行训练.应用训练完成的支持向量回归机,输入交通流量、占有率、平均速度,来预测下时段的交通流量.最后,以某城市道路的实时数据来对模型进行验证,预测结果表明了模型的有效性.

关键词:交通流短时预测;支持向量回归机;统计学习;人工智能

中图分类号:U491.112

文献标识码:A

Research on Methods of Short-Term Traffic Forecasting Based on Support Vector Regression

YAO Zhi-sheng, SHAO Chun-fu, GAO Yong-liang

(School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China)

Abstract: The paper proposes short-term traffic forecasting model based on support vector regression. First, the traffic volumes, occupancy-rate, average velocity at several preceding periods of time and upstream and downstream collected by RTMS are considered as input, traffic volumes at current period of time are considered as output. Second, the support vector regression is trained after selecting a kernel function. Finally, the traffic volumes being forecasted at several periods of time in the future are available by inputting the traffic volumes, occupancy-rate and average velocity necessary to the trained support vector regression. The paper also uses the real time data of certain urban road to test the efficiency of the proposed model and the result is satisfied.

Key words: short-term traffic flow forecasting; support vector regression; statistical learning; artificial intelligence

智能交通系统被认为是解决道路交通阻塞、减少汽车排气污染和交通事故等问题的有效方法之一^[1]. 交通控制与诱导系统是智能交通系统的组成部分,交通状态的短时预测是交通控制与诱导的基础,其研究受到广泛关注. 迄今,交通短时预测方法主要有统计回归法、状态估计法、时间序列法、神经

网络法、动态交通分配及交通模拟法等^[2-5]. 近年来又有专家不断提出新的算法,如尖点突变理论、状态空间模型、Kalman 滤波、改进的神经网络、非参数回归、小波、分形、混沌等非线性系统理论的方法^[6-10].

城市道路交通系统是一个时变、非线性、高维、非平稳、带有随机性的系统. 虽然神经网络能够根据

收稿日期:2005-05-27

基金项目:国家“十五”攻关计划资助项目(2005BA414B02);国家自然科学基金资助项目(50578009)

作者简介:姚智胜(1979—),男,福建泉州人,博士生. email: yzhisheng@163.com

邵春福(1957—),男,河北沧州人,教授,博士,博士生导师.

历史数据进行学习训练和经验积累,具有自适应能力的优点.但是它得到的结果是基于经验风险最小化的,需要有足够大的样本数据数量,可能还会出现学习问题以及求得局部极小解的问题.支持向量机(Support Vector Machine, SVM)以结构风险最小化为目标,以统计学习理论为理论基础,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.已有一些专家和学者将其应用在交通流量时间序列预测中,例如文献[11]提出基于统计学习理论的交通流量时间序列预测方法^[11],文献[12]探讨了支持向量机在交通流量预测中的应用^[12].本文作者综合考虑交通状态的影响因素,利用已有交通状态数据,提出基于支持向量回归机的交通状态短时预测方法.

1 预测模型

1.1 支持向量回归机的选取

支持向量机是借助于最优化方法解决机器学习问题的新工具.它以统计学习为理论基础,最初于20世纪90年代由Vapnick提出,近年来在其理论研究和算法实现都取得了突破性进展,开始成为克服“非线性和维数灾难问题”、“过学习问题”和“局部极小点问题”等传统困难的有力手段^[13].

支持向量回归机是支持向量机在回归估计问题中的应用.支持向量回归机主要有 ϵ -支持向量回归机、 ν -支持向量回归机和最小二乘支持向量回归机等. ϵ -支持向量回归机、 ν -支持向量回归机所选取的损失函数都是 ϵ -不敏感损失函数,针对实际问题也可选择其它形式的损失函数^[13].当选择不同的损失函数后,所要求解的最优化问题的目标函数和约束条件就发生了改变.

1.2 交通状态短时预测

交通状态预测指的是根据过去若干个时段的交通状态的时间、空间变化情况,结合其它影响因素,预测未来若干个时段内的交通状态估计值.表征交通状态的基本参数,如交通流量、平均速度、交通流密度和车道占有率,可以通过车辆检测器获得.这些数据中,交通流时间变化的数据主要有本路段过去若干个时段的交通参数及历史平均值等.反映交通流空间变化主要有本路段上下游当前及过去各时段的交通参数等.交通状态预测按照时段可分为日交通预测、小时交通预测、短时交通预测等.一般认为预测时段小于15 min的为短时交通预测.

1.3 预测模型

由于交通状态变化过程是一个实时、非线性、高

维、非平稳随机过程,随着统计时段的缩短,交通状态变化的随机性和不确定性越来越强.交通状态短时变化不仅跟本路段过去几个时段的交通状态有关,还受上下游交通状态的影响.交通状态短时预测要解决的问题就是如何从带有随机性和不确定性的交通状态变化中,根据车辆检测器得到的交通状态参数,结合其它影响因素,找出其中的规律性,建立预测模型,以预测未来几个时段的交通状态变化.

根据支持向量回归机的算法(以 ν -支持向量回归机算法^[13]为例)及交通状态短时预测的基本原理,我们提出如下基于支持向量回归机的预测模型:

(1)设当前时段为 t ,对路段 i 的 $t+1$ 时段的交通参数(以交通流量为例)进行预测.路段 i 共有相邻上下游路段 m 个,分别记为路段 $i+j$, $j=1,2,3,\dots,m$.路段 i 在 t 时段统计的交通流量、平均速度和车道占有率分别记为 $q_i(t)$ 、 $v_i(t)$ 、 $o_i(t)$.

(2)选定回溯系数 p ,构造训练集

$\{(x_1, y_1), \dots, (x_l, y_l)\}$,其中: y_1, y_2, \dots, y_l 为路段 i 过去几个时段已知交通流量 $q_i(t-l+1)$, $q_i(t-l+2), \dots, q_i(t)$, l 为训练集的数据组数. x_1, x_2, \dots, x_l 为过去几个时段本路段及上下游路段的交通状态参数的向量组合.其中, $x_i \in X = \mathbf{R}^{3(m+1)p}$,例如对应于 $q_i(t-l+1)$ 输出的输入为 x_1 如下所示向量:

$$\begin{aligned} &\{q_i(t-l-p+1), \dots, q_i(t-l)\}, \dots, \\ &\{q_{i+m}(t-l-p+1), \dots, q_{i+m}(t-l)\}; \\ &\{v_i(t-l-p+1), \dots, v_i(t-l)\}, \dots, \\ &\{v_{i+m}(t-l-p+1), \dots, v_{i+m}(t-l)\}; \\ &\{o_i(t-l-p+1), \dots, o_i(t-l)\}, \dots, \\ &\{o_{i+m}(t-l-p+1), \dots, o_{i+m}(t-l)\} \end{aligned}$$

组成的 $3(m+1)p$ 维向量.

(3)选择适当的正数 ν 、 C 和核函数 $K(x, x')$,构造并求解最优化问题

$$\begin{aligned} \max_{\alpha^{(*)} \in \mathbf{R}^{2l}} W(\alpha^{(*)}) &= \sum_{i=1}^l (\alpha_i^* - \alpha_i) y_i - \\ &\frac{1}{2} \sum_{j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(x_i, x_j), \\ \text{s.t. } &\begin{cases} \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, \\ \alpha_i^{(*)} \in [0, \frac{C}{l}], \quad i = 1, 2, 3, \dots, l, \\ \sum_{i=1}^l (\alpha_i + \alpha_i^*) \leq C \cdot \nu, \end{cases} \end{aligned}$$

得到最优解 $\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)$,相当于从

训练集中选出了支持向量.

(4)构造预测函数

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x_i, x) b^*.$$

其中, b^* 按下列方式计算.

选择 $\bar{\alpha}^{(*)}$ 的位于开区间 $(0, \frac{C}{l})$ 中的两个分量 $\bar{\alpha}_j$ 和 $\bar{\alpha}_k^*$, 令

$$b^* = \frac{1}{2} [y_j + y_k - (\sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_j) + \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_k))].$$

(5)最后,输入当前时段及过去几个时段本路段及上下游路段的交通参数向量组合 x , 利用预测函数 $f(x)$ 就可以预测出下时段的交通流量.

上述基于支持向量回归机的交通状态短时预测模型的结构图如图 1 所示.

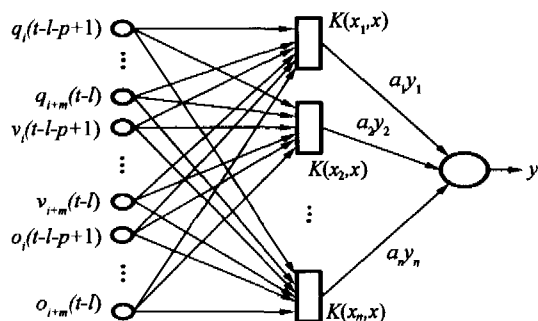


图 1 预测模型结构图

Fig.1 Chart of model structure

在实际应用模型的过程中,根据具体问题和所获得的资料情况,也可有选择地增加或减少输入向量的维数,例如,对于训练集的输入增加反映环境变化的参数向量、忽略平均速度或车道占有率对输出的影响等.

2 实例分析

实例分析采用远程交通微波检测器(RTMS)采集的城市道路的流量、速度、车道占有率数据,数据的统计间隔为 2 min. 为了验证上述模型的有效性,选定以 2 min 为一个时段,采用平峰时间段的数据进行分析,即利用周一到周四 9 点到 15 点以及周五 9 点到 12 点的数据构造训练集对支持向量回归机进行训练,预测周五 13 点到 15 点的流量,并与实际观测值相比较.

预测时采用 ν -支持向量回归机,通过已知的数据进行分组交叉验证试验,即把已知数据分成几组,以其中一组为预测对象,其余的数据组用作训练集,

根据预测结果的好坏来选择参数. 通过实验,最后选择参数 $\nu = 0.2, C = 100$,选用的核函数为 Gauss 径向基(RBF)核函数(其参数 σ 取 20). 计算时利用 Chih-Jen Lin 开发的 LIBSVM2.6 软件包. 构造训练集时以 $t-1$ 时段的流量 $q_i(t-1)$ 、速度 $v_i(t-1)$ 、车道占有率 $o_i(t-1)$ 及 $t-2$ 时段的流量 $q_i(t-2)$ 、速度 $v_i(t-2)$ 、车道占有率 $o_i(t-2)$ 为输入属性值,以 t 时段的流量 $q_i(t)$ 为输出属性值. 周五 13 点到 15 点共有 90 个时段,从 13:00 到 13:02 时段开始到 15:58 到 16:00 时段分别标号为 1, ..., 90 时段. 预测值和观测值的对比如图 2 所示.

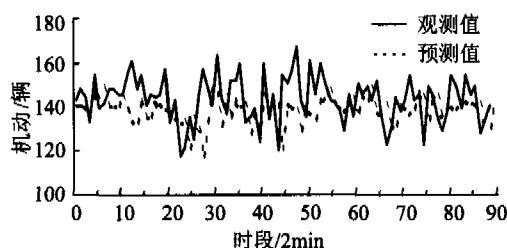


图 2 观测值与预测值对比示意图

Fig.2 Chart of comparing observed values with forecasted values

在 90 个交通流量预测值中,与实际观测值的平均绝对误差约为 9.9 辆,平均相对误差约为 6.84%. 其中,最大偏差出现在 13:54 到 13:56 时段,为 40 辆,相对误差达到 25.43%. 这是因为本次实例分析只考虑了过去两个时段的交通参数对模型输出的影响,忽略了过去第 3 个时段以上交通参数的影响,而该时段的前两个时段交通流量都比较低. 根据模型训练的结果,按照输入数据预测也应该比较低,但实际流量却比较高,因而出现了较大偏差. 解决的办法是训练模型时考虑更长时段的交通参数对模型输出的影响. 为了对比,我们还用同样的数据运用神经网络模型中应用最广泛的 BP 神经网络进行了同样的预测,神经网络采用四层结构,输入层有 6 个输入,输出层有 1 个输出,中间两层有 11 个神经元,性能函数为均方差误差性能函数,利用训练数据经过 20 000 次训练后,BP 神经网络预测结果的平均绝对误差约为 10.81 辆,平均相对误差约为 7.39%,从计算的两个误差指标来看均不如上面的计算结果. 实例分析的结果表明,上述模型对于随机性、不确定性较强的 2 min 时段的交通短时预测有相对比较满意的预测结果.

3 结语

本文作者利用支持向量回归机和交通状态短时预测的基本原理,综合考虑交通状态影响因素,利用

交通状态数据,建立基于支持向量回归机的交通状态短时预测模型,探讨了支持向量机在交通状态短时预测中的应用,并以道路的实时数据对模型进行验证.基于支持向量回归机的交通状态短时预测模型考虑了上下游交通状态的相互影响,是无模型的因果预测算法,预测结果表明了模型的有效性.

但是该模型还只是基于车辆检测器的单点预测方法,而在智能交通系统中进行控制和诱导需要的是道路网的交通状态准确而又快速的短时预测,因此怎样把该模型在保持预测精度又占用较少计算资源的条件下扩充为多点预测方法,使得该方法能够满足实际应用是下一步的研究内容.另外,该模型的参数优化、预测精度的提高、模型应用规模等方面还需要进行更加深入细致探讨和研究.

参考文献:

- [1] 邵春福,张魁麟,谷远利. 基于实时数据的网状城市快速路行驶时间预测方法研究[J]. 土木工程学报, 2003, 36(1): 16-20.
SHAO Chun-fu, ZHANG Kui-lin, GU Yuan-li. A Study of Route Travel Time Forecast Method Based on Real Data of Urban Expressway Network[J]. China Civil Engineering Journal, 2003, 36(1): 16-20. (in Chinese)
- [2] 王正武,黄中祥. 短时交通流预测模型的分析与评价[J]. 系统工程, 2003, 21(6): 97-100.
WANG Zheng-wu, HUANG Zhong-xiang. An Analysis and Discussion on Short-Term Traffic Flow Forecasting[J]. Systems Engineering, 2003, 21(6): 97-100. (in Chinese)
- [3] 贺国光,李宇,马寿峰. 基于数学模型的短时交通流预测方法探讨[J]. 系统工程理论与实践, 2000, 20(12): 51-56.
HE Guo-guang, LI Yu, MA Shou-fen. Discussion on Short-Term Traffic Flow Forecasting Methods Based on Mathematical Model[J]. Systems Engineering—Theory & Practice, 2000, 20(12): 51-56. (in Chinese)
- [4] Bart Van Arem, Howard R Kirby, Martie J, et al. Recent Advances and Application in the Field of Short-Term Traffic Forecasting[J]. International Journal of Forecasting, 1997, 13: 1-12.
- [5] Chrobok R, Kaumann O, Wahle J, ed al. Different Methods of Traffic Forecast Based on Real Data[J]. European Journal of Operational Research, 2004, 155: 558-568.
- [6] 宗春光,宋靖雁,任江涛,等. 基于相空间重构的短时交通流预测研究[J]. 公路交通科技, 2003, 20(4): 71-75.
ZONG Chun-guang, SONG Jing-yan, REN Jiang-tao, et al. Short-Term Traffic Flow Forecasting Research Based on Phase Space Reconstruction[J]. Journal of Highway and Transportation Research and Development, 2003, 20(4): 71-75. (in Chinese)
- [7] 贺国光,马寿峰,李宇. 基于小波分解与重构的交通流短时预测法[J]. 系统工程理论与实践, 2002, 22(9): 101-106.
HE Guo-guang, MA Shou-fen, LI Yu. Study on the Short-Term Forecasting for Traffic Flow Based on Wavelet Analysis[J]. Systems Engineering—Theory & Practice, 2002, 22(9): 101-106.
- [8] 张亚平,张起森. 尖点突变理论在交通流预测中的应用[J]. 系统工程学报, 2000, 15(3): 272-276.
ZHANG Ya-ping, ZHANG Qi-sen. The Application of Cusp Catastrophe Theory in The Traffic Flow Forecast[J]. Journal of Systems Engineering, 2000, 15(3): 272-276. (in Chinese)
- [9] 王宏杰,林良明,徐大淦,等. 基于改进 BP 网交通流动态时序预测算法的研究[J]. 交通与计算机, 2001, 19(3): 11-14.
WANG Hong-jie, LIN Liang-ming, XU Da-gan, et al. Research of Traffic Flow Dynamic Time Series Forecasting Algorithm Based on Improved BP Neural Network[J]. Computer and Communications, 2001, 19(3): 11-14. (in Chinese)
- [10] 宫晓燕,汤淑明. 基于非参数回归的短时交通流量预测与事件检测综合算法[J]. 中国公路学报, 2003, 16(1): 82-86.
GONG Xiao-yan, TANG Shu-ming. Integrated Traffic Flow Forecasting and Traffic Incident Detection Algorithm Based on Non-Parametric Regression[J]. China Journal of Highway and Transport, 2003, 16(1): 82-86. (in Chinese)
- [11] 丁爱玲. 基于统计学习理论的交通流量时间序列预测[J]. 交通与计算机, 2002, 20(2): 27-30.
DING Ai-ling. Traffic Flow Time Series Prediction Based on Statistics Learning Theory[J]. Computer and Communications, 2002, 20(2): 27-30. (in Chinese)
- [12] 王继生,高宝成,时良平. 支持向量机在交通量预测中的应用[J]. 信息技术, 2004, 28(4): 8-10.
WANG Ji-sheng, GAO Bao-cheng, SHI Liang-ping. Application of Support Vector Machines in Traffic Volume Forecast[J]. Information Technology, 2004, 28(4): 8-10. (in Chinese)
- [13] 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: 224-274.
DENG Nai-yang, TIAN Ying-jie. New Method of Data Mining—Support Vector Machines[M]. Beijing: Science Press, 2004: 224-274. (in Chinese)