

# 基于支持向量机回归的港口吞吐量非线性组合预测

武 骁 宗蓓华

(上海海事大学交通运输学院,上海 200135)

**摘要** 提出了一种基于支持向量机回归算法的港口吞吐量非线性组合建模预测方法,并运用该方法进行了港口吞吐量预测,同时将该预测结果与其他方法的预测结果进行了比较.结果表明,该方法具有很强的学习及泛化能力,在处理具有一定程度的不确定性的非线性系统的组合建模预测问题时具有很好的应用价值.

**关键词** 港口吞吐量;非线性组合预测;支持向量机回归

**中图分类号** :U652.1 **文献标识码** :A **文章编号** :1000-198X(2005)04-0482-03

影响港口吞吐量的因素很多,如港口自然条件、航运市场状况、世界经济形势和国家政策等.其中有一些因素是不确定的,有一些因素是随时间变化的.这给港口吞吐量的定量预测造成了困难.目前,预测港口吞吐量的方法很多,如回归分析法、灰色预测法、RBF 神经网络模型、组合预测等,但各种方法的预测结果差异较大.这从另一个侧面反映了港口吞吐量预测的复杂性.鉴于此,本文从非线性组合预测角度研究港口吞吐量预测问题并提出了一种基于支持向量机回归的港口吞吐量非线性组合预测方法.

## 1 支持向量机回归

支持向量机是 Vapnik 等<sup>[1,2]</sup>提出的一种基于统计学习和结构风险最小化原理的新型机器学习算法.与神经网络相比,支持向量机的训练算法不存在局部极小和维数灾难问题,具有自动设计模型复杂度和泛化能力强等优点,已被成功地用于优化控制<sup>[3]</sup>、模式识别<sup>[4]</sup>和金融预测<sup>[5,6]</sup>等领域.因此,本文将支持向量机回归算法用于港口吞吐量非线性组合预测.

### 1.1 支持向量机回归算法

设采样所得样本集为

$$X = \{(x_i, y_i) | i = 1, 2, \dots, m\} \quad (x_i \in \mathbf{R}^n, y_i \in \mathbf{R})$$

$x$  和  $y$  存在函数依赖关系

$$F = \{f | f: \mathbf{R}^n \rightarrow \mathbf{R}\}$$

函数回归问题可以描述为如何寻找一个函数  $f \in F$ , 使泛函

$$R(f) = \int L(y - f(x), x) dP(x, y)$$

最小. 其中  $L(y - f(x), x)$  为损失函数, 表示  $y$  和  $f(x)$  之间的偏差.

在支持向量机的回归算法中, 常用的线性不敏感损失函数定义为

$$L(y - f(x), x) = \begin{cases} 0 & (|y - f(x, \omega)| \leq \epsilon) \\ |y - f(x, \omega)| - \epsilon & (|y - f(x, \omega)| > \epsilon) \end{cases}$$

式中:  $\omega$ ——待定系数;  $\epsilon$ ——不敏感值,  $\epsilon > 0$ .

若样本点之间呈线性关系, 则回归关系为  $f(x) = \omega x + b$  ( $b$  为偏置量). 但由于大多数经济系统中的样本点之间呈非线性关系, 支持向量机回归算法是通过非线性函数  $\phi$  将每个样本点映射到高维特征空间( Hilbert 空间), 并在高维特征空间进行线性回归的, 因而可将非线性函数估计问题转化为高维特征空间中的线性函

数估计问题 即

$$f(x, \omega) = \omega \Phi(x) + b$$

根据 Vapnik 的结构风险最小化原则<sup>[17]</sup> ,引入结构风险函数  $R(\omega)$  ,函数  $f$  应使  $R(\omega)$  最小 ,即优化问题

$$\min R(\omega) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n L(y_i - f(x_i), x_i)$$

式中 : $\|\omega\|^2$ ——描述函数  $f$  复杂度的项 ; $C$ ——给定的正则化参数 ,其作用是在经验风险和模型复杂度之间取折中 .为了使上述优化问题有解 ,引入松弛变量  $\xi_i, \xi_i^*$  ,则优化问题转化为

$$\begin{aligned} \min R(\omega) &= \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s.t.} \quad &\begin{cases} f(x_i) - y_i \leq \xi_i + \epsilon \\ y_i - f(x_i) \leq \xi_i + \epsilon \end{cases} \quad (i = 1, 2, \dots, n) \\ &\xi_i, \xi_i^* \geq 0 \end{aligned}$$

将以上问题的求解转化为其对偶问题 ,得

$$\begin{aligned} \max \quad & -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^n \alpha_i (\epsilon - y_i) - \sum_{i=1}^n \alpha_i^* (\epsilon + y_i) \\ \text{s.t.} \quad &\begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \\ &K(x_i, x_j) = \Phi(x_i) \Phi(x_j) \end{aligned}$$

式中 : $K(x_i, x_j)$ ——核函数 ,是满足 Mercer 条件的任意对称函数 ; $\alpha_i, \alpha_i^*$ ——对偶变量 .

这样 ,支持向量机回归建模就归结为一个二次规划问题 ,求解上述二次规划问题就得到  $\alpha_i$  和  $\alpha_i^*$  的全局最优解 ,从而可以得到  $\omega = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \Phi(x_i)$  . 可以通过边界上的任一点来计算偏置量  $b$ <sup>[8]</sup> . 但为了提高偏置量计算的稳定性 ,使用边界点上的平均值 ,即

$$b = \text{mean} \left\{ \delta_k + y_k - \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x_k) \right\}$$

$\delta_k$  为预测误差 .

因此 , $f(x)$  可表示为

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

常用的核函数有 :多项式核函数  $K(x_i, x_j) = (x_i x_j + 1)^d$  ( $d = 1, 2, \dots, n$ ) ;径向基函数核函数 (RBF)  $K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$  ;感知器核函数 (MLP)  $K(x_i, x_j) = \tanh(\beta x_i x_j + b)$  .

1.2 基于支持向量机回归的港口吞吐量非线性组合预测原理

基于支持向量机回归的港口吞吐量非线性组合预测原理是 ,将各参加组合的吞吐量预测结果作为支持向量机的输入向量 ,将代表相应年份的吞吐量实际值作为支持向量机的输出 ,然后用一定的预测样本训练这个系统 ,使不同的输入向量得到相应的输出值 ,从而在各参加组合的港口吞吐量预测模型结果与吞吐量实际值之间建立一种非线性映射关系 ,经过学习及测试达到较高精度后 ,该系统便可作为港口吞吐量非线性预测的有效工具 .系统结构如图 1 所示 .

2 港口吞吐量的组合预测实例

为了验证基于支持向量机回归的非线性组合预测方法的有效性 ,本文以文献 9 中的吞吐量预测实例来说明该方法的应用效果 .文献 9 以线性回归模型、灰色预测模型和 RBF 神经网络模型作为组合项 ,利用 1978 ~ 1998 年的年度数据建立不变权组合预测模型并进行了外推预测及检验 .为了便于与文献 9 所采用的预测方法相比较 ,本文采用了与文献 9 相同的

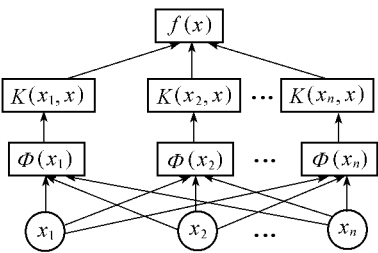


图 1 支持向量机回归结构  
Fig.1 Structure of support vector machine regression

各单项模型和数据,根据支持向量机回归算法进行港口吞吐量非线性组合预测(所采用的核函数为径向基函数,损失函数为线性不敏感损失函数)的效果及评价如表1所示。

表1 预测效果评价  
Table 1 Evaluation of effect of prediction

| 年份   | 实际<br>吞吐量<br>/万 t | 线性回归<br>预测模型  |                  |                 | 灰色预测模型        |                  |                 | RBF 神经网络<br>预测模型 |                  |                 | 不变权组合<br>预测模型 |                  |                 | 基于支持向量机回归的<br>非线性组合预测模型 |                  |                 |
|------|-------------------|---------------|------------------|-----------------|---------------|------------------|-----------------|------------------|------------------|-----------------|---------------|------------------|-----------------|-------------------------|------------------|-----------------|
|      |                   | 预测吞吐<br>量/万 t | $E_{MA}/$<br>万 t | $E_{MAP}/$<br>% | 预测吞吐<br>量/万 t | $E_{MA}/$<br>万 t | $E_{MAP}/$<br>% | 预测吞吐<br>量/万 t    | $E_{MA}/$<br>万 t | $E_{MAP}/$<br>% | 预测吞吐<br>量/万 t | $E_{MA}/$<br>万 t | $E_{MAP}/$<br>% | 预测吞吐<br>量/万 t           | $E_{MA}/$<br>万 t | $E_{MAP}/$<br>% |
| 1995 | 5 786.65          | 5 449.49      |                  |                 | 5 834.03      |                  |                 | 5 632.20         |                  |                 | 5 654.14      |                  |                 | 5 840.37                |                  |                 |
| 1996 | 6 188.33          | 6 141.85      | 198.90           | 3.16            | 6 069.41      | 142.36           | 2.17            | 6 428.25         | 115.69           | 1.89            | 6 263.64      | 121.58           | 1.90            | 6 374.34                | 77.24            | 1.24            |
| 1997 | 6 789.30          | 6 591.95      |                  |                 | 6 485.33      |                  |                 | 6 725.53         |                  |                 | 6 605.19      |                  |                 | 6 731.26                |                  |                 |
| 1998 | 6 818.40          | 7 033.02      |                  |                 | 6 917.57      |                  |                 | 6 823.00         |                  |                 | 6 912.77      |                  |                 | 6 829.59                |                  |                 |

从表1预测效果评价(平均绝对误差  $E_{MA}$  和绝对平均百分比误差  $E_{MAP}$ )可以看出,基于支持向量机回归的非线性组合预测取得了较好的预测效果,其评价指标优于文献[9]提出的不变权组合预测和各单项预测方法,这验证了基于支持向量机回归的非线性组合预测方法的有效性。

3 结 束 语

本文从非线性组合预测角度出发,运用基于支持向量机回归的非线性组合预测方法进行了港口吞吐量预测,并通过实例计算进行了对比。结果表明,本文所提出的方法在处理诸如港口吞吐量预测这种具有一定程度的不确定性系统的组合建模与预测问题时具有广泛的应用价值,对促进预测理论的进一步研究和发

展具有一定意义。

参考文献:

[1] VAPNIK V. The nature of statistical learning theory[M]. New York: Springer Verlag, 1995. 46—71.

[2] CORTES C, VAPNIK V. Support vector networks[J]. Machine Learning, 1995, 20(3): 273—297.

[3] SUYKENS J A K, VANDEWALLE J, DE MOOR B. Optimal control by least support vector machines[J]. Neural Networks, 2001, 14(1): 23—35.

[4] BUREGES C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121—167.

[5] VAN GESTEL T, SUYKENS J A K, BAESTAENS D E, et al. Financial time series prediction using least support vector machines within the evidence framework[J]. IEEE Transactions on Neural Networks, 2001, 12(4): 809—821.

[6] 董景荣. 汇率预报的非线性组合建模与预测方法研究[J]. 重庆师范学院学报(自然科学版), 2003, 20(3): 1—4.

[7] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32—42.

[8] MULLER K R, SMOLA A J, RATSCH G, et al. Predicting time series with support vector machines[A]. In: GERSTNER W, GERMOND A, eds. Proc of the International Conference on Artificial Neural Networks[C]. Berlin: Springer Verlag, 1997. 999—1004.

[9] 王学哲, 王相范. 关于几种经济预测模型的应用研究[J]. 哈尔滨商业大学学报, 2001, 17(2): 36—39.

Nonlinear combined prediction of port throughput based  
on support vector machine regression

WU Xiao, ZONG Bei-hua

(College of Transport & Communication, Shanghai Maritime University, Shanghai 200135, China)

**Abstract:** A nonlinear combined model for prediction of port throughput was presented based on the regression algorithm of support vector machines. The comparison of the predicted result of port throughput by the present method with those of other methods shows that the method developed in this paper is of strong capability in learning and mapping, and has high applicability in modeling for combined prediction of nonlinear systems with some uncertainties.

**Key words:** port throughput; nonlinear combined prediction; support vector machine regression