

文章编号:1673-9469(2009)02-0089-04

## 房地产投资环境的支持向量机评价

周书敬,曹立梅,李彦仓

(河北工程大学 土木工程学院,河北 邯郸 056038)

**摘要:**房地产投资环境是房地产投资成功与否的关键因素,因此做好房地产投资环境的评价工作十分重要。本文在介绍支持向量机回归理论的基础上,提出了基于支持向量机回归理论的房地产投资环境评价模型。通过数据仿真及与BP神经网络方法的比较,证明了该方法的有效性。

**关键词:**房地产;投资环境;指标;支持向量机回归;评价

**中图分类号:** TU293.3

**文献标识码:** A

### Real estate investment environment evaluation based on SVM

ZHOU Shu-jing, CAO Li-mei, LI Yan-cang

(College of Civil Engineering, Hebei University of Engineering, Handan 056038, China)

**Abstract:** The environment is the key to the success of real estate investment. Therefore, it is very important for the investment environment evaluation of real estate. Firstly, the basic theory of support vector machine for regression (SVR) was introduced, and then a SVM to evaluate the investment environment of real estate was built. Through simulating and comparing with the method of BP neural network, the method is proved effectiveness.

**Key words:** real estate; investment environment; index; SVR; evaluation

房地产投资环境是指在投资的一定区域(国家、地区、城市或街区)、一定时期内对投资所要达到的目标产生有利或不利影响的外部条件<sup>[1]</sup>。对投资环境进行评价,目前普遍采用的方法有:(1)初级因素打分法,包括:多因素等级尺度法、心理评估法、冷热因素分析法等。(2)多因素系统评估法,包括:多因素评估法、综合评判法等。(3)综合型定量法,包括:AHP层次分析法、模糊综合评判法、BP神经网络模糊综合评判法、灰色模糊综合评判法、灰聚类等<sup>[2]</sup>方法。这些方法都是根据国内外投资环境分析需要而探索出来的,其基本原理有较广的适应性。

统计学习理论(Statistical Learning Theory),简称SLT,是由V. N. Vapnik等人在20世纪70年代末提出的一种针对有限样本的统计理论。支持向量机方法(Support Vector Machine),简称SVM,是目前为止该理论最成功的实现<sup>[3,4]</sup>。SVM理论目前已经成功地处理回归问题(时间序列分析)和模式

识别(分类问题、判别分析)等问题。关于回归型支持向量机(Support Vector Regression,简称SVR),它可以按任意精度逼近非线性函数,具有全局极小点和收敛速度快的优点,已经被成功应用于天气预报、地下水位预报、电力系统负荷预测以及股票价格预测等领域。本文以支持向量机回归理论为基础,提出一种新的房地产投资环境评价方法,并以实例验证其可行性,为房地产投资环境评价提供了一种新方法。

### 1 基于SVR的房地产投资环境评价

#### 1.1 评价指标体系的建立

投资环境评价指标体系是对投资环境进行综合评价的依据和标准。它的设置应符合系统全面、简明科学、稳定可比、灵活可操作的原则。因此,本文立足于投资行业的实际情况,运用系统工程的理论和方法建立了一套结构齐全、内容丰富、

操作方便的投资环境评价体系,主要因素有自然环境、经济环境、政策环境、社会文化环境,这4个方面又有多级子因素,详见表1。

表1 房地产投资环境评价指标体系  
Tab.1 Index system of real estate investment environment evaluation

投资环境	指标体系
自然环境	基础设施 $X_1$
	区位交通 $X_2$
	环境状况 $X_3$
	土地状况 $X_4$
	地理位置 $X_5$
经济环境	房地产供求状况 $X_6$
	房地产中介服务 $X_7$
	地价 $X_8$
	居民收入水平 $X_9$
	经济增长水平 $X_{10}$
政策环境	同类楼盘的分布 $X_{11}$
	房地产政策 $X_{12}$
	政府办事效率 $X_{13}$
	城市规划控制 $X_{14}$
	税收 $X_{15}$
社会文化环境	区域人口数量 $X_{16}$
	区域人口素质 $X_{17}$
	家庭人口构成 $X_{18}$
	消费文化与观念 $X_{19}$
	社会治安因素 $X_{20}$

1.2 支持向量机回归理论<sup>[5]</sup>

设训练样本集  $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$   $x_i \in R^n, y_i \in R$

其中  $x_i$  是第  $i$  个学习样本的输入值,即第  $i$  类房地产投资环境评价指标属性值,其为  $n$  维向量  $x_i = (x_i^1, x_i^2, \dots, x_i^n)$ ,  $y_i$  为对应的目标输出值,即此类房地产投资环境的评价值。

1)线性回归问题。线性回归方程为  $f(x) = \langle \omega, x \rangle + b$  (1)

式中  $\omega \in R^n$ —权值向量; $b \in R$ —阈值。

先定义不敏感损失函数为  $\epsilon$ -insensitive 损失函数,其形式如下

$$L_\epsilon(y) = \begin{cases} 0, & |f(x) - y| \leq \epsilon \\ |f(x) - y| + \epsilon, & |f(x) - y| > \epsilon \end{cases} \quad (2)$$

即如果目标值  $y$  和估计函数值  $f(x)$  之差小于

$\epsilon$ ,则损失为0。

SVR 问题,可化为求解下列数学规划问题(引入松弛变量  $\zeta_i, \zeta_i^*$ ):

$$\min_{\omega, b, \zeta_i, \zeta_i^*} \|\omega\|^2/2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*) \quad (3)$$

$$((\omega \cdot x_i) + b) - y_i \leq \omega + \zeta_i$$

$$s. t. \quad y_i - ((\omega \cdot x_i) + b) \leq \omega + \zeta_i^* \quad (4)$$

$$\zeta_i, \zeta_i^* \geq 0, (i = 1, 2, \dots, n)$$

通过拉格朗日变换,得到其对偶问题为

$$\max_{\alpha, \alpha^*} \left[ -\frac{2}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \right] \quad (5)$$

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$$

$$s. t. \quad 0 \leq \alpha_i, \alpha_i^* \leq C, (i = 1, 2, \dots, n) \quad (6)$$

其中  $C$  为正常数,称为惩罚参数。

求解上述规划问题,得拉格朗日乘子  $\alpha_i, \alpha_i^*$ , 则回归方程(1)式中的系数

$$\omega = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i \quad (7)$$

因  $\alpha_i, \alpha_i^*$  不可能全为0,由 Karush - Kuhn - Tucker 条件知,当  $0 < \alpha_i < C$  时有  $\zeta_i = 0$ ,不等式  $((\omega \cdot x_i) + b) - y_i \leq \omega + \zeta_i$  就变成等式,可得  $b$ ; 同样当  $0 < \alpha_i^* < C$  时有  $\zeta_i^* = 0$ ,不等式  $y_i - ((\omega \cdot x_i) + b) \leq \omega + \zeta_i^*$  就变成等式,也可得  $b$ 。

2)非线性回归问题。由于本文讨论的房地产投资环境评价问题为非线性回归问题,故在此通过引入适当的核函数  $K(x, y)$ ,将输入样本空间非线性变换到另一特征空间,并在此空间中运用原空间的函数进行线性估计。常用的核函数有:

线性(Linear)核函数:  $K(x, y) = xy$ ; (8)

多项式(Polynomial)核函数:  $K(x, y) = (xy + 1)^d \quad d = 1, 2, \dots$  (9)

径向基(Radial Basis)核函数:  $K(x, y) = \exp\left[-\frac{\|x - y\|^2}{2\sigma^2}\right]$  (10)

Sigmoid 内积函数:  $K(x, y) = \tanh(v(xy) + c)$  ( $c$  为核参数) (11)

本文采用常用的径向基核函数。

非线性回归方程为

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

求解下列规划问题:

$$\max_{\alpha, \alpha^*} \left[ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) + \right]$$

$$\sum_{i=1}^n \gamma_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \quad (13)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, (i = 1, 2, \dots, n) \end{aligned} \quad (14)$$

1.3 评价原理及实现步骤

基于支持向量机的房地产投资环境评价模型,就是将影响房地产投资环境的指标属性值作为支持向量机的训练样本,通过调整模型和核函数参数训练并构造最理想评价机器。

步骤如下:

步骤1 建立房地产投资环境评价指标体系,本文采用的指标体系如表1。

步骤2 取一定数量的结果已有定论的项目作为样本,建立训练样本集。

步骤3 求解模型的参数  $\epsilon$ , 选择核函数参数  $\sigma$  和惩罚参数  $C$ 。

步骤4 输入指标属性值,对房地产投资环境进行评价。

2 案例分析

本文采用文献[6]中数据进行训练,如表2中所示。

把各房地产投资环境评价指标  $X_1$  到  $X_{20}$  的值作为 SVR 的输入,最后一列的得分作为 SVR 的期望输出,用 SVR 进行仿真。测试结果分为4个水平,1~0.8为优秀,0.8~0.7为良好,0.7~0.5为一般,0.5~0为差。选取前12个作为训练样本,其余4个作为检验样本。这里取不敏感损失函数度量值  $\epsilon = 0.0001$ ,由交叉验证选取惩罚参数  $C = 700$ ,径向基核函数参数  $\sigma = 1.52$ 。在 Matlab6.5 环境下用 SVR 进行仿真,然后与文献[6]中 BP 神经网络仿真结果比较,结果见表3。

表2 学习样本  
Tab.2 Learning samples

编号	评价指标																				得分
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	$X_{16}$	$X_{17}$	$X_{18}$	$X_{19}$	$X_{20}$	
1	0.7	1	1	0.7	0.7	1	0.7	0.7	1	1	0.7	1	0.7	0.7	0.7	0.7	1	1	1	1	0.861
2	0.5	0.7	0.5	0.7	0.5	0.5	0.5	0.7	0.5	0.7	0.7	0.5	0.7	0.5	0.7	0.7	0.7	0.7	0.7	0.5	0.604
3	0.5	0.5	0.3	0.5	0.3	0.3	0.3	0.5	0.3	0.5	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.3	0.3	0.1	0.34
4	0.5	0.5	0.5	0.5	0.7	0.3	0.3	0.5	0.5	0.7	0.7	0.5	0.3	0.5	0.7	0.7	0.7	0.7	0.3	0.5	0.488
5	0.5	0.5	0.5	0.7	0.7	0.5	1	0.7	1	0.7	0.7	1	1	0.7	0.7	0.7	1	0.7	0.7	0.1	0.713
6	1	0.7	1	1	1	1	1	0.7	1	1	0.7	1	1	0.7	0.7	0.7	1	0.3	0.7	1	0.931
7	1	0.7	1	1	1	1	1	0.7	1	1	0.7	1	1	0.7	0.5	0.7	1	1	0.7	0.1	0.81
8	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.5	0.7	0.7	0.5	0.5	0.7	0.7	0.7	0.7	0.7	0.7	1	0.5	0.641
9	0.7	0.7	1	0.7	0.7	0.5	0.7	0.5	0.7	0.7	0.7	0.5	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.5	0.683
10	0.7	0.7	1	1	0.7	0.5	1	0.7	1	1	0.7	1	1	0.7	0.7	0.7	1	0.7	1	0.5	0.827
11	0.7	0.5	0.5	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.7	0.7	0.5	0.7	0.7	0.7	1	0.647
12	0.7	0.7	1	0.7	0.7	0.5	1	0.7	1	0.7	0.7	1	1	0.7	0.7	0.7	1	0.7	0.7	0.5	0.727
13	0.7	0.7	0.3	0.5	0.3	0.5	0.3	0.5	0.5	0.5	0.7	0.5	0.3	0.7	0.7	0.5	0.7	0.3	0.3	0.1	0.46
14	0.7	1	1	1	0.7	1	1	0.7	1	1	0.7	1	0.7	0.7	0.7	0.7	0.7	0.7	1	0.5	0.817
15	0.7	0.7	0.5	1	0.7	1	1	0.7	1	1	0.7	1	1	0.7	0.7	0.7	0.7	0.7	0.7	0.5	0.766
16	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.5	0.5	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.3	1	0.1	0.63

表3 基于SVR和BP神经网络的训练结果比较

Tab.3 Based on the SVM and BP neural network training results

编 号	期望输出	SVR 方法		BP 网络方法	
		训练结果	相对误差 (%)	训练结果	相对误差 (%)
1	0.861	0.8609	0.0116	0.8612	0.0232
2	0.604	0.6035	0.0828	0.6029	0.1821
3	0.34	0.3396	0.1176	0.3408	0.2353
4	0.488	0.4883	0.0615	0.4894	0.2869
5	0.713	0.7146	0.2244	0.7185	0.7714
6	0.931	0.9328	0.1933	0.9193	1.257
7	0.81	0.8107	0.0864	0.8094	0.0741
8	0.641	0.6421	0.1716	0.6343	1.0452
9	0.683	0.6843	0.1903	0.689	0.8785
10	0.827	0.829	0.2418	0.8289	0.2297
11	0.647	0.6472	0.0309	0.6511	0.6337
12	0.727	0.7219	0.7015	0.7196	1.0179

表4 检验结果与分类比较

Tab.4 Comparison of test results and classification

编 号	期望输出	SVR 方法		BP 网络方法	
		检验结果	误差	检验结果	误差
13	0.46	0.4513	1.8913	0.4509	1.9783
14	0.817	0.8247	0.9425	0.8369	2.4357
15	0.766	0.7665	0.0652	0.7567	1.2141
16	0.63	0.6381	1.2857	0.6614	4.9841

从以上仿真结果可看出,用检验样本仿真评价的输出结果和期望输出基本一致,最大相对误差为 1.891 3%,最小相对误差为 0.065 2%,总相对误差较神经网络方法要小(见表 4),说明 SVM 的评估结果比 BP 网络方法更准确有效。

### 3 结束语

支持向量机方法计算简便,其建模只要通过对样本的学习即可获得因变量和自变量之间复杂的非线性关系。其最早是针对分类问题提出的,其后才逐渐推广到非线性回归估计中,而且取得了良好的效果。本文将其用在房地产投资环境评价中,避免了样本复杂及易陷入局部极小点等问题,是对房地产投资环境作出正确评价的一个简单有效的方法。

### 参考文献:

- [1] 方维慰. 房地产投资环境评价体系的构建及应用[J]. 南京航空航天大学学报, 2002, 4(4): 43 - 46.
- [2] 李继勇, 周书敬, 李彦苍. 基于信息熵的房地产投资组合模型[J]. 河北建筑科技学院学报, 2005, 22(2): 76 - 80.
- [3] VAPNIK V N. The nature of statistical learning theory[M]. New York: Spinger - Verlag, 1995.
- [4] VAPNIK V N. Statistical learning theory[M]. New York: John Wiley, 1998.
- [5] 许葆华, 李洪儒, 年海涛. 支持向量机在时间序列预测中的应用[J]. 微计算机信息化, 2008, 24 : 253 - 254.
- [6] 张新红. 基于神经网络的高技术项目投资风险综合评价模型[J]. 情报理论与实践, 2001, 24(5): 377 - 379.

(责任编辑 闫纯有)