

# 地统计学与支持向量机相融合的建筑工程造价预测<sup>①</sup>

刘 春

(四川建筑职业技术学院 网络管理中心, 德阳 618000)

**摘 要:** 为了精确对建筑工程造价进行预测, 根据建筑工程造价数据间的时间相关性和非线性变化特点, 构建了地统计学与支持向量机相融合的建筑工程造价预测模型. 首先收集建筑工程造价数据, 然后根据建筑工程造价数据的时间相关性, 采用地统计学估计时间序列的嵌入维数, 构建建筑工程造价的学习样本, 最后通过支持向量机建立工程造价预测模型, 并通过实例对性能进行测试与分析. 结果表明, 本文模型可以有效拟合建筑工程造价数据时间相关性, 获得高精度的建筑工程造价预测结果, 而且结果要明显优于其它模型.

**关键词:** 建筑工程造价; 嵌入维数; 地统计学; 时间相关性; 支持向量机

引用格式: 刘春. 地统计学与支持向量机相融合的建筑工程造价预测. 计算机系统应用, 2018, 27(4): 272-275. <http://www.c-s-a.org.cn/1003-3254/6305.html>

## Prediction of Construction Cost Based on Integration of Geostatistics and Support Vector Machines

LIU Chun

(Network Management Center, Sichuan College of Architectural Technology, Deyang 618000, China)

**Abstract:** In order to accurately predict the cost of construction project, according to the sample time dependent and nonlinear characteristics, this study has constructed a prediction model of construction project cost based on the geostatistics and support vector machine. Firstly, the cost data of construction engineering are collected, and then, the embedding dimension of time series is obtained by geostatistics according to time correlation of construction cost sample, and the construction cost learning samples of construction project, support vector machine is used to establish project cost prediction model, and test and analysis of the performance through the prediction example. The results show that the proposed model can effectively fit the construction cost of sample time correlation, and get higher accuracy of construction cost prediction, and the result is much better than the other models.

**Key words:** construction cost; embedding dimension; geostatistics; time correlation; support vector machines

### 1 引言

随着我国经济水平的不断提高, 人们住房、生活得到了大幅度的改善, 出现了许多大型建筑. 在建筑构建过程中, 工程造价是大家关心的重要问题, 其直接影响到工程的管理、成本以及其它方面, 因此建立高精度的建筑工程造价预测模型一直是人们关注的焦点<sup>[1-3]</sup>.

建筑工程造价建模常基于时间序列数据进行, 把建筑工程造价看作是一种按时间前后顺序组织在一起

的数据, 数据之间存在明显时序特性, 同时受到其它因素影响, 具有较强的非线性变化特点. 传统建筑工程造价预测模型有: 多元性线回归, 滑动平均模型等, 它们属于线性建模方法, 无法描述建筑工程造价的非线性变化特征, 建筑工程造价预测结果不太可靠<sup>[4,5]</sup>. 近些年, 人们将神经网络引入到了建筑工程造价的建模中, 由于具有非线性预测能力, 获得了比传统模型更高的建筑工程预测精度<sup>[6-8]</sup>. 神经网络要求建筑工程造价的

<sup>①</sup> 基金项目: 四川建筑职业技术学院 2015 年院级科研项目 (2015KJ07)

收稿时间: 2017-07-31; 修改时间: 2017-08-14; 采用时间: 2017-08-22; csa 在线出版时间: 2018-03-31

历史数据多,如果历史数据太少,会出现“过拟合”的建筑工程造价预测结果.支持向量机(Support Vector Machine, SVM)是一种非线性建模能力强的算法,克服了神经网络对训练样本规模的约束条件,预测性能要优于神经网络,为建筑工程造价提供了一种新的建模工具<sup>[9]</sup>.如何对历史数据间的时间相关性进行有效挖掘十分关键,常采用拓阶的方法实现,拓阶过程消耗时间长,而且属于线性拓阶方法,结果无法描述建筑工程造价的非线性变化特性.为了精确对建筑工程造价进行预测,根据对建筑工程造价样本时间相关性和非线性变化特点,构建了基于地统计学与支持向量机相融合的建筑工程造价预测模型,并通过建筑工程造价预测实例对性能进行测试与分析.

## 2 支持向量机与地统计学

基于支持向量机的建筑工程造价预测建模过程中,由于根据建筑工程造价数据间的时间相关性,由支持向量机自身无法确定表示时间相关性的嵌入维数,地统计学(Geostatistics, GS)<sup>[10,11]</sup>是一种时序特性分析法,可以将变程作为时间序列的阶数,即嵌入维数,而且工作过程简单、执行速度快,可以用于确定支持向量机的建筑工程造价时间序列拓阶.

### 2.1 支持向量机

设训练样本集合为:  $T = (x_i, y_i)$ ,  $i = 1, 2, \dots, m$ ,  $x_i$ 为输入向量,  $y_i$ 为输出向量,支持向量机通过学习对输入和输出向量间的关系进行拟合进行建模,拟合函数可以表示为:

$$f(x) = w\Phi(x) + b \quad (1)$$

式中,  $w$ 为权向量,  $b$ 为阈值,  $\Phi(\cdot)$ 为高维特征空间的映射函数.

为了使预测结果的风险最小化,建立如下的经验风险函数:

$$R_{\text{emp}} = \frac{1}{m} \sum_{i=1}^m L(y_i - f(x_i)) \quad (2)$$

式中,  $L$ 为损失函数,具体为

$$L(\cdot) = \begin{cases} 0, & \text{if } |y_i - f(x_i)| < \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{else} \end{cases} \quad (3)$$

式中,  $\varepsilon$ 是回归误差的阈值<sup>[12]</sup>.

对于小样本预测问题,基于经验风险原理进行建模可能导致预测结果差,为此支持向量机考虑了泛化

风险,得到如下回归函数:

$$f(x) = \sum_{i=1}^m w_i \Phi_i(x) + b \quad (4)$$

式(4)可以转化为求解以下函数规划问题:

$$\min \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \right) \quad (5)$$

约束条件为:

$$\text{s.t.} \begin{cases} y_k - f(x_k) \leq \varepsilon + \xi_i \\ f(x_k) - y_k \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \quad (6)$$

式中,  $\xi, \xi^*$ 为松弛变量,  $C$ 为惩罚参数.

引入拉格朗日算子  $\alpha_i$  得到式(6)的对偶问题,然后进行求解可得:

$$f(x) = \sum (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (7)$$

式中,  $K(\cdot)$ 为核函数,其满足如下条件:

$$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j)) \quad (8)$$

本文采用RBF函数,其定义为

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \quad (9)$$

在建筑工程造价进行建模时,相关研究表明,下一时刻建筑工程造价与前  $m$  个建筑工程造价相关,即时间序列的嵌入维数,用  $y_{k+1}$  表示当建筑工程造价,那么建筑工程造价的预测原理可以表示为:

$$y_{k+1} = f(y_k, y_{k-1}, \dots, y_{k-m}) \quad (10)$$

### 2.2 地统计学

地统计学(GS)是一种现代统计学理论,根据变异函数  $\gamma(h)$  刻画一个时间序列的结构.设时间序列为:  $Z(x_i)$ ,  $i = 1, 2, \dots, n$ ,  $\gamma(h)$  定义如下:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (11)$$

协方差函数定义为:

$$C(h) = E[Z(x)Z(x+h)] - m^2 \quad (12)$$

式中,  $h$  表示时间间隔距离.

## 3 GS-SVM 的建筑工程造价预测模型

(1) 收集建筑工程造价相关数据,得到数据形式为  $X = \{x_1, x_2, \dots, x_n\}$ , 并进行如下处理.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \times 0.8 + 0.1 \quad (13)$$

(2) 采用地统计学通过变程确定建筑工程造价的最优嵌入维数, 具体为: 根据不同  $h$ , 由式 (11) 得到不同的  $\gamma(h)$ , 刚开始,  $h$  与  $\gamma(h)$  之间是一种线性变化关系, 当  $h$  超过一定距离后,  $\gamma(h)$  得到第一个最大值, 该  $\gamma(h)$  的值为基台值, 此时的  $h$  为变程, 用  $\alpha$  表示, 即得到了时间序列的嵌入维值.

(3) 根据最优嵌入维数和建筑工程造价的影响因素构建支持向量机的输入向量, 与相应的建筑工程造价组成学习样本.

(4) 将建筑工程造价训练样本集输入支持向量机, 并设置支持向量机参数, 建立建筑工程造价预测模型. GS-SVM 的工程造价建模过程如图 1 所示.

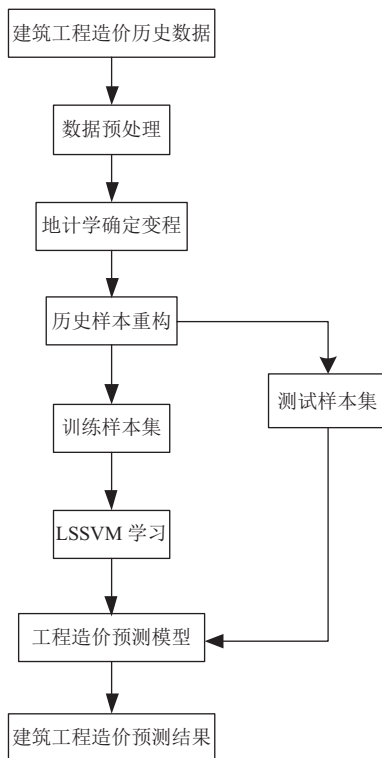


图 1 工程造价预测模型的工作流程

## 4 建筑工程造价预测实例

### 4.1 数据来源

选择某市 2000~2016 年的每月工程造价数据作为研究对象, 共获得 180 个样本, 采用单层面积、层数、层高、承载力、基深作为影响, 工程造价数据如图 2 所示, 以 130 个月数据作为训练样本, 建立建筑工程造

价预测模型, 其余 50 个样本对建筑工程造价预测模型的泛化能力进行测试.

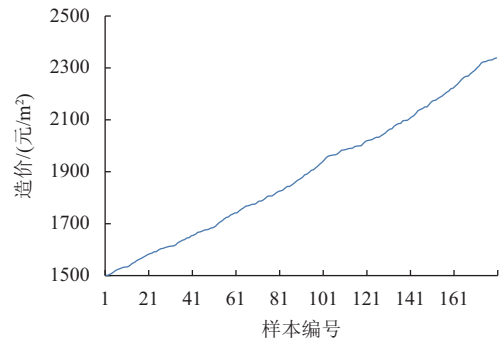


图 2 实验对象

从图 2 可知, 2000~2016 年的每月建筑工程造价数据具有明显上升变化态势, 为此建立建筑工程造价模型之前, 对原始数据进行去趋势平稳化法处理, 具体如下:

首先对 2000~2016 年的每月建筑工程造价数据取对数, 然后进行拟合, 即

$$\ln y_t = a + bt \quad (14)$$

根据式 (14) 得到去趋势平稳化后的 2000~2016 年的每月建筑工程造价数据:

$$y_t = \ln y_t - (a + bt) \quad (15)$$

后续建模基于平稳化后的建筑工程造价数据.

### 4.2 基于 GS 确定建筑工程造价数据嵌入维数

建筑工程造价不仅与单层面积、层数、层高、承载力、基深等因素有关, 而且与前一段时间的建筑工程造价相关, 因此, 建筑工程造价时间序列具有明显的时间关联特性, 对平稳化后的建筑工程造价数据进行地统计学分析, 得到建筑工程造价的半变异函数如图 3 所示. 从图 3 可知, 变程  $\alpha=5$ , 即建筑工程造价与前 5 个月的建筑工程造价相关, 因此嵌入维数为 5. 前 5 个月的建筑工程造价数据与单层面积、层数、层高、承载力、基深作为输入, 构建支持向量机的学习样本.

### 4.3 结果与分析

#### 4.3.1 GS-SVM 的预测结果

通过十折交叉验证法确定支持向量机的参数:  $C=240.80$ ,  $\sigma=1.784$ , GS-SVM 的建筑工程造价预测结果如图 4 所示. 从图 4 可知, GS-SVM 的建筑工程造价预测精度相当高, 能够较好描述建筑工程造价的变化特点, GS-SVM 可以应用于实际的建筑工程造价中.

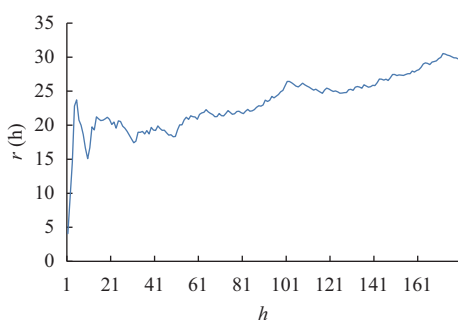


图3 建筑工程造价的半变异函数图

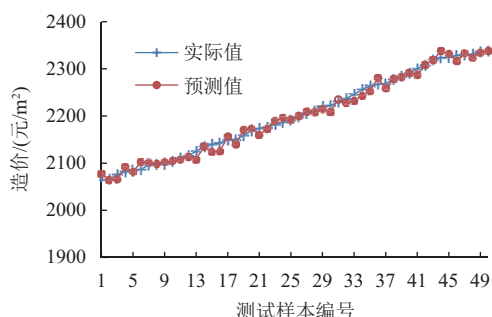


图4 GS-SVM的建筑工程造价预测结果

#### 4.3.2 与其它模型的预测结果

比较采用多元线性回归模型(MLR)、支持向量机(SVM)、地统计学与BP神经网络(GS-BPNN)作为对比模型,采用平均预测精度作为建筑工程造价预测结果好坏评价指标.统计所有模型的建筑工程造价预测精度,得到结果如表1所示.对表1的实验结果进行分析,可知:

(1)单一支持向量机(SVM)没有考虑建筑工程造价数据的时序动态特性,没有进行拓阶操作,只是简单采用影响因素作为输入变量,导致建筑工程造价预测精度低.

(2)多元线性回归模型(MLR)的建筑工程造价最低,说明其不能准确描述建筑工程造价的变化特点,获得了较高的建筑工程造价预测误差,没有什么实际应用价值.

(3)GS-BPNN的建筑工程造价平均预测精度得到了一定提升,这主要由于通过GS对建筑工程造价时间序列进行拓阶处理,描述了数据的时序动态特性,大幅度降低了建筑工程造价的预测误差,表明引入变程可以有效提高建筑工程造价预测精度.

(4)GS-SVM的建筑工程造价预测精度最高,这是因为采用支持向量机对建筑工程造价进行建模,克服了神经网络大样本要求,泛化能力优异,并通过GS描

述建筑工程造价数据的时序特性,建立理想的建筑工程造价预测模型,相比对比模型,建筑工程造价预测精度具有明显优势.

表1 建筑工程造价的预测精度对比

模型名称	建筑工程造价预测精度(%)
MLR	85.97
SVM	90.52
GS-BPNN	92.28
GS-SVM	94.64

## 5 结束语

针对建筑工程造价中的训练样本选择问题,以及非线性变化特点,综合利用地统计学和支持向量机的优点,设计了基于GS-SVM的建筑工程造价预测模型,通过地统计学快速确定时间序列的嵌入维数,然后通过支持向量机建立建筑工程造价的非线性预测模型,具体应用实例测试结果表明,GS-SVM大幅度提高了建筑工程造价的预测精度,具有广泛的应用前景.

## 参考文献

- 陈鹏,刘亚南.建筑工程造价管理的现状与对策.工业工程,2015,34(48):33-34.
- 张利荣.多基元模糊算法在工程估价中的应用.施工技术,2010,39(6):64-66.
- 胡六星.基于时间序列的建筑工程造价预测研究.太原理工大学学报,2012,43(6):706-709,714.
- 郭一斌,王红革,王翔.基于Vague集贴近度的工程项目投资快速估算方法.现代经济信息,2011,12(2):50-55.
- 孟俊娜,梁岩,房宁.基于BP神经网络的民用建筑工程造价估算方法研究.建筑经济,2015,36(9):64-68.
- 胡晓娟.多元线性回归模型参数的STLS估计法及其在工程造价预测中的应用.四川建筑科学研究,2016,42(4):142-147.
- 牛东晓,乞建勋,邢棉.建筑工程造价预测的变结构神经网络模型研究.华北电力大学学报,2001,28(4):1-4.
- 李杰.改进粒子群算法优化支持向量机的工程造价预测.计算机系统应用,2016,25(6):202-206.
- Li GQ, Deng M, Zhu JJ, et al. Spatial outliers detection considering distances among their neighbors. Journal of Remote Sensing, 2009, 13(2): 197-202.
- Petitgas P, Woillez M, Doray M, et al. A geostatistical definition of hotspots for fish spatial distributions. Mathematical Geosciences, 2016, 48(1): 65-77. [doi: 10.1007/s11004-015-9592-z]
- 刘莘,张绍良,王飞,等.基于地统计学的空间离群点检测算法的研究.计算机应用研究,2016,33(12):3700-3704. [doi: 10.3969/j.issn.1001-3695.2016.12.040]
- 王雪松.改进支持向量机的网络流量预测.计算机系统应用,2017,26(3):230-233.