

改进粒子群算法优化支持向量机的工程造价预测^①

李 杰

(延安职业技术学院, 农林建筑工程系, 延安 716000)

摘 要: 工程造价预测一直是工程管理研究中的重点, 针对工程造价预测中的支持向量机参数优化问题, 提出一种改进粒子群算法优化支持向量机的工程造价预测模型(IPSO-SVM)。首先收集工程造价数据, 并对其进行归一化处理, 然后采用支持向量机对工程造价的训练样本进行学习, 并采用改进粒子群算法对支持向量机的核函数参数进行优化, 最后采用 Matlab 2012 工具箱对工程造价进行仿真实验。实验结果表明, IPSO-SVM 有效提高工程造价的预测精度, 预测结果具有一定的实际应用价值。

关键词: 工程造价预测; 参数优化; 粒子群算法; 支持向量机

Project Cost Forecasting Based on Improved Particle Swarm Algorithm Optimizing Support Vector Machine

LI Jie

(Yanan vocational & Technical College, Yanan 716000, China)

Abstract: Project cost forecasting is a key point in the research on project management, in view of support vector machine parameter optimization problem in project cost forecasting, a new project cost forecasting model (IPSO-SVM) is proposed, which is based on the improved particle swarm optimizing supporting vector machine. Firstly, project cost data is collected and processed, and then support vector machine is used to learn for training samples in which improved particle swarm algorithm is used to optimize kernel function parameters of support vector machine, At last, the simulation experiment is used to test the performance of project cost forecasting by using Matlab 2012. The experimental results show that IPSO-SVM can effectively improve the forecasting accuracy of project cost, and the forecasting results have some practical application values.

Key words: project cost forecasting; parameter optimizing; particle swarm algorithm; support vector machine

随着城市规模的不断壮大, 建筑数量急剧上升, 造价预测是建筑工程管理中的一个重要研究内容, 其对工程计划、管理、成本具有十分重要的作用^[1]。工程造价预测是根据当前的数据, 采用一定的技术进行分析, 对建筑工程的造价进行估计, 估计结果可以有效降低建筑工程的成本, 对建设工程顺利完成十分关键, 因此设计性能优异的工程造价预测模型引起了人们的重视, 成为建筑工程管理研究中的热点^[2]。

为了提高工程造价预测的准确性, 提出一种改进粒子群算法优化支持向量机的工程造价预测模型(IPSO-SVM), 最后采用 Matlab 2012 工具箱对工程造价进行仿真实验, 实验结果表明, IPSO-SVM 提高了工

程造价的预测准确性, 预测结果具有一定的实际应用价值。

1 工程造价预测的研究现状

由于建筑工程造价具有十分重要的实际应用价值, 国内外研究人员对其展开了一系列的研究, 取得了一定的研究成果。最原始的预测方法为定额法, 该方法通过一些专家对工程造价进行估计, 估计时间相当长, 比较复杂, 而且估计结果的好坏与专家自身所具有的经验直接相关, 得到的结果包含一定的主观性, 参考价值不大^[4]。随后有学者提出了基于类比工程的工程造价预测方法, 根据相似工程的造价对当前工程造价

^① 收稿时间:2015-09-19;收到修改稿时间:2015-12-02 [doi:10.15888/j.cnki.csa.005210]

进行估计, 由于市场变化的无常, 估计的精度比较低^[5]. 文献[6]和文献[7]提出了基于模糊理论的工程造价预测方法, 使工程造价预测精度得到了改善, 但是其确定特征隶属度、调整系数比较困难, 影响其应用范围^[8]. 近年来, 一些学者提出基于回归分析法、时间序列法的建筑工程造价预测模型, 回归分析法分析影响因子与建筑工程造价之间的变化关系, 但是其考虑的影响因子个数有限, 当影响因素具有不确定性, 预测性能急剧下降; 时间序列法将工程造价数据看作一种按时间先后变化的序列, 通过分析序列之间的联系, 建立工程造价预测模型, 其建模效率高、易实现, 但其本质属于线性建模技术, 当工程造价具有非线性变化趋势时, 预测结果不可靠, 局限性比较明显^[8]. 随着非线性技术和机器学习理论的飞速发展, 一些学者将神经网络、支持向量机引入到工程造价预测建模中, 它们需求的工程造价先验经验少, 具有较强的自适应能力, 可以拟合工程造价与影响因素之间的映射关系, 成为当前主要的研究方向. 在实际应用中, 神经网络是一种基于“大数定理”的非线性建模方法, 其网络结构复杂、样本规模要求比较大, 不然容易出现“过拟合、欠学习”等缺陷, 预测结果不稳定^[9,10]. 支持向量机 (support vector machine, SVM) 是一种基于解决神经网络缺陷的机器学习算法, 平衡了样本规模和预测误差之间的关系, 对于小样本的工程造价数据, 仍然可以获得较好的预测结果^[11]. 然而将 SVM 应用于工程造价预测过程中, SVM 的核函数及参数选择对工程造价预测结果起着决定性的作用, 当前主要采用遗传算法、粒子群算法解决 SVM 的参数选择问题, 然而标准粒子群算法存在收敛速度慢、易得到局部最优参数, 对工程造价的建模和预测产生不利影响^[12].

2 改进粒子群算法

2.1 标准粒子群算法及存在的不足

Kennedy 等受到鸟群的飞行行为的启发, 基于进化计算理论提出了粒子群优化 (Particle Swarm Optimization, PSO) 算法, 相对于其它进化算法如遗传算法, 其易于实现、调整参数少, 在许多领域得到了广泛的应用^[13].

在 PSO 算法中, 每个待求解问题的解被比喻成为一个粒子, 它们通过适应度函数评价其性能数, 第 i 个粒子的位置和速度分为: $\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ 和

$\vec{v}_i = (v_{i1}, v_{i2}, \dots, v_{iD})$, 其最优位置为 $\vec{p}_i = (p_{i1}, p_{i2}, \dots, p_{iD})$, 粒子群的最优位置为 $\vec{p}_g = (p_{g1}, p_{g2}, \dots, p_{gD})$, 那么第 t 代第 i 个粒子状态更新方式为^[14]:

$$\vec{v}_{id}(t+1) = \omega \cdot v_{id}(t) + c_1 r_1 (p_{id} - x_{id}(t)) + c_2 r_2 (p_{gd} - x_{id}(t)) \quad (1)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (2)$$

式中, w 表示惯性因子; c_1 和 c_2 为非负常数, 表示学习因子.

令 $\varphi_1 = c_1 \text{rand}()$, $\varphi_2 = c_2 \text{rand}()$, $\bar{\varphi} = \varphi_1 + \varphi_2$, $\bar{\varphi} = 1 - \bar{\varphi}$, $\varphi_p = \varphi_1 \text{pbest} + \varphi_2 \text{gbest}$, 利用状态空间将其描述为:

$$Y(t+1) = AY(t) + B \quad (3)$$

$$\text{式中, } Y(t) = \begin{bmatrix} x(t) \\ v(t) \end{bmatrix}, A = \begin{bmatrix} \bar{\varphi} & w \\ -\varphi w & \end{bmatrix}, B = \begin{bmatrix} \varphi_p \\ \varphi_p \end{bmatrix}.$$

A 的特征方程为:

$$\lambda^2 - (w + \bar{\varphi})\lambda + w = 0 \quad (4)$$

那么 A 的两个特征值为:

$$e_{1,2} = \frac{w + \bar{\varphi} \pm \sqrt{\Delta}}{2} \quad (5)$$

式中, $\Delta = (w + \bar{\varphi})^2 - 4w$.

那么式(3)的显式表示为:

$$\begin{cases} x(t) = \frac{\varphi_p - [k_1 e_1^t (e_1 - w) + k_2 e_2^t (e_2 - w)]}{\varphi} \\ v(t) = k_1 e_1^t + k_2 e_2^t \end{cases} \quad (6)$$

对式(6)进行分析得到: $x(t)$ 和 $v(t)$ 收敛速度与 $\|e_{1,2}\|$ 密切相关, $\|e_{1,2}\|$ 与 w 的取值相联系, $w = 0.4$ 和 $w = 0.9$, 标准 PSO 算法对标准函数 Rose 函数进行求解时, 粒子的运动过程如图 1 所示, 从图 1 可以得出, 当初期 w 较大, 局部搜索能力较弱, 迭代后期 w 较小, 全局搜索能力变弱, 易陷入局部最优, 而且缺乏学习机制, 易陷入局部最优, 很难逃出, 出现早熟收敛现象.

2.2 粒子群算法的改进

为了克服标准 PSO 算法存在的不足, 本文对其进行相应改进, 提出一种改进粒子群 (IPSO) 算法. 对于第 i 个粒子来说, 当 $x_{id}^{(t+1)}$ 代替 $\text{pbest}_{id}^{(t+1)}$ 时, 即 $x_{id}^{(t+1)}$ 为最优位置, 粒子就认为速度 $v_{id}^{(t)}$ 具有一定的优越性,

对 $x_{id}^{(t+1)}$ 影响比较大, 则保持 $v_{id}^{(t)}$, 不然对 $v_{id}^{(t)}$ 进行修正. w 具体调整策略为: 粒子当前最优位置一旦产生变化, 那么采用式(7)调整 w 值, 不然采用式(8)调整 w 值.

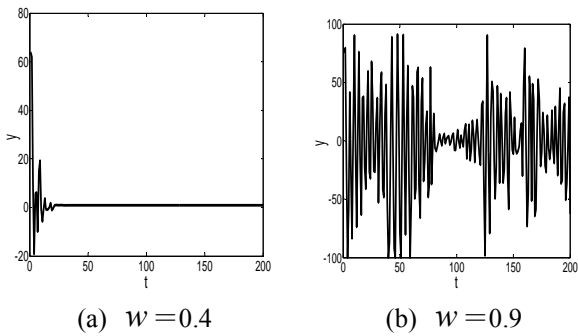


图 1 粒子的运动过程

$$w_{id} = \lambda_{1,id} + rand_{id} / 2.0, \quad pbest_{id}^{(t+1)} = x_{id}^{(t+1)} \quad (7)$$

$$w_{id} = \lambda_{2,id} + rand_{id} / 2.0, \quad pbest_{id}^{(t+1)} = pbest_{id}^{(t)} \quad (8)$$

由于 i_{best} 缺乏固定的学习机制, 为此本文也对其进行变异, 且满足下式:

$$p(v_{best,d}^{(t)}) = \exp[-(v_{best,d}^{(t)})^2 / 4(\sigma_d^{(t)})^2] / (\sigma_d^{(t)} \sqrt{2\pi}) \quad (9)$$

$$g_{best,d}^{(t+1)} = \begin{cases} g_{best,d}^{(t)} + v_{best,d}^{(t)} & f(g_{best,d}^{(t)} + v_{best,d}^{(t)}) \leq f(g_{best,d}^{(t)}) \\ g_{best,d}^{(t)} & f(g_{best,d}^{(t)} + v_{best,d}^{(t)}) > f(g_{best,d}^{(t)}) \end{cases} \quad (10)$$

$$\sigma_d^{(t)} = k \cdot g_{best,d}^{(t)} \quad (11)$$

式中, $p(v_{best,d}^{(t)})$ 为 i_{best} 变异的正态分布概率, $\sigma_d^{(t)}$ 愈小, i_{best} 更适用于局部搜索, 反之 i_{best} 更适用于全局搜索.

2.2 IPSO-SVM的工程造价预测模型

设训练集为 $\{(x_i, y_i) | i=1, \dots, l\}$, SVM 的回归函数可表示为:

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

选择 RBF 函数构建 SVM 的核函数, RBF 函数为:

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \quad (13)$$

式中, σ 为核函数宽度^[15].

最后, 支持向量机的回归函数为:

$$f(x) = \sum_{i=1}^l \alpha_i \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) + b \quad (14)$$

大量研究表明, 参数 C 和 σ 对工程造价的预测精度影响较大, C 值过大或过小会出现过学习或欠学习的缺陷, 核参数 σ 过大或过小会影响 SVM 的泛化性能. 参数 C 和 σ 优化目标是提高工程造价预测准确性, 采用工程造价精度(G)作为参数 C 和 σ 优化目标函数, 即有:

$$\begin{aligned} & \max_M G(M) \\ & s.t. \\ & \begin{cases} C_{\min} \leq C \leq C_{\max} \\ \sigma_{\min} \leq \sigma \leq \sigma_{\max} \end{cases} \end{aligned} \quad (15)$$

本文采用 IPSO 算法优化 SVM 的参数 C 和 σ . 每个粒子代表组合模型的最优参数(C, σ), 经过不断搜索, 找到粒子群的最优位置, 即为(C, σ), 然后 SVM 根据(C, σ)进行训练, 建立工程造价预测模型. 具体如图 2 所示.

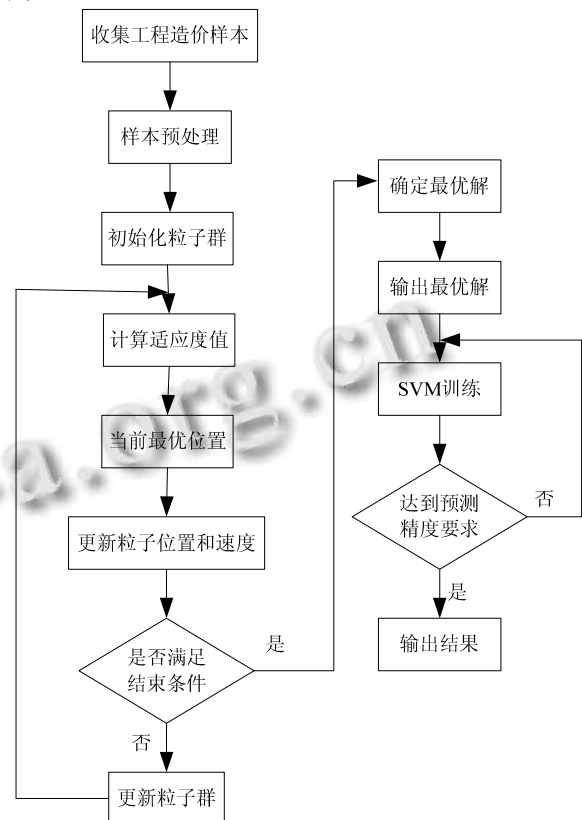


图 2 IPSO-SVM 的工程造价预测模型

3 仿真实验

3.1 数据来源

为了分析 IPSO-SVM 的工程造价预测结果, 在 Intel 4 核 3.0 GHz CPU, 8 GB 的内存, 1TB 的硬盘,

Window 8 的个人计算机上,采用 Matlab 2012 工具箱进行仿真实验.选取延安市 2013-2014 年工程造价典型数据的 100 组作为仿真对象,工程造价如图 3 所示.

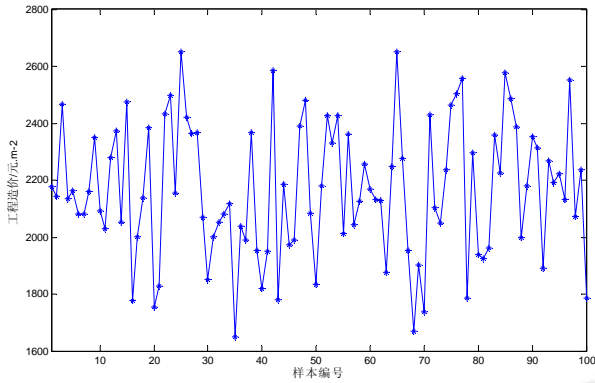


图3 仿真实验数据

3.2 结果与分析

3.2.1 IPSO-SVM 的预测性能分析

将图3中前50个的工程造价数据输入到SVM进行训练,采用IPSO算法选择SVM最优参数,最大迭代次数500, $c_1=c_2=2$, C 的取值范围为[0.1 100], σ 的取值范围为[0.1 10],得到SVM最优参数为 $C=1.57$, $\sigma=0.149$,IPSO-SVM对后50个工程造价数据的预测结果如图4所示. IPSO-SVM的工程造价预测值与工程造价预测的实际值非常接近,可以准确描述工程造价变化趋势,拟合偏差相当小,得到了十分理想的拟合效果.

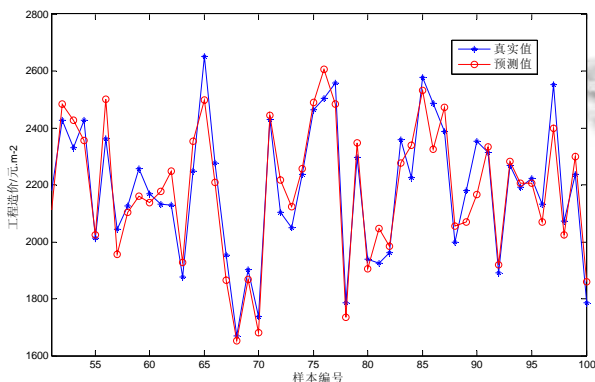


图4 IPSO-SVM 的预测结果

3.2.2 与经典模型的性能比较

将IPSO-SVM的预测结果与经典工程造价预测结果的结果进行比较,经典工程造价预测模型分为:多元线性回归分析、时间序列法、BP神经网络、PSO算

法优化SVM,统计它们的RMSE和MAPE,所有工程造价预测模型的各项预测指标值如表1所示,从表1可以得到如下结论:

(1) 多元线性回归分析模型和时间序列模型的预测误差较大,这主要是由于它们不能反映工程造价与影响因素之间的联系,预测精度比较低.

(2) IPSO-SVM的各项预测指标值要优于BP神经网络,这由于BP神经网络要求训练样本大,而在实际应用中收集工程造价样本数量有限,这样BP神经网络的拟合效果比较差,导致个别数据的预测值与真实值相差比较远.

(3) 相对于PSO-SVM,IPSO-SVM的各项预测指标值也更优,IPSO-SVM拥有更高的工程造价预测准确性,这说明本文对标准粒子群算法进行改进后克服了标准粒子群算法难以找到局部最优参数的缺陷,找到了更合理的SVM参数,可以更好描述工程造价数据的内在变化规律.

表1 与经典模型的预测性能对比

模型名称	RMSE	MAPE(%)
多元线性回归	124.09	9.22
时间序列法	120.18	8.75
BP神经网络	100.24	6.84
PSO-SVM	90.72	5.12
IPSO-SVM	81.76	3.39

4 结语

IPSO-SVM的工程造价预测模型首先对标准粒子群算法存在的问题进行改进,然后以实际工程造价作为SVM的期望输出,影响因素作为SVM的输入向量,挖掘工程造价数据序列的变化特点,以实际工程造价数据进行仿真实验,结果表明,IPSO-SVM的工程造价预测效果要优于当前经典的工程造价预测模型,为了工程造价建模与预测进行了有益探索.

参考文献

- 1 闻惠萍.建筑工程单方造价快速估算方法综述.建筑与工程,2008,9(7):40-45.
- 2 张利荣.多基元模糊算法在工程估价中的应用.施工技术,2010,22(6):64-66.
- 3 张利荣,郑健华.基于计价基元的工程估价方法研究.施工技术,2011,23(6):92-95.
- 4 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.
- 5 郑周练,邓绍江,高师娴,等.建筑工程造价的模糊确定.重庆大学学报(自然科学版),2001,24(1):55-58.
- 6 浦春,孙政顺,赵世敏.建设项目投资估算模型分析.计算机仿真,2007,25(5):135-137.
- 7 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.
- 8 易欣.基于 BP 网络的建筑工程造价估算.华东交通大学学报,2012,29(3):828-831.
- 9 周丽萍,胡振锋.BP 神经网络在建筑工程估价中的应用.西安建筑科技大学学报(自然科学版),2005,37(2):262-264.
- 10 雷雨.基于 BP 神经网络和 vague 集贴近度的工程造价估算研究.工程经济,2013,20(3):9-13.
- 11 胡伟勋.工程造价估算模型研究与应用.中南林业科技大学学报(自然科学版),2011,31(8):163-166.
- 12 张红梅,卫志农,龚灯才,等.基于粒子群支持向量机的短期电力负荷预测.继电器,2006,34(3):28-31.
- 13 柏丹丹,和敬涵,王小君,田文奇.自适应粒子群支持向量机风速组合预测模型.太阳能学报,2015,36(4):792-797.
- 14 李欣然,靳雁霞.权重自适应调整的混沌量子粒子群优化算法.计算机系统应用,2012,21(8):127-130.
- 15 孙志刚.蚁群优化支持向量机的物流需求预测.计算机系统应用,2013,22(5):107-110.