

DOI: 10.19951/j.cnki.1672-9331.20220624001

文章编号: 1672-9331(2023)03-0137-09

引用格式: 彭军龙, 胡珂, 王梦瑶, 等. 基于SSA-LSSVM的住宅工程造价预测研究[J]. 长沙理工大学学报(自然科学版), 2023, 20(3): 137-145.

Citation: PENG Junlong, HU Ke, WANG Mengyao, et al. Research on residential project cost prediction based on SSA-LSSVM [J]. J Changsha Univ Sci Tech (Nat Sci), 2023, 20(3): 137-145.

# 基于SSA-LSSVM的住宅工程造价预测研究

彭军龙, 胡珂, 王梦瑶, 彭超

(长沙理工大学 交通运输工程学院, 湖南 长沙 410114)

**摘要:**【目的】提高建筑工程造价预测的准确性, 进而为项目前期投资决策提供重要依据。【方法】针对工程实践中小样本数据和工程造价特征指标之间高维、非线性关系的特点, 构建基于麻雀搜索算法(sparrow search algorithm, SSA)的优化最小二乘支持向量机(least squares support vector machine, LSSVM)造价预测模型SSA-LSSVM。首先, 通过主成分分析法对住宅工程造价样本的输入指标数据进行处理, 减少数据冗余; 其次, 采用SSA算法对LSSVM模型中的正则化参数 $c$ 和核函数参数 $\sigma$ 进行优化, 以弥补LSSVM模型参数确定困难的缺陷; 最后, 将处理后的数据导入所构建的模型进行训练和预测, 并通过相关系数、平均绝对百分比误差和均方根误差对模型的预测性能进行评价。【结果】与LSSVM模型、用灰狼优化算法优化的LSSVM模型和反向传播神经网络模型相比, SSA-LSSVM模型具有更好的泛化能力和更高的预测精度。【结论】本研究所构建的模型可以比较精准、高效地对实际住宅工程造价进行预测, 同时可为工程前期投资决策提供一定参考。

**关键词:**住宅工程; 造价预测; 主成分分析; 麻雀搜索算法(SSA); 最小二乘支持向量机(LSSVM)

中图分类号: TU723.3

文献标志码: A

## 0 引言

建筑工程造价预测是项目可行性研究的重要内容, 准确的造价预测可协助建设单位实现更加科学的投资决策。在工程建设项目前期可获取的工程信息有限、设计深度不够等问题将会导致造价预测的精度和效率较低, 不能满足建筑市场精细化发展的实际需要<sup>[1]</sup>。因此, 快速、准确地进行建筑工程造价预测具有实践意义。

为提高工程造价预测的精度与科学性, 国内外学者们采用不同的方法构建工程造价预测模型。传统造价预测方法包括移动平均法、时间序列分析法和简单回归分析法等<sup>[2]</sup>。由于建设工程造价的影响因素多且复杂, 这类传统造价预测方法在实际应用过程中的精度有待进一步提高<sup>[3]</sup>。随着人工智能技术的蓬勃发展, 建筑工程造价预

测方法越来越智能、高效。机器学习方法因其具有良好的非线性映射能力, 被引入到了工程造价预测领域, 如学者们将粒子群算法<sup>[4]</sup>(particle swarm optimization, PSO)、人工神经网络(artificial neural network, ANN)和反向传播(back propagation, BP)神经网络<sup>[5]</sup>等智能机器学习算法应用于造价预测建模<sup>[6]</sup>。李玲玲<sup>[7]</sup>将人工神经网络与Adaboostda迭代算法、代价敏感方法相结合对建筑工程造价进行预测, 提高了ANN的泛化能力; 何利娟<sup>[8]</sup>基于灰色理论构建多变量GM(1, N)预测模型对哈密市的工程实例进行造价预测, 该模型能够很好地反应变量间的动态关系; 张晓东等<sup>[9]</sup>先对市政工程造价影响因素进行主成分分析, 将主成分得分作为输入变量, 再采用贝叶斯正则化算法构建带外源输入的非线性自回归(nonlinear autoregressive with exogenous inputs, NARX)神经网络模型进行市政工程造价预测;

收稿日期: 2022-06-24; 修回日期: 2022-08-12; 接受日期: 2022-08-22

基金项目: 国家自然科学基金资助项目(51578080); 湖南省自然科学基金资助项目(2021JJ30746)

通信作者: 彭军龙(1976—)(ORCID: 0000-0002-5585-0569), 男, 副教授, 主要从事工程项目管理方面的研究。

E-mail: 375135287@qq.com

KIM等<sup>[10]</sup>提出了案例推理(case-based reasoning, CBR)与遗传算法的混合预测模型,利用遗传算法优化特征权重,建立了建筑前期投资预估模型;ZHAO等<sup>[11]</sup>将遗传算法与建筑信息模型(building information modeling, BIM)辅助分析相结合对项目成本进行了预测;YE<sup>[12]</sup>基于240组高层住宅项目的竣工结算数据,将PSO算法与BP神经网络相结合对工程造价进行了预测。

上述智能算法相较于传统造价预测方法,预测的准确度有所提高,对样本数的要求有所降低<sup>[8]</sup>,但这些算法仍存在较易出现局部最优的缺陷,必须在大样本数据训练的情况下才能得到较优结果<sup>[13]</sup>。为提高预测精度,本研究采用非线性拟合能力好、泛化能力强、能较好地避免陷入局部最优解、对小样本数据表现突出的最小二乘支持向量机(least squares support vector machine, LSSVM)对住宅工程进行造价预测<sup>[14]</sup>。此外,通过麻雀搜索算法(sparrow search algorithm, SSA)对LSSVM模型中的正则化参数和核函数参数进行优化计算,找到合适的参数,使LSSVM具有更快的收敛速度、更强的泛化能力以及更高的预测精度,建立了基于SSA-LSSVM的住宅造价预测模型,并通过工程实例分析以及与其他模型的预测结果进行比较,选择最优预测模型,以期对住宅工程项目前期投资决策提供参考。

## 1 相关理论及住宅造价预测模型的建立

### 1.1 LSSVM

LSSVM是在标准支持向量机(support vector machine, SVM)基础上的一种重要改进,它选取误差平方和作为经验损失,用等式约束取代原算法中的不等式约束,不再需要求解复杂的二次规划问题,降低了求解难度,优化了算法的整体收敛速度,提高了计算结果的精度<sup>[15-16]</sup>。LSSVM的实现过程如下:

设定训练集  $S = \{(x_i, y_i), x_i \in \mathbf{R}^n, y_i \in \mathbf{R}\}$ , 其中,  $x_i$  为输入项;  $y_i$  为输出项;  $i = 1, 2, \dots, N$ ;  $\mathbf{R}$  为实数集;  $\mathbf{R}^n$  为  $n_1$  维实数集;  $N$  为训练样本的数量。

回归估计函数为:

$$f(x) = \omega^T \varphi(x) + b \quad (1)$$

式中:  $\omega$  为权值系数向量;  $\varphi(x)$  为非线性映射函数;  $b$  为偏置项。

根据结构风险最小化原则,可以将LSSVM的优化目标函数可以转化为:

$$\begin{cases} \min J(\omega, e) = \frac{1}{2} \|\omega\|^2 + \frac{c}{2} \sum_{i=1}^N e_i^2 \\ \text{s.t. } y_i = \omega^T \varphi(x_i) + b + e_i \end{cases} \quad (2)$$

式中:  $J$  为损失函数;  $e_i$  为误差变量;  $c$  为正则化参数。

通过构造拉格朗日乘子  $\alpha_i$  求解上述问题:

$$L(\omega, b, e, \alpha) = J(\omega, e) - \sum_{i=1}^N \alpha_i [\omega^T \varphi(x_i) + b + e_i - y_i] \quad (3)$$

式中:  $L$  为拉格朗日函数。

根据卡罗需-库恩-塔克(Karush-Kuhn-Tucher, KKT)条件,对式(3)中的各变量求偏导数:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^N \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = ce_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \omega^T \varphi(x_i) + b + e_i - y_i = 0 \end{cases} \quad (4)$$

通过求解,消掉上述方程中的  $\omega$  和  $e_i$ , 最终得到LSSVM回归函数为:

$$f(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \quad (5)$$

式中:  $K(x, x_i)$  为核函数。

常用的核函数有线性核函数、Sigmoid函数、径向基核函数(radial basis function, RBF)等。RBF具有预先设定的参数少、对实际问题泛化学习能力强的特点,且可以非线性映射,已经应用于许多实际问题当中。因此,RBF是LSSVM模型核函数的首选。RBF的具体表达式如下:

$$K(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\sigma^2}\right) \quad (6)$$

式中:  $\sigma$  为参数,表示核函数距中心点的宽度。

1.2 SSA

SSA 是 XUE 等<sup>[17]</sup>受麻雀种群在觅食和反捕食过程中的合作行为启发,在 2020 年提出的一种新型仿生智能优化算法。麻雀种群由发现者、追随者和预警者组成,根据适应度数值的改变,发现者和追随者可以相互转化。另外,需在麻雀种群中选择一部分个体作为预警者,其需要时刻警惕环境中的危险,并在察觉危险时发出预警。它们根据各自的更新规则不断更新位置,具体如下:

初始的麻雀种群位置可表示为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix} \quad (7)$$

式中: $n$ 为麻雀种群数量; $d$ 为数据维度。

通常情况下,发现者是种群中适应度较高的个体,担负着为麻雀种群寻找食物的责任。发现者的位置更新公式如下:

$$x_{ij}^{(t+1)} = \begin{cases} x_{ij}^{(t)} \exp\left(-\frac{i}{\alpha \cdot i_{\max}}\right), & R_2 < S_T \\ x_{ij}^{(t)} + QL, & R_2 \geq S_T \end{cases} \quad (8)$$

式中: $j=1, 2, \dots, d$ ;  $t$ 为当前迭代次数; $x_{ij}^{(t)}$ 为在第  $t$  次迭代中第  $i$  个发现者在第  $j$  维中的位置; $\alpha \in (0, 1)$ ,为均匀随机数; $Q$ 为服从正态分布的随机数; $i_{\max}$ 为最大迭代数; $L$ 为每一个元素皆为 1 的  $1 \times d$  维矩阵; $S_T$ 为安全阈值; $R_2$ 为预警值。当  $R_2 < S_T$  时,说明此时环境安全,发现者会扩大食物搜寻范围;当  $R_2 \geq S_T$  时,说明麻雀已经意识到危险,这时所有麻雀会迅速飞往安全区域进行觅食。

追随者的适应度较低,在整个觅食过程中它们都会监视发现者,并利用发现者来获取食物。当  $i > n/2$  时,表明第  $i$  个追随者未搜寻到食物,处于极度饥饿的状态,因此它需要前往其他区域搜寻食物,以获得更多能量。追随者的位置更新公式为:

$$x_{ij}^{(t+1)} = \begin{cases} Q \cdot \exp\left(\frac{x_{\text{worst}}^{(t)} - x_{ij}^{(t)}}{i^2}\right), & i > \frac{n}{2} \\ x_p^{(t+1)} + |x_{ij}^{(t)} - x_p^{(t+1)}| \cdot A^+ \cdot L, & i \leq \frac{n}{2} \end{cases} \quad (9)$$

式中: $x_p^{(t)}$ 为该次迭代时发现者占据的最优位置; $x_{\text{worst}}^{(t)}$ 为第  $t$  次迭代时种群的全局最劣位置; $A$ 为  $1 \times d$  维矩阵,其元素都被随机分配为 1 或 -1; $A^+ = A^T(AA^T)^{-1}$ 。

预警者的位置更新公式为:

$$x_{ij}^{(t+1)} = \begin{cases} x_{\text{best}}^{(t)} + \beta \cdot |x_{ij}^{(t)} - x_{\text{best}}^{(t)}|, & f_i > f_g \\ x_{ij}^{(t)} + \lambda \cdot \frac{|x_{ij}^{(t)} - x_{\text{worst}}^{(t)}|}{(f_i - f_w) + \varepsilon}, & f_i = f_g \end{cases} \quad (10)$$

式中: $x_{\text{best}}^{(t)}$ 为第  $t$  次迭代时种群的全局最优位置; $\beta$ 为步长参数,并服从标准正态分布  $N[0, 1]$ ; $\lambda \in [-1, 1]$ ,为一个随机数; $\varepsilon$ 为一个极小常数,加入  $\varepsilon$ 是为了防止出现  $f_i - f_w = 0$  的情况; $f_i, f_g, f_w$  分别为当前个体适应度、全局最优适应度和全局最差适应度。当  $f_i > f_g$  时,表示麻雀正处于种群的边缘地带,易被天敌捕食;当  $f_i = f_g$  时,表示处于种群中心位置的麻雀发觉了危险,需要靠近其他麻雀个体以保证自身安全。

1.3 住宅工程造价预测模型的建立

由 LSSVM 的建模过程可知,在住宅工程造价预测中,选择合适的 LSSVM 正则化参数  $c$  和核函数参数  $\sigma$  是精确预测的关键,故在预测之前运用 SSA 对 LSSVM 的  $c$  和  $\sigma$  进行参数寻优<sup>[18]</sup>。基于 SSA-LSSVM 的住宅工程造价预测模型的计算流程如图 1 所示。

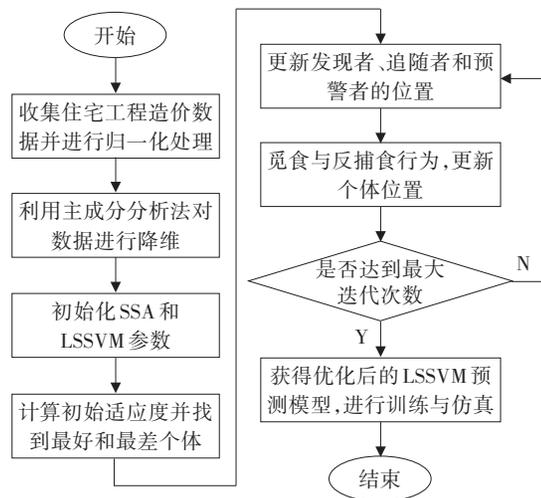


图 1 基于 SSA-LSSVM 的住宅工程造价预测模型的计算流程

Fig. 1 Algorithm flow of the residential project cost prediction model based on SSA-LSSVM

具体步骤为:

step 1: 获取住宅工程造价的相关数据,确定训练样本和测试样本数量,统一进行量化和归一化处理。

step 2: 利用主成分分析法对住宅工程造价的输入指标数据进行降维,并分析得出主成分。

step 3: 设置 SSA-LSSVM 模型的初始参数, 包括种群规模、预警者占种群的比例、最大迭代次数等。

step 4: 计算每只麻雀的初始适应度  $f_i$ , 并根据  $f_i$  的大小进行排序。

step 5: 根据式(8)~(10)更新麻雀种群发现者、追随者和预警者的位置, 计算麻雀新位置的适应度, 并更新  $f_g$  以及  $f_w$ 。

step 6: 判断是否达到停止条件, 若达到则输出全局最优参数, 否则转至 step 4。

step 7: 将  $c$  和  $\sigma$  赋值给 LSSVM, LSSVM 利用最优参数进行造价预测。

## 2 工程特征指标的选取及量化处理

关键特征指标的确定是构建预测模型的前提<sup>[1]</sup>。本研究采用文献分析法, 以“工程造价预

测”“cost prediction”“project investment estimate”等为关键词, 在知网、Science Direct 和 Web of Science 选取具有代表性的、与住宅工程造价预测有关的 50 篇文献, 根据造价影响因素出现的频数及其对造价的影响程度, 结合案例的实际情况<sup>[19]</sup>, 选定了 15 个特征指标, 分别为工期  $X_1$ 、地上层数  $X_2$ 、地下层数  $X_3$ 、建筑面积  $X_4$ 、檐口高度  $X_5$ 、三级钢筋质量占比  $X_6$ 、主体结构类型  $X_7$ 、基础类型  $X_8$ 、钢筋市价  $X_9$ 、混凝土市价  $X_{10}$ 、抗震等级  $X_{11}$ 、建筑外形  $X_{12}$ 、门窗类型  $X_{13}$ 、装修标准  $X_{14}$ 、安装工程水平  $X_{15}$ 。

对于定量数据直接按实际数值输入; 对于定性数据, 则采用特征哈希的思想, 将每个定性指标的编码设为相邻自然数。考虑到钢筋和混凝土的采购价格随市场供需关系的变化而变动较大, 故采用组距分组法将其划分为不同的价格区间。特征指标量化表见表 1。

表 1 特征指标量化表

Table 1 Characteristic index quantification table

定性指标编码	主体结构类型	基础类型	钢筋市价/(元·t <sup>-1</sup> )	混凝土市价/(元·m <sup>-3</sup> )	抗震等级	建筑外形	门窗类型	装修标准	安装工程水平
1	框架剪力墙	独立	≤3 800	≤350	一	塔式	塑钢	毛坯房	简单
2	剪力墙	筏板	(3 800, 4 100]	(350, 400]	二	板式	铝合金	简单	一般
3	框架	满堂	(4 100, 4 400]	(400, 450]	三	异形	实木	精良	良好
4	砖混	桩	(4 400, 4 700]	(450, 500]	四		防火		较好
5	短肢剪力墙	条形	>4 700	>500					完备

## 3 SSA-LSSVM 住宅造价预测模型的应用

### 3.1 数据来源

本研究从某工程造价信息数据平台、相应企

业等获取样本数据, 共收集了 50 个完整度较高的已竣工住宅项目的数据作为样本, 对住宅工程造价预测模型进行仿真分析。50 个样本的工程特征指标量化处理结果以及单方造价见表 2。

表 2 工程特征指标量化处理结果以及单方造价

Table 2 Quantitative processing results of engineering characteristic indexes and cost per square meter

样本	$X_1/d$	$X_2/$ 层	$X_3/$ 层	$X_4/m^2$	$X_5/m$	$X_6/%$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	单方造价/(元·m <sup>-2</sup> )
1	512	27	3	11 273.51	81.60	56.20	1	3	2	3	1	2	3	2	3	1 809.18
2	720	33	1	17 621.16	99.00	70.90	3	1	3	4	1	1	2	3	4	2 126.34
3	627	26	1	14 721.56	75.40	75.55	3	3	3	3	1	2	3	3	4	2 660.57
4	523	19	2	7 241.17	57.45	55.60	2	2	3	3	2	2	2	2	3	1 820.64
5	594	32	2	11 710.61	96.00	96.20	2	3	3	4	1	2	3	2	4	2 123.25

表 2 (续)  
Table 2 (Continued)

样本	$X_1$ / d	$X_2$ / 层	$X_3$ / 层	$X_4$ / m <sup>2</sup>	$X_5$ / m	$X_6$ / %	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	单方造价/ (元·m <sup>-2</sup> )
6	523	19	2	17 446.49	57.45	70.30	2	2	3	3	2	2	3	2	3	1 804.69
7	750	18	1	16 948.58	52.50	97.50	1	2	1	3	1	2	2	2	2	1 475.68
8	300	8	2	8 696.46	26.40	97.20	2	3	4	4	1	1	3	1	4	2 403.23
9	547	16	2	14 376.25	44.80	71.97	3	4	3	3	2	1	2	1	4	2 038.61
10	800	30	1	12 273.92	88.80	94.40	1	4	4	5	1	2	3	3	5	2 612.69
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
41	397	11	1	6 671.00	32.20	72.60	2	4	3	4	1	2	3	2	4	1 986.23
42	450	9	1	7 527.30	28.90	61.60	1	4	2	4	1	2	1	2	3	1 646.17
43	651	28	2	12 467.00	79.20	38.20	1	2	2	3	1	2	4	3	4	1 868.37
44	594	32	1	14 875.25	92.80	95.10	1	4	4	4	1	1	4	3	3	2 493.40
45	532	11	1	17 183.26	31.90	59.70	3	3	2	2	2	2	3	2	2	1 452.33
46	540	18	1	11 197.80	53.20	72.88	1	3	3	3	1	2	4	2	3	1 780.42
47	620	32	1	15 357.44	87.00	92.35	1	3	3	4	1	2	2	3	4	2 171.02
48	710	15	2	18 306.29	45.00	70.30	1	3	3	3	2	2	3	3	3	1 732.94
49	340	10	0	7 927.00	32.10	77.36	2	4	3	3	1	2	2	3	4	1 989.34
50	512	27	3	11 273.51	81.60	58.60	1	3	3	3	1	2	3	2	3	1 889.18

3.2 主成分分析法

主成分分析法是一种经典的特征提取方法,它将相关性较高的变量转化成相互独立的主成分,同时能尽可能多地保留原始数据承载的信息<sup>[20]</sup>。一般来讲,初步选定的、会影响住宅工程造价预测结果的因素过多且存在一定的相关性,因此本研究对 50 个样本的指标数据进行主成分分析来降低数据维度,以提高模型的预测精度。

使用 SPSS 24.0 软件对样本数据进行分析,结果见表 3。由表 3 可知,KMO(Kaiser-Meyer-Olkin)的值为 0.734,大于 0.7; Bartlett 球形检验的显著性水平  $P < 0.001$ ,说明样本数据之间存在较强的相关性,对工程特征指标进行主成分分析在技术上是可行的。

表 3 KMO 的值和 Bartlett 球形检验结果

Table 3 KMO value and results of the Bartlett's sphericity test

KMO 值	Bartlett 的球形检验		
	近似卡方	$d_f$	$P$
0.734	471.071	105	<0.001

分析计算得出各主成分的特征值和贡献率,结果见表 4。一般来讲,可提取累积贡献率超过 85% 的因子,因此本研究提取前 8 个主成分,其累积贡献率达到了 85.984%,用这 8 个主成分取代初

始特征指标作为预测模型的新的输入数据集。

表 4 主成分的特征值与贡献率

Table 4 The eigenvalues and contribution rates of principal components

主成分	特征值	贡献率/%	累积贡献率/%
1	4.724	31.493	31.493
2	2.524	16.829	48.322
3	1.217	8.111	56.433
4	1.208	8.056	64.489
5	0.985	6.566	71.055
6	0.848	5.653	76.708
7	0.752	5.013	81.720
8	0.640	4.264	85.984
9	0.554	3.690	89.674
10	0.509	3.396	93.070
11	0.396	2.643	95.714
12	0.300	2.001	97.715
13	0.198	1.320	99.035
14	0.142	0.947	99.982
15	0.003	0.018	100.000

3.3 模型预测结果分析

基于 MATLAB 2019 平台构建住宅工程造价预测模型,并进行仿真分析。结合住宅工程造价预测过程的特点,经过多次试验,设置了 SSA-LSSVM 的参数,结果见表 5。

表5 SSA-LSSVM的参数设置

Table 5 Parameter settings of SSA-LSSVM

参数	取值
种群规模	20
最大迭代次数	50
$c$	[0.001, 500]
$\sigma$	[0.001, 100]
安全值 $S_T$	0.6
发现者占种群的比例	0.7
预警者占种群的比例	0.2

将样本数据中的40组数据用于训练基于SSA-LSSVM的住宅工程造价预测模型,用余下的10组数据组成测试集。为了减小试验中偶然性和随机性的影响,取10次试验结果的均值作为最终的预测结果。通过计算SSA优化前LSSVM模型单方造价的预测值与实际值的绝对误差和相对误差,可以对SSA的优化效果进行定量评价,具体计算结果见表6。

表6 LSSVM与SSA-LSSVM单方造价预测结果对比

Table 6 Comparison of cost prediction results per square meter between LSSVM and SSA-LSSVM

测试 样本	实际值/ (元·m <sup>-2</sup> )	LSSVM			SSA-LSSVM		
		预测值/ (元·m <sup>-2</sup> )	绝对误差/ (元·m <sup>-2</sup> )	相对误差/ 差/%	预测值/ (元·m <sup>-2</sup> )	绝对误差/ (元·m <sup>-2</sup> )	相对误差/ 差/%
1	1 986.23	2 139.85	153.62	7.73	2 046.68	60.45	3.04
2	1 646.17	1 711.36	65.19	3.96	1 606.95	-39.22	-2.38
3	1 868.37	1 979.35	110.98	5.94	1 954.13	85.76	4.59
4	2 493.40	2 585.07	91.67	3.68	2 534.59	41.19	1.65
5	1 452.33	1 629.08	176.75	12.17	1 487.46	35.13	2.42
6	1 780.42	1 695.49	-84.93	-4.77	1 764.57	-15.85	-0.89
7	2 171.02	2 117.18	-53.84	-2.48	2 147.34	-23.68	-1.09
8	1 732.94	1 572.47	-160.47	-9.26	1 784.42	51.48	2.97
9	1 989.34	2 152.86	163.52	8.22	2 029.13	39.79	2.00
10	1 889.18	1 807.76	-81.42	-4.31	1 835.52	-53.66	-2.84

由表6可知,由LSSVM预测模型得到的单方造价绝对误差的最大值为176.75元/m<sup>2</sup>,最小值为-160.47元/m<sup>2</sup>,相对误差的最大值为12.17%,最小值为-9.26%;由SSA-LSSVM预测模型得到的单方造价绝对误差的最大值为85.76元/m<sup>2</sup>,最小值为-53.66元/m<sup>2</sup>,相对误差的最大值为4.59%,最小值为-2.84%。根据以上数据可知SSA-LSSVM预测模型的预测精度更高,且该模型造价预测的相对误差在±5%以内,远低于实际工程造价快速估算的误差要求,因此该模型可用于住宅工程造价的预测。

同时,LSSVM预测模型相对误差的分布区间为[-9.26%, 12.17%],极差为12.17% - (-9.26%) = 21.43%;SSA-LSSVM预测模型相对误差的分布区间为[-2.84%, 4.59%],极差为4.59% - (-2.84%) = 7.43%。通过上述分析可知,用SSA优化的LSSVM模型的预测精度更稳定。

### 3.4 模型适用性分析

为了进一步验证SSA-LSSVM住宅工程造价预测模型的适用性,本研究对某个实际住宅项目进行了造价预测分析。该项目的特征指标见表7。

表7 某住宅项目特征指标值

Table 7 Characteristic index values of the residential project

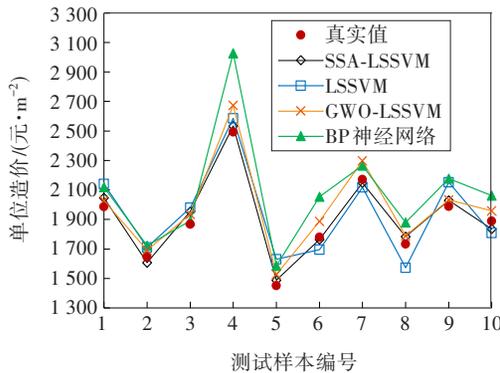
$X_1/d$	$X_2/\text{层}$	$X_3/\text{层}$	$X_4/\text{m}^2$	$X_5/\text{m}$	$X_6/\%$	$X_7$	$X_8$
538	23	1	24 482.75	69	77.36	1	3
$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	单方造价/ (元·m <sup>-2</sup> )
3	3	2	2	3	1	2	1 933.32

将该项目的特征指标值代入已经过训练和测试的SSA-LSSVM造价预测模型中,计算得到该住宅项目工程造价的预测结果为2 009.11元/m<sup>2</sup>,其与实际造价的误差为3.92%,满足前期投资估算的精度要求。因此,SSA-LSSVM工程造价预测模型对样本之外的实际住宅项目同样适用,这也验证

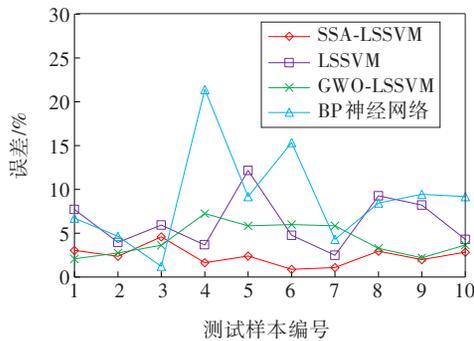
了该模型在住宅工程造价预测方面的适用性。

### 3.5 模型对比分析

为验证 SSA-LSSVM 预测模型的有效性和优越性,本研究用未优化的 LSSVM 模型、用灰狼优化算法 (grey wolf optimizer, GWO) 优化的 LSSVM 模型和 BP 神经网络 3 种模型与该模型进行分析对比。各预测模型的样本数据、输入变量和输出变量均与 SSA-LSSVM 模型的相同。为了和 SSA 保持同步,将 GWO 的参数设置为:狼群的大小为 20 只,最大迭代次数为 50,狼群权重由 2 线性下降到 0;将 BP 神经网络的参数设置为:输入层、隐含层和输出层的节点数分别为 8、7、1,收敛误差为 0.001,训练次数为 1 000。各模型预测结果如图 2 所示。



(a) 单方造价



(b) 预测误差

图 2 各模型预测结果及其误差对比

Fig. 2 Comparison of prediction results and errors of each model

由图 2 可知,本研究所构建的 SSA-LSSVM 模型的预测结果较其他模型更接近实际造价,说明其有较好的预测性能。为了进一步对比模型的预测效果,选用常见的均方根误差  $e_{RMS}$ 、平均绝对百分比误差  $e_{MAP}$  以及相关系数  $R^2$  3 种性能评价指标,对上述模型的预测性能进行对比,结果见表 8。

表 8 不同模型预测性能的对比

Table 8 Comparison of prediction performance of different models

模型	$e_{RMS}$	$e_{MAP}/\%$	$R^2$
LSSVM	122.082	6.25	0.907
SSA-LSSVM	48.377	2.39	0.984
GWO-LSSVM	91.561	4.36	0.952
BP 神经网络	240.138	10.61	0.667

由表 8 可知,SSA-LSSVM 模型的  $e_{RMS}=48.377$  和  $e_{MAP}=2.39\%$ ,都比 LSSVM 模型、GWO-LSSVM 模型和 BP 神经网络的小,且其相关系数  $R^2$  更接近于 1,这说明本研究所构建的 SSA-LSSVM 模型的预测结果具有更高的精度,且能很好地提取输入指标和输出指标之间复杂的非线性关系。综上所述,SSA-LSSVM 预测模型较其他 3 个模型更适用于住宅工程造价的预测。

## 4 讨论

1) 本研究通过文献研究法,并结合工程实例,确定了 15 个住宅造价特征指标作为输入指标,将单方造价作为输出指标。使用主成分分析法对输入数据进行处理,由计算结果确定了 8 个主成分作为住宅工程造价预测模型新的输入数据。然后,将 SSA-LSSVM 模型与 LSSVM、GWO-LSSVM 和 BP 神经网络模型通过均方根误差、平均绝对百分比误差以及相关系数 3 种预测性能评价指标进行对比,发现 SSA-LSSVM 模型具有更高的预测精度和回归预测性能,能很好地满足住宅工程前期造价估算对精度的要求。

2) 以往基于数学统计方法得到的建筑工程造价投资估算的精度欠佳,故本研究提出了由 SSA 与 LSSVM 组成的组合算法 SSA-LSSVM,该算法在电力负荷预测、水质预测和二氧化碳排放预测等领域表现出优异的预测性能,但尚无将 SSA-LSSVM 应用于住宅造价预测的研究。用 SSA 优化 LSSVM 的正则化参数  $c$  和核函数参数  $\sigma$ ,克服了选取 LSSVM 参数的困难,使 LSSVM 具有更快的收敛速度、更强的泛化能力,提高了 LSSVM 对住宅工程造价预测的精度和稳定性。

3) 本研究通过文献研究法确定了影响住宅造价的特征指标,其选取过程存在一定的主观性,且可能会忽略主要特征指标,故特征指标体系的确定

立流程仍有优化空间。

## 5 结论

为快速、准确地对住宅工程造价进行预测,本研究将 SSA 和 LSSVM 相结合,建立了基于 SSA-LSSVM 的住宅工程造价预测模型。在使用该模型时,先使用主成分分析法对住宅样本数据进行降维处理,再将降维处理后的样本数据导入模型进行训练和预测。在工程实践中,该模型能够在小样本情况下提取住宅造价数据之间复杂的非线性关系。仿真试验结果表明:该模型的平均绝对百分比误差仅为 2.39%,满足工程快速投资估算的精度要求,使用该模型可提高工作效率,对实际工程前期的投资估算具有一定的参考意义,可提高投资决策的科学合理性。

未来可加强对不同住宅种类的研究,继续细化住宅建筑的类型,如高层、中高层、低层住宅等,使建筑造价预测模型更具针对性。

### [参考文献]

- [1] 董娜,卢泗化,熊峰.大数据背景下基于 ABC-SVM 的建筑工程造价预测[J].技术经济,2021,40(8):25-32.  
DONG Na, LU Sihua, XIONG Feng. Cost prediction in construction project based on ABC-SVM under the background of big data[J]. Journal of Technology Economics, 2021, 40(8): 25-32.
- [2] ABU HAMMAD A A, ALI S M A, SWEIS G J, et al. Statistical analysis on the cost and duration of public building projects[J]. Journal of Management in Engineering, 2010, 26(2): 105-112. DOI: 10.1061/(ASCE)0742-597X(2010)26:2(105).
- [3] BAYRAM S, AL-JIBOURI S. Efficacy of estimation methods in forecasting building projects' costs[J]. Journal of Construction Engineering and Management, 2016, 142(11): 5016012.1. DOI: 10.1061/(ASCE)CO.1943-7862.0001183.
- [4] KHALAF T Z, ÇAĞLAR H, ÇAĞLAR A, et al. Particle swarm optimization based approach for estimation of costs and duration of construction projects[J]. Civil Engineering Journal, 2020, 6(2): 384-401. DOI: 10.28991/cej-2020-03091478.
- [5] 张虎,颜可珍,朱向平.基于 BP 神经网络的双层体系模量反演研究[J].中外公路,2020,40(6):37-40. DOI: 10.14048/j.issn.1671-2579.2020.06.008.  
ZHANG Hu, YAN Kezhen, ZHU Xiangping. Modulus back-calculation of two-layer system via BP neural network[J]. Journal of China & Foreign Highway, 2020, 40(6): 37-40. DOI: 10.14048/j.issn.1671-2579.2020.06.008.
- [6] 王德美,陈慧,肖之鸿,等.基于数据挖掘的住宅工程造价预测[J].土木工程与管理学报,2021,38(1):175-182. DOI: 10.14048/j.issn.1671-2579.2020.06.008.  
WANG Demei, CHEN Hui, XIAO Zhihong, et al. Prediction of housing project cost based on data mining [J]. Journal of Civil Engineering and Management, 2021, 38(1): 175-182. DOI: 10.14048/j.issn.1671-2579.2020.06.008.
- [7] 李玲玲.基于改进神经网络的建筑工程造价预测[D].天津:天津大学,2019. DOI: 10.27356/d.cnki.gtjdu.2019.002461.  
LI Lingling. Prediction of construction cost based on improved neural network[D]. Tianjin: Tianjin University, 2019. DOI: 10.27356/d.cnki.gtjdu.2019.002461.
- [8] 何利娟.基于灰色理论的建筑工程造价预测和应用研究[D].乌鲁木齐:新疆大学,2020. DOI: 10.27429/d.cnki.gxjdu.2020.000337.  
HE Lijuan. Prediction and application of construction cost based on grey theory[D]. Urumqi: Xinjiang University, 2020. DOI: 10.27429/d.cnki.gxjdu.2020.000337.
- [9] 张晓东,杨圣祥.基于 PCA 与 NARX 的市政工程造价组合预测[J].控制工程,2017,24(12):2485-2490. DOI: 10.14107/j.cnki.kzgc.170785.  
ZHANG Xiaodong, YANG Shengxiang. Forecasting the cost of municipal engineering based on PCA and NARX [J]. Control Engineering of China, 2017, 24(12): 2485-2490. DOI: 10.14107/j.cnki.kzgc.170785.
- [10] KIM S, SHIM J H. Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry[J]. Canadian Journal of Civil Engineering, 2014, 41(1): 65-73. DOI: 10.1139/cjce-2013-0223.
- [11] ZHAO L, ZHANG W, WANG W S. Construction cost prediction based on genetic algorithm and BIM[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(7): 2059026. DOI: 10.1142/S0218001420590260.
- [12] YE D. An algorithm for construction project cost forecast based on particle swarm optimization-guided BP neural network[J]. Scientific Programming, 2021, 2021: 4309495. DOI: 10.1155/2021/4309495.
- [13] 张模蕴,肖国安.基于优化灰色模型的湖南省粮食产量预测方法改进研究[J].湘潭大学学报(哲学社会科学版),2020,44(3):118-122. DOI: 10.13715/j.cnki.jxupss.2020.03.020.  
ZHANG Moyun, XIAO Guoan. Study on the improvement of grain yield prediction method in Hunan Province based on the optimized grey model[J]. Journal of Xiangtan University (Philosophy and Social Sciences), 2020, 44(3): 118-122. DOI: 10.13715/j.cnki.jxupss.2020.03.020.
- [14] 姜海龙,李潼清,程浩,等.基于 PSO-LSSVM 的高压真空断路器弹簧机构机械故障诊断[J].高压电器,2019, 55(12): 248-255. DOI: 10.13296/j.1001-1609.hva.2019.12.037.  
JIANG Hailong, LI Tongqing, CHENG Hao, et al.

- Mechanical fault diagnosis for spring mechanism of high-voltage vacuum circuit breaker based on PSO-LSSVM[J]. *High Voltage Apparatus*, 2019, 55(12): 248–255. DOI:10.13296/j.1001-1609.hva.2019.12.037.
- [15] 杨邓,杨俊杰,胡晨阳,等. 基于改进LSSVM的短期电力负荷预测[J]. *电子测量技术*, 2021, 44(18): 47–53. DOI:10.13296/j.1001-1609.hva.2019.12.037.
- YANG Deng, YANG Junjie, HU Chenyang, et al. Short-term power load forecasting based on improved LSSVM[J]. *Electronic Measurement Technology*, 2021, 44(18): 47–53. DOI: 10.13296/j. 1001-1609. hva. 2019.12. 037.
- [16] ZHAO H R, HUANG G, YAN N. Forecasting energy-related CO<sub>2</sub> emissions employing a novel SSA-LSSVM model: considering structural factors in China[J]. *Energies*, 2018, 11(4): 781. DOI: 10.3390/en11040781.
- [17] XUE J K, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. *Systems Science & Control Engineering*, 2020, 8(1): 22–34. DOI:10.1080/21642583.2019.1708830.
- [18] 全卫国,郭超宇,赵如意. 基于改进麻雀算法优化LSSVM的再循环箱浆液密度预测模型[J]. *电子测量技术*, 2022, 45(1): 70–76. DOI: 10.19651/j. cnki. emt. 2108057.
- TONG Weiguo, GUO Chaoyu, ZHAO Ruyi. Prediction model of slurry density in recycling tank based on LSSVM optimized by improved sparrow algorithm[J]. *Electronic Measurement Technology*, 2022, 45(1): 70–76. DOI:10.19651/j.cnki.emt.2108057.
- [19] 蒋红妍,白雨晴. 基于灰关联的PSO-BP神经网络的高层住宅造价估算[J]. *工程管理学报*, 2019, 33(1): 29–33. DOI:10.13991/j.cnki.jem.2019.01.006.
- JIANG Hongyan, BAI Yuqing. High-rise residential cost estimation based on grey correlation analysis and PSO-BP neural network[J]. *Journal of Engineering Management*, 2019, 33(1): 29–33. DOI: 10.13991/j. cnki. jem.2019.01.006.
- [20] 刘伟军,赵威. 基于RF-PSO-LSSVM的高层建筑项目工期风险预测[J]. *长沙理工大学学报(自然科学版)*, 2021, 18(2): 49–56. DOI: 10.19951/j. cnki. cslgdxzbk. 2021.02.007.
- LIU Weijun, ZHAO Wei. Risk prediction of high-rise building project duration based on RF-PSO-LSSVM[J]. *Journal of Changsha University of Science & Technology (Natural Science)*, 2021, 18(2): 49–56. DOI: 10.19951/j.cnki.cslgdxzbk.2021.02.007.

## Research on residential project cost prediction based on SSA-LSSVM

PENG Junlong, HU Ke, WANG Mengyao, PENG Chao

(School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha 410114, China)

**Abstract:** [Purposes] This paper aims to improve the accuracy of construction cost prediction, and provide an important basis for the early investment decision of the project. [Methods] For the characteristics of high-dimensional nonlinear relationships between small sample data and the characteristic indexes of project cost in engineering practice, a cost prediction model (SSA-LSSVM) based on the sparrow search algorithm (SSA) optimized least squares support vector machine (LSSVM) is constructed. Firstly, the input index data of residential project cost is processed by principal component analysis to reduce data redundancy. Secondly, the SSA is used to optimize the regularization parameter  $c$  and the kernel function parameter  $\sigma$  in the LSSVM model to help determine the parameters of the LSSVM model. Finally, the processed data is imported into the constructed model for training and prediction, and the model prediction performance is evaluated by three metrics, that is correlation coefficient, mean absolute percentage error and root-mean-square error. [Findings] SSA-LSSVM model has better generalization ability and prediction accuracy compared with LSSVM model, LSSVM model optimized by grey wolf algorithm and back propagation neural network. [Conclusions] The model built in this study can accurately and efficiently predict the actual residential project cost, which can provide some reference for the pre-project investment decision.

**Key words:** residential project; cost prediction; principal component analysis; sparrow search algorithm (SSA); least squares support vector machine (LSSVM)

**Manuscript received:** 2022-06-24; **revised:** 2022-08-12; **accepted:** 2022-08-22

**Foundation item:** Project (51578080) supported by the National Natural Science Foundation of China; Project (2021JJ30746) supported by Natural Science Foundation of Hunan Province

**Corresponding author:** PENG Junlong (1976—) (ORCID: 0000-0002-5585-0569), male, associate professor, research interest: project management. E-mail: 375135287@qq.com

(责任编辑:石月珍;校对:刘平;英文编辑:刘至真)