

# 大数据背景下基于ABC-SVM的建筑工程造价预测

董娜, 卢泗化, 熊峰

(四川大学建筑与环境学院, 成都610065)

**摘要:** 建筑工程项目决策阶段信息量少, 精准高效的造价预测是科学决策的关键。为了提高项目前期工程造价预测的精度, 探讨如何利用历史项目大数据及机器学习进行新建建筑工程项目的造价预测至关重要。本文首先通过文献研究确定了建筑工程决策阶段造价的主要影响因素, 然后利用人工蜂群算法(ABC)对支持向量机(SVM)参数即惩罚因子 $C$ 和核函数参数 $g$ 进行优化计算, 最终构建了基于ABC-SVM的建筑工程造价预测模型。最后以某工程造价数据平台上的84个建筑工程项目为数据源进行模型验证, 结果显示, 与GRID-SVM模型和BP神经网络模型相比, 本文所提的ABC-SVM模型的预测精度更高, 具有更好的适用性。

**关键词:** 大数据; 造价预测; 人工蜂群算法; 支持向量机

**中图分类号:** F224.3    **文献标志码:** A    **文章编号:** 1002—980X(2021)08—0025—08

## 一、引言

造价预测是项目可行性研究的重要组成部分(陈一飞等, 2011), 科学高效的造价预测, 可以协助业主进行多方案比选, 实现更加合理的投资决策(Fang和Moon, 2017)。在工程实践中, 项目前期可获取的信息有限, 业主和承包商常根据历史经验确定工程造价, 效率和精度都有待提高(胡文发等, 2018)。为了探讨如何更加准确地预测工程造价, 学者们利用不同方法建立了造价预测模型。这些方法大致可以分为两类: 传统预测方法和机器学习方法。

传统方法必须明确预测变量与自变量之间的关系(Xu和Moon, 2013), 最典型的代表是回归分析, 由于其运用过程简单快捷, 能直观反映变量间数学关系, 被广泛应用于预测领域, 如Lowe et al(2006)利用多元回归模型, 在设计阶段项目信息不充足的情况下进行造价预测, 平均预测误差MAPE为19.3%。马可可等(2021)基于回归分析法构建标准化的自来水厂工程造价模型。陈小龙和王立光(2009)以小高层住宅建筑为例, 利用多元线性回归分析法建立工程造价的快速估算模型。但该方法在不确定因素繁多且样本数据较少时预测效果较差(段鹏, 2008)。为进一步提高工程造价预测精度, 部分学者基于过去积累的大量建筑工程典型案例, 在多元回归模型上融合基于案例推理方法, 如Jin et al(2011)、Jin et al(2014)<sup>1</sup>和Ji et al(2010)学者分别提出了一种基于案例推理(CBR)的建筑工程项目造价估算模型, 通过修正历史相似案例来实现预测, 与多元线性回归模型相比, 虽然该方法在预测精度上有一定程度的提高, 但面临计算类别变量相似度困难的挑战。

近年来, 信息技术的快速发展为工程造价预测提供了信息集成平台, 数据的增长和获取更加快速便捷(荆浩等, 2018)<sup>121</sup>。随着大数据和人工智能技术的不断发展, 基于工程大数据和机器学习, 充分挖掘繁杂数据背后隐藏的规律进行准确、快速的工程造价预测具有重要的现实意义(王德美等, 2021)<sup>176</sup>。机器学习方法中以BP神经网络和支持向量机(SVM)最为广泛应用于造价预测领域。其中, BP神经网络具有很强的自适应、非线性映射能力(凌云鹏等, 2012), 周丽萍和胡振锋(2005)、王成军和左新慧(2010)及孙安黎等(2018)学者提出了基于BP神经网络的工程造价预测模型, 但其需要大量样本数据作为支撑(刘婧和叶青, 2013)且学习过程难以理解。而SVM, 由Vapnik-Chervonenkis(VC)维理论和结构风险最小原理作为理论支撑, 在解决小样本、非线性问题中能表现出许多特有的优势(孙涵等, 2011)<sup>2003</sup>, 能获得好的预测结果, 较好的解决神经

**收稿日期:** 2020—03—02

**项目基金:** 四川省科学技术协会支撑项目“大数据背景下的建设项目智慧评价”(19H0469)

**作者简介:** 董娜, 博士, 四川大学建筑与环境学院副教授, 研究方向: 建筑信息化及工程项目管理; 卢泗化, 四川大学建筑与环境学院硕士研究生, 研究方向: 建筑信息化及工程项目管理; 熊峰, 博士, 四川大学建筑与环境学院教授, 研究方向: 工程结构与BIM技术的应用。

网络存在的问题(肖启荣,2016),备受学者青睐,王德美等(2021)<sup>175</sup>选取SVM构建住宅工程造价预测模型;Fan和Sharma(2021)基于SVM和最小二乘支持向量机(LSSVM)建立工程造价预测模型,研究表明基于SVM的预测模型具有更高的预测精度,结果具有鲁棒性。除此之外,部分学者还分别将SVM应用于能源需求预测及供应链需求预测中(孙涵等,2011;荆浩等,2018),进一步证明该方法可靠性。SVM方法在具体应用中成功的关键在于如何设置影响算法的核心参数(彭光金等,2011)。目前,常用的SVM参数优化方法主要有遗传算法(GA)(Zhao et al,2011)、蚁群算法(ACO)(赵新建等,2011)及粒子群算法(PSO)(Lin et al,2019)。但以上方法在寻优过程中会不同程度陷入局部最优解,无法达到最优预测效果(于明和艾月乔,2012)。人工蜂群(ABC)算法是一种新型群智能优化算法,能在很大程度上避免陷入局部最优解问题(李璟民和郭敏,2015),陈芳等(2021)、Kuo et al(2018)等学者的研究成果表明基于ABC算法优化的SVM模型具有更满意的预测结果。

因此本文旨在充分利用工程大数据的基础上,利用SVM进行建筑工程的造价预测,通过ABC算法进行模型中参数的优化计算,提出基于ABC-SVM的造价预测模型,并将预测结果同基于网格搜索算法优化的SVM模型(GRID-SVM)和BP神经网络模型的预测结果进行比较,寻找基于大数据和机器学习的建筑工程的造价最优预测模型。

## 二、建筑工程项目造价影响因素识别

变量确定是预测模型建立的前提。现有文献表明建筑工程造价的影响因素众多,经过梳理汇总见表1。本文根据影响因素出现频数,同时考虑到不同年份建筑材料价格的变化及电梯数量对工程造价影响的必要性,最终确定建筑工程项目决策阶段造价的主要影响因素为总建筑面积、结构形式、地上层数、户数、梯户比、工程地点及材价时间。

将上述确定的造价影响因素作为变量,可分为连续型、离散型、分类型。在建立建筑工程造价预测模型时,将各影响因素作为自变量,将工程造价作为因变量,探讨如何根据既定的主要影响因素预测工程造价。最后工程造价预测中的所有变量描述见表2。

表1 工程造价影响因素文献统计

作者	工程造价影响因素
李文华和李文玉(1999)	建筑面积、结构特征、层数、首层层高、标准层高、平均户面积、沿高、开间、抗震裂缝、地耐力、基础类型、外墙厚等
秦中伏等(2016)	地上和地下建筑面积、结构类型、地上和地下层数、地上和地下层高(平均)、基础类型、混凝土均价增长率、钢筋均价增长率等
梁喜和刘雨(2017)	建筑面积、标准层面积、结构类型、建筑层数、建筑层高、抗震等级、平面形状、基础类型及埋深等
胡晓娟(2016)	建筑物高度、标准层面积、建筑层数等
Jin et al(2014) <sup>5</sup>	建筑面积、结构类型、层数、户数、电梯数量、架空层数量、户型、走廊类型等
Dursun和Stoy(2016)	建筑面积、层数、地上和地下层数、平均层高等、内外墙面积、土壤情况等
Koo et al(2011)	建筑面积、地上和地下层数、户数、地区、多户住房类型(出租、出售)等
An et al(2007)	建筑面积、层数、单元数、单元面积、地区、屋顶类型等
Ji et al(2019)	建筑面积、结构类型、层数、单元数、层高、屋顶类型等

表2 工程造价预测变量描述

变量名称	变量类型	变量设置	单位
总建筑面积	连续型	$x_1$	m <sup>2</sup>
结构形式	分类型	$x_2$	—
地上层数	离散型	$x_3$	层
户数	离散型	$x_4$	户
梯户比	连续型	$x_5$	—
工程地点	分类型	$x_6$	—
材价时间	分类型	$x_7$	年
工程造价	连续型	$y$	元

注:“—”表示无单位。

## 三、优化的SVM建筑工程项目造价预测模型构建

### (一)SVM理论基础

SVM是近年来提出的一种机器学习算法,它由VC维理论和结构风险最小原理作为理论支撑,可分为支

持向量分类机和支撑向量回归机。支撑向量回归机可以根据有限的样本信息,成功处理回归问题并可推广于预测领域内(孙涵等,2011)<sup>2003</sup>,具有非线性拟合效果好,预测精度高等优点。因此,本文采用支撑向量回归机建立预测模型。

对非线性回归问题,SVM可通过非线性映射将样本数据映射到高维空间中,在高维空间中对数据进行线性回归。高维空间中回归函数如式(1)所示:

$$f(x_j) = w^T \phi(x_j) + b \quad (1)$$

其中: $w$ 为样本的权向量; $\phi(x_j)$ 为输入空间到高维空间的非线性映射; $b$ 为样本的常值偏差。

假设能容忍 $f(x)$ 与 $y$ 之间的最大偏差为 $\varepsilon$ ,仅当 $f(x)$ 与 $y$ 之间的偏差的绝对值大于 $\varepsilon$ 时,才计算损失。因此SVM的优化模型可写为式(2):

$$\begin{aligned} & \min \frac{1}{2} \|w\|^T + C \sum_{j=1}^S (\xi_j + \xi_j^*) \\ & \text{s.t.} \begin{cases} y(x_j) - w^T \phi(x_j) - b \leq \varepsilon + \xi_j \\ w^T \phi(x_j) + b - y(x_j) \leq \varepsilon + \xi_j^* \\ \xi_j \geq 0 \\ \xi_j^* \geq 0 \end{cases} \end{aligned} \quad (2)$$

其中: $S$ 为样本数量; $y(x_j)$ 为第 $j$ 个样本的实际工程造价值; $\xi_j, \xi_j^*$ 均为松弛变量; $C$ 为非负常数,表示对超出误差 $\varepsilon$ 的样本的惩罚程度。

引入Lagrange乘子: $\alpha_j, \alpha_j^* \geq 0, \gamma_j, \gamma_j^* \geq 0$ ,可得到Lagrange函数和其对偶问题,再对其对偶问题求解,可以得到式(3)的回归函数:

$$f(x) = \sum_{j=1}^n (\alpha_j - \alpha_j^*) K(x, x_j) + b \quad (3)$$

其中: $\alpha_j$ 和 $\alpha_j^*$ 为Lagrange乘子; $K(x, x_j)$ 是内积函数,也称为核函数。SVM可以选择不同的核函数,在本文中,选择应用广泛的高斯径向核函数,如式(4)所示:

$$K(x, x_j) = \exp(-g \|x_j - x_k\|^2) \quad (4)$$

其中: $x_j$ 和 $x_k$ 分别为不同的输入数据向量; $g$ 表示核函数的参数,且有 $g > 0$ 。

## (二)人工蜂群算法优化SVM参数

在利用SVM建立建筑工程造价预测模型时,除选择核函数为高斯径向核函数外,还需要确定其中两个重要参数,即惩罚因子 $C$ 和核函数参数 $g$ ,它们的取值将直接影响到模型的预测误差(刘渝根和陈超,2019)。人工蜂群算法(ABC)是模拟蜂蜜的采蜜过程而提出的一种新型智能优化算法(高雷阜等,2016),能有效克服局部最优解,与遗传算法和粒子群算法相比,具有控制参数少的优点(Karaboga和Akay,2009)。因此,为了提高造价预测的精确度,本文将采用ABC对上述两个重要参数进行寻优。

参数寻优的流程具体如下所示:

(1)ABC算法参数设置:蜂群的蜜蜂总数 $NP$ 、食物源数量 $NP/2$ 、最大搜索次数 $limit$ 、最大迭代次数 $max\ Cycle$ 、待优化参数个数 $D(D=2)$ 及待优化参数的上下界,并根据下式(5)随机生成初始可行解 $x_{ij}(i=1, 2, 3, \dots, NP/2; j=1, 2, \dots, D)$ ,同时利用式(6)计算解的适应度。

$$x_{ij} = l_j + \text{rand}(0, 1)(u_j + l_j) \quad (5)$$

$$fit_i(x_i) \begin{cases} 1/[1 + f_i(x_i)], f_i(x_i) \geq 0 \\ 1 + \text{abs}[f_i(x_i)], f_i(x_i) < 0 \end{cases} \quad (6)$$

其中: $\text{rand}(0, 1)$ 表示生成一个在数字0~1之间的随机数; $u_j, l_j$ 分别是 $x_{ij}$ 的上界值和下界值; $f_i(x_i)$ 是第 $i$ 个解的目标函数,在本文中,目标函数为预测结果的平均平方误差; $fit_i(x_i)$ 为第 $i$ 个解的适应值; $\text{abs}$ 表示绝对值符号。

(2)采蜜蜂工作:在开始阶段,采蜜蜂根据式(7)产生一个新食物源 $v_{ij}$ ,并按照式(6)计算新食物源的适应度,若新食物源的适应度优于旧食物源,则用新食物源替代旧食物源;反之,则保留旧食物源。

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{ij}) \quad (7)$$

其中： $x_{ij}$  为新可行解 ( $k=1, 2, 3, \dots, NP/2$ , 且  $k \neq i$ )； $\phi_{ij}$  为  $[-1, 1]$  之间均匀分布的随机数。

(3) 跟随蜂工作：在采蜜蜂完成搜寻过程后，采蜜蜂在招募区将食物源的信息与跟随蜂分享。跟随蜂根据式(8)计算得出的概率  $P_i$  来选择食物源，然后在区间  $[-1, 1]$  内随机产生一个数，若食物源的概率值大于该随机数，则跟随蜂由式(7)在该食物源附近产生一个新食物源，并计算对应的适应度，若新食物源适应度优于旧食物源，则旧食物源将被替代，反之，则保留旧食物源。

$$P_i = \frac{fit_i(x_i)}{\sum_{i=1}^{NP/2} fit_i(x_i)} \quad (8)$$

(4) 侦查蜂工作：在跟随蜂完成搜寻过程后，如某一食物源经过  $limit$  次循环仍然没有被进一步更新，则该食物源会被舍弃，同时该食物源对应的采蜜蜂将转为侦查蜂，侦查蜂按式(5)产生一个新的食物源去代替它。

(5) 若运算迭代次数达到  $max\ Cycle$  或最优适应度满足精度要求，则整个 ABC 算法运算结束，返回的值即为优化后的参数，否则将重复上述计算。

### (三) 基于 ABC-SVM 的建筑工程项目造价预测流程

结合上述 ABC 算法优化 SVM 参数流程，构建基于 ABC-SVM 的建筑工程造价预测模型，其具体步骤如下所示：

(1) 将样本的 80% 作为训练集、剩下的 20% 作为测试集，并利用式(9)、式(10)将自变量与因变量均归一化处理到  $[0, 1]$  之间：

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

$$y'_i = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (10)$$

其中： $x_i, x'_i$  为归一化处理前、后的自变量的值； $y_i, y'_i$  为归一化处理前、后的因变量的值。

(2) 利用 ABC 算法对 SVM 预测模型的惩罚因子  $C$  和核函数参数  $g$  进行寻优计算。

(3) 利用 ABC 算法优化后的 SVM 模型对测试集数据进行预测，并将输出结果进行反归一化处理，即可得到测试集数据的造价预测结果。

## 四、ABC-SVM 建筑工程造价预测模型应用

### (一) 数据来源

工程造价预测模型的数据来源于某工程造价数据平台，从中获取了 2010—2017 年已完工住宅项目，根据数据的完整性，最终整理后得到 84 个有效样本。收集的案例数据中的工程造价皆为项目的竣工结算价格。样本数据概况见表 3，工程造价影响因素取值范围见表 4。

表 3 建筑工程样本概况

样本类型	样本数量	样本工程造价(元)
住宅建筑	84	757780.47~73776140.29

数据来源：<http://djt.datacost.cn/login.html#/>。

表 4 工程造价影响因素取值范围

变量名称	取值范围
总建筑面积	561.05~36715.11
结构形式	框剪结构(1)、剪力墙结构(2)、框架结构(3)、砖混结构(4)、底框砖混结构(5)
地上层数	5~34
户数	12~577
梯户比	0~5
工程地点	成都崇州市(1)、都江堰(2)、高新区(3)、金牛区(4)、锦江区(5)、青羊区(6)、双流区(7)、新都区(8)、遂宁市(9)
材价时间	2009—2017 年

注：本文采用数字编码形式来表示“结构形式”和“工程地点”两个分类变量的取值，取值范围中括号内的数字即为对应取值的编码，如“框剪结构”用数字“1”表示。

## (二) ABC优化SVM参数结果

本文的ABC参数寻优及预测工作均在MatlabR2018上实现,建模过程中采用了台湾林智仁教授开发的Libsvm3.1工具箱。ABC算法具体的参数设置为:蜂群的蜜蜂总数 $NP = 20$ ;食物源数量 $NP/2 = 10$ ;最大搜索次数 $limit = 100$ ;最大迭代次数 $max\ Cycle = 10$ ;待优化参数个数 $D = 2$ 及待优化参数范围为 $[0.0005, 100]$ 。经过ABC优化后,SVM模型的 $C$ 和 $g$ 分别为73.9306和0.0199。

## (三) 模型预测结果分析

将优化后的参数带入SVM预测模型中,由68组训练集案例数据建立造价预测模型,并用剩余16组案例数据进行预测。最终,ABC-SVM模型的训练结果和案例验证预测结果分别如图1和图2所示。图1为训练集预测值与实际值的拟合效果,图2为测试集预测值与实际值的对比。

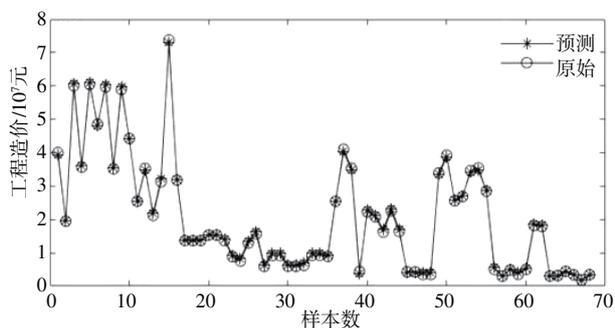


图1 ABC-SVM模型训练集拟合图

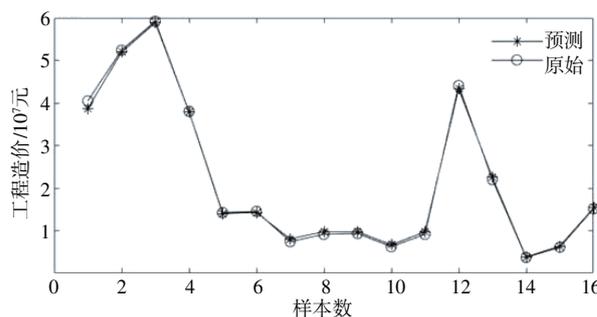


图2 ABC-SVM模型测试集预测结果对比

结合预测结果图1和图2可知,ABC-SVM模型对训练集拟合效果很好,回归拟合度 $R^2 = 0.999$ ,且对测试集的预测效果也很好,16个测试集案例样本中,预测误差范围为0.13%~9.01%,误差在5%以内的数据占75%。由此可以看出,该模型对于建筑工程的造价预测是适用的,且具有很强的预测推广能力,能够有效提高建筑工程造价预测的精确度。

## (四) ABC-SVM模型在实际建筑工程中的应用

为了进一步验证ABC-SVM建筑工程造价预测模型的适用性,本文对工程造价数据平台来源外的一个实际住宅工程项目进行了造价预测分析。该住宅项目位于成都市双流区,具体的项目参数描述见表5。

将该项目自变量参数带入前述已经通过训练和测试的ABC-SVM建筑工程造价预测模型中,预测结果见表6。

由表6可以看出,该住宅项目案例预测结果的误差为5.08%,在验证集的预测误差0.13%~9.01%范围之内,且误差小、精度高。因此,ABC-SVM建筑工程造价预测模型对来源于工程造价数据平台之外的建筑工程项目同样是适用的,且具有很好的预测结果,验证了ABC-SVM模型对于建筑工程造价预测的适用性。

表5 实际案例项目的参数描述

项目名称	总建筑面积	结构类型	地上层数	户数	梯户比	工程地点	材价时间	工程造价
成都市某住宅工程项目	16076.69	2	31	188	3	7	2017	29315299.3

表6 案例预测结果

项目名称	工程造价真实值/元	工程造价预测值/元	误差绝对值	误差相对值
成都市某住宅工程项目	29315299.3	30803579.59	1488280.29	5.08%

## 五、多模型对比

为了对比本文提出的ABC-SVM预测模型的性能,采用libsvm自带参数寻优的GRID-SVM及BP神经网络对建筑工程造价进行预测,预测结果见表7。各预测模型的样本数据、输入变量和输出变量均与ABC-SVM相同。BP神经网络参数设置为:输入层、隐藏层和输出层的节点数分别为7、6、1,收敛误差为0.001,迭代次数为1000。

表 7 三种预测模型结果对比

测试集编号	真实值/元	预测值/元		
		ABC-SVM	GRID-SVM	BP 神经网络
1	40478707.64	38677233.59	37705096.03	37957719.56
2	52395334.09	51878128.25	51834184.72	60135424.05
3	59175993.35	58795029.70	57091425.20	66650479.73
4	38003875.51	37956024.46	41645745.46	37953669.40
5	14248047.91	14031050.19	14874672.93	11770029.23
6	14552385.47	14275889.11	15027426.15	11770029.23
7	7447704.36	8122135.35	8853749.53	7354483.06
8	9157339.98	9766544.47	10175184.30	7732284.32
9	9353193.99	9804544.76	10205479.81	7695910.74
10	6234743.11	6705471.19	7718051.31	5278089.93
11	9158247.13	9767278.87	10175645.86	7732284.32
12	44069788.23	43261592.10	44955790.13	47673265.62
13	21990329.29	22510816.70	22128218.90	23846480.79
14	3732731.18	3857201.01	4601892.27	3119026.96
15	6131785.37	6408796.88	5541255.67	4638160.97
16	15176739.38	15321070.76	15401486.98	11452033.97

为了对 3 种预测模型的结果有更加清晰直观的对比分析,本文利用式(11)、式(12)计算测试集预测结果的均方根误差  $e_{RMSE}$  和平均相对误差  $e_{MAPE}$ ,并通过这两种指标对模型的预测性能进行评价,结果见表 8。

$$e_{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_{i\_predict} - y_{i\_ture})^2}{N}} \tag{11}$$

$$e_{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_{i\_predict} - y_{i\_ture}}{y_{i\_ture}} \right| \tag{12}$$

其中: $N$ 为测试集的样本数量, $y_{i\_ture}$ 为样本的造价真实值, $y_{i\_predict}$ 为样本的造价预测值。

由表 8 可知,ABC-SVM 模型的  $e_{RMSE}=123909.86$  和  $e_{MAPE}=3.59\%$ ,远低于 GRID-SVM 模型、BP 神经网络模型的  $e_{RMSE}$  和  $e_{MAPE}$ ,表明 ABC-SVM 模型的预测结果更接近实际,具有更高的预测精度,同时也表明与另外两个模型相比,ABC-SVM 模型更适用于建筑工程的造价预测。

表 8 三种预测模型性能对比

预测模型	ABC-SVM	GRID-SVM	BP 神经网络
$e_{RMSE}$	123909.86	291376.20	623374.76
$e_{MAPE}$	3.59%	7.83%	13.61%

## 六、结论

决策阶段建筑工程高效精准的造价预测是保证项目顺利实施的关键,以往研究提出了基于各种数学方法的建筑工程造价预测模型,但它们在预测性能上表现欠佳。因此本文提出了一种大数据背景下基于 ABC-SVM 的建筑工程造价预测模型。首先通过文献研究确定了影响建筑工程项目决策阶段造价的主要因素,作为预测模型的输入集。其次,采用 ABC 算法优化 SVM 的相关参数,提出了基于 ABC-SVM 的建筑工程造价预测模型,并利用某工程造价数据平台的数据对模型进行训练。然后,利用测试集对训练好的模型进行验证,并通过一个实际的住宅项目进一步验证了模型的适用性。最后,将 ABC-SVM 模型预测结果与 GRID-SVM 和 BP 神经网络模型的预测结果对比。结果表明:①本文确定的建筑工程项目造价主要影响因素,适用于所有建筑工程项目的前期造价预测;②ABC-SVM 模型的  $e_{RMSE}=123909.86$  和  $e_{MAPE}=3.59\%$ ,远低于 GRID-SVM 模型、BP 神经网络模型的  $e_{RMSE}$  和  $e_{MAPE}$ ,表明 ABC-SVM 模型较其他两个预测模型具有更高的预测精度,更适用于建筑工程项目的造价预测;③ABC-SVM 模型对来源于工程造价数据库平台之外的建筑工程项目同样是适用的,且具有很好的预测结果;④在大数据背景下,随着建筑工程项目造价数据的不断累积,会为 ABC-SVM 模型提供更多的数据支持,将进一步提高模型的适用性和准确性。

### 参考文献

- [ 1 ] 陈芳,程献宝,黄安民,等,2021.基于人工蜂群算法优化 SVM 的 NIR 杉木弹性模量预测[J].林业科学,57(1): 161-168.
- [ 2 ] 陈小龙,王立光,2009.基于建筑设计参数分析模型的工程造价估算[J].同济大学学报(自然科学版),37(8):

- 1115-1121.
- [3] 陈一飞, 唐祥忠, 李芳成, 2011. 基于实例推理的工程造价估算知识重用模型[J]. 计算机应用研究, 28(8): 2937-2940.
- [4] 段鹏, 2008. 建设工程造价预测方法研究[D]. 重庆: 重庆大学.
- [5] 高雷阜, 高晶, 赵世杰, 2016. 人工蜂群算法优化SVR的预测模型[J]. 计算机工程与应用, 52(11): 55-59, 76.
- [6] 胡文发, 常永霞, 何新华, 2018. 建筑工程项目工期的影响因素与预测模型[J]. 土木工程学报, 51(2): 108-117.
- [7] 胡晓娟, 2016. 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[J]. 四川建筑科学研究, 42(4): 142-147.
- [8] 荆浩, 刘垭, 唐金环, 2018. 基于多变量支持向量机的供应链需求预测分析[J]. 系统工程, 36(11): 125-130.
- [9] 李璟民, 郭敏, 2015. 人工蜂群算法优化支持向量机的分类研究[J]. 计算机工程与应用, 51(2): 151-155.
- [10] 李文华, 李玉文, 1999. 基于人工神经网络的建筑工程造价预测方法研究[J]. 中国管理科学(4): 29-34.
- [11] 梁喜, 刘雨, 2017. 基于模糊神经网络的建筑工程造价预测模型[J]. 技术经济, 36(3): 109-113.
- [12] 凌云鹏, 阎鹏飞, 韩长占, 等, 2012. 基于BP神经网络的输电线路工程造价预测模型[J]. 中国电力, 45(10): 95-99.
- [13] 刘婧, 叶青, 2013. 采用BP和RBF神经网络的厦门市工程造价预测模型[J]. 华侨大学学报(自然科学版), 34(5): 576-580.
- [14] 刘渝根, 陈超, 2019. 基于人工蜂群算法优化支持向量机的接地网腐蚀速率预测模型[J]. 电力自动化设备, 39(5): 189-193, 207.
- [15] 马可可, 王祺, 周律, 等, 2021. 标准化水厂建设工程造价模型[J]. 南水北调与水利科技(中英文), 19(1): 191-197.
- [16] 彭光金, 司海涛, 俞集辉, 等, 2011. 改进的支持向量机算法及其应用[J]. 计算机工程与应用, 47(18): 218-221.
- [17] 秦中伏, 雷小龙, 翟东, 等, 2016. 基于SVM和LS-SVM的住宅工程造价预测研究[J]. 浙江大学学报(理学版), 43(3): 357-363.
- [18] 孙安黎, 向春, 伍焯熙, 2018. 基于BP神经网络的输电线路工程造价预测模型研究[J]. 现代电子技术, 41(2): 79-82.
- [19] 孙涵, 杨普容, 成金华, 2011. 基于Matlab支持向量回归机的能源需求预测模型[J]. 系统工程理论与实践, 31(10): 2001-2007.
- [20] 王成军, 左新慧, 2010. 基于BP神经网络的建筑工程估价预测模型[J]. 山西财经大学学报, 32(S2): 327, 330.
- [21] 王德美, 陈慧, 肖之鸿, 等, 2021. 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, 38(1): 175-182.
- [22] 肖启荣, 2016. 基于现代统计学理论的工程造价估算模型研究[J]. 内蒙古师范大学学报(自然科学汉文版), 45(6): 802-805.
- [23] 于明, 艾月乔, 2012. 基于人工蜂群算法的支持向量机参数优化及应用[J]. 光电子·激光, 23(2): 374-378.
- [24] 赵新建, 沈友文, 徐俊, 2011. 基于ACO算法的SVM核函数的参数优化[J]. 计算机工程与科学, 33(10): 126-130.
- [25] 周丽萍, 胡振锋, 2005. BP神经网络在建筑工程估价中的应用[J]. 西安建筑科技大学学报(自然科学版)(2): 262-264, 296.
- [26] AN S H, KIM G H, KANG K I, 2007. A case-based reasoning cost estimating model using experience by analytic hierarchy process[J]. Building and Environment, 42(7): 2573-2579.
- [27] DURSUN O, STOY C, 2016. Conceptual estimation of construction costs using the multistep ahead approach[J]. Journal of Construction Engineering and Management, 142(9): 04016038.
- [28] FAN M, SHARMA A, 2021. Design and implementation of construction cost prediction model based on SVM and LSSVM in industries 4.0[J]. International Journal of Intelligent Computing and Cybernetics, 14(2): 145-157.
- [29] FANG S T, ZHAO T Y, ZHANG Y, 2017. Prediction of construction projects' costs based on fusion method[J]. Engineering Computations, 34(7): 2396-2408.
- [30] JI C Y, HONG T H, HYUN C T, 2010. CBR revision model for improving cost prediction accuracy in multifamily housing projects[J]. Journal of Management in Engineering, 26(4): 229-236.
- [31] JI S H, AHN J, LEE H S, et al, 2019. Cost estimation model using modified parameters for construction projects[J]. Advances in Civil Engineering, 6: 1-10.
- [32] JIN R Z, CHO K M, HYUN C T, et al, 2011. MRA-based revised CBR model for cost prediction in the early stage of construction projects[J]. Expert Systems with Applications, 39(5): 5214-5222.
- [33] JIN R Z, HAN S W, HYUN C T, et al, 2014. Improving accuracy of early stage cost estimation by revising categorical variables in a case-based reasoning model[J]. Journal of Construction Engineering and Management, 140(7): 04014025.
- [34] KARABOGA D, AKAY B, 2009. A comparative study of artificial bee colony algorithm[J]. Applied Mathematics and Computation, 214(1): 108-132.
- [35] KOO C W, HONG T H, HYUN C T, 2011. The development of a construction cost prediction model with improved prediction capacity using the advanced CBR approach[J]. Expert Systems with Applications, 38(7): 8597-8606.

- [36] KUO R J, LIHUANG S B, ZULVIA F E, et al, 2018. Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction[J]. Knowledge and Information Systems, 55(1): 253-274.
- [37] LIN T Y, YI T, ZHANG C, et al, 2019. Intelligent prediction of the construction cost of substation projects using support vector machine optimized by particle swarm optimization[J]. Mathematical Problems in Engineering, 2019: 1-10.
- [38] LOWE D J, EMSLEY M W, HAEDING A, 2006. Predicting construction cost using multiple regression techniques[J]. Journal of Construction Engineering and Management, 132(7): 750-758.
- [39] XU J W, MOON S W, 2013. Stochastic forecast of construction cost index using a cointegrated vector autoregression model [J]. Journal of Management in Engineering, 29(1): 10-18.
- [40] ZHAO M Y, FU C, JI L P, et al, 2011. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes[J]. Expert Systems With Applications, 38(5): 5197-5204.

## Cost Prediction in Construction Project Based on ABC-SVM under the Background of Big Data

Dong Na, Lu Sihua, Xiong Feng

(School of Architecture and Environment, Sichuan University, Chengdu 610065, China)

**Abstract:** The construction project decision-making stage is lack of information, accurate and efficient cost prediction is the key to scientific decision-making. In order to improve the accuracy of pre-project construction cost prediction, it is very important to discuss how to use the big data of historical projects and machine learning to predict the cost of new construction projects. Firstly, the main influencing factors of the cost in the decision-making stage of construction engineering were determined through literature research. Then, the artificial bee colony algorithm(ABC) was used to optimize the support vector machine(SVM) parameters, namely penalty factor  $C$  and kernel function parameter  $g$ . Finally, the construction cost prediction model based on ABC-SVM was built. And then, 84 construction projects from a construction cost data platform were used as data sources for model validation. The results showed that, compared with GRID-SVM model and BP neural network model, the ABC-SVM model has higher prediction accuracy and better applicability.

**Keywords:** big data; cost prediction; artificial bee colony algorithm; support vector machine