

引用格式: 叶恺慧. 基于支持向量回归的变电站工程造价预测 [J]. 南方能源建设, 2024, 11(增刊 1): 100-105. YE Kaihui. Substation engineering cost prediction based on support vector regression [J]. Southern energy construction, 2024, 11(Suppl. 1): 100-105. DOI: [10.16516/j.ceec.2024.S1.15](https://doi.org/10.16516/j.ceec.2024.S1.15).

# 基于支持向量回归的变电站工程造价预测

叶恺慧<sup>✉</sup>

(中国能源建设集团广东省电力设计研究院有限公司, 广东 广州 510663)

**摘要:** [目的]在电力建设精益化投资发展约束下,提升变电站造价精准管控水平成为供电企业日益关注的问题。为了解决目前变电站工程造价影响因素繁多且复杂、工程造价难以准确预测的问题,文章提出一种基于机器学习算法的变电站工程造价预测模型。[方法]首先基于层次分析法、典型项目分析、问卷德尔菲法、相关系数法从历史变电站工程造价的数据中筛选重要造价影响因素,广泛调研并收集变电站工程造价相关数据,形成可供预测模型检验测试的大样本训练集,其次基于交叉验证与贝叶斯优化算法对支持向量回归模型进行关键参数寻优,探索造价误差较小的模型参数,最后利用寻优之后确定的支持向量回归模型进行造价预测并开展实证校验。[结果]结果显示,支持向量回归模型在保障了训练集较好拟合效果的同时,模型的泛化能力更强,在变电站工程造价总结算价格及各分部分项工程费用预测上取得了较好的准确度。[结论]通过本模型方法的运用,能形成对变电站设计阶段造价的科学预测与有效管控能力,可为实现变电站工程造价精准预测提供方法参考。

**关键词:** 造价预测; 变电站工程; 影响因素; 机器学习; 支持向量机

中图分类号: TM7; F426.61

文献标志码: A

文章编号: 2095-8676(2024)S1-0100-06

DOI: [10.16516/j.ceec.2024.S1.15](https://doi.org/10.16516/j.ceec.2024.S1.15)

OA: <https://www.energychina.press/>



论文二维码

## Substation Engineering Cost Prediction Based on Support Vector Regression

YE Kaihui<sup>✉</sup>

(China Energy Engineering Group Guangdong Electric Power Design Institute Co., Ltd., Guangzhou 510663, Guangdong, China)

**Abstract:** [Introduction] Under the constraint of lean investment in power construction, precise cost control of substation engineering has become an increasingly concerned issue for power suppliers. To address current difficulties in cost prediction due to the large number and great complexity of influencing factors, this paper proposes a substation engineering cost prediction model based on machine learning algorithms. [Method] Firstly, important influencing factors were selected from historical substation construction cost data using methods such as the analytic hierarchy process, analysis of typical projects, the Delphi method, and the correlation coefficient. Relevant data on substation engineering costs were collected through extensive investigation to form a substantial training dataset for model validation and testing. Then, key parameters in the Support Vector Regression (SVR) model were optimized using cross-validation and the Bayesian optimization algorithm to minimize prediction errors. Finally, the optimized SVR model was used for cost prediction, and an empirical validation was conducted. [Result] The results show that the SVR model not only demonstrates a robust fit to the training data but also excels in generalizability. It achieves good accuracy in predicting the total settlement prices of substation engineering costs as well as the costs of various sub-projects. [Conclusion] This approach enables scientific forecasting and effective management of construction costs during the substation design phase. It can offer a methodological reference for precise cost predictions in substation engineering projects.

**Key words:** cost prediction; substation engineering; influencing factors; machine learning; support vector machine

2095-8676 © 2024 Energy China GEDI. Publishing services by Energy Observer Magazine Co., Ltd. on behalf of Energy China GEDI.

This is an open access article under the CC BY-NC license (<https://creativecommons.org/licenses/by-nc/4.0/>).

## 0 引言

随着能源需求增长和电力系统扩张,变电站工程更加复杂化与多样化,准确预测变电站工程造价对于项目决策、预算规划和资源配置意义重大<sup>[1]</sup>。

目前关于变电站工程造价预测的研究主要集中在对机器学习算法的优化上,而忽视了在预测模型建立前的重要影响因素筛选阶段,可能导致预测结果的不准确和不可靠。如果没有结合项目的实际情况对影响因素进行分析,再对这些因素进行合理筛选和权重分配,而仅依赖机器学习算法的优化可能会导致模型过于复杂、过拟合或忽略了关键因素的作用,导致模型的泛化能力较差,难以指导实际的造价管控<sup>[2]</sup>。

为解决上述问题,本文章在详细分析项目现状的基础上使用多种方法筛选因素,基于支持向量回归方法对变电站工程的造价进行预测。首先,文章通过分析实际项目的结算资料和借鉴参考文献,确定了影响变电站工程造价的指标体系。针对这一指标体系,综合考虑相关系数法、典型项目分析得到了主要的影响因素。随后,文章利用交叉验证和贝叶斯算法对支持向量回归模型的参数进行寻优,以保证模型具有足够的精度和稳定性。最后,文章基于优化后的支持向量回归模型对变电站工程造价进行了预测。预测结果与造价实际值对比偏差较小,且相较于其他模型综合表现更好,能有效提升配变电站工程造价的预测精度,利于实现造价精准化管控。

## 1 方法与数据

### 1.1 支持向量回归

#### 1) 原理

支持向量回归(Support Vector Regression, SVR)是一种基于支持向量机(Support Vector Machine, SVM)的函数回归方法,其数学基础源于统计学理论,被广泛应用于机器学习领域<sup>[3]</sup>。其核心思想类似于 SVM,旨在通过最小化样本点与超平面之间的总偏差来寻找全局最优解,从而实现结构风险最小化。SVR 方法得出的结果仅与支持向量有关,这意味着仅使用支持向量作为训练样本与使用全部数据进行训练得到的效果是相同的。这种特性使得 SVR 能够利用增量学习方法处理大规模样本数据,同时在样本具

有典型性时,只需相对较少的训练样本即可获得良好的结果。相较于复杂算法如神经网络,SVR 需要的样本量较少,且能有效处理高维数据,避免了神经网络经验风险较高的缺点。在时间序列预测、成本估计和函数逼近等领域,SVR 已经取得了成功的应用。因此,在小样本变电站造价预测等案例中,SVR 是一个合适的选择。

#### 2) 参数选择

支持向量回归的性能受多种因素影响,其中之一是核函数的选择<sup>[4]</sup>。在实践中,线性核函数和高斯核函数应用最为广泛。一般来说,当样本集的特征维度较高且样本数量充足时,使用线性核函数能有效减少庞大的计算开销;而当样本数量较少或特征维度较低时,则通常选择高斯核函数,因其具有较广的映射维度、参数较少且计算简单的优点。支持向量回归的核心思想在于将数据映射到高维空间中,以实现数据在该空间中的线性分割。

在支持向量回归模型中,平衡模型复杂度和训练误差的关键在于惩罚因子  $c$  的选择。而核函数中的  $\gamma$  参数主要反映了训练样本的范围特性。 $c$  和  $\gamma$  的取值直接影响了支持向量回归的学习能力。当  $c$  和  $\gamma$  取值较大时,可能导致模型过度学习,即在训练集上表现出很高的准确性,但在测试集上表现出较大误差;而当  $c$  和  $\gamma$  取值较小时,则可能导致模型欠学习,即在训练集和测试集上都表现出较高的误差<sup>[5]</sup>。因此,为了进一步提升支持向量回归的分类效果并降低预测误差,本文采用贝叶斯优化方法对参数  $c$  和  $\gamma$  进行优化调整。

#### 3) 模型评价指标

在机器学习和数据科学中,模型评价非常重要<sup>[6]</sup>,其好坏决定了算法的应用和可靠性,甚至直接影响商业价值。而 RMSE(Root-Mean-Square Error)是一种常用的模型评价指标,用于衡量实际值与预测值之间的误差。

通常情况下,低的 RMSE 值表示模型具有较高的精度。与其他指标相比, RMSE 能够更好地反映预测值  $y_i^*$  和实际值  $y_i$  之间的误差,因此经常被用来验证模型的准确性,其计算方式如公式(1)所示。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2} \quad (1)$$

本实验中的预测对象为总结算价格、建筑工程费、安装工程费、设备购置费的单位均为百万(元),在数据处理时仅进行单位转换使得所有数据都处于同样数量级水平,在模型训练中虽进行了归一化处理,但在计算均方根误差(RMSE)之前重新对数据进行了反归一化,因此本实验中的 RMSE 的单位为百万(元),实际意义为预测值与实际值之差。

## 1.2 数据

### 1) 数据来源

数据均来自于已结算的变电工程文件。在考虑前期的文献调查结果和实际数据可获得性的基础上,构建了变电工程造价影响因素,并从文件中对相关数据进行了手动收集、整理和分类工作,从而支持后续的影响因素分析以及支持向量回归模型的训练。

### 2) 数据的预处理

数据预处理模块旨在对原始数据进行清洗、归约和降维等操作,以消除干扰因素,提取关键特征,从而获得更高质量的数据集用于模型训练。这样可以提升回归预测效果,并确保在预测时采用相同的处理方式,以保证模型的一致性。通过数据预处理,可以提高机器学习模型的准确性和稳定性,同时缩短训练时间,提高效率和经济性。数据预处理的主要操作包括数据清洗、特征处理和数据集划分等步骤。

由于变电站项目的原始数据主要是人工收集整理,因此存在大量的漏填、错填等问题。为了解决这些问题,首先需要补缺失值、异常值和无效值进行处理,并制定一套处理标准。我们可以采用程序化的方式自动处理这类数据。对于缺失值,我们需要根据情况进行处理。例如,对于一些特征如出线回路数等,由于本来就存在大量为 0 的数据,当这些特征缺失时,我们默认将其补充为 0;对于全站面积等类数据以较为齐全的近似数据如全站建筑面积作为填补值;对于本期主变台数和本期主变价格等可以合并的数据做合并处理得到本期主变总价。对于每个特征,根据历史数据设定上下限阈值(针对数值型数据)或是数据格式(针对非数值型数据),通过自动判断剔除不符合要求的异常值和无效数据。

为了满足模型输入的定量数值要求,需要对非数值型数据进行转换。在原始数据中,存在大量的中文数据,例如地形地貌等,这些数据无法直接用于

模型处理。为了解决这个问题,我们可以设定映射表,将非数值数据替换为对应的数值或等级,同时尽可能地保留原始数据之间的相关性<sup>[7]</sup>。

在对各个特征数据进行实际分析后,我们发现大部分数据的数量级都可以控制在 100 以内。为了避免模型计算时出现病态情况,我们对数值较大的数据进行量纲转换。例如,将结算价格从以万为单位转换为以百万为单位,将全站面积从以平方米为单位转换为以百平方米为单位。

在机器学习中,为了能够训练出一个泛化性能良好的模型,一般需要将已有的数据划分为训练集、验证集和测试集 3 部分。考虑本课题获取的变电站造价数据有限,采取简单随机划分,将数据划成 80% 训练集和 20% 测试集,并启用交叉验证,从有限的数据库中尽可能挖掘多的信息,避免出现局部的极值。

## 2 实证分析

本文实验部分主要使用 MATLAB 软件进行代码实现和图像绘制,其中参数的寻优使用 MATLAB 中的回归学习器完成,所有的模型都在 windows 环境下进行训练和测试。实验选用的样本为南方某市供电局竣工时间在 2014—2022 年间 110kV 以及 220kV 的真实变电站工程历史数据,在去除了部分异常样本后共有 13 个样本,主要为新建主变工程。为了取得最佳的预测效果,我们在针对不同的预测目标保留了不同的指标用于造价试预测,经过模型的实际测试,得到了用于造价预测的具体指标。

本文通过对比多种算法的结果得到了最适合的寻优方法。具体的,本文使用 MATLAB 软件进行参数寻优设置,将预处理好的变电站工程造价数据导入到软件当中,选择支持向量回归模型,设置基础训练参数。将数据导入后选择响应变量和解释变量,选择验证方案“留出法验证”,留出百分比为 20%。在将解释变量、被解释变量设置好后,选择可优化 SVM,打开参数寻优界面,分别选择贝叶斯优化、网格搜索和随机搜索 3 种方法,为了防止寻优时间过长,设置迭代次数为 150 次,网格分区数量使用默认值,不同算法的寻优结果如表 1 所示,最终选择了误差和训练时间都更好的贝叶斯优化进行寻优。

表 1 算法寻优结果对比

Tab. 1 Comparison of algorithm optimization results

寻优算法	R方	训练时间/s	RMSE
贝叶斯优化	0.944 81	91.821	2.425 7
网格搜索	0.264 6	1 858.6	8.854 5
随机搜索	0.943 87	56.721	2.446 2

### 3 结果和分析

#### 3.1 支持向量回归结果

在 Matlab 中使用回归学习器训练支持向量回归模型用于变电站工程造价的预测,通过贝叶斯优化进行寻优,在训练 10 次模型后我们发现模型的误差较为稳定,如表 2 所示。不同编号的模型结果差距不大, RMSE 均在 0.864 5 左右(即 86 万元造价偏差),预测误差较小。

表 2 支持向量回归模型结果

Tab. 2 Results of SVR model

模型编号	R方	模型类型	RMSE
1	0.998 74	SVM	0.555 52
2	0.997 03	SVM	0.853 2
3	0.995 48	SVM	1.052 6
4	0.996 24	SVM	0.960 23
5	0.998 13	SVM	0.676 76
6	0.995 62	SVM	1.037 2
7	0.995 21	SVM	1.084 4
8	0.997 24	SVM	0.822 17
9	0.996 39	SVM	0.941 43
10	0.998 21	SVM	0.662 12

以 RMSE 相对最小的模型 1 为例,本文使用测试集数据进行了预测,对比造价预测值与实际值。通过对某市如意(新风)输变电、110 kV 雅宝输变电两个实际工程的总结算进行预测,可以看到其总结算预测造价值分别为 4 465 万元和 4 449 万元,而实际工程的真实造价分别为 4 387 万元和 4 440 万元,预测偏差仅为 1.8% 和 0.2%,表明整体的预测结果与实际造价较为接近,预测效果较好。

#### 3.2 分项造价预测模型误差分析

上文对构建了以工程结算为目标的总造价预测

模型,部分将以总结算的各个分项,包括建筑工程、安装工程、设备购置费、其他费用 4 项为预测目标,运用同样流程分别构建分项预测模型,各分项预测模型的误差结果如表 3 所示。

表 3 分项预测误差对比(百万元)

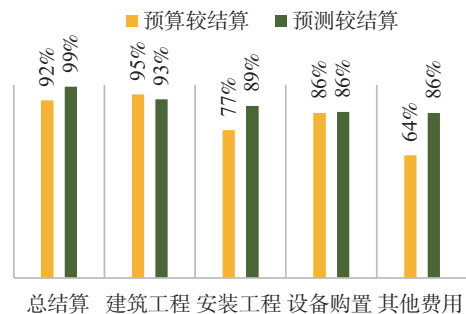
Tab. 3 Comparison of prediction errors by Item/RMB 1,000,000

预测目标	RMSE
总结算	0.556
建筑工程	1.156
安装工程	0.766
设备购置	1.879
其他费用	1.785

从表 3 中可以发现,各分项预测模型在测试集样本上的预测值和实际值之间的平均差距基本都较总结算偏大,其中建筑工程和设备购置的误差相对安装工程和其他费用偏大,可能由于分项在造价中的占比不同。

#### 3.3 结果综合准确率对比

为了比较支持向量预测模型的准确性,我们计算了对应测试集样本中总体竣工结算价、建筑工程费、安装工程费、设备购置费和其他费用的预算较竣工结算的平均准确率(即预算准确率=1-(预算-结算)/结算),并将其与支持向量回归模型的准确率(即模型预测准确率=1-(模型预测值-结算)/结算)进行对比,结果如图 1 所示。可以发现,本报告预测模型对总结算的准确率较初设概算提高了约 7%,同时在安装工程、其他费用的预测准确率也有不同程度的提高,虽然建筑工程、设备购置的准确率并无提升,但从整体来看模型预测取得了较好的效果。因



注:项目结算、概算数据来自项目总结算报告,样本为对应模型中的测试集样本。

图 1 模型预测准确率对比

Fig. 1 Comparison of model prediction accuracies

此,基于预测模型的造价预测结果可以作为项目初设阶段进行精准造价管控的参考依据。

#### 4 结论

本文针对南方某市实际变电站工程造价数据超额的问题,筛选了主要影响工程造价的因素,并将其作为造价预测的参考。鉴于可获得的数据量较少且存在较大的不确定性,本文选择了支持向量回归模型来进行造价的预测。总的来说,主要内容如下:

首先为了体现模型预测的准确度,我们进行了多元线性回归和 BP 神经网络的拟合,并将两种模型在训练集和测试集上的表现与支持向量回归的结果进行比较,结果显示,虽然另外两种方法在训练集上的拟合效果更好,但支持向量回归模型在保障了训练集较好拟合效果的同时,在测试集上的表现好于另外两种方法,模型的泛化能力更强,更具备实际意义。随后本文进行了支持向量回归模型的构建,对模型构建的关键步骤如预测因素的确定、数据的预处理、参数的寻优和模型的导出进行了详细说明。最后,我们对模型的预测结果进行了分析,对总结算价格的分项进行预测,并比较了预测的准确率,基于支持向量回归模型的准确率均得到了不同程度的提升,预测取得了较好的效果,有助于项目造价的精准管控。

该研究存在以下不足:首先,由于变电站造价数据搜集困难,导致本文选取的样本量相对较小,一定程度上影响了模型的可推广性。其次,指标数据的选取不够全面,需要深入挖掘影响造价合理性的因素,以扩充和完善数据集,从而为模型的预测提供更全面的数据基础。因此,接下来的研究可以在以下 3 个方面进行提高和完善:

1)增加样本容量:样本容量的大小直接影响模型的可推广性和结论的可靠性。因此,需要进一步搜集和整理已有的变电站造价资料,充分挖掘可获得的数据,以扩充变电站造价数据的样本容量。

2)完善指标选择:指标选择的全面性和合理性是构建预测模型的基础,也是进行因素分析的前提。因此,需要深入挖掘现有资料中的有价值的信息,并采用适当的模型方法进行指标的优选和组合,以提高变电站工程造价预测的准确性。

3)改进支持向量回归模型及其参数选择方式:通过改进该模型与参数选择方式,以减少计算量、提高参数选择的效果。可以尝试更换核函数,或者尝试其他模型,并引入随机噪声等方式来增强模型的泛化能力。同时,可以进行多种模型的结果对比,以得到最优的预测模型。

#### 参考文献:

- [1] 高观浩,王雨佳,黄佳雯,等. 电力工程造价精准管控大数据平台构建研究 [J]. *建筑经济*, 2023, 44(增刊 2): 270-274. DOI: 10.14181/j.cnki.1002-851x.2023S2270.  
GAO G H, WANG Y J, HUANG J W, et al. Research on construction of big data platform for accurate cost control of power engineering [J]. *Construction economy*, 2023, 44(Suppl.2): 270-274. DOI: 10.14181/j.cnki.1002-851x.2023S2270.
- [2] 刘景江,郑畅然,洪永森. 机器学习如何赋能管理学研究?——国内外前沿综述和未来展望 [J]. *管理世界*, 2023, 39(9): 191-215. DOI: 10.19744/j.cnki.11-1235/f.2023.0106.  
LIU J J, ZHENG C R, HONG Y M. How can machine learning empower management research?-a domestic-foreign frontier review and future prospects [J]. *Journal of management world*, 2023, 39(9): 191-215. DOI: 10.19744/j.cnki.11-1235/f.2023.0106.
- [3] 周晓剑,顾翔. 基于贝叶斯支持向量回归机的稳健参数设计 [J]. *统计与决策*, 2023, 39(24): 23-28. DOI: 10.13546/j.cnki.tjyjc.2023.24.004.  
ZHOU X J, GU X. Robust parameter design based on Bayesian support vector regression machine [J]. *Statistics & decision*, 2023, 39(24): 23-28. DOI: 10.13546/j.cnki.tjyjc.2023.24.004.
- [4] 丁世飞,张子晨,郭丽丽,等. 孪生支持向量回归机研究进展 [J]. *电子学报*, 2023, 51(4): 1117-1134. DOI: 10.12263/DZXB.20220703.  
DING S F, ZHANG Z C, GUO L L, et al. Survey on twin support vector regression [J]. *Acta electronica sinica*, 2023, 51(4): 1117-1134. DOI: 10.12263/DZXB.20220703.
- [5] 梁皓,孙丽,陈姝莽,等. 基于支持向量机算法的地震事件分类研究——以东北地区为例 [J]. *地球物理学报*, 2023, 66(12): 5030-5040. DOI: 10.6038/cjg2022Q0829.  
LIANG H, SUN L, CHEN S Q, et al. Research on seismic event classification based on SVM algorithm: an application in Northeast China [J]. *Chinese journal of geophysics*, 2023, 66(12): 5030-5040. DOI: 10.6038/cjg2022Q0829.
- [6] 王荣颖,卞鸿巍,刘文超. 基于 RMSE 的惯导系统随机误差影响分析 [J]. *海军工程大学学报*, 2017, 29(6): 18-23,27. DOI: 10.7495/j.issn.1009-3486.2017.06.004.  
WANG R Y, BIAN H W, LIU W C. Analysis of stochastic errors

impact in SINS based on RMSE [J]. *Journal of Naval University of Engineering*, 2017, 29(6): 18-23,27. DOI: 10.7495/j.issn.1009-3486.2017.06.004.

[7] 李颜平, 吴刚. 基于典型数据集的数据预处理方法对比分析 [J]. *沈阳工业大学学报*, 2022, 44(2): 185-192. DOI: 10.7688/j.issn.1000-1646.2022.02.11.

LI Y P, WU G. Comparative analysis of data preprocessing methods based on typical data set [J]. *Journal of Shenyang University of Technology*, 2022, 44(2): 185-192. DOI: 10.7688/j.issn.1000-1646.2022.02.11.

作者简介:



叶恺慧

叶恺慧 (通信作者)

1990-, 女, 硕士, 经济师, 从事电力技术经济专业方面工作(e-mail) yekaihui1127@163.com。

(编辑 赵琪)

# 广 告

- 封面图片：“山东枣菏高速公路交能融合（源网荷储一体化）示范工程”----- 封一
- “南方能源建设”微信订阅号----- 封二
- 中国能源建设集团广东省电力设计研究院有限公司氢能技术中心----- 封三
- 中国能源建设集团广东省电力设计研究院有限公司----- 封四