

8-20-2020

A combinational forecasting method for predicting the cost of an overhead line reconstruction project

Min YU

Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co , Ltd , Hangzhou 310000 , China

Yuanxiang WANG

College of Electrical Engineering, Zhejiang University, Hangzhou 310027 , China

Yuan YAN

College of Electrical Engineering, Zhejiang University, Hangzhou 310027 , China

Xiaoyong YANG

Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co , Ltd , Hangzhou 310000 , China

Xiaohong XIA

Zhejiang Zhetianyun Engineering Cost Consultation Co , Ltd , Hangzhou 310000 , China

See next page for additional authors

Follow this and additional works at: <https://jepst.researchcommons.org/journal>

Recommended Citation

YU, Min; WANG, Yuanxiang; YAN, Yuan; YANG, Xiaoyong; XIA, Xiaohong; and WEN, Fushuan (2020) "A combinational forecasting method for predicting the cost of an overhead line reconstruction project," *Journal of Electric Power Science and Technology*. Vol. 35: Iss. 1, Article 3.

DOI: 10.19781/j.issn.16739140.2020.01.003

Available at: <https://jepst.researchcommons.org/journal/vol35/iss1/3>

This Article is brought to you for free and open access by Journal of Electric Power Science and Technology. It has been accepted for inclusion in Journal of Electric Power Science and Technology by an authorized editor of Journal of Electric Power Science and Technology.

A combinational forecasting method for predicting the cost of an overhead line reconstruction project

Authors

Min YU, Yuanxiang WANG, Yuan YAN, Xiaoyong YANG, Xiaohong XIA, and Fushuan WEN

架空线路改造工程造价的组合预测方法

俞 敏¹, 王愿翔², 闫 园², 杨小勇¹, 夏晓红³, 文福拴²

(1. 国网浙江省电力有限公司经济技术研究院, 浙江 杭州 310000; 2. 浙江大学电气工程学院, 浙江 杭州 310027;
3. 浙江浙天工程造价咨询有限公司, 浙江 杭州 310000)

摘 要: 架空线路改造工程造价预测是项目管控的重要环节, 为提高其预测精度, 提出一种组合预测方法。首先, 将架空线路改造工程分解为一些子工程, 分别预测各子工程造价, 之后进行集成。接着, 利用主成分分析法并结合专家经验, 筛选出影响子工程造价的关键因素, 之后, 采用基于遗传算法优化的支持向量机和极限梯度提升算法对子工程造价分别进行预测。然后, 借鉴博弈论中的 Shapley 值理论确定组合预测模型中的权重, 得到组合预测模型。最后, 用实际项目数据对所提出的组合预测方法进行验证, 仿真结果表明, 与采用单个预测模型相比, 所构造的组合预测模型获得了更高的预测精度。

关 键 词: 架空线路改造工程; 造价预测; 组合预测; 基于遗传算法的支持向量机; 极限梯度提升算法; Shapley 值
DOI: 10.19781/j.issn.1673-9140.2020.01.003 中图分类号: TM93 文章编号: 1673-9140(2020)01-0024-07

A combinational forecasting method for predicting the cost of an overhead line reconstruction project

YU Min¹, WANG Yuanxiang², YAN Yuan², YANG Xiaoyong¹, XIA Xiaohong³, WEN Fushuan²

(1. Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310000, China;
2. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China;
3. Zhejiang Zhetianyun Engineering Cost Consultation Co., Ltd., Hangzhou 310000, China)

Abstract: The cost prediction of an overhead line reconstruction project represents an important part of management and control of the project, and to improve the prediction accuracy, a combinational forecasting method is presented. An overhead line reconstruction project is first decomposed into several subprojects depending on the characteristics of the project, and the cost prediction could be made first for each subproject and then integrated. The principal component analysis and expertise of domain experts are combined to get the key factors of the subprojects. The genetic algorithm based support vector machine and extreme gradient boosting algorithm are then used for cost prediction respectively. The Shapley value theory in game theory is next employed to determine the weights of combinational forecasting so as to attain an appropriate combinational forecasting model. Finally, an actual overhead line reconstruction project is employed to demonstrate the presented combinatorial forecasting model, and more accurate prediction result

is attained, compared with those attained by the genetic algorithm based support vector machine and extreme gradient boosting system independently.

Key words: overhead line reconstruction project; cost prediction; combinational forecasting; genetic algorithm based support vector machine; extreme gradient boosting algorithm; Shapley value

提高架空线路改造工程造价的预测精度,确保相关工程效益,是以投资效益为导向的管理理念的重要体现。随着国家电网公司“走出去”的战略实施,需要建立面向国际化的更精准的架空线路改造工程造价预测方法。传统的根据定额概预算确定工程造价的方法无法充分利用历史数据,已经不能满足市场经济发展的需求;尤其是在项目中后期阶段,造价管理部门很难通过定额概预算方法来判别项目设计方案以及施工技术的先进性和合理性。因此,有必要对架空线路改造工程造价预测进行研究。

在电力工程造价预测方面,已有一些研究报道,文献[1]采用多层前馈神经网络对输电线路造价进行预测,将用文字表述的影响因素进行量化处理,分析了隐含层中的神经元数量对模型预测精度的影响;文献[2]利用模糊聚类方法对样本进行分类,将具有相同规律的历史样本归为一类,同时应用基于粒子群优化的支持向量回归模型进行电力工程造价预测;文献[3]采用数据挖掘技术对影响输电工程造价的因素进行偏相关分析,应用 K-means 聚类方法进行降噪分析,最后利用支持向量机对输电工程造价进行预测;文献[4]采用帕累托分析方法对影响 220 kV 变电工程造价的因素进行分类,对细分后的独立影响因素进行多元线性回归分析从而对造价进行预测。文献[1-4]虽采用不同的方法进行电力工程造价预测,但均采用了单一的预测模型,没有综合利用各种预测模型的优点实现组合预测。基于机器学习中的集成方法,建立组合预测模型,可以避免各个单一预测模型的不足,取得更准确的预测结果。在组合预测模型中,如何适当确定各种预测方法的权重至关重要,在这方面已有一些研究报道。文献[5-6]采用层次分析法确定权重;文献[7]根据均方误差、平均绝对误差以及平均绝对相对误差这 3 个预测误差评估指标分别确定组合预测模型的权系数向量,再对这些权系数应用层次分析法确定最终权系数。上述文献采取的层次分析法主观性较强。文献[8]采用粗糙集理论,用模糊隶属度来模拟预测模

型的权重;文献[9]采用拟合优度赋权方法,基于预测结果与实际值的吻合程度确定权重;文献[10]采用熵权法,某种预测模型的预测误差序列的变异程度越大,则赋予该预测模型的权重越小,以此为基础构造了预测模型;文献[11]借鉴博弈论中的 Shapley 值理论,依据 Shapley 值将组合预测模型的误差分配给相关模型,从而确定权重。

在上述背景下,参考文献[11]中的方法,采用 Shapley 值确定组合预测模型中各个预测模型的权重,构建架空线路改造工程造价组合预测模型,以期取得比较准确的预测结果。

架空线路改造工程可以先分解为交、直流输电工程,之后再细分为土方工程、基础工程、架线工程、杆塔工程及附件工程;可通过先对各子工程进行造价预测,之后再集成得到总输电工程的造价预测结果。该文以架空线路改造工程中的附件工程为例,首先利用主成分分析法(principal components analysis, PCA)确定附件工程的影响因素权重并排序,并结合专家经验采用主客观相结合的方式筛选出关键影响因素。随后,采用基于遗传算法的支持向量机(genetic algorithm based support vector machine, GA-SVM)和近年来提出的极限梯度提升(extreme gradient boosting, XGBoosting)机器学习算法分别对工程造价进行预测。之后,采用 Shapley 值理论确定权重系数,进而得到组合预测模型。最后,用实际架空线路改造工程造价数据对所提组合预测方法进行验证,仿真结果表明与采用单一预测模型相比,采组合预测模型可以获得更高的预测精度。

1 关键影响因素的确定

影响架空线路改造工程造价的因素繁多,且与造价之间的相关性程度难以直接观察。在构建的造价预测模型中,如果包括的因素太多,预测精度未必高,甚至有可能导致预测精度下降。因此,有必要对

影响因素进行分析,从中选取关键因素。通过对各个影响因素与造价的相关性分析,可得到其对造价的影响程度,在此基础上再结合专家经验,即可选出关键影响因素。采用这种主客观相结合的方法确定关键影响因素,可以降低数据噪声,弥补架空线路改造工程样本较少的问题,从而取得更准确的预测结果。

作为常用的多元统计分析方法,主成分分析(PCA)方法的基本思想是用一组新的综合指标来代替原来具有一定线性相关性的多个指标,使这些新的综合指标尽可能多地反映原指标的信息且彼此间互不相关^[12]。用PCA筛选指标时,以主成分的方差贡献率为指标权重,对各主成分线性组合中的系数进行加权平均归一化处理,也可理解为求取综合评价函数的过程。该文采用著名的统计产品与服务解决方案(statistical product and service solutions,SPSS)软件来实现。下面说明确定权重的过程。

1) 确定决策矩阵系数。

利用SPSS软件可得到初始因子载荷系数 a_{ij} 与各主成分的特征值 λ_j ,与决策矩阵系数 u_{ij} 关系为

$$u_{ij} = \frac{a_{ij}}{\sqrt{\lambda_j}}, i=1,2,\dots,n; j=1,2,\dots,m \quad (1)$$

式中 n 为原指标个数, m 为降维后得到的主成分个数。

各主成分 F_j 与原始指标 w_i 的关系可描述为

$$\begin{cases} F_1 = u_{11}w_1 + u_{12}w_2 + \dots + u_{1n}w_n \\ F_2 = u_{21}w_1 + u_{22}w_2 + \dots + u_{2n}w_n \\ \vdots \\ F_m = u_{m1}w_1 + u_{m2}w_2 + \dots + u_{mn}w_n \end{cases} \quad (2)$$

2) 确定权重系数。

构建综合评价函数:

$$F_Z = \sum_{j=1}^m \frac{\sigma_j}{\sum_{j=1}^m \sigma_j} F_j = \sum_{j=1}^m \frac{\sigma_j}{\sum_{j=1}^m \sigma_j} \sum_{i=1}^n u_{ji}w_i = \sum_{i=1}^n K_i w_i \quad (3)$$

式中 σ_j 为第 j 个主成分的方差贡献率, K_i 为原始指标 w_i 的指标权重。

归一化后即可得到最终权重系数 K'_i :

$$K'_i = \frac{K_i}{\sum_{i=1}^n K_i} \quad (4)$$

2 造价预测模型/方法

2.1 基于遗传算法优化的支持向量机(GA-SVM)

支持向量机(support vector machine, SVM)是在统计学习理论范畴内发展起来的一种机器学习方法。SVM是1964年提出的,在20世纪90年代后得到了快速发展,并逐步衍生出了一系列的改进和扩展算法,可以有效求解非线性回归和模式识别等诸多问题。

SVM以统计学理论的VC维(vapnik-chervonenkis dimension)理论和结构风险最小化为原则,通过特征映射将在低维特征空间线性不可分的样本集映射到线性可分的高维特征空间^[13]。SVM利用核函数使计算在输入空间完成,避免了高维空间的内积运算,降低了模型复杂性,且具有良好的泛化能力和全局寻优能力。架空线路改造工程样本数量有限,关键影响因素与造价之间的关系复杂,而SVM模型适于求解非线性、高维及小样本输入的预测问题。

SVM模型的预测精度在很大程度上取决于误差惩罚参数 C 和核参数 g 的选择。 C 用于平衡样本的错分误差和算法复杂度之间的关系,通过调整置信范围,以最大化模型的泛化能力^[14]。 g 对映射函数有影响,并与特征空间维度相关,其可改变数据在高维空间中的分布特性。该文采用遗传算法^[14]优化 C 和 g ,以提高模型预测精度。

2.2 极限梯度提升机器学习算法(XGBoosting)

XGBoosting是一种基于梯度提升决策树(gradient boosting decision tree, GBDT)的集成学习算法^[15],通过对一组弱分类器的迭代计算实现准确分类。

在传统的集成学习算法中,通过改变训练样本的权重,逐步改善多个分类器的性能,最终获得最优分类器。在每一轮训练结束后,降低被正确分类的训练样本权重,增大被错误分类的样本权重,经过多次训练后,可得到多个简单的分类器,通过对这些分类器进行组合获得最终模型。XGBoosting在传统集成学习算法的基础上,利用中央处理器的并行计算能力,能够高效处理稀疏数据,灵活实现分布式并行计算,同时还在目标函数中引入正则化项以控制

模型的复杂度,防止出现过拟合风险。采用这些措施后,在保证计算精度的前提下,明显提升了计算速度。

XGBoosting 以分类和回归树(classification and regression tree, CART)作为基分类器。含有 K 个决策树的 XGBoosting 的树集成模型可描述为

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (5)$$

式中 $\hat{y}_i$ 为样本 i 的目标变量(因变量)的预测值, x_i 为相对应的输入变量(自变量), f_k 为第 k 个决策树的预测函数。

树集成优化模型可描述为

$$\text{Min } G = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (6)$$

式中 $l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$ 为损失函数,即均方误差; $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$ 为正则化项,其中 γ 为复杂度参数, T 为树中叶子节点个数, λ 为固定系数, ω 为叶子节点量化权重向量。

正则化项用于控制模型的复杂度,以防止过拟合问题。

求解式(6)所描述的优化问题,就是求解所有 CART 树的结构。由于对所有树进行训练难度较大,文献[16]提出了一种新的思路:保留已训练好的树模型不变,每次添加一个新的树到模型中,通过迭代计算最终得到目标量的预测值。具体求解公式:

$$\begin{aligned} \hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\ &\dots \\ \hat{y}_i^{(t)} &= \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \end{aligned} \quad (7)$$

式中 $\hat{y}_i^{(t)}$ 为 t 轮时目标量的预测值; $\hat{y}_i^{(t-1)}$ 为保留的 $t-1$ 轮的目标量的预测值; $f_t(x_i)$ 为第 t 轮加入的新的预测函数。

在实际应用时,可采用优化方法如贪心算法确定每个新子树的分割节点,选取 $f_t(x_i)$ 函数加入到目标函数中,并求取最小化目标函数的优化解,至此就完成了对预测模型的训练。

3 组合预测模型

基于遗传算法的支持向量机(GA-SVM)和极限梯度提升(XGBoosting)算法各有特点,二者均为基于统计理论的机器学习方法。相比而言,GA-SVM 具有更好的泛化能力与全局寻优能力,也更适合处理高维问题。不过,对于多元回归问题,GA-SVM 对于数据样本规模很大的问题计算效率不高^[17]。XGBoosting 通过采用优化算法,降低了问题的计算复杂度,且借助引入正则项在相当程度上可避免模型过拟合问题,在计算速度和处理大规模数据集方面具有明显优势,已在很多领域得到应用。

在上述背景下,为综合利用 GA-SVM 和 XGBoosting 2 种方法的优点,该文构建基于这 2 种方法的组合预测模型,并借鉴博弈论中的 Shapley 值理论来确定组合预测模型中二者的权重。

Shapley 值理论最早用于研究合作博弈中的利益分配问题。假设集合 N 包含 n 个成员,每个成员应得的利益分配可由 Shapley 值确定,计算公式为

$$\varphi_i = \sum_{\{S|i \in S\}} w(|S|) [V(S) - V(S \setminus \{i\})] \quad (8)$$

式中 S 为集合 N 的含有第 i 个成员的子集;

$$w(|S|) = \frac{(n-|S|)! (|S|-1)!}{n!}; |S| \text{ 为集合 } S \text{ 所包含的参与者个数}; V(S) \text{ 为集合 } S \text{ 的收益}; V(S \setminus \{i\})$$

为集合 S 除去成员 i 以外的收益。

在确定组合预测模型中的权重时,借鉴 Shapley 值理论在收益分配中的应用,视误差为收益,确定组合预测模型总误差在所含预测方法之间的分配,并以此为基础确定各预测方法的权重。该文以各个预测模型的均方误差为基础确定权重,具体步骤如下^[11]。

1) 求取 Shapley 值。

$$E'_i = \sum_{\{E|i \in E\}} w(|E|) [V(E) - V(E \setminus \{i\})] \quad (9)$$

式中 E'_i 为第 i 个预测模型分得的均方误差量;

$$E \text{ 为包含第 } i \text{ 种预测模型的组合预测模型}; w(|E|) = \frac{(n-|E|)! (|E|-1)!}{n!}; |E| \text{ 为 } E \text{ 包含的预测模型个数}; n \text{ 为单一预测模型总数}; V(E) \text{ 为 } E \text{ 的均方误差};$$

$V(E \setminus \{i\})$ 为 E 中除去预测模型 i 以后的均方误差。

2)计算权重。

某个预测模型获得的误差分配值越大,其预测精度越低,在组合预测模型中的权重就越小。基于这样的原则,确定预测模型 i 的权重 λ_i ,即

$$\lambda_i = \frac{1}{n-1} \cdot \frac{E-E'_i}{E}, i=1,2,\cdots,n \tag{10}$$

最终的组合预测模型可描述为

$$Y = \sum_{i=1}^n \lambda_i Y_i \tag{11}$$

式中 Y 和 Y_i 分别为组合模型和模型 i 预测结果。

4 算例与结果

架空线路改造工程通常可拆解为杆塔工程、地线工程、附件工程等子工程。以 33 个实际架空线路改造工程的附件子工程的造价为例,并将 33 个总样本分为 30 个训练样本与 3 个测试样本,对 GA-SVM、XGBoosting 和组合预测模型进行评估。

首先识别出影响该附件子工程造价的 15 个因素:单位刚性跳线数量、单位间隔棒数量、单位合成绝缘子数量、回路数、单位普通绝缘子数量、普通绝缘子单价、单位单相线夹数量、单位防振锤数量、单位挂线金具数量、地形系数、耐张塔比例、防振锤单价、间隔棒单价、合成绝缘子单价、刚性跳线单价。之后,采用 SPSS 软件对该工程的这 15 个影响因素数据进行分析,得到 KMO(kaiser-meyer-olkin)校验指数为 0.761,Bartlett 球体检验显著性指数为 0,适合进行主成分分析。由式(1)~(4)可得到 15 个影响因素的权重排序,表 1 中列出了前 7 个影响因素,它们的累积方差百分比高达 94.17%,明显超出一般所采用的 85%的阈值,因此选为主成分。综合考虑主成分分析结果和专家经验,以表 1 中列出的 7 个因素作为关键影响因素,即预测模型的输入变量。

表 1 关键影响因素/输入变量

Table 1 Key impacting factors / input variables			
序号	工程项目属性特征	序号	工程项目属性特征
x_1	单位刚性跳线数量	x_5	单位普通绝缘子数量
x_2	单位间隔棒数量	x_6	普通绝缘子单价
x_3	单位合成绝缘子数量	x_7	单位单相线夹数量
x_4	回路数		

4.1 采用 GA-SVM 模型的造价预测

给定参数:遗传算法所允许的最大进化代数为 200,种群最大数量为 20;误差惩罚参数 C 的变化范围为 $[0,100]$,其初始值为 0;核参数 g 的变化范围为 $[0,1000]$,其初始值为 0。

计算得到的 GA 的适应度曲线如图 1 所示, C 和 g 的最优值分别为 63.129 3 和 0.006 237 1。将 C 和 g 的最优值代入 SVM 中求解,得到训练样本的拟合结果与实际值吻合较好,如图 2 所示;相关系数 $R=92.147\ 3\%$,说明模型拟合精度较高。GA-SVM 对测试样本的预测结果如表 2 所示。

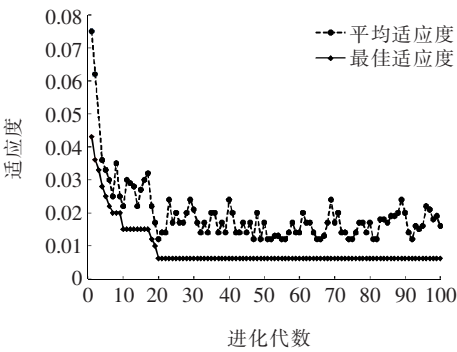


图 1 遗传算法求解参数 C 与 g 的适应度曲线

Figure 1 Fitness curve of the genetic algorithm for determining parameters C and g

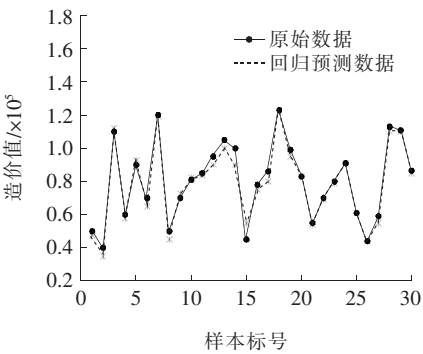


图 2 GA-SVM 对于训练样本的拟合结果

Figure 1 Fitting results of GA-SVM for training samples

表 2 GA-SVM 对测试样本的预测结果

Table 2 Predicted results of GA-SVM for test samples				
测试样本序号	造价实际值/元	造价预测值/元	绝对误差/元	相对误差/%
1	96 200	92 807.31	3 392.69	3.53
2	134 700	137 302.40	2 602.40	1.93
3	110 800	109 786.07	1 013.93	0.92

4.2 采用 XGBoosting 模型的造价预测

给定 XGBoosting 模型的参数如下:学习率为 0.1;树最大深度为 6,用于限制模型复杂度;子节点最小样本权重和为 4,作为树分裂过程终止条件。

XGBoosting 预测模型对于测试样本的预测结果如表 3 所示。

表 3 XGBoosting 模型对于测试样本的预测结果

Table 3 Prediction results of XGBoosting for test samples

测试样本序号	造价实际值/元	造价预测值/元	绝对误差/元	相对误差/%
1	96 200	97 173.64	973.64	1.01
2	134 700	129 533.72	5 166.28	3.84
3	110 800	114 651.41	3 851.41	3.48

4.3 采用组合预测模型的造价预测

构建 GA-SVM 和 XGBoosting 的组合预测模型。首先确定这 2 个预测模型的权重。给定 GA-SVM 和 XGBoosting 预测模型分别为预测模型 1 和 2。对这 2 个预测模型的均方误差进行计算,得到 $E\{1\}=6\,436\,958.823$, $E\{2\}=14\,157\,269.04$, 则组合预测的误差待分摊值 $E\{1,2\}=(E\{1\}+E\{2\})/2=10\,297\,113.93$ 。

由式(9)求得 GA-SVM 和 XGBoosting 预测模型的 Shapley 值分别为 $E'_1=2\,576\,803.713$ 和 $E'_2=7\,720\,310.217$,组合预测模型的 Shapley 值为 $E'=E'_1+E'_2=10\,297\,113.93$ 。 E' 与 $E\{1,2\}$ 的数值相同,计算结果正确。

由式(10)可得 GA-SVM 和 XGBoosting 这 2 个预测模型的权重系数分别为 $\lambda_1=0.75$ 和 $\lambda_2=0.25$ 。最终得到的组合预测模型为 $Y=0.75Y_1+0.25Y_2$ 。

组合预测模型对于测试样本的预测结果如表 4 所示。

表 4 组合预测模型对于测试样本的预测结果

Table 4 Prediction results of the combinational forecasting model for test samples

测试样本序号	造价实际值/元	造价预测值/元	绝对误差/元	相对误差/%
1	96 200	93 898.90	2 301.10	2.39
2	134 700	135 360.23	660.23	0.49
3	110 800	111 002.40	202.40	0.18

采用 3 种误差指标即均方误差(mean square error,MSE)、平均绝对误差(mean absolute error,MAE)和平均绝对相对误差(mean absolute percent error,MAPE)评估 3 种预测模型的性能,结果如表 5 所示。

表 5 3 种预测模型的预测误差对比

Table 5 Comparisons of prediction errors among three forecasting models

预测模型	MSE	MAE	MAPE/%
GA-SVM	6 436 958.82	2 336.34	2.12
XGBoosting	14 157 269.04	3 330.44	2.77
组合模型	1 923 984.46	1 054.58	1.02

从表 5 可以看出,在这 3 种预测模型中,组合模型得到的预测结果所对应的 MSE、MAE、MAPE 这 3 个误差指标都最小且具有明显优势,即预测精度明显高于 GA-SVM 和 XGBoosting 这 2 个预测模型。

5 结语

针对架空线路改造工程造价问题,构建了一种组合预测模型,以提高预测精度。首先采用主成分分析法与专家经验相结合,筛选出影响工程造价的关键因素作为预测模型的特征量(输入量)。接着,采用了基于遗传算法优化的支持向量机(GA-SVM)和极限梯度提升算法(XGBoosting)分别进行造价预测,在此基础上构建了这 2 种方法的组合预测模型并用 Shapley 值确定二者的权重。最后,以实际架空线路改造工程的附件子工程造价预测为例,通过仿真计算验证了基于 GA-SVM 和 XGBoosting 的组合预测模型的预测精度明显高于分别采用这 2 种模型进行预测的结果。

参考文献:

[1] 凌云鹏,闫鹏飞,韩长占,等. 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,2012,45(10): 95-99.
LING Yunpeng, YAN Pengfei, HAN Changzhan, et al. BP neural network based cost prediction model for transmission projects[J]. Electric Power, 2012, 45(10): 95-99.
[2] 冯瀚,刘冰旂,张玉鸿,等. 基于 FCM 和 PSO-SVM 的电力工程造价预测模型研究[J]. 华东电力,2014,42(12): 2713-2716.

- FENG Han, LIU Bingyi, ZHANG Yuhong, et al. Predicting model of power engineering cost based on the FCM and PSO-SVM[J]. East China Electric Power, 2014, 42(12): 2713-2716.
- [3] 彭光金, 俞集辉, 崔荣, 等. 基于数据挖掘技术的输电工程造价估算[J]. 工业工程与管理, 2009, 14(3): 90-95.
- PENG Guangjin, YU Jihui, CUI Rong, et al. Cost estimation of power transmission projects based on data mining technology[J]. Industrial Engineering and Management, 2009, 14(3): 90-95.
- [4] 王守民. 基于关键影响因素的220kV变电工程本体造价预测模型[J]. 中国科技纵横, 2011, 12(1): 169.
- WANG Shoumin. 220 kV substation project cost prediction based on key factors[J]. China Science and Technology Overview, 2011, 12(1): 169.
- [5] 郭艳飞, 任雪桂, 鞠力, 等. 基于层次分析法的综合能源系统能效评估方法研究及应用[J]. 电力科学与技术学报, 2018, 33(4): 121-128.
- GUO Yanfei, REN Xuegui, JU Li, et al. The comprehensive efficiency evaluation method for integrated energy system based on AHP[J]. Journal of Electrical Power Science and Technology, 2018, 33(4): 121-128.
- [6] 李春生, 陈光辉. 基于最大信息压缩指标与层次分析法的电力负荷组合预测模型[J]. 电力科学与技术学报, 2008, 23(1): 56-59.
- LI Chunsheng, CHEN Guanghui. Combination model for power system load forecast based on maximal information compress index and analytic hierarchy process[J]. Journal of Electrical Power Science and Technology, 2008, 23(1): 56-59.
- [7] 王丰效. 不同优化准则统计组合预测权系数优化[J]. 数学的实践与认识, 2013, 43(13): 135-139.
- WANG Fengxiao. Weight coefficient optimization of combination forecast by different criteria[J]. Mathematics in Practice and Theory, 2013, 43(13): 135-139.
- [8] 钟波, 肖智, 周家启. 组合预测中基于粗糙集理论的权值确定方法[J]. 重庆大学学报: 自然科学版, 2002, 25(7): 127-130.
- ZHONG Bo, XIAO Zhi, ZHOU Jiaqi. Determination to weighting coefficient of combination forecast based on rough set theory[J]. Journal of Chongqing University: Natural Science Edition, 2002, 25(7): 127-130.
- [9] 刘思, 白桦, 张凯, 等. 基于拟合优度赋权的农网用电需求组合预测[J]. 电力建设, 2015, 36(8): 55-60.
- LIU Si, BAI Ye, ZHANG Kai, et al. Power demand combination forecasting of rural power network based on goodness of fit empowerment[J]. Electric Power Construction, 2015, 36(8): 55-60.
- [10] 陈华友. 熵值法及其在确定组合预测权系数中的应用[J]. 安徽大学学报: 自然科学版, 2003, 44(4): 1-6.
- CHEN Huayou. Entropy method and application to determine weights of combination forecasting[J]. Journal of Anhui University: Natural Science Edition, 2003, 44(4): 1-6.
- [11] 秦国真. 云南能源需求影响因素分析及预测[D]. 昆明: 云南财经大学, 2011.
- [12] 商立群, 王守鹏. 改进主成分分析法在火电机组综合评价中的应用[J]. 电网技术, 2014, 38(7): 1928-1933.
- SHANG Liqun, WANG Shoupeng. Application of improved principal component analysis in comprehensive assessment on thermal power generation units[J]. Power System Technology, 2014, 38(7): 1928-1933.
- [13] 卢艳超, 张馨. 基于核主成分分析模型的输电线路工程造价评价指标[J]. 电力建设, 2012, 33(7): 73-77.
- LU Yanchao, ZHANG Xin. Cost evaluation index of transmission line project based on kernel principal component analysis model[J]. Electric Power Construction, 2012, 33(7): 73-77.
- [14] 徐莹, 王嘉阳, 苏华英. 基于遗传算法的支持向量机在径流中长期预报中的应用[J]. 水利与建筑工程学报, 2014, 12(5): 42-45+72.
- XU Ying, WANG Jiayang, SU Huaying. Application of support vector machine based on genetic algorithm to mid-term and long-term run-off prediction[J]. Journal of Water Resources and Architecture Engineering, 2014, 12(5): 42-45+72.
- [15] CHEN Tianqi, Guestrin C. XGBoost: a scalable tree boosting system[C]// Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. San Francisco, CA, USA: ACM, 2016.
- [16] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232.
- [17] 李应红, 尉询楷. 支持向量机和神经网络的融合发展[J]. 空军工程大学学报: 自然科学版, 2005, 6(4): 70-73.
- LI Yinghong, WEI Xunkai. Fusion development of support vector machines and neural networks[J]. Journal of Air Force Engineering University: Natural Science Edition, 2005, 6(4): 70-73.