

doi: 10. 11830/ISSN. 1000-5013. 201606010



应用灰关联分析的 PSO-SVR 工程造价预测模型

王佼^{1,2}, 刘艳春¹

(1. 辽宁大学 商学院, 辽宁 沈阳 110036;
2. 东北电力大学 经济管理学院, 吉林 吉林 132012)

摘要: 为了准确预测与控制工程造价水平, 提出一种基于灰关联分析(GRA)与粒子群优化(PSO)的支持向量回归机(SVR)组合预测模型. 将 GRA 提取的工程造价主要指标向量输入 PSO-SVR 模型预测造价, 采用 PSO 优化的 SVR 模型进行工程造价预测, 对比分析 PSO-SVR 模型和其他智能模型, 对某一地区相同输电工程进行造价预测. 结果表明: 基于灰关联分析的 PSO-SVR 模型的造价预测效果更理想, 预测精度更高.

关键词: 工程造价; PSO-SVR 预测模型; 粒子群优化算法; 灰关联分析

中图分类号: TU 723. 3 **文献标志码:** A **文章编号:** 1000-5013(2016)06-0708-06

Prediction Model for Construction Cost Based on Grey Relational Analysis PSO-SVR

WANG Jiao^{1,2}, LIU Yanchun¹

(1. School of Business, Liaoning University, Shenyang 110036, China;
2. School of Economics and Management, Northeast Dianli University, Jilin 132012, China)

Abstract: In order to accurately predict and control construction cost, we propose a forecasting model based on grey relational analysis (GRA) and support vector regression (SVR) integrated with particle swarm optimization (PSO). Key indicators of construction cost are firstly extracted using grey relational analysis (GRA) and then input into the PSO-SVR model to make predictions. The construction costs of the same electricity transmission projects predicted by the PSO-SVR model and other intelligent models were compared. The results show that the PSO-SVR model based on GRA is more accurate.

Keywords: construction cost; PSO-SVR prediction model; partial swarm algorithm; grey relational analysis

随着输电工程建设规模不断扩大, 工程造价有效管控已经成为输电工程(架空输电线路工程)建设必须解决的关键问题. 输电工程是由线路长度、电压等级、输送容量、地形、气象、导线型号、杆塔选型、钢材用量、线材耗量及价格等诸多因素组成的一个非线性复杂系统^[1]. 因此, 输电工程造价估算实际上是针对这个多变量、非线性复杂系统的求解过程. 目前, 国内输电项目工程造价影响因素的研究主要集中于定性分析. 戚安邦^[2]主要从电力工程项目主体、项目本身及环境 3 方面进行具体分析, 列出影响要素. 王丽霞^[3]以全过程管理为指导, 详细分析了输电工程各阶段存在的问题. 董军^[4]在输电投资模式差异化下, 分别从可靠性、经济性和环保性 3 个方面对输电工程造价影响因素进行较为系统、全面的研究. 窦文雷^[5]从输变电工程建设的实际情况出发, 对电力工程造价超支等方面存在的问题进行研究, 并提出一些

应对措施和建议. 上述文献均未对导致工程造价偏差的原因进行深入量化研究, 进而无法得出对输电工程造价进行有效评估的较为客观、准确的分析结果. 现阶段随着国内外学者对智能算法的深入研究, 出现了以 BP 为代表的神经网络、支持向量回归机(SVR)及遗传算法(GA)优化下的支持向量回归机(GA-SVR)等预测模型, 但各个模型在实际工程造价预测中都有不足之处. BP 神经网络由于要求大量的训练样本数据而导致网络模型训练时间较长, 且容易出现局部最优化等问题^[6]. SVR 虽然能够较好地解决小样本、高维数、非线性及局部最优化等实际问题, 但针对输电工程预测的特殊性, 单一利用支持向量回归机建模进行造价预测时, 由于模型参数设置存在着盲目性, 从而导致预测误差较大^[7]. 刘爱国等^[8]发现 GA-SVR 虽然可以在一定程度上对 SVR 的参数进行优化, 但却存在遗传算法自身的交叉率、变异率等复杂参数设置问题. 本文提出一种基于基于灰关联分析(GRA)与粒子群优化(PSO)的支持向量回归机混合算法, 构建输电工程造价预测模型.

1 灰关联分析

灰关联分析(GRA)的概念是由灰色系统理论所提出, 通过灰关联分析可以判定各子系统间密切联系的程度, 从而某个系统发展变化态势可由灰关联分析法供量化的标准, 其基本分析有如下 5 个步骤.

步骤 1 比较矩阵的确定. 设参考数列(又称系统特征序列)为 $x'_0 = \{x'_0(k) | k=1, 2, \dots, n\}$; 比较数列(又称相关因素序列)为

$$x'_i = \{x'_i(k) | k=1, 2, \dots, n\}, \quad i=1, 2, \dots, m. \tag{1}$$

步骤 2 无量纲化的数据处理. 因为系统中各因素指标代表着不同的意义, 所以指标单位不一致. 这就导致了数据的量纲差异较大, 无法准确比较, 或在对比时难以获得较为科学的结论. 因此, 在进行灰关联分析时, 通常要进行无量纲化数据处理. 无量纲化数据处理, 即

$$x_i(k) = x'_i(k)/x_i, \quad i=0, 1, 2, \dots, m; \quad k=1, 2, \dots, n. \tag{2}$$

步骤 3 灰关联系数 $\xi_{0,i}(k)$ 的确定. 所谓灰关联系数, 指对于参考数列 x_0 有若干个比较数列 x_1, x_2, \dots, x_m , 每一个比较数列与参考数列在各个因素指标点上都有关联程度值, 即

$$\xi_{0,i}(k) = \frac{\{\min_i \min_k |x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|\}}{\{|x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|\}}. \tag{3}$$

式(3)中: ρ 为分辨系数, 通常取 0.5; $i=1, \dots, m; k=1, 2, \dots, n$.

步骤 4 计算关联度. 关联系数为比较数列与参考数列在各个因素指标点上的关联程度值, 因此, 它不止一个值. 但是过于分散的信息不利于各子系统间进行整体性比较, 所以有必要采用求平均值的方法来将各因素指标点上的关联系数集中到一个值, 并将它视为比较数列与参考数列间关联程度的数量表示, 即关联度为

$$r_{0,i} = (1/n) \sum_{k=1}^n \xi_{0,i}(k), \quad i=1, \dots, m; \quad k=1, 2, \dots, n. \tag{4}$$

步骤 5 关联度优势排序. 关联度 $r_{0,i}$ 越大, 说明比较数列 x'_i 与参考数列 x'_0 变化趋势更接近, 或者该因素序列指标对特征序列指标影响更强. 为精确筛选出主要影响因素指标, 可根据研究需要设定一个阈值 r , 如果 $r_{0,i} > r$, 则视为目标关联; 若相反, 则认为虚假关联. 根据 3-5-8 因素相关原则, r 设定为强相关阈值 0.8.

2 PSO-SVR 预测模型

利用支持向量回归机估算时, 核函数的类型选取、核函数的参数 σ 及惩罚系数 C 的设定都很重要. 因此, 有必要调整这些重要参数, 以求达到获取最佳推广能力的目标. SVR 参数选择问题实际上相当于一个优化求解过程, 位于搜索空间中的每个点都可能成为最佳模型一个解^[9], 再通过能力推广对预测值进行评估. 因此, 最优参数的求解问题实质上就是误差最小化泛化问题. PSO 算法是一种进化算法, 可应用于一切 GA 能应用的场合, 在编码和寻优策略上, PSO 要比 GA 更加简单、有效.

2.1 粒子群优化算法

假设在一个 N 维空间进行搜索^[10], 可用两个 N 维向量表示粒子 i 的信息, 粒子 i 的位置与速度分

别为

$$\begin{aligned} \boldsymbol{x}_i &= (x_{i,1}, x_{i,2}, \cdots, x_{i,N})^T, \\ \boldsymbol{v}_i &= (v_{i,1}, v_{i,2}, \cdots, v_{i,N})^T. \end{aligned}$$

当粒子 i 找到两个最优解后,更新自己的位置和速度,即

$$v_{i,d}^{k+1} = \omega v_{i,d}^k + C_1 \times \text{rand}_1^k \times (\text{Pbest}_{i,d}^k - x_{i,d}^k) + C_2 \times \text{rand}_2^k \times (\text{Gbest}_{i,d}^k - x_{i,d}^k). \tag{5}$$

$$x_{i,d}^{k+1} = x_{i,d}^k + v_{i,d}^{k+1}. \tag{6}$$

式(5),(6)中: $v_{i,d}^k$ 是粒子 i 在第 k 次迭代中第 d 维速度; $x_{i,d}^k$ 是粒子 i 在第 k 次迭代中第 d 维的当前位置; $i=1,2,3,\cdots,N$,表示种群大小; ω 为惯性权值; C_1 和 C_2 为学习因子(控制加速系数),合适的 C_1 和 C_2 既可加快收敛又不易陷入局部最优,通常在 $[0,2]$ 取值; rand_1^k 和 rand_2^k 是介于 $[0,1]$ 之间的随机数; $\text{Pbest}_{i,d}^k$ 是粒子 i 在第 d 维的个体极值点的位置; $\text{Gbest}_{i,d}^k$ 是整个种群在第 d 维的全局极值点的位置.

最大速度 v_{\max} 决定了问题空间搜索的力度,粒子的每一维速度 $v_{i,d}$ 都会被限制在 $[-v_{d,\max}, +v_{d,\max}]$ 之间,假设搜索空间的第 d 维定义为区间 $[-x_{d,\max}, +x_{d,\max}]$,则有 $v_{d,\max} = \delta_{x_{d,\max}}$. 其中:每一维都用相同的设置方法.

式(5),(6)计算粒子自身最优位置和群体最优位置可表示为

$$\text{Pbest}_{i,d}^{k+1} = \{ x_{i,d}^{k+1} \mid f(x_{i,d}^{k+1}) < f(\text{Pbest}_{i,d}^k) \}, \tag{7}$$

$$f(\text{Gbest}_{i,d}^k) = \min \{ f(\text{Pbest}_{i,d}^k) \}, \quad i = 1, 2, \cdots, N. \tag{8}$$

2.2 基于 PSO 优化的支持向量回归机

PSO-SVR 算法依据粒子群群体寻优的思想加速支持向量回归机寻找最优参数值,具体有如下 5 个步骤.

步骤 1 创建初始样本训练集. 若共有 n 个工程样本,则 $D_i \{ (x_i, y_i), i=1,2,\cdots,n \}$. 选取粒子群的初始种群规模 N 和设定控制加速系数 C_1 和 C_2 ,以及在合理范围下生成粒子的初始位置与速度,并利用 PSO 算法对 SVR 的重要参数 C 与 σ 进行优选.

步骤 2 训练 SVR. 通过样本训练集训练 SVR,计算出各个粒子的适应度函数值,并将各个粒子所经历过的最佳位置 $\text{Pbest}_{i,d}^k$ 与该适应度函数值进行比较. 如果 $\text{Pbest}_{i,d}^k$ 劣于该适应度函数值,则可将此适应度函数值作为新的适应度函数值. 为保证适应度函数的稳定性,采用平均相对误差作为适应度函数的值,即

$$f(\text{Gbest}_{i,d}^k) = (1/N) \sum_{i=1}^N | (y_i - y'_i) / y_i |. \tag{9}$$

式(9)中: N 为样本训练集中样本点数目; y_i 和 y'_i 分别为第 i 个样本的实际值与预测值.

步骤 3 适应度函数值的比较. 将每一个粒子的适应度函数值与所有群粒子的适应度函数值相比较,如果群粒子的适应度函数值大于每一个粒子的适应函数,则用全局最优位置 $\text{Gbest}_{i,d}^k$ 将被当前粒子的最优位置 $\text{Pbest}_{i,d}^k$ 所取代. 同时根据式(5),(6),分别对粒子的位置与速度进行调整.

步骤 4 判断是否终止计算. 如果满足终止条件,则结束寻优搜索,同时,输出 SVR 的最优参数;若是不满足条件,则需要重复步骤 2.

步骤 5 最优参数代入模型. 将经过 PSO 训练获得的最优 C 和 σ 代入 SVR 模型中,重新进行样本训练学习,得到较为理想的 SVR 预测模型.

3 仿真分析

3.1 输电工程造价的影响因素

结合以往相关文献的研究,并依据电气、结构及技经等相关专业技术经济指标性质,初步筛选出回路数、导线分裂数、导线截面、线路长度、输送容量、线材量、塔材量、地线量、杆塔基数、风速、覆冰、地形、电压等级、线材价格和塔材价格 15 个影响输电工程的造价指标^[11].

由于受到电力工程样本数据可获取性的限制,通过对现有 220 kV 输电工程样本数据整理清洗,研究输电工程造价主要影响因素. 因此,这里暂不考虑电压等级因素对工程造价的影响. 其次,由于地形是电力工程造价系统分析中的必要因素,依据国家电网颁布的《输电工程典型造价(220 kV 输电线路分

册)》,将地形按实际情况分为平地、河网泥沼、丘陵、一般山地和高山大岭 5 类地形。

通常情况下,不同的地形会对电力工程造价产生的影响差异较大,而实地工程建设中的地形多为混合型^[12]。在保证研究结论科学性的前提下,将依据《输电工程典型造价(220 kV 输电线路分册)》中列出的 9 种典型方案^[13],以每一个典型方案中平地作为基准,将地形因素对输电工程造价的影响由地形综合系数指标反映,从而有效地提升工程造价影响因素分析的准确性,即

$$\begin{aligned} \text{地形综合系数} = & P_1 + P_2 \times (Z_2/Z_1) + P_3 \times (Z_3/Z_1) + \\ & P_4 \times (Z_4/Z_1) + P_5 \times (Z_5/Z_1). \end{aligned} \tag{10}$$

式(10)中: $P_i(i=1,2,3,4,5)$ 代表典型工程中 5 种地形所占比例; $Z_i(i=1,2,3,4,5)$ 代表典型工程中 5 种地形所对应的典型造价。

3.2 PSO-SVR 系统的输入向量

以华北电网系统内已竣工投产的 29 个 220 kV 输电线路工程为样本,如表 1 所示。表 1 中: X_0 为静态投资。由表 1 可知:灰关联分析中的参考数列变量、比较数列变量在量纲上不一致,且其数值变化差异较大,不适宜直接进行灰关联分析;否则,严重影响主要因素提取结果。

表 1 原始工程样本数据

Tab.1 Sample data of original engineering

工程编号	X_0 /万元	输送容量/kW	线路长度/km	线材量/t	...	线材价格/万元·t ⁻¹
1	4 384.374 2	146 468	18.75	317.87	...	1.373
2	2 696.711 5	62 772	21.77	132.12	...	1.680
3	3 636.148 4	146 268	16.00	246.83	...	1.358
4	5 358.362 1	104 610	50.46	421.00	...	1.924
⋮	⋮	⋮	⋮	⋮	⋮	⋮

根据式(2),对数据变量进行无量纲化处理,具体结果如表 2 所示。利用 GM 软件将经过处理的数据变量进行灰关联分析,即将上文初步分析识别出来的 14 个输电工程造价影响因素作为工程特征参数进行关于工程造价-静态投资的灰关联分析。

表 2 归一化工程样本数据

Tab.2 Normalized engineering sample data

工程编号	X_0 /万元	输送容量/kW	线路长度/km	线材量/t	...	线材价格/万元·t ⁻¹
1	1.000 000	1.000 000	1.000 000	1.000 000	...	1.000 000
2	0.615 073	0.428 571	1.161 067	0.415 642	...	1.000 000
3	0.829 343	1.000 000	0.853 330	0.776 512	...	0.989 075
4	1.222 150	0.714 217	2.691 200	1.324 441	...	1.401 311
⋮	⋮	⋮	⋮	⋮	⋮	⋮

将关联度在 0.8 以上的 11 个输电工程造价影响因素确定为造价主要影响因素,并作为 PSO-SVR 预测系统输入量,分析结果如表 3 所示。

表 3 因素变量灰关联分析

Tab.3 Grey relational analysis for factor variables

X	X_1 /kW	X_2 /km	X_3	X_4	X_5	X_6 /m ²	X_7
关联度	0.985 7	0.983 6	0.980 6	0.975 6	0.974 9	0.974 5	0.970 6
X	X_8 /m·s ⁻¹	X_9 /mm	X_{10}	X_{11}	X_{12} /万元·t ⁻¹	X_{13} /万元·t ⁻¹	X_{14}
关联度	0.870 3	0.867 5	0.865 3	0.865 0	0.698 5	0.691 2	0.486 8

表 3 中: X 为因素指标; X_1 为输送容量; X_2 为线路长度; X_3 为线材量; X_4 为塔材量; X_5 为回路数; X_6 为导线截面; X_7 为地形综合系数; X_8 为风速; X_9 为覆冰; X_{10} 为杆塔基数; X_{11} 为导线分裂数; X_{12} 为剔除线材价格; X_{13} 为塔材价格; $X_{14}(t)$ 为地线量 3 个因素指标。

由表 3 可知:地线多用于输电线路引雷,避免线路遭遇雷击而受损,通常在输电线路工程间使用量相差不大,所以地线量与工程造价关联度小。同时,考虑由于相近年份工程耗材价格波动不大,所以研究相近年份输电工程造价时,价格因素与工程造价变化情况的关联度相对较小。

利用 MATLAB7.8 中加载 Libsvm 工具箱,由于该工作箱的工作界面程序中包含归一化函数 Tr-

amnmax 函数和反归一化函数 Postmnmx 函数,所以该系统会自动将原始数据变量做归一化处理输入系统运行,再将结果做反归一化处理后输出系统.将灰关联分析所提取出的 11 个造价主要影响因素原始数据直接录入 SVR 程序系统,如表 4 所示.由表 4 可知:输入集共 11 个因素指标构成了 29×11 矩阵;输出集静态投资造价(万元)1 个因素指标构成 29×1 列向量.

表 4 输入输出属性集
Tab. 4 Input and output set for characters

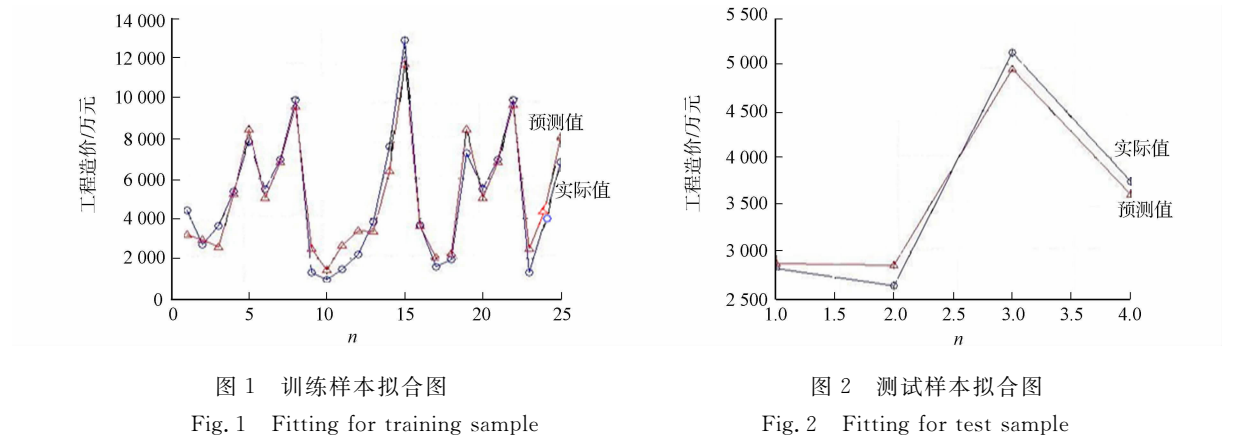
样本序号	因素指标					Y
	X ₁	X ₂	X ₃	...	X ₁₁	
1	146 468	18.75	317.87	...	1	4 384.374
2	62 772	21.77	132.12	...	1	2 696.712
3	146 268	16.00	246.83	...	1	3 636.148
⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	146 468	20.00	309.00	...	1	4 140.370

选取 25 个工程造价数据为学习样本,剩余 4 个工程造价数据为测试样本,利用 PSO 优化支持向量回归机(SVR)的参数,将学习样本输入 PSO-SVR 预测网络,得到稳定的模型,再利用测试样本在稳定的网络模型中得出的预测结果,与真实测试集输出结果进行比较,即可完成整个预测过程.

3.3 PSO 优化 SVR 参数

采用 PSO 对支持向量回归机(SVR)的惩罚系数 C 和径向基核函数(RBF)的参数 σ 进行寻优.初始化粒子群的各项参数,设 PSO 规模是 20,解空间为二维分别对应 C 和 σ ,控制加速系数 C_1 和 C_2 分别等于 1.5 和 1.7,参数 C 的变化范围是 $[0,50]$,参数 σ 的取值区间为 $[0,1]$.那么,模型参数对应的 Scope 阵是 $[0,50;0,1]$.为寻找适合的最大进化代数 T_{\max} 和交叉验证折数 V ,经过多次试验获得 T_{\max} 适合值为 100, V 适合值为 5,此时训练集与测试集拟合程度分别如图 1,2 所示.

由图 1,2 可知:拟合效果较为理想. SVR 模型中的惩罚系数 C 和径向基核函数的参数 σ ,经过 PSO 寻优后分别为 36.29 和 0.01.



3.4 不同模型预测效果对比分析

为了进一步测试 PSO-SVR 模型的预测效果,利用相同的样本数据分别进行 BP 神经网络模型预测、GA-SVR 模型预测和 PSO-SVR 模型预测.测试结果如表 5 所示.

表 5 模型预测效果对比
Tab. 5 Comparison of effect model predictions

工程编号	实际造价/万元	BP 预测		GA 预测		POS-SVR 预测	
		预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%
26	2 807.4	2 113.2	-24	3 120.2	11.0	2 868.2	2.16
27	2 624.3	1 198.1	-54	3 030.6	15.0	2 847.7	8.51
28	5 117.8	4 330.3	15	4 815.6	-5.9	4 941.7	-3.44
29	3 740.4	2 462.5	34	3 670.9	-1.8	3 598.5	-3.79
平均绝对值相对误差			31		8.4		4.47

由表 5 可知:PSO-SVR 造价预测结果与真实造价值的相对误差绝对值均在 10%以内,且无论是相对误差比较还是误差均值比较,都明显优于 BP 预测结果,极大地提高了预测精度;同时,也较 GA-SVR 预测效果更加理想.

4 结论

1) 运用灰关联分析法对各工程特征参数进行灰关联分析并排序,确定 PSO-SVR 的输入向量,构建基于灰关联分析的 PSO-SVR 输电工程造价预测模型. 结果显示:静态投资工程造价的相对误差绝对值最大为 8.51%,最小为 2.16%,说明 11 个工程特征参数的提取可以较好地表征工程造价,降低了估算误差. 仿真结果证明基于 GRA 的 PSO-SVR 预测模型对输电工程造价的估算和审查是科学有效的.

2) 利用 PSO 优化算法对支持向量回归机进行参数优化,避免 SVR 模型参数选择的盲目性. 与现今较为广泛应用的 BP 神经网络模型及 GA-SVM 混合算法预测模型相比较,基于灰关联分析的 PSO-SVR 模型使得工程造价计算与预测工作量大幅减少,效率大大提高,其预测结果更加理想,估算造价精度更高.

参考文献:

[1] 彭光金,俞集辉. 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.

[2] 戚安邦. 多要素项目集成管理方法研究[J]. 南开管理评论,2002,5(6):70.

[3] 王丽霞. 输变电工程项目造价管理研究[D]. 北京:华北电力大学,2011:1-6.

[4] 董军. 输电项目评价研究现状综述 [J]. 华东电力,2010,7(3):313-318.

[5] 窦文雷. 输变电工程造价控制[J]. 农业科技与装备,2012,32(11):51-52.

[6] DING Shifei. An optimizing method of RBF neural network based on geneticalgorithm[J]. Neural Computing and Applications,2012(2):333-336.

[7] NIU Dongxiao. Research on neural network prediction of power transmission and transformation project cost based on GA-RBF and PSO-RBF[J]. Applied Mechanics and Materials,2014(6):234-231.

[8] 刘爱国,薛云涛. 基于 GA 优化 SVM 的风电功率的超短期预测[J]. 电力系统保护与控制,2015,24(2):90-95.

[9] 邵信光,杨慧中. 基于粒子群优化算法的支持向量机参数选择及其应用[J]. 控制理论与应用,2006,23(5):740-748.

[10] 彭光金,司海涛. 改进的支持向量机算法及其应用[J]. 计算机工程与应用,2011,47(18):218-221.

[11] 王佼. 500 kV 架空输电线路工程造价主要影响因素分析[J]. 东北电力大学学报,2012,32(5):9-11.

[12] 凌云鹏,阎鹏飞. 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.

[13] 刘振亚. 输电工程典型造价: 220 kV 输电线路分册[M]. 北京:中国电力出版社,2006:7.

(责任编辑: 陈志贤 英文审校: 方德平)