

基于逐步回归的 AdaBoost-SVR 模型在海上风电项目造价预测中的应用

周子东¹, 李东伟¹, 李国胜², 汪 群², 陈晓路³, 管春雨³

(1. 中国电建集团华东勘测设计研究院有限公司, 杭州 310014; 2. 华能国际电力股份有限公司, 北京 100031;

3. 华能江苏清洁能源分公司, 南京 210009)

摘 要: 为克服海上风电造价预测模型中自变量存在的多重共线性对建模的影响,同时针对现阶段海上风电实际成本数据较少,样本数量不多的现状,提出一种基于逐步回归的 AdaBoost-SVR 造价预测模型。逐步回归方法可有效提取重要自变量,避免多重共线性的影响。AdaBoost-SVR 模型可在小样本的条件下保证模型的精度,同时提高泛化能力。工程实例分析表明该模型可有效提高预测精度。

关键词: 海上风电; 逐步回归; 自适应增强算法; 支持向量机

中图分类号: TK89

文献标志码: A

0 引 言

风能作为一种清洁的可再生能源,在中国越来越受到重视,风力发电的布局也逐步在由陆地走向海洋。与相对成熟的陆上风电相比,海上风电造价高、风险大,因此有效地把控项目造价对于项目前期投资决策具有重要的意义^[1]。

为对投资项目的整体造价有一个合理预测,必须找出引起造价变化的因素与工程造价之间的内在联系与统计规律,在这方面回归分析是很常用的工具。以单台风电机组基础造价为例,在项目前期阶段资料缺乏的情况下一般可取水深、波高、风速、上部风电机组重量及钢材价格等作为自变量,由于自变量数量较多且相互之间常存在多重相关性,直接对全部自变量采用多元回归建立统计模型时易导致因子系数不合理。通过逐步回归法(Stepwise)对每引入一个变量后的模型进行 F 检验,只保留最显著的变量,使得筛选后保留在模型中的变量既是重要的,又无严重多重共线性。但回归模型是一种统计模型,在样本数量较少的情况下模型精度有限,对于中国海上风电造价研究而言,实际可供研究的样本数量较少,直接采用逐步回归建模可能存在精度不够且部分参数的意义无法解

释的问题。支持向量回归(support vector regression, SVR)是一种新颖的小样本学习方法,它基本上不涉及概率测度及大数定律等,因此不同于现有的统计方法,对小样本拟合也有较好的效果,更加适应目前海上风电造价样本数量不够的现状。同时结合自适应增强(adaptive boosting, AdaBoost)算法对 SVR 训练的弱学习器进一步优化,在保证海上风电造价预测模型精度的前提下,进一步提高泛化能力。

1 逐步回归

在一个多元回归模型中,从专业角度通常会挑选出较多的自变量,但并不是所有的自变量都与因变量有显著关系,有些自变量的作用可以忽略。此外若自变量之间存在多重共线性会使模型估计失真或难以准确估计,因此有必要利用数学方法从中筛选出显著的自变量子集。逐步回归法就是一种在实际问题中常用的方法^[2]。

逐步回归的基本思想是将变量逐个引入,引入变量的条件是偏回归平方和经检验是显著的,同时在每引入一个新变量后,对已选入的变量要进行逐个检验,将不显著变量剔除,这样保证最后所得的变量子集中的所有变量都是显著的。这样经若干步以后便得到“最优”变量子集^[3-4]。

收稿日期: 2017-11-14

基金项目: 中国华能集团公司科技项目(HNKJ16-H25)

通信作者: 周子东(1991—),男,硕士、工程师,主要从事工程经济咨询与管理方面的工作及研究。zhouzd450815@163.com

2 AdaBoost-SVR 模型

2.1 支持向量回归(SVR)

支持向量机(SVM)的理论基础来源于统计学习理论的 VC 维理论和结构风险最小原理,由前苏联科学家 Vladimir Vapnik 和 Alexey Chervonenkis 分别在 1963 和 1995 年首先提出。SVM 本身是针对经典二分类问题提出的,其核心思想是确定一个最优超平面,使得样本与超平面间的距离最大化。支持向量回归(SVR)是支持向量在函数回归领域的应用,与 SVM 不同,SVR 所寻求最优超平面不是使 2 类样本点分的最开,而是使所有样本点离超平面的总偏差最小,求最优超平面等价于求最大间隔^[5-6]。

对于非线性情况,设样本 (x_i, y_i) , $i=1, 2, 3, \dots, n$, $x_i \in \mathbf{R}^n$ 为输入量, $y_i \in \mathbf{R}$ 为输出量。首先将输入量 x 通过映射 $\varphi: \mathbf{R}^n$ 过 H 映射到高维特征空间 H 中,然后在高维空间中利用式(1)建立样本的线性回归函数,且式中需要确定 ω 和 b ^[7-8]:

$$f(x) = \omega \psi(x) + b \quad (1)$$

机器学习在学习过程中需要对误差有一个度量,称之为惩罚函数,标准支持向量机采用的惩罚函数为 ε 不敏感损失函数:

$$c[x, y; f(x)] = \max\{0, |y - f(x)| - \varepsilon\} \quad (2)$$

即:

$$\begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i \\ f(x_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (3)$$

式中, ξ_i 、 ξ_i^* ——松弛因子,当拟合误差超过误差限时, ξ_i 和 ξ_i^* 都大于 0,否则取 0。

则求解式(1)中 ω 和 b ,即转化为最优化问题:

$$\begin{aligned} \min & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s.t.} & \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i \\ f(x_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

式中, $f(x)$ ——预测值; y ——实测值; $\varphi(x)$ ——映射函数; ε ——误差限; C ——惩罚因子。

式(4)中第 1 项使拟合函数更为平坦,提高泛化能力;第 2 项减小误差,这是一个凸二次优化问题,引入 Lagrange 函数得到对偶形式:

$$\begin{aligned} \max_{\alpha, \alpha^*} & \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varepsilon + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i \right] \\ \text{s.t.} & \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases} \end{aligned} \quad (5)$$

求解式(5)得到最优解为 $\alpha_i = [\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n]$, $\alpha_i^* = [\alpha_1^*, \alpha_2^*, \alpha_3^*, \dots, \alpha_n^*]$ 。

最后回归函数为:

$$f(x) = \omega^* \psi(x) + b^* = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b^* \quad (6)$$

式中, $K(x_i, x) = \Phi(x_i) \cdot \Phi(x)$ 为核函数。

核函数的作用是计算出经过某个变换后在高维空间里的原低维空间向量的内积值。常见的核函数有多项式核、径向基核及 Sigmoid 核等。

2.2 AdaBoost 算法

AdaBoost (adaptive boost) 算法是根据 Boosting 算法改造而成,该算法由美国加州大学的教授 Yoav Freund 和美国普林斯顿大学的教授 Robert Schapire 在 1995 年共同提出。AdaBoost 算法中心思想就是对指定的对象,通过多次较弱的学习训练(弱学习器),而不是一次单独的学习训练(强学习器)建立映射模型。

该算法是通过改变样本数据的分布来实现的,它根据每次训练集中各个样本的学习是否正确以及上一次样本学习的权值总和来确定这次每个样本的权值。对于学习错误的样本,增加其对应的权重;而对于学习正确的样本,降低其权重,这样学习错误的样本就被凸显出来,从而得到一个新的样本分布。将修正权值后的新数据赋给下一个学习器进行训练,又可以得到一个新的样本分布。以此类推,经过 T 次循环,得到 t 个弱学习器,把这 t 个弱学习器按它们所得到的权重叠加起来,得到最终想要的强学习器^[9]。

2.3 AdaBoost-SVR 模型

支持向量回归有很多优点,但其性能很大程度取决于核函数及参数的选择。在样本数量较少的情况下为保证泛化能力,惩罚因子 C 值不应取得过大,此时的 SVR 模型可视作“弱学习器”处理。针对于任意的弱学习器,AdaBoost 算法都能够提高其精度,将二者结合可解决如下问题:单独使用 SVR 进行样本学习,模型性能取决于所选核函数和核参数。而使用 SVR 作为 AdaBoost 的弱学习器,可降低核函数与核参数的选择所带来的影响^[10]。

可按照以下步骤来建立海上风电造价的 AdaBoost-SVR 预测模型:

步骤 1:选取的 m 组数据作为训练样本,并给每一组样本赋初始权值 $D(i) = 1/m$;

步骤 2:根据选用的核函数,选择合适的核参数,用 SVR 作为弱学习器进行训练,得到第 t 个基学习器 $h(t)$;

步骤 3: 计算第 t 个基学习器 $h_t(i)$ 输出值与实际值的误差, 并累加超过误差限项的权值 $D_t(i)$, 得到预测误差 ε_t ;

步骤 4: 计算学习器的权重。根据该基学习器 $h_t(i)$ 的预测误差 ε_t 计算权重 α_t , 权重计算公式为:

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (7)$$

步骤 5: 测试数据权重调整, 下一轮的样本权重为:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t}, & \text{误差} < \text{误差限} \delta_t \\ e^{\alpha_t}, & \text{误差} \geq \text{误差限} \delta_t \end{cases} \quad (8)$$

式中, Z_t ——归一化系数, 使 $\sum_{i=1}^m D_{t+1}(i) = 1$ 。

即:

$$Z_t = \sum_{i=1}^m D_t(i) \times \begin{cases} e^{-\alpha_t}, & \text{误差} < \text{误差限} \delta_t \\ e^{\alpha_t}, & \text{误差} \geq \text{误差限} \delta_t \end{cases} \quad (9)$$

步骤 6: 训练 T 轮后得到了强学习函数 $H(x)$:

$$H(x) = \frac{\sum_{t=1}^T \alpha_t}{\sum_{t=1}^T \alpha_t} h_t(x) \quad (10)$$

3 工程实例分析

海上风电项目的组成较复杂, 为便于分析, 可将整

个海上风电分为若干个模块来进行造价研究, 即风电机组基础模块及风电机组模块、海上升压站模块、陆上集控中心模块、其他费用模块等。本节以风电机组基础模块为例, 首先采用逐步回归方法选取显著影响造价水平的自变量, 然后利用 AdaBoost-SVR 方法来建立风电机组基础造价的预测模型, 其余模块的预测模型可参照建立。

3.1 逐步回归

首先从专业角度初选单台风电机组基础造价预测模型的影响因子即自变量: 1) 水深; 2) 波高; 3) 风速; 4) 离岸距离; 5) 单台风电机组容量; 6) 基础上部风电机组重量; 7) 钢材价格。然后选择了 6 个实际工程样本, 并将各自工程特性参数列出, 详见表 1。表 1 中 7 个工程特性, 编号为 $x_1 \sim x_7$ (水深~钢材信息价), 即风电机组基础造价回归模型的自变量, y 为每个样本的风电机组基础造价, 即因变量, 则回归模型可设为:

$$f(x) = \sum_{i=1}^n a_i x_i + c \quad (11)$$

式中, $f(x)$ ——目标函数; a_i ——待确定的回归系数; n ——自变量数量; c ——常数项。

表 1 风电机组基础造价回归模型建模样本

Table 1 Training samples for cost forecast model of wind turbine foundation

特征编号	特征名称	样本 1	样本 2	样本 3	样本 4	样本 5	样本 6
x_1	水深/m	2.5	7.5	12.0	9.0	11.0	8.0
x_2	全年有效波高均值/m	0.56	0.56	0.56	1.80	1.80	0.97
x_3	年均风速/ $\text{m} \cdot \text{s}^{-1}$	7.49	7.49	7.49	7.00	7.00	6.50
x_4	离岸距离/km	27	25	30	11	14	25
x_5	风电机组容量/MW	4.0	4.2	5.0	4.0	3.0	4.0
x_6	风电机组质量/t	484	513	705	423	350	474
x_7	钢材价/ $\text{元} \cdot \text{t}^{-1}$	2487	2487	2487	2856	2856	2916
y	每台造价/万元	1009	1152	1307	1130	994	1339

式(11)中包括常数项共有 8 个参数, 并不需要全部进入回归模型, 采用 Matlab 的 Stepwise 函数进行逐步回归分析, 最终选出 5 个参数项: a_1 、 a_3 、 a_5 、 a_7 和 b , 可引入建立回归模型, 得到的模型复相关系数 R 为 0.9759, 均方根误差为 12.90, 模型系数见表 2。

根据逐步回归建模的结果, a_1 、 a_3 、 a_5 、 a_7 对应的自变量为水深、年平均风速、单机容量和钢材信息价, 这 4 项是筛选出的对因变量最具解释力的因子, 由此建立的回归模型也具有一定的精度。但由于总体建模的样本数量较少, 若用于预测存在一定不确定性, 因此考虑

利用筛选出的 4 个自变量建立 AdaBoost-SVR 模型。

表 2 风电机组基础造价回归模型系数

Table 2 Regression model coefficient of cost forecast model

系数	a_1	a_3	a_5	a_7	b
数值	15.08	-547.82	153.29	-0.72	6260.30

回归模型为:

$$f(x) = 15.08x_1 - 547.82x_3 + 153.29x_5 - 0.72x_7 + 6260.30 \quad (12)$$

模型的拟合结果与样本数据的对比如表 3 所示。

表3 回归模型拟合数据与样本数据对比

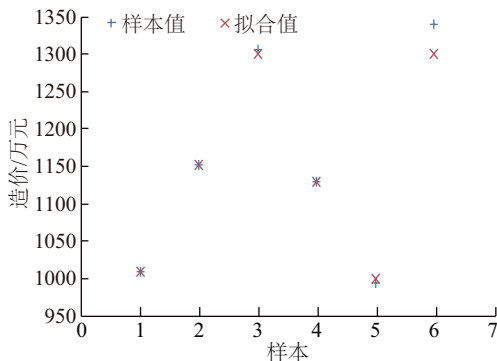
Table 3 Comparison of regression model's fitting data and sample data

样本 编号	样本 数据/万元	拟合 数据/万元	误差/ 万元	误差率/ %
1	1008.69	1021.87	13.18	-1.31
2	1152.46	1127.95	-24.51	2.13
3	1306.77	1318.45	11.68	-0.89
4	1130.01	1123.50	-6.51	0.58
5	993.62	1000.38	6.77	-0.68
6	1339.08	1339.21	0.13	-0.01

3.2 AdaBoost-SVR 模型

仍采用 3.1 节中的 6 个样本,自变量选择水深、年平均风速、单机容量和钢材信息价,按照 2.3 节所述的 AdaBoost-SVR 步骤来建立模型。选取径向基函数作为 SVR 的核函数(RBF),模型中需要确定的参数有 2 个:惩罚因子 C 和核参数 σ^2 。许多有关 SVR 的研究表明,对于 RBF-SVR 当 C 值固定后, σ 的值对模型的拟合效果有较大的影响。为避免 C 值过大导致过拟合,设定参数范围为 $C \in [20, 25]$,步长 0.5;设置 $\sigma \in [15, 25]$,步长 1。采用网格搜索法来搜索一个较优的参数值,得到寻优结果为 $C = 22.6$, $\sigma = 21$ 。然后固定 C 值,根据 σ 的寻优结果设定 $\sigma \in [20, 22]$,步长为 1,来设置 3 个基学习器。

接下来进行第 1 次基学习器训练,参数设定为 $C = 22.6$, $\sigma = 20$ 。首先,对训练样本赋予初始权值,一共有 6 组样本,因此初始权值 $D_1(m) = 1/6 (m = 1, 2, 3, \dots, 6)$ 。设定误差限 $\delta_i = 6$ 万元。把上述 6 个计算方案相应的参数值组合作为 SVR 模型的输入,对应的有限元计算结果为输出,训练 SVR 模型,模型拟合结果如图 1 所示。

图1 基学习器 h_1 拟合结果Fig. 1 Fitting result of basic learning machine h_1

第 1 次训练结束后,计算每一组输出值的误差 ε ,并与设定的误差限比较,共有 2 组数据的误差 ε 超过误差限 δ_i ,因此可得到该基学习器的预测误差 $\varepsilon_1 = 1/6 \times 2 = 0.33$,然后将预测误差 ε_1 代入式(7)计算该学习器的权重,得到权重 $\alpha_1 = 0.347$,由此得到第 1 个基学习器。在进行下一轮训练之前,先要对样本的权重进行更新,根据式(9)计算归一化系数 $Z_1 = 0.943$,然后按照式(8)计算第 2 次训练的样本权重,经计算对于在第 1 个基学习器中满足要求的样本其下一轮权重 $D_2(m) = 0.125$,对于学习不满足要求的样本下一轮权重 $D_2(m) = 0.250$,与第 1 个基学习器的权值相比,学习正确的样本权值被减小,相应错误的被放大。随后保持 C 值不变,参数 σ 加上步长 1 取为 21,进行第 2 个基学习器的训练,重复上述过程。最后一共得到 3 个基学习器,每个学习的权重及其在最终的强学习器中所占比重如表 4 所示。

表4 基学习器权重

Table 4 Weight of base learning machine

基学习器编号	计算权重	在强学习器中比重
$h_1(x)$	0.347	0.25
$h_2(x)$	0.255	0.18
$h_3(x)$	0.805	0.57

最后得到的强学习器,即 AdaBoost-SVR 模型:

$$H(x) = 0.25h_1(x) + 0.18h_2(x) + 0.57h_3(x) \quad (13)$$

将上述 6 组样本的自变量代入 $H(x)$ 进行计算得到模型的拟合值,并计算与实际值之间的误差,与逐步回归和 SVR 方法的拟合误差做对比,如图 2 所示。

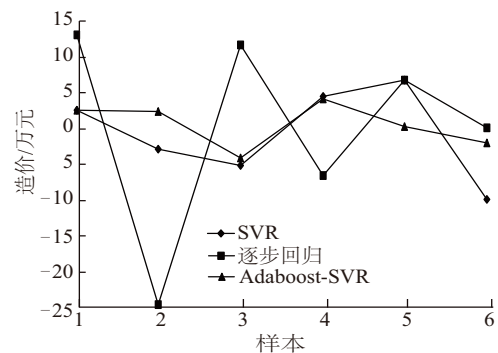


图2 逐步回归、SVR、AdaBoost-SVR 这 3 种模型拟合误差对比

Fig. 2 Fitting error of three kinds of model: stepwise, SVR and AdaBoost-SVR

经计算,逐步回归、SVR 和 AdaBoost-SVR 这 3 种模型拟合值的均方根误差 $RMSE$ 分别为 12.90、5.76 和

2.62。由此可知对于此样本来说,3 种模型中逐步回归的拟合误差大,效果相对较差。AdaBoost-SVR 在 SVR 模型的基础上进一步提高了精度,拟合效果最好。

衡量一个模型的优劣,除了拟合精度,泛化能力也是一个重要的衡量指标。另选取某地区 3 个未参与建模的实际项目的风电机组基础造价样本,用于验证模型的预测效果,样本信息如表 5 所示。

将上述 3 个样本用逐步回归、SVR 和 AdaBoost-

SVR 的 3 种模型进行测试,得到 3 种模型输出预测值与对应实际值的误差,如表 6 所示。由对比表 6 分析可知,3 种模型中逐步回归预测模型的预测结果误差较大,基本在约 10%,性能最差;单一 SVR 模型较逐步回归有了较大的提升,说明在小样本情况下,泛化能力也能得到保证;而 AdaBoost-SVR 模型在单一 SVR 模型的基础上利用 AdaBoost 方法进一步提高了泛化能力,误差相对较小,效果最好。

表 5 风电机组基础造价模型预测样本

Table 5 Test samples for cost forecast model of wind turbine foundation

特征编号	特征名称	测试样本 1	测试样本 2	测试样本 3
x_1	水深/m	5.5	10.0	16.5
x_3	年平均风速/ $\text{m}\cdot\text{s}^{-1}$	7.64	7.41	7.42
x_5	风电机组容量/MW	4.0	4.0	4.0
x_7	钢材价/ $\text{元}\cdot\text{t}^{-1}$	2278	2687	2688
y	每台风力机基础造价/万元	1023.33	911.76	1061.62

表 6 3 种模型风电机组基础造价预测误差对比

Table 6 Prediction error of three kinds of cost forecast model

样本序号	样本实际值	逐步回归模型		SVR 模型		AdaBoost-SVR 模型	
		预测值/万元	误差/%	预测值/万元	误差/%	预测值/万元	误差/%
1	1023.33	1135.25	10.94	1106.15	8.09	985.53	-3.69
2	911.76	1035.20	13.54	963.77	5.70	949.85	4.18
3	1061.62	1127.27	6.18	1015.67	-4.33	1032.94	-2.70

4 结 论

逐步回归方法在面对模型参数多、存在多重共线性的情况下,有着独特的优势,可有效提取重要自变量,简化模型,避免多重共线性对模型精度的影响,但在小样本的情况下,模型的拟合精度和泛化能力一般;AdaBoost-SVR 在小样本的情况下也可保证良好的拟合精度,同时还拥有良好的泛化能力。本文将 2 种方法结合,利用工程实例,首先用逐步回归方法对模型的自变量进行筛选,提取主要自变量,然后利用 AdaBoost-SVR 方法建立海上风电风电机组基础造价的预测模型,最后选取 3 个测试样本来验证模型的可靠性,得到的结果表明 AdaBoost-SVR 模型的精度和泛化能力都优于逐步回归模型和单一 SVR 模型,具有一定应用价值。

[参考文献]

[1] 黄琳,黄静波. 海上风电场建设成本及风险分析[J].

水力发电, 2012, 38(12): 81-83.
HUANG L, HUANG J B. Analysis on construction cost and risks of offshore wind farm[J]. Water power, 2012, 38(12): 81-83.
[2] 李小春,施颢. 基于逐步回归的偏最小二乘法在大坝位移监控中的应用[J]. 现代测绘, 2016, 39(1): 7-10.
LI X C, SHI H. Application of partial least squares regression based on stepwise regression to dam displacement monitoring[J]. Modern surveying and mapping, 2016, 39(1): 7-10.
[3] 周吕,文鸿雁,李超. 基于灰色系统与逐步回归及二者组合的大坝变形分析模型研究[J]. 城市勘测, 2012 (4): 135-138.
ZHOU L, WEN H Y, LI C. Research of dam deformation analysis model based on grey system, stepwise regression and the combined[J]. Urban geotechnical investigation & surveying, 2012(4): 135-138.

- [4] 傅蜀燕, 欧正峰. 基于逐步回归的 BP 网络混合模型在大坝变形分析中的应用[J]. 人民珠江, 2014, 35(3): 46-48.
FU S Y, OU Z F. Application of BP neural network hybrid model based on stepwise regression in analysis to dam displacement[J]. Pearl river, 2014, 35(3): 46-48.
- [5] 燕乔, 宋志诚, 张利雷, 等. 基于支持向量机的面板堆石坝堆石料参数反演分析[J]. 中国农村水利水电, 2014(10): 123-125.
YAN Q, SONG Z C, ZHANG L L, et al. The application of inversion analysis based on support vector machines (SVM) of rock fill parameters[J]. China rural water and hydropower, 2014(10): 123-125.
- [6] 王泉, 郑东健, 范振东, 等. 基于 PCA-SVR 模型的大坝裂缝早期预报研究[J]. 人民长江, 2015, 46(5): 35-38.
WANG Q, ZHENG D J, FAN Z D, et al. Study on early warning of dam crack based on PCA-SVR model[J]. Yangtze river, 2015, 46(5): 35-38.
- [7] 袁兴国, 程琳. 基于多输出支持向量机和遗传算法的岩土参数反演方法[J]. 三峡大学学报(自然科学版), 2014, 36(2): 50-55.
YUAN X G, CHENG L. Geotechnical parameter inversion analysis method based on multi-output support vector machines and genetic algorithm[J]. Journal of China Three Gorges University (natural sciences), 2014, 36(2): 50-55.
- [8] 周志聪, 祁广云. 基于支持向量机对稻米淀粉含量的回归预测研究[J]. 黑龙江八一农垦大学学报, 2014, 26(6): 88-92.
ZHOU Z C, QI G Y. Study of regression prediction based on support vector machine to rice starch content[J]. Journal of Heilongjiang August First Land Reclamation University, 2014, 26(6): 88-92.
- [9] 牛艳庆, 胡宝清. 基于模糊 AdaBoost 算法的支持向量回归机[J]. 模糊系统与数学, 2006(2): 140-145.
NIU Y Q, HU B Q. Support vector machines based on fuzzy AdaBoost algorithm[J]. Fuzzy systems and mathematics, 2006(2): 140-145.
- [10] 胡德华, 郑东健, 付浩雁. AdaBoost-BP 模型在大坝变形预测中的应用[J]. 三峡大学学报(自然科学版), 2015, 37(5): 5-8.
HU D H, ZHENG D J, FU H Y. Application of AdaBoost-BP model to dam deformation prediction[J]. Journal of China Three Gorges University (natural sciences), 2015, 37(5): 5-8.

APPLICATION OF AdaBoost-SVR MODEL BASED ON STEPWISE REGRESSION IN COST FORECAST OF OFFSHORE WIND POWER

Zhou Zidong¹, Li Dongwei¹, Li Guosheng², Wang Qun², Chen Xiaolu³, Guan Chunyu³

(1. Power China Huadong Engineering Corporation Limited, Hangzhou 310014, China;

2. Huaneng Power International Inc., Beijing 100031, China; 3. Huaneng Jiangsu Clean Energy Corporation, Nanjing 210009, China)

Abstract: This paper presents an AdaBoost-SVR model combined with stepwise regression, in the condition of lack of actual cost of offshore wind project, which solves the Multi-collinearity problems among the indices in offshore wind power cost forecast model. Stepwise regression method can effectively extract important independent variables, which avoid the influence of multi-collinearity. AdaBoost-SVR model ensures the precision of the model, under the condition of small sample, and improves the generalization ability. The engineering example show that the model can effectively improve the prediction accuracy.

Keywords: offshore wind power; stepwise regression; AdaBoost algorithm; support vector machine