

基于 GA 优化 SVM 的风电功率的超短期预测

刘爱国, 薛云涛, 胡江鹭, 刘路平

(南昌大学信息工程学院, 江西 南昌 330031)

摘要: 研究风电功率预测技术对于减轻其输出电能的随机性对电力系统的影响具有重要意义。首先结合风电监控系统数据库中的历史功率数据和环境参数形成样本数据, 同时采用遗传算法优化该模型的核函数类型、核函数参数及错误惩罚因子等参数, 建立了 GA-SVM 模型, 提高了模型参数组合优化选择的效率和预测精度。最后结合实例验证, 并与标准 SVM 方法和 BP 神经网络方法比较。预测效果表明: 所提出的 GA-SVM 优化模型在超短期风电功率预测上具有更优的学习能力和泛化能力。

关键词: 风电场功率预测; 支持向量机; 遗传算法; 超短期预测

Ultra-short-term wind power forecasting based on SVM optimized by GA

LIU Aiguo, XUE Yuntao, HU Jianglu, LIU Luping

(School of Information Engineering, Nanchang University, Nanchang 330031, China)

Abstract: Research on wind power prediction technology is of great significance for mitigating the effects of randomness on the output power of the power generated by the system. First of all, the sample data is formed combined with wind power monitoring system of history in the database data and environmental parameters, at the same time genetic algorithm is used to optimize the model parameters, such as kernel function type, the kernel function parameter and error warning factor. The GA-SVM model is established which improves the efficiency of the model parameter combination optimization choice and the prediction precision. At last, based on the example verification, the standard SVM method and BP neural network method are compared. Prediction results show that the proposed GA-SVM optimization model on the ultra-short-term wind power prediction has a better learning ability and generalization ability.

Key words: wind farm power prediction; SVM; genetic algorithm; ultra-short-term prediction

中图分类号: TM715 文献标识码: A 文章编号: 1674-3415(2015)02-0090-06

0 引言

风电系统输出功率的波动性和间歇性对电网的电能质量及其稳定性将产生较大的影响, 且给电网调度带来了困难与挑战, 在目前建设坚强智能电网的背景下, 风电场功率预测是有效的解决途径^[1-3]。

文献[4-6]分析了风电场的动态特性, 并从机理分析进行物理建模, 但整个建模过程复杂且在实际风场中部分参数测取困难, 所以实际预测精度难以保证。文献[7]通过从分段函数和整体建模两个角度比较各种模型的准确程度, 得到了基于灰色模型的风电功率特性曲线的函数模型, 但仅适合在线运行, 另外, 为了减小突变点的误差, 需要收集更多的风速时间序列数据。文献[8]提出了运用聚类方法对历史风速数据进行处理, 实现了历史数据的自动分类, 并按照相似度最大的原则, 选择合适的训练样本,

运用时间序列方法建立风速预测模型, 预测精度虽有提高, 但对数据的挖掘技术提出了更高的要求。文献[9]从分析风速序列的非线性和非平稳性特征出发, 将一种基于聚类经验模态分解(EEMD)和最小二乘支持向量机(LSSVM)的组合预测模型引入到风速预测中, 但其没有考虑核函数的选取对LSSVM预测精度的影响。文献[10]构造了用于风速序列预测分析的迟滞神经网络, 并采用卡尔曼滤波方法将其与ARMA模型相融合, 实现了风速序列的混合预测, 但只能在有效范围内减小了单一预测机理造成的误差积累。文献[11]提出了一种基于连续时间段聚类的支持向量机风电功率预测方法, 在准确度和稳定性上都有了一定提升, 但其指定分类方案需随应用环境而不断改变。文献[12]总结了目前国内外对风电功率波动特性的不同描述方法和量化指标, 归纳了多台风机组成的风场对输出功率波动的平滑

作用和规律, 但是其量化指标对持续时间的考虑较少, 没有重视波动持续时间的影响, 因此对风电功率在不同时间尺度上的波动情况分析还存在一定的缺陷。文献[13]针对光伏电站日前小时短期出力预测问题, 提出一种基于集合经验模态分解和支持向量机的EEMD-SVM 组合模型预测方法, 但分解方法依赖经验判断, 给预测精度带来了一定的不确定度。目前常用的预测方法大都是统计模型与物理模型的结合, 统计模型在短期和超短期预测中有着较大的优势, 又由于物理模型往往模式固定, 训练时间长, 对硬件要求高, 对其优化较困难。故如果对风电场功率短期及超短期预测的统计模型算法进行研究, 预测精度和预测的适应度会有相应改善。

针对目前的研究现状, 本文结合支持向量机非线性拟合及小样本学习的优势, 并针对 SVM 参数的寻找仅凭借经验在寻优方面的问题, 提出了采用遗传算法优化模型核函数、核函数参数和错误惩罚因子等参数的方法, 建立了 GA-SVM 超短期风电预测模型。以期提高其运算速度与预测精度。

1 SVM 方法原理

支持向量机的回归预测模型最大的优点就是能有效地克服常用预测方法预测结果偏差过大和存在学习、维数偏高带来的计算问题以及局部极值等问题, 可为风电功率预测系统提供一个可靠的预测模型。

1.1 支持向量机回归基本理论

回归问题即是根据一个新的输入数据样本 x , 推断得到它对应的实数输出 y 。规定的样本集合为 $\{(x_1, y_1), \dots, (x_l, y_l)\}$ 。其中: $x_i \in R^n, y_i \in R, i = 1, 2, 3, \dots, l$ 。寻找 R^n 上的一个函数 $f(x)$, 继而用 x 所对应的 y 值。

SVM 估计回归的基本思想是通过一个非线性的映射将输入空间的数据 x 映射到高维特征空间 G 中, 并在这个空间进行线性回归。

给定的样本数据值为 $\{x_i, y_i\}, i = 1, 2, 3, \dots, s(x_i \in R^n, y_i \in R)$ 。其中 y_i 为期望值; s 为数据点的总数。

SVM 通过引入损失函数来解决回归问题。本文采用函数 $y = f(x) = [w\phi(x)] + b$ 。对优化目标取极值

$$\begin{aligned} \min Q = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \zeta_i) \\ \text{s.t.} \left\{ \begin{array}{l} y_i - [w\phi(x_i)] - b \leq \varepsilon + \xi_i \\ [w\phi(x_i)] + b - y_i = \varepsilon_i + \zeta_i \\ \xi_i, \zeta_i \geq 0, i = 1, l, s \end{array} \right. \quad (1) \end{aligned}$$

式中: C 为错误惩罚因子; ξ_i 和 ζ_i 为松弛因子; ε 为损失函数。损失函数可用稀疏数据点表示决策函数。引入具有良好效果的损失函数:

$$L_\varepsilon(y) = \begin{cases} 0 & |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon & |f(x) - y| \geq \varepsilon \end{cases} \quad (2)$$

引入拉格朗日乘子 a_i 和 b_i , 把凸优化问题简化为最大化二次型。

$$\begin{aligned} \max W(a_i, b_i) = & \sum_{i=1}^n y_i (a_i - b_i) - \varepsilon \sum_{i=1}^n (a_i + a'_i) - \\ & \frac{1}{2} \sum_{i,j=1}^n (a_i - a'_i)(a_j - a'_j)(x_i - x'_j) \quad (3) \\ \text{s.t.} \left\{ \begin{array}{l} \sum_{i=1}^n a_i = \sum_{i=1}^n a'_i \\ 0 \leq a_i, a'_i \leq C \end{array} \right. \quad i = 1, 2, \dots, n \end{aligned}$$

式中: C 用于控制模型的复杂度和逼近误差的折中, C 越大, 拟合度越好; ε 用于控制回归逼近误差和模型的泛化能力。

用支持向量回归解决非线性问题时, 主要采用的程序是经过某种变换函数将非线性的数据变量映射到高维数据空间, 然后再对这个高维的特征空间进行线性回归。而实现这种从低维到高维的转换是通过核函数 $k(x, x_i) = \phi(x)\phi(x_i)$ 实现的。常用的核函数如表 1 所示。引入核函数代替内积计算后, 优化目标整合为

$$\begin{aligned} \max W(a_i, b_i) = & \sum_{i=1}^n y_i (a_i - b_i) - \varepsilon \sum_{i=1}^n (a_i + a'_i) - \\ & \frac{1}{2} \sum_{i,j=1}^n (a_i - a'_i)(a_j - a'_j)k(x_i - x'_j) \quad (4) \\ \text{s.t.} \left\{ \begin{array}{l} \sum_{i=1}^n a_i = \sum_{i=1}^n a'_i \\ 0 \leq a_i, a'_i \leq C \end{array} \right. \quad i = 1, 2, \dots, n \end{aligned}$$

表 1 常用的核函数
Table 1 Common kernel function

序号	名称	表达式
1	线性核函数	$k(u, v) = (u \cdot v)$
2	多项式核函数	$k(u, v) = (r(u \cdot v) + \text{coef}0)^d$
3	REF 核函数	$k(u, v) = \exp(-r u - v ^2)$
4	Sigmoid 核函数	$k(u, v) = \tanh(r(u - v) + \text{coef}0)$

1.2 SVM 预测方法

对于时间序列 $\{x_1, x_2, x_3, \dots, x_n\}, i = 1, 2, 3, \dots, n$, $\{x_n\}$ 是预测的目标值, 建立自相关输入与输出 $y_n = \{x_n\}$ 之间的映射关系: $R^m \rightarrow R$, m 为嵌入维数。得到用于支持向量机学习的样本为

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-m} & x_{n-m+1} & \cdots & x_n \end{bmatrix}, Y = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_n \end{bmatrix}$$

回归函数为

$$y_i = \sum_{j=1}^{n-m} (a_i - a'_j) k(\overline{a_i} - \overline{a_j}) + b, j = m+1, \dots, n \quad (5)$$

第一步预测为

$$x_{n+1} = \sum_{i=1}^{n-m} (a_i - a'_i) k(\overline{a_i} - \overline{a_{n-m+1}}) + b \quad (6)$$

式中: $\overline{a_{n-m+1}} = \{x_{n-m+1}, x_{n-m+2}, \dots, x_n\}$, 得到一个样本

$$\overline{a_{n-m+2}} = \{x_{n-m+2}, x_{n-m+3}, \dots, x_n, x_{n+1}\}, \text{ 因此 } x_{n+2} = \sum_{i=1}^{n-m} (a_i - a'_i) k(x_i, x_{n-m+2}) + b.$$

经推导可得

$$x_{n+l} = \sum_{i=1}^{n-m} (a_i - a'_i) k(\overline{a_i} - \overline{a_{n-m+l}}) + b \quad (7)$$

式中, $\overline{a_{n-m+l}} = \{x_{n-m+l}, x_{n-m+l+1}, \dots, x_{n+l-1}\}$ 。

2 基于遗传算法优化 SVM 的实现

支持向量机的核函数是将非线性可分样本转换到线性可分的特征空间, 不同的核选择使 SVM 产生的分类超平面不同, 因此核函数的改变会使 SVM 产生较大的差异性, 对 SVM 的性能有直接的影响。而核函数参数 r 的改变实际上是改变映射函数参数, 改变了函数关系, 从而也改变样本映射特征空间的复杂程度, 所以 SVM 性能的优劣较大程度也受到核函数参数 r 的影响^[14-15]。同时, 惩罚因子 C 是实现在错分样本比例和算法复杂程度之间的折衷, 在确定的特征子空间中调节学习机器置信范围和经验风险的比例, 对 SVM 的泛化推广能力也有较大的影响^[16-17]。基于以上分析, 针对当前 SVM 参数选择技术存在的不足, 本文试图实现同时对 SVM 核函数、核函数参数和错误惩罚因子参数组合的选择和优化, 以期提高参数选择速度和效率, 提高 SVM 学习能力和泛化能力^[18]。由于遗传算法是一种借鉴生物界自然选择和自然遗传机制的随机搜索算法, 被广泛用于解决优化问题。它的优越性主要表现在: 具有群体搜索和内在启发式随机搜索的特性, 不易陷入局部最优; 遗传算法固有的并行性和并行计算的能力, 非常适用于大规模并行计算。本文选取该算法对 SVM 参数组合进行优化。总结归纳 GA(Genetic Algorithm)优化 SVM 预测求解过程的实现流程图如图 1。

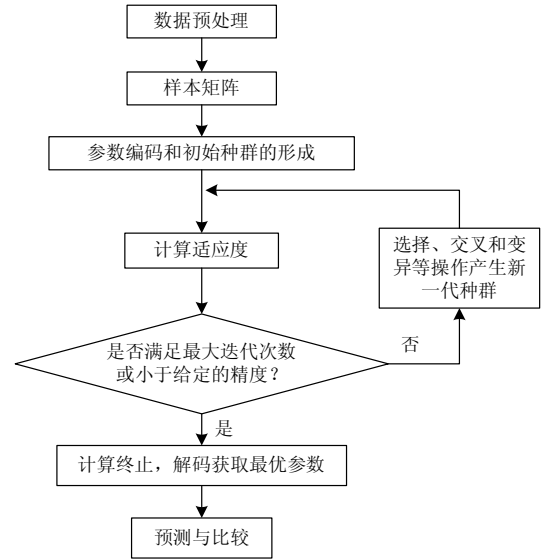


图 1 优化及预测流程图

Fig. 1 Optimization and forecasting flowchart

第一步: 数据预处理。对于风电历史功率, 通过小波分析去噪声处理不良数据和缺失数据。对于环境参数, 本文采集风速, 风向, 大气温度、大气压等。

第二步: 形成样本矩阵。构造模型训练样本、训练目标、测试样本和测试目标, 并写成满足运行需要的矩阵形式。其中, 为不忽略众多环境因素对预测的影响, 把风速、风向、大气温度、大气压一起作为模型的输入数据, 即训练样本和测试样本, 而以历史功率数据作为预测输出值, 即训练目标和测试目标^[19]。

第三步: 编码和初始种群的产生。在本文中, 对支持向量机的核函数类型、惩罚因子和径向基核参数采用了二进制编码方式对其进行编码, 且初始种群是随机产生的。

在编码设计时核函数选型编码为其所对应的序号采用两位二进制编码, 对应核函数类型如表 1 所示的四种类型; 核函数参数 r 和惩罚因子 C 在取值范围采用二进制编码, 编码分别为 t_1 位和 t_2 位的二进制串, 将 $t_2 + t_1 + 2$ 位二进制编码组合就得到个体染色体基因串, 如图 2。

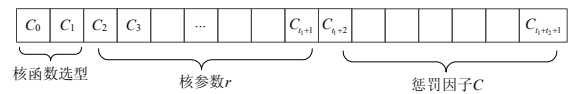


图 2 基因串结构示意图

Fig. 2 Gene string structure

第四步: 计算适应度。定义测试样本的平均相对百分比误差 $MAPE$ 作为适应度函数, 评价种群。

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{\left| \hat{p}_i - p_i \right|}{p_i} \times 100\% \quad (8)$$

式中: N 表示预测个数; p_i 表示实际功率值; \hat{p}_i 表示预测功率值。并根据个体的适应度值和选择原则选择进入下一代的个体。

第五步: 采用交叉、变异等遗传算子对当前一代种群进行处理后产生下一代群体。

第六步: 判断终止条件是否满足。若满足则进入下一步, 否则转到第四步继续执行; 继而得到使得模型最优化的一组支持向量机参数。

第七步: 预测与比较。使用对应的支持向量机的最优参数来实现预测, 并用平均相对百分比误差

$MAPE$ 评价预测性能, 最后与传统方法进行综合比较。

3 实例分析

预测系统的样本数据包括风电监控系统数据库中的历史功率数据和环境参数。风电机组数据取自国内某风电场采集与监控系统(SCADA), 采集数据以 15 min 为记录间隔, 将 2012 年 8 月 24 日至 2012 年 9 月 10 日期间的数据作为训练集, 将 2012 年 9 月 11 日至 2012 年 9 月 17 日的数据作为测试集。预测提前 15 min 风电场超短期出力, 并实现滚动预测。最后与该风力发电场预测时刻相同时间点的实测运行数据对比, 采用平均相对误差(MAPE)评价分析模型的精确度。如表 2 所示。

表 2 标准 SVM 与 GA-SVM 预测结果对比

Table 2 Comparison between standard SVM and GA-SVM prediction results

预测类型	SVM 参数			训练正确率/%	预测准确率/%
	核函数类型	核函数参数	错误惩罚因子		
经验选择	1		10	81.5	84.33
	2	1	10	92	85.45
	3	10	1	74	73.75
	4	1	10	67.4	94.53
GA 优化	2	7.35	14.91	99.5	99.47

从表 2 可以看出, 基于遗传算法优化支持向量机模型的预测结果, 无论是训练准确率还是预测准确率, 都优于经验选择参数的标准支持向量机, 可以认为优化支持向量机的性能从整体上要优于标准支持向量机。图 3 为 GA-SVM 的预测结果图。

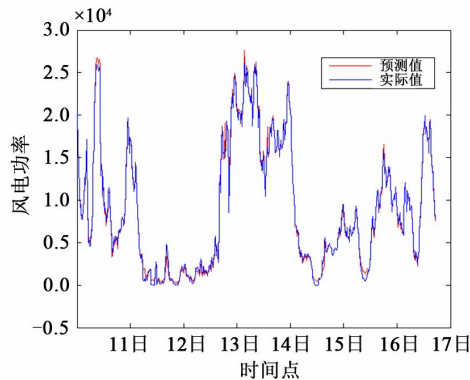


图 3 风能发电功率预测曲线

Fig. 3 Power generation forecasting curve

为了验证采用 GA-SVM 算法的优越性, 在保证实验数据不变情况下, 同时分别采用标准 SVM 和

BP 神经网络进行风电功率预测。得到预测数据与实际风电值的误差如图 4 所示。

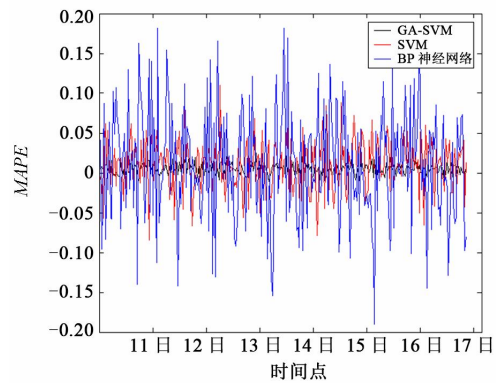


图 4 预测误差曲线

Fig. 4 Forecasting error curve

由图 4 可知, SVM 和 GA-SVM 要预测的稳定度较高, 而且经过 GA 优化后的 GA-SVM 预测精度也有明显提高。为了精确计算出各情形中风电功率预测的误差值, 本文采用了平均相对百分比误差 $MAPE$ 评价预测性能, 如表 3 所示。

在以上三种方法的风电功率中, 通过对平均相对误差计算可以看出, 经 GA 优化后的 SVM 预测

模型的预测能力是最好的, 其平均相对百分比误差要比 SVM 要低出约 4.0951%, 较 BP 神经网络约低 11.835%。

表 3 各预测方法误差
Table 3 Forecasting errors

方法	GA-SVM	SVM	BP 神经网络
MAPE /%	0.530 4	4.625 5	12.365 4

4 结论

本文针对如何降低风电并网对电网的安全经济运行影响的问题, 采用风电监控系统数据库中的历史风电功率数据和环境参数作为实验数据, 并考虑到 SVM 的核函数、核函数参数 r 和错误惩罚因子 C 等难以确定的问题, 提出了 GA 优化后的 GA-SVM 预测模型。结合实例验证, 并与标准 SVM 方法和传统 BP 神经网络方法比较, 得出本文提出的 GA-SVM 参数优化模型的学习能力和泛化能力更优, 平均相对百分比误差要比 SVM 要低出约 4.0951%, 较 BP 神经网络减少约 11.835%。

参考文献

- [1] 张坤, 毛承雄, 陆继明, 等. 基于储能的直驱风力发电系统的功率控制[J]. 电工技术学报, 2011, 26(7): 7-14.
ZHANG Kun, MAO Chengxiong, LU Jiming, et al. Power control strategy of directly driven wind turbine with energy storage system[J]. Transactions of China Electrotechnical Society, 2011, 26(7): 7-14.
- [2] 田云飞, 张立涛, 徐昊亮. 大规模风电并网对电力系统的影响及应对措施[J]. 电气自动化, 2013, 35(3): 54-56.
TIAN Yunfei, ZHANG Litao, XU Haoliang. Impact of large scale grid-connected wind power on power system and countermeasures[J]. Electric Automation, 2013, 35(3): 54-56.
- [3] 李加升, 黄文清, 戴瑜兴. 基于自定义阈值函数的小波去噪算法[J]. 电力系统保护与控制, 2008, 36(19): 21-24.
LI Jiasheng, HUANG Wenqing, DAI Yuxing. Wavelet-based power quality disturbances de-noising by customized thresholding[J]. Power System Protection and Control, 2008, 36(19): 21-24.
- [4] 崔杨, 穆钢, 刘玉, 等. 风电功率波动的时空分布特性[J]. 电网技术, 2011, 35(2): 110-114.
CUI Yang, MU Gang, LIU Yu, et al. Spatiotemporal distribution characteristic of wind power fluctuation[J]. Power System Technology, 2011, 35(2): 110-114.
- [5] 林卫星, 文劲宇, 艾小猛, 等. 风电功率波动特性的概率分布研究[J]. 中国电机工程学报, 2012, 32(1): 38-46.
LIN Weixing, WEN Jinyu, AI Xiaomeng, et al. Probability density function of wind power variations[J]. Proceedings of the CSEE, 2012, 32(1): 38-46.
- [6] 曹娜, 赵海翔, 任普春, 等. 风电场动态分析中风速模型的建立及应用[J]. 中国电机工程学报, 2007, 27(36): 68-72.
CAO Na, ZHAO Haixiang, REN Puchun, et al. Establish and application of wind speed model in wind farm dynamic analysis[J]. Proceedings of the CSEE, 2007, 27(36): 68-72.
- [7] 李俊芳, 张步涵, 谢光龙, 等. 基于灰色模型的风速-风电功率预测研究[J]. 电力系统保护与控制, 2010, 38(19): 151-159.
LI Junfang, ZHANG Buhan, XIE Guanglong, et al. Grey predictor models for wind speed-wind power prediction[J]. Power System Protection and Control, 2010, 38(19): 151-159.
- [8] 方江晓, 周晖, 黄梅. 基于统计聚类分析的短期风电功率预测[J]. 电力系统保护与控制, 2011, 39(11): 67-73.
FANG Jiangxiao, ZHOU Hui, HUANG Mei. Short-term wind power prediction based on statistical clustering analysis[J]. Power System Protection and Control, 2011, 39(11): 67-73.
- [9] 王贺, 胡志坚, 张翌晖. 基于聚类经验模态分解和最小二乘支持向量机的短期风速组合预测[J]. 电工技术学报, 2014, 29(4): 237-245.
WANG He, HU Zhijian, ZHANG Yihui. A hybrid model for short-term wind speed forecasting based on ensemble empirical mode decomposition and least squares support vector machines[J]. Transactions of China Electrotechnical Society, 2010, 29(4): 237-245.
- [10] 修春波, 任晓, 李艳晴. 基于卡尔曼滤波的风速序列短期预测方法[J]. 电工技术学报, 2014, 29(2): 253-259.
XIU Chunbo, REN Xiao, LI Yanqing. Short-term prediction method of wind speed series based on Kalman filtering fusion[J]. Transactions of China Electrotechnical Society, 2014, 29(2): 253-259.
- [11] 丁志勇, 杨苹, 杨曦, 等. 基于连续时间段聚类的支持向量机风电功率预测方法[J]. 电力系统自动化, 2012, 36(14): 131-135.
DING Zhiyong, YANG Ping, YANG Xi, et al. An analysis on impacts and characteristics of stator voltage of DFIG-based wind turbine generator under grid short

- circuit[J]. Automation of Electric Power Systems, 2012, 36(14): 131-135.
- [12] 廖旎焕, 胡智宏, 马莹莹, 等. 电力系统短期负荷预测方法综述[J]. 电力系统保护与控制, 2011, 39(1): 147-152.
- LIAO Nihuan, HU Zhihong, MA Yingying, et al. Review of the short-term load forecasting methods of electric power system[J]. Power System Protection and Control, 2011, 39(1): 147-152.
- [13] 茆美琴, 龚文剑, 张榴晨, 等. 基于 EEMD-SVM 方法的光伏电站短期出力预测[J]. 中国电机工程学报, 2013, 33(34): 17-24.
- MAO Meiqin, GONG Wenjian, CHANG Liuchen, et al. Short-term photovoltaic forecasting based on EEMD-SVM combined method[J]. Proceedings of the CSEE, 2013, 33(34): 17-24.
- [14] FUKUSHI D, SHICHIRI M, SUGIYAMA S, et al. Scanning near-field optical/automatic force microscopy detection of fluorescence in Situ hybridization signals beyond the optical limit[J]. Exp Cell Res, 2003, 289: 237-244.
- [15] HAPELLE O, VAPNIK V, BOUSQET O, et al. Choosing multiple parameters for support vector machines[J]. Machine Learning, 2002, 46(1): 131-159.
- [16] HUANG C M, LEE Y J, LIN D K J, et al. Model selection for support vector machine via uniform design[J]. Computational Statistics Data Analysis, 2007, 52(1): 335-346.
- [17] KUMAR A. Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in National electricity market[J]. IET Generation, Transmission & Distribution, 2010, 4(1): 310-315.
- [18] 夏可青, 赵明奇, 李扬. 用于多目标无功优化的自适应遗传算法[J]. 电网技术, 2006, 30(13): 55-60.
- XIA Keqing, ZHAO Mingqi, LI Yang. A self-adaptive genetic algorithm for multi-objective reactive power optimization[J]. Power System Technology, 2006, 30(13): 55-60.
- [19] 蒋小亮, 蒋传文, 彭明鸿. 基于时间连续性及季节周期性的风速短期组合预测方法[J]. 电力系统自动化, 2010, 34(15): 75-79.
- JIANG Xiaoliang, JIANG Chuanwen, PENG Minghong. A short-term combination wind speed forecasting method Considering seasonal periodicity and time-continuity[J]. Automation of Electric Power Systems, 2010, 34(15): 75-79.
-
- 收稿日期: 2014-04-18; 修回日期: 2014-10-15
- 作者简介:
- 刘爱国 (1965-), 男, 副教授, 研究方向为电力系统分析与经济运行. E-mail: liuaiguo@ncu.edu.cn