

# 基于混核 LSSVM 的批特征风功率预测方法

刘畅<sup>1,2</sup> 郎劲<sup>3,4</sup>

**摘要** 针对风电场风功率预测问题, 利用历史风功率、气象数据和测风塔实时数据等相关信息, 提出了带有批特征的混核最小二乘支持向量机 (Hybrid kernel least squares support vector machine, HKLSSVM) 方法, 建立风电场风功率预测模型. 为了增强模型的适应性, 设计改进的差分进化算法对模型参数进行优化, 并利用稀疏选择方法来选取合适的训练样本集, 缩短建模时间, 保证预测模型精度. 根据风场风机的地理位置分布情况, 提出批划分的建模策略, 对相近地理位置的风机进行组批, 替代传统风场风功率预测方法. 通过风场中实际数据进行测试, 实验结果表明与其他预测方法相比, 本文提出的方法能够提高预测精度和效率, 减少风电波动性对电网的影响, 从而提高电网的安全性和可靠性.

**关键词** 风功率预测, 批特征, 混核最小二乘支持向量机, 差分进化, 稀疏选择

**引用格式** 刘畅, 郎劲. 基于混核 LSSVM 的批特征风功率预测方法. 自动化学报, 2020, 46(6): 1264–1273

**DOI** 10.16383/j.aas.c180103

## Wind Power Prediction Method Using Hybrid Kernel LSSVM With Batch Feature

LIU Chang<sup>1,2</sup> LANG Jin<sup>3,4</sup>

**Abstract** For the wind power prediction problem in a wind farm, this paper collects some related data such as historical wind power data, meteorological data, and wind speed data sampled by anemometer tower. Then, a wind power prediction method with batch feature is proposed, which is based on hybrid kernel least squares support vector machine (HKLSSVM). It is used to establish the wind power prediction model in the wind farm. To enhance the model's adaptability, an improved differential evolution algorithm is designed to optimize the model parameters, and a sparse selection method is used to select the appropriate training samples set. Thus, the modeling time is shortened and the prediction model accuracy is guaranteed. According to the location distribution of wind turbines in the wind farm, a modeling strategy based on batch partition is proposed, some wind turbines at similar locations can be clustered by batch strategy, which is used instead of the traditional wind power prediction methods in the wind farm. The proposed model is tested through the real data in the wind farm. Experimental results show that the proposed method can improve the accuracy and efficiency of wind power prediction compared with other prediction methods, and can reduce the effect of the wind fluctuation. Hence it can ensure the safety and reliability of the power grid.

**Key words** Wind power prediction, batch feature, hybrid kernel least squares support vector machine (HKLSSVM), differential evolution (DE), sparse selection

**Citation** Liu Chang, Lang Jin. Wind power prediction method using hybrid kernel LSSVM with batch feature. *Acta Automatica Sinica*, 2020, 46(6): 1264–1273

收稿日期 2018-02-26 录用日期 2018-05-07  
Manuscript received February 26, 2018; accepted May 7, 2018  
国家重点研究发展计划基金 (2016YFB0901900), 国家自然科学基金重点国际合作项目 (71520107004), 流程工业综合自动化国家重点实验室基础研究项目基金 (2013ZCX02), 111 引智基地基金 (B16009) 资助

Supported by National Key Research and Development Program of China (2016YFB0901900), the Major International Joint Research Project of National Natural Science Foundation of China (71520107004), the State Key Laboratory of Synthetical Automation for Process Industrial Fundamental Research Funds (2013ZCX02), and the 111 Project (B16009)

本文责任编辑 孙秋野

Recommended by Associate Editor SUN Qiu-Ye

1. 智能工业数据解析与优化教育部重点实验室 (东北大学) 沈阳 110819 2. 东北大学工业与系统工程研究所 沈阳 110819 3. 东北大学辽宁省制造系统与物流优化重点实验室 沈阳 110819 4. 东北大学流程工业综合自动化国家重点实验室 沈阳 110819

1. Key Laboratory of Data Analytics and Optimization for

随着世界经济的快速发展, 未来对能源的需求会大幅增加. 考虑到传统化石能源不断的消耗对生态环境造成的严重影响, 绿色和可再生能源如风能、太阳能、地热能、潮汐能等已成为各国政府努力发展方向, 在我国的“十三五规划”中也提出了对能源安全、绿色生产方面的要求. 风电作为一种清洁能源, 能够缓解能源危机, 并能够部分代替化石燃料的使用, 减少温室气体的排放<sup>[1-3]</sup>. 由于风场风速具有

Smart Industry (Northeastern University), Ministry of Education, Shenyang 110819 2. Institute of Industrial and Systems Engineering, Northeastern University, Shenyang 110819 3. Liaoning Key Laboratory of Manufacturing System and Logistics, Northeastern University, Shenyang 110819 4. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819

随机性与波动性,使得发电调度运行的不确定性显著增加,甚至威胁电力系统的安全运行.因此,高精度的风电功率预测技术对于电力系统的调度、运行和风电场的生产和维护以及提高风电场利用效率都具有十分重要的意义.

对于风电场风功率的预测一般可分为间接和直接两种方式,一种是先预测风速,再预测风功率;另一种是直接预测风功率.间接的方法是根据风速的变化拟合风场功率特性曲线.但随着科技的不断发展,各种多源化的数据信息为风功率预测提供了更多的参考.基于数据的风功率预测主要是利用风电场的风功率、风速、地形地貌、数值天气预报等多种数据来建立风电场输出功率的预测模型,在保证风电场机组的设备正常运行的状态下,可以直接通过输入的特征数据预测出风电场未来的有功功率.

一般来说,风功率的预测分为3个不同的阶段:短期预测、中期预测、长期预测<sup>[4-6]</sup>.目前研究人员多数采用数据解析的方法对风场不同阶段的风功率进行预测,例如:自回归滑动平均(Auto-regressive and moving average, ARMA)<sup>[7]</sup>、极限学习机(Extreme learning machine, ELM)<sup>[8]</sup>、深度神经网络(Deep neural network, DNN)<sup>[9]</sup>等.此外,一些混合人工智能方法也广泛应用在风功率预测中.Pousinho等<sup>[10]</sup>提出了一种混合的PSO-ANFIS方法对短期风功率进行预测.Haque等<sup>[11]</sup>结合小波变换和模糊神经网络来预测风场风功率.Heinermann等<sup>[12]</sup>提出一种机器学习集成方法对风功率进行预测,将决策树与支持向量回归机相结合,不仅获得了较高的预测精度,而且减少了运算时间.此外,一些时空模型<sup>[13]</sup>在最近几年也得到了广泛关注.

由于风场风速与周围环境的变化情况有关,因此,现场积累的数据具有高维度特征.针对于这类高维数据的特点<sup>[14]</sup>,一些基于核函数的预测方法得到了有效的验证<sup>[15-16]</sup>,并由此发展为多核函数之间的相互学习<sup>[17-18]</sup>.例如,Ouyang等<sup>[19-20]</sup>分别提出了多核函数学习以及多模型混合学习的风功率预测方法.Khosravi等<sup>[21]</sup>结合了人工神经网络,支持向量回归以及自适应神经推理系统对风速和风向进行预测.Yuan等<sup>[22]</sup>利用一种基于重力搜索方法(Gravity search algorithm, GSA)来优化最小二乘支持向量机(Least squares support vector machine, LSSVM)模型的参数,使得预测结果有一个更高的准确率.Zhang等<sup>[23]</sup>将固有时间尺度分解(Intrinsic time-scale decomposition, ITD)与LSSVM相结合,提高了风功率的预测精度.

基于以上的研究内容,本文采用基于混核的最小二乘支持向量机(Hybrid kernel least squares support vector machine, HKLSSVM)方法对风功

率进行预测.为了增强模型的自适应性,设计改进的差分进化(Improved differential evolution, IDE)算法对模型的参数进行优化,同时利用稀疏选择策略来选取合适的训练数据集,以缩短模型训练时间.针对于风电场风机的地形地貌特点,提出了批划分的策略,对风场风功率进行实时预测,有效地解决了现场实际风功率预测精度不高的难题.

## 1 风功率预测方法

在最小二乘支持向量机的建模过程中,核函数用来将低维特征空间中的非线性问题映射为高维特征空间中的线性问题.核函数的特征对于最小二乘支持向量机的学习和预测都有很重要的影响.本文将高斯核函数与多项式核函数相结合,采用HKLSSVM方法对风功率进行预测.为了得到更好的模型参数,对DE算法进行了改进,在变异过程中加入了新的变异算子,设计了IDE算法对预测模型进行参数优化.为了减小训练样本个数,利用一种稀疏选择策略,对模型的训练样本进行筛选,从而找到合适的训练样本集.最后,根据风场风机的分布情况,提出了批划分的策略,达到预测整个风场的总风功率的目的.

### 1.1 HKLSSVM

LSSVM<sup>[24-25]</sup>最早由Suykens等提出,它是基于支持向量机(Support vector machine, SVM)的一种改进.LSSVM与SVM的区别在于,前者将复杂的二次规划问题转换为线性方程组问题进行求解,并将SVM优化问题里的不等式约束替换为等式约束,提高了算法的运行速度,降低了计算的复杂性.

基于LSSVM的优势,本文设计的HKLSSVM方法步骤如下:

**步骤1.**利用构造的最优决策函数,将非线性估计函数转化为高维特征空间中线性估计函数:

$$f(\mathbf{x}_i) = \mathbf{w}^T \times \mathbf{x}_i + b, \quad i = 1, 2, \dots, N \quad (1)$$

其中, $N$ 为样本的个数, $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,D})$ 表示输入数据, $D$ 为输入数据的特征数, $f(\mathbf{x}_i)$ 为输出值, $\mathbf{w}^T$ 为回归系数, $b$ 为偏差.

**步骤2.**利用结构风险最小化原理,将回归问题表示成带有约束的优化问题

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} J(\mathbf{w}, \xi) &= \frac{1}{2} \|\mathbf{w}\|^2 + \frac{\gamma}{2} \sum_{i=1}^N \xi_i^2 \\ \text{s.t.} \quad y_i &= \mathbf{w}^T \times \varphi(\mathbf{x}_i) + b + \xi_i, \\ \gamma &\geq 0 \end{aligned} \quad (2)$$

其中, $\gamma$ 为惩罚系数, $\xi_i$ 为松弛变量, $\varphi(\mathbf{x}_i)$ 是一个从低维向高维的映射.

步骤 3. 构造式 (2) 的拉格朗日函数, 即

$$L(\mathbf{w}, b, \xi_i, \alpha_i) = J(\mathbf{w}, \xi_i) - \sum_{i=1}^N \alpha_i \{ \mathbf{w}^T \varphi(\mathbf{x}_i) + b + \xi_i - y_i \} \quad (3)$$

其中,  $\alpha_i$  为拉格朗日乘子.

步骤 4. 利用  $L$  分别对回归系数  $\mathbf{w}$ 、偏差  $b$ 、松弛变量  $\xi_i$  以及拉格朗日乘子  $\alpha_i$  进行求导, 如式 (4) 所示.

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i \varphi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma \xi_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \mathbf{w}^T \varphi(\mathbf{x}_i) + b + \xi_i - y_i = 0 \end{cases} \quad (4)$$

步骤 5. 消去  $\mathbf{w}$  和  $\xi_i$  后, 可得如下线性方程组:

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \Omega + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (5)$$

其中,  $\mathbf{1} = [1, 1, \dots, 1]^T$ ,  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ ,  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ ,  $\Omega = K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$ .

步骤 6. 最终得到的预测模型为

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (6)$$

核函数  $K(\mathbf{x}_i, \mathbf{x}_j)$  主要实现了样本从输入空间到特征空间的映射, 其主要分为两类: 局部核函数和全局核函数. 局部核函数对于样本附近小邻域内的数据具有较强的学习能力, 但对于远距离的样本点缺乏泛化能力; 而全局核函数与之相反, 其泛化性能较强, 但学习能力较弱. 因此, 将两种核函数进行线性组合可以更好地发挥核函数的性能, 并且在改变原映射空间的情形下, 能够保证混核函数的有效性.

本文利用高斯核函数的优势, 选取其为预测模型的局部核函数

$$K_1(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right) \quad (7)$$

其中,  $\sigma$  为高斯核函数中的核宽度参数.

结合多项式核函数的特点, 选取其为预测模型的全局核函数

$$K_2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \cdot \mathbf{x}_j + 1)^d, \quad d \in \mathbf{N}^* \quad (8)$$

其中,  $d$  为多项式核函数参数.

最后, 将  $K_1(\mathbf{x}_i, \mathbf{x}_j)$  与  $K_2(\mathbf{x}_i, \mathbf{x}_j)$  相结合, 以混核函数形式作为 LSSVM 的核函数

$$K(\mathbf{x}_i, \mathbf{x}_j) = a_1 K_1(\mathbf{x}_i, \mathbf{x}_j) + a_2 K_2(\mathbf{x}_i, \mathbf{x}_j) \quad (9)$$

其中,  $a_1$  和  $a_2$  分别为混核函数系数, 且  $0 \leq a_1 \leq 1$ ,  $0 \leq a_2 \leq 1$ ,  $a_1 + a_2 = 1$ .

### 1.2 IDE

为了得到准确的 HKLSSVM 模型参数, 基于差分进化 (Differential evolution, DE)<sup>[26]</sup> 的搜索机制, 本文设计了改进的差分进化 (IDE) 算法分别对 HKLSSVM 的模型参数  $\gamma$ 、 $\sigma$ 、 $d$ , 以及  $a_1$  进行优化, IDE 方法的具体步骤如下:

步骤 1. 初始化相关参数设置, 确定种群规模  $NP$ , 个体维数为  $M$  (即需要分别对  $\gamma$ 、 $\sigma$ 、 $d$  和  $a_1$  进行优化).  $U_q$  和  $L_q$  分别为个体取值的上界和下界, 其中,  $q = 1, 2, \dots, M$ . IDE 的交叉概率因子为  $CR$ , 变异算子为  $F_1$ , 最大迭代次数为  $g_{\max}$ .

步骤 2. 在可行域  $[L_q, U_q]$  范围内, 随机产生  $NP$  个解作为初始种群, 种群中的每个个体为  $\mathbf{v}_p^g = (v_{p,1}^g, v_{p,2}^g, \dots, v_{p,M}^g)$ , 其中,  $p = 1, 2, \dots, NP$ ,  $q = 1, 2, \dots, M$ ,  $g$  为当前代数.

步骤 3. 以模型的预测值与实际值均方根误差最小为目标, 计算每个个体的自适应函数值  $F(\mathbf{v}_p^g)$

$$F(\mathbf{v}_p^g) = \min \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - f(\mathbf{x}_i, \mathbf{v}_p^g))^2} \quad (10)$$

其中,  $Y_i$  为实际值,  $f(\mathbf{x}_i, \mathbf{v}_p^g)$  为预测值,  $N$  为训练数据个数.

步骤 4. 采用改进的变异方式对种群中每个个体进行变异与交叉操作, 加入了新的变异算子  $F_2$ , 得到交叉后的个体

$$u_{p,q}^g = \begin{cases} v_{p,q}^g + F_1 \times (v_{r_1,q}^g - v_{r_2,q}^g) + \\ \quad F_2 \times (v_{\text{best},q}^g - v_{r_3,q}^g), & \text{若 } \text{rand}_{p,q}^g(0, 1) \leq CR \\ v_{p,q}^g, & \text{否则} \end{cases} \quad (11)$$

$$F_2 = F_1 \times \frac{g}{g_{\max}} \quad (12)$$

其中,  $v_{\text{best},q}^g$  为当前种群中最好个体,  $r_1$ 、 $r_2$ , 以及  $r_3$  分别是在  $NP$  内随机选择的三个互不相同的整数, 同时也不等于目标个体的索引值. 当生成的随机数  $\text{rand}_{p,q}^g(0, 1)$  小于等于交叉系数  $CR$  时, 子代个体中的元素就赋值为对应的变异个体, 否则该元素赋值为对应父代个体元素.

**步骤 5.** 判断交叉后的个体是否超出了边界范围, 如超出边界范围则进行修复, 使其在可行域内随机产生, 作为新的交叉后的个体; 否则, 执行步骤 6.

**步骤 6.** 交叉操作后得到的子代个体要与对应的父代目标个体进行自适应函数值的比较, 二者中较优的一个将作为新的父代目标个体进入后续的迭代搜索过程, 具体选择方式为

$$\mathbf{v}_p^{g+1} = \begin{cases} \mathbf{u}_p^g, & \text{若 } F(\mathbf{u}_p^g) \leq F(\mathbf{v}_p^g) \\ \mathbf{v}_p^g, & \text{否则} \end{cases} \quad (13)$$

其中,  $\mathbf{u}_p^g = (u_{p,1}^g, u_{p,2}^g, \dots, u_{p,M}^g)$ .

**步骤 7.** 判断终止条件是否满足. 当迭代次数达到最大迭代次数  $g_{\max}$  时, 满足算法的终止条件, 算法终止运行, 输出最优解 (即 HKLSSVM 的最佳参数). 否则, 转至步骤 4, 执行  $g+1$ , 继续进行迭代.

### 1.3 稀疏选择

由于 LSSVM 式 (4) 中  $\alpha_i = \gamma \xi_i$ , 导致解向量具有较弱的稀疏性. 基于文献 [27] 提出的稀疏剪枝策略, 本文设计一种带有稀疏选择的 IDE-HKLSSVM (Sparse IDE-HKLSSVM, SIDE-HKLSSVM) 算法, 从原始数据集中选取合适的训练样本集. 具体步骤如下:

**步骤 1.** 利用  $N$  个样本数据训练 IDE-HKLSSVM 模型.

**步骤 2.** 对  $|\alpha_i|$  由低到高进行排序, 设置稀疏阈值  $\lambda$ , 对训练样本进行稀疏选择, 如果  $|\alpha_i| \leq \lambda$ , 则删掉对应的训练样本; 否则, 保留原有的训练样本.

**步骤 3.** 利用 IDE-HKLSSVM 方法重新训练稀疏后的训练样本集.

**步骤 4.** 判断是否满足稀疏选择迭代次数  $\theta$ , 如果满足则稀疏选择策略停止; 如不满足, 则继续执行步骤 2.

从以上稀疏策略可知, 每一次稀疏选择的迭代过程都会删掉部分的训练数据, 使得新的训练数据集具有稀疏性. 并且在每次选择过程中, IDE 都会优化 HKLSSVM 的模型参数, 保证了稀疏后的模型具有较高的预测精度.

### 1.4 风场批划分和批样机的选取

传统策略对实际风功率进行预测时分为两种方式: 1) 利用风场的整体信息来对风场风功率进行预测; 2) 对每台风机进行风功率预测, 再对所有机组风功率进行预测求和, 从而实现风场的总风功率预测. 前者利用了测风塔数据对整个风场的风功率进行预测, 并且以风场的气象数据代替所有风机的气象数据, 从而忽略了每台机组的运行特征及参数. 在实际建设风电场时, 由于风机分布不均匀、地理位置

不同、环境条件和气候变化的影响因素, 产生的风功率会有较大的差别, 导致预测精度不高. 后者可归结为单机求和 (Single turbine summation, STS) 方法, 虽然其考虑到每一台机组的功率变化情况, 保障了自身的预测精度. 但是, 实际风场中机组分布较分散, 并且风机数量大, 每台风机的相关数据不完备, 导致单机求和策略不易操作.

基于上述问题, 本文提出了一种带有批特征的风功率预测方法. 即把地理位置相近和具有相同运行参数的风机划分为同一个批进行预测. 由于风电场划分批的数量不同, 因此每个批内的风机数量也不相同. 依据风场风机的地理位置和历史输出功率、气象数据等信息, 先对批内风机进行聚合批处理, 然后对各个批进行风功率预测, 这样就可以把整个风电场的风功率预测转化为对风场各个批的预测. 批样机的选取过程是先统计各自批内的每台风机的发电情况, 将接近于各自批内所有风机发电量平均值的风机选为批样机. 再利用 IDE-HKLSSVM 方法对每台批样机的风功率进行预测, 然后将批样机预测结果乘以该批样机所处批内的风机数目, 得到各区域内的风功率预测值, 最后将其进行求和, 即可得到整个风电场在未来时段的风功率预测值, 具体公式为

$$Q_o = Q'_o \times h_o, \quad o \in \{1, \dots, E\} \quad (14)$$

$$Q = Q_1 + \dots + Q_E \quad (15)$$

其中,  $Q_o$  为每个区域内的风功率总和,  $Q'_o$  为每个区域内的批样机风功率,  $h_o$  为对应的区域内的风机个数,  $Q$  为风场的总风功率预测值,  $E$  为批样机的总个数.

### 1.5 BSIDE-HKLSSVM 方法

图 1 为本文提出的带有批特征的 SIDE-HKLSSVM (Batch SIDE-HKLSSVM, BSIDE-HKLSSVM) 方法, 具体描述如下:

首先对现场采集的数据进行预处理, 即剔除异常的数据, 对于缺失的数据进行插值补全, 再对处理后的数据进行归一化, 通过 BSIDE-HKLSSVM 方法对风场的风速和风功率分别进行预测. 为了确保获取最佳的模型参数, 提出 IDE 算法对模型参数进行优化. 并利用稀疏选择策略对训练数据集进行筛选. 最后将风场风机进行批划分处理, 即将风场风机划分为多个区域, 对每个区域内的风功率进行预测, 进而得到风场总的风功率.

## 2 实验验证

本实验的数据来自于某风电场的实际数据, 其中以 4 月 1~7 日的部分数据作为训练数据, 以 8~

10 日的数据作为测试数据, 每隔 15 分钟进行一次测试, 实验主要包括两部分: 风速预测和风电功率预测.

法实现. 运行实验的计算机 CPU 主频为 3.4 GHz, 内存为 16 GB, 操作系统为 Windows 7.

### 2.2 聚合批结果

图 2 为某风场的风机分布图, 由于有限电的情况发生, 并且存在一些故障的风机, 其中, 23#、35#、36#、39#、41#, 以及 42# 风机停止运行. 因此在本实验中将不考虑这些风机的划分.

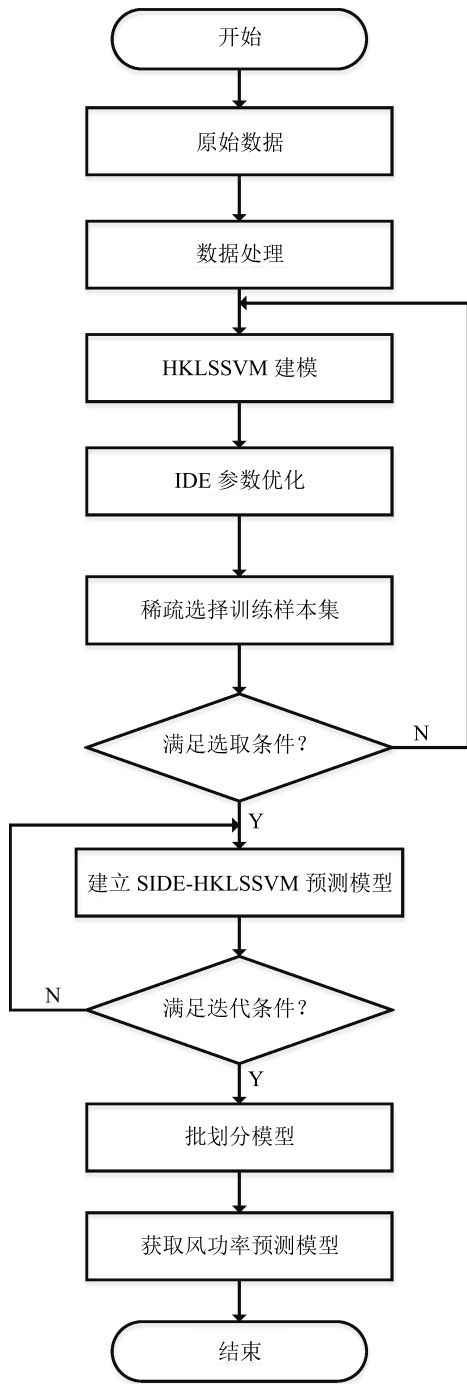


图 1 BSIDE-HKLSSVM 流程图

Fig. 1 The flowchart of BSIDE-HKLSSVM

### 2.1 实验设置

本实验是在 Microsoft Visual Studio 2008 平台上进行操作, 以 C++ 为计算机编程语言进行算

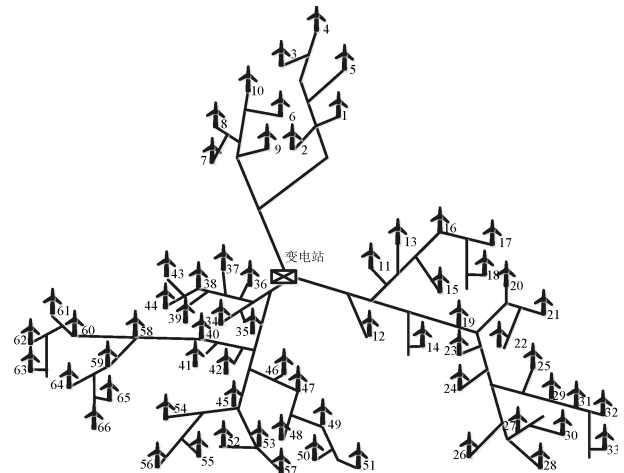


图 2 某风场的风机分布

Fig. 2 The distribution of wind turbines in the wind farm

以整个风场 66 台风机中正常运行的 60 台无故障无数据缺失风机为例, 考虑到每台风机的地理位置分布情况, 对所有风机聚合成批. 依据省调选取的样机个数, 从所有风机中选出 6 台风机作为批样机, 统计各自批内的每台风机的发电情况, 以接近于各自批内所有风机发电量平均值的风机作为各自批样机, 通过数据分析, 6 台批样机分别为 9#、18#、31#、37#、52#, 以及 59#. 以下为批样机的选取过程:

1) 1#~10#, 共 10 台风机. 获取每台风机的发电量数值, 并求得 10 台风机的平均发电量为 142.60 kW·h. 其中, 9# 风机的发电量为 152.46 kW·h, 最接近于批内的平均发电量, 所以选 9# 风机作为批样机. 依据此方法选取以下其他批样机.

2) 11#~22#, 共 12 台风机. 选取的批样机为 18# (其中平均发电量为 99.3275 kW·h, 18# 为 98.93 kW·h).

3) 24#~33#, 共 10 台风机. 选取的批样机为 31# (其中平均发电量为 121.036 kW·h, 31# 为 123.32 kW·h).

4) 34#、37#、38#、40#、43#, 44#, 共 6 台风机. 选取的批样机为 37# (其中平均发电量为 139.2233 kW·h, 37# 为 138.26 kW·h).

5) 45#~57#, 共 13 台风机. 选取的批样机为 52# (其中平均发电量为 102.6562 kW·h, 52# 为 105.29 kW·h).

6) 58#~66#, 共 9 台风机. 选取的批样机为 59# (其中平均发电量为 114.8989 kW·h, 59# 为 119.03 kW·h).

### 2.3 数据处理

以风速预测实验为例, 首先剔除历史数据中的异常数据, 并对实际数据进行插值补全, 再将处理后的数据进行归一化, 从中选出当前时刻风速、当前时刻温度、当前时刻湿度、当前时刻压力、当前时刻风向; 输出为下一时刻风速. 对于风功率预测实验, 从中选出当前时刻空气湿度、当前时刻温度、当前时刻风向、当前时刻气压、当前时刻风速、当前时刻的功率作为输入信息, 将下一时刻功率作为输出信息. 最后, 选取对应批中的风功率预测模型进行训练.

以风功率预测实验为例, 初始化原始数据矩阵

$$\begin{pmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \end{pmatrix} = \begin{pmatrix} z_{1,1} & \cdots & z_{1,7} \\ \vdots & \ddots & \vdots \\ z_{N,1} & \cdots & z_{N,7} \end{pmatrix} \quad (16)$$

其中,  $z_{i,1} \in \{z_{1,1}, \dots, z_{N,1}\}$  表示当前时刻空气湿度,  $z_{i,2} \in \{z_{1,2}, \dots, z_{N,2}\}$  表示当前时刻温度,  $z_{i,3} \in \{z_{1,3}, \dots, z_{N,3}\}$  表示当前时刻风向,  $z_{i,4} \in \{z_{1,4}, \dots, z_{N,4}\}$  表示当前时刻气压,  $z_{i,5} \in \{z_{1,5}, \dots, z_{N,5}\}$  表示当前时刻风速,  $z_{i,6} \in \{z_{1,6}, \dots, z_{N,6}\}$  表示当前时刻风功率,  $z_{i,7} \in \{z_{1,7}, \dots, z_{N,7}\}$  表示下一时刻风功率, 且  $i \in N$ ,  $N$  为训练数据个数.

对数据进行归一化处理,  $x_{i,k}$  为  $z_{i,k}$  归一化后的结果

$$x_{i,k} = \frac{z_{i,k} - \bar{z}_k}{\sqrt{\frac{\sum_{i=1}^N (z_{i,k} - \bar{z}_k)^2}{N}}} \quad (17)$$

其中,  $i = 1, 2, \dots, N$ ,  $k = 1, 2, \dots, 7$ ,  $\bar{z}_k$  表示矩阵中第  $k$  列数据均值, 得到归一化后的数据矩阵

$$\begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{pmatrix} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,7} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,7} \end{pmatrix} \quad (18)$$

### 2.4 评价指标

为验证算法的有效性, 本文分别以均方根误差 (Root mean square error, RMSE)、最大误差 (Maximum error, MAXE)、平均相对误差 (Average relative error, ARE)、以及风场风功率预测准

确度 (Prediction accuracy rate, PAR) 作为评价指标对所提算法进行评价, 具体公式为

$$\text{RMSE} = \sqrt{\frac{1}{v} \sum_{t=1}^v e_t^2} \quad (19)$$

$$\text{MAXE} = \max(e_t), \quad t = 1, \dots, v \quad (20)$$

$$\text{ARE} = \frac{1}{v} \sum_{t=1}^v \left( \frac{e_t}{Y_t} \right) \quad (21)$$

$$\text{PAR} = \left( 1 - \sqrt{\frac{1}{v} \sum_{t=1}^v \left( \frac{Y_t - Y'_t}{Cap} \right)^2} \right) \times 100\% \quad (22)$$

其中,  $v$  为 288 (相当于 3 天内每 15 分钟采样的总测试数据),  $e_t = |Y_t - Y'_t|$ ,  $Y_t$  为  $t$  时段实际风功率,  $Y'_t$  为  $t$  时段预测风功率,  $Cap$  为风电场运行装机容量, 这里  $Cap$  为 99 MW.

### 2.5 参数优化性能分析

本文优化参数设置范围如表 1 和表 2 所示. 其中表 1 是 IDE 方法的参数设置, 表 2 显示了 BSIDE-HKLSSVM 方法的参数设置范围.

表 1 IDE 的参数设置  
Table 1 Parameters setting of IDE

参数	$NP$	$M$	$g_{\max}$	$\theta$	$CR$	$F_1$
取值	100	4	20	10	0.7	0.3

表 2 BSIDE-HKLSSVM 的参数设置范围  
Table 2 Parameters setting scope of BSIDE-HKLSSVM

参数	$\lambda$	$\sigma$	$\gamma$	$d$	$a_1$
范围	[0.05, 0.1]	[1, 50]	(0, 1 000]	[1, 10]	(0, 0.0001]

图 3 为相同迭代次数下 IDE 和 DE<sup>[26]</sup> 独立运行 20 次后的平均风速预测模型收敛图, 其中 IDE 的最终收敛值为 0.013768, DE 的最终收敛值为 0.013833.

从图 3 可以看出, 由于初始解的影响, IDE 在前期的优化过程中并没有 DE 效果好, 但是在第 6 代之后, IDE 找到了更小的自适应函数值. 实验结果表明: 通过后期的迭代, IDE 能找到更合适的风速预测模型参数.

图 4 为相同迭代次数下 IDE 和 DE 独立运行 20 次后的平均风功率预测模型收敛图, 其中 IDE

的最终收敛值为 0.016346, DE 的最终收敛值为 0.016663.

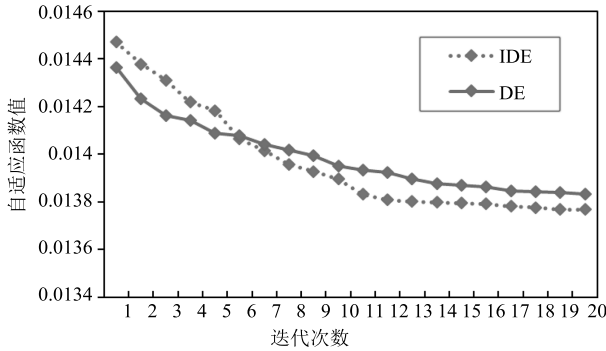


图 3 IDE 与 DE 的风速预测模型收敛比较

Fig. 3 The convergence comparison of wind speed prediction model between IDE and DE

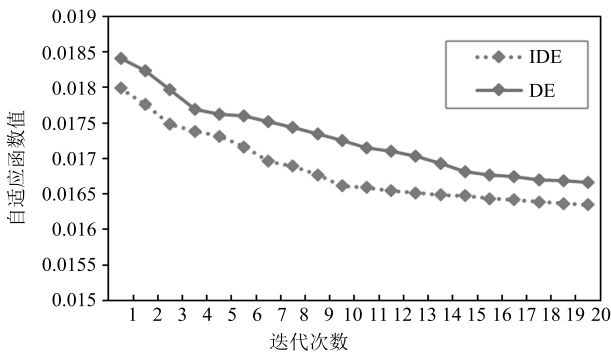


图 4 IDE 与 DE 的风功率预测模型收敛比较

Fig. 4 The convergence comparison of wind power prediction model between IDE and DE

从图 4 可以看出, 与 DE 进行比较, 在 20 次迭代的过程中, IDE 都能够找到更小的自适应函数值. 基于上述的实验结果可以得出结论, 在保证一定代数的前提下, IDE 能够找到合适的风功率预测模型参数, 从而提高模型的自适应性.

## 2.6 风速预测结果

图 5 为风速实际值与 SIDE-HKLSSVM 的风速预测值的比较结果, 其中采样时间间隔为 15 分钟. 从图 5 可以看出, SIDE-HKLSSVM 能够取得较小预测误差, 并且预测结果接近实际的风速曲线. 此外, 针对于风速预测实验, 本文还与其他改进策略以及 SVR<sup>[28]</sup>、ELM<sup>[29]</sup> 进行了比较. 从表 3 可以看出, 在三种性能指标比较下, SIDE-HKLSSVM 的均方根误差为 0.61, 最大误差为 2.07 m/s, 平均相对误差为 0.07, 均为最小的预测误差结果, 说明了 SIDE-HKLSSVM 方法能够达到实际风速预测效果. 在混核方法的比较中, 与 IDE-LSSVM 相

比, IDE-HKLSSVM 的预测误差更小, 说明了混核方法具有较好的预测性能. 为了说明稀疏策略的有效性, 本文对训练模型的计算时间进行了统计, SIDE-HKLSSVM 训练模型时需要 28.6 s, IDE-HKLSSVM 训练耗时为 64.5 s. 说明了提出的稀疏策略能够缩短模型的训练时间, 并且没有影响预测的精度, 提高了训练模型的效率.

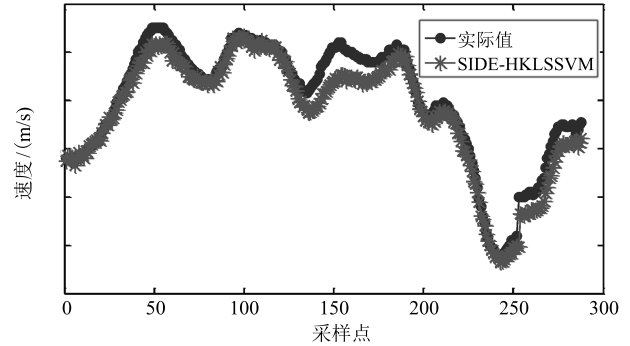


图 5 风场风速预测曲线

Fig. 5 The prediction curve of wind speed in the wind farm

表 3 风场风速预测结果

Table 3 Prediction results of wind speed in the wind farm

方法	RMSE	MAXE	ARE
		(m/s)	
SIDE-HKLSSVM	<b>0.61</b>	<b>2.07</b>	<b>0.07</b>
IDE-HKLSSVM	0.83	2.16	0.10
IDE-LSSVM	1.14	3.17	0.12
SVR	2.14	3.76	0.25
ELM	1.21	2.35	0.13

## 2.7 风功率预测结果

针对于风功率预测实验, 本文主要结合了批划分的思想进行设计, 表 4 和表 5 分别给出了批划分策略对批样机风功率的预测结果. 从中可以看出, 对不同的批样机进行风功率预测, BSIDE-HKLSSVM 方法与 BIDE-HKLSSVM 方法相比, 稀疏策略能够在训练样本减小的情况下, 保持较好的预测精度.

图 6 为风功率实际值与 BSIDE-HKLSSVM 的风功率预测值的比较结果, 其中采样时间间隔为 15 分钟. 从图 6 可以看出, BSIDE-HKLSSVM 的预测方法能够取得较小的预测误差, 得到的预测结果与实际风场风功率相近. 表 6 为 BSIDE-HKLSSVM 与其他对比方法的风场风功率预测比较结果. 为了进一步验证算法的有效性, 在本实验中还添加了 STS 风场风功率预测方法, 以及专家经验 (Expert

experience, EE) 的风功率预测方法进行比较. 从表 6 可以看出, BSIDE-HKLSSVM 的各项性能指标均优于其他方法, 并且风场风功率的预测准确度能够达到 95.4%, 均高于其他方法的预测准确度.

表 4 基于 BSIDE-HKLSSVM 方法的批样机风功率预测结果

Table 4 Prediction results of wind power based on BSIDE-HKLSSVM method for batch turbines

批样机	RMSE	MAXE (MW)	ARE
9#	0.18	0.63	0.35
18#	0.15	0.61	0.42
31#	0.12	0.46	0.51
37#	0.14	0.50	0.70
52#	0.13	0.53	0.90
59#	0.12	0.63	0.44

表 5 基于 BIDE-HKLSSVM 方法的批样机风功率预测结果

Table 5 Prediction results of wind power based on BIDE-HKLSSVM method for batch turbines

批样机	RMSE	MAXE (MW)	ARE
9#	0.16	0.60	0.32
18#	0.14	0.62	0.36
31#	0.18	0.44	0.92
37#	0.12	0.51	0.50
52#	0.13	0.51	0.94
59#	0.12	0.64	0.41

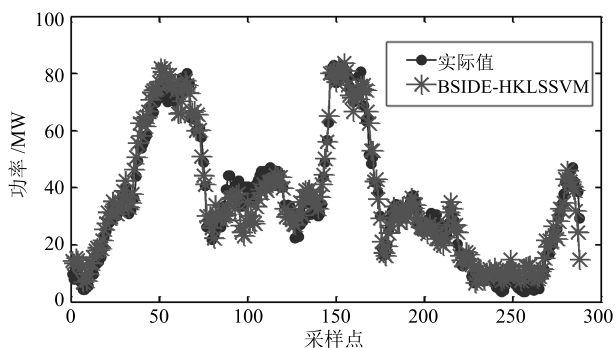


图 6 风场风功率预测曲线

Fig. 6 The prediction curve of wind power in the wind farm

从表 6 中还可以得出结论, 人工经验的风功率预测误差较大. 通过单核和混核的实验结果可以看

出, 混核的方法优于单核的方法. 通过最终的统计分析可以看出, 带有批划分策略的方法的预测精度均在 95% 以上, 能够达到更好的预测效果. 另外, 在计算时间上, 如表 7 所示, 结合稀疏策略的预测模型, 与对应的无稀疏策略的方法相比, 均能够获得较短的训练时间. 综合以上实验统计结果, 我们可以看出 BSIDE-HKLSSVM 能够高效地对风场风功率进行预测, 并且能够满足实际风场风功率预测的需求.

表 6 风场风功率预测结果

Table 6 Prediction results of wind power in the wind farm

方法	RMSE	MAXE (MW)	ARE	PAR (%)
BSIDE-HKLSSVM	<b>4.57</b>	<b>17.36</b>	<b>0.21</b>	<b>95.4</b>
BIDE-HKLSSVM	4.84	18.86	0.22	95.1
SIDE-HKLSSVM	6.79	21.94	0.22	93.1
IDE-HKLSSVM	7.44	25.98	<b>0.21</b>	92.5
IDE-LSSVM	8.35	26.95	0.26	91.6
STS	8.09	22.01	0.33	91.8
SVR	7.53	27.34	0.22	92.4
ELM	7.93	27.17	0.33	92.0
EE	13.08	36.05	0.33	86.8

表 7 带有稀疏策略的风场风功率训练模型时间比较

Table 7 Time comparisons of training model with sparsity strategy for the wind power in the wind farm

方法	时间 (s)
BSIDE-HKLSSVM	<b>382.27</b>
BIDE-HKLSSVM	474.07
SIDE-HKLSSVM	<b>40.71</b>
IDE-HKLSSVM	64.93

## 2.8 实验分析

针对风场数据的特征, 本文对提出的预测模型进行了分析, 图 7 和图 8 分别显示了 BSIDE-HKLSSVM 方法对风场中某台风机的风速和风功率的预测误差分布情况, 从图中我们可以看出风速和风功率的预测误差均服从于均值在 0 左右的正态分布. 为定量说明正态分布性, 本文对预测误差进行了 K-S 检验. 它是一种基于经验累积分布函数的检验方法, 具有分布无关性, 适用于任何连续分布. 通过 K-S 检验我们可以得出结论: BSIDE-HKLSSVM 方法的风速预测误差渐进显著性指标值为 0.295, 风功率预测误差渐进显著性指标值为 0.856, 这些渐进



显著性指标值均大于 0.05, 说明本文提出的预测模型适用于该实验数据。

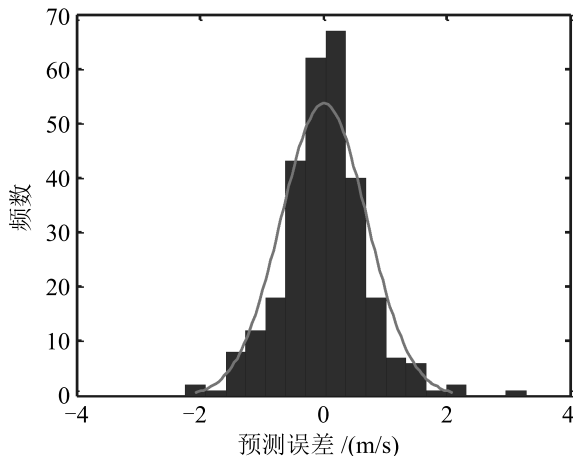


图 7 风速预测误差分布

Fig. 7 The distribution of wind speed prediction errors

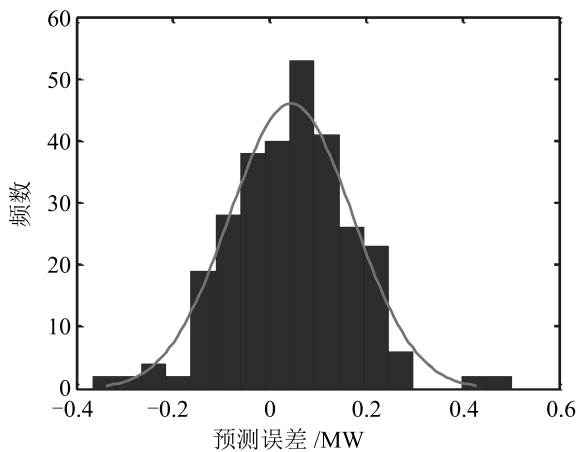


图 8 风功率预测误差分布

Fig. 8 The distribution of wind power prediction errors

### 3 结束语

本文提出带有批特征的混核最小二乘支持向量机方法对风场风功率进行预测. 在建立模型的过程中, 为提高模型的适应性, 设计改进的差分进化算法对模型参数进行优化, 并利用稀疏选择方法对模型的训练数据集进行筛选, 以缩短预测模型的训练时间. 为了使得风场风功率预测更加科学, 提出了批划分策略, 对风场的风机进行合理的划分. 从实验结果可以看出, 与其他预测方法相比, 本文提出的改进方法能够对风场风功率进行准确预测. 并且具有较高的预测精度, 可为后续的电力调度提供参考.

### References

1 Marí L, Nabona N. Renewable energies in medium-term power planning. *IEEE Transactions on Power Systems*,

2015, **30**(1): 88–97

2 Lang Jin, Tang Li-Xin. Unit commitment problem for wind turbines power generation with batching characteristics consideration. *Acta Automatica Sinica*, 2015, **41**(7): 1295–1305

(郎劲, 唐立新. 考虑风力发电批特征的电力机组调度问题. *自动化学报*, 2015, **41**(7): 1295–1305)

3 Moeini-Aghtaie M, Farzin H, Fotuhi-Firuzabad M, Amrollahi R. Generalized analytical approach to assess reliability of renewable-based energy hubs. *IEEE Transactions on Power Systems*, 2017, **32**(1): 368–377

4 Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. *Renewable and Sustainable Energy Reviews*, 2014, **34**: 243–254

5 Croonenbroeck C, Ambach D. A selection of time series models for short- to medium-term wind power forecasting. *Journal of Wind Engineering and Industrial Aerodynamics*, 2015, **136**: 201–210

6 Colak I, Sagiroglu S, Yesilbudak M. Data mining and wind power prediction: A literature review. *Renewable Energy*, 2012, **46**: 241–247

7 Yuan X H, Tan Q X, Lei X H, Yuan Y B, Wu X T. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. *Energy*, 2017, **129**: 122–137

8 Wan C, Xu Z, Pinson P, Dong Z Y, Wong K P. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, 2014, **29**(3): 1033–1044

9 Qureshi A S, Khan A, Zameer A, Usman A. Wind power prediction using deep neural network based meta regression and transfer learning. *Applied Soft Computing*, 2017, **58**: 742–755

10 Pousinho H M I, Mendes V M F, Catalão J P S. A hybrid PSO-ANFIS approach for short-term wind power prediction in Portugal. *Energy Conversion and Management*, 2011, **52**: 397–402

11 Haque A U, Mandal P, Meng J L, Srivastava A K, Tseng T L, Senjyu T. A novel hybrid approach based on wavelet transform and fuzzy ARTMAP networks for predicting wind farm power production. *IEEE Transactions on Industry Applications*, 2013, **49**(5): 2253–2261

12 Heinermann J, Kramer O. Machine learning ensembles for wind power prediction. *Renewable Energy*, 2016, **89**: 671–679

13 Ye L, Zhao Y N, Zeng C, Zhang C H. Short-term wind power prediction based on spatial model. *Renewable Energy*, 2017, **101**: 1067–1074

14 Liu Qiang, Qin S Joe. Perspectives on big data modeling of process industries. *Acta Automatica Sinica*, 2016, **42**(2): 161–171

(刘强, 秦泗钊. 过程工业大数据建模研究展望. *自动化学报*, 2016, **42**(2): 161–171)

- 15 Tao Jian-Wen, Wang Shi-Tong. Kernel support vector machine for domain adaptation. *Acta Automatica Sinica*, 2012, **38**(5): 797–811  
(陶剑文, 王士同. 领域适应核支持向量机. 自动化学报, 2012, **38**(5): 797–811)
- 16 Shi Yong, Li Pei-Jia, Wang Hua-Dong. L2-Loss large-scale linear nonparallel support vector ordinal regression. *Acta Automatica Sinica*, 2019, **45**(3): 505–517  
(石勇, 李佩佳, 汪华东. L2 损失大规模线性非平行支持向量顺序回归模型. 自动化学报, 2019, **45**(3): 505–517)
- 17 Gao Ming-Zhe, Xu Ai-Qiang, Tang Xiao-Feng, Zhang Wei. Analog circuit diagnostics method based on multi-kernel learning multiclass relevance vector machine. *Acta Automatica Sinica*, 2019, **45**(2): 434–444  
(高明哲, 许爱强, 唐小峰, 张伟. 基于多核多分类相关向量机的模拟电路故障诊断方法. 自动化学报, 2019, **45**(2): 434–444)
- 18 Zhang Kai-Jun, Liang Xun. An improved domain multiple kernel support vector machine. *Acta Automatica Sinica*, 2014, **40**(10): 2288–2294  
(张凯军, 梁循. 一种改进的显性多核支持向量机. 自动化学报, 2014, **40**(10): 2288–2294)
- 19 Ouyang T H, Zha X M, Qin L, Xiong Y, Xia T. Wind power prediction method based on regime of switching kernel functions. *Journal of Wind Engineering and Industrial Aerodynamics*, 2016, **153**: 26–33
- 20 Ouyang T H, Zha X M, Qin L. A combined multivariate model for wind power prediction. *Energy Conversion and Management*, 2017, **144**: 361–373
- 21 Khosravi A, Koury R N N, Machado L, Pabon J J G. Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sustainable Energy Technologies and Assessments*, 2018, **25**: 146–160
- 22 Yuan X H, Chen C, Yuan Y B, Huang Y H, Tan Q X. Short-term wind power prediction based on LSSVM-GSA model. *Energy Conversion and Management*, 2015, **101**: 393–401
- 23 Zhang L L, Li M S, Ji T Y, Wu Q H. Short-term wind power prediction based on intrinsic time-scale decomposition and LS-SVM. In: Proceedings of the 2016 IEEE Innovative Smart Grid Technologies — Asia (ISGT-Asia). Melbourne, Australia: IEEE, 2016. 41–45
- 24 Vapnik V N. *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 1995.
- 25 Liu C, Tang L X, Liu J Y, Tang Z H. A dynamic analytics method based on multistage modeling for a BOF steelmaking process. *IEEE Transactions on Automation Science and Engineering*, 2019, **16**(3): 1097–1109
- 26 Storn R, Price K. Differential Evolution — A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces, Technical Report TR-95-012, International Computer Science Institute, Berkeley, USA, 1995.
- 27 Suykens J A K, van Gestel T, de Brabanter J, de Moor B, Vandewalle J. *Least Squares Support Vector Machines*. Singapore: World Scientific, 2003.
- 28 Drucker H, Burges C J, Kaufman L, Smola A J, Vapnik V. Support vector regression machines. In: Proceedings of the 1997 Advances in Neural Information Processing Systems. Denver, CO, USA: NIPS, 1997. 155–161
- 29 Huang G B, Zhou H M, Ding X J, Zhang R. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, **42**(2): 513–529



**刘畅** 智能工业数据解析与优化教育部重点实验室(东北大学)、东北大学工业与系统工程研究所博士研究生. 主要研究方向为机器学习和智能优化. 本文通信作者. E-mail: lc1987328@126.com  
(**LIU Chang** Ph. D. candidate at the Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education, and the Institute of Industrial and Systems Engineering, Northeastern University. His research interest covers machine learning and intelligent optimization. Corresponding author of this paper.)



**郎劲** 东北大学辽宁省制造系统与物流优化重点实验室、东北大学流程工业综合自动化国家重点实验室讲师. 主要研究方向为可再生能源, 风力发电和电力系统. E-mail: langjin@ise.neu.edu.cn  
(**LANG Jin** Lecturer at Liaoning Key Laboratory of Manufacturing System and Logistics, and the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University. Her research interest covers renewable energy, wind power, and power systems.)