

基于优化核极限学习机的风电功率时间序列预测

李军 李大超

Wind power time series prediction using optimized kernel extreme learning machine method

Li Jun Li Da-Chao

引用信息 Citation: *Acta Physica Sinica*, 65, 130501 (2016) DOI: 10.7498/aps.65.130501

在线阅读 View online: <http://dx.doi.org/10.7498/aps.65.130501>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2016/V65/I13>

您可能感兴趣的其他文章

Articles you may be interested in

自调整平滑区间粒子滤波平滑算法

Auto-adjust lag particle filter smoothing algorithm for non-linear state estimation

物理学报.2016, 65(4): 040502 <http://dx.doi.org/10.7498/aps.65.040502>

小管径气液两相流空隙率波传播的多尺度相关性

Multi-scale cross-correlation characteristics of void fraction wave propagation for gas-liquid two-phase flows in small diameter pipe

物理学报.2016, 65(1): 010501 <http://dx.doi.org/10.7498/aps.65.010501>

基于参数切换算法的混沌系统吸引子近似及其电路设计

Approximations of chaotic attractors and its circuit design based on the parameter switching algorithm

物理学报.2015, 64(20): 200508 <http://dx.doi.org/10.7498/aps.64.200508>

语音信号序列的 Volterra 预测模型

Volterra prediction model for speech signal series

物理学报.2015, 64(20): 200507 <http://dx.doi.org/10.7498/aps.64.200507>

基于改进教学优化算法的 Hermite 正交基神经网络混沌时间序列预测

Hermite orthogonal basis neural network based on improved teaching-learning-based optimization algorithm for chaotic time series prediction

物理学报.2015, 64(20): 200506 <http://dx.doi.org/10.7498/aps.64.200506>

基于优化核极限学习机的风电功率时间序列预测*

李军† 李大超

(兰州交通大学自动化与电气工程学院, 兰州 730070)

(2016年1月18日收到; 2016年4月15日收到修改稿)

针对时间序列预测, 在单隐层前馈神经网络的基础上, 基于进化计算的优化策略, 提出了一种优化的核极限学习机 (optimized kernel extreme learning machine, O-KELM) 方法. 与极限学习机 (extreme learning machine, ELM) 方法相比, 核极限学习机 (kernel extreme learning machine, KELM) 方法无须设定网络隐含层节点的数目, 以核函数表示未知的隐含层非线性特征映射, 通过正则化最小二乘算法计算网络的输出权值, 它能以极快的学习速度获得良好的推广性. 在 KELM 的基础上, 分别将遗传算法、模拟退火、微分演化三种进化算法用于模型的结构输入选择、正则化系数以及核参数的优化选取, 以进一步提高网络的性能. 将 O-KELM 方法应用于标准 Mackey-Glass 混沌时间序列预测及某地区的风电功率时间序列预测实例中, 在同等条件下, 还与优化的极限学习机 (optimized extreme learning machine, O-ELM) 方法进行比较. 实验结果表明, 所提出的 O-KELM 方法在预测精度上优于 O-ELM 方法, 表明了其有效性.

关键词: 核极限学习机, 优化方法, 时间序列, 预测

PACS: 05.45.Tp, 88.50.Mp

DOI: 10.7498/aps.65.130501

1 引言

随着风电场的风电装机容量的不断攀升, 风电在电力系统中高度渗透. 然而由于风能的随机性和间歇性, 风电场的发电输出功率往往很难控制, 会随着外界能量的变化而发生涨落、波动, 对电网造成冲击, 给电网计划和调度带来困难与挑战. 大规模风电的并网也给电网的稳定运行以及电能的可靠性等方面带来了挑战. 精确、可靠的风电功率预测^[1,2]对于优化电网运行的成本和提高电力系统的可靠性极其重要.

基于时间尺度以及所需做出不同运行决策的考虑, 风电功率的预测可以分为超短期、短期、中期和长期预测等^[3,4]. 风电功率的预测方法通常可分为物理方法、统计方法、空间相关性预测方法、组合预测方法等^[5,6]. 统计方法旨在通过对过去时间序列的统计分析, 找出输入数据与风电输出功率

率之间的联系, 在超短期、短期预测中应用非常广泛. 传统的统计方法是诸如 persistence 模型^[7]的时间序列方法, 它假定风电功率的预测值是最最近的测量值. 另一类统计方法是自回归积分滑动平均 (ARIMA) 模型^[8], Kalman 滤波等^[9], 但是这些基于线性回归模型的预测方法不能表示风电功率时间序列的非线性特性. 更先进的一类统计方法, 如神经网络^[10]、支持向量机 (SVM)^[11,12]等, 能从过去的时间序列中描述出输入与输出的非线性联系, 与 ARIMA 和 persistence 方法相比, 在风电功率时间序列的超短期与短期预测中能给出更好的预测结果.

极限学习机 (extreme learning machine, ELM) 是 Huang 等^[13]提出的一种基于单隐层前馈神经网络 (single-hidden layer feedforward neural networks, SLFNs) 的快速学习方法, 其特点是随机选择 SLFNs 的隐含层节点及相应的节点参数, 在训练过程中仅需通过正则化最小二乘算法调节网络

* 国家自然科学基金 (批准号: 51467008) 资助的课题.

† 通信作者. E-mail: lijun691201@mail.lzjtu.cn

的输出权值. 因此, 它能以极快的学习速度获得良好的网络泛化性能. 文献 [14] 提出一种基于多目标优化的改进极限学习机, 将训练误差和输出层权值的均方最小化同时作为优化的目标函数, 采用带精英策略的遗传算法对极限学习机的输入层到隐层的权值和阈值进行优化, 将该算法应用于摇摆工况下自然循环系统不规则复合型流量脉动的多步滚动预测, 取得了很好的应用效果. 结合进化计算的思想, 文献 [15] 给出一种优化的极限学习机 (O-ELM) 方法, 通过优化选择输入变量、隐含层节点的配置和偏置参数、正则化系数等, 进一步提高了网络的学习性能, 在 Boston 房价预测等实例应用中取得了很好的效果. 文献 [16] 给出了一种基于 ELM 的风电功率短期概率预测方法, 也取得了成功的应用.

在 ELM 的基础上, 当隐含层特征映射未知时, 若以核函数取代, 可形成一种核极限学习机 (KELM) 方法 [17], 其优点是无须预先确定隐含层节点的数目. 与 ELM 相比, KELM 方法在网络的训练学习过程中, 仅需选择适当的核参数与正则化系数, 通过矩阵运算, 获得网络的输出权值. 文献 [18] 提出一种基于多核极限学习机的方法, 应用于多元混沌时间序列预测中, 可取得较好的预测效果. 文献 [19] 提出了一种基于快速留一交叉验证的在线 KELM 方法, 在混沌时间序列预测及连续搅拌釜式反应器的过程辨识中取得了成功的应用. 鉴于 O-ELM 方法在 ELM 网络的参数选择方面的突出优点, 在文献 [15] 基础上, 本文提出一种优化的核极限学习机 (O-KELM) 方法, 基于遗传算法 (GA)、微分演化 (DE) 和模拟退火 (SA) 三种进化算法, 在网络训练过程中优化选择 KELM 的结构输入变量、核参数与正则化系数, 以获得更好的网络性能. 将所提出的 O-KELM 方法应用于 Mackey-Glass 混沌时间序列预测及某地区的超短期风电功率时间序列预测实例中, 在同等条件下与 O-ELM 方法相比较, 从预测精度方面比较预测模型的性能, 以进一步验证本文方法的有效性.

2 KELM 方法

KELM 是 ELM 的非线性延伸, 本节在 ELM 的基础上, 给出 KELM 的学习方法.

对于含有 L 个隐层节点的 SLFNs, 给定 N 组

训练样本数据集 $(\mathbf{x}_j, \mathbf{t}_j) \in \mathbb{R}^n \times \mathbb{R}^m$, 即 $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T$, $\mathbf{t}_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T$, 则其网络节点的输出为

$$y(\mathbf{x}_j) = \sum_{i=1}^L \beta_i G(\mathbf{x}_j; \mathbf{w}_i, b_i) = \mathbf{o}_j, \quad j = 1, \dots, N. \quad (1)$$

若 SLFNs 能以零误差逼近训练样本, 即 $\sum_{j=1}^N \|\mathbf{o}_j - \mathbf{t}_j\| = 0$, 则存在 $\beta_i, \mathbf{w}_i, b_i$, 使

$$\sum_{i=1}^L \beta_i G(\mathbf{x}_j; \mathbf{w}_i, b_i) = \mathbf{t}_j, \quad j = 1, \dots, N, \quad (2)$$

其中 β_i 为第 i 个隐含层节点和输出层节点之间的权值向量. 改写为矩阵形式

$$\mathbf{H}\mathbf{B} = \mathbf{T}, \quad (3)$$

且有

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} G(\mathbf{x}_1; \mathbf{w}_1, b_1) & \cdots & G(\mathbf{x}_1; \mathbf{w}_L, b_L) \\ \vdots & \cdots & \vdots \\ G(\mathbf{x}_N; \mathbf{w}_1, b_1) & \cdots & G(\mathbf{x}_N; \mathbf{w}_L, b_L) \end{bmatrix}_{N \times L},$$

$$\mathbf{B} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m},$$

其中, \mathbf{B} 为输出权值矩阵; \mathbf{H} 为隐含层节点的输出矩阵, 其第 j 行是与输入 \mathbf{x}_j 相关的隐含层特征映射 $\mathbf{h}(\mathbf{x}_j) = [G(\mathbf{x}_j; \mathbf{w}_1, b_1), \dots, G(\mathbf{x}_j; \mathbf{w}_L, b_L)]$, 即 $\mathbf{x}_j : \mathbf{h}(\mathbf{x}_j)$.

如果激活函数 $G(\mathbf{x}_j; \mathbf{w}_i, b_i)$ 在任意区间上无限可微, 且 SLFNs 隐含层节点及节点参数可以随机产生, 则从插值的思想看, 隐含层节点的最大数目 L 应小于训练样本数目 N . 当 L 小于 N 时, 由文献 [13] 可知, SLFNs 仍可以极小的训练误差逼近训练样本. 与使用梯度下降算法训练所有网络权值的 SLFNs 不同, ELM 的训练相当于求解线性方程组 $\mathbf{H}\mathbf{B} = \mathbf{T}$ 的最小二乘解 \mathbf{B} , 即

$$\|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_L, b_1, \dots, b_L)\hat{\mathbf{B}} - \mathbf{T}\| = \min_{\beta_i} \|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_L, b_1, \dots, b_L)\mathbf{B} - \mathbf{T}\|. \quad (4)$$

(4) 式中, 为提高网络的稳定性与泛化能力, 按照 Tikhonov 正则化及岭回归的思想, 给定正则化系数 η , 则输出权值的最小二乘解为

$$\hat{\mathbf{B}} = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \eta\mathbf{I})^{-1}\mathbf{T}. \quad (5)$$

因此, ELM 的相应输出函数为 $\mathbf{y}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\hat{\mathbf{B}}$.

考虑特征映射函数 $\mathbf{h}(\mathbf{x})$ 未知的情形, 此时需要将核函数引入到 ELM 中. 定义核矩阵 $\mathbf{\Omega}_{\text{ELM}} = \mathbf{H}\mathbf{H}^T$, 其元素为 $\mathbf{\Omega}_{\text{ELM}}(i, j) = \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$, 那么借助 (5) 式, 网络输出可表示为

$$\mathbf{y}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\hat{\mathbf{B}} = \mathbf{h}(\mathbf{x})\mathbf{H}^T(\eta\mathbf{I} + \mathbf{H}\mathbf{H}^T)^{-1}\mathbf{T} \\ = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T (\eta\mathbf{I} + \mathbf{\Omega}_{\text{ELM}})^{-1}\mathbf{T}, \quad (6)$$

(6) 式中, 核函数 $K(\mathbf{x}_i, \mathbf{x}_j)$ 的类型通常可选择径向基函数核, 即

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\gamma^2}\right).$$

3 O-KELM 方法及其学习算法

在 KELM 方法的学习算法实现中, 核参数 γ 与正则化系数 η 的选择如果不恰当, 会影响模型的性能, 通常在实验中可通过交叉验证的方法进行确定. 另外, 如果不恰当地选择网络的输入, 即 KELM 网络结构的选择不当也会影响模型的性能. 将进化计算算法与 KELM 相结合, 对模型的输入变量、核参数 γ 与正则化系数 η 进行优化选择, 则可在一定程度上有效解决上述问题. 因此, 为进一步提升网络的学习性能, 本节提出一种 O-KELM 方法.

在 O-KELM 方法实现过程中, 将最大化如下的适应度评价函数, 即

$$J_{\text{fitness}} = \frac{1}{1 + E_{\text{RMSE}}(y, y_d)}, \\ E_{\text{RMSE}}(y, y_d) = \sqrt{\frac{1}{N} \sum_{j=1}^N [y(j) - y_d(j)]^2}, \quad (7)$$

其中, $E_{\text{RMSE}}(y, y_d)$ 是均方根误差, $y(j)$ 是模型的预测输出, $y_d(j)$ 是期望输出.

待优化的决策变量是一个实数变量和二进制变量的组合, 优化的决策变量, 即参数是网络的核参数 γ , 正则化系数 η , 输入变量集合等, 它们

共同构成一个种群个体, 即 $\mathbf{r}_k = [s_1, \dots, s_n, \gamma, \eta]$, $k = 1, \dots, m$ 其中, 模型的输入维数为 n , 二进制变量 $s_a \in \{0, 1\}$, $a = 1, \dots, n$ 决定模型的第 j 维输入是应保留还是舍弃, m 表示种群个体的数目.

种群中的个体的优化, 即 $\mathbf{r}_k = [s_1, \dots, s_n, \gamma, \eta]$, $k = 1, \dots, m$ 的优化涉及到实数及二进制变量, 为应用优化算法, 首先将决策变量转化为位于 $[0, 1]$ 区间值之间的连续变量, 且为实数; 其次, 需要在计算每个个体的评价函数之前, 将个体中的待优化的所有决策变量转化为它们的真实值, 若个体 k 的第 l 个决策变量的真实值为实数时, 变换为

$$z_{k_l} = (z_l^{\max} - z_l^{\min})r_{k_l} + z_l^{\min}, \quad (8)$$

其中, z_{k_l} 为实数型变量, $z_l^{\min} \leq z_{k_l} \leq z_l^{\max}$ 表示实数型变量的上、下界. 若个体 k 的第 l 个变量的真实值为二进制数时, 变换为整数型变量, 即

$$z_{k_l} = \text{round}(r_{k_l}), \quad (9)$$

其中, 函数 $\text{round}(\cdot)$ 是指四舍五入到最近的整数.

具体应用时, 核参数 γ 为正实数, 需通过 (8) 式进行转换, 正则化系数 η 同样需通过 (8) 式进行转换, 二进制变量 s_a , $a = 1, \dots, n$, 则需通过 (9) 式进行转换.

基于进化计算的 GA, DE 及 SA 三种优化算法, 可对 KELM 中的以上决策变量进行优化, 形成 O-KELM 方法的三种不同算法: GA-KELM, DE-KELM, SA-KELM.

3.1 GA-KELM 算法

GA 是一种基于群体搜索和优化的算法, 它模拟自然界的生物进化过程, 对于解决工程中的优化问题简单而有效^[20,21]. 具体的 GA-KELM 算法实现描述如下.

步骤 1 种群初始化操作. 初始种群 \mathbf{R} 通过均匀分布随机产生如下:

$$\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_m], \quad (10)$$

$$\mathbf{r}_k = [r_{k_1}, r_{k_2}, \dots, r_{k_q}]^T, \quad (11)$$

r_{k_l} 是个体 \mathbf{r}_k 的第 l 个决策变量, 且 $0 \leq r_{k_l} \leq 1$, $k = 1, 2, \dots, m$, m 为种群的数量大小, $l = 1, 2, \dots, q$, q 表示待优化的决策变量数目.

步骤 2 由 (7) 式的适应度函数, 计算每一代种群中的个体的适应度.

步骤3 根据个体适应度值的大小, 经过联赛选择 (tournament selection) 算法^[21]操作, 从匹配池中选择双亲.

具体实现时, 从匹配池中随机挑选5个个体, 具有最好适应度值的个体被选择作为双亲, 并以概率 p_c 进行均匀交叉操作, 以产生新一代种群.

步骤4 以概率 p_m 进行自适应均匀变异操作.

具体实现时, 在每一代的进化中, 对于每个新个体, 产生一个随机数 p , 若 $p < p_m$, 其中 p_m 为变异概率, 则该新个体需要进行变异操作. 该变异操作是一个两步操作算子, 其中第1步是均匀变异, 新个体中的任一元素被随机选择后, 用区间值 $[0, 1]$ 之间的均匀分布随机数替代, 第2步是对变异后的新个体添加随机调节因子. 若新个体为 $\mathbf{r}_k = [r_{k_1}, r_{k_2}, \dots, r_{k_q}]^T$, 其第2个元素 r_{k_2} 被选择替换, 则变异后的个体为 $\mathbf{r}_k^1 = [r_{k_1}, r'_{k_2}, \dots, r_{k_q}]^T$. 再将 \mathbf{r}_k^1 中的所有元素均乘以一个随机数 $\beta_{k_l} - 1 \leq \beta_{k_l} \leq 1$, 再乘以变异因子 ξ_k , $0 \leq \xi_k \leq 1$, 即 $\mathbf{r}_k^c = \xi_k [\beta_{k_1} r_{k_1}, \beta_{k_2} r'_{k_2}, \dots, \beta_{k_q} r_{k_q}]^T$. 变异后的最终个体应为 $\mathbf{r}_k^2 = \text{rem}(\mathbf{r}_k^1 + \mathbf{r}_k^c)$. 其中, 求余函数 $\text{rem}(\cdot)$ 表示个体中的每一个变量除以1之后的剩余数, 这样则可以保证变异后的个体中每个变量的取值位于区间值 $[0, 1]$ 之间.

步骤4判断是否达到最大进化代数, 若是, 则算法终止, 否则, 迭代次数加1, 转步骤2.

在GA-KELM算法中, 通过自适应均匀变异操作, 补偿了种群在某一位可能缺失的基因, 保持了种群的多样性, 阻止算法陷入局部极小, 尽可能地扩大了搜索空间的范围, 以保证GA可以搜索到空间中所有点.

3.2 DE-KELM算法

DE是一种新型具有优异性能的全局优化算法^[22]. 作为一种基于种群的元启发式方法, 新的候选个体依赖于任意两个群体成员, 通过添加不同权重的微分向量, 且传递给第三个成员而产生.

DE-KELM与GA-KELM的算法实现类似, 种群初始化操作如(10)和(11)式所示, 产生初始种群. 对于每一代种群中的个体 $\mathbf{r}_k = [r_{k_1}, r_{k_2}, \dots, r_{k_q}]^T$, 随机选取3个不同个体 \mathbf{r}_ρ , \mathbf{r}_τ 和 \mathbf{r}_ϑ , 计算两两之间的微分向量, 继而乘以一个尺度因子, 添加给第三个被选择的个体. 即新的候选个体为

$$\mathbf{r}'_k = \mathbf{r}_\rho + \lambda(\mathbf{r}_\tau - \mathbf{r}_\vartheta), \quad (12)$$

其中 λ 为尺度因子.

另外, 在每一代的进化中, 考虑多样性, 可创建个体 $\mathbf{u}_k = [u_{k_1}, u_{k_2}, \dots, u_{k_q}]^T$, 其每一个元素 u_{k_l} , $l = 1, 2, \dots, q$ 为

$$u_{k_l} = \begin{cases} r'_{k_l}, & \text{if } U_{[0,1]} \leq C \text{ or } l = \delta_k, \\ r_{k_l}, & \text{otherwise,} \end{cases} \quad (13)$$

其中, $U_{[0,1]}$ 是位于 $[0, 1]$ 之间的均匀分布随机数, C 是可接受的概率, $\delta_k \in \{1, \dots, m\}$ 是个体 k 的随机索引数. 若种群中的新个体 \mathbf{u}_k 的性能优于 \mathbf{r}_k , \mathbf{r}_k 将被替换之. 通过(7)式计算适应度, 判断是否达到最大进化代数, 若是, 则算法终止.

3.3 SA-KELM算法

SA算法是一种基于Monte Carlo迭代求解, 用于全局寻优的元启发式随机搜索算法. 在搜索策略上, 它将组合优化问题与统计物理学中的热平衡问题类比, 模拟固体退火过程的物理机制, 使得固体在给定温度下达到热平衡的过程. 在退火降温过程中, SA算法逐步形成最低能量的基态, 即最稳定的结构状态.

在SA-KELM算法具体实现中, 定义 k 为代数, $T_k \in R^+$ 为温度, 每次迭代时, 温度的下降将满足 $T_{k+1} = \mu T_k$, 退火因子 $\mu \in (0, 1)$.

另外, 随机选取种群中的候选个体 $\mathbf{r}_k = [r_{k_1}, r_{k_2}, \dots, r_{k_q}]^T$, 对其施加摄动产生一个近邻解 \mathbf{v}_k , 近邻解以一定的概率 C 被接受为新的个体, 即

$$C = \begin{cases} 1, & \text{if } \psi(\mathbf{v}_k) \leq \psi(\mathbf{r}_k), \\ \exp\left(\frac{\psi(\mathbf{v}_k) - \psi(\mathbf{r}_k)}{T_k}\right), & \text{otherwise,} \end{cases} \quad (14)$$

其中 $\psi = E_{\text{RMSE}}(y, y_d)$. 因此, $\psi(\mathbf{v}_k)$, $\psi(\mathbf{r}_k)$ 分别是由(7)式计算适应度函数值时, 个体 \mathbf{v}_k , \mathbf{r}_k 的均方根误差函数.

4 应用实例

本节为验证优化KELM方法的有效性, 将其分别应用于混沌时间序列预测和风电功率时间序列预测实例中. 具体优化时, 核参数初始化时其取

值范围为 $\gamma \in (0, 600]$, 正则化系数初始化取值的范围为 $\eta \in [0, 100]$. 评价指标分别利用均方根误差 E_{RMSE} 、平均绝对值误差 E_{MAE} 和正则化均方误差 E_{NMSE} , 即

$$E_{\text{MAE}} = \frac{1}{N} \sum_{j=1}^N |y(j) - y_d(j)|, \quad (15)$$

$$E_{\text{NMSE}} = \frac{1}{N\sigma^2} \sum_{j=1}^N (y(j) - y_d(j))^2, \quad (16)$$

其中, $y_d(j)$ 为待预测时间序列各点的实际输出, $y(j)$ 为相应的模型预测输出, N 为预测样本点数, σ^2 为待预测时间序列的方差, 若 $E_{\text{NMSE}} = 1$, 意味着预测输出是时间序列的均值.

4.1 Mackey-Glass 混沌时间序列的直接多步预测

自 Mackey 和 Glass 发现时滞系统中的混沌现象以来, 时滞混沌系统就常用来检验非线性系统模型的性能, 其中, Mackey-Glass 混沌时间序列 [23] 由下述方程产生.

$$\frac{dy(t)}{dt} = \frac{0.2y(t-\tau)}{1+y^{10}(t-\tau)} - 0.1y(t), \quad (17)$$

其中, τ 为时滞参数, $\tau > 17$ 呈现混沌特性, τ 越大, 混沌程度越高. 由 (17) 式所产生的 Mackey-Glass 混沌时间序列, 可以通过四阶龙格-库塔法结合初始条件得出数值解, 其初始条件 $y(0) = 1.2$, $\tau = 17$. 针对该时间序列, 利用本文方法建立提前 6 步预测的模型. 首先, 从时间序列 $y(t)$ 中, 按照如下格式, 可抽取 1000 对输入输出数据

$$[y(t-18), y(t-12), y(t-6), y(t); y(t+6)],$$

其中 $t = 118-1117$. 即预测模型的输入为 $[y(t-18), y(t-12), y(t-6), y(t)]$, 待预测的输出为 $y(t+6)$. 前 500 个数据用于模型的训练, 其余的 500 个数据用于测试.

在 GA-KELM, DE-KELM, SA-KELM 方法中, GA 与 DE 方法的初始种群数是 100, 最大的进化代数为 250. GA 中的交叉概率 p_c 为 0.4, 变异概率 p_m 为 0.1. DE 中的尺度因子 λ 选择一个位于区间 $[0.5, 2]$ 的随机数. SA 中的最大迭代次数为 500, 初始温度 T_0 设置为 100, $\mu = 0.95$. 对比的 GA-ELM, DE-ELM, SA-ELM 方法中, 隐含层节点数初始化的最大值 $L = 200$.

表 1 O-KELM 方法与 O-ELM 方法的性能指标对比
Table 1. Evaluation of performance indicator based on O-KELM method and O-ELM method.

预测方法	RMSE	MAE	NMSE	L	γ	η
GA-ELM	4.570×10^{-2}	1.090×10^{-2}	9.900×10^{-3}	112	—	1.353×10^{-1}
DE-ELM	1.330×10^{-2}	0.094×10^{-2}	7.888×10^{-4}	140	—	5.777×10^{-3}
SA-ELM	3.120×10^{-2}	1.070×10^{-2}	4.400×10^{-3}	125	—	1.477×10^{-1}
GA-KELM	3.900×10^{-3}	3.000×10^{-3}	6.047×10^{-5}	—	6.160×10^{-2}	1.001×10^{-2}
DE-KELM	3.600×10^{-3}	2.800×10^{-3}	5.584×10^{-5}	—	6.131×10^{-2}	1.011×10^{-2}
SA-KELM	4.600×10^{-3}	3.400×10^{-3}	9.280×10^{-5}	—	8.370×10^{-2}	1.058×10^{-2}

经过训练后, 表 1 给出在测试集上, 不同 O-KELM 方法与不同 O-ELM 方法的预测性能指标对比, 包括不同 O-ELM 方法的隐层节点数目 L 、正则化系数 η , 不同 O-KELM 方法的核参数 γ 、正则化系数 η 的优选结果. 从表 1 的统计结果显然看出, O-KELM 方法的 RMSE, MAE, NMSE 值均比 O-ELM 方法好, 且具有较小的预测误差, 其中 DE-KELM 方法的预测精度最高.

图 1 给出了在测试集上基于 O-KELM 方法及 O-ELM 方法, 从 $t = 624-1123$ 的提前 6 步预测输

出和系统真实输出值. 图 2 给出了不同预测方法从 $t = 624-1123$ 的提前 6 步预测输出与系统真实输出的各点预测误差分布. 由图 1 和图 2 可以看出, O-KELM 方法的预测输出的曲线几乎完全重合于系统的真实输出的曲线, 二者的差别在很小的数量级上才能区分.

图 3 进一步给出了 GA-KELM 方法的 RMSE 随进化代数变化的收敛曲线, 可以看出, GA-KELM 方法的收敛速度较快, 经过约 55 次的进化代数后趋于平稳, 具有较好的性能.

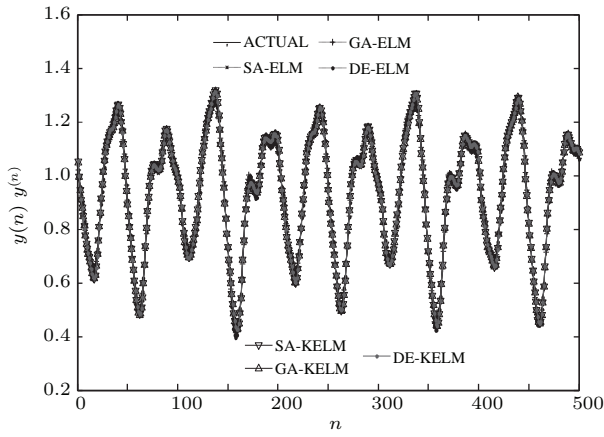


图1 O-KELM方法与O-ELM方法对Mackey-Glass时间序列预测的结果比较

Fig. 1. Comparison of prediction results produced by proposed O-KELM method and O-ELM method for Mackey-Glass time series in testing set.

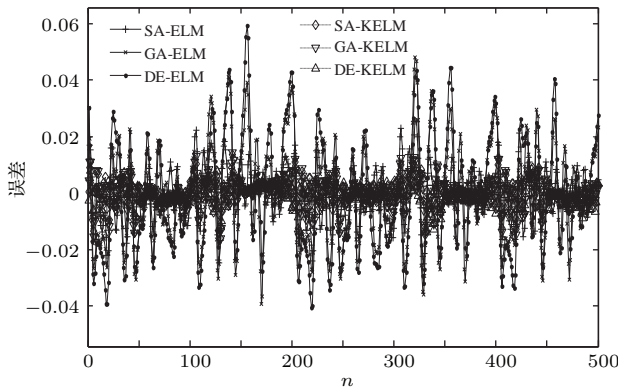


图2 O-KELM与O-ELM方法的预测误差比较曲线

Fig. 2. Comparison of prediction error based on O-KELM method and O-ELM method.

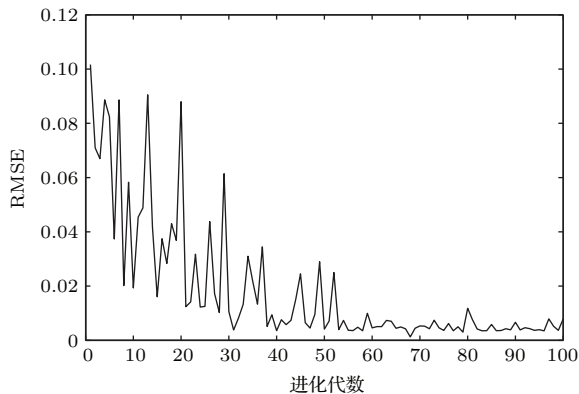


图3 GA-KELM方法的收敛曲线

Fig. 3. Convergence curve of GA-KELM method.

从4.1节的实验结果显然看出,与O-ELM方法相比较O-KELM方法预测效果明显占优,预测精度提高了约一个数量级以上,其中DE-KELM方法具有最好的预测性能。

4.2 风电功率时间序列超短期预测实验

本小节的实例实验采用实际的风场数据,选取美国国家可再生能源实验室(NREL)所提供的西部风资源Western数据集[24]。该数据集的数据通过10 min的时间间隔,2 km的空间间隔进行采样,每个网格点的距地面约100 m高度以上的10个风机的风电功率输出可通过SCORE统计技术模拟,同时也复制了风电功率输出具有随机性的特性。

实例的实验数据取自位于科罗拉多州丹佛市以西10英里位置,68个网格点,即680个同一类型风机组成的风电场,采样数据为在同一时间的68个网格点所提供的风电功率实测数据的平均值。选取2006年6月1日至10日连续10天,采样间隔为10 min的历史风电功率数据,前一半数据用于训练,其余的用于测试。

本节实验中,在O-KELM方法中,GA与DE方法的初始种群数,最大的进化代数,GA方法的交叉概率 p_c ,变异概率 p_m ,SA方法的最大的迭代次数初始温度设置,参数 μ ,DE方法的参数 F 选取等均与4.1节相同。在优化的O-ELM方法中,隐含层节点数 L 的初始化设置最大数值为500。

风电功率的超短期预测将按照时间序列建模的方式进行,即

$$y(t + D) = f(x_t), \quad \forall t = \Delta \cdots l, \quad (18)$$

其中 D 为预测步长, x_t 由历史风电功率值 $(y_{t-1}, y_{t-2}, \dots, y_{t-\Delta})$ 构成多维输入向量, Δ 表示预测模型的嵌入维数。在本节实验中,测试数据选取从2006年6月6日至10日的历史风电功率数据,初始的嵌入维数 Δ 选取为15,分别进行提前10 min的直接单步预测,提前20,30,40 min的直接多步预测,即 $D = 1, \dots, 4$,预测模型 $f(\cdot)$ 可分别采用优化KELM及优化ELM方法构建。

经过训练后,表2给出了GA-ELM方法分别进行提前10,20,30和40 min预测时,基于GA优化KELM网络结构的结果,即针对网络输入进行结构选择的结果,其中二进制变量为1表示该输入变量被选择,为0表示删除该输入。从表2可以看出,在KELM预测模型的构建过程中,存在一些不相关的输入,因此,通过O-KELM方法,选择输入变量,即选择网络的结构是可行的。

表2 基于GA-KELM方法优化后的KELM网络的输入变量选择
Table 2. Selection of input variables of the optimized KELM network by GA-KELM method.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
10 min	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0
20 min	1	0	1	0	0	0	0	0	1	0	1	0	0	0	1
30 min	1	0	1	1	0	0	0	1	0	0	1	0	0	0	1
40 min	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0

在测试集上, 不同O-KELM方法与不同O-ELM方法分别进行提前10, 20, 30和40 min预测时的 E_{MAE} 指标对比, 以及不同O-ELM方法的隐层节点数目 L 、正则化系数 η , 不同O-KELM方法的核参数 γ 、正则化系数 η 的优选数值结果分别由表3—表6给出.

表3 提前10 min预测的O-KELM方法与O-ELM方法的性能指标对比

Table 3. Evaluation of performance indicator for 10-minute ahead prediction using O-KELM method and O-ELM method.

预测方法	MAE	L	γ	η
GA-ELM	0.9628	78	—	2.1223
DE-ELM	0.9673	87	—	5.0793
SA-ELM	1.1608	150	—	19.107
GA-KELM	0.4207	—	599.34	2.762×10^{-2}
DE-KELM	0.4655	—	588.97	1.518×10^{-2}
SA-KELM	0.4773	—	473.38	4.829×10^{-2}

表4 提前20 min预测的O-KELM方法与O-ELM方法的预测性能指标对比

Table 4. Evaluation of performance indicator for 20-minute ahead prediction using O-KELM method and O-ELM method.

预测方法	MAE	L	γ	η
GA-ELM	1.0962	210	—	3.1028
DE-ELM	1.3106	230	—	0.6145
SA-ELM	2.1290	192	—	1.2767
GA-KELM	0.9405	—	154.32	1.006×10^{-2}
DE-KELM	1.0066	—	121.22	1.016×10^{-2}
SA-KELM	1.0399	—	195.99	1.173×10^{-2}

为了更直观地表明算法的性能, 在测试集上, 本文的GA-KELM, DE-KELM, SA-KELM方法与用于对比的GA-ELM, DE-ELM, SA-ELM方法分别进行提前10, 20, 30和40 min预测的NMSE由图4给出. 基于O-KELM方法进行预测的RMSE则由图5给出. 从图4和图5的结果显然看出, O-KELM方法的RMSE, NMSE值均比O-ELM方法

好, 具有较小的预测误差. 其中, 提前10, 20, 40 min预测时, GA-KELM方法具有最小的RMSE值, 提前30 min预测时, DE-KELM方法具有最小的RMSE及NMSE值.

表5 提前30 min预测的O-KELM方法与O-ELM方法的预测性能指标对比

Table 5. Evaluation of performance indicator for 30-minute ahead prediction using O-KELM method and O-ELM method.

预测方法	MAE	L	γ	η
GA-ELM	1.3962	167	—	4.902×10^{-1}
DE-ELM	1.2306	204	—	9.407×10^{-1}
SA-ELM	2.7869	135	—	9.120×10^{-2}
GA-KELM	1.3606	—	120.43	3.038×10^{-2}
DE-KELM	1.0678	—	168.12	1.033×10^{-2}
SA-KELM	1.3510	—	185.25	1.607×10^{-2}

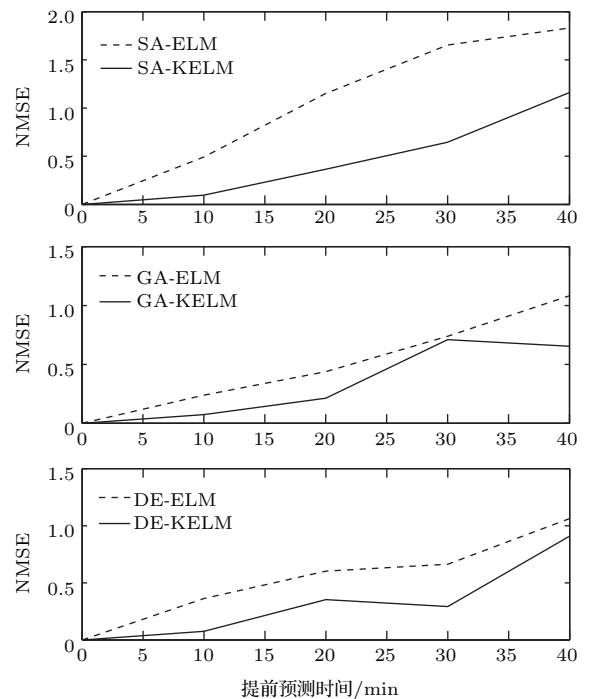


图4 不同预测步长下的O-KELM方法与O-ELM方法的NMSE曲线

Fig. 4. Curves of NMSE results using O-KELM and O-ELM method at different prediction horizon.

表6 提前40 min 预测的 O-KELM方法与O-ELM方法的预测性能指标对比

Table 6. Evaluation of performance indicator for 40-minute ahead prediction using O-KELM method and O-ELM method.

预测方法	MAE	L	γ	η
GA-ELM	1.8102	174	—	3.6009
DE-ELM	1.7106	218	—	2.1616
SA-ELM	3.1097	131	—	4.977×10^{-1}
GA-KELM	1.6202	—	81.856	2.183×10^{-2}
DE-KELM	1.6771	—	142.84	1.033×10^{-2}
SA-KELM	1.8459	—	100.32	3.997×10^{-2}

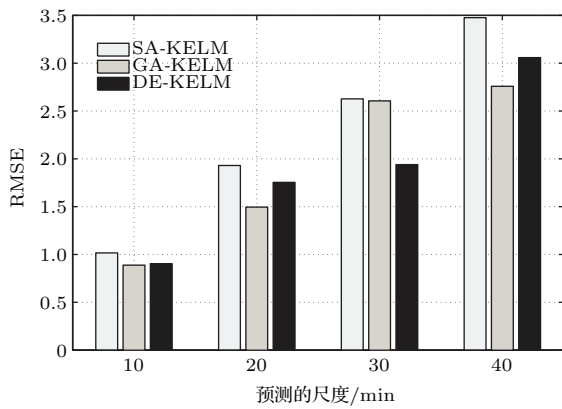


图5 不同预测步长下的 O-KELM 方法的 RMSE 直方图
Fig. 5. Histogram of RMSE results using O-KELM method at different prediction horizon.

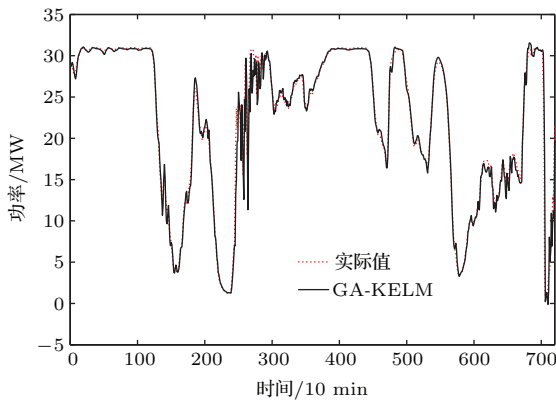


图6 (网刊彩色) 基于 GA-KELM 方法的提前 10 min 预测的风电功率实际值与预测值曲线
Fig. 6. (color online) Comparison of prediction result using GA-KELM method and actual value of wind power for 10-minute ahead prediction.

图6进一步给出了在测试集上基于 GA-KELM 方法进行提前 10 min 预测时, 风电功率实际输出与预测输出对比的预测效果图; 图7进一

步给出了在 2006 年 8 日 22 时至 9 日 8 时的时间段进行提前 10 min 预测时, 不同的 O-KELM 方法与 O-ELM 方法的预测输出与实际输出的各点预测误差对比曲线. 由图6和图7的预测效果可以看出, O-KELM 方法显示出很好的预测效果.

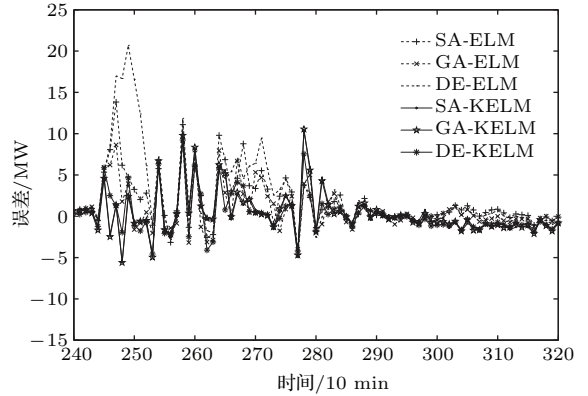


图7 提前 10 min 预测的 O-KELM 方法与 O-ELM 方法的预测误差比较曲线
Fig. 7. Comparison of prediction error between O-KELM method and O-ELM method for 10-minute ahead prediction.

5 结 论

与 ELM 方法相比, KELM 方法无需确定隐含层节点的数目, 在 ELM 的特征映射函数未知时, 应用核函数的技术, 基于正则化最小二乘算法获取输出权值的解. 因此, KELM 具有更好的逼近精度和泛化能力. KELM 方法中的核参数、正则化系数通常由交叉验证方法确定, 为进一步改进 KELM 方法的学习性能, 本文给出一类 O-KELM 方法, 分别基于 GA, SA, DE 三种进化计算策略, 优化选择 KELM 的核参数与正则化系数, 与此同时, 还优化选择 KELM 结构的输入, 剔除不相关的输入变量. 所提出的 O-KELM 方法分别应用到 Mackey-Glass 混沌时间序列预测及某地区的超短期风电功率时间序列预测的实例中, 在同等条件下, 与 O-ELM 方法进行了比较, 以验证本文方法的有效性. 实验结果表明, 本文方法在预测精度上明显优于 O-ELM 方法, 显示出其在风电功率超短期预测中的应用潜力.

参考文献

[1] Wang X C, Guo P, Huang X B 2011 *Energy Procedia* 12 770

- [2] Zhang G Y, Wu Y G, Zhang Y, Dai X L 2014 *Acta Phys. Sin.* **63** 138801 (in Chinese) [章国勇, 伍永刚, 张洋, 代贤良 2014 物理学报 **63** 138801]
- [3] Foley A M, Leahy P G, Marvuglia A, McKeogh E J 2012 *Renew. Energy* **37** 1
- [4] Colak I, Sagiroglu S, Yesilbudak M 2012 *Renew. Energy* **46** 241
- [5] Jung J, Broadwater R P 2014 *Renew. Sustain. Energy Rev.* **31** 762
- [6] Yang Z L, Feng Y, Xiong D F, Yang Z, Zhang X, Zhang J 2015 *Smart Grid* **3** 1 (in Chinese) [杨正瓴, 冯勇, 熊定方, 杨钊, 张玺, 张军 2015 智能电网 **3** 1]
- [7] Potter C W, Negnevitsky M 2006 *IEEE Trans. Power Syst.* **21** 965
- [8] Kavasseri R G, Seetharaman K 2009 *Renew. Energy* **34** 1388
- [9] Louka P, Galanis G, Siebert N, Kariniotakis G, Katsafados P, Kallos G, Pytharoulis I 2008 *J. Wind Eng. Ind. Aerodyn.* **96** 2348
- [10] Liu R Y, Huang L 2012 *Autom. Electr. Power Syst.* **36** 18 (in Chinese) [刘瑞叶, 黄磊 2012 电力系统自动化 **36** 18]
- [11] Seo I Y, Ha B N, Lee S W, Jang M J, Kim S O, Kim S J 2012 *J. Energy Power Eng.* **6** 1605
- [12] Zeng J W, Wei Q 2012 *IEEE Trans. Sustain. Energy* **3** 255
- [13] Huang G B, Zhu Q Y, Siew C K 2006 *Neurocomputing* **70** 489
- [14] Chen H Y, Gao P Z, Tan S C, Fu X K 2014 *Acta Phys. Sin.* **63** 200505 (in Chinese) [陈涵瀛, 高璞珍, 谭思超, 付学宽 2014 物理学报 **63** 200505]
- [15] Matias T, Souza F, Araújo R, Antunes C H 2014 *Neurocomputing* **129** 428
- [16] Wan C, Xu Z, Pinson P, Dong Z Y, Wong K P 2014 *IEEE Trans. Power Syst.* **29** 1033
- [17] Huang G B, Zhou H M, Ding X J, Zhang R 2012 *IEEE Trans. Syst. Man, and Cybern. B: Cybern.* **42** 513
- [18] Wang X Y, Han M 2015 *Acta Phys. Sin.* **64** 070504 (in Chinese) [王新迎, 韩敏 2015 物理学报 **64** 070504]
- [19] Zhang Y T, Ma C, Li Z N, Fan H B 2014 *J. Shanghai Jiaotong Univ. (Sci.)* **48** 641 (in Chinese) [张英堂, 马超, 李志宁, 范洪波 2014 上海交通大学学报 **48** 641]
- [20] Haupt R L, Haupt S E 2004 *Practical Genetic Algorithms* (2nd Ed. with CD-ROM) (New York: John Wiley & Sons) pp27–66
- [21] Mohamed M H 2011 *Neurocomputing* **74** 3180
- [22] Feoktistov V 2006 *Differential Evolution: In Search of Solutions* (Secaucus, NJ, USA: Springer-Verlag New York, Inc.) pp1–82
- [23] Machev M C, Glass L 1977 *Science* **197** 287
- [24] Potter C W, Lew D, McCaa J, Cheng S, Eichelberger S, Gritmit E 2008 *Wind Eng.* **32** 325

Wind power time series prediction using optimized kernel extreme learning machine method*

Li Jun[†] Li Da-Chao

(School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

(Received 18 January 2016; revised manuscript received 15 April 2016)

Abstract

Since wind has an intrinsically complex and stochastic nature, accurate wind power prediction is necessary for the safety and economics of wind energy utilization. Aiming at the prediction of very short-term wind power time series, a new optimized kernel extreme learning machine (O-KELM) method with evolutionary computation strategy is proposed on the basis of single-hidden layer feedforward neural networks. In comparison to the extreme learning machine (ELM) method, the number of the hidden layer nodes need not be given, and the unknown nonlinear feature mapping of the hidden layer is represented with a kernel function. In addition, the output weights of the networks can also be analytically determined by using regularization least square algorithm, hence the kernel extreme learning machine (KELM) method provides better generalization performance at a much faster learning speed. In the O-KELM, the structure and the parameters of the KELM are optimized by using three different optimization algorithms, i.e., genetic algorithm (GA), differential evolution (DE), and simulated annealing, meanwhile, the output weights are obtained by a least squares algorithm just the same as by the ELM, but using Tikhonov's regularization in order to further improve the performance of the O-KELM. The utilized optimization algorithms of the O-KELM are respectively used to select the set of input variables, regularization coefficient as well as hyperparameter of kernel function. The proposed method is first applied to the direct six-step prediction for Mackey-Glass chaotic time series, under the same condition as the existing optimized ELM method. From the analysis of the simulation results it can be verified that the prediction accuracy of the proposed O-KELM method is increased by about one order of magnitude over that of the optimized ELM method. Furthermore, the DE-KELM algorithm can achieve the lowest root mean square error (RMSE). The O-KELM method is then applied to real-world wind power prediction instance, i.e., the Western Dataset from NERL. The 10-minute ahead single-step prediction as well as 20-minute ahead, 30-minute ahead, 40-minute ahead multi-step prediction for wind power time series are respectively implemented to evaluate the O-KELM method. Experimental results of each of the short-term wind power time series predictions at different time horizons confirm that the proposed O-KELM method tends to have better prediction accuracy than the optimized ELM method. Moreover, the GA-KELM algorithm outperforms other two O-KELM algorithms at future 10-minute, 20-minute, 40-minute ahead prediction in terms of the RMSE value. The DE-KELM algorithm outperforms other algorithms at future 30-minute ahead prediction in terms of the normalized mean square error and the RMSE value. The results from these applications demonstrate the effectiveness and feasibility of the proposed O-KELM method. Therefore, the O-KELM method has a potential future in the field of wind power prediction.

Keywords: kernel extreme learning machine, optimization method, time series, prediction

PACS: 05.45.Tp, 88.50.Mp

DOI: 10.7498/aps.65.130501

* Project supported by the National Natural Science Foundation of China (Grant No. 51467008).

[†] Corresponding author. E-mail: lijun691201@mail.lzjtu.cn