

DOI: 10.7667/PSPC171782

基于IKFCM与多模态SSO优化SVR的光伏发电短期预测

黄予春¹, 曹成涛², 顾海³

(1. 国网河南省电力公司漯河供电公司, 河南 漯河 462000; 2. 华南理工大学, 广东 广州 510640;
3. 哈尔滨工业大学, 黑龙江 哈尔滨 150001)

摘要: 为提高短期光伏发电功率预测精度和降低气候等因素对预测结果的影响, 提出了一种基于 IKFCM 与多模态 SSO 优化 SVR 的光伏发电功率短期预测方案。首先采用改进的 KFCM(Improved KFCM, IKFCM)聚类方法对训练样本集进行处理, 通过引入紧致离散聚类有效性指数, 在提高 IKFCM 聚类准确率的同时实现了自动划分训练样本集, 有效降低了样本数据差异对预测性能的影响。然后构建与训练样本集分类一一对应的 SVR 预测模型, 并采用多模态 SSO 优化(Multi-mode SSO, MSSO)算法对 SVR 模型参数进行优化, 进而得到不同分类的最优 SVR 参数组合。最后, 运用 MSSO 优化 SVR 模型对测试数据进行预测评估。仿真结果表明, 该方案实现了不同天气下短期光伏发电功率准确预测, 而且同其他预测算法相比预测精度提高了 25.2%~37.8%。

关键词: 光伏发电功率; 核模糊 C-均值聚类; 群居蜘蛛优化; 支持向量回归(SVR)

Short-term photovoltaic power generation forecasting scheme based on IKFCM and multi-mode social spider optimization SVR

HUANG Yuchun¹, CAO Chengtao², GU Hai³

(1. Luohe Power Supply Company, State Grid Henan Electric Power Company, Luohe 462000, China; 2. South China University of Technology, Guangzhou 510640, China; 3. Harbin Institute of Technology, Harbin 150001, China)

Abstract: In order to improve the accuracy of short-term PV power prediction and reduce the influence of climate factors on forecasting results, a short-term photovoltaic power generation forecasting scheme based on Improved Kernel Fuzzy C-Means (IKFCM) and multi-mode social spider optimization SVR is proposed. Firstly, the improved KFCM (IKFCM) clustering method is used to process the training sample set. By introducing intra class scatter clustering validity index, the automatic training sample set separation is realized and the clustering accuracy of IKFCM is improved, thereby the effect of sample data difference on prediction performance is effectively reduced. Then, the SVR prediction models corresponding to training samples set classifications one to one are built, and the Multi-modal SSO (MSSO) optimization algorithm is used to optimize the parameters of SVR model, which helps to obtain the optimal SVR parameters combination for each SVR model. Finally, the MSSO optimization SVR model is used to predict the test data. Simulation results show that, the scheme can realize accurate short-term PV power prediction for different weather conditions, and compared with other prediction algorithms, the prediction accuracy is improved by 25.2%~37.8%.

This work is supported by China Spark Program (No. 2015GA780024).

Key words: forecasting of photovoltaic power generation; kernel fuzzy C-means clustering; social spider optimization; support vector regression (SVR)

0 引言

当前, 随着光伏产业规模的不断扩大以及发电技术的不断成熟, 太阳能发电已经成为新的可再生

能源发电增长点^[1]。由于光照具有昼夜周期特性, 这使得光伏电站是一种间歇型电源^[2], 而且光伏发电功率受太阳辐射强度、气象、环境条件以及光伏组件温度等因素的影响, 呈现出很强的间断性、随机性和波动性, 这些给并入电网系统功率平衡和调度管理等带来了很大的挑战^[3]。因此, 高效准确的

光伏发电功率预测机制对电网安全经济运行具有重要意义。

常见的短期光伏发电功率预测方法可以分为间接预测法和直接预测法两类, 其中间接预测法建模过程相对复杂, 模型预测精度不高并且抗干扰能力较差^[4]; 直接预测方法主要采用智能学习理论, 通过分析研究天气预报、历史运行情况等数据中的规律, 实现对未来一段时间光伏功率准确预测的目的^[5-6]。支持向量回归、极端学习机^[7]、BP 神经网络等智能学习理论在预测分类等方面展现的良好性能为光伏发电短期预测提供了新的研究思路, 学者们也围绕提高光伏发电功率算法的预测精度以及建模准确性等方面进行了一系列研究: 王昕^[4]提出了一种联合 PNN、PCA 与 SS-SVR 的光伏发电功率预测方法, 仿真实验也验证了该方法的有效性, 但是简单人为地将天气划分为 4 种不同类型, 并没有充分考虑样本数据差异对预测性能的影响; TENG^[8]等建立了基于 BP 神经网络的短期功率预测模型, 通过对天气类型等输入量的聚类识别, 实现了分季节光伏发电功率预测, 但是该方法容易陷入局部最优; 文献[9]提出了一种基于 GA 和 RBF 神经网络的光伏发电功率预测算法, 通过合理选取相似日、太阳辐射强度和温度等模型输入变量, 在提高光伏功率预测精度的同时有效避免了功率波动问题, 但是该算法对历史同期数据要求较高, 而且需要大量的样本数据, 并不适用于新建立的光伏电站; 文献[10]提出一种基于 PCA 和 GA 优化的 BP 神经网络的光伏发电功率预测算法, 利用主成分分析法和遗传算法确保了预测模型具有较好的预测性能, 但是遗传算法收敛效率低的缺陷影响了出力的预测结果。总体来看, 采用直接预测法面临的最主要的难题是如何提高预测算法的预测精度和建模的准确性。

为此, 本文提出了一种基于 IKFCM(Improved KFCM)与多模态 SSO^[11](MSSO)优化 SVR 的光伏发电功率短期预测方案。该方案采用改进的 KFCM 聚类算法^[12]对训练样本集自动分类划分, 得到不同隶属度样本集分类; 利用 MSSO 优化 SVR 模型分别对每个子类进行预测, 最终得到光伏发电功率预测结果。该方案不需要大量训练样本集, 具有较高的预测精度, 而且预测模型能够有效降低样本数据差异对预测性能的影响, 仿真实验也证明了该方案的有效性。

1 改进 KFCM 聚类方法

1.1 KFCM

FCM 是目前应用较为广泛的聚类算法之一^[12]。

考虑数据样本集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $\mathbf{x}_k \in R^S (k = 1, 2, \dots, n)$, FCM 目的是将 \mathbf{X} 划分成 C 个模糊子类 ($C \in \{2, 3, \dots, n-1\}$), 使得同类内数据具有明显相似性, 不同类间数据具有显著差异性。聚类结果可以表示为聚类中心集合 $\mathbf{V} = \{\mathbf{v}_i\} (i = 1, 2, \dots, C)$ 和隶属度矩阵 $\mathbf{U} = [\mu_{ik}]_{C \times n}$, 其中 μ_{ik} 为样本 \mathbf{x}_k 对第 i 个子类的隶属度, FCM 实质为聚类函数最优化问题, 即

$$\min J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^C \sum_{k=1}^n \mu_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (1)$$

式中, m 为模糊加权指数。分别对 μ_{ik} 、 \mathbf{v}_i 求偏导数有

$$\mu_{ik} = \frac{\|\mathbf{x}_k - \mathbf{v}_i\|^{-2/(m-1)}}{\sum_{j=1}^C \|\mathbf{x}_k - \mathbf{v}_j\|^{-2/(m-1)}}, \mathbf{v}_i = \frac{\sum_{k=1}^n \mu_{ik}^m \mathbf{x}_k}{\sum_{k=1}^n \mu_{ik}^m} \quad (2)$$

$$\text{s.t. } \sum_{i=1}^C \mu_{ik} = 1, \mu_{ik} \in [0, 1], \sum_{k=1}^n \mu_{ik} \in (0, 1) \quad (3)$$

FCM 利用式(2)经过多次迭代最终实现样本集分类, 由于采用欧氏距离作为相似度量, 导致 FCM 仅适用于处理具有较好紧致度和离散度的数据, 为了扩展 FCM 应用范围, 引入核 FCM(KFCM)聚类算法, 即

$$\min J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^C \sum_{k=1}^n \mu_{ik}^m \|\Phi(\mathbf{x}_k) - \Phi(\mathbf{v}_i)\|^2 \quad (4)$$

令 $K(\mathbf{x}, \mathbf{v}) = \Phi^T(\mathbf{x})\Phi(\mathbf{v})$ 并对式(2)~式(4)推导有

$$\min J(\mathbf{U}, \mathbf{V}) = 2 \sum_{i=1}^C \sum_{k=1}^n \mu_{ik}^m [1 - K(\mathbf{x}_k, \mathbf{v}_i) + K(\mathbf{v}_i, \mathbf{v}_i)] \quad (5)$$

$$\mu_{ik} = [1 - K(\mathbf{x}_k, \mathbf{v}_i)]^{-\frac{1}{m-1}} / \left(\sum_{j=1}^C [1 - K(\mathbf{x}_k, \mathbf{v}_j)]^{-\frac{1}{m-1}} \right) \quad (6)$$

$$\mathbf{v}_i = \left(\sum_{k=1}^n \mu_{ik}^m K(\mathbf{x}_k, \mathbf{v}_i) \mathbf{x}_k \right) / \left(\sum_{k=1}^n \mu_{ik}^m K(\mathbf{x}_k, \mathbf{v}_i) \right) \quad (7)$$

研究表明, 聚类数目需要事先确定以及模糊加权指数 m 对聚类结果影响较大仍是 KFCM 亟需解决的主要问题之一。

1.2 改进 KFCM(IKFCM)

针对聚类数目事先确定以及 KFCM 算法聚类准确率差的缺陷, 引入紧致离散聚类有效性指数, 在提高改进 KFCM(IKFCM)算法聚类准确率的同时实现了聚类子类数目的自动划分。

定义 1 紧致离散聚类有效性指数。对于 IKFCM 算法, 定义类内离散度聚类有效性指数 V_D 为

$$V_D = \frac{\frac{1}{n} \sum_{i=1}^C \sum_{j=1}^n \mu_{ik}^m \|\mathbf{v}_i - \mathbf{x}_j\|^2 + \max_{1 \leq j \leq n} \|\mathbf{x}_j - \bar{\mathbf{v}}_i\|^2}{\min_{i \neq j} \|\mathbf{v}_i - \mathbf{v}_j\|^2 + \delta} \quad (8)$$

式中, δ 为惩罚因子。从式(8)可以看出, V_D 反映了类内紧致度与类间离散度的比值, V_D 取值越小表明聚类结果越好。

结论 1 采用紧致离散聚类有效性指数的 IKFCM 算法具有较好的鲁棒性。

证明: V_D 代表了不同分类情况下 IKFCM 聚类结果的好坏, 对于极端情况, 即 $C \rightarrow n$ 或 $m \rightarrow \infty$ 有:

1) 当 $C \rightarrow n$ 时, \mathbf{v}_i 为样本本身, 即

$$\lim_{C \rightarrow n} \sum_{i=1}^C \sum_{j=1}^n \mu_{ik}^m \|\mathbf{v}_i - \mathbf{x}_j\|^2 = 0. \text{ 此时有}$$

$$\lim_{C \rightarrow n} V_D = \left(\max_{1 \leq j \leq n} \|\mathbf{x}_j - \bar{\mathbf{x}}_i\|^2 \right) / \left(\min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|^2 + \delta \right) \quad (9)$$

2) 当 $C \rightarrow n$ 与 $m \rightarrow \infty$ 时, 所有聚类中心重合,

即 $\lim_{m \rightarrow \infty} \|\mathbf{v}_i - \mathbf{v}_j\|^2 = 0$ 。此时有

$$\lim_{\substack{m \rightarrow \infty \\ C \rightarrow n}} V_D = \left(\max_{1 \leq j \leq n} \|\mathbf{x}_j - \bar{\mathbf{x}}_i\|^2 \right) / \delta \quad (10)$$

从式(9)、式(10)可以看出, V_D 对不同 C 或 m 都具有较好的鲁棒性, 能够获取较为理想的聚类结果。证毕。

基于紧致离散聚类有效性指数的 IKFCM 算法能够实现对本样本集聚类的自动子类划分, 其工作过程如图 1 所示。

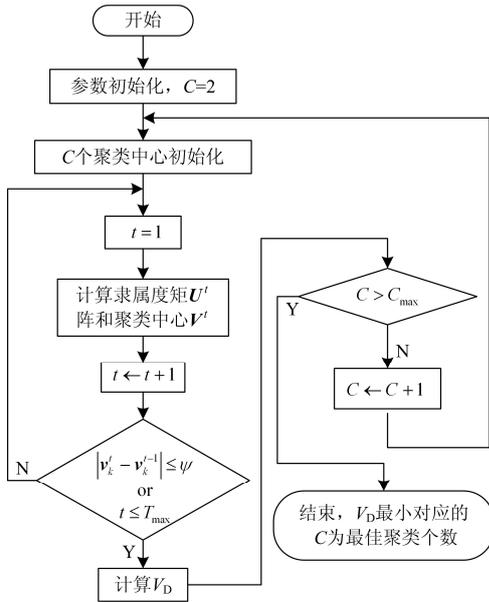


图 1 IKFCM 工作流程示意图

Fig. 1 Workflow schematic diagram of IKFCM

2 多模态 SSO 优化 SVR

2.1 支持向量回归(SVR)

SVR^[14]隶属于小样本学习机范畴, 在函数逼近、模式识别等领域展现出了良好性能。SVR 实质为最优化问题^[15-16], 其拟合线可以描述为

$$y = \boldsymbol{\omega}^T \cdot \boldsymbol{\phi}(\mathbf{x}) + b \quad (11)$$

式中: $\boldsymbol{\phi}(\mathbf{x})$ 为空间映射函数; $\boldsymbol{\omega}$ 、 b 为函数系数。

对于问题样本数据 (\mathbf{x}_i, y_i) , 引入松弛变量 ξ_i (ξ_i^*), 此时 SVR 拟合问题转化为

$$\min \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + C_s \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (12)$$

$$\text{s.t. } y_i - \boldsymbol{\omega}^T \mathbf{x}_i - b \leq \varepsilon + \xi_i, \boldsymbol{\omega}^T \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*$$

式中, ε 、 C_s 分别为不敏感损失参数和惩罚系数。

利用拉格朗日多项式处理式(12), 有

$$\max -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\mathbf{x}_i \cdot \mathbf{x}_j) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varepsilon \quad (13)$$

$$\text{s.t. } \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C_s]$$

式中, $K(\cdot)$ 为内积核函数, 本文选取 $K(\cdot)$ 为 Gaussian 径向基核函数, 结合 KKT 条件, 式(13)中只有少量 $(\alpha_i - \alpha_i^*) \neq 0$, 故式(13)可以表示为

$$y = \sum_{i \in \text{sv}} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (14)$$

由式(14)可见, 当设定 C_s 、 ε 和 σ 数值后就可以得到 SVR 的具体表达形式, 进而实现对数据样本的准确预测。因此, 合理选取 C_s 、 ε 和 σ 对改善 SVR 预测性能具有重要意义^[17], 本文采用多模态 SSO 算法优化 SVR 模型相关参数, 以获取最佳 C_s 、 ε 和 σ 参数组合。

2.2 多模态 SSO 优化算法(MSSO)

群居蜘蛛优化(SSO)算法作为近期提出的一种智能启发计算技术, 得到了越来越多学者的关注^[18-19], 本文在 SSO 的基础上, 引入“多模态种群”概念, 将蜘蛛种群动态划分成不同规模的子族群, 每个子族群内分别设计不同的迭代更新方式, 在提高种群多样性的同时, 提高了算法全局寻优能力(基本 SSO 原理参考相关文献, 本文不再赘述)。

1) 多模态子族群划分

对于蜘蛛种群(以雌性蜘蛛子群为例), 根据子群内蜘蛛个体适应度值划分为 3 个不同子族群 F_f^1 、 F_f^2 和

F_f^3 , 规模分别为 P_f^1 、 P_f^2 和 P_f^3 。

$$P_f^1 = \left[P_{f,\min}^1 + \alpha (P_{f,\max}^1 - P_{f,\min}^1) \right] (e^{\ln 2t/T_{\max}} - 1) \quad (15)$$

$$P_f^2 = \beta P_f^1, \quad P_f^3 = P_f - P_f^1 - P_f^2 \quad (16)$$

式中: P_f 为雌性蜘蛛子群规模; $P_{f,\max}^1$ 、 $P_{f,\min}^1$ 为 P_f^1 的最大值和最小值; α 、 β 为控制系数; F_f^1 、 F_f^3 分别由子群内适应度最优和最差的个体组成, 从式(15)、式(16)可以看出, 随着算法运算次数的不断增加, F_f^1 和 F_f^2 的规模不断扩大, 更多的蜘蛛个体以“多模态”更新策略迭代进化, 算法全局收敛性能明显改善。

2) “色散”更新策略

对于 F_f^1 内个体, 模拟光学“色散”现象(如图 2 所示), 提出“色散”更新策略, 其工作过程可以描述为: 有蜘蛛个体 $F_i \in F_f^1$, 其第 k 维变量 $F_{ik} \in [F_{ik}^{\min}, F_{ik}^{\max}]$, 以 F_{ik}^{\min} 与 F_{ik}^{\max} 的连线作为 x 轴, 并选取中间位置为坐标原点 O , 在 O 处按照“图 1”所示位置放置一个正三棱镜 ΔABC (棱边长 $2a$), 从光源 O' 处发出一束单色光射向 O 处, 在棱镜内发生折射, 并从棱镜边 AB 点 E 处射出, 在射出光线上选取点 O'' , O'' 在 x 轴上的投影即为更新后的 F_{ik}^{new} (入射光线与折射光线物理量定义见图 1)。

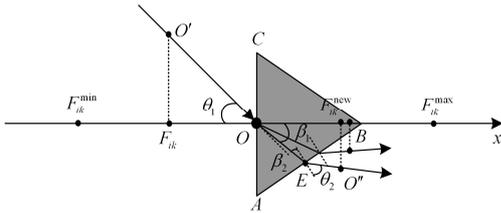


图 2 “色散”更新策略示意图

Fig. 2 Schematic diagram of "dispersion" update strategy

结论 2 蜘蛛个体 F_i 采取“色散”更新策略后, 得到新的个体 $F_i^{\text{new}} (F_{i1}^{\text{new}}, \dots, F_{ik}^{\text{new}}, \dots, F_{iN}^{\text{new}})$, 且满足:

$$F_{ik}^{\text{new}} = (F_{ik}^{\min} + F_{ik}^{\max})/2 + x_E (\|OO'\|, \eta_1) + \lambda \|OO'\| \cos(\theta_2 (\|OO'\|, \eta_1) - \pi/3) \quad (17)$$

式中: η_1 为折射率; λ 为控制系数; $x_E (\|OO'\|, \eta_1)$ 、 $\theta_2 (\|OO'\|, \eta_1)$ 分别表示为变量 $\|OO'\|$ 和 η_1 的函数; x_E 为 E 点横坐标。根据光学折射定律, 可以证明结论 2 的可行性。

从结论 2 可以看出, 当确定折射率 η_1 时, 就可以得到 F_{ik}^{new} 。为了进一步反映“色散”更新策略与群体内个体适应度间的关系, 定义 η_1 为

$$\eta_1 = (\sum J(F_i)) / (P_f^1 \cdot \max \{J(F_i)\}), F_i \in F_f^1 \quad (18)$$

式中, $J(\cdot)$ 为个体适应度值。当个体全部维度完成更新后, 若 $J(F_{ik}^{\text{new}}) \leq J(F_i)$ 则用新的个体替代原个体, 否则随机生成新的个体。从式(17)、式(18)可以看出, “色散”更新策略很大程度地提高了算法局部搜索范围, 有效避免了算法陷入局部最优解, 增强了算法全局寻优能力。

3) “自适应交叉变异”更新策略

对于 F_f^3 内个体, 设计“自适应交叉变异”更新策略, 其工作过程可以描述为: 有蜘蛛个体 $F_j (F_{j1}, \dots, F_{jk}, \dots, F_{jN}) \in F_f^1$ 、雌性种群当前最优个体 $F_b (F_{b1}, \dots, F_{bk}, \dots, F_{bN})$, F_j 逐维度以自适应交叉概率 p_i 进行更新:

$$F_{jk}^{\text{new}} = \begin{cases} F_{jk} & \text{rand}(0,1) < p_i \\ F_{bk} & \text{else} \end{cases} \quad (19)$$

$$p_i = \omega (e^{\ln 2t/T_{\max}} - 1) \quad (20)$$

式中, ω 为控制系数。从式(19)、式(20)可以看出, 随着算法迭代次数增加, 个体交叉变异概率逐渐增大, 因此有更多维度向最优个体学习, 提高了算法收敛速度。

4) 算法计算复杂度

MSSO 种群初始化计算复杂度为 $O(F)$, F_f^1 子族群进化计算复杂度为 $O(F \cdot \log N)$, F_f^2 子族群进化计算复杂度为 $O(F)$ (F_f^2 采用基本 SSO 算法更新策略), F_f^3 子族群进化计算复杂度为 $O(F \cdot \log N)$, 因此算法每迭代依次算法计算复杂度为 $O(F) + 2O(F \cdot \log N)$, 当种群完成 T_{\max} 次迭代后, MSSO 算法的计算复杂度为 $T_{\max} \cdot (O(F) + 2 \cdot O(F \cdot \log N))$ 。

2.3 多模态 SSO 优化 SVR(MSSO-SVR)实现

本文利用 MSSO 算法来优化 SVR 模型参数, 以获取最佳 SVR 参数组合。

定义 2 编码方式。对于 MSSO-SVR 定义蜘蛛个体编码方式为 $F_i(\varepsilon_i, \sigma_i, C_{s,i})$, 且满足:

$$\varepsilon_i \in [\varepsilon_{\min}, \varepsilon_{\max}], \sigma_i \in [\sigma_{\min}, \sigma_{\max}], C_{s,i} \in [C_{s,\min}, C_{s,\max}] \quad (21)$$

定义 3 适应度函数。对于 MSSO-SVR 参数优化问题, 适应度函数定义为

$$\min J = \left[\sum_{i=1}^n (\hat{y}_i - y_i)^2 \right] / n \quad (22)$$

式中, \hat{y}_i 、 y_i 分别为预测值和实测值。MSSO 优化 SVR 参数实现原理如图 3 所示。

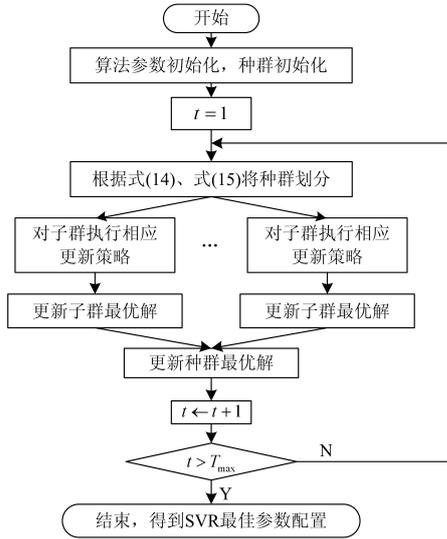


图 3 MSSO 优化 SVR 参数实现流程图

Fig. 3 Parameter implementation flow chart of MSSO-SVR

3 光伏发电短期预测

3.1 影响光伏发电的因素

在光伏电站实际应用过程中, 系统输出功率 P_s 约为^[8]

$$P_s = \eta_{pv} S I_r [1 - 0.005(T_c + 25)] \quad (23)$$

式中, η_{pv} 、 S 、 I_r 和 T_c 分别为光伏阵列的转换效率、总面积、接收的太阳辐射强度以及光伏组件板温度。对于设定好的光伏电站, 影响其发电功率的主要因素为气候条件, 综合大量文献研究, 本文选取云团记录数(C_l)、降水量(P_{re})、风速(W_v)、气压(P_a)、日照强度(I_r)、相对湿度(H_u)、风向(W_d)和温度(T_m)等 8 个气候条件作为主要因素进行分析。

定义 4 训练数据样本集。对于 MSSO-SVR 预测模型, 其训练数据样本集 $S_n = \{(x_i, y_i)\}$ 为

$$x_i = (C_{l,i}, P_{re,i}, P_{a,i}, I_{r,i}, H_{u,i}, W_{d,i}, T_{m,i}), y_i = P_{s,i} \quad (24)$$

式中, 这 8 个主要影响因素可以从气象部门获取。

3.2 光伏发电短期预测

本文采用基于 IKFCM 与 MSSO-SVR 的光伏发电功率短期预测机制对某时刻系统的发电功率进行预测。设需要预测的时间序列为 $\Delta t = \{d_1, \dots, d_K\}$, 通常 d_k ($k=1, 2, \dots, K$) 为 1 自然天数, 并且分别对每个整点时刻光伏功率进行预测。对于 d_k , 选取其对应历史时刻前后 n 天内整点时刻观测数据作为训练样本集, 采用 IKFCM 聚类对训练样本集进行自动聚类划分, 得到具有不同隶属度的样本集分类, 在

此基础上采用 MSSO-SVR 模型分别对每个聚类分类进行训练, 以获取最佳 $(\varepsilon_{best}, \sigma_{best}, C_{s,best})$, 最后分别寻找 d_k 每个整点时刻对应的最相似分类, 并得到所需预测结果。图 4 给出了光伏发电短期预测机制示意图。

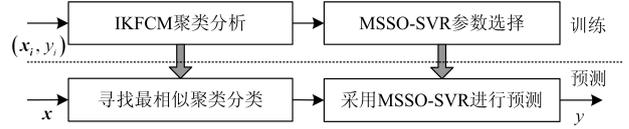


图 4 MSSO-SVR 工作示意图

Fig. 4 Schematic diagram of MSSO-SVR

图 4 给出了基于 IKFCM 与多模态 SSO 优化 SVR 的光伏发电短期预测实现流程图, 其实现过程可以描述为: 训练阶段, 对于训练数据样本集 (x_i, y_i) , 利用 IKFCM 算法对样本数据进行聚类分析, 得到多个聚类分类, 此时利用 MSSO-SVR 对所有聚类分类进行训练, 得到每个分类对应的最佳 SVR 参数配置; 真实数据预测阶段, 对于待预测数据 x , 采用与其最相似的聚类分类对应的 MSSO-SVR 模型进行预测, 最终得到所需预测结果 \hat{y}_i 。可以预见, 基于 IKFCM 与多模态 SSO 优化 SVR 的光伏发电短期预测方法具有较高的预测精度, 是因为首先对于训练样本集, 采用 IKFCM 对训练样本进行聚类分析, 使得具有更多相似特点的样本归为一类, 例如, 通过 IKFCM 可以将不同天气下的样本数据进行有效分割, 从而降低了数据差异性对预测结果的影响; 其次, 针对 IKFCM 得到的不同聚类, 分别采用 MSSO-SVR 进行预测分析, 由于 SVR 存在难以确定参数的缺陷, 利用 MSSO 进行优化处理, 从而得到最佳参数配置; 最后, 对于预测阶段, 通过找到待预测时刻对应的最相似分类, 并利用该分类对应的 MSSO-SVR 模型进行光伏发电功率预测, 从而得到所需预测结果。IKFCM 突出的自适应聚类性能够得到更好的训练样本分类结果, 而基于 MSSO 的 SVR 算法确保了预测结果的准确性。

4 实验仿真

选取文献[9]列举的光伏发电系统进行数据分析, 其中海拔 595 m、纬度 42.19 °C、经度 122.70 °C, 装机总量 15 kW, 采样频率 20 min。分别选取 6 月和 12 月每天整点时刻光伏发电功率作为预测对象, 并选取去年 5 月至 7 月以及 11 月至来年 1 月每天整点时刻气象数据和发电功率作为训练样本集数据。利用 Matlab 仿真平台对光伏短期发电功率进行预测分析, IKFCM 以及 MSSO-SVR 相关参数设置如

下: $C_{\max} = 20$ 、 $m = 3$ 、 $P = 200$ 、 $T_{\max} = 500$ 、 $P_{f,\max}^1 = 80$ 、 $P_{f,\min}^1 = 20$ 、 $\alpha = 0.6$ 、 $\beta = 0.2$ 、 $10 \leq C_s \leq 100$ 、 $0.1 \leq \varepsilon \leq 0.5$ 、 $0.01 \leq \sigma \leq 10$ 。为了进一步分析 MSSO-SVR 预测精度, 分别设计均值绝对误差 \overline{MAE} 、均方误差 $RMSE$ 和均等系数 \overline{EC} ^[20] 3 个评价指标。

$$\overline{MAE} = \left(\sum_{i=1}^n |\hat{y}_i - y_i| \right) / n, \quad RMSE = \left[\sum_{i=1}^n (\hat{y}_i - y_i)^2 \right] / n \quad (25)$$

$$\overline{EC} = 1 - \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2} / \left[\sqrt{\sum_{i=1}^n y_i^2} + \sqrt{\sum_{i=1}^n \hat{y}_i^2} \right] \quad (26)$$

式中, \overline{EC} 代表了预测值与真实值的拟合度, 取值越高表明预测结果越好。

4.1 仿真结果分析

采用本文提出的光伏发电短期预测算法对不同月份测试样本进行实验, 实验过程中对 6 月和 12 月每天整点时刻都进行预测, 表 1 给出了有关参数优化及评价指标对比结果。为了更加直观地给出预测对比结果, 选取 6 月份不同天气下 24 小时预测结果(6 月 4 号晴天、6 月 10 号多云、6 月 15 号阴天、6 月 20 号中雨), 图 5 给出了这 4 种代表性天气条件下的光伏发电功率预测结果, 图 6 给了上述 4 种代表性天气条件相对应的 MSSO 收敛曲线。

表 1 MSSO-SVR 参数优化结果

Table 1 MSSO-SVR parameter optimization results

| | C | \overline{MAE} | $RMSE$ | \overline{EC} |
|------|-----|------------------|--------|-----------------|
| 6 月 | 15 | 2.54 | 5.78 | 0.96 |
| 12 月 | 17 | 3.97 | 9.96 | 0.95 |

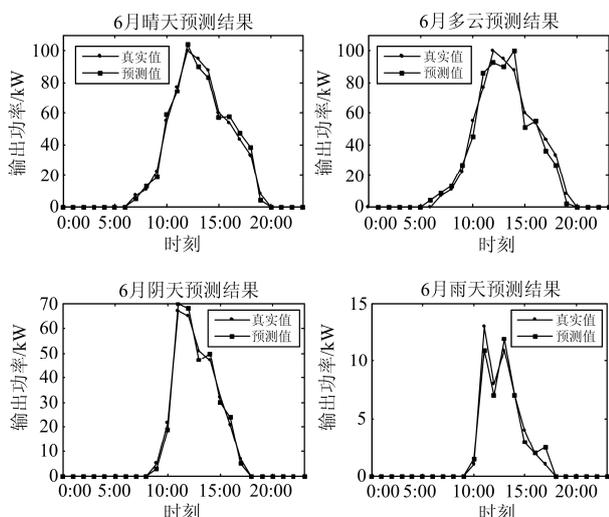


图 5 4 种代表性天气下光伏发电预测结果

Fig. 5 Forecasting results of photovoltaic power generation under 4 typical weather conditions

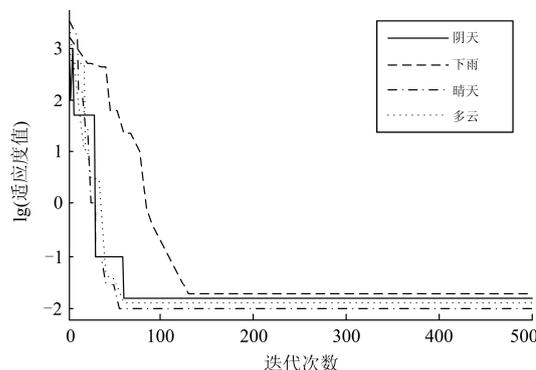


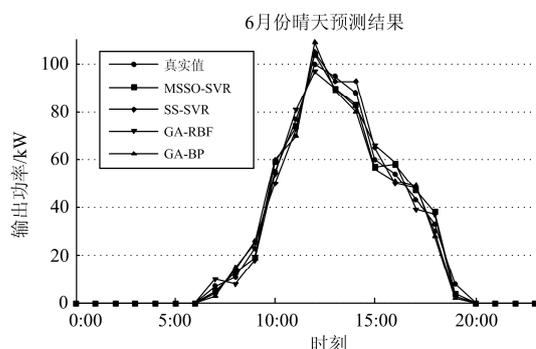
图 6 MSSO 优化 SVR 收敛曲线

Fig. 6 SVR convergence curve under MSSO

从表 1 可以看出, IKFCM 将 6 月和 12 月分别对应的训练数据样本集分成 15 和 17 个分类, 进一步细化了样本之间的不同属性, 从而有效降低了样本差异性对预测结果的影响, 提高了预测结果的可信度。对于预测结果, 无论是 6 月还是 12 月, 预测结果的 \overline{MAE} 和 $RMSE$ 指标都在合理范围内, 特别地对 \overline{EC} 指标都在 0.95 以上, 这表明预测值与真实值具有很高的拟合度。从图 5 可以看出, 无论对于哪种气候条件, 都能够给出准确的预测结果, 并且真实值与预测值曲线贴合很紧密。从图 6 可以看出, MSSO 算法只需要迭代 150 次左右就可以收敛于最优解, 表明该算法具有较快的收敛速度, 而且收敛精度较高。

4.2 对比实验分析

为了进一步对比分析 MSSO-SVR 性能, 分别选取文献[5]提出的 SS-SVR 方法、文献[8]提出 GA-RBF 方法以及文献[10]提出的 GA-BP 预测模型进行对比实验。实验过程中对 6 月和 12 月每天整点时刻都进行预测, 表 2 给出了不同预测方法评价指标对比。同 4.1 节实验类似, 给出 6 月 4 号晴天、6 月 10 号多云、6 月 15 号阴天、6 月 20 号中雨 4 种典型天气下不同预测方法对比结果(如图 7 所示)。



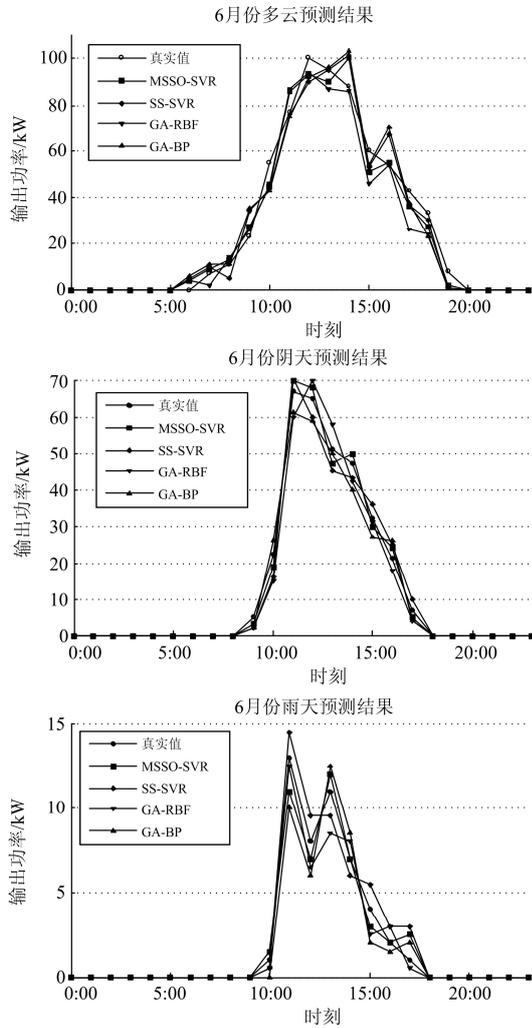


图 7 不同预测方法光伏发电预测结果

Fig. 7 Forecasting results of photovoltaic power generation by different forecasting methods

表 2 不同算法评价指标对比结果

Table 2 Comparison results of evaluation indexes of different algorithms

| 预测方法 | 评价指标 | | | |
|------|----------|------|-------|-------|
| | MAE | RMSE | EC | |
| 6 月 | MSSO-SVR | 3.12 | 7.24 | 0.954 |
| | SS-SVR | 4.34 | 8.05 | 0.931 |
| | GA-RBF | 5.90 | 9.14 | 0.928 |
| | GA-BP | 7.77 | 11.46 | 0.895 |
| 12 月 | MSSO-SVR | 2.71 | 4.64 | 0.950 |
| | SS-SVR | 4.15 | 12.43 | 0.913 |
| | GA-RBF | 4.86 | 14.26 | 0.887 |
| | GA-BP | 5.07 | 15.43 | 0.857 |

从表 2 可以看出, 无论是 6 月份还是 12 月份, MSSO-SVR 预测评价指标都要好于其他三种算法, 其中预测均值绝对误差 MAE 提高了 25.2%~37.8%,

而对于均等系数 \overline{EC} 指标, MSSO-SVR 均达到了 0.95 以上, 而其他算法基本在 0.90 左右, 最差的是 GA-BP, 其 \overline{EC} 最大值才只有 0.895, 明显小于其他三种算法。从图 7 可以看出, 对于不同的气候条件, MSSO-SVR 预测曲线更贴近真实值曲线, 这表明 MSSO-SVR 预测值与实测值之间具有更高的拟合度, 预测精度更高。

4.3 结果分析

从 4.1 节和 4.2 节仿真实验结果来看, 基于 IKFCM 与多模态 SSO 优化 SVR 的光伏发电功率短期预测精度更高, 更具有实用性。其中, MSSO-SVR 预测模型具有较高的预测精度, 对于绝大多数预测样本数据, MSSO-SVR 预测均等系数都达到了 0.95 以上, 而且 MSSO-SVR 模型预测性能要好于常见的其他几种预测方法, 特别的 MSSO-SVR 模型预测均值绝对误差平均提高了约 30.2%, 预测均等系数提高了约 5.5%。这是因为:

1) IKFCM 改进聚类方法的引入, 实现了对训练样本集的“高质量”划分, 为测试数据提供了更加精确的参考样本, 从而有效降低了样本差异性对预测结果的影响, 而且该方法对训练样本集要求不高, 使得方法扩展性更强。

2) 多模态 SSO(MSSO)算法的引入, 很大程度地改善了基本 SSO 算法全局寻优能力, 为 SVR 参数配置提供了新的思路, 而且参数配置更优, 进一步提高了 MSSO-SVR 预测模型的准确性, 从而使得预测结果具有更高的精度和可信度。

3) IKFCM 与 MSSO 有机结合, 为预测数据提供了更加合理的对应训练样本和更加精确的预测工具, 从而保证了预测结果的有效性。

5 结论

提出了一种联合 IKFCM 与多模态 SSO 优化 SVR 的光伏发电短期预测方案, 该方案综合了模糊聚类 and 智能学习理论, 并利用群体启发式计算技术快速寻优的优势, 提高了光伏发电短期预测精度和可靠性, 具有一定的推广意义。下一步将重点围绕算法预测可靠性和可移植性等方面进行研究。

参考文献

[1] 余志强, 王淳, 胡奕涛, 等. 并网光伏发电置信容量评估[J]. 电力系统保护与控制, 2016, 44(7): 122-127.
YU Zhiqiang, WANG Chun, HU Yitao, et al. Capacity credit evaluation of grid-connected photovoltaic generation[J]. Power System Protection and Control, 2016, 44(7): 122-127.

[2] HE Jinghan, LIU Lin, LI Wenli, et al. Development and

- research on integrated protection system based on redundant information analysis[J]. Protection and Control of Modern Power Systems, 2016, 1(1): 108-120. DOI: 10.1186/s41601-016-0024-y.
- [3] JIANG Yu, CHEN Xingying, YU Kun, et al. Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm[J]. Journal of Modern Power Systems and Clean Energy, 2018, 6(2): 223-234.
- [4] 王昕, 黄柯, 郑益慧, 等. 基于 PNN/PCA/SS-SVR 的光伏发电功率短期预测方法[J]. 电力系统自动化, 2016, 40(17): 156-162.
WANG Xin, HUANG Ke, ZHENG Yihui, et al. Short-term forecasting method of photovoltaic output power based on PNN/PCA/SS-SVR[J]. Automation of Electric Power Systems, 2016, 40(17): 156-162.
- [5] 王新普, 周想凌, 邢杰, 等. 一种基于改进灰色 BP 神经网络组合的光伏出力预测方法[J]. 电力系统保护与控制, 2016, 44(18): 81-87.
WANG Xinpu, ZHOU Xiangling, XING Jie, et al. A prediction method of PV output power based on the combination of improved grey back propagation neural network[J]. Power System Protection and Control, 2016, 44(18): 81-87.
- [6] 陈志宝, 丁杰, 周海, 等. 地基云图结合径向基函数人工神经网络的光伏功率超短期预测模型[J]. 中国电机工程学报, 2015, 35(3): 561-567.
CHEN Zhibao, DING Jie, ZHOU Hai, et al. A model of very short-term photovoltaic power forecasting based on ground-based cloud images and RBF neural network[J]. Proceedings of the CSEE, 2015, 35(3): 561-567.
- [7] 郭志强, 王沁, 万亚东, 等. 基于综合性评估的无线链路质量分类预测机制[J]. 计算机研究与发展, 2013, 50(6): 1227-1238.
GUO Zhiqiang, WANG Qin, WAN Yadong, et al. Wireless link quality classification prediction mechanism based on comprehensive evaluation[J]. Journal of Computer Research and Development, 2013, 50(6): 1227-1238.
- [8] TENG Xianliang, GAO Zonghe, ZHANG Yingyuan, et al. Key technologies and the implementation of wind, PV and storage co-generation monitoring system[J]. Journal of Modern Power Systems and Clean Energy, 2014, 2(2): 104-113.
- [9] 叶林, 陈政, 赵永宁, 等. 基于遗传算法-模糊径向基神经网络的光伏发电功率预测模型[J]. 电力系统自动化, 2015, 39(16): 16-22.
YE Lin, CHEN Zheng, ZHAO Yongning, et al. Photovoltaic power forecasting model based on genetic algorithm and fuzzy radial basis function neural network[J]. Automation of Electric Power Systems, 2015, 39(16): 16-22.
- [10] 许童羽, 马艺铭, 曹英丽, 等. 基于主成分分析和遗传优化 BP 神经网络的光伏输出功率短期预测[J]. 电力系统保护与控制, 2016, 44(22): 90-95.
XU Tongyu, MA Yiming, CAO Yingli, et al. Short term forecasting of photovoltaic output power based on principal component analysis and genetic optimization of BP neural network[J]. Power System Protection and Control, 2016, 44(22): 90-95.
- [11] MA Ailong, ZHONG Yanfei, ZHANG Liangpei. Adaptive multiplicative memetic fuzzy clustering algorithm for remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(8): 4202-4217.
- [12] CUEVAS E, ZALDIVA D C M. A swarm optimization algorithm inspired in the behavior of the social spider[J]. Expert System with Applications, 2013, 40(16): 6374-6384.
- [13] 朴尚哲, 超木日力格, 于剑. 模糊 C 均值算法的聚类有效性评价[J]. 模式识别与人工智能, 2015, 28(5): 452-461.
PIAO Shangzhe, CHAOMU Rilige, YU Jian. Clustering validity evaluation of fuzzy C mean algorithm[J]. Pattern Recognition and Artificial Intelligence, 2015, 28(5): 452-461.
- [14] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer, 1995.
- [15] 殷豪, 董朕, 陈云龙. 基于 CEEMD 和膜计算优化支持向量机的风速预测[J]. 电力系统保护与控制, 2017, 45(21): 27-34.
YIN Hao, DONG Zhen, CHEN Yunlong. Wind speed forecasting based on complementary ensemble empirical mode decomposition and support vector regression optimized by membrane computing optimization[J]. Power System Protection and Control, 2017, 45(21): 27-34.
- [16] LIU T, CERPA A E. Foresee (4C): wireless link prediction using link features[C] // Proc of the 10th International Conference on Information Processing in Sensor Networks, Piscataway NJ: IEEE, 2011: 294-305.
- [17] CHENG M Y, HOANG N D, ROY A F V, et al. A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion[J]. Engineering Application of Artificial Intelligence, 2012, 25(4): 744-752.
- [18] ERIK C, CIENFUEGOS M. A new algorithm inspired in the behavior of social-spider for constrained optimization[J]. Expert System with Applications, 2014, 41(1): 412-425.
- [19] LOHOKARE M R, PATTNAIK S S, PRANIGRAHI B K, et al. Accelerated biogeography-based optimization with neighborhood search for optimization[J]. Applied Soft Computing, 2013, 13(5): 2318-2342.
- [20] CAO J, LI K, STOJMENOVIC I. Optimal power allocation and load distribution for multiple heterogeneous multicore server processors across clouds and data centers[J]. IEEE Transactions on Computers, 2014, 63(1): 45-58.

收稿日期: 2017-12-07; 修回日期: 2018-03-21

作者简介:

黄予春(1971—), 女, 双学位, 高级工程师, 主要研究方向为二次及继电保护; E-mail: 842669898@qq.com

曹成涛(1981—), 男, 博士, 副教授, 研究方向为智能控制技术等;

顾海(1970—), 男, 博士, 副教授, 研究方向为电力系统控制等。

(编辑 许威)