

融合聚类及随机配置网络的短期光伏功率预测*

韩莹 朱宏宇 李琨

(辽宁工程技术大学电气与控制工程学院 葫芦岛 125105)

摘要:为了降低天气因素对光伏发电功率的输出值预测精度的影响,从聚类分析和信号分解两方面入手,提出了一种融合聚类算法(KDGMM),改进的变分模态分解(VMD)与随机配置网络(SCN)的预测模型。首先通过 KDGMM 聚类将气象数据划分成晴天、阴天和雨天,针对阴天难以准确预测的问题,采用灰色关联度分析(GRA)选择相似日,其次引入莱维飞行北方苍鹰优化算法(LNGO)优化 VMD 得到最优参数,从而降低阴天光伏功率的非平稳性。最后构建 SCN 预测模型对光伏功率数据进行预测,输出其预测结果。通过实验分析,所提方法的均方根误差(RMSE)和平均绝对百分比误差(MAPE)仅为 1.44 和 1.3%,拟合优度指标 R^2 高达 0.99,与其他预测方法相比,本文所提方法有较高的预测精度。

关键词: 光伏功率预测;融合聚类;北方苍鹰优化算法;莱维飞行策略;变分模态分解;随机配置网络

中图分类号: TM615

文献标识码: A

国家标准学科分类代码: 470.40

Short-term PV power prediction by fusion of clustering and SCN

Han Ying Zhu Hongyu Li Kun

(Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: In order to reduce the influence of weather factors on the prediction accuracy of the output value of photovoltaic power generation, it is proposed a prediction model incorporating the clustering algorithm (KDGMM), the improved variational modal decomposition (VMD) and the stochastic configuration network (SCN), starting from both cluster analysis and signal decomposition. Firstly, the meteorological data are classified into sunny, cloudy and rainy days by KDGMM clustering, and for the problem that it is difficult to predict accurately on cloudy days, gray correlation analysis (GRA) is used to select similar days, and secondly, the Lévy northern goshawk optimization (LNGO) algorithm is introduced to optimize VMD to get the optimal parameters, so as to reduce the non-smoothness of PV power on cloudy days. Finally, the SCN prediction model is constructed to predict the PV power data and output its prediction results. Through experimental analysis, the root mean square error (RMSE) and the mean absolute percentage error (MAPE) of the proposed method are only 1.44 and 1.3%, and the R^2 index for goodness of fit is as high as 0.99. Compared with other prediction methods, the proposed method has higher prediction accuracy.

Keywords: photovoltaic power prediction; integrated clustering; northern gohawk optimization algorithm; Lévy flight strategy; variational modal decomposition; stochastic configuration network

0 引言

由于光伏发电具有一定的不平稳性和非连续性,为了使它的预测精度得到提升,国内外学者将光伏发电预测方法分为线性预测法^[1]、非线性预测法^[2]。前者在处

理随机变化的数据往往不准确,只能适合于天气比较稳定情况下的光伏发电功率预测^[3];后者典型的如基于支持向量机(support vector machine, SVM)模型^[4]的光伏功率预测。

不同天气情况会对光伏功率预测精度产生较大的影响。张雨金等^[5]利用 K-means 聚类算法将气象数据划分

成不同的天气类型,再与 SVM 模型结合,完成对光伏功率的预测。由于传统的 K-means 聚类算法对噪声和异常值比较敏感,文献[6]通过比较 K-medoids 与 K-means 两种聚类算法,证明 K-medoids 能够有效降低对此类数据的敏感程度。此外,Gogebakan^[7]提出了高斯混合模型(gaussian mixture model,GMM)聚类算法来估计聚类的数目。然而,只采用单一聚类方法很容易忽略算法本身的不确定性,造成分类效果较差。武新章等^[8]将集成聚类应用在光伏功率预测中,实验结果表明采用多种聚类集成的方法明显好过单一聚类算法,光伏发电功率的预测精度得到了明显的提升。

光伏数据自身具有一定的波动性,和风电功率相似,对其进行预测前都需要采用信号分解实现数据处理。Zhang 等^[9]利用经验模态分解(empirical mode decomposition,EMD)算法使风电功率数据的稳定性得到了加强,有效地提高了其预测精度。He 等^[10]采用集成经验模态分解(ensemble empirical mode decomposition,EEMD)将复杂的、不规则的风电功率时间序列转化为一组相对易于分析的子序列,然而 EEMD 没有从本质上解决 EMD 的模式混叠问题。完全集成经验模态分解(complete ensemble empirical mode decomposition,CEEMDAN)是一种 EEMD 的改进算法,能够降低白噪声的影响。王贺等^[11]将 CEEMDAN 应用在风速短期预测建模中,取得了较好的预测效果。变分模态分解^[12](variational mode decomposition,VMD)是一种完全非递归分解方法,其分解效果优于 EEMD 等方法^[13]。但是 VMD 方法不能自动调节其参数,需预先进行人工设置模态分解层数和其他关键参数。

通过上述分析,本文提出基于 KDGMM-LNGO-VMD-SCN 的短期光伏功率预测模型。与单一的 K-medoids、GMM 等聚类算法相比,采用融合聚类算法(K-medoids-Gaussian mixture model,KDGMM)能够更准确地划分天气类型。针对阴天光伏功率预测精度较低的问题,通过灰色关联度分析(grey relational analysis,GRA)选出与预测日气象特征相似的日期作为相似日。再采用莱维飞行北方苍鹰优化算法(Lévy northern goshawk optimization,LNGO)优化 VMD 参数的分解方法对光伏功率序列进行平稳化处理。最后将处理后的光伏功率数据代入随机配置网络(stochastic configuration networks,SCN)预测模型中,得出光伏发电功率的预测值。将本文所提出方法分别与反向传播(back propagation,BP)神经网络^[14]、回声状态网络^[15](echo state network,ESN)和径向基函数^[16](radical basis function,RBF)神经网络等进行比较,仿真实验证明本文所提出方法的预测精度更高。

1 数据处理

1.1 筛选气象特征

本文选取宁夏回族自治区某光伏电站实际的数据,包含了光伏功率和太阳辐射、相对湿度等气象特征。通过计算皮尔逊相关系数^[17]对气象特征进行筛选,比较各气象特征与光伏功率之间的相关性强弱,将相关性较强的一些气象特征保留。

对于一个 m 维线性连续变量,皮尔逊相关系数的计算公式为:

$$r = \frac{\sum_{i=1}^m (x_i - \bar{x})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (t_i - \bar{t})^2}}$$

(1)

式中: m 为某一气象特征数据的个数; x_i 为第 i 个气象特征,如太阳辐射强度、云层不透明度、相对湿度等; t 为光伏发电功率; $\bar{x}$ 和 $\bar{t}$ 分别表示 2 个变量的均值。光伏发电功率与气象特征的关系如表 1 所示。

表 1 皮尔逊相关系数的分布情况

Table 1 Distribution of Pearson correlation coefficients		
序号	气象特征	皮尔逊相关系数
1	方位角	-0.04
2	露点温度	0.08
3	云层不透明度	-0.26
4	大气可降水量	0.07
5	相对湿度	-0.40
6	天顶角	-0.82
7	太阳总水平辐射	0.99
8	降雪深度	-0.09
9	地面气压	-0.05
10	高度 10 m 风向	0.05
11	高度 10 m 风速	0.15

皮尔逊相关系数的值在 $[-1,1]$ 之间浮动,正值表示该气象特征与光伏发电功率呈正相关,反之为负相关,二者之间的相关性的强弱取决于对应的绝对值大小,绝对值越大表示其相关性就越强^[18]。因此,本文筛选皮尔逊相关系数绝对值较大的气象特征。由表 1 可以清晰地看出 6 号和 7 号气象特征的相关性较大,3 号和 5 号气象特征次之。因此,本文最终保留了太阳总水平辐射、云层不透明度、相对湿度和天顶角 4 种气象特征,为接下划分天气类型提供了聚类样本。

1.2 融合聚类算法

KDGMM 聚类是一种基于 K-medoids 融合的 GMM 聚类算法。由于 K-medoids 算法的聚类数目 K 值需要人工预先设定,为了合理地确定聚类数目 K 值的大小,本文选用轮廓系数(silhouette coefficient, SC)、CH(Calinski-

Harabasz) 指数和 DB (Davies-Bouldin) 指数作为评估聚类好坏的指标,通过计算这些评估指标来确定其 K 值的大小,再将由 K-medoids 算法得到的聚类中心点作为 GMM 聚类的初始中心,能够克服 GMM 聚类对初始聚类中心敏感的缺陷。采用 KDGMM 算法对经过皮尔逊相关系数筛选出的 4 种气象数据进行分类,进而划分出天气类型。计算公式如下。

$$\begin{cases} s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \\ SC = \frac{\sum_{i=1}^N s(i)}{N} \end{cases} \quad (2)$$

$$CH = \frac{tr(B_k)(N - K)}{tr(W_k)(K - 1)} \quad (3)$$

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j, i, j \in [1, k]} \frac{\overline{S_i} + \overline{S_j}}{d_{ij}} \quad (4)$$

比较 K-means、K-medoids、GMM、KDGMM 聚类相应的评价指标如表 2~4 所示。

表 2 K 值对应的轮廓系数

Table 2 Silhouette coefficients corresponding to K values

聚类算法	$K=2$	$K=3$	$K=4$	$K=5$
K-means	0.52	0.54	0.46	0.48
K-medoids	0.54	0.64	0.41	0.46
GMM	0.41	0.62	0.21	0.37
KDGMM	0.68	0.73	0.53	0.62

通过表 2~4 可以看到,当聚类数目 $K=3$ 时,上述 4 种聚类算法的各项评估指标更为优异;并且与 K-means、K-medoids、GMM 3 种聚类算法相比,本文提出的 KDGMM 算法的聚类效果最好。

$$\varepsilon_q(u) = \frac{\min q \min u \mid Y_q(u) - Y_o(u) \mid + \rho \max q \max u \mid Y_q(u) - Y_o(u) \mid}{\mid Y_q(u) - Y_o(u) \mid + \rho \max q \max u \mid Y_q(u) - Y_o(u) \mid} \quad (5)$$

式中: $Y_o(u)$ 、 $Y_q(u)$ 分别为预测日与第 q 个历史日的第 u 个气象因素特征向量, ρ 通常取 0.5。

预测日与第 q 个历史日的灰色关联度 r_q 的计算公式为:

$$r_q = \frac{1}{m} \sum_{k=1}^m \varepsilon_q(u) \quad (6)$$

预测时,优先计算与预测日相邻近的历史日和预测日之间的灰色关联度,最终选取 i 个历史发电日 ($r_i \geq 0.8$) 作为阴天的相似日样本。

2 光伏发电功率预测模型

2.1 变分模态分解

由于阴天时对应的光伏发电功率具有较明显的波动

表 3 K 值对应的 DB 指数

Table 3 DB index corresponding to K values

聚类算法	$K=2$	$K=3$	$K=4$	$K=5$
K-means	0.76	0.68	1.11	0.86
K-medoids	0.74	0.63	0.89	0.78
GMM	1.22	0.79	1.83	1.37
KDGMM	0.65	0.61	0.87	0.71

表 4 K 值对应的 CH 指数

Table 4 CH index corresponding to K values

聚类算法	$K=2$	$K=3$	$K=4$	$K=5$
K-means	781.41	801.43	692.61	745.86
K-medoids	912.35	1 215.34	760.11	844.03
GMM	779.78	1 018.91	480.31	692.61
KDGMM	2 404.73	2 793.42	2 506.88	2 721.22

通过 KDGMM 聚类将气象数据分成 3 种样本并结合当地气候完成对天气类型的划分。其中,第 1 类样本,云层不透明度较低,相对湿度较低,太阳辐照强度较高,天顶角较小,此类样本对应的天气为晴天;第 2 类样本,云层不透明度较高,相对湿度适中,太阳辐照强度适中,天顶角适中,为阴天天气;第 3 类样本,云层不透明度适中,相对湿度较高,太阳辐照强度较低,天顶角较大,为雨天天气。

1.3 灰色关联度分析

由于在阴天条件下,气象因素容易发生突变,使光伏功率预测难度增大,本文采用相似日的选择以及数据平滑处理等方法来提高预测精度。通过灰色关联度分析^[19]来选择相似日。

各历史发电日与待预测日的关联系数的计算公式如下:

性,需要对其进行平滑处理。本文使用 VMD^[20] 对阴天相似日的光伏功率信号 $f(t)$ 进行分解,将其分为不同的本征模函数 (intrinsic mode function, IMF) 分量 v_s , 和它们的中心频率 ω_s 。

VMD 算法有两个参数需要人为提前设置,它们分别是分解层数 S 和二次罚因子 α 。如果 S 的取值较小,会出现模态混叠现象;若 S 取值设置偏大,会发生过度分解并产生虚假信号。当 α 较小时,会造成信号缺少;当 α 较大时,也会导致信号混叠,影响后面的预测精度。因此,如何确定 VMD 的分解参数成为阴天相似日样本信号分解的关键。

2.2 改进的北方苍鹰优化算法

1) 北方苍鹰优化算法

北方苍鹰优化算法 (northern goshawk optimization, NGO) 是由 Dehghani 等^[21]于 2021 年提出, 该算法是基于北方苍鹰的狩猎行为而衍生出来的。具体过程如下:

(1) 第 1 阶段为全局搜索。北方苍鹰在该阶段随机攻击一只猎物, 此行为是用来找到最优区域。该阶段可由下面的数学公式来表达:

$$\begin{cases} P_i = A_k, i = 1, 2, \dots, N \\ k = 1, 2, \dots, i - 1, i + 1, N \end{cases} \quad (7)$$

$$\begin{cases} a_{i,j}^{new,p_1} = a_{i,j} + q(p_{i,j} - Ea_{i,j}), O_{p_i} < O_i \\ a_{i,j}^{new,p_1} = a_{i,j} + q(p_{i,j} - Ep_{i,j}), O_{p_i} \geq O_i \end{cases} \quad (8)$$

$$\begin{cases} A_i = A_i^{new,p_2}, O_i^{new,p_1} < O_i \\ A_i = A_i, O_i^{new,p_1} \geq O_i \end{cases} \quad (9)$$

式中: P_i 为第 i 只苍鹰要捕捉的猎物位置; O_{p_i} 为目标函数值; k 是在 $[1, N]$ 内且不为 i 的随机整数。 $a_{i,j}^{new,p_1}$ 为第 i 只苍鹰的猎物在第 j 维度的值, $O_{i,j}^{new,p_1}$ 为对应的目标函数值; q 和 E 是用来生成随机的 NGO 行为的随机参数, 设定 $q \in [0, 1]$, E 的大小通常等于 1 或 2。

(2) 第 2 阶段为局部搜索。猎物在受到北方苍鹰攻击后, 会尝试逃亡, 然而, 猎物在任何情况下都能被北方苍鹰并追捕到。具体计算公式如下:

$$a_{i,j}^{new,p_2} = a_{i,j} + R(2r - 1)a_{i,j} \quad (10)$$

$$R = 0.02 \left(1 - \frac{t}{T} \right) \quad (11)$$

$$\begin{cases} A_i = A_i^{new,p_2}, O_i^{new,p_2} < O_i \\ A_i = A_i, O_i^{new,p_2} \geq O_i \end{cases} \quad (12)$$

式中: t 为当前迭代次数, T 则表示当前迭代次数和最大迭代次数, R 为攻击半径; $a_{i,j}^{new,p_2}$ 是在该阶段更新后第 i 只苍鹰在第 j 维度的位值, $O_{i,j}^{new,p_2}$ 表示相应的目标函数值。

2) LNGO 优化算法

为了解决 NGO 算法迭代寻优时容易陷入局部极值的问题, 本文引入莱维 (Lévy) 飞行策略^[22]来增强 NGO 算法的全局搜索能力。Lévy 飞行表达式如下:

$$\sigma = \left[\frac{\Gamma(1 + \beta) \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1 + \beta}{2}\right) \cdot \beta \cdot 2^{(\beta-1)/2}} \right] \quad (13)$$

$$levy = 0.01 \cdot \frac{r_1 \cdot \sigma}{|r_2|^{\frac{1}{\beta}}} \quad (14)$$

式中: $levy$ 为飞行步长, β 为常数, 通常取 1.5, r_1, r_2 为 0~1 的随机数。

本文采用改进的北方苍鹰优化算法 (LNGO) 是将

Lévy 飞行策略应用在 NGO 算法追击逃亡阶段, 融合 Lévy 飞行策略后的表达式如下:

$$a_{i,j}^{new} = levy \cdot a_{i,j} + R(2r - 1)a_{i,j} \quad (15)$$

$$R = 0.02 \left(1 - \frac{t}{T} \right) \quad (16)$$

2.3 基于 LNGO 算法的 VMD 参数优化

由于采用人工预先设定 VMD 参数的方式很容易造成过分解不合理的问题, 进而影响光伏功率的预测精度。故本文采用改进的北方苍鹰优化算法对光伏功率信号进行 VMD 分解以达到最佳的分解效果。并选择包络熵极小值作为适应度函数, VMD 分解层数 S 和二次罚因子 α 的寻优范围^[23]分别取为 $S \in [2, 15]$, $\alpha \in [200, 3\ 000]$, 其优化流程如图 1 所示。

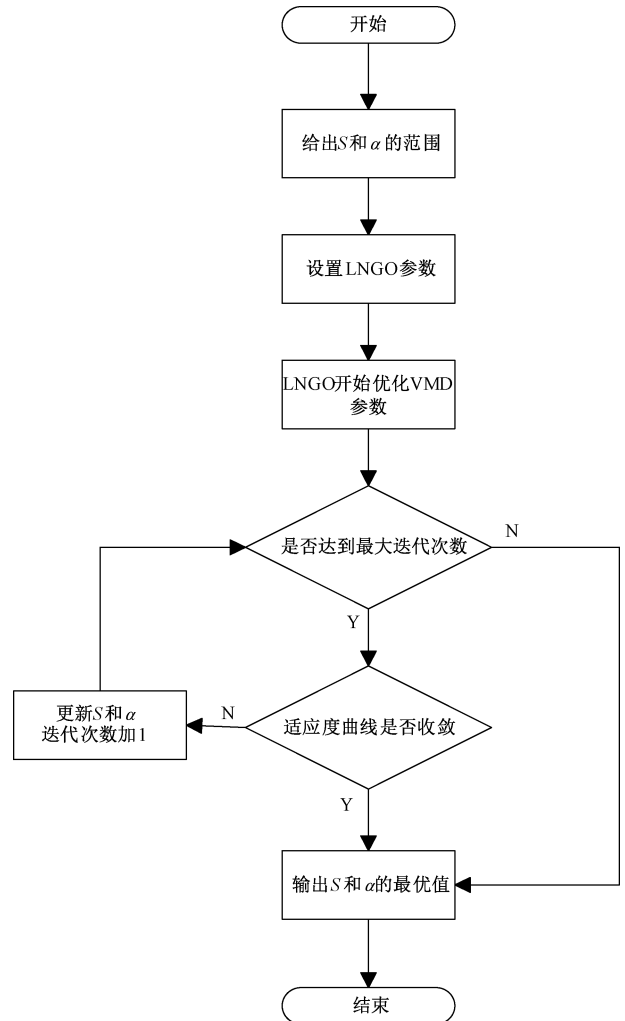


图 1 LNGO 优化 VMD 参数流程

Fig. 1 LNGO optimized the VMD parameters process

2.4 随机配置网络

SCN^[24]是一种增量式的单隐藏层前馈神经网络, 根据训练效果, 该网络可以自适应生成其隐藏层结构。在

构造网络的过程中,隐含层节点参数随机在一个动态范围内完成取值,该网络的输出权值一般是采用全局最小二乘法对其求解。SCN 模型结构如图 2 所示。

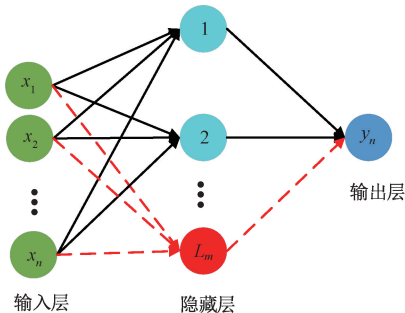


图 2 随机配置网络模型

Fig. 2 Network model of SCN

针对提供的目标函数 $f: R^d \rightarrow R^m$, 假设开始确定了一个带有 $L-1$ 个隐含层节点的 SCNs 模型, 其基本的映射关系为:

$$f_{L-1} = \sum_{j=1}^m \beta_j \varphi_j(\omega_j^T U + \theta_j), L = 1, 2, \dots, L_m \quad (17)$$

式中: $f_0 = 0$; $\beta_j = [\beta_{j,1}, \beta_{j,2}, \dots, \beta_{j,m}]^T$ 为第 j 个隐藏层节点输出的权重矩阵; $U = [u_1, u_2, \dots, u_i]^T$ 代表其对应的输入特征矩阵; ω_j 是第 j 个隐藏层节点输入的权重, θ_j 为相应的阈值矩阵; $\varphi(\cdot)$ 表示 Sigmoid 激活函数。

当模型开始进行计算时, 计算输出值与真实值的差值。当误差无法达到设定值的要求时, 随机生成 φ_L 和 θ_φ , 并计算更新输出权值 β_L , 隐藏层节点数加 1, 得到模型新的输出为 $f = f_{L-1} + \beta_L \varphi_L$, 目的是为了实现残差的修正, 使得 $\lim_{L \rightarrow \infty} \|f - f_L\| = 0$ 。

2.5 预测模型的构建

本文采用 KDGMM-LNGO-VMD-SCN 的光伏发电功率短期预测模型, 首先计算皮尔逊相关系数对气象特征进行筛选, 其次采用 KDGMM 对筛选后的气象数据进行聚类分析, 划分出晴天、阴天和雨天 3 类, 针对阴天情况, 通过灰色关联度分析选取相似日, 采用 LNGO-VMD 分解方式来降低原始光伏发电功率序列的波动性, 最后选择 SCN 网络模型对分解后的数据进行预测并叠加, 输出预测结果。光伏发电功率预测流程如图 3 所示。

3 算例分析

3.1 实验仿真

本文选取宁夏自治区某光伏电站的数据, 其中, 光伏发电数据为 2019 年 8~10 月的数据, 由于在深夜光伏发电系统基本不工作, 没有功率输出。因此, 选取的时间段为 9:30~19:30, 其功率采样间隔为 15 min。预测日为 8

月 28 日(晴天)、9 月 21 日(阴天)、10 月 22 日(雨天)。

为了更准确地区分不同天气类型, 经过计算皮尔逊相关系数筛选出前面章节介绍的 4 种气象因素, 选择该地区 2019 年对应的气象数据作为聚类样本, 采用 KDGMM 聚类将其分成 3 种样本进而完成对天气类型的划分。聚类结果如图 4 所示。

图 4 为采用 KDGMM 聚类划分 3 种典型天气的聚类结果, 由于选用的是 4 种气象数据, 无法在同一个仿真图中实现 4 维可视化。将 4 种气象数据以排列组合的方式 ($C_4^3 = 4$) 得到 4 个仿真图来展示聚类结果。

晴天和雨天的光伏功率训练集选择该月相同天气条件下的历史发电日。在阴天情况下, 通过计算历史发电日与预测日气象特征之间的灰色关联度来选取相似日, 得到训练样本, 利用 LNGO 优化 VMD 得到最优解, 分解层数 $S=6$, 二次惩罚因子 $\alpha=1350$ 。通过 LNGO-VMD 对阴天的相似日样本进行分解处理, 使得到的子序列的平稳性显著增强。

将采集处理后的 370 个光伏功率数据进行 VMD 分解, 将原始信号分解得到 6 个 IMF 分量。经 VMD 分解后的 IMF 分量, 相对于原光伏功率曲线, 其波动幅度和稳定性均较好。为验证改进 NGO 算法优化 VMD 参数的性能, 比较 LNGO 和原始 NGO 的适应度值曲线, 如图 5 所示。

由图 5 可以看出, LNGO-VMD 的适应度最优值出现在第 5 代其值为 0.607, 而 NGO-VMD 的适应度最优值出现在第 8 代其值为 0.594。说明 LNGO 算法的收敛速度更快, 相应的适应度值也更低。

3.2 评估指标

为了全面评估预测方法的测试性能, 本文选择均方根误差 (root mean square error, RMSE)、平均绝对百分比误差 (mean absolute percentage error, MAPE)、拟合优度 (goodness of fit) 指标 R^2 作为衡量光伏功率预测精度的评估指标, 具体公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - T_i)^2} \quad (18)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{t_i - T_i}{T_i} \right| \times 100\% \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (t_i - T_i)^2}{\sum_{i=1}^n (t_i - \bar{T})^2} \quad (20)$$

式中: t_i 为光伏发电功率的实际值, T_i 为光伏发电功率的预测值。拟合优度 R^2 的值越接近 1, 则说明模型的拟合性能越优。

3.3 预测结果分析

为了选择合适的网络模型完成光伏发电功率的预

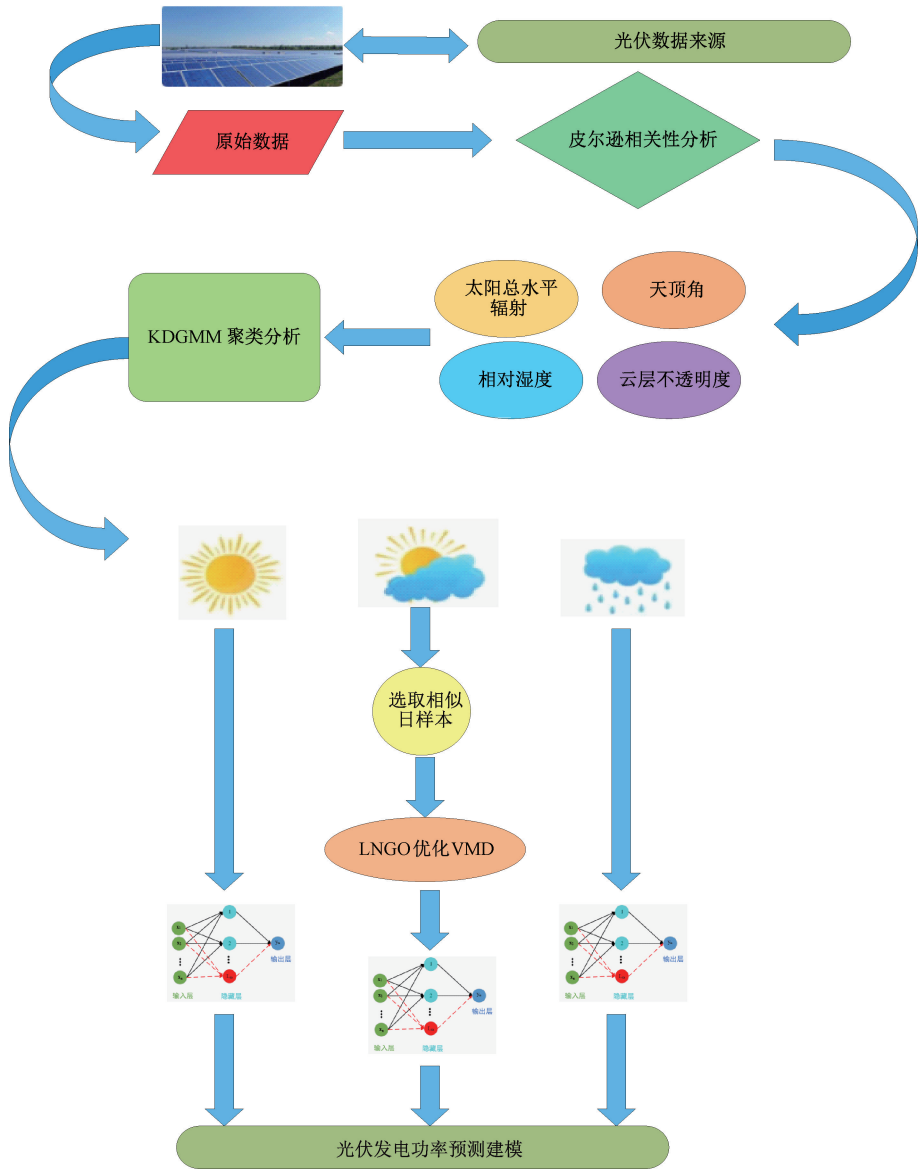


图 3 光伏发电功率预测流程

Fig. 3 Photovoltaic power prediction process

测,本文选择 BP、RBF、ESN 和 SCN 这 4 种网络模型来对比其预测性能。通过预测 8~10 月某一天的光伏发电功率,用来比较不同模型的预测性能,分别计算不同模型的 RMSE、MAPE、 R^2 ,预测结果如图 6 所示,对应的评估指标如表 5 所示。

表 5 预测模型对应的评估指标

Table 5 Evaluation indexes corresponding to prediction models

预测模型	RMSE	MAPE/%	R^2
BP	14.87	10.5	0.85
RBF	14.19	10.2	0.86
ESN	11.94	9.9	0.87
SCN	10.51	8.4	0.89

从图 6 可以清晰看出,SCN 模型的预测曲线与光伏发电功率实际曲线最为接近。由于 SCN 不需要事先设定权值和偏置,摆脱了神经网络容易陷入局部最优的不足,并且 SCN 的预测效果明显优于 BP 和 RBF。然而,由表 5 可知,相较于 ESN,SCN 模型的 RMSE 减少了 1.43, MAPE 下降了 1.5%,也就是说 ESN 的预测精度与 SCN 网络很接近,此外,采用单独的神经网络模型对光伏功率进行预测。因此需要在不同天气情况下对 SCN 与 ESN 这两种模型做进一步分析。

本文采用 KDGMM 聚类算法来划分天气类型,相同天气类型下,将对应的光伏发电功率数据代入 SCN 预测模型中,通过对比 KDGMM-SCN 与 GMM-SCN、K-means-

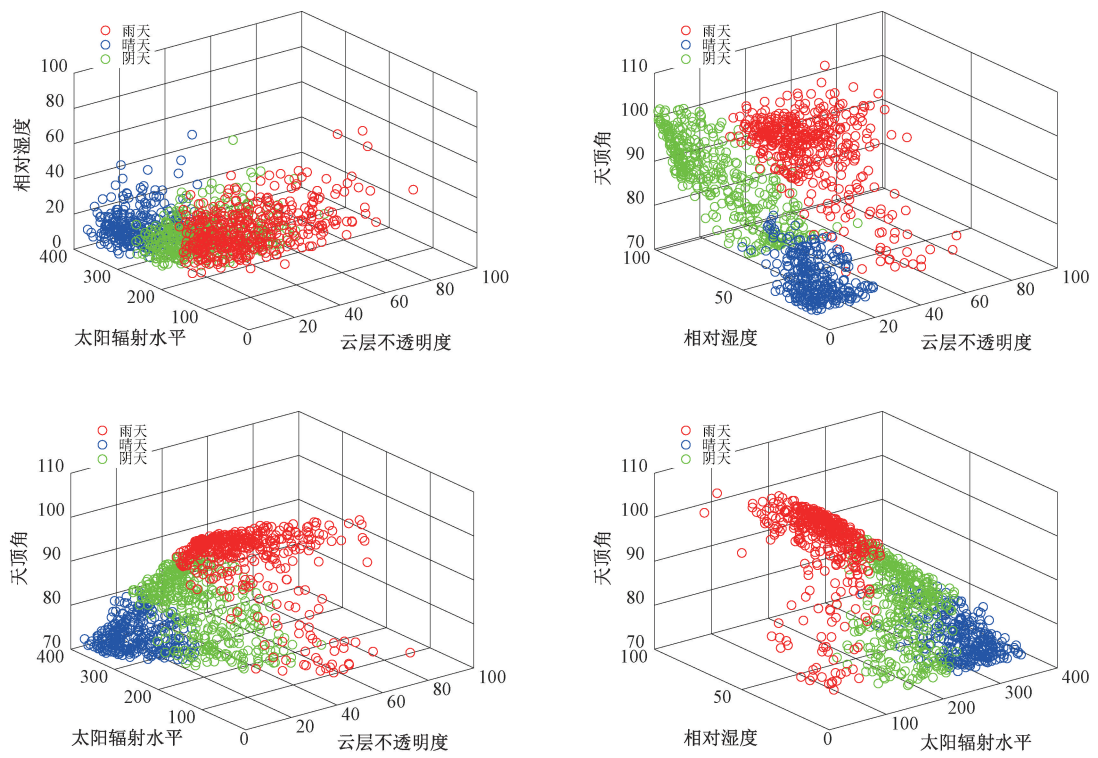


图 4 KDGM 聚类算法划分的天气类型

Fig. 4 Weather types divided by KDGM clustering algorithm

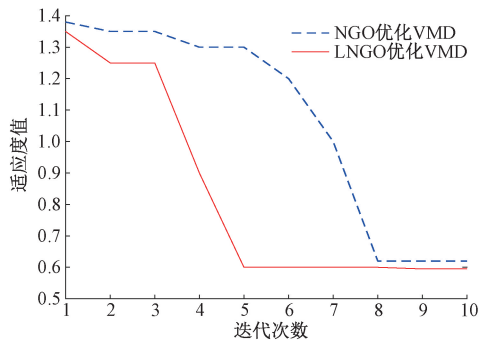


图 5 LINGO-VMD 与 NGO-VMD 参数寻优对比

Fig. 5 Parameters optimization between LINGO-VMD and NGO-VMD

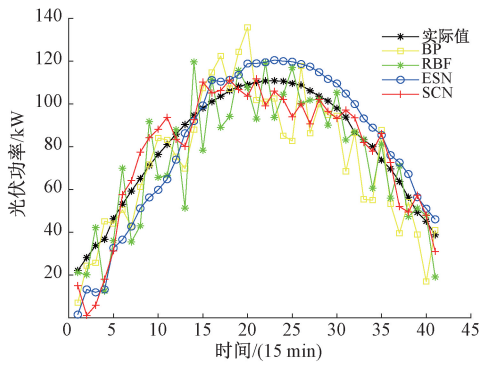


图 6 不同网络模型的预测结果对比

Fig. 6 Prediction results of different network models

SCN、K-medoids-SCN、KDGM-ESN 4 种预测方法的性能差异。预测结果如图 7~9 所示, 评估指标如表 6~8 所示。

表 6 晴天评估指标对比

预测方法	RMSE	MAPE/%	R^2
K-means-SCN	6.92	5.1	0.93
GMM-SCN	8.11	7.5	0.91
K-medoids-SCN	5.41	4.4	0.96
KDGM-ESN	3.24	2.4	0.98
KDGM-SCN	1.89	1.4	0.99

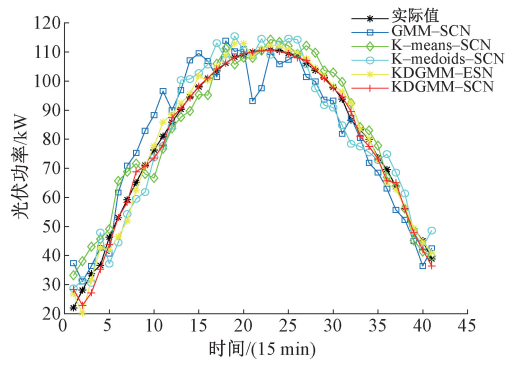


图 7 晴天预测结果

Fig. 7 Sunny weather prediction results

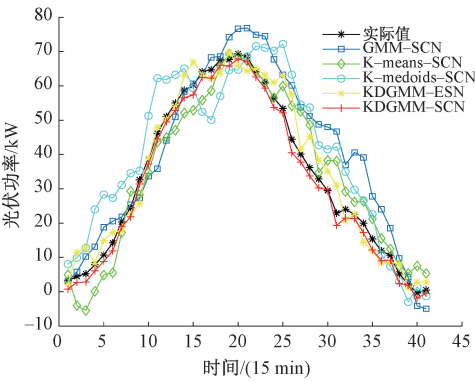


图 8 雨天预测结果
Fig. 8 Rainy weather prediction results

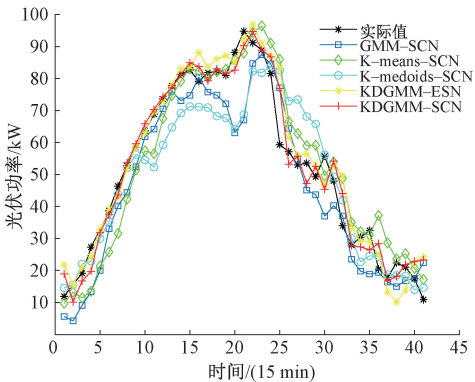


图 9 阴天预测结果
Fig. 9 Cloudy weather prediction results

表 7 雨天评估指标对比
Table 7 Comparison of rainy weather evaluation indexes

预测方法(雨天)	RMSE	MAPE/%	R ²
K-means-SCN	10.88	12.8	0.87
GMM-SCN	10.17	15.1	0.78
K-medoids-SCN	7.55	11.3	0.89
KDGMM-ESN	5.71	9.2	0.94
KDGMM-SCN	2.14	1.8	0.99

表 8 阴天评估指标对比
Table 8 Comparison of cloudy weather evaluation indexes

预测方法(阴天)	RMSE	MAPE/%	R ²
K-means-SCN	11.98	15.4	0.78
GMM-SCN	9.91	12.7	0.85
K-medoids-SCN	8.89	10.2	0.88
KDGMM-ESN	6.66	6.7	0.93
KDGMM-SCN	5.79	5.8	0.95

通过上述图表可以看出,在晴天条件下,采用 KDGM-SCN 模型对比其他 3 种预测模型,预测精度得

到进一步提高。MAPE 只有 1.3%,拟合优度 R^2 几乎为 1,在雨天条件下,采用 KDGM-SCN 模型预测光伏发电功率极大地提高了预测精度,这是因为采用 KDGM 融合聚类算法克服了聚类在区分雨天和阴天时不准确,导致其预测精度较低的缺点。并且,该预测精度基本符合实际要求。尽管 KDGM-SCN 模型相较于 GMM-SCN、K-medoids-SCN 和 KDGM-ESN 这 3 种预测方法得到的预测误差小, R^2 更接近于 1。但是在阴天情况下,该模型的 RMSE 为 5.79,MAPE 为 5.8%,对比晴天和雨天的预测误差还是比较大。因为阴天存在一些云层遮挡,照在地面由光伏电站接受的太阳辐射量会减少,气象因素容易发生突变,进而影响光伏发电功率的输出值。

为了降低阴天条件下光伏发电的波动性带来的不利影响,在选取相似日样本的基础上,本文采用 VMD 分解来降低功率信号的非平稳性。将方法 a(KDGM-NGO-VMD-SCN)与方法 b(KDGM-SCN)方法 c(KDGM-EMD-SCN)、方法 d(KDGM-EEMD-SCN)、方法 e(KDGM-VMD-SCN)进行比较。不同分解状态下的预测结果如图 10 所示,相应的评估指标如表 9 所示。

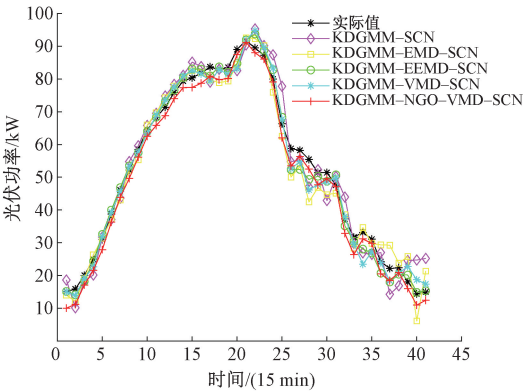


图 10 不同信号分解下的预测结果
Fig. 10 Prediction results under different signal decomposition

表 9 不同信号分解状态下预测的评估指标
Table 9 Evaluation indexes of prediction under different signal decomposition states

预测方法	RMSE	MAPE/%	R ²
方法 a	5.82	5.8	0.95
方法 b	5.13	5.1	0.96
方法 c	4.66	4.3	0.97
方法 d	3.56	3.8	0.98
方法 e	3.11	3.4	0.98

从表 9 可以看出方法 e(KDGM-VMD-SCN)的预测效果优于其他 4 种预测方法,这是因为采用 VMD 分解克服 EMD 分解导致模式混叠的问题和 EEMD 分解产生的端点效应问题,有效地降低了阴天光伏数据的不平稳性,

提高了预测精度。

为了更全面地比较在阴天条件下不同预测方法的预测性能,将本文方法(KDGMM-LNGO-VMD-SCN)与方法 1(KDGMM-LNGO-VMD-SCN)、方法 2(KDGMM-LNGO-VMD-BP)、方法 3(KDGMM-LNGO-VMD-RBF)、方法 4(KDGMM-LNGO-VMD-ESN)进行对比,得到阴天的最终预测结果如图 11 所示。

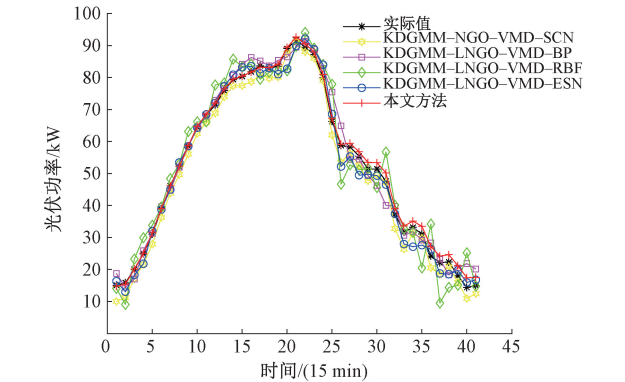


图 11 阴天的最终预测结果

Fig. 11 Overcast final prediction results of cloudy weather

将 5 种预测方法分别重复运行 20 次,记录最优值、最差值、平均值和标准差,如表 10~13 所示。

表 10 上述预测模型运行 20 次评估指标的最优值
Table 10 The above predicted models run 20 times for the best value of evaluation index

预测模型	RMSE	MAPE/%	R ²
方法 1	3. 11	3. 4	0. 98
方法 2	6. 75	7. 3	0. 92
方法 3	6. 16	6. 6	0. 93
方法 4	4. 83	4. 7	0. 96
本文方法	1. 44	1. 3	0. 99

表 11 上述预测模型运行 20 次评估指标的最差值
Table 11 The above predicted models run 20 times for the worst value of evaluation index

预测模型	RMSE	MAPE/%	R ²
方法 1	7. 18	5. 9	0. 93
方法 2	12. 63	13. 6	0. 76
方法 3	8. 92	10. 5	0. 88
方法 4	6. 12	6. 5	0. 94
本文方法	4. 66	3. 9	0. 97

由表 10~13 可知,无论是对比最优值、最差值还是平均值、标准差,本文所采用的预测方法在阴天条件下光伏发电的预测精度最高,这是因为采用 VMD 分解可以克服 EMD 分解导致模态混叠的问题和 EEMD 分解产生的

端点效应问题。此外,LNGO-VMD 克服了 VMD 人工确定模态分解数目的不确定性,相较于 BP、RBF 和 ESN 几种预测模型,SCN 模型的学习效率更高,预测效果更优。

表 12 上述预测模型运行 20 次评估指标的平均值
Table 12 The above predicted model run 20 times for average value of evaluation index

预测模型	RMSE	MAPE/%	R ²
方法 1	4. 29	4. 5	0. 97
方法 2	10. 89	13. 3	0. 79
方法 3	6. 45	6. 4	0. 94
方法 4	5. 24	5. 5	0. 95
本文方法	3. 39	3. 7	0. 98

表 13 上述预测模型运行 20 次评估指标的标准差
Table 13 The above predicted models run 20 times for standard deviation of evaluation index

预测模型	RMSE	MAPE/%	R ²
方法 1	0. 432 5	0. 004 3	0. 011 6
方法 2	1. 869 9	0. 021 4	0. 458 8
方法 3	0. 772 9	0. 008 5	0. 021 3
方法 4	0. 679 5	0. 006 4	0. 017 2
本文方法	0. 346 1	0. 003 8	0. 003 8

为了进一步验证本文方法的预测性能,将其与参考文献中的 K-means-SVM,EMD-LSTM 两种预测方法在晴天(8 月 28 日)、雨天(9 月 21 日)、阴天(10 月 22 日)3 种不同天气条件下进行对比,预测结果如图 12~14 所示,对应的评价指标如表 14~16 所示。

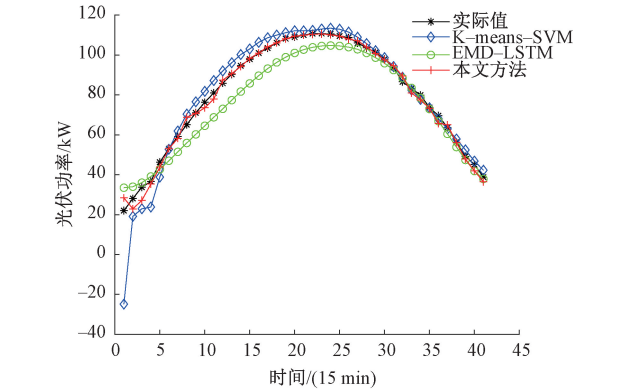


图 12 8 月 28 日预测结果

Fig. 12 Predict result on August 28th

表 14 8 月 28 日预测评估指标
Table 14 Evaluation index on August 28th

预测模型	RMSE	MAPE/%	R ²
EMD-LSTM	7. 32	5. 5	0. 92
K-means-SVM	4. 64	4. 12	0. 95
KDGMM-SCN	2. 36	1. 9	0. 99

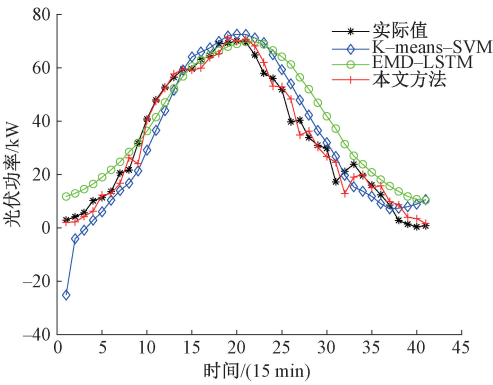


图 13 9 月 21 日预测结果
Fig. 13 Predict result on September 21st

表 15 9 月 21 日预测评估指标
Table 15 Evaluation index on September 21st

预测模型	RMSE	MAPE/%	R ²
EMD-LSTM	9.25	10.7	0.87
K-means-SVM	8.15	11.6	0.89
KDGMM-SCN	2.59	2.3	0.99

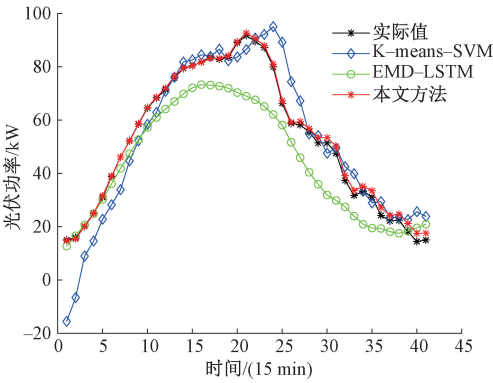


图 14 10 月 22 日预测结果
Fig. 14 Predict result on October 22nd

表 16 10 月 22 日预测评估指标
Table 16 Evaluation index on October 22nd

预测模型	RMSE	MAPE/%	R ²
EMD-LSTM	11.28	12.4	0.81
K-means-SVM	10.44	10.1	0.85
本文方法	1.44	1.3	0.99

综上所述,在晴天和雨天条件下,KDGMM-SCN 的各评估指标均优于 K-means-SVM 和 EMD-LSTM 这两种预测方法。相比晴天,在雨天情况,K-means-SVM 的预测误差明显低于采用 EMD-LSTM,这是因为光伏发电受不同天气条件影响较大,采用 KDGMM 聚类划分天气类型后,提高了预测精度。针对阴天情况下,比较其他两种预测

方法,本文预测方法得到的预测精度更高。

4 结 论

由于采用单独的 K-medoids 和 GMM 聚类算法在对气象数据样本进行天气类型分类时,难以有效区分阴天和雨天天气,因此提出了一种 KDGMM 的融合聚类算法,通过实验表明该方法划分,增强了聚类效果。针对阴天情况光伏发电功率预测精度较低的问题,首先采用灰色关联度分析来选取相似日来实现训练集和测试集的合理划分,然后选择莱维飞行策略的北方苍鹰优化算法优化变分模态分解的分解层数和二次惩罚因子的方法来实现对光伏发电阴天数据进行平滑化处理,最后将经过处理后的功率数据代入随机配置网络模型对其完成短期预测。通过实验仿真,本文所采用方法的预测精度最高,验证了本文提出方法的优越性。

由于随机配置网络的权值和偏置是随机分配的,不需要人为设定,增加了其网络优化难度。因此,对随机配置网络模型进行改进来提高阴天条件下光伏发电功率的预测精度是下一步的重点研究工作。

参考文献

[1] YAO L, HABERMAN S J, ZHANG M. Penalized best linear prediction of true test scores[J]. Psychometrika, 2019, 84(1): 186-211.

[2] RAJ J S, ANANTHI J V. Recurrent neural networks and nonlinear prediction in support vector machines [J]. Journal of Soft Computing Paradigm (JSCP), 2019, 1(1): 33-40.

[3] KUO W C, CHEN C H, CHEN S Y, et al. Deep learning neural networks for short-term PV power forecasting via sky image method[J]. Energies, 2022, 15(13): 4779.

[4] PAN M, LI C, GAO R, et al. Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization[J]. Journal of Cleaner Production, 2020, 277: 123948.

[5] 张雨金, 杨凌帆, 葛双冶, 等. 基于 Kmeans-SVM 的短期光伏发电功率预测[J]. 电力系统保护与控制, 2018, 46(21): 124-130.

ZHAN Y J, YANG L F, GE SH Y, et al. Short-term photovoltaic power prediction based on Kmeans-SVM[J]. Power System Protection and Control, 2018, 46(21): 124-130.

[6] 潘超, 李润宇, 蔡国伟, 等. 考虑风速属性约简聚类的组合预测模型[J]. 电网技术, 2022, 46(4): 1355-1362.

- PAN CH, LI R Y, CAI G W, et al. A combined prediction model considering approximate clustering of wind speed attributes [J]. *Power Grid Technology*, 2022, 46(4): 1355-1362.
- [7] GÖGEBAKAN M. A novel approach for gaussian mixture model clustering based on soft computing method [J]. *IEEE Access*, 2021, 9: 159987-160003.
- [8] 武新章, 王泽宇, 代伟, 等. 基于异质聚类与 Stacking 的双集成光伏发电功率预测 [J]. *电网技术*, 2023, 47(1): 275-284.
- WU X ZH, WANG Z Y, DAI W, et al. Dual-integrated photovoltaic power prediction based on heterogeneous clustering and Stacking [J]. *Power Grid Technology*, 2023, 47(1): 275-284.
- [9] ZHANG F, GUO Z, SUN X, et al. Short-term wind power prediction based on EMD-LSTM combined model [C]. *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, 2020, 514(4): 042003.
- [10] HE Y, WANG Y. Short-term wind power prediction based on EEMD-LASSO-QRNN model [J]. *Applied Soft Computing*, 2021, 105: 107288.
- [11] 王贺, 陈燕峰, 熊敏, 等. 融合 CEEMDAN 和 ICS-LSTM 的短期风速预测建模 [J]. *电子测量与仪器学报*, 2022, 36(4): 17-23.
- WANG H, CHEN H F, XIONG M, et al. Short-term wind speed prediction modeling by integrating CEEMDAN and ICS-LSTM [J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36(4): 17-23.
- [12] 王千军, 李自立. 基于 VMD-WTD 的高频地波雷达电离层杂波抑制 [J]. *国外电子测量技术*, 2022, 41(12): 20-25.
- WANG Q J, LI Z L. High frequency ground wave radar ionospheric clutter suppression based on VMD-WTD [J]. *Foreign Electronic Measurement Technology*, 2022, 41(12): 20-25.
- [13] LIU H, MI X, LI Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM [J]. *Energy Conversion and Management*, 2018, 159: 54-64.
- [14] 丁璨, 李江, 袁召, 等. 基于 NSGA-II 和 BP 神经网络的杯状纵磁触头结构优化设计 [J]. *电工技术学报*, 2022, 37(23): 6074-6082.
- DING C, LI J, YUAN ZH, et al. Optimal design of cup-shaped longitudinal magnetic contact structure based on NSGA-II and BP neural network [J]. *Journal of Electrical Engineering Technology*, 2022, 37(23): 6074-6082.
- [15] RIBEIRO V H A, REYNOSO-MEZA G, SIQUEIRA H V. Multi-objective ensembles of echo state networks and extreme learning machines for streamflow series forecasting [J]. *Engineering Applications of Artificial Intelligence*, 2020, 95: 103910.
- [16] 王亚锋, 安坤, 孟江. 基于 RBF 神经网络的迟滞非线性模型预测控制 [J]. *电子测量技术*, 2021, 44(23): 42-47.
- WANG Y F, AN K, MENG J. Hysteresis nonlinear model predictive control based on RBF neural network [J]. *Electronic Measurement Technology*, 2021, 44(23): 42-47.
- [17] 刘若男, 辛义忠, 李岩. 基于皮尔逊相关系数的动态签名验证方法 [J]. *仪器仪表学报*, 2022, 43(7): 279-287.
- LIU R N, XIN Y ZH, LI Y. A dynamic signature verification method based on Pearson correlation coefficient [J]. *Chinese Journal of Scientific Instrument*, 2022, 43(7): 279-287.
- [18] 孟安波, 陈嘉铭, 黎湛联, 等. 基于相似日理论和 CSO-WGPR 的短期光伏发电功率预测 [J]. *高电压技术*, 2021, 47(4): 1176-1184.
- MENG AN B, CHEN J M, LI ZH L, et al. Short-term photovoltaic power prediction based on similar day theory and CSO-WGPR [J]. *High Voltage Technology*, 2021, 47(4): 1176-1184.
- [19] 侯松宝, 王侃宏, 石凯波, 等. 基于相似日和主成分分析的光伏发电系统短期出力预测 [J]. *可再生能源*, 2018, 36(1): 15-21.
- HOU S B, WANG K H, SHI K B, et al. Short-term output prediction of photovoltaic power generation systems based on similar day and principal component analysis [J]. *Renewable Energy*, 2018, 36(1): 15-21.
- [20] DING J, XIAO D, LI X. Gear fault diagnosis based on genetic mutation particle swarm optimization VMD and probabilistic neural network algorithm [J]. *IEEE Access*, 2020, 8: 18456-18474.
- [21] DEHGhani M, HUBALOVSKY S, TROJOVSKY P. Northern goshawk optimization: A new swarm-based algorithm for solving optimization problems [J]. *IEEE Access*, 2021, 9: 162059-162080.
- [22] PALYULIN V V, BLACKBURN G, LOMHOLT M A, et al. First passage and first hitting times of Lévy flights and Lévy walks [J]. *New Journal of Physics*, 2019, 21(10): 103028.
- [23] WANG D, ZHU L, YUE J, et al. Application of

variational mode decomposition based on particle swarm optimization in pipeline leak detection[J]. Engineering Research Express, 2020, 2(4): 045036.

- [24] WANG D H, LI M. Stochastic configuration networks: Fundamentals and algorithms[J]. IEEE Transactions on Cybernetics, 2017, 47(10): 3466-3479.

作者简介



韩莹, 2005 年于辽宁工程技术大学获得学士学位, 2008 年于辽宁工程技术大学获得硕士学位, 2020 年于东北大学获得博士学位, 现为辽宁工程技术大学副教授, 主要研究方向为时间序列分析与应用、故障预测与健康管理等。

E-mail: hyfengyan@163.com

Han Ying received her B. Sc. degree from Liaoning Technical University in 2005, M. Sc. degree from Liaoning Technical University in 2008 and Ph. D. degree from Northeastern University in 2020, respectively. Now she is an associate professor in Liaoning Technical University. Her main research interests include time series analysis and application and prognostics health management.



朱宏宇, 现为辽宁工程技术大学电气工程专业在读硕士研究生, 主要研究方向为光伏功率预测、智能优化算法等。

E-mail: 1697615087@qq.com

Zhu Hongyu now is a M. Sc. candidate of Electrical Engineering, Liaoning Technical University. His main research interests include PV power prediction and intelligent optimization algorithm.



李琨(通信作者), 2005 年于山东科技大学获得学士学位, 2008 年于辽宁工程技术大学获得硕士学位, 2013 年于东北大学获得博士学位, 现为辽宁工程技术大学教授, 主要研究方向为系统建模、优化与控制、智能巡检技术等。

E-mail: phdlikun@163.com

Li Kun (Corresponding author) received his B. Sc. degree from Shandong University of Science and Technology in 2005, M. Sc. degree from Liaoning Technical University in 2008 and Ph. D. degree from Northeastern University in 2013, respectively. Now he is a professor in Liaoning Technical University. His main research interests include system modeling, optimization and control, intelligent inspection technology.