

# 基于天气状态模式识别的 SSA-BP 神经网络 光伏电站功率及碳减排量预测

胡浚惠<sup>1,2</sup>, 丁伟<sup>1,2</sup>, 曹敬<sup>1,2</sup>, 陈时熠<sup>3,4</sup>, 李梦阳<sup>1,2</sup>, 姚钦才<sup>3,4</sup>

(1.国电南瑞科技股份有限公司, 江苏 南京 211106; 2.国电南瑞南京控制系统有限公司, 江苏 南京 211106;  
3.东南大学 能源与环境学院, 江苏 南京 210096; 4.东南大学 碳中和科学技术研究院, 江苏 南京 210096)

**摘要:** 文章提出了一种基于天气状态模式识别并结合 SSA-BP (Sparrow Search Algorithm-Back Propagation) 预测光伏出力的方法。首先,在分析辐照度、温度、风速等参数变化规律基础上,基于高斯混合模型,针对专业天气类型开展分类,获得类晴、类雨和类阴 3 种典型的广义天气;然后,将数据作为 SSA-BP 神经网络输入,对光伏电站出力分类进行预测;最后,结合碳核算方法学对光伏发电项目碳减排量进行核算。结果表明:利用分类识别和改进的 SSA-BP 神经网络,在 3 种天气类型预测中平均相对误差分别为 0.195,0.243,0.310;SSA-BP 与其他模型相比,平均相对误差降低了 17.8%~66.7%。此外,预测 CO<sub>2</sub> 减排量与实际核算值相对误差为 3.37%,亦表现出良好预测效果。

**关键词:** 光伏发电; 模式识别; SSA-BP 神经网络; 功率预测; 天气状态

**中图分类号:** TK51; TP181 **文献标志码:** A **文章编号:** 1671-5292(2024)07-0877-09

## 0 引言

由于气象等多因素影响,光伏发电量存在随机性、间歇性和波动性<sup>[1]</sup>,这些不确定性会对电力系统产生冲击,严重影响电力系统安全和稳定运行。如何提高光伏发电预测精度成为光伏并网亟须解决的关键问题<sup>[2]</sup>。

国内外已陆续开展了光伏发电功率预测的相关研究。基于历史数据,不同的机器学习算法可用于光伏发电功率预测。文献[3]开展了基于马氏距离相似度量预测光伏功率研究,该方法可提高超短期预测精度,但模型训练过程效率不高。文献[4]建立了一种基于自注意机制和多任务学习预测模型,但两种模型存在网络结构复杂、模型并行性能差的问题。文献[5]提出了一种基于 K 均值算法和支持向量机的短期光伏发电功率预测方法,该方法具有一定预测精度,但实际聚类效果不佳。文献[6]提出一种基于改进骨干差分进化算法优化最小二乘支持向量机的光伏发电功率预测模型,模型能够准确预测短期光伏发电功率,但模型易陷入局部最优。文献[7]针对影响光

伏处理属性较多的问题,采用主成分分析降维,然后结合长短期记忆网络,该方法可较好地反映时序数据的动态特性,预测精度较高,但该模型并行性能差,计算耗时较长。相关文献为光伏发电预测准确性提升提供了理论支持,获得了一定成果。前期研究也发现,不同天气状态下采用单一模型开展光伏发电功率预测的准确性相对较低<sup>[8]</sup>。通过天气状态识别,在不同类型天气模式下独立预测是提高光伏发电预测精度的一条重要途径。

本文在分析辐照度变化等规律基础上,提出了一种基于高斯混合模型的天气状态模式识别方法,并建立 SSA(Sparrow Search Algorithm)优化改进的 BP(Back Propagation)光伏发电预测模型。通过与常规支持向量机(Support Vector Machines, SVM)、BP 模型对比,评估 SSA-BP 模型的光伏发电预测效果,并结合碳减排核算方法学计算碳减排量。结果表明,该方法对光伏发电功率和碳减排量预测较为精准,为光伏发电及其碳减排量预测理论和应用提供有力指导。

收稿日期: 2023-05-30。

基金项目: 国电南瑞南京控制系统有限公司科技项目(524609220030)。

作者简介: 胡浚惠(1978-),女,硕士,高级工程师,研究方向为新能源发电及电力系统自动化。

E-mail: huxunhui@sgepri.sgcc.com.cn

通信作者: 曹敬(1983-),男,硕士,高级工程师,研究方向为新型电力系统及其自动化。E-mail: caojing2@sgepri.sgcc.com.cn

## 1 特征选择与数据处理

### 1.1 特征选择

影响光伏发电的自然环境因素较多<sup>[9]</sup>,本文选用了太阳辐照度、温度、风速和大气压力4个典型特征。其中:太阳辐照度是光伏发电最重要的影响因素<sup>[10]</sup>;温度直接影响光伏电池的电流输出,光伏发电对其敏感性也较高<sup>[11]</sup>;风速可影响光伏电池散热,从而影响发电效率<sup>[12]</sup>;大气压力通过间接影响其他气象因素从而影响光伏发电效率<sup>[13]</sup>。

### 1.2 数据清洗处理

本文采用基于正态分布统计学理论的3sigma原则,将超过3倍标准差的数据视为异常值,进而将其清洗或剔除,以保证数据的准确性和可靠性。首先计算数据集中相关自然环境因素(太阳辐照度、温度、风速、大气压力)特征的均值和标准差;然后根据该原则进行数据清洗。

$$U = \mu + 3\sigma \quad (1)$$

$$L = \mu - 3\sigma \quad (2)$$

式中: $U$ 为上界; $L$ 为下界; $\mu$ 为特征均值; $\sigma$ 为特征标准差。

对于数据集中的每个特征,若其数值超过上限界或低于下限界,则被视为异常值,需清洗或剔除。通过数据清洗,可有效减少数据集中的异常值,提高模型精度和可信度。

### 1.3 天气状态模式识别

经数据清洗处理后,使用高斯混合模型聚类法对数据进行分类处理。高斯混合模型是一种基于概率分布的聚类算法。它假设数据点来自于若干个高斯分布混合而成,每个高斯分布对应一个聚类,通过估计每个高斯分布的参数(均值和协方差矩阵),确定每个数据点的类别,实现无监督条件下数据分类处理。

高斯混合模型假设数据点 $x$ 来自于 $K$ 个高斯分布的混合,具体步骤<sup>[14]</sup>如下。

①初始化高斯混合模型参数,随机初始化每个高斯分布的均值向量 $\mu_k$ 、协方差矩阵 $\Sigma_k$ 和先验概率 $\pi_k$ ,其中 $\sum_{k=1}^K \pi_k = 1, K$ 为高斯分布的数量。

②计算每个数据点 $x_i$ 属于每个高斯分布的概率 $p(z_{ik}=1|x_i)$ ,其中 $z_{ik}$ 是一个 $K$ 维的二值向量,

表示第 $i$ 个数据点属于第 $k$ 个聚类的概率为1,其余为0。

③更新高斯分布参数 $\mu_k, \Sigma_k, \pi_k$ 。具体地,对每个高斯分布 $k$ ,计算其对应的先验概率 $\pi_k$ 。 $\pi_k$ 为所有数据点被分配给第 $k$ 个高斯分布的概率之和,即:

$$\pi_k = \frac{1}{N} \sum_{i=1}^N p(z_{ik}=1|x_i) \quad (3)$$

然后,计算每个高斯分布的均值向量 $\mu_k$ 。 $\mu_k$ 为所有被分配到第 $k$ 个高斯分布的数据点 $x_i$ 的加权平均值,即:

$$\mu_k = \frac{\sum_{i=1}^N p(z_{ik}=1|x_i)x_i}{\sum_{i=1}^N p(z_{ik}=1|x_i)} \quad (4)$$

最后,计算每个高斯分布的协方差矩阵 $\Sigma_k$ 。 $\Sigma_k$ 为所有被分配到第 $k$ 个高斯分布的数据点 $x_i$ 的加权协方差矩阵,即:

$$\Sigma_k = \frac{\sum_{i=1}^N p(z_{ik}=1|x_i)(x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^N p(z_{ik}=1|x_i)} \quad (5)$$

④重复步骤②和步骤③,直到收敛。一般来说,可以根据对数似然函数的收敛情况判断算法是否已经收敛,也可以设定一个最大迭代次数控制算法运行。

⑤将每个数据点 $x_i$ 分配给概率最大的高斯分布,即 $z_i = \operatorname{argmax}_k p(z_{ik}=1|x_i)$ 。

⑥最终得到 $K$ 个聚类,每个聚类包含被分配到同一个高斯分布的数据点。

### 1.4 归一化处理与数据集的选取

当处理数据集时,需要进行归一化处理。本文采用最小最大值归一化方法。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

式中: $x'$ 为归一化后的值; $x$ 为原始数据集中的值; $x_{\max}, x_{\min}$ 分别为数据集中最大值、最小值。

处理后,数据集中的所有值都将在0~1内。

## 2 预测模型

### 2.1 基于SSA-BP神经网络的天气预测模型

本文基于BP神经网络的预测模型实现光伏

发电功率的回归预测,如图 1 所示。

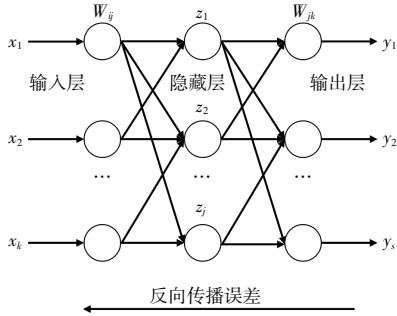


图 1 BP 神经网络原理图

Fig.1 Schematic diagram of BP Neural Network

BP 神经网络是一种由输入层、隐藏层和输出层组成的多层网络。输入层用于接收外部输入,隐藏层和输出层则包含多个神经元,每个神经元都配备一个激活函数和一系列权重。传统的 BP 神经网络存在易陷入局部最优、收敛速度慢等问题,对此,本文引入 SSA 算法优化改进 BP 神经网络。SSA 算法模拟了麻雀觅食时的搜索行为,具有全局搜索能力,可以避免陷入局部最优解,提高全局最优解寻优能力;SSA 算法能加速收敛过程,降低 BP 神经网络训练时间成本,提高计算效率<sup>[15,16]</sup>。

SSA 算法具体实现方法如下。

①初始并更新发现者位置

在 SSA 算法中,高适应度值的个体具有更高搜索优先级。高适应度值的个体不仅承担了为整个种群寻找食物的责任,还能够为其他个体提供探索方向,进而提高整个种群的觅食效率。发现者位置更新规则为

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{-i}{a \cdot M}\right) & R_2 < ST \\ X_{i,j}^t + Q \cdot L & R_2 \geq ST \end{cases} \quad (7)$$

式中: $X_{i,j}^t$  为  $t$  时刻的  $i,j$  个体的位置; $Q$  为服从正态分布的随机数; $L$  为单位行向量; $a$  为 0~1 内的随机值; $M$  为最大迭代次数; $R_2$  为预警值, $R_2 \in [0, 1]$ ;  $ST$  为安全值, $ST \in [0.5, 1]$ 。

②更新加入者位置

当某个个体发现食物时,周围其他个体会被吸引过来,聚集在此处一起寻找食物。该过程表示为

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^t - X_{i,j}^t}{i}\right) & i > \frac{n}{2} \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot H + L & i \leq \frac{n}{2} \end{cases} \quad (8)$$

式中: $X_{worst}^t$  为适应度最低的个体值; $H$  为包含 1 和 -1 的行向量。

③更新警惕者位置

当整个种群面临压力和威胁时,处于群体边缘的个体会寻求更加安全的栖息地,以确保生存环境的适宜性。它们会积极探索周围环境,寻找更好的生存条件,采取迅速的转移和迁徙策略。

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot |X_{i,j}^t - X_{best}^t| & f_i > f_{best} \\ X_{i,j}^t + r \cdot \frac{X_{i,j}^t - X_{worst}^t}{(f_i - f_{worst}) + \epsilon} & f_i = f_{best} \end{cases} \quad (9)$$

式中: $\beta$  为服从正态分布的随机数; $r$  为 -1~1 内的随机数,表达个体适应度值; $\epsilon$  为无穷小量; $f_{best}$  和  $f_{worst}$  分别为当前全局最佳和最差的适应度值; $f_i$  为当前个体的适应度。

BP 神经网络训练过程包括前向传播和反向传播两个阶段。在前向传播阶段,输入结合气象特征以及历史发电量的光伏发电数据,经过输入层、隐藏层计算,最终输出结果。在反向传播阶段,首先计算输出层误差,然后将误差信号反向传播到隐藏层,并根据误差信号来更新每个神经元权重。引入 SSA 算法的目的即是强化该过程。具体而言,SSA-BP 神经网络将 BP 神经网络中的阈值和权重转化为 SSA 算法搜索空间中的个体。其中,麻雀个体适应度值与 BP 神经网络的误差函数相关,适应度值高的个体具有更高的搜索优先级,因此能更快地找到全局最优解。通过迭代更新,SSA 能够逐渐调整 BP 神经网络的阈值和权重,使其逼近最优状态。

基于天气状态模式识别的 SSA-BP 神经网络预测模型全流程如图 2 所示。

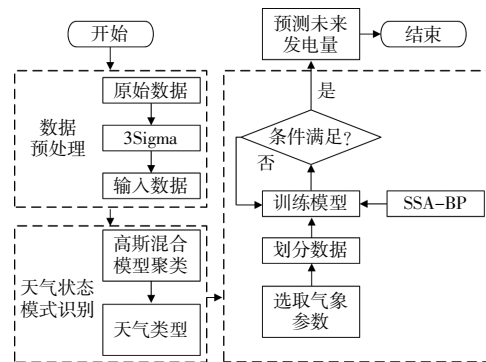


图 2 SSA-BP 算法流程

Fig.2 SSA-BP algorithm flow

具体步骤如下:

①收集历史天气数据和光伏发电功率数据,对数据进行预处理;

②以太阳辐射强度、温度等气象因素,将历史数据按照天气状态进行分类;

③对于每个类别,选取出最相关的气象参数,包括太阳辐射度、温度、大气压力、风速等,将其数据分成训练集和测试集;

④使用 SSA-BP 神经网络对每个类别的数据进行训练,并进行模型优化,例如调整神经网络神经元数、学习率等;

⑤使用训练好的模型对测试集进行预测,并计算预测误差,判断是否达到预测精度标准;如果没达到,则重新训练模型;

⑥对于每个类别,选择最优模型,并使用其对未来光伏发电功率进行预测;

⑦最后,将预测结果进行整合,获得未来一段时间内光伏发电功率预测值。

## 2.2 模型参数选择

根据经验公式,神经网络隐藏层可由以下计算式确定:

$$y = \sqrt{n+m} + a \quad (10)$$

式中: $y$  为隐含层数目; $m, n$  分别为输出量及输入量个数; $a$  为 1~10 的常量。

本研究中输入层个数为 4, 输出层个数为 1, 经测试,获得最优隐藏层数为 7。神经网络模型设置了最大训练次数为 500,学习速率为  $5 \times 10^{-3}$ ,训练目标精度为  $10^{-5}$ 。

## 2.3 预测评价指标

平均绝对误差(Mean Absolute Error, MAE)是预测值与真实值之间差异的绝对值的平均值。MAE 的表达式为

$$MAE = \frac{1}{N} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

式中: $y_i$  为真实值; $\hat{y}_i$  为预测值。

平均相对误差(Mean Relative Error, MRE)是一种用于衡量预测值与真实值之间差异的统计指标,其表达式为

$$MRE = \frac{1}{N} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (12)$$

式中: $n$  为样本数量。

决定系数( $R^2$ )是衡量模型对数据方差的解释能力的指标, $R^2$  值越接近 1, 表示模型性能越好。 $R^2$  的表达式为

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

式中: $\bar{y}$  为真实值的平均值。

## 3 碳减排量核算方法

本文选择适用于小型光伏项目的《CMS-002-V01 联网的可再生能源发电(第一版)》方法学规定要求核算碳减排。该标准是中国可再生能源行业权威性文件,规范了可再生能源发电项目建设、运营和碳减排量核算。根据该规定,项目碳减排量=基准线排放量-项目排放量-泄露量<sup>[7]</sup>。

## 4 分析与讨论

### 4.1 光伏发电功率预测

本研究采用西北某光伏电站一处光伏阵列发电数据集作为神经网络训练集及测试集。该光伏阵列装机时间为 2017 年,功率为 118 kW,光伏材料为单晶硅,地面固定安装。

数据集包含 2018 年 1 月-2022 年 5 月电站的光伏发电功率输出数据、太阳斜面总辐射、环境温度及风速 4 种类型的数据。数据采集时间为 7:00-18:00,数据采集间隔为 15 min,总计  $45 \times 1619$  个光伏发电数据。本研究采用 2018-2021 年  $45 \times 1461$  个数据作为模型的训练集,2022 年 5 月  $45 \times 31$  个数据作为模型的测试集。对历史数据进行归一化处理,利用高斯混合模型聚类算法对气象数据进行聚类,类别为 3,混合系数为 0.7,采用极大似然估计法训练迭代次数 100 次,获得的气象数据如表 1 所示。

表 1 天气数据聚类结果  
Table 1 Clustering of weather data

广义天气类型	I	II	III	
专业气象类型	晴、 晴间多云、 多云间晴	小雨、中雨、大雨、 雷阵雨、小雪、中 雪、大雪、中到大 雪、沙尘暴	多云、阴、 阴间多云、 多云间阴、 雾	
训练集	数量/d 占比/%	597 41	168 11	701 48
测试集	数量/d 占比/%	13 42	4 13	14 45

分析聚类结果的气象数据,并结合中国国家气象局天气类型标准以及西北地区气象特征,聚类获得的天气类型 I, II, III 分别为广义的类晴、类雨及类阴。根据聚类结果,分别选取 3 种天气情

况做测试。

针对 3 种天气类型,选取 3 个典型工作日作为光伏发电预测的预测日,气象数据变化趋势如图 3 所示。

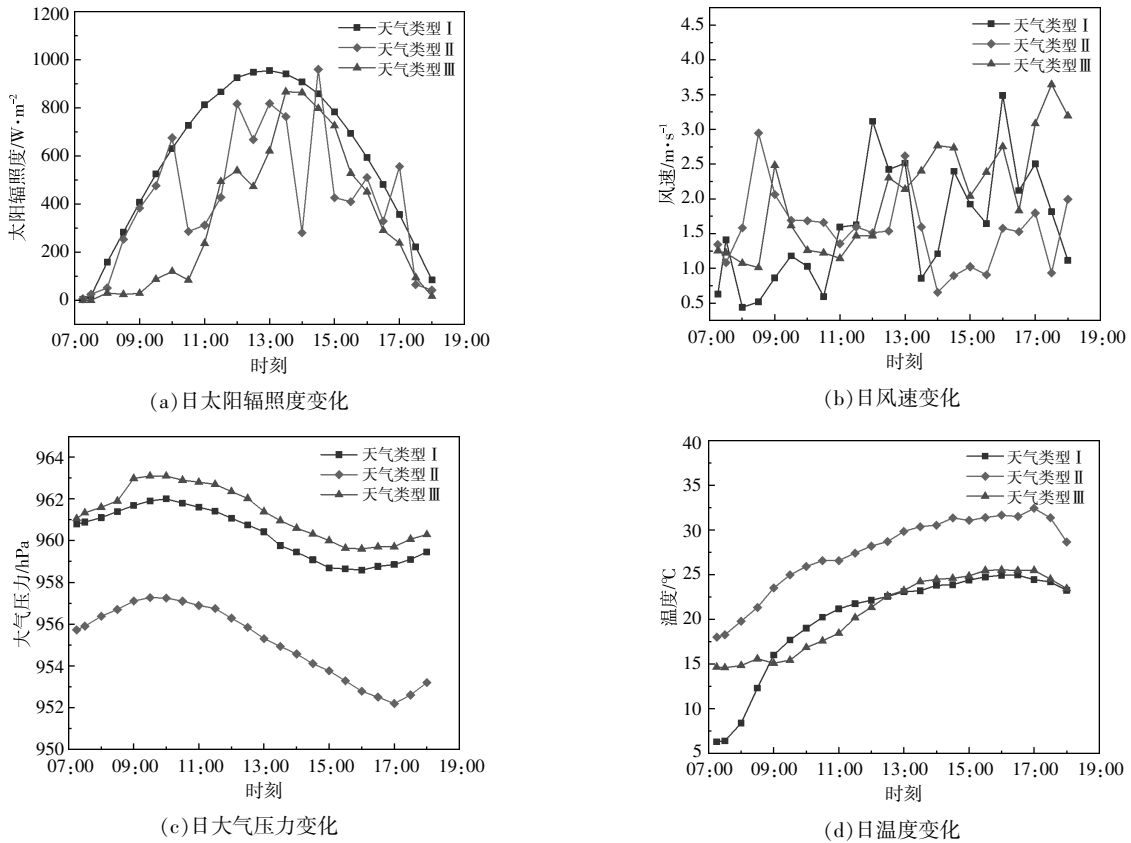
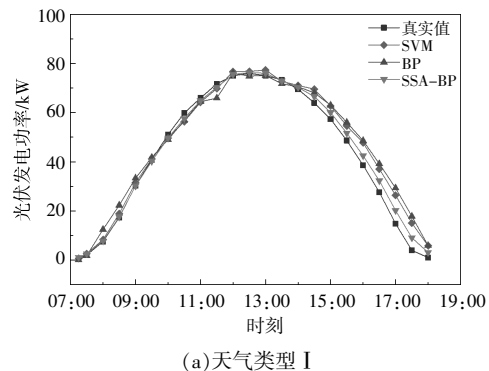


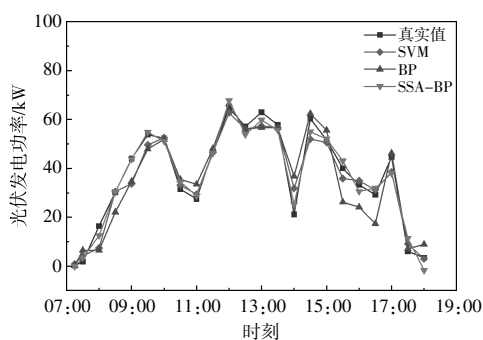
图 3 预测日气象参数变化

Fig.3 Variations of meteorological parameters on forecast days

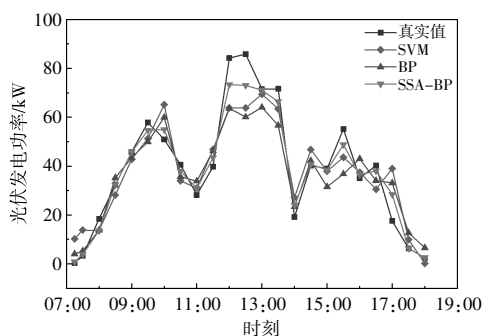
图 3(a)描述了太阳辐射度在 1 d 内的变化趋势。可以看到,太阳辐射度在中午时最高,而在早晚较低。与天气类型 I 相比,天气类型 II 和天气类型 III 因为受自然因素影响,太阳辐射度波动比较大。图 3(b)描述了 1 d 内风速变化情况。可以看到,1 d 内风速变化无固定规律,总体呈现较大波动。图 3(c)描述了预测日大气压力变化情况。可以看出,不同类型天气下大气压力变化趋势类似,1 d 内大气压力呈现波动性,10:00 左右达到峰值,16:00 左右达到最低值。这可能是地球自转和气压系统变化等因素造成。在天气类型 II 中,大气压力相对较小。图 3(d)描述了预测日温度变化情况。可以看出,1 d 内温度基本呈现上升的趋势,这可能是太阳辐射和地表反射热量增加导致。在 17:00 左右,温度达到峰值。

经天气状态模式识别,采用 SSA-BP 神经网络对 3 种天气下光伏发电功率进行预测,并与采用 SVM、常规 BP 神经网络预测进行对比,预测结果如图 4 所示。由图 4 可知,总体上天气类型 I 中光伏发电功率变化较为平稳,3 种模型的预测效果都很好。天气类型 II 由于雨水干扰,光伏发电





(b)天气类型 II



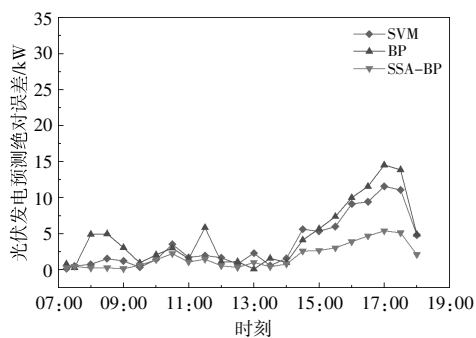
(c)天气类型 III

图4 不同天气类型下光伏发电预测对比

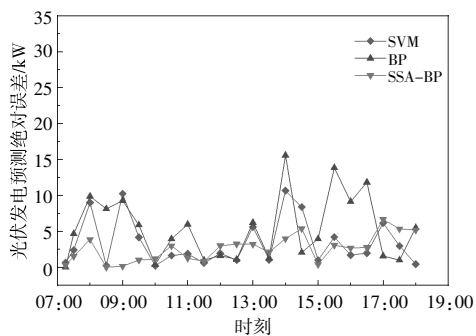
Fig.4 PV power generation forecasts under different weather types

功率变化较为剧烈。天气类型 III 由于云层遮挡导致太阳辐照度变化,光伏发电功率也呈现波动状态。

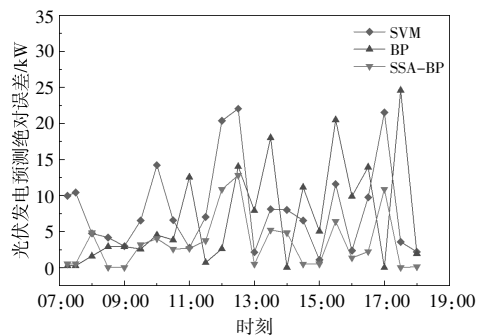
3 种天气的预测绝对误差如图 5 所示。



(a)天气类型 I



(b)天气类型 II



(c)天气类型 III

图 5 光伏发电预测绝对误差对比

Fig.5 Absolute measurement errors of PV power generation

3 种天气类型中,总体上天气类型 I 中预测误差最小,但是常规 BP 算法从 15:00-18:00 有较大的预测误差,可能是由于太阳辐照度在神经网络模型中权重过大所致,这一现象经 SSA 算法优化后有明显改善。

预测结果的评价指标如表 2 所示。

表 2 预测日评价指标

Table 2 Evaluation indicators of prediction days

天气类型	算法	$R^2$	MAE/kW	MRE
I	SVM	0.976	3.860	0.442
	BP	0.967	4.759	0.585
	SSA-BP	0.994	1.870	0.195
II	SVM	0.783	7.021	0.668
	SSA-BP	0.959	2.866	0.243
III	SVM	0.825	6.078	0.377
	BP	0.722	8.755	0.596
	SSA-BP	0.946	3.767	0.310

对比 3 种模型预测结果可以发现,SSA-BP模型的预测效果最好。在天气类型 I 的预测中,SSA-BP 模型的  $R^2$  达到了 0.994,比 SVM 和 BP 算法相对提升了 0.018 和 0.027;MAE 降低到 1.870 kW,比 SVM 和 BP 算法的误差降低了 1.990 kW 和 2.889 kW;MRE 比 SVM 和 BP 下降了 55.9% 和 66.7%。

在天气类型 II 的预测中,SSA-BP 模型的  $R^2$  为 0.959,比 SVM 和 BP 算法相对提升了 0.176 和 0.155;MAE 为 2.866 kW,比 SVM 和 BP 算法提升了 4.155 kW 和 3.882 kW;MRE 比 SVM 和 BP 下降了 63.6%和 38.6%。

在天气类型 III 的预测中,相比前两种天气类型,由于气象变化的复杂性和不确定性,SSA-BP

模型表现相对较差,  $R^2$  为 0.946, 但是仍然比 SVM 和 BP 算法提升了 0.121 和 0.224, 尤其相对传统 BP 算法,  $R^2$  提升幅度更大; MAE 为 3.767 kW, 比 SVM 和 BP 算法提升了 2.311 kW 和 4.988 kW; MRE 比 SVM 和 BP 算法下降了 17.8% 和 48.0%。

这表明 SSA-BP 模型在天气类型 I 和 II 的预测中具有更高的准确性和稳定性, 且预测误差较小。在天气类型 III 的预测中, 虽然 SSA-BP 模型的准确性对比其在其他天气类型预测中有所下降, 但相比 SVM 和 BP, 预测更准确。在 3 种天气类型中相较于其他预测模型, SSA-BP 神经网络预测值  $R^2$  最高提升了 31%, MRE 最高降低了 66.7%, SSA-BP 展示出较好的预测精度与天气类型适应性。

采用 SSA-BP 神经网络对该发电项目 2022 年 5 月光伏发电量进行预测, 预测结果和真实数据对比如图 6 所示。

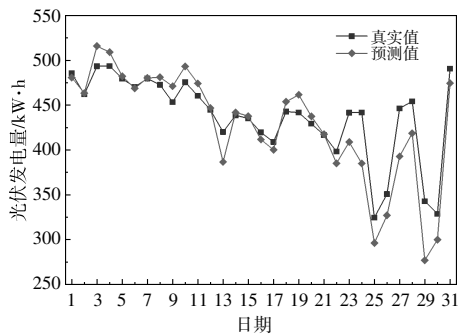


图 6 光伏发电量预测对比

Fig.6 Comparison of PV power generation prediction

对预测数据与实际数据进行一个月的累计, 2022 年 5 月该光伏项目预测发电量为 13.54 MW·h, 实际发电量为 13.23 MW·h, 误差百分比为 2.31%。结果表明, SSA-BP 预测模型具有较高的精度和准确性, 能够为实际光伏发电应用提供可靠的参考。

#### 4.2 碳减排量核算结果

除功率预测外, 基于光伏功率和实际出力还可开展碳减排量预测。2022 年度西北地区电量边际排放因子 (Operation Margin, OM) 和容量边际排放因子 (Build Margin, BM) 分别为 0.892 2 tCO<sub>2</sub>/(MW·h) 和 0.440 7 tCO<sub>2</sub>/(MW·h), 其组合边际排放因子为 0.779 3 tCO<sub>2</sub>/(MW·h)<sup>[18]</sup>。在 2022 年 5 月, 该光伏阵列项目实际获得的减排量为 10.67 tCO<sub>2</sub>e, 通过模式识别的 SSA-BP 预测功率下二氧

化碳减排量为 10.31 tCO<sub>2</sub>e。累计预测值与实际值的相对误差仅为 3.37%。这一结果表明, SSA-BP 神经网络与碳核算方法学结合可获得较为准确的碳减排量预测结果, 能够为碳市场中光伏项目交易碳量期望提供重要参考。

#### 5 结束语

本文提出了基于天气模式识别并结合 SSA-BP 神经网络的光伏发电功率预测方法, 可有效提高光伏发电功率预测的准确度。天气模式识别可实现天气分类预测, 提高预测准确度; SSA 算法能够优化 BP 神经网络阈值和权重, 使其趋于最优状态进而提高神经网络预测精度, 尤其是复杂天气类型下的精度。与 SVM 以及常规 BP 神经网络模型相比, 基于高斯混合模型分类下的 SSA-BP 神经网络模型, 典型工况下  $R^2$  最高提升了 31%, MRE 最高降低了 66.7%, 该方法对于实际光伏功率预测具有重要的指导意义和应用价值。用该光伏发电模型预测结合碳减排核算, 结果误差仅为 3.37%, 可为碳市场中光伏碳减排量期望提供参考。

#### 参考文献:

- [1] 李富彪, 庞秀岚, 马超. 光伏电站组件和运行方式组合及其容量优化配置研究[J]. 可再生能源, 2022, 40(1): 34-40.
- [2] Li Y, Xu J, Yuan C. A study of photovoltaic power prediction[J]. Applied Mechanics & Materials, 2015, 733: 203-206.
- [3] 杨茂, 冯帆. 基于马氏距离相似度量的光伏功率超短期预测方法的研究[J]. 可再生能源, 2021, 39(2): 175-181.
- [4] Ju Y, Li J, Sun G. Ultra-short-term photovoltaic power prediction based on self-attention mechanism and multi-task learning[J]. IEEE Access, 2020(3): 44821-44829.
- [5] 张雨金, 杨凌帆, 葛双治, 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): 118-124.
- [6] 张涛, 朱瑞金, 扎西顿珠. 基于改进骨干差分进化算法优化 LSSVM 的短期光伏发电功率预测[J]. 热力发电, 2021, 50(5): 102-107.
- [7] 宋绍剑, 李博涵. 基于 LSTM 网络的光伏发电功率短期预测方法的研究[J]. 可再生能源, 2021, 39(5): 594-602.

- [8] Gao M, Li J, Hong F, et al. Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM[J]. *Energy*, 2019, 187: 115838.
- [9] 赵滨滨, 王莹, 王彬, 等. 基于 ARIMA 时间序列的分布式光伏系统输出功率预测方法研究[J]. *可再生能源*, 2019, 37(6): 820-823.
- [10] 姜文玲, 赵艳青, 王勃, 等. 基于 NWP 辐照度斜面转换的光伏功率预测方法 [J]. *山东大学学报 (工学版)*, 2021, 51(5): 114-121.
- [11] Xiao W, Dai J, Wu H, et al. Effect of meteorological factors on photovoltaic power forecast based on the neural network [J]. *RSC Advances*, 2017, 88: 55846 - 55850.
- [12] Gökmen N, Hu W, Hou P, et al. Investigation of wind speed cooling effect on PV panels in windy locations[J]. *Renewable Energy*, 2016, 90: 283-290.
- [13] Amajama J. Effect of air pressure on the output of photovoltaic panel and solar illuminance (or intensity) [J]. *International Journal of Scientific Engineering and Applied Science*, 2016, 2(8): 139-144.
- [14] Afkhami R G, Azarnia G, Tinati A M. Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals [J]. *Pattern Recognition Letters*, 2016, 70(15): 45-51.
- [15] 魏鹏飞, 樊小朝, 史瑞静, 等. 基于改进麻雀搜索算法优化支持向量机的短期光伏发电功率预测[J]. *热力发电*, 2021, 50(12): 74-79.
- [16] 尹德鑫, 张达敏, 蔡朋宸, 等. 改进的麻雀搜索优化算法及其应用 [J]. *计算机工程与科学*, 2022, 44(10): 1844-1851.
- [17] 中华人民共和国国家发展和改革委员会办公厅. CMS-002-VO1 联网的可再生能源发电(第一版)[EB/OL]. <https://www.ccchina.org.cn/archiver/cdmcn/UpFile/Files/Default/20130311171148538669.pdf>, 2013-03-05.
- [18] 中华人民共和国生态环境部. 2019 年度减排项目中国区域电网基准线排放因子[EB/OL]. <https://www.mee.gov.cn/ywgz/ydqhbh/wsqtz/202012/W020201229610353340851.pdf>, 2020-12-29.

## Prediction of photovoltaic power plant output and related carbon reduction based on SSA-BP neural network with pattern recognition

Hu Xunhui<sup>1,2</sup>, Ding Wei<sup>1,2</sup>, Cao Jing<sup>1,2</sup>, Chen Shiyi<sup>3,4</sup>, Li Mengyang<sup>1,2</sup>, Yao Qincan<sup>3,4</sup>

(1.NARI Technology Co.,Ltd., Nanjing 211106, China; 2.NARI Technology Nanjing Control System Co.,Ltd., Nanjing 211106, China; 3.School of Energy and Environment, Southeast University, Nanjing 210096, China; 4.Institute of Science and Technology for Carbon Neutrality, Southeast University, Nanjing 210096, China)

**Abstract:** This paper proposes a method to predict the photovoltaic output based on weather state pattern recognition and SSA-BP, which is more accurate than traditional single models under different weather conditions. Firstly, the historical data was cleaned using the 3sigma algorithm to obtain the data that can reflect the output of photovoltaic power plants and the regularity of weather changes. Then, based on the analysis of the parameters such as irradiance, temperature, and wind speed, Gaussian mixture models were applied to classify the professional weather types and three typical generalized weather types were obtained. Furthermore, the data was used as SSA-BP neural network input to predict the futuristic photovoltaic power plant output. Finally, the carbon accounting method was used to calculate the carbon emission reduction of the photovoltaic power generation project. The experimental results show that through classification recognition and the optimized SSA-BP neural network, the mean relative errors in the prediction for the three weather types are 0.195, 0.243 and 0.310, respectively. Compared with other predication models, the relative errors are reduced by 17.8%~66.7%. In addition, the relative error between the predicted carbon dioxide emission reduction and actual value is only 3.37%. The model proposed in this work shows satisfactory prediction results.

**Keywords:** photovoltaic power; pattern recognition; Sparrow Search Algorithm-Backpropagation (SSA-BP); power prediction; weather conditions