

# 基于SSA优化的太阳能发电功率预测混合模型

王子毅, 冯翼

扬州大学数学科学学院, 江苏 扬州

收稿日期: 2024年6月1日; 录用日期: 2024年6月21日; 发布日期: 2024年6月30日

## 摘要

本文通过建立基于麻雀搜索算法(SSA)优化的太阳能发电混合模型, 对太阳能发电功率进行预测。首先, 通过改进的完全经验模态分解(ICEEMDAN)技术对太阳能发电数据进行预处理, 将原始数据分解为固有模态函数(IMFs)和残余部分, 从而有效去除噪声和非平稳成分。然后, 将预处理后的IMFs和残余数据输入到双向长短时记忆网络(BiLSTM)进行建模, 进一步使用ODE-LSTM技术稳定地学习长时间依赖关系。为提升该预测模型的性能, 采用SSA进行超参数优化。实验结果表明, 通过SSA优化的ODE-BiLSTM模型的均方误差(MSE)为116.78、平均绝对误差(MAE)为206.38, 判定系数( $R^2$ )为0.7527。在这些指标表现上均优于传统的ARIMA、单一LSTM及随机森林模型, 展现出更高的预测精度和鲁棒性。本研究验证了混合模型和优化算法在太阳能发电预测领域的潜力, 为智能电网中新能源的高效管理提供了有力支持。

## 关键词

经验模态分解, 神经微分方程, 电力系统预测, 长短时记忆网络

# A Hybrid Model for Solar Power Prediction Based on SSA Optimization

Ziyi Wang, Yi Feng

School of Mathematical Sciences, Yangzhou University, Yangzhou Jiangsu

Received: Jun. 1<sup>st</sup>, 2024; accepted: Jun. 21<sup>st</sup>, 2024; published: Jun. 30<sup>th</sup>, 2024

## Abstract

In this paper, solar power generation power is predicted by establishing a hybrid model for solar power generation based on the optimisation of Sparrow Search Algorithm (SSA). In the first place, the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) technique is utilized for preprocessing solar power generation data, decomposing the original data into Intrinsic Mode Functions (IMFs) and residuals to effectively remove noise

and non-stationary components. Subsequently, the preprocessed IMFs and residual data are input into a Bidirectional Long Short-Term Memory network (BiLSTM) for modeling, with further steadily learning for long-term dependencies achieved through ODE-LSTM technology. To enhance the performance of the predictive model, SSA is employed for hyper parameter optimization. Experimental results demonstrate that the SSA-optimized ODE-BiLSTM model outperforms traditional ARIMA, standalone LSTM, and Random Forest models in terms of Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared ( $R^2$ ) metrics, exhibiting superior predictive accuracy and stability. This study confirms the potential of hybrid models and optimization algorithms in the field of solar power generation forecasting, providing robust support for efficient management of renewable energy within smart grids.

## Keywords

Empirical Mode Decomposition, Neural Differential Equations, Power System Prediction, Long and Short Term Memory Networks

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

党的十九届六中全会提出, 聚焦生态文明建设, 贯彻新发展理念, 制定严格的环境保护制度法规, 坚守人与自然和谐共生的底线, 在针对我国能源禀赋形成的刚性约束之下, 以协调为内生特点, 将绿色视作今后发展的普遍形态, 不断推动国内经济发展模式朝着低碳绿色的方向前进, 实现我国的“双碳”目标[1]。为落实“碳达峰碳中和”的能源战略目标, 预计到2060年, 全社会用电量将超过16万亿kW·h, 新能源发电装机达到50亿kW, 新能源发电量占比将由目前的8%提高到60%以上[2]。太阳能发电作为一种清洁可再生能源, 在全球能源转型和减少碳排放方面具有重要的战略意义。随着可持续发展目标和全球环境保护意识的提升, 太阳能发电厂已成为现代能源体系的重要组成部分。然而, 太阳能发电的间歇性和不稳定性使得精准预测其发电量对电力系统的稳定运行和能源管理至关重要。准确的太阳能发电预测不仅能够帮助电力系统平衡供需、优化电网运营, 还可以减少备用发电的需求, 提高经济效益, 并降低因太阳能资源不确定性引起的风险。因此, 通过时间序列预测方法对太阳能发电数据进行分析, 为实现可再生能源的高效利用提供了重要的技术支撑。本文旨在探讨如何利用时间序列预测技术对太阳能发电数据进行建模, 以期实现对太阳能发电的准确预测, 从而提升电力系统的可持续性和运营效率。

## 2. 文献综述

国内外光伏发电量预报技术主要可分为如下3类[3]。

一是仿真预报法, 利用太阳总辐射的预报值和光伏发电模拟模型[4], 通过模拟太阳能电池的I/V特性曲线进行预测。该方法基于标准条件下的模型公式, 能较好地模拟电池的实际运行状态。二是原理预报法, 基于光伏发电的物理原理及光电转换效率模型, 建立太阳总辐射和发电量之间的关系[5]。考虑辐射度、温度等因素, 输入太阳总辐射预报值计算光伏发电量。三是动力-统计预报法, 结合历史气象数据(如天气情况、太阳总辐射)和发电数据, 使用统计学模型(如多元回归、神经网络)构建预测模型, 并将数值天气预报结果作为输入, 预测发电量。事实上, 新能源发电的随机性和不确定性使得通过简单的多元回归准确地预测新能源功率非常困难。近年来, 以数据驱动为核心的人工智能技术得到了快速发展。机器

学习是一种从数据中自动学习规律并做出预测的方法, 其主要优势是能够处理大量的数据, 并通过学习数据的规律来预测功率输出[6]。

Manoja Kumar Behera等人(2018)采用优化的极限学习机(ELM)技术[7], 对太阳能光伏发电进行短期预测。以极限学习机为基础的预测模型在减少误差方面具有较高的稳定性。通过应用不同粒子群优化技术(PSO、CRPSO、APSO), 将输入权重和偏置更新到最佳状态。比较表明, APSO优化的ELM模型在均方根误差、平均绝对误差和平均绝对百分比误差方面均优于其他模型。对于真实数据预测, 该模型实现了更精确的结果。刘园园(2018)提出了一种自适应K-均值与门控循环单元(GRU)结合的短期预测模型[8], 以解决太阳能光伏发电预测中的间歇性和随机性问题。通过先用自适应K-均值对数据进行聚类, 再用GRU进行发电量预测。研究在晴天、多云和雨天三种典型天气下进行了测试, 比较自适应K-均值GRU模型和其他传统模型的预测效果。结果表明, 在不同天气情况下, 该模型均显著减少了预测误差。与其他模型相比, 误差减少了近15%到20%。

Zhuang Wei等(2024)提出了一种结合图卷积网络(GCN)和Informer的新型预测框架, 用于中期光伏发电预测[9]。该模型通过图卷积网络提取特征信息, 并利用Informer的多输出能力实现长序列的快速预测。实验结果表明, GCN-Informer模型在均方误差和平均绝对误差方面表现优于其他预测模型, 特别在处理复杂的时间序列数据时, 展现出更高的精度和实时性。

这些论文共同表明, 综合先进的机器学习技术与优化算法, 可以显著提高太阳能等新能源的短期预测精度, 为智能电网中的能源管理提供有力支持。

### 3. 模型结构

#### 3.1. 改进的完全经验模态分解(ICEEMDAN)

改进的完全经验模态分解(ICEEMDAN)技术由Marcelo A. Colominas等人在2014年提出[10]。这种技术解决了传统EMD方法中的频率混叠和伪模式问题。通过添加白噪声, 它能够实现相邻尺度之间的频率连续性, 并逐渐减少频率混叠的效果。主要步骤如下:

1. 通过EMD确定局部平均值的实现:

$$s^i = s + \beta_0 E_1(w^i)$$

其中 $i$ 是添加的噪声序号,  $s$ 是原始信号,  $s^i$ 是待分解的信号,  $w^i$ 表示白噪声,  $E_1(w^i)$ 表示白噪声的第一分量。

2. 计算第一次残差 $R_1$ :

$$R_1 = M(s^i)$$

3. 计算第一个模态 $IMF_1$ :

$$IMF_1 = s - R_1$$

4. 计算第二次残差和第二个模态:

$$R_2 = M(R_1 + \beta_1 E_2(w^i))$$

$$IMF_2 = R_1 - R_2$$

5. 计算第 $k$ 次残差和模态:

$$R_k = M(R_{k-1} + \beta_k E_k(w^i))$$

$$IMF_k = R_{k-1} - R_k$$

6. 对于后续的阶段, 重复步骤4。

总的来说, ICEEMDAN强有力的改进了传统EMD分解最终趋势含有残余特征, 在一个分解模式中有不同尺度震荡等缺陷, 它通过加入白噪声来克服传统EMD方法中的问题, 采用多个步骤迭代计算出多个内禀模态函数(IMF)和残差, 逐步分解信号以分析其内在的频率特征。

### 3.2. SSA 优化算法

麻雀搜索算法(Sparrow Search Algorithm, 简称SSA)是一种于2020年提出的优化算法[11], 灵感源自麻雀的觅食和反捕食行为。相较于进化差分算法[12], 记忆粒子群优化[13], 鲸鱼优化算法[14], SSA以其简单的原理、强大的鲁棒性、快速的收敛速度、较少的参数调整需求、卓越的搜索能力和出色的效果等优势而备受瞩目。该算法及其变体在解决连续优化问题时表现出色。

在一个 $D$ 维的解空间中, 每只麻雀的位置可以表示为 $x = \{x_1, x_2, \dots, x_D\}$ , 其适应度值用 $f = f\{x_1, x_2, \dots, x_D\}$ 来表示。假设麻雀的总数量为 $N$ , 在每一代中会选择表现最优的 $PN$ 只麻雀作为生产者。而剩下的 $N-PN$ 只麻雀则扮演拾荒者的角色。

在该算法中, 个体适应度值较高的麻雀在寻找食物时会比其他麻雀更具优势。生产者通常能够找到丰富的食物, 并且为拾荒者提供方向。因此, 在每次迭代中, 生产者位置的更新遵循以下公式:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot \text{iter max}}\right), R_2 < ST \\ X_{i,j}^t + Q \cdot L, R_2 \geq ST \end{cases}$$

其中,  $X_{ij}$ 表示第 $i$ 只麻雀在第 $j$ 维度中的位置;  $t$ 代表当前迭代次数;  $\alpha$ 是一个随机数, 取值范围为 $(0,1]$ ;  $\text{iter}_{\max}$ 是一个常数;  $R_2$ 和 $ST$ 分别代表警报值和安全阈值, 其中 $R_2 \in [0,1]$ ,  $ST \in [0.5,1]$ ;  $Q$ 是一个随机分布值;  $L$ 是一个 $1 \times d$ 的矩阵, 所有元素都为1。如果 $R_2$ 小于 $ST$ , 说明没有捕食者威胁, 生产者会进行更广泛的搜索; 如果 $R_2$ 大于等于 $ST$ , 则说明有麻雀发现捕食者并发出警报, 促使整个种群前往更安全的区域觅食。拾荒者的位置更新则由以下公式确定:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{\text{worst}}^t - X_{i,j}^t}{\alpha \cdot \text{iter max}}\right), i > n/2 \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot A^+, \text{otherwise} \end{cases},$$

其中,  $X_{\text{worst}}$ 表示当前全局最差的位置;  $X_p^{t+1}$ 表示第 $t+1$ 次迭代时生产者的最佳位置;  $A$ 是一个 $1 \times d$ 的矩阵, 其元素为随机选择的1或-1, 且需满足 $A^+ = A^T (AA^T)^{-1}$ 。如果 $i > \frac{n}{2}$ , 则表示该拾荒者适应度较低, 因未获取到足够的食物而饥饿, 此时需要迁移到其他地方觅食。

当麻雀察觉到危险时, 会放弃当前食物并进行预警行为, 其位置更新公式为:

$$X_{i,j}^{t+1} = \begin{cases} X_{\text{best}}^t + \beta \cdot |X_{i,j}^t - X_{\text{best}}^t|, f_i > f_g \\ X_{i,j}^t + K \cdot \left(\frac{|X_{i,j}^t - X_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon}\right), f_i = f_g \end{cases},$$

其中, 随机测度 $K \in [-1,1]$ ;  $\beta$ 是服从正态分布的步长控制参数, 平均值为0;  $\varepsilon$ 是一个最小常数, 防止分母为零的错误;  $X_{\text{best}}$ 表示当前全局最佳位置;  $f_i$ 代表当前麻雀个体的适应度值;  $f_w$ 和 $f_g$ 分别是当前全局最差和最好的适应度值。

当  $f_i > f_g$  时, 意味着麻雀位于种群的边缘, 容易受到捕食者的攻击; 而当  $f_i = f_g$  时, 说明位于种群中央的麻雀已发现危险, 为避免被捕食的风险, 它们会尽量靠近其他麻雀。

### 3.3. ODE-LSTM

在研究不规则采样的时间序列数据时, 循环神经网络(RNN)是一种适合的工具。然而, 当涉及到长时间依赖时, 它们经常遇到梯度消失或爆炸的问题。这种现象同样影响到基于神经微分方程的 ODE-RNN 模型。为了克服这一挑战, Mathias Lechner 等人提出了一种创新的解决方案——ODE-LSTM [15]。标准 RNN 中隐藏状态的变化可通过以下常微分方程表示:

$$\frac{\partial h}{\partial t} = f_{\theta}(x_{t+T}, h_t, T) - \tau h,$$

其中,  $x_{t+T}$  表示输入序列,  $h_t$  是 RNN 的隐藏状态,  $\tau$  为阻尼系数。时间滞后  $T$  决定了输入数据的采样时间。然而, 由于在训练过程中梯度可能出现消失或爆炸, 模型难以学习长时间依赖。解决方案是利用 LSTM 独特的门控机制, 能够确保误差在记忆路径中保持一致的传播。其输入更新由以下公式定义:

$$z_{t+1} = \tanh(W_z x_{t+1} + R_z h_t + b_z)。$$

输入门、遗忘门和输出门分别通过:

$$i_{t+1} = \sigma(W_i x_{t+1} + R_i h_t + b_i)$$

$$f_{t+1} = \sigma(W_f x_{t+1} + R_f h_t + b_f + 1)$$

$$o_{t+1} = \sigma(W_o x_{t+1} + R_o h_t + b_o)$$

这些门控机制使得 LSTM 模型可以记住或遗忘输入信息。记忆单元和隐藏状态更新分别为:

$$c_{t+1} = z_{t+1} \cdot \dot{i}_{t+1} + c_t \cdot f_{t+1}$$

$$h_{t+1} = \tanh(c_{t+1}) \cdot o_{t+1}$$

为了应对时间序列中任意间隔的数据, ODE-LSTM 使用自适应步长求解器来处理 LSTM 的输出状态:

$$h_i = \text{ODESolve}(f_{\theta}, h_{i-1}, h'_i, t_i - t_{i-1}),$$

ODE 层的引入确保 LSTM 记忆和隐藏状态与连续时间架构相结合, 能够更好地处理长时间依赖的学习任务。为分析不规则采样的复杂时间序列数据提供了一种稳健的解决方案。

### 3.4. 基于 SSA 优化的混合预测模型

首先使用改进的集合经验模态分解法(ICEEMDAN)对太阳能发电数据进行预处理。ICEEMDAN 通过逐层分解信号, 将原始数据分解为不同频率的固有模式函数(IMFs)和残余部分。各 IMF 分量代表不同的时间特征, 保留原始数据的完整信息。通过这种方式可以有效地去除噪声和非平稳成分, 保留关键趋势, 然后将固有模式函数与本来筛选的特征整合, 提高后续的预测精度。

然后经过预处理后的 IMFs 和残余数据作为输入, 分别被送入双向长短时记忆网络(BiLSTM)进行建模。相比传统的 LSTM, BiLSTM 能够从前向后两个方向捕捉序列的时间依赖关系, 获取更丰富的特征信息。这有助于捕获发电功率随时间变化的趋势和周期性特征, 提升预测的准确性。为了解决双向 LSTM 在长时间依赖预测中的梯度消失或爆炸问题, 研究使用 ODE-LSTM 将 BiLSTM 的输出融入连续时间架构。ODE-LSTM 通过将隐状态的变化建模为常微分方程, 可以稳定地学习长时间跨度的依赖关系。

为了优化模型性能并提升预测的准确性, 对模型采用麻雀优化算法(SSA)进行超参数优化。SSA 模拟



麻雀的觅食和防御行为, 通过种群之间的信息传递和个体巡游寻找最佳食物源, 它可以有效寻找到 ODE-BiLSTM 的最佳超参数组合, 使模型在预测太阳能发电功率时表现更为准确。

## 4. 模型实证检验

### 4.1. 数据处理

本文使用的数据集为“Solar Power Generation Data”, 该数据集是由 Kaggle 平台用户 Aniket Datar 提供的公开数据集(<https://www.kaggle.com/datasets/anikannal/solar-power-generation-data>), 旨在帮助研究人员和工程师分析太阳能发电功率的变化规律并预测未来的功率输出。数据集涵盖了两个太阳能发电站的详细数据: 发电站 1 (Plant 1) 和发电站 2 (Plant 2), 记录了它们的发电数据及相关的气象信息。该数据集非常适合用于构建和评估太阳能发电预测模型。数据集变量具体情况如表 1 所示。

Table 1. Dataset variables

表 1. 数据集变量

变量	描述
DATE_TIME	记录数据的时间戳
PLANT_ID	电厂标识符
SOURCE_KEY	发电逆变器的标识符
DC_POWER	逆变器产生的直流电功率
AC_POWER	逆变器输出的交流电功率
DAILY_YIELD	当天发电量
TOTAL_YIELD	累计发电量
AMBIENT_TEMPERATURE	环境空气温度
MODULE_TEMPERATURE	太阳能模块温度
IRRADIATION	接受到的太阳能辐照强度

数据预处理分为以下步骤, 以确保数据质量和分析的有效性:

(1) 处理缺失数据:

由于数据逻辑可能会对分析造成偏差, 我们使用三次样条插值法来估算缺失的数据点。具体公式为:

$$S(x) = a_3x^3 + a_2x^2 + a_1x + a_0,$$

其中,  $a_0, a_1, a_2$  和  $a_3$  是样条函数系数, 这些系数通过已知数据点解算得到, 以确保插值曲线在数据点间平滑过渡。

(2) 异常值处理:

对数据集进行异常值分析, 识别可能由设备故障或数据记录错误引起的数据点。异常值通过 IQR 方法识别, 定义为超出四分位数范围 1.5 倍 IQR 的值。异常值将被替换为相邻值的三次样条插值结果。

(3) 数据标准化:

为了消除不同量纲的影响, 使用 Z-score 标准化方法对数据进行规范化处理。该公式为:

$$z = \frac{(x - \mu)}{\sigma},$$

其中,  $x$  是原始数据点,  $\mu$  是平均值,  $\sigma$  是标准差。此步骤后, 所有特征值将转换为以 0 为中心, 标准差为 1 的分布, 便于后续模型分析。

为了确保模型在处理光伏发电影响因素时不受过度拟合的影响, 提高其预测精度, 并同时减少计算

负担, 特征选择显得至关重要。在这一过程中, 我们需要有效地筛选出与预测变量相关性较高的特征。皮尔森系数法和斯皮尔曼系数法是两种常用的统计方法, 用于分析两组变量之间的相关程度。斯皮尔曼系数法由于其适用于变量中含有定序数据的情况而广泛受到应用, 而皮尔森系数则更适用于处理服从正态分布的数据。鉴于本文所涉数据并未显著服从正态分布, 我们选择采用斯皮尔曼系数法来计算每个特征与预测变量之间的相关性。

$$\rho(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

通过绘制特征相关系数的热力图, 我们可以清晰地观察到各个特征与预测变量之间的关联程度, 如图 1。

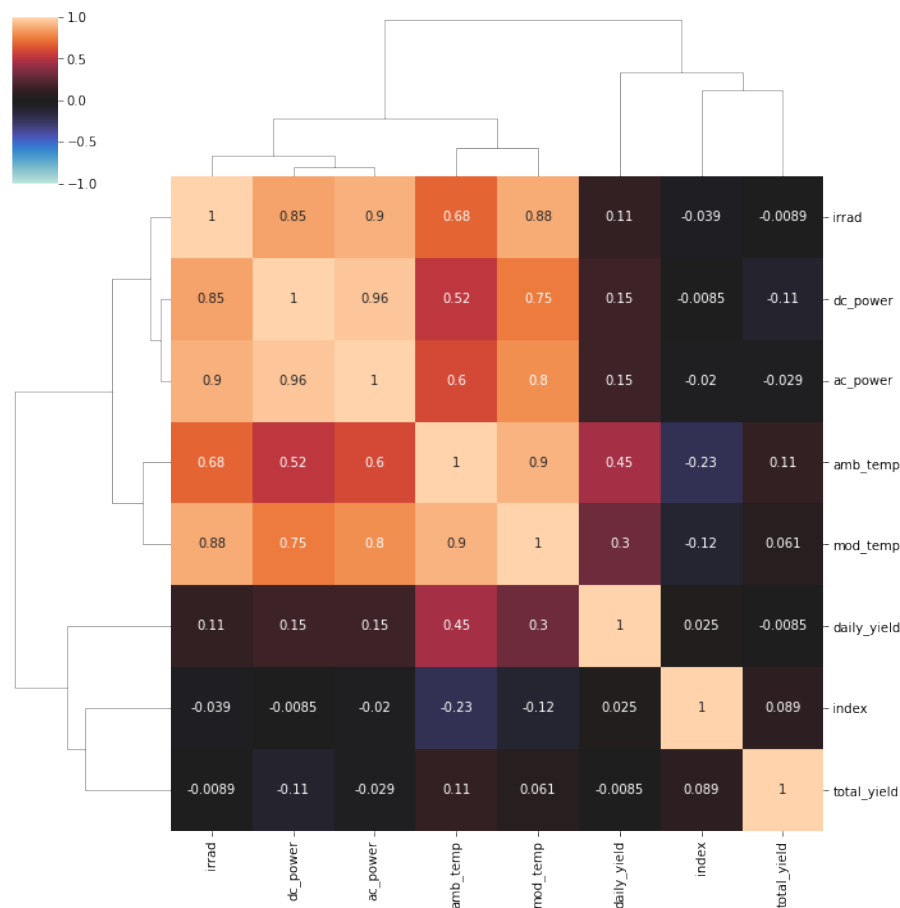


Figure 1. Cluster heatmap  
图 1. 聚类热图

图 1 配合层次聚类(在左侧和顶部的树状图)用于显示不同特征之间的相关性。图中颜色从蓝色(表示非常低的相关性, 即接近 -1)到红色(表示非常高的相关性, 即接近 1)。dc\_power 和 ac\_power 之间有非常高的相关性(0.96), 这是合理的, 因为它们都与产生的电量有关。mod\_temp 和 amb\_temp 也有很高的相关性(0.9), 可能因为外部环境温度直接影响模块温度。上述两组相关性高的数据每组只保留一个防止特征携带重复的信息, 导致模型过拟合。

## 4.2. 实验设置

首先使用改进的完全经验模态分解处理预测变量, 分解出了 11 个 IMF, 通过残差分析, 即在去除某些 IMF 后, 分析剩余信号的特性, 以确定这些 IMF 是否包含重要信息。如果剩余信号失去了重要的信息或特性, 那么这些 IMF 可能是关键的。由此筛选了 6 个重要的 IMF 如图 2 所示。

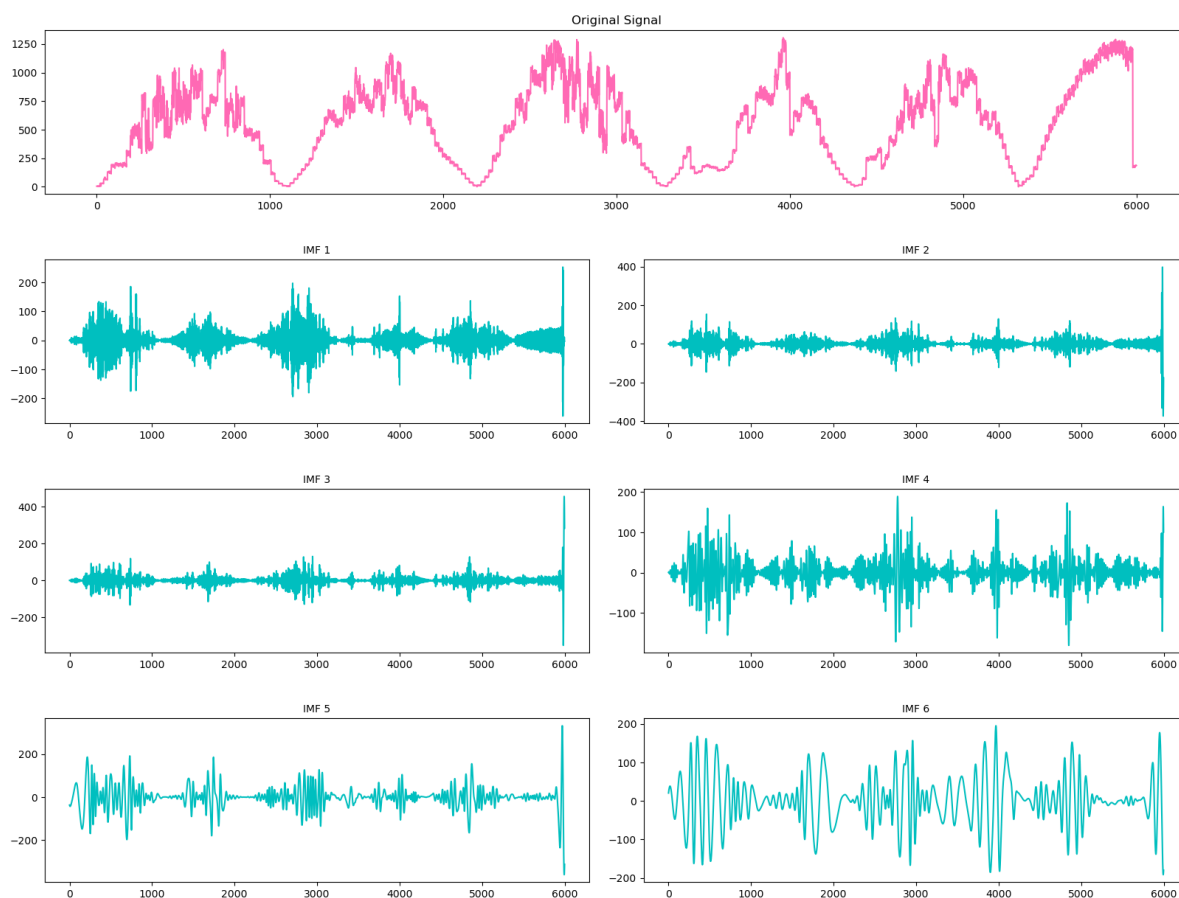


Figure 2. IMF decomposition

图 2. IMF 分解

然后使用麻雀优化算法进行参数寻优, 设置观察者和预警者概率分别为 0.2, 0.1, 利用 python 编程求得最优参数如表 2 所示。

Table 2. Parameter data

表 2. 参数信息

Parameter	value
BiLSTM_layers	3
BiLSTM_neurons_list	{51, 87, 106}
hidden_dim	408
num_layers	2
num_heads	2
learning_rate	0.0013750005425864223



### 4.3. 结果分析

通过 python 编程实现基于 ODE-BiLSTM 混合模型的结果, 如图 3 显示的是训练集和测试集在训练过程中的均方误差(MSE)损失值。根据图中的数据, 可以进行几点分析。从图中可以看到 MSE 总体呈现下降趋势, 训练损失和测试损失都有明显的下降趋势, 特别是在前 10 个 epoch。这表明模型在学习过程中逐渐适应了数据。但是测试损失的波动比训练损失大。这种现象通常表明模型对训练数据过拟合, 对于未见过的测试数据则不够泛化。不过, 在这个数据集中, 尽管测试损失有波动, 整体趋势仍然是下降的, 这是一个好的迹象。在大约 30 个 epoch 之后, 训练损失和测试损失趋于平稳, 但测试损失的波峰在后期有所上升。同时给出与基线模型: ARIMA 和单一 LSTM 模型以及随机森林模型的预测结果对比, 见表 3。显然本文提出的 SSA 优化的 ODE-BiLSTM 模型在三个指标均取得较好的效果。

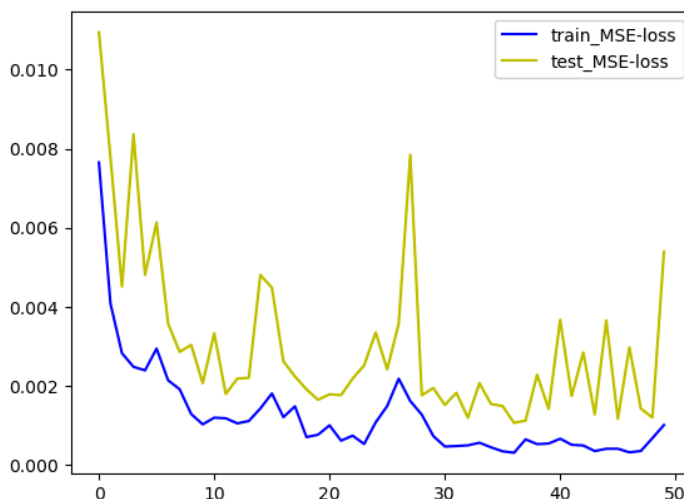


Figure 3. Training iteration  
图 3. 训练迭代

Table 3. Comparison of results between the ODE-BiLSTM model and baseline models  
表 3. ODE-BiLSTM 模型与基线模型对比

数量	MAE	RMSE	R <sup>2</sup>
ARIMA	419.33	424.82	0.4565
LSTM	287.69	309.94	0.4982
Random Forest	233.92	241.79	0.5234
ODE-BiLSTM	106.38	116.78	0.7527

## 5. 总结

基于SSA优化的太阳能发电混合模型功率预测研究在本文中得到了深入探讨和验证。首先, 通过改进的完全经验模态分解(ICEEMDAN)技术对数据进行预处理, 将其分解成多个固有模式函数(IMF)和残差, 极大地提高了数据的预测准确性。然后, 利用麻雀搜索算法(SSA)优化的ODE-LSTM模型对预处理后的数据进行建模。通过SSA的觅食和预警行为, 成功找到了模型的最佳超参数组合, 使预测模型在准确性和鲁棒性方面得到了显著提升。模型结果分析表明, 本文提出的SSA优化的ODE-BiLSTM混合模型在均方误差(MSE)、平均绝对误差(MAE)和判定系数(R<sup>2</sup>)等指标上均优于传统ARIMA、单一LSTM及随机森林模型, 展现出更高的预测精度和稳定性。本研究证实了混合模型和优化算法在太阳能发电预测领域的

潜力, 为实现智能电网中新能源功率的高效管理提供了有力支持。未来可以考虑进一步引入更多优化算法, 并结合实际情况, 进一步提高模型的泛化能力和适应性。

## 基金项目

江苏省高等学校大学生创新创业训练计划项目(No. 202311117106Y)。

江苏高校品牌专业建设工程资助项目(数学与应用数学, PPZY2015B109)。

## 参考文献

- [1] 刘滢茹, 李明媛, 宋炜. 双碳约束、能源消费结构与全要素能源效率[J]. 统计与决策, 2024, 40(2): 83-88.
- [2] 舒印彪. 发展新型电力系统助力实现“双碳”目标[J]. 中圆电力企业管理, 2021(7): 8-9.
- [3] 李芬, 陈正洪, 成驰, 段善旭. 太阳能光伏发电量预报方法的发展[J]. 气候变化研究进展, 2011, 7(2): 136-142.
- [4] 周德佳, 赵争鸣, 吴理植, 袁立强, 孙晓瑛. 其于仿真模型的太阳能光伏电池阵列特性的分析[J]. 清华大学学报(自然科学版), 2007(7): 1109-1112, 1117.
- [5] Perpiñan, O., Lorenzo, E. and Castro, M.A. (2006) On the Calculation of Energy Produced by a PV Grid-Connected System. *Progress in Photovoltaics: Research and Applications*, **15**, 265-274. <https://doi.org/10.1002/pip.728>
- [6] 朱琼峰, 李家腾, 乔颖, 史梦洁, 王朝亮. 人工智能技术在新能源功率预测的应用及展望[J]. 中国电机工程学报, 2023, 43(8): 3027-3048.
- [7] Behera, M.K., Majumder, I. and Nayak, N. (2018) Solar Photovoltaic Power Forecasting Using Optimized Modified Extreme Learning Machine Technique. *Engineering Science and Technology, an International Journal*, **21**, 428-438. <https://doi.org/10.1016/j.jestch.2018.04.013>
- [8] Liu, Y. (2022) Short-term Prediction Method of Solar Photovoltaic Power Generation Based on Machine Learning in Smart Grid. *Mathematical Problems in Engineering*, **2022**, Article ID: 8478790. <https://doi.org/10.1155/2022/8478790>
- [9] Zhuang, W., Li, Z., Wang, Y., Xi, Q. and Xia, M. (2024) Gen-informer: A Novel Framework for Mid-Term Photovoltaic Power Forecasting. *Applied Sciences*, **14**, Article 2181. <https://doi.org/10.3390/app14052181>
- [10] Colominas, M.A., Schlotthauer, G. and Torres, M.E. (2014) Improved Complete Ensemble EMD: A Suitable Tool for Biomedical Signal Processing. *Biomedical Signal Processing and Control*, **14**, 19-29. <https://doi.org/10.1016/j.bspc.2014.06.009>
- [11] 陈路倩. 基于麻雀搜索算法优化 LSTM 的超短期风速预测[D]: [硕士学位论文]. 南昌: 江西财经大学, 2022.
- [12] Slowik, A. and Kwasnicka, H. (2020) Evolutionary Algorithms and Their Applications to Engineering Problems. *Neural Computing and Applications*, **32**, 12363-12379. <https://doi.org/10.1007/s00521-020-04832-8>
- [13] Petalas, Y.G., Parsopoulos, K.E. and Vrahatis, M.N. (2007) Memetic Particle Swarm Optimization. *Annals of Operations Research*, **156**, 99-127. <https://doi.org/10.1007/s10479-007-0224-y>
- [14] Mirjalili, S. and Lewis, A. (2016) The Whale Optimization Algorithm. *Advances in Engineering Software*, **95**, 51-67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
- [15] Lechner, M. and Hasani, R. (2020) Learning Long-Term Dependencies in Irregularly-Sampled Time Series. arXiv: 2006.04418.