

DOI:10.3876/j.issn.1000-1980.2025.06.020

基于分时段 CDTW 选取相似日和 CSSA 优化 LSTM 的短期光伏功率预测模型

王 瑞¹, 杨源浩¹, 逯 静²

(1. 河南理工大学电气工程与自动化学院, 河南 焦作 454003; 2. 河南理工大学计算机科学与技术学院, 河南 焦作 454003)

摘要: 针对光伏功率预测模型中直接使用历史数据集作为输入数据难以满足预测精度的问题, 提出了一种基于分时段卷积动态时间规整 (CDTW) 选取相似日和混沌麻雀搜索算法 (CSSA) 优化长短期记忆网络 (LSTM) 的短期光伏功率预测模型 (CSSA-LSTM 模型)。该模型在数据预处理后, 通过皮尔逊相关性分析选出影响功率的重要气象因素, 避免不相关数据的干扰; 根据待测日总辐射序列进行时段划分, 对各时段采用 CDTW 进行相似时段的选取并重构为相似日集, 同时利用 CSSA 对 LSTM 模型进行超参数寻优, 自适应搜寻最佳超参数。我国南部地区某光伏电站 2021 年的实测数据仿真分析结果表明, 相比普通相似日和采用麻雀搜索算法优化参数的 LSTM 模型, 本文提出的相似日选取方法和 CSSA-LSTM 模型具有较高的预测精度和鲁棒性。

关键词: 光伏功率; 相似日; 长短期记忆网络; 动态时间规整; 卷积神经网络; 混沌麻雀搜索算法

中图分类号: TM615

文献标志码: A

文章编号: 1000-1980(2025)06-0166-09

A short-term photovoltaic power prediction model based on time-segmented CDTW for selecting similar days and CSSA-optimized LSTM

WANG Rui¹, YANG Yuanhao¹, LU Jing²

(1. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454003, China;

2. School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China)

Abstract: To address the issue that directly using historical datasets as input in photovoltaic (PV) power prediction models struggles to meet the required prediction accuracy, a short-term PV power prediction model, namely CSSA-LSTM model was proposed based on time-segmented convolutional dynamic time warping (CDTW) for similar day selection and the chaotic sparrow search algorithm (CSSA) for long short-term memory (LSTM) optimization. After preprocessing the data, the model identified key meteorological factors affecting power generation through Pearson correlation analysis, thereby avoiding interference from irrelevant data. It then divided time periods based on the total solar radiation sequence of the day to be predicted, applied the CDTW algorithm to each time period for selecting similar time segments, and reconstructed these segments into a set of similar days. Meanwhile, CSSA was used to optimize the hyperparameters of the LSTM model, enabling adaptive search for the optimal hyperparameters. Simulation analysis results using measured data from 2021 of a PV power station in southern China show that compared with traditional similar day methods and LSTM models optimized by the sparrow search algorithm (SSA), the similar day selection method and CSSA-LSTM model proposed in this study exhibit higher prediction accuracy and robustness.

Key words: photovoltaic power; similar day; long short-term memory network; dynamic time warping; convolutional neural network; chaotic sparrow search algorithm

太阳能作为一种安全、清洁无污染的可再生能源得到了愈来愈多的关注^[1]。由于气象和环境条件的不确定性,光伏发电系统具有较大的波动性,对电能并网运行和电网调度具有极大的影响^[2]。因此精确的光

基金项目: 国家自然科学基金项目 (62273133); 河南省科技攻关项目 (222102210120)

作者简介: 王瑞 (1977—), 男, 副教授, 硕士, 主要从事电力系统分析研究。E-mail: wangrui@hpu.edu.cn

通信作者: 逯静 (1980—), 女, 副教授, 硕士, 主要从事深度学习和物联网技术研究。E-mail: lujing@hpu.edu.cn

引用本文: 王瑞, 杨源浩, 逯静. 基于分时段 CDTW 选取相似日和 CSSA 优化 LSTM 的短期光伏功率预测模型[J]. 河海大学学报(自然科学版), 2025, 53(6): 166-174.

WANG Rui, YANG Yuanhao, LU Jing. A short-term photovoltaic power prediction model based on time-segmented CDTW for selecting similar days and CSSA-optimized LSTM[J]. Journal of Hohai University(Natural Sciences), 2025, 53(6): 166-174.

伏功率预测技术对于维护电网系统安全和稳定、为电网调度提供参考依据具有重要的现实意义和研究价值^[3-4]。

目前,国内外学者对光伏发电功率预测已开展大量研究,根据预测模型的不同,光伏发电功率预测方法主要有物理方法和统计方法。物理方法是按照光伏电池发电原理,利用数值天气预报的气象数据结合光伏系统的参数,通过建立物理模型直接计算得到光伏发电功率。物理方法依赖详细的地理信息和气象数据,模型抗干扰能力差,鲁棒性不强。统计方法是通过发电站的历史数据,使用不同算法建立输入与输出的映射模型来实现对未来光伏发电功率的预测,具有建模简单的特点^[5-6],常见的预测模型有反向传播神经网络^[7]、最小二乘支持向量机^[8]、极限学习机^[9]、长短时记忆网络(long short-term memory, LSTM)^[10]等模型。

当前学者们主要采用统计方法研究历史数据间的映射关系来实现功率预测,而映射关系的相关性往往对模型的预测精度有很大影响,这使得优化模型的训练集质量成为提升预测精度的关键措施之一,光伏发电功率有着很强的日周期性,因此考虑构建相似日集便成为一种常用且有效的方法。孟安波等^[11]通过结合皮尔逊相关系数与模糊 C 均值(fuzzy C-means, FCM)聚类的加权模糊聚类方法选取相似日,考虑了不同气象因素的权值,但其聚类中心是基于所有的数据样本进行迭代更新的,未充分考虑待测日本身的影响,具有一定的局限性。杨茂等^[12]构建了基于马氏距离寻找相似日进行光伏功率预测的模型,以待测日本身为主体,且克服了欧氏距离缺乏考虑变量相关性的缺点,对提升预测精度有一定的作用,但马氏距离是数据间的空间距离,未考虑数据中的冗余信息。张慧娥等^[13]通过灰色关联分析选取相似日,然后搭建模型进行光伏功率预测,这一方法虽然保证了历史日与待测日数据曲线上的相似性,但未考虑天气条件具备的突变性,难以兼顾不同时段的变化。

近年来,基于 LSTM 的预测模型已经广泛应用于光伏功率预测领域。相较于其他如反向传播神经网络、线性回归、支持向量机和极限学习机等模型, LSTM 模型具有更好的预测精度,但存在模型参数需要手动调整和调整困难的问题,而采用启发式算法优化模型参数具有良好的效果。麻雀搜索算法(sparrow search algorithm, SSA)^[14]因其具有较好的稳定性和寻优能力而被广泛应用于模型参数优化,但存在陷入局部最优化的风险。

针对现有相似日方法缺乏对冗余信息和天气转变的考虑而影响预测精度的问题,本文构建了基于卷积动态时间规整(convolutional dynamic time warping, CDTW)选取相似日、采用混沌麻雀搜索算法(chaotic sparrow search algorithm, CSSA)优化 LSTM 的短期光伏功率预测模型(CSSA-LSTM 模型),通过引入 CDTW 筛选相似日集,采用经 CSSA 优化参数后的 LSTM 模型进行训练预测,从而提高模型预测光伏功率的精度。

1 相似日选取方法

1.1 气象特征选择

相似日是通过历史日和待测日气象特征数据的相似性选取的,但影响光伏功率的因素有很多,且对功率的影响程度各不相同,使用相关性强的因素能更准确地捕捉功率与影响因素之间的关系。本文采用皮尔逊相关系数法选择重要的气象因素^[15]。

本文采用的气象数据包括总辐射、散射辐射、温度、湿度和气压数据,与发电功率的皮尔逊相关系数分别为 0.81、0.81、0.56、-0.11 和 -0.11,可见总辐射和散射辐射是相关性最强的因素,温度中等相关,湿度与气压的相关性最低。故选择将总辐射、散射辐射和温度作为相似日选取和模型训练的气象因素。

1.2 待测日时段划分

普通相似日的选取是以整个待测日序列为对象寻找相似日,时间跨度较长,但同一天不同时段内的天气条件会有着不同程度的变化,这会直接影响相似日的选取结果,直接以一日的时间跨度寻找的相似日难以兼顾不同时段的相似度。故本文采用以时段划分的方式进行相似日的数据筛选,以减少天气条件变化对所寻相似日数据集的影响,更能体现与待测日的相似度。

假设时间序列为 $X = \{x_1, x_2, \dots, x_n\}$,分段后可表示为

$$X = \begin{cases} f_1(t, w_1) + e_{1t} & 1 \leq t \leq t_1 \\ f_2(t, w_2) + e_{2t} & t_1 < t \leq t_2 \\ \vdots & \vdots \\ f_k(t, w_k) + e_{kt} & t_{k-1} < t \leq n \end{cases} \quad (1)$$

式中: $w_1, w_2, \dots, w_k$ 为时间区间的端点坐标; $f_1(t, w_1), f_2(t, w_2), \dots, f_k(t, w_k)$ 为时间区域内的线性函数; $e_{1t}, e_{2t}, \dots, e_{kt}$ 为对应分段与实际序列之间的误差; t 为分段后的函数自变量。

时段划分采用自上而下算法,该算法是一种对时间序列进行自适应分段的算法,其工作原理是考虑时间序列的每一种可能的划分,并在最佳位置将其拆分为两段,如两段的近似误差低于指定的阈值,则算法递归地继续拆分时间序列,直到所有时段的误差都低于阈值。

因为温度不具有突变性,难以及时反映天气的波动情况,而总辐射与散射辐射之间又具有强相关性,其规律基本一致,故通过自上而下算法对待测日的总辐射序列进行递归分段,分段点表示序列趋势开始发生转变,记录各个分段点,进而分别对历史日中各对应时段的数据计算序列相似度。

1.3 序列相似度计算

传统相似度计算方法为直接对原始时间序列进行点对点的相似度计算,难以体现序列局部特征信息的相似度。本文采用由卷积神经网络(convolutional neural network, CNN)和动态时间规整(dynamic time warping, DTW)组合而成的 CDTW 相似度计算方法计算序列相似度。CNN 具有过滤干扰、强化特征的作用,适合在比较相似度之前提取数据特征和过滤冗余信息,减少在相似度比较过程中的干扰^[16-17]。通过 CNN 将待测日气象特征和历史日气象特征卷积和池化后的序列作为 DTW 的输入数据进而得到代表相似度的规整路径累计距离(CDTW 距离)^[18-19]。CDTW 计算流程如图 1 所示。

结合相关性分析,将影响功率的主要因素数据归一化处理后分别通过 CDTW 计算历史日与待测日各个特征之间的相似度距离,并进行加权求和得到最终的 CDTW 距离:

$$D = V_1 D_1 + V_2 D_2 + V_3 D_3$$
$$V_1 = G_1/G \quad V_2 = G_2/G \quad V_3 = G_3/G$$

其中

式中: D 为经过加权后的最终用来比较历史日与待测日相似度的 CDTW 距离; V_1, V_2, V_3 为总辐射、散射辐射和温度根据相关系数分配的权重; D_1, D_2, D_3 为总辐射、散射辐射和温度对应相似度的 CDTW 距离; G_1, G_2, G_3 分别为总辐射、散射辐射和温度与光伏功率的相关性系数, $G = G_1 + G_2 + G_3$ 。

1.4 相似日构建

相似日构建步骤如下:

- a. 对数据进行归一化处理,消除影响因素量纲的影响,通过皮尔逊相关性分析选取与光伏功率相关性强的气象因素。
- b. 将待测日气象特征序列划分为多个时段,对各个时段分别进行相似时段的选取,减少一日中因气象条件变化而导致序列相似度降低的问题。
- c. 对待测日和历史数据集中相同时段的序列进行 CDTW 距离计算并排序。
- d. 对排序后不同时段的数据进行时间尺度上的拼接,重构数据集,在相似日集中选取相似度最高的前 15 日作为相似日集。

2 CSSA-LSTM 模型

2.1 LSTM

LSTM 是在循环神经网络(recurrent neural network, RNN)的基础上加以改进的神经网络,它主要解决了 RNN 处理长序列数据时易出现梯度爆炸或梯度消失的问题^[20-21]。两者结构相似,均由输入层、隐藏层和输出层构成,LSTM 通过引入门控机制来控制信息的累积速度,可有选择性地添加新的信息或遗忘部分信息,基于此来改善 RNN 的长期依赖问题。LSTM 的具体原理详见文献^[22-23]。

将通过皮尔逊相关性分析筛选后的气象特征和光伏功率组成的向量作为 LSTM 模型的输入数据,光伏功率作为输出数据,采用单步滚动预测进行模型的训练和预测,以前 m 个时刻点的气象数据和功率数据作为模型的训练输入数据、 $m+1$ 时刻点的功率作为训练输出数据的方式进行标签划分,以预测值和真实值的均方根误差(RMSE)作为损失函数进行迭代训练,经过大量仿真,设 $m=7$ 时预测效果最好。

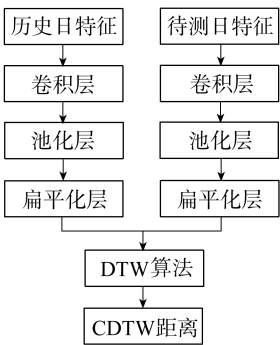


图 1 CDTW 计算流程
Fig. 1 CDTW calculation process

2.2 CSSA

LSTM 模型在进行预测训练时,所设超参数对模型训练的预测结果影响较大,若人为设置超参数,只能凭借经验来完成,耗费精力且难以设置一个最佳的参数,故采用人工智能算法优化 LSTM 模型的超参数。

SSA 是受自然界中麻雀觅食过程的启发而提出的一种智能优化算法,具有寻优能力强、收敛速度快和稳定性好的优点。在觅食过程中,麻雀种群主要分为发现者和加入者,并带有侦察预警机制。但是 SSA 存在易陷入局部最优的问题,故本文在 SSA 的基础上引入了 Tent 混沌搜索和高斯变异,可提高初始解的质量,增加种群的多样性,增强算法的全局搜索能力^[24]。

- CSSA 优化 LSTM 模型超参数步骤如下:
- a. 初始化 CSSA 参数,设置 LSTM 模型待优化超参数的范围,并选择 RMSE 作为 LSTM 模型的适应度函数。
 - b. 利用 Tent 混沌序列初始化种群。
 - c. 计算适应度值,并记录最优适应度和最劣适应度对应的位置。
 - d. 更新发现者和加入者位置,通过侦察预警机制再次更新位置。
 - e. 一次迭代结束,计算每只麻雀适应度值 f_i 和平均适应度值 f_{avg} 。
 - f. 若 $f_i < f_{avg}$,则对其进行高斯变异;若 $f_i \geq f_{avg}$,则对其进行混沌扰动。
 - g. 计算适应度值,求得最优位置。
 - h. 判断是否满足算法结束条件,若满足,输出最优结果;若不满足,则返回步骤 d。
 - i. 将 CSSA 迭代的最优结果作为 LSTM 模型的超参数设置,构建预测模型。

2.3 模型构建及预测流程

基于分时段 CDTW 选取相似日并构建 CSSA-LSTM 模型的具体流程如图 2 所示。

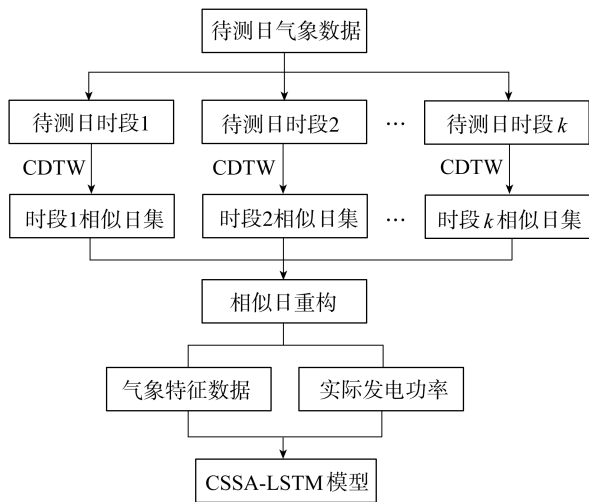


图 2 分时段 CDTW 选取相似日与 CSSA-LSTM 模型构建流程

Fig.2 Similar day selection via time-segmented CDTW and construction of CSSA-LSTM model

- 模型预测步骤如下:
- a. 对数据进行预处理,分析相关系数进行特征选择。
 - b. 根据气象序列划分时段,然后通过 CDTW 选取各时段的相似时段并重构为相似日数据集。
 - c. 构建 LSTM 模型并使用 CSSA 优化其超参数。
 - d. 使用相似日集的气象特征数据和发电功率数据训练模型并对待测日作出预测。
 - e. 对预测结果进行分析。

3 算例验证

3.1 数据简介

采用我国南方某光伏发电厂的历史数据集作为测试对象,该数据集包含 2021 年全年的实际发电功率数据和每日的总辐射、散射辐射、温度、湿度和气压数据。数据采样间隔为 15 min,全天共 96 个数据,考虑到光

光伏发电主要时间段为白天,故采用每天 7:00—18:45 共 48 个数据作为仿真数据。对历史数据中不同的天气类型各随机选取一天作为待测日,将根据待测日筛选的相似日集作为该日的训练集,分别对 10 月 30 日(晴)、10 月 26 日(多云)和 11 月 21 日(阴雨)进行建模仿真,并通过多次预测取平均值来对 CSSA-LSTM 模型的有效性进行验证分析,采用 RMSE、平均绝对误差(MAE)和决定系数(R^2)作为评价指标来评价模型的预测效果。

3.2 相似日对比分析

根据第 1.2 节方法,采用总辐射序列进行时段划分,结果如图 3 所示,红色曲线为总辐射原数据,蓝色折线为自上而下算法分段后的效果,其拐点为分段时刻点。10:00 之前,数值增长较快,代表上午太阳升高的阶段;10:00—11:30 增长幅度开始放缓,对应着临近正午的时段;11:30—13:00 辐射度开始下降,13:00—17:00 辐射度下降速度变快;17:00 之后辐射度为 0,表示日落后的黄昏阶段。不同阶段代表着不同的气象规律,故按以上时间点进行时段划分,并对所划分时段进行 CDTW 相似度的比较选取。

为更直观地展示本文相似日选取方法中 CNN 和时段划分的作用,将 DTW、CDTW 和分时段 CDTW 3 种方法选取的相似日(以 3 个相似日为例)进行对比分析,结果如图 4 所示。

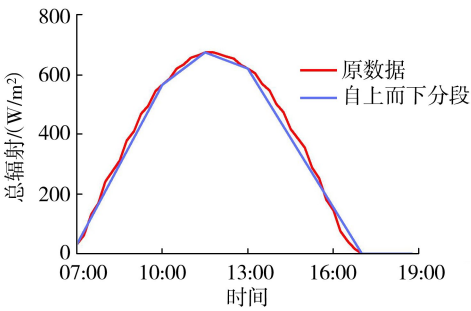


图 3 自上而下算法划分时段
Fig. 3 Top-down algorithm for dividing time periods

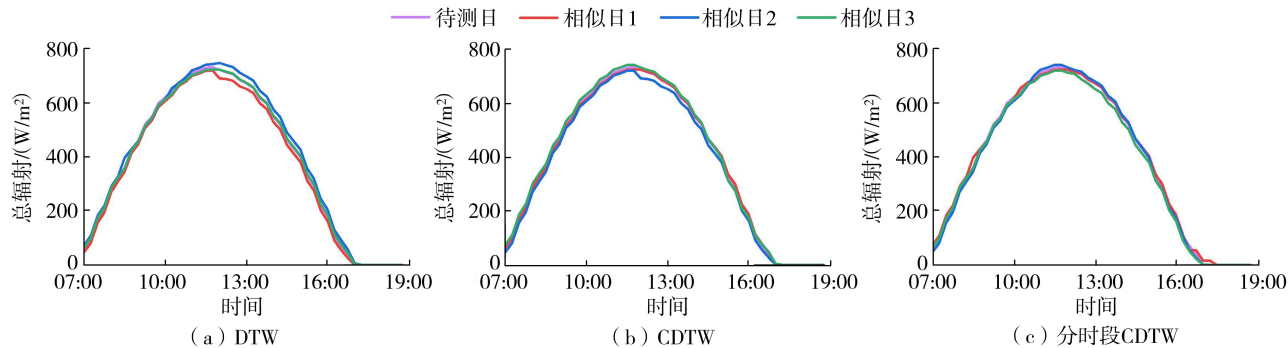


图 4 不同方法选取的相似日

Fig. 4 Similar day selection results of different methods

图 4 中 3 个相似日按照与待测日计算的相似度从大到小排序为相似日 1、相似日 2、相似日 3,总辐射作为最主要影响因素,其相似情况基本能代表整体情况。由图 4(a)可以看出,在 DTW 选取的相似日中,相似日 2 和相似日 3 反而比相似日 1 更加贴合待测日的曲线,这是因为该算法并没有严格在相同时刻进行距离比较而导致的,DTW 提高了序列比较的灵活性但也一定程度降低了相似度。在对特征进行 CNN 提取后进行 DTW 相似日选取,相似日 1 更加贴合待测日的总辐射曲线,曲线相似程度有了很大提升,这得益于 CNN 卷积核对序列的卷积效果,相对于 DTW,CDTW 提高了对局部特征的敏感度并且扩展了原来的时间序列,可以避免 DTW 寻找最短路径时两个不同时刻的间隔过大;在此基础上引入分时段,分时段 CDTW 对不同规律的时段单独进行相似度的选取并重构相似日,其相似日选择效果也因此而得到提升。

以 LSTM 模型为预测模型,采用不同方法选取相似日的模型预测结果曲线和误差对比如图 5 和表 1 所示,由图 5 可以看出,采用 DTW 选取的相似日数据训练模型,较直接预测的精度有明显的提升,原因是通过相似日数据的选取,加强了训练模型所用数据与待测日数据之间的相关性。而采用 CDTW 选取相似日更是为了强化这种相关性,通过 CNN 突出历史数据中的特征,进行待测日信息的特征与历史数据信息特征的比对选取更具相关性的相似日数据,其 RMSE 和 MAE 较直接预测分别降低了 9.425%和 20.804%。在 CDTW 方法的基础上,分时段 CDTW 通过对一天的时间序列进行分段,对每个时段分别选取相似时段,再构建相似日数据集,可以更好应对一天中某一时段天气骤变的情况,其 RSME 和 MAE 较 CDTW 方法分别下降了 7.466%和 9.396%,验证了本文相似日选取方法的有效性。

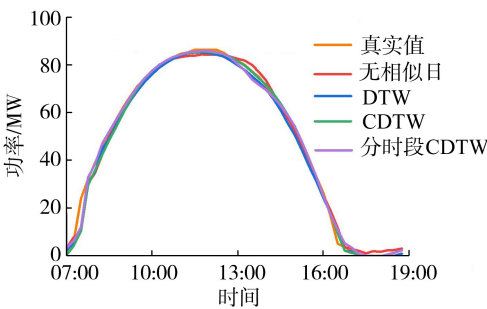


图 5 LSTM 模型晴天预测曲线

Fig. 5 Prediction curves of LSTM model under sunny condition

3.3 优化算法对比分析

为保证仿真对比的严谨性,SSA 和 CSSA 的参数设置保持一致,设置种群数为 20,迭代次数为 30,发现者比例为 0.2,警戒者比例为 0.2,预警值为 0.8;LSTM 模型超参数寻优范围设置如下:学习率为 0.001~0.01,第一、第二层隐藏节点数均为 1~100。不同寻优算法的收敛曲线如图 6 所示,可以看出,经过加入混沌扰动和高斯变异后的 CSSA 初始解质量更高,在第一次迭代时便能寻找到一个较优的位置,而后又能很快跳出局部最优,增加了全局搜索的能力。

采用分时段 CDTW 选取的相似日数据,将 SSA 和 CSSA 对 LSTM 模型 3 个超参数的寻优结果(学习率、第一和二层隐藏节点数分别为 0.002、15、84 和 0.006、45、81)分别代入 LSTM 模型,得到 SSA-LSTM 模型和 CSSA-LSTM 模型并进行预测对比分析,结果如图 7 和表 2 所示。由图 7 和表 2 可以看出,通过 SSA 寻优超参数可以更好地激发 LSTM 模型性能,而采用 CSSA 优化 LSTM 模型超参数较 SSA 有了进一步的提升,CSSA-LSTM 模型比 SSA-LSTM 模型的 RMSE 和 MAE 分别降低了 7.528%和 6.044%,较 LSTM 模型的 RMSE 和 MAE 分别降低了 16.180%和 22.332%,表明混沌扰动和高斯变异可以有效避免算法局部最优的情况,有利于寻找到更优的超参数让模型发挥更好的预测能力。

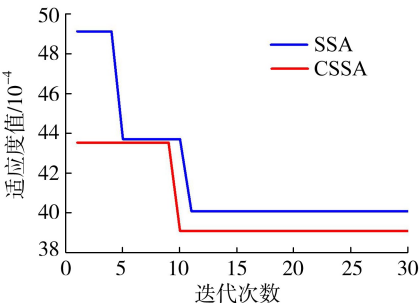


图 6 优化算法寻优 LSTM 模型超参数适应度值

Fig. 6 Fitness value in LSTM model hyperparameter optimization by optimization algorithm

3.4 不同天气类型泛化性分析

为验证本文模型(即 CSSA-LSTM 模型)在不同天气类型下的有效性,以 10 月 26 日(多云)和 11 月 21 日(阴雨)为待测日进行仿真分析,图 8 和表 3 分别为两种天气类型的预测曲线和预测误差。

由图 8 可以看出,多云和阴雨天的预测效果明显不如晴天,这是因为曲线波动幅度变大而增加了预测难度,根据预测结果来看,利用分时段 CDTW 选取的相似日进行模型训练和预测的效果最好,这是因为经过卷积提取局部特征后比较相似度,能够兼顾临近点对相似度的影响,从而可以降低部分数据因存在干扰信息对相似度比较的影响。而通过分时段可以缩短比较序列的长度,在天气条件发生转变的时候,可以针对各时段的气象情况分别在各时段选出更优的相似区间,提高复杂天气下所选相似日的相似度。如表 3 所示,在多云天气下, RMSE 和 MAE 较普通相似日分别降低了 4.762%和 4.416%;在阴雨天气下, RMSE 和 MAE 较普通相

表 1 LSTM 模型晴天预测效果对比

Table 1 Comparison of prediction effects of LSTM model under sunny condition

相似日选取方法	RMSE/MW	MAE/MW	R^2
无相似日	2.780	1.841	0.992
DTW	2.638	1.745	0.993
CDTW	2.518	1.458	0.994
分时段 CDTW	2.330	1.321	0.995

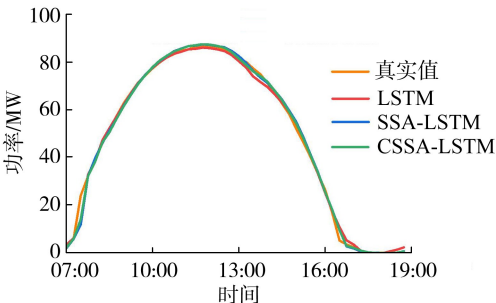


图 7 不同寻优算法模型预测曲线对比

Fig. 7 Comparison of prediction curves of different optimization algorithm models

表 2 不同寻优算法模型预测效果对比

Table 2 Comparison of prediction effects of different optimization algorithm models

模型	RMSE/MW	MAE/MW	R^2
LSTM	2.330	1.321	0.995
SSA-LSTM	2.112	1.092	0.996
CSSA-LSTM	1.953	1.026	0.997

似日分别降低了 6.421%和 5.099%。在此基础上结合 CSSA 优化 LSTM 模型超参数提高模型的预测能力,预测精度进一步得到提高,验证了 CSSA-LSTM 模型在不同天气类型下具有同样的适用性。

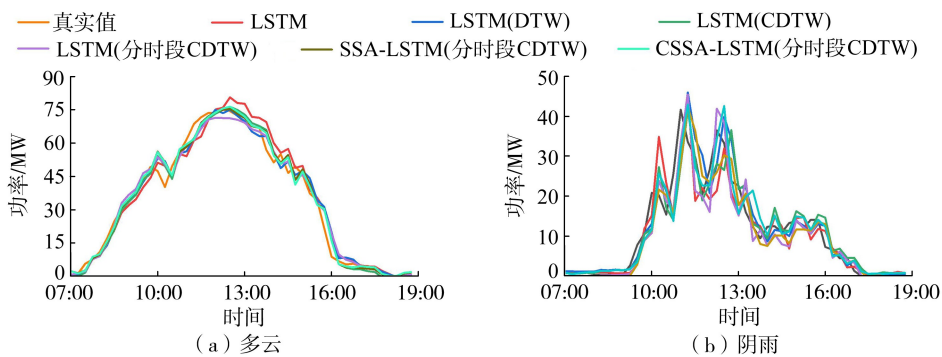


图 8 不同天气预测曲线对比

Fig. 8 Comparison of prediction curves under different weather conditions

表 3 不同天气预测效果对比

Table 3 Comparison of prediction effects under different weather conditions

天气	模型	相似日选取方法	RMSE/MW	MAE/MW	R^2
多云	LSTM	无相似日	4.963	3.642	0.965
	LSTM	DTW	4.032	2.853	0.977
	LSTM	CDTW	3.855	2.866	0.979
	LSTM	分时段 CDTW	3.840	2.727	0.979
	SSA-LSTM	分时段 CDTW	3.828	2.653	0.980
	CSSA-LSTM	分时段 CDTW	3.785	2.644	0.982
阴雨	LSTM	无相似日	5.242	3.484	0.789
	LSTM	DTW	4.797	3.236	0.824
	LSTM	CDTW	4.629	3.163	0.836
	LSTM	分时段 CDTW	4.489	3.071	0.846
	SSA-LSTM	分时段 CDTW	4.313	3.001	0.858
	CSSA-LSTM	分时段 CDTW	4.260	2.977	0.861

3.5 不同相似日选取方法对比分析

为了进一步讨论本文分时段 CDTW 选取相似日方法对提升模型预测精度的有效性,与 FCM 聚类选取相似日和灰色关联度选取相似日进行对比分析,结果如图 9 和表 4 所示。由图 9 和表 4 可以看出,晴天的数据规律性强,不同相似日选取方法均有着良好的预测精度,由于分时段 CDTW 引入了 CNN 和分时段的处理方法,其选取的相似日集使得预测效果能进一步得到提升, RMSE 和 MAE 较 FCM 聚类分别降低了 20.867%和 42.777%,较灰色关联度分别降低了 16.288%和 37.969%。多云天气数据的波动性变大,使得预测难度提高,但分时段 CDTW 仍能优化相似日集进而提升预测精度, RMSE 和 MAE 较 FCM 聚类分别降低了 14.289%和 25.897%,较灰色关联度分别降低了 8.905%和 14.902%。阴雨天数据波动性更加剧烈,在 10:00—13:00 时段内数据波动较大,在 13:00—16:00 时段内波动幅度变小,明显存在着波动性不同的特点,通过分时段

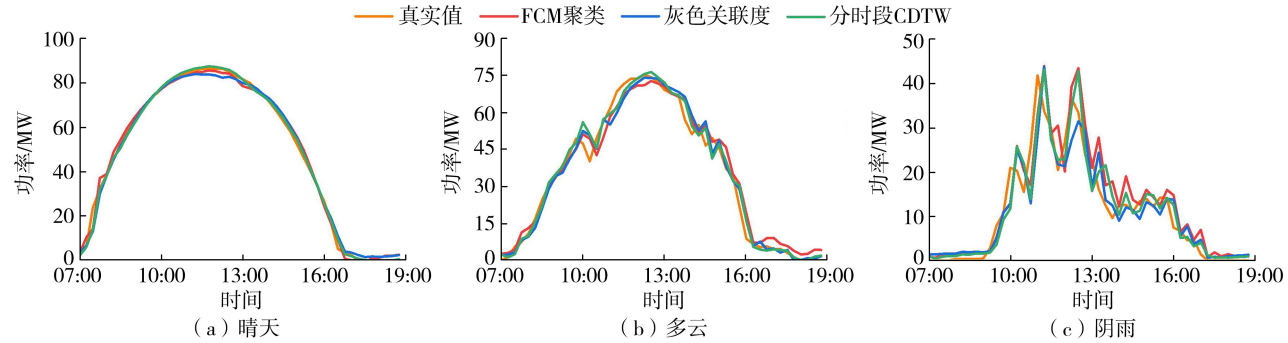


图 9 不同相似日选取方法预测曲线对比

Fig. 9 Comparison of prediction curves under different similar day selection methods

CDTW 选取相似日可以兼顾不同时段不同波动的特点来提升相似日集的质量进而提高预测精度, RMSE 和 MAE 较 FCM 聚类分别降低了 10.972% 和 14.331%, 较灰色关联度分别降低了 5.773% 和 4.919%。可见不同天气下的预测精度均可以得到有效提升, 验证了本文相似日选取方法的有效性。

3.6 泛化性对比分析

为了进一步验证 CSSA-LSTM 模型的泛化性, 选取江苏某电站 11 月连续 7 天数据进行对比分析, 结果如图 10 所示, 可见分时段 CDTW 选取相似日可以有效减小预测误差, 在此基础上通过 CSSA 优化 LSTM 模型超参数后的预测精度可以得到进一步提升, 验证了在不同数据集集中, CSSA-LSTM 模型均能有效降低误差, 具有较好的预测性能。

4 结 语

针对构建相似日集时对历史数据气象特征和不同时段的气象规律利用有限的问题, 本文提出了基于分时段 CDTW 选取相似日和 CSSA 优化 LSTM 的短期光伏功率预测模型 (CSSA-LSTM 模型), 实例分析表明, 引入 CDTW 选取相似日可有效利用 CNN 特征提取特点, 结合 DTW 对时间序列相似度的比较, 提高待测日与历史日数据的特征比较能力, 更加有效地体现序列的相似度; 采用分时段的方法筛选相似日, 将不同时段具有不同气象特征的现象分别进行相似时段选取, 可以应对短时间内天气突变的情况, 提高相似日数据的合理性; 采用 CSSA 优化 LSTM 模型超参数, 提高了模型的预测性能。分时段 CDTW 选取相似日和 CSSA-LSTM 模型在不同气候环境下预测精度相比普通相似日及 LSTM 模型都有较大的提升, 具有良好的预测能力和鲁棒性。

参考文献:

[1] 吴硕. 光伏发电系统功率预测方法研究综述[J]. 热动力工程, 2021, 36(8): 1-7. (WU Shuo. Review of power forecasting methods of photovoltaic power generation system[J]. Journal of Engineering for Thermal Energy and Power, 2021, 36(8): 1-7. (in Chinese))

[2] 李军徽, 张嘉辉, 李翠萍, 等. 参与调峰的储能系统配置方案及经济性分析[J]. 电工技术学报, 2021, 36(19): 4148-4160. (LI Junhui, ZHANG Jiahui, LI Cuiping, et al. Configuration scheme and economic analysis of energy storage system participating in grid peak shaving[J]. Transactions of China Electrotechnical Society, 2021, 36(19): 4148-4160. (in Chinese))

[3] LIU Hongfei, GAO Qian, MA Pengcheng. Photovoltaic generation power prediction research based on high quality context ontology and gated recurrent neural network[J]. Sustainable Energy Technologies and Assessments, 2021, 45: 101191.

[4] 郭怿, 明波, 黄强, 等. 考虑输电功率平稳性的水-风-光-储多能互补日前鲁棒优化调度[J]. 电工技术学报, 2023, 38(9): 2350-2363. (GUO Yi, MING Bo, HUANG Qiang, et al. Day-ahead robust optimal scheduling of hydro-wind-pv-storage complementary system considering the steadiness of power delivery[J]. Transactions of China Electrotechnical Society, 2023, 38(9): 2350-2363. (in Chinese))

[5] 朱琼锋, 李家腾, 乔骥, 等. 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): 3027-3047. (ZHU Qiongfang, LI Jiateng, QIAO Ji, et al. Application and prospect of artificial intelligence technology in renewable energy forecasting[J]. Proceedings of the CSEE, 2023, 43(8): 3027-3047. (in Chinese))

[6] LI Hongxia, LI Jianlin, MI Yang. Research on photovoltaic power prediction technology based on machine learning[J]. Journal of Physics: Conference Series, 2021, 2087(1): 012004.

[7] 朱诚, 王昭敏, 隆锋, 等. 基于 ABC-BP 神经网络的地铁盾构地表沉降预测[J]. 河海大学学报(自然科学版), 2023, 51(4): 72-80. (ZHU Cheng, WANG Zhaomin, LONG Feng, et al. Prediction of ground settlement of subway shield based on ABC-BP neural network[J]. Journal of Hohai University (Natural Sciences), 2023, 51(4): 72-80. (in Chinese))

表 4 不同相似日选取方法预测效果对比
Table 4 Comparison of prediction effects under different similar day selection methods

天气	相似日选取方法	RMSE/MW	MAE/MW	R^2
晴天	FCM 聚类	2.468	1.793	0.994
	灰色关联度	2.333	1.654	0.994
	分时段 CDTW	1.953	1.026	0.997
多云	FCM 聚类	4.416	3.568	0.972
	灰色关联度	4.155	3.107	0.976
	分时段 CDTW	3.785	2.644	0.982
阴雨	FCM 聚类	4.785	3.475	0.825
	灰色关联度	4.521	3.131	0.844
	分时段 CDTW	4.260	2.977	0.861

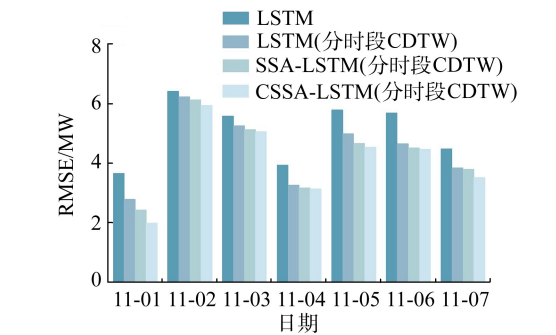


图 10 其他数据集预测误差对比
Fig. 10 Comparison of prediction errors across other datasets

- [8] 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. (ZHANG Tao,ZHU Ruijin,ZHAXI Dunzhu. Short-term photovoltaic power prediction based on IBBDE-LSSVM[J]. Thermal Power Generation,2021,50(5):102-107. (in Chinese))
- [9] 张红涛,韩婧,谭联,等.基于FCM和SSA-ELM的超短期风功率预测[J].工程科学与技术,2020,52(6):234-241. (ZHANG Hongtao,HAN Jing,TAN Lian,et al. Ultra-short-term wind power prediction based on combination of FCM and SSA-ELM[J]. Advanced Engineering Sciences,2020,52(6):234-241. (in Chinese))
- [10] 吉铎格,李慧,刘思嘉,等.基于MIE-LSTM的短期光伏功率预测[J].电力系统保护与控制,2020,48(7):50-57. (JI Xinge,LI Hui,LIU Sijia,et al. Short-term photovoltaic power forecasting based on MIE-LSTM[J]. Power System Protection and Control,2020,48(7):50-57. (in Chinese))
- [11] 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 Anbo,CHEN Jiaming,LI Zhanlian,et al. Short-term photovoltaic power generation prediction based on similar day theory and CSO-WGPR[J]. High Voltage Engineering,2021,47(4):1176-1184. (in Chinese))
- [12] 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. (YANG Mao,FENG Fan. Ultra-short-term prediction of PV power based on similar days of Mahalanobis distance[J]. Renewable Energy Resources,2021,39(2):175-181. (in Chinese))
- [13] 张慧娥,刘大贵,朱婷婷,等.基于相似日和动量法优化BP神经网络的光伏短期功率预测研究[J].智慧电力,2021,49(6):46-52. (ZHANG Huie,LIU Dagui,ZHU Tingting,et al. Short-term PV power prediction based on BP neural network optimized by similar daily and momentum method[J]. Smart Power,2021,49(6):46-52. (in Chinese))
- [14] XUE Jiankai,SHEN Bo. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering,2020,8(1):22-34.
- [15] 李斌,张一凡,颜世烨,等.基于改进极限学习机ELM的光伏发电预测方法研究[J].热能动力工程,2022,37(10):207-214. (LI Bin,ZHANG Yifan,YAN Shiye,et al. Research on photovoltaic power generation prediction method based on improved extreme learning machine[J]. Journal of Engineering for Thermal Energy and Power,2022,37(10):207-214. (in Chinese))
- [16] 汤德清,朱武,侯林超.基于CNN-LSTM-XGBoost模型的超短期光伏功率预测[J].电源技术,2022,46(9):1048-1052. (TANG Deqing,ZHU Wu,HOU Linchao. Ultra-short-term photovoltaic power prediction based on CNN-LSTM-XGBoost model[J]. Chinese Journal of Power Sources,2022,46(9):1048-1052. (in Chinese))
- [17] 裴艳宇,杨小彬,传金平,等.一维卷积神经网络特征提取下微震能级时序预测[J].工程科学学报,2021,43(7):1003-1009. (PEI Yanyu,YANG Xiaobin,CHUAN Jinping,et al. Time series prediction of microseismic energy level based on feature extraction of one-dimensional convolutional neural network[J]. Chinese Journal of Engineering,2021,43(7):1003-1009. (in Chinese))
- [18] 吕佳朋,史贤俊,肖支才.基于DTW时序距离的故障可诊断性评估方法[J].兵工学报,2024,45(3):997-1009. (LYU Jiapeng,SHI Xianjun,XIAO Zhicai. Diagnosability evaluation method based on DTW temporal distance[J]. Acta Armamentarii,2024,45(3):997-1009. (in Chinese))
- [19] 袁慧,谭章禄,王福浩.一种高效的相似性度量方法及其分类效果研究[J].中国科学:技术科学,2022,52(7):1096-1110. (YUAN Hui,TAN Zhanglu,WANG Fuhao. An efficient similarity measurement method and its classification effect[J]. Scientia Sinica: Technologica,2022,52(7):1096-1110. (in Chinese))
- [20] 刘雨佳,樊艳芳,白雪岩,等.基于特征交叉机制和误差补偿的风力发电功率短期预测[J].电工技术学报,2023,38(12):3277-3288. (LIU Yujia,FAN Yanfang,BAI Xueyan,et al. Short-term wind power prediction based on feature crossover mechanism and error compensation[J]. Transactions of China Electrotechnical Society,2023,38(12):3277-3288. (in Chinese))
- [21] 刘倩,胡强,杨凌帆,等.基于时间序列的深度学习光伏发电模型研究[J].电力系统保护与控制,2021,49(19):87-98. (LIU Qian,HU Qiang,YANG Lingfan,et al. Deep learning photovoltaic power generation model based on time series[J]. Power System Protection and Control,2021,49(19):87-98. (in Chinese))
- [22] 程井,孔垂穗,邹科辉.基于LSTM的泵闸工程混凝土施工期温度场预测[J].水利水电科技进展,2023,43(2):76-81. (CHENG Jing,KONG Chuisui,ZOU Kehui. Temperature field prediction during concrete construction period of pump and sluice project based on LSTM[J]. Advances in Science and Technology of Water Resource,2023,43(2):76-81. (in Chinese))
- [23] 周兰庭,邓思源,柳志坤,等.基于SSA-LSTM-GF的混凝土坝变形预测方法[J].河海大学学报(自然科学版),2023,51(2):73-80. (ZHOU Lanting,DENG Siyuan,LIU Zhikun,et al. Deformation prediction method of concrete dam based on SSA-LSTM-GF[J]. Journal of Hohai University (Natural Sciences),2023,51(2):73-80. (in Chinese))
- [24] 吕鑫,慕晓冬,张钧,等.混沌麻雀搜索优化算法[J].北京航空航天大学学报,2021,47(8):1712-1720. (LYU Xin,MU Xiaodong,ZHANG Jun,et al. Chaos sparrow search optimization algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics,2021,47(8):1712-1720. (in Chinese))