

DOI:10.19651/j.cnki.emt.2518764

DDPG-LSTM 算法在动态多峰场景下 光伏 MPPT 研究*

李义博¹ 袁金丽¹ 负智¹ 郑森潇¹ 郭志涛^{1,2}

(1. 河北工业大学电子信息工程学院 天津 300130; 2. 河北工业大学创新研究院(石家庄) 石家庄 050299)

摘要: 针对光伏系统在局部遮挡、光照突变等复杂动态环境下,出现多极值特性导致最大功率点难以准确追踪和功率波动问题,本文将深度确定性策略梯度连续动作空间优化能力与长短期记忆网络时序特征优势相融合,提出一种基于深度强化学习的 DDPG-LSTM 算法。通过设计分层奖励函数实现功率追踪、动作平滑性和系统稳定性的多目标协同优化。在 MATLAB/Simulink 平台搭建光伏系统仿真模型。实验表明:在多极值和动态环境变化条件下,DDPG-LSTM 算法可稳定跳出局部极值且在最大功率点附近无显著振荡,平均追踪效率达 98% 以上,验证了 DDPG-LSTM 算法在动态环境中的高效性与鲁棒性,为光伏系统智能控制及可再生能源高效利用提供了理论支持。

关键词: 光伏系统;最大功率点追踪;深度强化学习;DDPG-LSTM;局部遮挡

中图分类号: TN01 **文献标识码:** A **国家标准学科分类代码:** 470.4047

Research on DDPG-LSTM algorithm for photovoltaic maximum power point tracking in dynamic multi-peak scenarios

Li Yibo¹ Yuan Jinli¹ Yun Zhi¹ Zheng Senxiao¹ Guo Zhitao^{1,2}

(1. College of Electronic Information Engineering, Hebei University of Technology, Tianjin 300130, China;

2. Innovation Research Institute, Hebei University of Technology (Shijiazhuang), Shijiazhuang 050299, China)

Abstract: To address the challenges of multiple peaks and power fluctuations in maximum power point tracking (MPPT) for photovoltaic systems under complex dynamic environments characterized by partial shading, rapid irradiance fluctuations, and temperature variations, a novel deep reinforcement learning-based algorithm, termed DDPG-LSTM, is proposed. The algorithm integrates the continuous action space optimization capability of the Deep Deterministic Policy Gradient and the temporal feature extraction advantage of Long Short-Term Memory networks. Hierarchical reward mechanisms are designed to achieve multi-objective collaborative optimization, balancing power tracking, action smoothness, and system stability. A simulation model of the photovoltaic system is built on the MATLAB/Simulink platform, and experimental results demonstrate that under multi-peak shading and dynamic environmental conditions, the DDPG-LSTM algorithm stably escapes local optima with negligible oscillations near the maximum power point, achieving an average tracking efficiency exceeding 98%. The robustness and adaptability of the proposed method in dynamic environments are validated, providing theoretical support for the intelligent control of photovoltaic systems and the efficient utilization of renewable energy.

Keywords: photovoltaic systems; maximum power point tracking; deep reinforcement learning; DDPG-LSTM; partial shading conditions

0 引言

当前全球能源消费仍以化石燃料为主导,其大规模使用加剧了环境污染与气候变化^[1]。在众多可再生能源中,

太阳能因其清洁性和可持续优势成为研究热点^[2]。光伏系统的输出功率受环境温度及光照强度影响显著^[3]。为最大程度利用太阳能资源,需通过最大功率点追踪(maximum power point tracking, MPPT)技术使光伏发电系统始终在

收稿日期:2025-05-08

* 基金项目:河北工业大学创新研究院(石家庄)石家庄市科技合作专项基金(SJZZXB23005, SJZZXC24011)项目资助

最大功率点处运行,提高系统的发电效率^[4]。传统MPPT方法,如扰动观察法^[5](P&O)和增量电导法^[6](incremental conductance, INC),在呈现单峰特性时表现较好,但在局部遮挡(partial shading condition, PSC)条件下光伏电池呈现的多峰值特性会大幅降低传统算法的追踪效率^[7]。为解决PSC条件下的多极值以及在最大功率点附近的振荡问题,优化算法被引入MPPT技术^[8]。文献[9]将万有引力优化算法用于提升并网光伏系统的最大功率点跟踪性能,提高了最大功率点提取效率和动态响应速度。文献[10]则提出了一种基于长短期记忆网络(long short-term memory, LSTM)的深度学习算法,提升100 kW双级并网太阳能光伏系统的MPPT性能,突破了优化算法难以处理大数据的局限。

然而上述算法仅为一次性优化过程,无法在训练过程中积累经验或自适应调整,难以处理动态变化的环境^[11]。深度强化学习(deep reinforcement learning, DRL)方法通过融合强化学习与深度学习的优势,使智能体能够对高维输入空间的数据进行高效处理;其交互式的学习过程,能够通过环境的反馈不断调整策略,在动态环境中具有极强的适应能力,成为解决MPPT控制问题的强大而潜在的工具^[12]。迄今为止,深度强化学习已经成功应用游戏、计算机视觉^[13]、智能电网^[14]等多个领域。文献[15]引入了推出了深度Q-Network(deep q-network, DQN)技术,该技术使用卷积神经网络逼近Q值函数,从而稳定训练过程。DQN技术只适用于离散的动作空间,在MPPT控制技术中,连续的动作空间能对光伏系统进行更精细的控制从而减小功率振荡,Avila等^[16]基于深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法构建深度确定性策略框架,通过神经网络近似动作空间实现精细化控制。

本文提出了一种基于深度强化学习方法—DDPG-LSTM算法。DDPG算法采用深度神经网络作为值函数和策略函数的逼近器,能够更精准地拟合高维输入空间中的Q值和策略输出;LSTM网络能有效捕捉数据时序依赖性,利用DDPG经验回放机制能使LSTM能够充分利用历史数据提取其中时序特征,帮助算法跳出光伏电池P-V曲线的局部极值点,增强对环境变化的适应能力,使MPPT控制在多极值和动态环境中表现更稳定。本文研究内容如下:

1) 基于MATLAB/Simulink平台搭建复杂条件下光伏系统仿真模型,分析光伏电池阵列在不同环境下P-V曲线变化情况,为MPPT算法优化设计提供理论依据。

2) 设计强化学习模型的状态空间、动作空间、奖励函数。针对连续状态和动作空间下的光伏系统模型,设计合适的奖励函数,通过分层次、多角度的反馈机制引导策略逐步优化,在保证系统高功率输出的同时兼顾系统稳定性。

3) 在DDPG网络结构基础上引入LSTM网络提出一种新的算法,用于从光伏系统观测的状态中提取环境状态

特征,利用LSTM特有的门控机制使算法充分利用历史信息进行策略优化,使光伏系统在局部遮挡条件下实现更高效的最大功率点追踪。

1 光伏系统仿真模型搭建

本章通过MATLAB/Simulink构建可真实反映在多极值以及环境动态变化下的光伏系统仿真模型,并分析不同环境条件下光伏系统输出功率的变化情况,为后续优化控制算法的设计奠定坚实的理论基础。图1为该仿真模型框图,其核心组成包括光伏电池阵列、DC-DC转换器、MPPT控制算法。MPPT控制算法是系统核心控制单元,用于采集光伏组件的电压、电流等信息并生成相应的占空比控制信号。DC-DC转换器根据控制算法生成的PWM信号来动态调整占空比,使光伏电池阵列始终工作在最大功率点,最大化光伏系统的发电效率。

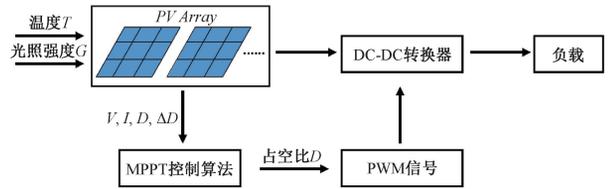


图1 光伏系统模型框图

Fig. 1 Diagram of the photovoltaic system model

1.1 光伏电池模型

光伏电池基于光生伏特效实现光能向电能的转换。PN结在光照下产生空穴-电子对,形成光生载流子。在P-N结内建电场作用驱动载流子定向迁移形成光生电压,接入外部负载电路后产生电流。光伏电池的等效电路模型如图2所示,由电流源、二极管和电阻构成^[17],这些元件共同描述了光伏电池在不同环境条件下的输出特性。

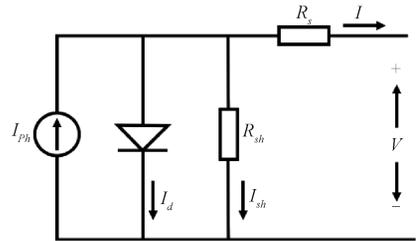


图2 光伏电池等效电路模型

Fig. 2 Equivalent circuit model of photovoltaic cell

根据基尔霍夫定律和二极管特性,其输出电流:

$$I = I_{ph} - i_d \left(\exp \left[\frac{q(V + IR_s)}{AKT} \right] - 1 \right) - \frac{(V + IR_s)}{R_{sh}} \quad (1)$$

$$I_{ph} = [I_{sc} + K(T_c - T_r)] \times \frac{G}{G_{STC}} \quad (2)$$

公式各参数意义及数值如表1所示。上述公式建立了光伏电池的基本数学模型,描述了其伏安特性随环境参数

变化,这种动态特性会导致系统最大输出功率呈现显著变化。本文使用 ACS-335-M 型号光伏电池对光伏系统进行仿真,标准环境 (standard testing condition, STC) 为 $T = 25\text{ }^{\circ}\text{C}$, $G = 1\ 000\ \text{W}/\text{m}^2$ 。标准环境下其参数如表 2 所示。图 3 光伏电池 P-V 曲线表明:温度不变时,输出功率随光照强度增加而显著提升;光照强度不变时,温度越低,光伏电池输出功率越大;相较于温度,光照强度变化对光伏电池输出功率的影响更为显著。

表 1 公式各参数的意义及数值

Table 1 Definitions and numerical values of each parameter in the formula

参数	意义	数值
I_{ph}	光伏电池板产生的电流	—
I_{sc}	短路电流 (STC)	—
i_d	二极管反向饱和电流	—
q	电荷常数	$1.602 \times 10^{-19}\ \text{C}$
A	二极管理想因子	—
K	玻尔兹曼常数	$1.38 \times 10^{-23}\ \text{J}/\text{K}$

表 2 光伏电池参数

Table 2 Photovoltaic cell parameters

参数	数值
开路电压 (V_{oc})	49.9 V
短路电流 (I_{sc})	9 A
最大功率点电压 (V_{mpp})	41.5 V
最大功率点电流 (I_{mpp})	8.07 A
最大功率 (P_{mpp})	334.905 W

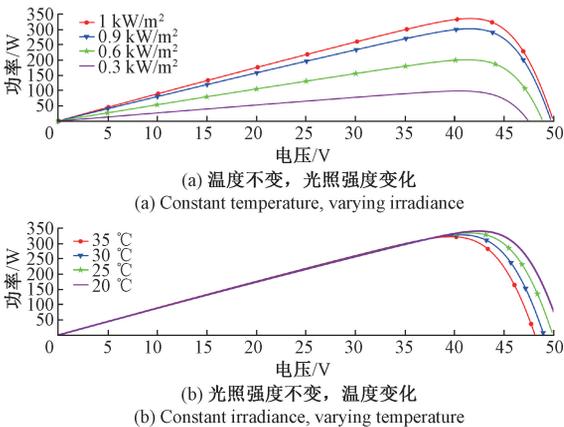


图 3 光伏电池 P-V 不同条件下输出特性曲线

Fig. 3 P-V curves of photovoltaic cell

1.2 光伏系统受遮挡下输出特性分析

本文使用 6 个光伏电池,为分析不同遮挡方式下光伏电池的输出特性,按照 3 串联、2 并联的方式组成光伏电池阵列,如图 4 所示。本文将光伏电池按遮挡方式分为无遮

挡、行遮挡和列遮挡,Env1 表示无遮挡环境;Env2、Env3 和 Env8 为列遮挡环境;Env4~Env7 为行遮挡环境,各环境具体参数设置如表 3 所示 ($T = 25\text{ }^{\circ}\text{C}$, 光照强度单位: W/m^2)。

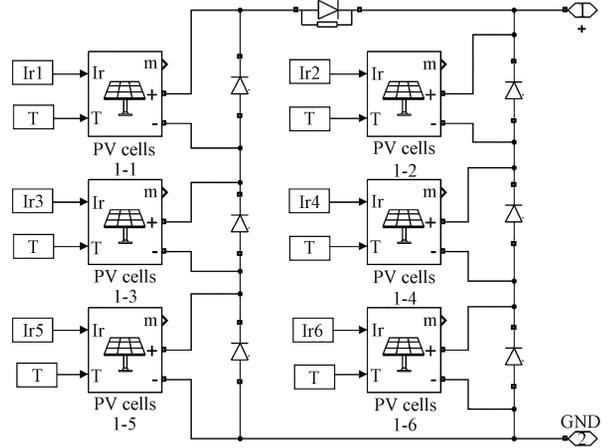


图 4 光伏电池阵列模型

Fig. 4 Photovoltaic cell array model

表 3 光伏电池阵列仿真环境

Table 3 Simulation environment for photovoltaic array

环境	遮挡方式	PV1	PV2	PV3	PV4	PV5	PV6
Env1	无遮挡	900	900	900	900	900	900
Env2	列遮挡	900	500	900	500	900	500
Env3		400	500	400	500	400	500
Env4		900	900	900	900	300	300
Env5	行遮挡	900	900	600	600	300	300
Env6		900	900	600	600	500	500
Env7		900	900	350	350	300	300
Env8		700	700	600	600	600	600

不同的遮挡方式会改变其输出特性:在行遮挡情况下,由于热斑效应需要增加旁路二极管以保护 PV 系统^[18]。当光伏电池受遮挡时旁路二极管会分担部分电流,使光伏电池阵列的 P-V 曲线出现多个局部极值点。图 5 与图 6 显示了 Env4~Env7 逐行遮挡方式下获得的 P-V 曲线。在行遮挡情况下,随着遮挡行数增加,P-V 曲线中的极值点数量同步增长,遮挡行数 b 与极值点数量 a 满足 $a = b + 1$ 的关系。

在列遮挡情况下,每列光伏电池的支路电流相同,即使遮挡列数增加,P-V 曲线仍保持单峰,遮挡程度仅影响峰值大小。由图 5 知,在列遮挡环境下,光伏阵列的 P-V 曲线呈单峰特性,遮挡程度越严重峰值越小。

2 深度强化学习算法设计

2.1 强化学习模型

强化学习是一种无模型在线优化方法,其中强化学习代理是能与环境产生交互并产生动作的智能体 (Agent),通马尔可夫决策过程与环境交互学习获得最优控制策略。

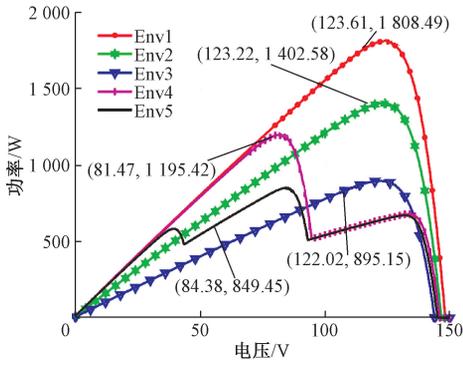


图5 Env1~Env5环境条件下P-V曲线及其最大功率点

Fig. 5 P-V curves and maximum power points under environmental conditions Env1 to Env5

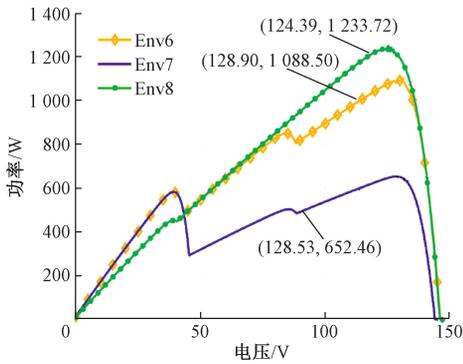


图6 Env6~Env8环境条件下P-V曲线及其极值点

Fig. 6 P-V curves and extremum points under environmental conditions Env6 to Env8

状态空间、动作空间和奖励函数构成强化学习的核心三元素。本研究光伏电池的关键输出特性是电压 V 与电流 I ，二者变化直接决定光伏电池的输出功率，观察电压和电流的实时数据可以及时评估系统当前的功率输出及运行状态。占空比 D 作为控制动作的核心特征，既是DC-DC转换器的直接控制信号，也是调整光伏电池最大功率点的直接手段。为避免控制信号的过度调整或振荡，引入占空比的变化量 ΔD 帮助算法控制调整的平滑性和稳定性。因此，本文设计的四维状态空间维度为：

$$S = \{V, I, D, \Delta D\} \quad (3)$$

本文采用连续动作空间，其动作设置为占空比的变化量 ΔD 。通过调整 ΔD ，使代理依据光伏系统当前状态进行微调，逐步逼近最大功率点。同时限制每次更新的幅度，避免因过度调整引起的振荡，确保追踪过程更加稳定。动作空间范围应该包含正值、负值和零，使光伏系统既可提升或降低功率，也能维持现有控制参数，从而保持动作输出的稳定性。因此，本文的动作空间设计如下：

$$A = \{\Delta D \mid -0.03 \sim 0.03\} \quad (4)$$

为实现功率跟踪、控制约束与动作平滑的多目标协同优化，本文设计由5个部分组成的奖励函数，通过多层次反

馈机制引导代理优化控制策略，增强光伏系统在动态环境变化下的鲁棒性。各部分奖励具体设计如下：

$$r = \sum_{i=1}^5 \omega_i r_i \quad (\sum_i |\omega_i| = 1) \quad (5)$$

$$r_1 = \frac{P_{t+1}}{P_{MPP}} \quad (6)$$

$$r_2 = \begin{cases} \left(\frac{P_{t+1}}{P_{MPP}}\right)^2, & \Delta P \geq -\delta_1 \\ 0, & \Delta P < -\delta_1 \end{cases} \quad (7)$$

$$r_3 = \begin{cases} 1, & |P_{MPP} - P_{t+1}| \leq |P_{MPP} - P_t| \\ 0, & \text{其他} \end{cases} \quad (8)$$

$$r_4 = \begin{cases} 1, & 0 \leq D \leq 1 \\ 0, & \text{其他} \end{cases} \quad (9)$$

$$r_5 = 1 - \frac{\beta |\Delta D|}{\beta_{\max} \Delta D_{\max}} \quad (\beta = \beta_{\min} + (\beta_{\max} - \beta_{\min}) \frac{t}{T}) \quad (10)$$

其中， r_i 为各部分奖励函数， ω_i 为各奖励函数所占权重。 P_{MPP} 表示光伏电池阵列标准条件下的最大功率； $\Delta P = P_{t+1} - P_t$ ；表示当前状态下输出功率与上一状态输出功率的差值； δ_1 表示最大功率点附近的小面积区域。

r_1 表示当前功率与最大功率的比例，用于衡量跟踪至最大功率点的程度，帮助代理优先探索高功率区域，提高系统在最大功率点附近运行的概率； r_2 表示功率增加或功率波动处于合理范围内时给予较高奖励，加速策略优化速度；同时当功率下降超出阈值 δ_1 时取消奖励，以规避功率的剧烈波动，增强光伏系统的稳定性； r_3 表示鼓励代理对正确方向的探索和收敛，避免策略陷入振荡或反向偏离； r_4 表示防止占空比的非法动作输出，避免代理产生无效的动作，确保训练过程的安全性和稳定性； r_5 表示对连续时间步下占空比 D 间的剧烈变化进行惩罚，惩罚系数 β 随时间步线性增长，允许初期较大动作探索，后期抑制动作大范围波动，提升系统运行的整体稳定性。

经仿真实验验证， r_1 与 r_2 对算法追踪精度具有决定性影响，直接衡量输出功率的大小与稳定性，将其赋予最高权重0.25； r_3 和 r_4 保证算法的稳定性，减小振荡，其权重为0.2； r_5 对算法输出约束，权重不宜过高，设置为0.1。

2.2 DDPG-LSTM算法

1) 长短期记忆网络

本研究中光伏系统具有非马尔可夫特性，电压、电流、占空比等状态量都具有明显的时间相关性——当前状态不仅取决于当下的环境，也受到前一时刻的动态变化影响。传统DDPG以静态状态输入为主，难以处理数据的时序依赖关系，容易将局部极值点误认为是最大功率点。LSTM是一种特殊循环神经网络结构，其独特的门控机制可保留时间序列中较长跨度的历史信息，捕捉输入数据的长期依赖关系，辅助代理优化决策策略。其网络结构如图7所示。

输入数据 x_t 为来自状态空间观测的数据，通过细胞状

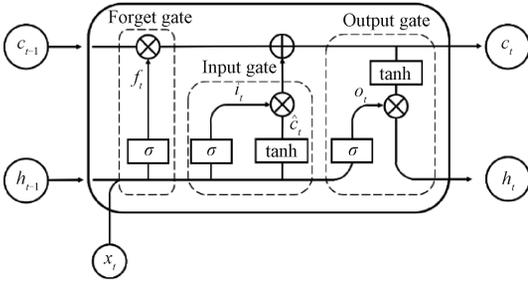


图 7 LSTM 网络结构

Fig. 7 LSTM cell architecture

态来确定数据之间的时间关系。忘记门 f_t 的输入是上一步的隐藏状态 h_{t-1} 和当前输入数据 x_t ，其经 sigmoid 函数控制保留信息的比例，其计算公式为：

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (11)$$

\hat{C}_t 表示候选细胞状态，将当前时刻输入的信息通过 \tanh 函数处理后生成新的候选信息。输入门 i_t 通过 sigmoid 函数控制有多少候选信息被加入到细胞状态中。其计算公式为：

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (12)$$

$$\hat{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (13)$$

经过遗忘门和输入门的数据处理，该 LSTM 单元的细胞状态 C_t 计算公式为：

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \quad (14)$$

输出门控制当前单元细胞状态 C_t 有多少信息应该传递到下一单元。其计算公式为：

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (15)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (16)$$

2) DDPG-LSTM

DDPG 是一种基于 Actor-Critic 架构的强化学习算法，特别适用于处理连续动作空间任务。actor 网络输出确定性连续动作值，critic 网络评估状态-动作对的预期累计奖励。本研究通过引入 LSTM 网络，增强处理数据长期依赖关系的能力，为 DDPG 的策略网络提供更丰富的输入，以提升算法在多极值环境中决策的准确性，DDPG-LSTM 算法框图如图 8 所示。该算法采用经验回放机制，通过与环境持续交互获取经验元组 $\{s_t, a_t, r_t, s_{t+1}\}$ ，并存储于有限容量的经验回放池 R 中。当 R 达到容量上限时，采用先进先出策略进行数据更新。训练过程中，从 R 中随机采样批量经验元组，每条经验元组可多次利用，直至其对策略影响衰减至无效。该机制可以重复利用历史样本打破数据相关性，降低局部经验过拟合风险，有助于算法保持较高稳定性和训练效率。

actor-critic 网络结构如图 9 所示，网络计算过程如下：actor 网络的输入层接收光伏系统电压、电流等数据，经 LSTM 层提取时序特征 $h_t^{(a)}$ ，通过全连接层及非线性变换

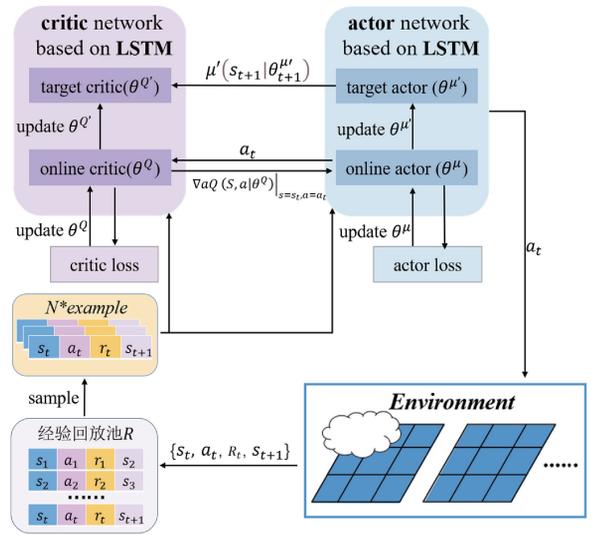


图 8 DDPG-LSTM 算法框图

Fig. 8 Diagram of the DDPG-LSTM algorithm

生成策略函数 $\mu(s_t | \theta_t^{\mu'})$ 。为增强探索能力并防止训练初期过早的收敛，在动作输出叠加噪声 N_t ，最终动作输出为：

$$a_t = \mu(s_t | \theta_t^{\mu'}) + N_t \quad (17)$$

critic 网络由状态和动作双分支结构构成，其中状态分支同样利用 LSTM 提取时序特征，动作分支将动作 a_t 经过一层全连接层映射为动作特征 $h_t^{(a)}$ ，然后，状态分支的特征 $h_t^{(s)}$ 与动作分支的特征 $h_t^{(a)}$ 在加法层融合：

$$\tilde{h}_t = h_t^{(s)} + h_t^{(a)} \quad (18)$$

再经过 ReLU 激活和全连接层，评论家网络输出当前状态-动作对的价值评估，即 Q 值。DDPG-LSTM 算法的具体流程如算法 1 所示。

算法 1

- 输入：状态空间中 4 维变量： $V, I, D, \Delta D$
 初始化：初始化 LSTM 网络参数 θ^L 、隐藏状态 h_t ；
 初始化 actor network 参数 $\theta^{\mu}, \theta^{\mu'}$ ；
 初始化 critic network 参数 $\theta^Q, \theta^{Q'}$ ；
 初始化经验回放池 R
 1 for epoch=1, ..., N do:
 2 初始化占比 D 、经验回放池 R；
 3 for t=1, ..., T do:
 4 通过 LSTM 网络输出特征 $h_t^{(a)}, h_t^{(s)}$ ；
 5 根据 $h_t^{(a)}$ 由 actor online network 输出策略 $\mu(s_t | \theta_t^{\mu'})$ ；
 6 加入噪声 N_t 形成动作输出 a_t ；
 7 计算 R_t ，形成经验元组 $\{s_t, a_t, r_t, s_{t+1}\}$ 存入 R 中；
 8 从 R 中采样 N 批次经验元组，由目标 critic 网络计算样本 Q 值：

$$y_t = r_t + \gamma Q'(s_{t+1}, \mu_{t+1} | \theta^{Q'})$$

算法1(续)

Algorithm 1 (continued)

9 特征融合 $\tilde{h}_t = h_t^{(s)} + h_t^{(a)}$ 由在线 critic 网络计算实际 Q 值:

$$y = Q(s_t, a_t | \theta^Q) = Q(\tilde{h}_t, a_t | \theta^Q)$$

10 critic 网络目标最小化 y_i 与 y 的均方误差,梯度下降方式更新参数 θ^Q :

$$L(\theta^Q) = E[(y_i - Q(s_t, a_t | \theta^Q))^2]$$

11 actor 网络目标最大化 Q 值,梯度上升方式更新参数 θ^μ :

$$\nabla_{\theta^\mu} J(\theta^\mu) = \frac{1}{N} \sum_i \nabla_{\theta^\mu} \mu(s_i | \theta^\mu) \nabla_a Q(s_i, a | \theta^Q) \Big|_{a=\mu(s_i)}$$

12 滑动方式更新目标网络参数 θ^Q 、 $\theta^{\mu'}$:

$$\theta^Q \leftarrow \tau_Q + (1 - \tau_Q) \theta^Q$$

$$\theta^{\mu'} \leftarrow \tau_\mu + (1 - \tau_\mu) \theta^{\mu'}$$

13 end

14 计算训练集总回报值

15 end

3 仿真实验与结果分析

本文使用 Matlab 2023a/Simulink 仿真平台搭建光伏系统,DDPG-LSTM 算法训练使用强化学习工具箱实现。每个 episode 的仿真时长为 0.5 s,采样时间为 0.01 s,时间步长为 50。训练总集数为 5 000 个 episode。表 4 为 DDPG-LSTM 算法参数设置。

为验证本文提出的 DDPG-LSTM 算法在追踪效率、功率振荡等方面的综合性能,对按行遮挡、按列遮挡、环境动态变化三种不同遮挡方式下的光伏系统的最大功率点追踪问题进行测试及结果分析,并将 DDPG-LSTM 算法与 DDPG^[19]、INC^[6]和 GSA^[9]算法进行性能比较。

3.1 按行遮挡环境下算法性能分析

由表 3 可知,Env4 和 Env7 分别为光伏电池阵列受遮挡一行和两行的实验环境,根据光伏电池行遮挡分析知 Env4 条件下 P-V 曲线呈现两个极值点,Env7 则呈现 3 个极值点。两种环境下 4 种算法功率追踪曲线和占空比追踪曲线如图 10 所示。在 Env4 中,INC 和 DDPG 算法在追踪稳定后仍在最大功率附近出现约 80 W 功率波动,INC 占

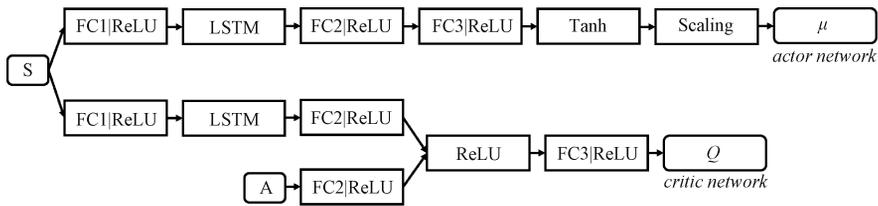


图9 Actor-critic 网络结构

Fig. 9 Actor-critic network structure

表4 DDPG-LSTM 算法参数设置

Table 4 Parameter configuration of the DDPG-LSTM

参数	数值
Experience replay size	1×10^6
Mini batch size	256
Discount factor (γ)	0.9
Sequence length	10
τ_Q	1×10^{-3}
τ_μ	1×10^{-4}
Variance	0.5
Decay of Variance	1×10^{-4}
Smoothing factor	1×10^{-4}

空比曲线波动更为严重,频繁调节占空比会导致 DC-DC 转换器中的开关器件产生较大的开关损耗和热损耗,长时间运行会对器件造成不可逆损伤。在 Env7 中,INC 算法追踪的最大功率为 502 W 左右,由图 6 中该环境极值曲线图可知在第三极大值点处停止寻优,将局部极值点误认为最大功率点;DDPG 算法追踪功率一直在 470~575 W 之

间来回波动,由图 6 可知其在第二极值点和第三极值点之间反复寻优。GSA 算法在多极值条件下追踪效率比较低,这两种环境下均未能追踪到最大功率点,在 Env7 中 GSA 算法将第二极值点误认为最大功率点输出。而 DDPG-LSTM 算法在两种环境种均能成功跳出局部极值,功率和占空比追踪曲线收敛后均无明显振荡,并稳定维持在最大功率点附近稳定输出,其追踪效率分别高达 99.5%、99.07%。在 Env7 环境中更是比 INC 和 DDPG 算法分别提升 22.4%、16.9% 的追踪效率。实验结果表明,DDPG-LSTM 算法在多极值场景下表现出优异的追踪性能。

3.2 按列遮挡环境下算法性能分析

由表 3 可知,Env2 为第一列光伏电池阵列受遮挡、Env3 为两列光伏电池阵列均受遮挡的实验环境,根据光伏电池列遮挡分析知两种环境下的 P-V 曲线中均仅存在一个极值点,即最大功率点。两种环境下 4 种算法功率追踪曲线和占空比追踪曲线如图 11 所示。由图可知 4 种算法在单极值点情况下基本都在最大功率点附近输出,但 INC 和 DDPG 算法存在不同程度的功率波动。DDPG-LSTM 算法在收敛后始终稳定在 1 377 W 输出,且占空比曲线也

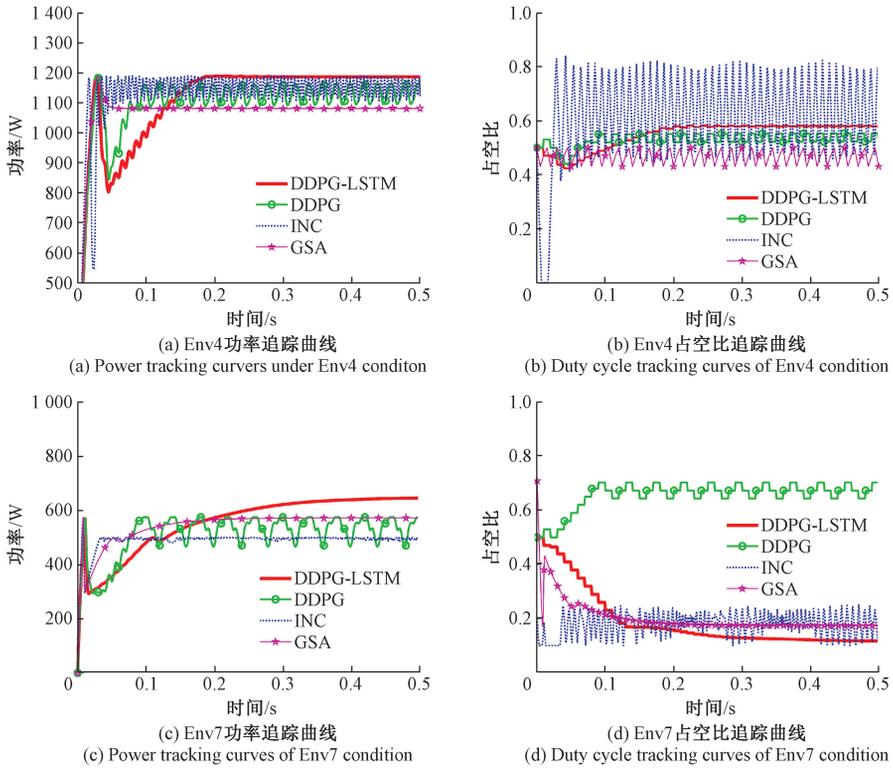


图 10 Env4 和 Env7 下四种算法追踪曲线

Fig. 10 Tracking curves of four algorithms under Env4 and Env7 conditions

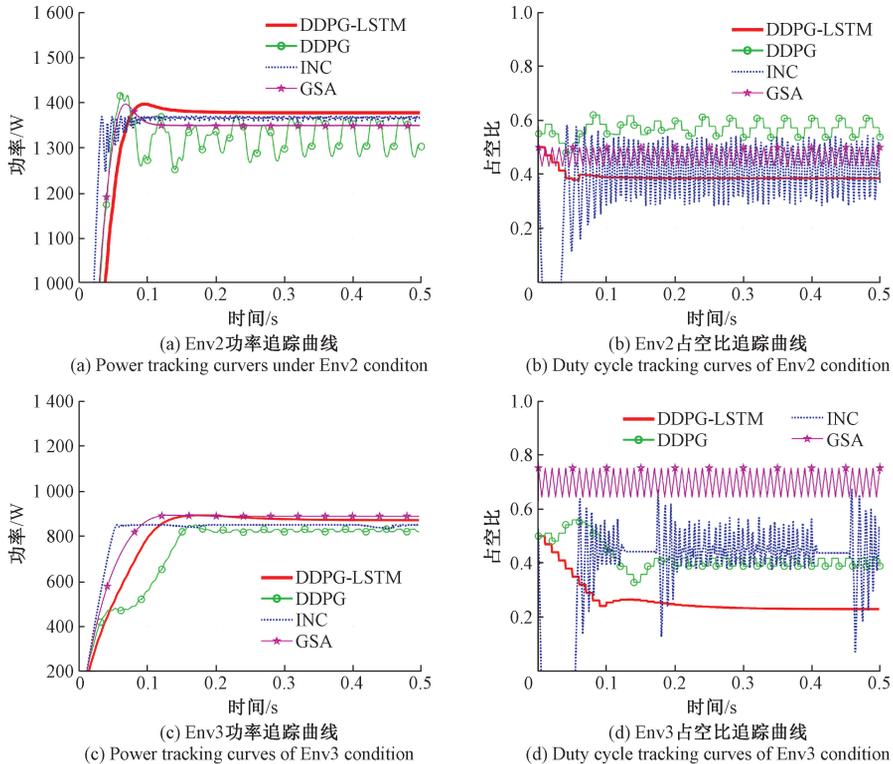


图 11 Env2 和 Env3 下四种算法追踪曲线

Fig. 11 Tracking curves of four algorithms under Env2 and Env3 conditions

无明显振荡,其追踪效率分别比 INC、DDPG、GSA 提高了 0.79%、3.85%、2.06%。在 Env3 条件下,INC 与 DDPG 算法功率波动较小,但 DDPG-LSTM 的追踪效率仍然比 INC、DDPG 提高了 2.79%、5.7%。GSA 算法追踪效率效率最高,但是占空比出现快速波动,稳定性较差。实验表明,单极值环境时,DDPG 追踪功率效果最差,INC、GSA 居中,DDPG-LSTM 在功率追踪效率和稳定性均表现出了优异性能。

3.3 光照强度动态变化下算法性能分析

为验证 DDPG-LSTM 算法在光照强度动态变化条件下的性能鲁棒性,实验在 5 s 内依次设置了 4 种不同遮挡环境进行测试。Env1 为无遮挡,Env8 为列遮挡,Env6 和 Env7 为行遮挡。每种环境持续 1 s,实验环境变化依次为 Env1、Env6、Env8、Env7。图 12 展示了 4 种算法在环境动态变化时的功率追踪曲线与占空比曲线。由图 12 可知,INC 算法虽能基本能够找到最大功率点,但存在小范围功率波动,在多极值环境中尤为明显。同时,占空比曲线始终处于较为严重的振荡状态。DDPG 算法同样能找到各环境的最大功率点,但功率波动幅度较大,其在 Env7 中,该算法陷入该环境的第三极值点,追踪效率大幅下降。GSA 算法参数实时调整能力低,在动态变化环境中表现不佳,仅在 Env6 中追踪效率较高,其他环境中存在着至少 80 W 的功率损失,追踪效果较差。相比之下,DDPG-

表 5 各环境四种算法追踪效率

Table 5 Tracking efficiencies of four algorithms under various environmental conditions %

环境	DDPG-LSTM	DDPG	INC	GSA
Env1	99.57	* 96.68	* 97.75	95.72
Env2	98.22	* 94.37	97.43	96.16
Env3	97.65	91.95	94.86	99.21
Env4	99.50	* 94.62	* 96.45	90.40
Env6	99.26	* 76.04	* 74.84	99.21
Env7	99.07	* 82.14	76.63	87.79
Env8	98.83	99.28	99.43	90.67

注:“*”表示存在功率波动

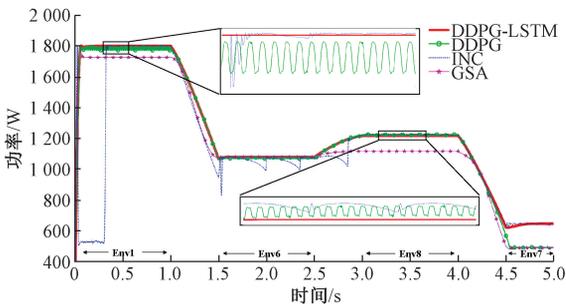
LSTM 算法在环境动态变化下表现出良好的追踪效果,在开始追踪时仅用 0.16 s 就完成收敛,相比 INC 算法提升了 60% 的追踪速度。在动态环境变化时,DDPG-LSTM 算法依然能够稳定跳出局部极值,始终保持 99% 左右的追踪精度,充分验证了其在动态环境下卓越的鲁棒性。表 5 统计了不同环境下各算法的追踪效率。

4 结 论

本文针对多极值以及环境动态变化的复杂条件下光伏系统的最大功率追踪问题,提出了一种基于深度强化学习的 DDPG-LSTM 算法。该算法充分利用了 DDPG 在连续动作空间处理上的优势和 LSTM 对时序特征提取的能力,实现了动态多峰场景下对光伏系统最大功率点的精准追踪。强化学习中多层次设计的奖励函数兼顾了高功率输出与系统鲁棒性,解决了传统方法在动态环境下的频繁振荡问题。实验结果表明,所提出的 DDPG-LSTM 算法在追踪效率和功率波动控制的关键指标上均优于 INC、DDPG 和 GSA 算法。由表 5 可知,在不同的实验场景下,该算法均能有效跳出局部极值,稳定实现高于 98% 的追踪效率,且在最大功率点附近无显著振荡,显著降低了系统能量损耗和器件损耗。未来的工作可围绕硬件部署,通过现场验证确保稳定高效追踪最大功率点,实现 DDPG-LSTM 算法在光伏系统的工程化应用。

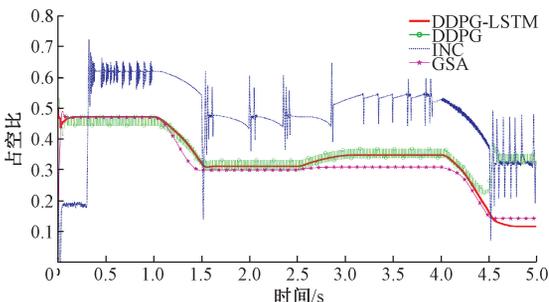
参考文献

- [1] QUERESHI S, JADHAO P R, PANDEY A, et al. Overview of sustainable fuel and energy technologies[J]. Sustainable Fuel Technologies Handbook, 2021: 3-25, DOI:10.1016/b978-0-12-822989-7.00001-9.
- [2] 张严,王亚君,余佳琪.基于神经网络-自适应模糊的光伏 MPPT 算法[J].国外电子测量技术,2022,41(4): 62-69.
ZHANG Y, WANG Y J, YU J Q. Photovoltaic MPPT algorithm based on ANN-adaptive FLC [J]. Foreign Electronic Measurement Technology, 2022,41(4):62-69.



(a) 光照强度动态变化功率曲线

(a) Power tracking curves under dynamically varying irradiance conditions



(b) 光照强度动态变化占空比曲线

(b) Duty cycle tracking curves under dynamically varying irradiance conditions

图 12 光照强度动态变化下四种算法追踪曲线

Fig. 12 Tracking curves of four algorithms under dynamically varying irradiance conditions

- [3] 魏业文, 邱天一. 基于改进鸚鵡算法的光伏多峰值 MPPT 控制[J]. 电子测量技术, 2024, 47(12): 83-90.
WEI Y W, QIU T Y. PV multi-peak MPPT control based on improved parrot algorithm [J]. Electronic Measurement Technology, 2024, 47(12): 83-90.
- [4] BELHACHAT F, LARBES C. A review of global maximum power point tracking techniques of photovoltaic system under partial shading conditions[J]. Renewable and Sustainable Energy Reviews, 2018, 92: 513-553.
- [5] BÜYÜK M, İNCI M. Improved drift-free P&O MPPT method to enhance energy harvesting capability for dynamic operating conditions of fuel cells [J]. Energy, 2023, 267: 126543.
- [6] GUPTA A K, PACHAURI R K, MAITY T, et al. Effect of various incremental conductance MPPT methods on the charging of battery load feed by solar panel[J]. IEEE Access, 2021, 9: 90977-90988.
- [7] WORKU M Y, HASSAN M A, MARAABA L S, et al. A comprehensive review of recent maximum power point tracking techniques for photovoltaic systems under partial shading[J]. Sustainability, 2023, 15(14): 11132.
- [8] ZAFAR M H, KHAN N M, MIRZA A F, et al. A novel meta-heuristic optimization algorithm based MPPT control technique for PV systems under complex partial shading condition [J]. Sustainable Energy Technologies and Assessments, 2021, 47: 101367.
- [9] PATTNAYAK S K, CHOUDHURY S, NAYAK N, et al. Maximum power tracking & harmonic reduction on grid PV system using chaotic gravitational search algorithm based MPPT controller [J]. 2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE), 2020: 1-6.
- [10] YOUNAS U, KULAKSIZ A A, ALI Z. Deep learning stack LSTM based MPPT control of dual stage 100 kWp grid-tied solar PV system[J]. IEEE Access, 2024, 12: 77555-77574.
- [11] GE ZH H, LI X SH, XU F, et al. An improved distributed maximum power point tracking technique in photovoltaic systems based on reinforcement learning algorithm[J]. IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 2023, 5(1): 167-178.
- [12] WADEHRA A, BHALLA S, JAISWAL V, et al. A deep recurrent reinforcement learning approach for enhanced MPPT in PV systems [J]. Applied Soft Computing, 2024, 162: 111728.
- [13] KAHN G, VILLAFLORES A, DING B, et al. Self-supervised deep reinforcement learning with generalized computation graphs for robot navigation[C]. 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 5129-5136.
- [14] ZHANG D X, HAN X Q, DENG CH Y. Review on the research and practice of deep learning and reinforcement learning in smart grids [J]. CSEE Journal of Power and Energy Systems, 2018, 4(3): 362-370.
- [15] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [16] AVILA L, DE PAULA M, TRIMBOLI M, et al. Deep reinforcement learning approach for MPPT control of partially shaded PV systems in smart grids[J]. Applied Soft Computing, 2020, 97: 106711.
- [17] 王军龙, 杨欢红, 沈森, 等. IBES 算法在并联 Boost 电路 MPPT 系统中的应用[J]. 电子测量技术, 2023, 46(15): 1-9.
WANG J L, YANG H H, SHEN M, et al. Application of IBFS algorithm in parallel Boost circuit MPPT system [J]. Electronic Measurement Technology, 2023, 46(15): 1-9.
- [18] YOUSAF M Z, KOONDHAR M A, ZAKI Z A, et al. Improved MPPT of solar PV systems under different environmental conditions utilizes a novel hybrid PSO [J]. Renewable Energy, 2025, 244: 122709.
- [19] PHAN B C, LAI Y C, LIN C E. A deep reinforcement learning-based MPPT control for PV systems under partial shading condition[J]. Sensors, 2020, 20(11): 3039.

作者简介

李义博, 硕士研究生, 主要研究方向为深度强化学习。

E-mail: yibo_li2001@163.com

袁金丽, 副教授, 主要研究方向为智能信息处理、计算机视觉、机器学习。

E-mail: 2005051@hebut.edu.cn

侯智, 硕士研究生, 主要研究方向为嵌入式与人工智能。

E-mail: 942775630@qq.com

郑森潇, 硕士研究生, 主要研究方向为人工智能与深度学习。

E-mail: 2752665362@qq.com

郭志涛(通信作者), 教授, 主要研究方向为物联网、智能传感器、智能优化算法。

E-mail: 2002089@hebut.edu.cn