

黄显峰, 冉超越, 周文, 等. 基于深度强化学习算法的水光互补优化调度研究[J]. 水利水电技术(中英文), 2025, 56(4): 235-247. DOI: 10.13928/j.cnki.wrahe.2025.04.019

HUANG Xianfeng, RAN Chaoyue, ZHOU Wen, et al. Research on water-light complementary optimal scheduling based on deep reinforcement learning algorithm [J]. Water Resources and Hydropower Engineering, 2025, 56(4): 235-247. DOI: 10.13928/j.cnki.wrahe.2025.04.019

基于深度强化学习算法的水光互补优化调度研究

黄显峰¹, 冉超越¹, 周文¹, 李旭²

(1. 河海大学 水利水电学院, 江苏 南京 210098; 2. 华能澜沧江水电股份有限公司, 云南 昆明 650214)

摘要:【目的】水光互补优化调度中光伏出力具有波动性、随机性、间歇性等特点, 其求解空间通常较高维、复杂且为连续空间, 水光互补优化调度问题中涉及的多种连续控制决策问题。【方法】深度强化学习算法中的深度确定性策略梯度(DDPG)算法适合处理求解空间连续、复杂的问题, 首先对水光互补问题进行强化学习建模, 并基于水光互补机理考虑需调、可调概念设置环境、动作、奖励函数与惩罚函数, 采用DDPG算法进行优化。对比分析只使用初始DDPG算法和遗传算法的优化结果, 讨论模型的适用性和有效性。以澜沧江上游大型水光互补基地为例, 设置三个梯级水电站配置方案, 三种水文代表年, 开展实例研究。【结果】(1)采用DDPG算法运行速度较快, 在考虑需调、可调水光互补机理建模后光伏消纳电量达到129.93亿kWh, 为三种模型最高。(2)来水越枯, 光伏消纳能力越强; 梯级水电站装机862万kW时, 平水年到枯水年光伏消纳能力只提高1%, 此时能够最大限度利用水电系统的互补能力。(3)在丰水期的光伏消纳能力较低, 三个方案光伏消纳率分别为77.43%、79.85%、89.39%。【结论】深度强化学习算法在水光互补优化调度中展现出快速收敛的优势, 将“需调”和“可调”机理融入强化学习建模能够显著提升光伏消纳效果, 实现更优的资源利用, 有效提升了水光互补系统的运行效率与光伏电量消纳能力。该方法在清洁能源基地的容量配置和运行调度领域有较好效果, 为后续清洁能源系统的扩展应用奠定了理论与实践基础。

关键词: 水光互补; 强化学习; DDPG; 优化调度; 影响因素; 水电站

DOI: 10.13928/j.cnki.wrahe.2025.04.019

开放科学(资源服务)标志码(OSID):

中图分类号: TV213.9

文献标志码: A

文章编号: 1000-0860(2025)04-0235-13



Research on water-light complementary optimal scheduling based on deep reinforcement learning algorithm

HUANG Xianfeng¹, RAN Chaoyue¹, ZHOU Wen¹, LI Xu²

(1. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, Jiangsu, China;

2. Huaneng Lancang River Hydropower Inc., Kunming 650214, Yunnan, China)

Abstract: [Objective] Photovoltaic output in water-light complementary optimal scheduling is characterized by volatility,

收稿日期: 2024-10-16; 修回日期: 2024-11-27; 录用日期: 2024-11-28; 网络出版日期: 2024-12-04

基金项目: 国家自然科学基金项目(52179012); 中国华能集团科技项目“澜沧江西藏段千万千瓦清洁能源基地建设水光互补关键技术研究”(HNKJ20-H20)

作者简介: 黄显峰(1980—), 男, 副教授, 硕士研究生导师, 博士, 主要从事水资源规划及利用、河流管理与保护等方面研究。E-mail: xfluang@hhu.edu.cn

通信作者: 冉超越(2000—), 男, 硕士研究生, 主要从事水光互补调度等方面研究。E-mail: 978928716@qq.com

©Editorial Department of Water Resources and Hydropower Engineering. This is an open access article under the CC BY-NC-ND license.

randomness, and intermittency. Its solution space is typically high-dimensional, complex, and continuous. A variety of continuous control decision-making problems are involved in water-light complementary optimal scheduling. [Methods]The Deep Deterministic Policy Gradient (DDPG) algorithm in the deep reinforcement learning algorithm was suitable for solving continuous and complex problems in the solution space. The water-light complementary problem was modeled using reinforcement learning. Based on the water-light complementary mechanism, the concepts of “demand for adjustment” and “capacity for adjustment” were considered to set up the environment, actions, reward function, and penalty function. The DDPG algorithm was then used for optimization. The applicability and effectiveness of the model were assessed by comparing and analyzing the optimization result using only the initial DDPG algorithm and those using the genetic algorithm. Taking the large-scale water-light complementary base in the upper reaches of the Lancang River as an example, three cascade hydropower station configuration schemes and three representative hydrological years were set up for a case study. [Results]The analysis indicated that: (1) the DDPG algorithm performed faster. After considering the mechanisms of “demand for adjustment” and “capacity for adjustment”, the photovoltaic power consumption reached 12.993 billion kWh, which was the highest among the three models. (2) The lower the water inflow, the stronger the photovoltaic consumption capacity was. When the installed capacity of cascade hydropower stations was 8.62 million kW, the photovoltaic consumption capacity only increased by 1% from normal year to dry year. At this time, the complementary capacity of hydropower system could be maximized. (3) The photovoltaic consumption capacity was relatively low during the wet season, and the photovoltaic consumption rates of the three schemes were 77.43%, 79.85%, and 89.39%, respectively. [Conclusion]The deep reinforcement learning algorithm demonstrates the advantage of rapid convergence in the water-light complementary optimal scheduling. Integrating mechanisms of “demand for adjustment” and “capacity for adjustment” into reinforcement learning modeling can significantly enhance the photovoltaic consumption efficiency, achieve better resource utilization, and effectively improve the operation efficiency of the water-light complementary system and the photovoltaic power consumption capacity. This method shows promising result in the capacity configuration and operation scheduling of clean energy base, providing a theoretical and practical foundation for the future expansion and application of clean energy system.

Keywords: water-light complementary; reinforcement learning; DDPG; optimal scheduling; influencing factors; hydropower stations

0 引言

光伏发电具有环保、可再生、长寿命、低维护等优点,但也有波动性、随机性、间歇性等特点,影响能源利用效率。水力发电可以弥补光伏发电不确定性,提高清洁能源发电效率和可持续性^[1-4],不仅有助于协同推进减污降碳^[5],还能有效加强水资源高质量利用^[6-7]。

在水光互补优化调度系统特性研究中, YANG 等^[8]考虑了光伏系统与水电系统的容量配置。TAN 等^[9-10]将多能互补目标由最大化水能资源利用效率转变为水能高效利用与灵活性调节并举。周永斐等^[11]从电力系统出发,考虑互补系统要求构建模型。朱燕梅等^[12-13]等进一步将随机波动性解析为整体与局部指标,黄显峰等^[14]在考虑短期的互补策略的情况下提出了时间尺度为中长期的互补调度模型。鲜于虎成等^[15]考虑了水光互补中的短期弃电与波动风险,提出了中长期双层嵌套优化调度方法。对水光互补特性机理的归纳研究,在水光互补系统调度策略、模型使用、容量配置上作出贡献。

针对互补系统的目标选择与计算方法,梁艺缤等^[16-18]将经济目标作为求解算法目标带入求解,在电网协调的研究上,宋柯^[19]将电网波动性作为目标进行优化,罗彬等^[20]考虑光伏出力不确定性,以整体可消纳电量期望最大为目标,用混合整数线性规划进行求解。LI 等^[21-23]考虑了水光互补发电量、经济、生态目标。在对水光互补系统目标求解中,常用算法主要是基于运筹学理论的传统算法与基于启发式方法的算法,前者主要是动态规划算法及其改进型^[24-25],虽然动态规划能适应水库调度问题的需求,但其需要全局搜索以得到最优解,运行缓慢,在水光互补调度特别是包括梯级水库的水光互补系统中,复杂度上升极易陷入“维数灾”。基于启发式方法的算法应用广泛,包括遗传算法^[26]、粒子群算法^[27-28]等,相较于传统优化算法能更好发挥计算机效能,取得了更好的决策结果以及更快的求解速度。其中 NSGA-II 是一种经典的多目标优化算法,广泛应用水库调度等领域的多目标优化问题^[29-31]。其主要优势在于能够生成高质量的 Pareto 前沿,具有较好的收敛性。然而, NSGA-II 模型在处理高维度复杂问题时存在不足,尤

其是当问题维度增加时, 计算时间增加, 解的分布可能局限于局部最优解, 导致算法的效率和优化效果降低。

深度强化学习是将深度学习技术与强化学习相结合, 使用深度神经网络来近似值函数或策略函数, 解决复杂环境和大规模状态空间下的强化学习问题^[32-34]。在水库优化调度领域, LI 等^[35]采用强化学习方法计算水库实时调度问题, WU 等^[36]在水电站水库的供水、生态、发电多目标优化调度问题上验证了深度强化学习的有效性。胡鹤轩等^[37]提出一种基于 n 步 Q -learning 的水库群优化调度方法, 对强化学习在水库调度问题中做出有效探索。李文武等^[38-40]均通过 DQN 算法进行求解, 也有效解决了梯级水库优化调度的维数灾问题。此外在非常规水资源优化配置^[41]、水库防洪模型^[42]等问题中, 深度强化学习算法在设置合理参数的情况下能有较好训练结果。然而, 这些研究通常将动作空间进行离散化处理。尽管这种方法在强化学习构建完善的情况下是有效的, 但在现实工程中, 包括水库优化调度等问题过程的求解空间是一个连续的区间变量。具体而言, 在水光互补系统中, 水电和光伏出力的互补过程复杂, 需要同时考虑水力发电和光伏发电的协调。求解空间采用离散化出力会导致局部最优情况频发, 在多维空间中的离散化也会导致求解速度缓慢, 从而限制了优化调度的效果。

深度确定性策略梯度算法, 即 DDPG 算法是一种深度强化学习算法, 适合解决连续动作空间的决策问题^[43]。它结合了深度神经网络和经验回放等技术, 适用于处理复杂的环境和连续的动作空间^[44], 在实际应用中, DDPG 算法已被广泛应用于电网多能互补运行等领域, 展现了其在处理多种能源协同优化问题上的潜力^[45], 仍缺少与水光互补中梯级水库调度模型深度结合且考虑具体调度机理建模的研究。水光互补调度模型的复杂性和动态变化性要求算法能够灵活且高效地处理连续的调度决策, 而 DDPG 算法具有这一潜力。

为此, 采用适用于求解连续空间问题的深度确定性策略梯度算法 (DDPG), 并考虑水光互补调度机理构建水光互补优化调度环境、动作空间, 对比初始 DDPG 模型与 NSGA-II 模型测试效果, 并设置三种水电装机方案为算例分析结果。

1 模型建立

1.1 目标函数

水光合计发电量最大

$$W = \max \sum_{m=1}^M \left\{ \sum_{i=1}^I \sum_{d=1}^D N_{i,m,d}^{a,h} T \Delta t + W_m^p \right\} \quad (1)$$

式中, Δt 为计算时段长度, $\Delta t = 1 \text{ h}$; T 为日内小时数, $T = 24$; M 为年内月份, $M = 12$; D 为当月天数; I 为梯级水电站数; $N_{i,m,d}^{a,h}$ 为第 i 个电站在第 m 月第 d 天的水光互补后的月均水电出力 (万 kW); W_m^p 为第 m 月的光伏消纳电量 (万 kWh); W 为年水光合计发电量 (万 kWh)。

1.2 约束条件

水量平衡约束

$$V_{i,t} = V_{i,t-1} + (q_{i,t}^r - q_{i,t}^x) \Delta t - E_{i,t} S_{i,t} - D_{i,t} \quad (2)$$

式中, $V_{i,t-1}$, $V_{i,t}$ 为第 i 个电站在 t 时段初、末期的水库蓄水量 (m^3); $q_{i,t}^r$ 为第 i 个水库在 t 时段的入库流量 (m^3/s); $q_{i,t}^x$ 为第 i 个水库在 t 时段的下泄流量 (m^3/s); $E_{i,t}$ 为第 i 个电站在 t 时段蒸发深度 (m); $S_{i,t}$ 为第 i 个电站在 t 时段水库面积 (m^2); $D_{i,t}$ 为第 i 个电站在 t 时段渗漏损失水量 (m^3)。

水位限制

$$Z_{i,t,\min} \leq Z_{i,t} \leq Z_{i,t,\max} \quad (3)$$

式中, $Z_{i,t,\min}$ 为第 i 个电站在 t 时段允许运行最低水位 (m); $Z_{i,t,\max}$ 为第 i 个电站在 t 时段允许最高运行水位 (m)。

发电流量限制

$$q_{i,t,\min} \leq q_{i,t} \leq q_{i,t,\max} \quad (4)$$

式中, $q_{i,t,\min}$ 为第 i 个电站在 t 时段各水库的最小生态流量与水轮机要求的最小引水流量的较小值 (m^3/s); $q_{i,t,\max}$ 为第 i 个电站在 t 时段最大允许引水流量 (m^3/s)。

电站出力约束

$$N_{m,d,t}^{h,l} \leq A q_{i,t} H_{i,t} \leq N_{m,d,t}^{h,u} \quad (5)$$

式中, $N_{m,d,t}^{h,u}$ 为第 m 月第 d 天第 t 时段瞬时梯级电站出力上限 (万 kW); $N_{m,d,t}^{h,l}$ 为第 m 月第 d 天第 t 时段瞬时梯级电站出力下限 (万 kW); A 为出力系数; $q_{i,t}$ 为发电流量; $H_{i,t}$ 为发电水头。

通道约束

$$N_{m,d,t}^a < N_{m,d,t}^c \quad (6)$$

式中, $N_{m,d,t}^a$ 为第 m 月第 d 天第 t 时段的水光互补后的水光合计出力 (万 kW); $N_{m,d,t}^c$ 为单位时段通道限制 (万 kW)。

上下游水力联系

$$q_{i,t}^r = q_{i-1,t}^x + q_{i-1,i,t}^j \quad (7)$$

式中, $q_{i-1,i,t}^j$ 是第 $i-1$ 个水库与第 i 个水库在 t 时段

的区间来水流量(m^3/s); $q_{i-1,t}^x$ 是第 $i-1$ 个水库在 t 时段的下泄流量(m^3/s); $q_{i-1,t}^x$ 是第 i 个水库在 t 时段在 t 时段的入库流量(m^3/s)。

2 基于深度强化学习算法的水光互补优化调度

2.1 深度确定性策略梯度算法

深度确定性策略梯度(DDPG)算法使用经验回放机制, 使用一个确定性策略网络(Actor)直接输出具体的动作值, 而不是输出一个动作的概率分布。这种确定性策略符合连续动作空间, 因为在连续空间中, 动作值是无限且需要精确控制的。DDPG 算法还结合了深度神经网络和策略梯度方法, 能够有效地训练复杂的神经网络模型。引入了 AC(Actor-Critic)算法框架, 包括 Actor 网络、Target-Actor 网络、Critic 网络、Target-Critic 网络 4 个网络, 如图 1 所示。

Actor 网络与环境交互, 输入环境的状态 S , 输出动作 A , 输出动作的策略函数为 $\mu(S|\theta)$, 并将环境产生的经验数据 (S, S', A, R) 存储在回放缓冲区中, Critic 网络则是一个价值函数网络, 输入环境的状态 S 和动作 A , 输出选择当前动作所期望获得的回报值得分, 即当前时刻的预测 Q 值, 其计算公式为

$$Q(S, A|w) = R(S, A) + \gamma \max_{A'} Q^*(S', A') \quad (8)$$

式中, $R(S, A)$ 表示在状态 S 下采取动作 A 获得的即时奖励(reward); $Q^*(S, A)$ 表示状态 S 下采取动作 A 后遵循最优策略的动作值函数。

Target-Actor 网络用于生成下一阶段状态 S' , 动作 A' , 而 Target-Critic 网络通过输入下一阶段状态 S' 和动作 A' 后, 得到下一阶段 Q' 值, 然后通过公式得到目标预测的目标 Q 值 Q_t

$$Q_t = R + \gamma Q'(S', A'|w') \quad (9)$$

式中, R 为奖励函数; γ 为衰减系数; w' 为价值网络参数。

Actor 网络作用是不断优化参数 θ , 使其输出的动作 A 的策略 $\mu(S|\theta)$ 能让 Critic 网络输出的 Q 值最大, 需要计算梯度, 进行梯度上升处理

$$\nabla J(\theta) = \nabla_{\mu_{\theta}} Q(S, A|w) \times \nabla_{\theta} \mu(S|\theta) \quad (10)$$

而 Critic 网络的作用就是不断优化其参数 w , 使得 Q 值与 Q_t 值的相差减小, 即最小化

$$L(w) = \min_{E_{S, A, r, S'}} (Q(S, A|w) - Q_t(S', A'|w')) \quad (11)$$

最小化这一过程, 同样可以通过对参数 w 求导进行梯度下降, 如下

$$\nabla J(w) = \nabla_w E_{S, A, r, S'} (Q(S, A|w) - Q_t(S', A'|w')) \quad (12)$$

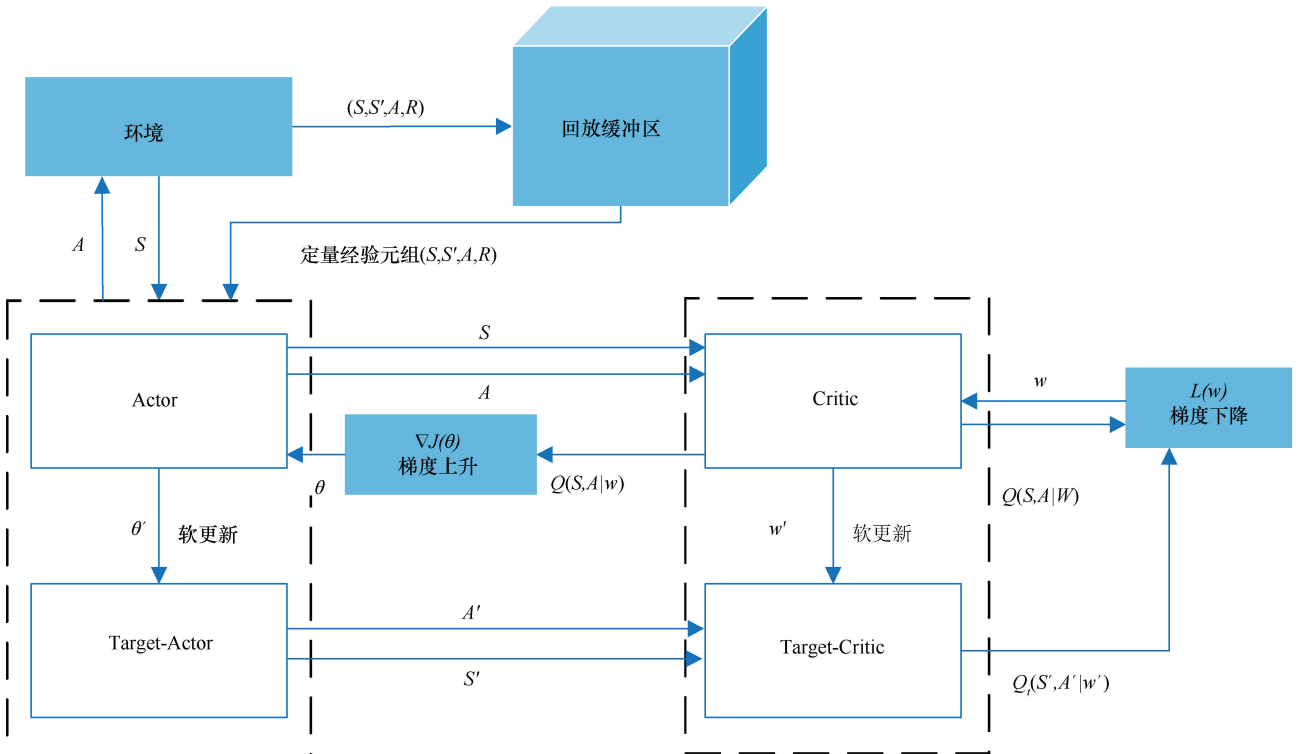


图 1 DDPG 模型流程

Fig. 1 Flowchart of DDPG model

这两者的更新通过回放缓冲机制进行, 将定期从经验回放缓冲区中随机采样一批经验元组 (S, S', A, R) , 提取的动作和状态作为输入值。

由于参数 w 更新频繁, 对 Q 值的估计有剧烈的波动或发散的可能性。所以为了保证 Critic 网络准确地学习和估计 Q 值, 与真实的长期回报 Q_t 值一致, 从而稳定模型, Target-Actor 网络与 Target-Critic 网络的参数 θ' 与 w' 的更新会缓慢而稳定, 方式是采用软更新 (Soft Update)。

Target-Actor 网络定期复制 Actor 网络中的参数 θ 为 θ' , 以减小参数的变动幅度, 从而降低训练过程中的波动性, 参数 θ' 更新公式为

$$\theta' \leftarrow \tau\theta + (1 - \tau)\theta' \quad (13)$$

式中, τ 为更新系数, 通常在 0.001 到 0.005 之间。

Target-Critic 网络同样软更新 (Soft Update) 的方式, 更新下一阶段参数 w' , 公式为

$$w' \leftarrow \tau w + (1 - \tau)w' \quad (14)$$

2.2 基于深度强化学习算法的水光互补优化调度模型智能体设计

为达到目标函数(1), 梯级电站建模按照“以水定电”模式, 即设置调度期始末水位, 输入天然入库流量, 求解达到目标函数极值的梯级电站系统的运行情况。模型环境变量设计包括梯级电站环境变量和光伏变量, 将第 d 天第 t 时段初水库水位 $Z_{i,t}$, 时段末的水位 $Z_{i,t+1}$ 及光伏实时出力 $N_{m,d,t}^h$ 作为强化学习算法的状态空间 S_t 。

水光互补过程引入需调、可调机理量化日内光伏消纳情况, 当输电通道上限确定时, 需调是指在梯级出力上下限限制下, 保证日内水力发电效益不受损失的情况下, 为消纳超出用电通道的光电, 梯级出力可以减小的出力, 可调是指在梯级出力上下限限制下, 保证日内水力发电效益不受损失的情况下, 水电可以加大出力能力^[14]。其中需调 $G_{m,d,t}$ 计算公式为

$$G_{m,d,t} = \begin{cases} 0, & N_{m,d,t}^c < N_{m,d,t}^{h,p} \\ (N_{m,d,t}^c - N_{m,d,t}^{h,p})\Delta t, & N_{m,d,t}^c \geq N_{m,d,t}^{h,p} \end{cases} \quad (15)$$

式中, $N_{m,d,t}^{h,p}$ 为第 m 月第 d 天第 t 时段水光互补前系统的水光合计出力 (万 kW); $N_{m,d,t}^c$ 为单位时段通道限制 (万 kW); Δt 为计算时段长度, $\Delta t = 1$ h。

可调 $F_{m,d,t}$ 计算公式为

$$F_{m,d,t} = \begin{cases} 0, & N_{m,d,t}^{h,u} < N_{m,d,t}^{h,p} \\ (N_{m,d,t}^{h,u} - N_{m,d,t}^{h,p})\Delta t, & N_{m,d,t}^{h,u} \geq N_{m,d,t}^{h,p} > N_{m,d,t}^{h,l} \end{cases} \quad (16)$$

式中, $N_{m,d,t}^{h,p}$ 为梯级电站在第 m 月第 d 天第 t 时段水光互补前的水光合计出力 (万 kW); $N_{m,d,t}^{h,u}$ 为第 m 月第 d 天第 t 时段瞬时梯级电站出力上限 (万 kW); $N_{m,d,t}^{h,l}$ 为第 m 月第 d 天第 t 时段瞬时梯级电站出力下限 (万 kW); Δt 为计算时段长度, $\Delta t = 1$ h。

如图 2 所示, 当可调 \geq 需调时, 梯级水电的互补能力能够完全的消纳光伏出力, 可通过将需调、可调数值进行强化学习建模, 在求解目标函数最大水光合计发电量的同时, 保证光伏消纳率的提高。在日内水光互补后, 生成实时水光互补系统出力 $W_{m,d,t}$ 。结合以上分析可以为强化学习模型设置环境变量 $S_t = [Z_{i,t}, Z_{i,t+1}, N_{m,d,t}^{h,p}, G_{m,d,t}, F_{m,d,t}, N_{m,d,t}^{h,u}, N_{m,d,t}^{h,l}, W_{m,d,t}]$ 。

不同于经典的马尔可夫决策过程, DDPG 模型状态转移没有明确的转移概率, 即 $P(S_{t+1} | S_t, A_t)$, 这意味着 DDPG 在给定状态 S_t 和动作 A_t 的情况下, 下一个状态 S_{t+1} 是确定的, 而不是基于概率的。因此, DDPG 不依赖显式的转移概率, 而是直接学习确定性策略, 即 Actor 网络根据状态 S_t 通过参数 θ 生成确定的动作 A_t 。所以 DDPG 状态转移主要通过水量平衡方程、上下游水力联系计算, 动作 A_t 将包括本时段的平均下泄流量 $q_{i,t}$, 即 $A_t = [q_{i,t}]$, 该动作位于连续的约束空间内, 初始设置发电流量约束, 限制电站出力约束函数定义动作空间, 即可通过水量平衡方程、上下游水力联系完成 $Z_{i,t}, Z_{i,t+1}, N_{m,d,t}^{h,u}, N_{m,d,t}^{h,l}, N_{m,d,t}^{h,p}$ 的更新, 之后会计算可调与需调, 当时间尺度到达 24 时, 完成日内状态转移, 生成逐小时水光互补后的水光消纳电量 $W_{m,d,t+1}$ 。

奖励函数设计的目标是最大化水光系统的发电量, 需要与目标函数(1)对应, 故直接设置奖励函数 $R_{m,d,t}$ 代表时段水光系统出力值为

$$R_{m,d,t} = \sum_{t=1}^T W_{m,d,t} \Delta t \quad (17)$$

式中, $W_{m,d,t}$ 为第 m 月第 d 天第 t 时段水光合计发电量 (万 kWh); Δt 为计算时段长度, $\Delta t = 1$ h; T 为日内小时数, $T = 24$ 。

对强化学习算法而言, 约束函数并不像传统的优化问题那样显式处理, 除限制动作空间范围外, 常以惩罚函数的形式在智能体训练中进行量化处理, 以水位限制作为惩罚函数 $P_{1,m,d,t}$, 设置一个极大的惩罚数值以训练智能体不会出现不满足约束条件例如水光合计电量小于梯级瞬时下限的状况

$$P_{1,m,d,t} = \begin{cases} 0, & \text{满足约束要求} \\ -10^{10}, & \text{else} \end{cases} \quad (18)$$

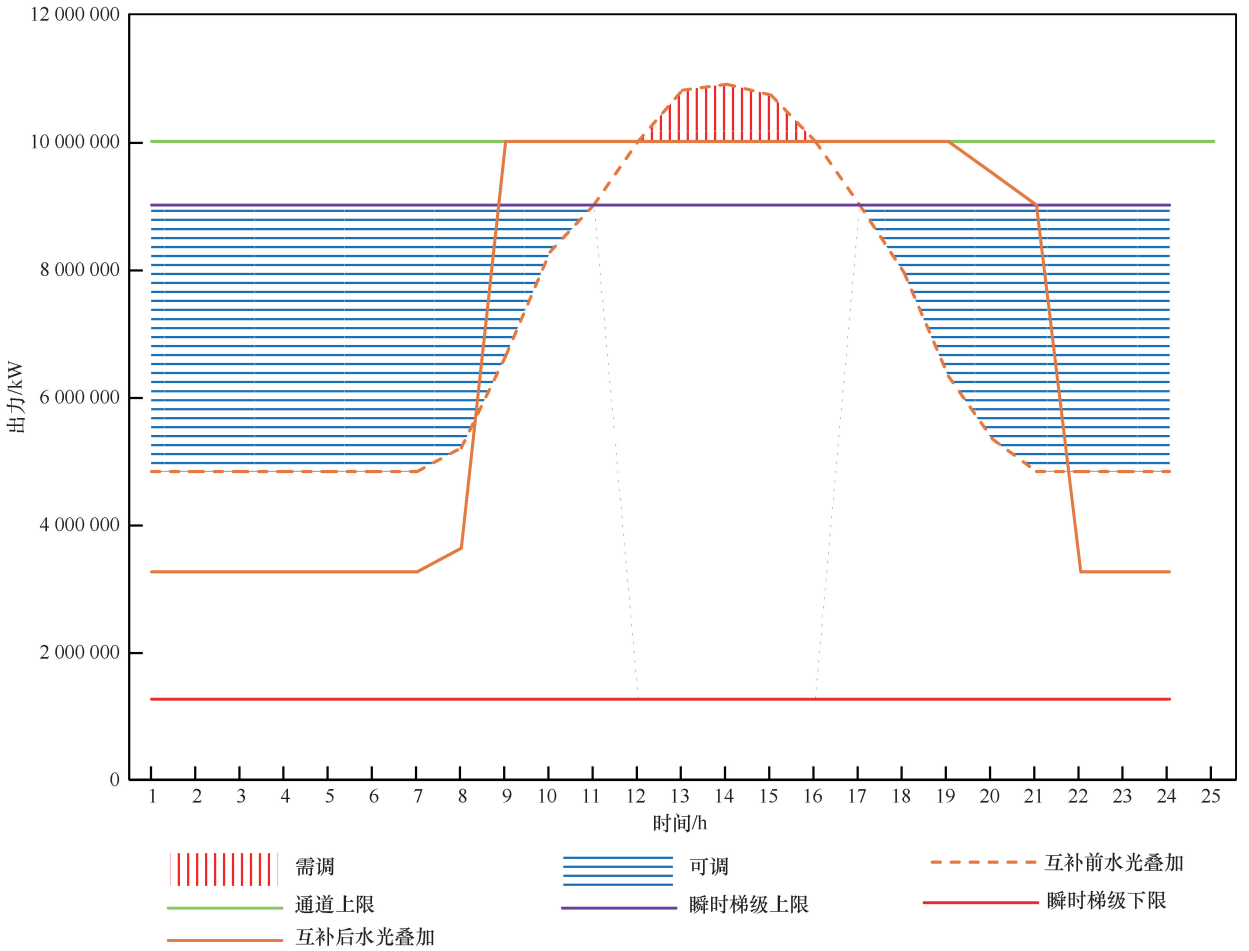


图2 水光互补过程示意

Fig. 2 Schematic diagram of water-light complementary process

此外可调 \geq 需调时,才能保证光伏的完全消纳,故设置惩罚函数 $P_{2,m,d}$ 为

$$P_{2,m,d} = F_{m,d} - G_{m,d} F_{m,d} < G_{m,d} \quad (19)$$

2.3 优化结果检验

NSGA-II 算法在水光互补领域已有较多应用,丁士东等^[46]将波动最小与效益最高作为目标对水光互补系统进行求解,马靖宇等^[47]考虑系统稳定性和效益最大作为目标求解,体现了 NSGA-II 算法一定的收敛性。为验证考虑水光互补机理的深度强化学习 DDPG 模型计算结果的有效性与准确性,对照模型将选用初始 DDPG 模型以及最大年水光合计发电量、最大光伏消纳电量的双目标 NSGA-II 优化模型,其中 NSGA-II 光伏消纳电量最大目标公式为

$$W^p = \max \sum_{d=1}^D \sum_{t=1}^T \sum_{m=1}^M W_{m,d,t}^p \quad (20)$$

$$W_{m,d,t}^p = \begin{cases} (N_{m,d,t}^c - N_{m,d,t}^h) \Delta t, & N_{m,d,t}^{h,p} > N_{m,d,t}^c \\ N_{m,d,t}^p \Delta t, & N_{m,d,t}^{h,p} < N_{m,d,t}^c \end{cases} \quad (21)$$

式中, W^p 为全年光伏消纳电量(万 kWh); $W_{m,d,t}^p$ 为第 m 月第 d 天第 t 时段光伏消纳电量(万 kWh); $N_{m,d,t}^{h,p}$ 为第 m 月第 d 天第 t 时段水光互补前的水光合计出力(万 kW); $N_{m,d,t}^c$ 为单位时段通道限制(万 kW); $N_{m,d,t}^p$ 为第 d 天第 t 时段水光互补前的系统光伏出力(万 kW); $N_{m,d,t}^h$ 为第 d 天第 t 时段水光互补前的系统水电出力(万 kW); M 为该年内的月数, $M=12$; D 为当月天数; Δt 是计算时段长度, $\Delta t=1\text{ h}$; T 为日内小时数, $T=24$ 。

而 NSGA-II 作为遗传算法,种群与迭代次数设置对计算结果起到较大影响,故设置种群数为 100,迭代次数设置 1 000 代,结果如图 3 所示,此时解集稳定,不再有较大变化。

初始 DDPG 模型不考虑可调、需调进行建模,环境变量 $S_t = [Z_{i,t}, Z_{i,t+1}, N_{m,d,t}^h, N_{m,d,t}^u, N_{m,d,t}^l, W_{m,d,t}]$ 直接对水光出力进行叠加并舍弃超出通道上限的出力,其状态转移同样水量平衡方程、上下游水力联系完成。奖励函数 $R_{m,d,t}$ 与惩罚函数 $P_{1,m,d,t}$ 不变,

表 1 不同模型年内发电量

Table 1 Annual power generation of different models

模型	光伏发电量/亿 kWh	水电发电量/亿 kWh	总发电量/亿 kWh	计算时间/s
DDPG(考虑可调需调)	129.93	471.91	601.84	67
DDPG(初始)	123.17	487.39	610.56	59
NSGA-II	118.09	503.78	621.87	1183

重新设置惩罚函数及弃光函数 $P_{3,m,d,t}$, 其公式为

$$P_{3,m,d,t} = \begin{cases} 0, & N_{m,d,t}^{h,p} < N_{m,d,t}^c \\ (N_{m,d,t}^c - N_{m,d,t}^{h,p})\Delta t, & N_{m,d,t}^{h,p} > N_{m,d,t}^c \end{cases} \quad (22)$$

式中, $N_{m,d,t}^c$ 为单位时段通道限制(万 kW); $N_{m,d,t}^{h,p}$ 为第 m 月第 d 天第 t 时段水光互补前的水光合计出力(万 kW); Δt 是计算时段长度, $\Delta t=1\text{ h}$ 。

计算结果如表 1 所列, 两个 DDPG 模型的计算时间显著低于 NSGA-II 模型。采用可调、需调建模的 DDPG 模型在光伏发电量上比初始 DDPG 增加 6.76 亿 kWh, 比 NSGA-II 模型高出 11.84 亿 kWh。如图 4 所示, 尽管 NSGA-II 在梯级电站发电量和总发电量上较大, 但其在设置始末水位限制条件后, 过快达到较高的水头, 而且水位波动较大, 这会使发电流量增加, 在发电通道一定的情况下挤占光伏发电空间, 减少光伏消纳量。

如图 3 所示, NSGA-II 算法通过种群迭代优化水位, 易陷入局部最优, 且种群分布较窄。初始 DDPG 虽通过惩罚函数减少弃光, 光伏消纳效果略优于 NSGA-II, 但仍未涉及水光互补的内在机理, 限制了光伏消纳率的进一步提升。相比之下, 考虑“可调、需调”机制的 DDPG 模型更好地实现了光伏消纳目标, 有效提高了水光互补系统的发电量。

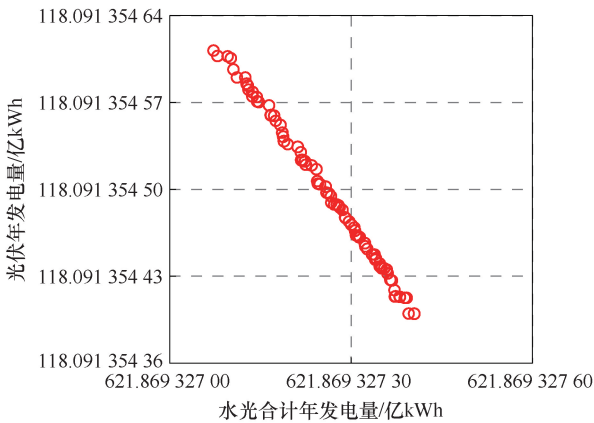


图 3 NSGA-II 双目标寻优

Fig. 3 NSGA-II dual-objective optimization

综上, DDPG 模型展现了更高的求解速度与有效性, 而在水光互补调度中, 考虑可调、需调机制的底

层机理模型不仅优化了水光合计发电量, 也在模型内部充分体现了光伏消纳目标。

3 实例研究

3.1 研究区概况

所选取澜沧江上游河段 371 km, 平均年径流量达 211.6 亿 m^3 , 多年平均太阳能年总辐射量达 5 800~6 700 MJ/m^2 , 具有丰富的水光资源。水光互补系统选取历史实测数据为基准, 其中气象数据选取 Solargis 数据库近十年中各月平均值作为作为典型气象年, 水文径流量选取实测数据中的典型丰水年、平水年与枯水年。

上游水光互补系统有光伏电站装机 1 000 万 kW, 梯级水电站设置三种配置方案进行对比计算, 其中方案一为 1 000 万 kW、方案二为 952.5 万 kW、方案三为 862 万 kW, 梯级水电站具体参数如表 2 所列。

表 2 梯级电站水电站主要工程参数

Table 2 Key engineering parameters of cascade hydropower stations

电 站	装机容量/万 kW			调节库容/亿 m^3	正常蓄水位/m
	方案一	方案二	方案三		
班 达	150	150	150	3.34	3 054
如 美	260	260	220	24.33	2 895
邦 多	72	72	64	0.12	2 605
古 学	210	210	190	0.45	2 535
曲孜卡	40.5	40.5	38	0.13	2 345
古 水	252	220	200	6.34	2 267

3.2 调度结果分析

将水光数据输入 DDPG 模型进行训练。在训练开始时, 模型的累积收益值呈现较大幅度的波动, 这表明模型在对环境进行全局搜索, 处于探索学习阶段。在经过 500 次迭代后, 累计收益值迅速上升, 并在 600 次迭代后趋于稳定。模型在不同方案, 不同代表年均能收敛稳定, 显示出良好的训练效果(见图 5)。

经过优化训练, 水光互补后梯级电站与光伏电站消纳电量, 每月水光资源消纳率如图 6 所示和表 3 所列。由表 3 可知, 方案梯级水电站配置容量越大, 梯级水电站及系统总发电量越高, 但光伏消纳电量越小。三个方案的光伏消纳率均高于 80%, 且方案

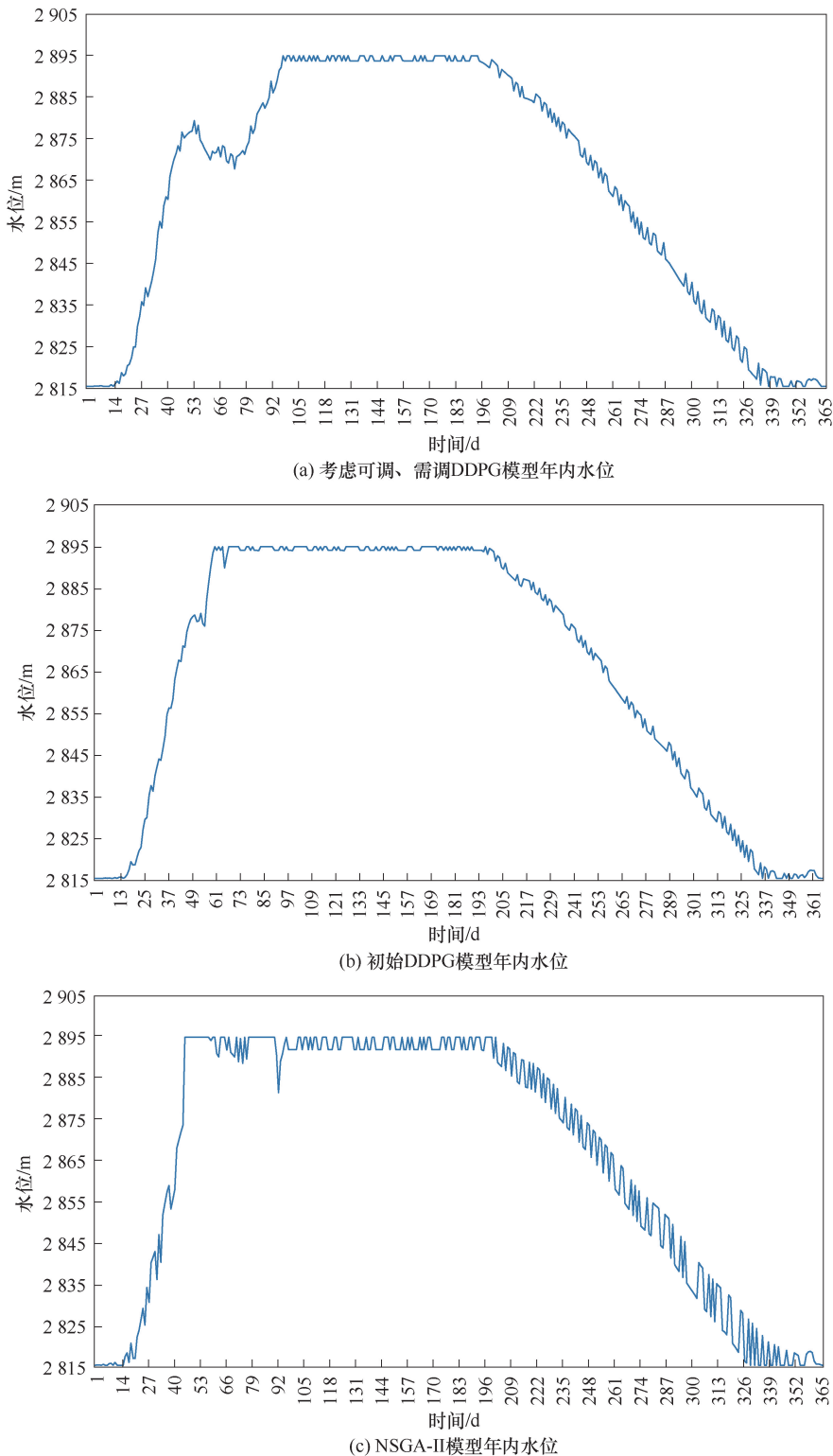


图4 不同模型年内水位

Fig. 4 Annual water levels of different models

三的平均光伏消纳率为 93.68%，大幅高于方案一和方案二，这是因为在水光互补过程中，梯级电站会对光伏出力的波动性进行互补，但当互补损失的梯级水电站发电量远远小于互补增加的光伏发电量时，水光

互补系统会保证目标函数设置的总发电量最大从而弃光。

丰水年到平水年，光伏资源消纳率提升最大，三个方案分别提升了 5.8%、7.3% 和 5.9%，平水年到枯水年这一范围，方案三提升 1%，小于方案一和方案二的 3.71% 和 3.08%，这是由于方案三在此区间光伏消纳能力已达水光系统互补能力上限，无法有过大的提升，表明方案三在平水年就能达到光伏消纳能力为峰值。对比三个方案，862 万 kW 梯级装机能更充分地利用光伏资源。

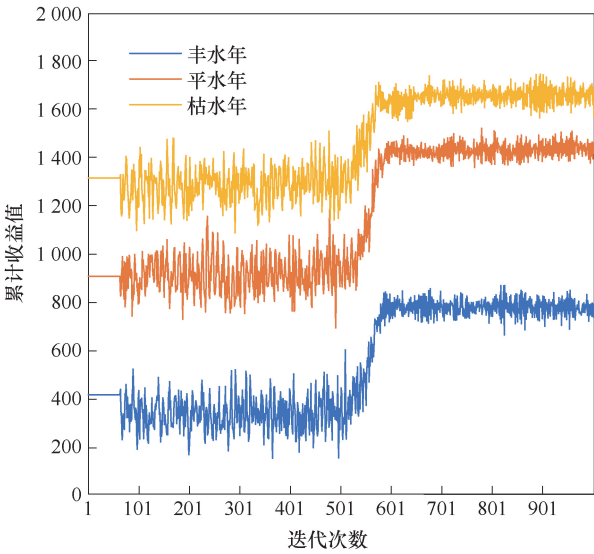
同一方案而言，年内在来水情况越枯，光伏消纳率越高。如图 6 所示，分析年内逐月发电量及光伏消纳率可得汛期径流来水情况偏丰，水电量过大导致挤占通道，此时弃光增加，枯水期来水较少，能够消纳更多光伏出力。

4 结果讨论

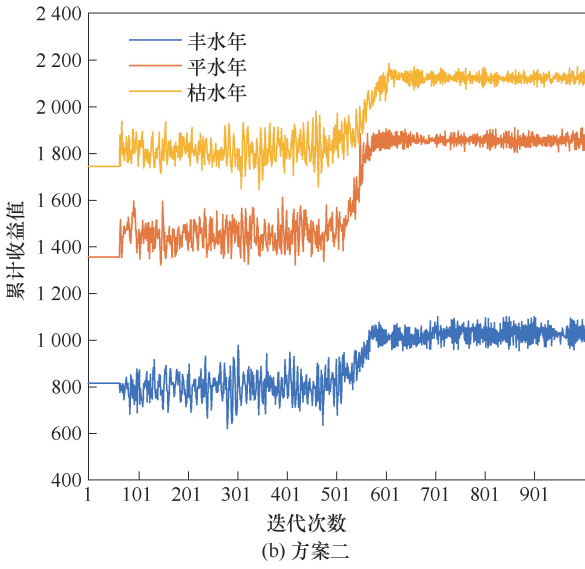
在相同算例条件下，枯水年的光伏消纳能力更强，这是由于来水量较少时，水电和光伏系统的总发电量离通道上限较远，水库蓄水压力减小，调节能力增强，使水电系统可以更灵活地平衡电网负荷和光伏发电的波动性。当系统在水光互补前的出力已接近通道上限时，尽管梯级电站的出力上限提升，但调节能力受限，导致光伏弃光现象增加，尤其在丰水年的汛期。此特性为水光互补系统的容量配置和调度优化提供了重要参考。

基于深度强化学习 (DDPG) 算法的水光互补调度优化方案能在连续、复杂的解空间中高效工作，与传统优化方法相比，不依赖于线性

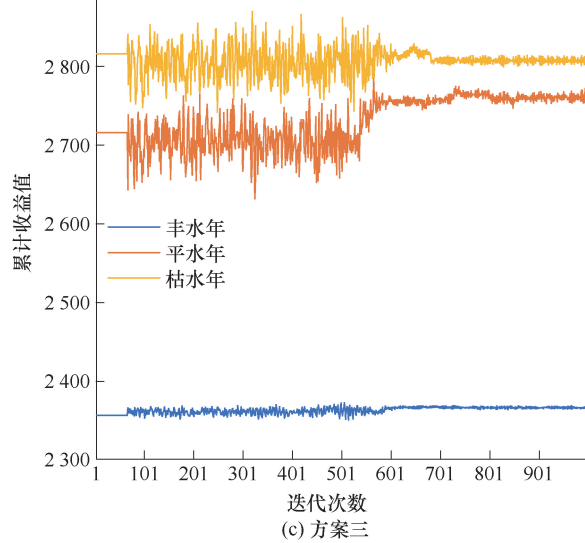
化或过多简化假设，解决了复杂约束条件下的全局优化问题。即便不深入建模水光互补机制，该算法仍能快速生成结果；而在融入物理机理建模后，DDPG 算法显著提升了光伏消纳目标的实现效果，有助于获得



(a) 方案一



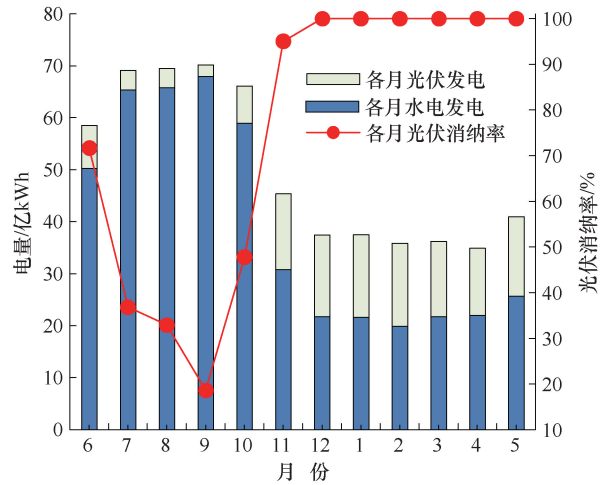
(b) 方案二



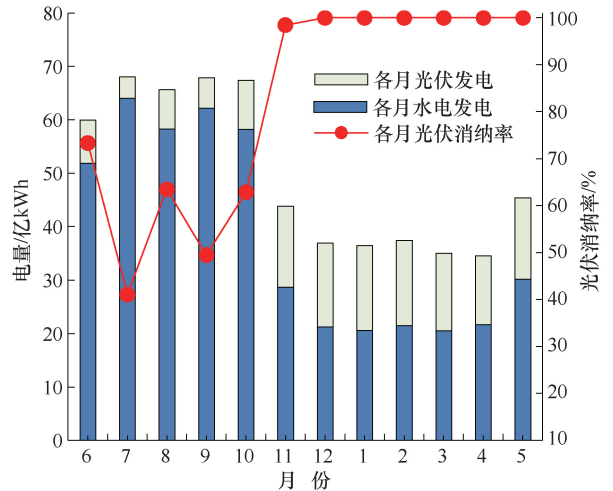
(c) 方案三

图5 不同方案累计收益值迭代对比

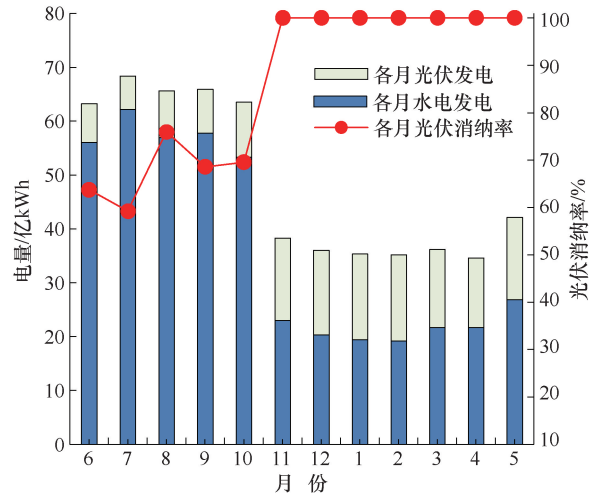
Fig. 5 Iterative comparison of cumulative profit values of different schemes



(a) 丰水年逐月发电量及光伏消纳率



(b) 平水年逐月发电量及光伏消纳率



(c) 枯水年逐月发电量及光伏消纳率

图6 三个代表年逐月发电量及光伏消纳率

Fig. 6 Monthly power generation and photovoltaic consumption rates of three representative years

更优的调度结果, 并提高计算效率。

目前在水库调度、水光互补调度领域应用强化

表3 三个方案发电量及光伏消纳率

Table 3 Power generation and photovoltaic consumption rates of three schemes

	代表年份	光伏电量 /亿 kWh	光伏消纳量 /亿 kWh	梯级电站发 电量/亿 kWh	互补总发电量 /亿 kWh	互补光伏 消纳率/%	多年平均 利用率/%
方案一	丰水年	167.80	129.93	471.91	601.84	77.43	82.53
	平水年	167.80	139.67	458.83	598.50	83.23	
	枯水年	167.80	145.90	438.14	584.04	86.94	
方案二	丰水年	167.80	133.99	464.13	598.12	79.85	85.76
	平水年	167.80	146.27	448.89	595.16	87.17	
	枯水年	167.80	151.44	426.60	578.04	90.25	
方案三	丰水年	167.80	150.00	443.52	593.52	89.39	93.68
	平水年	167.80	159.95	426.17	586.12	95.32	
	枯水年	167.80	161.63	410.87	572.50	96.32	

学习的研究较少, 而采用强化学习算法处理于水光互补系统的调度问题, 为应对非线性、多变量和不确定性问题提供了全新技术手段。将物理机理建模与深度强化学习算法相结合, 突破了传统机器学习算法对领域知识理解不足的限制, 针对特定问题的效果能有较大提升。适用于多种水光互补场景, 特别是梯级水库调度和非平稳光伏出力条件下的调度优化需求。同时, 该方法对不同时间尺度(如日内、月度、年度)均表现出较强的适应性, 能够根据具体需求灵活调整优化目标, 例如光伏消纳优先、削峰填谷或经济性优化等。

尽管在水光互补调度问题中结合物理模型机理的DDPG算法有高效性和适用性, 但在更大规模的多能互补系统中, 现有DDPG算法模型可能会出现局限性, 需要考虑水风光储火多种能源形式重新设计更高效的机理模型。此外虽然引入了物理机理建模以增强模型的可解释性, 但可能无法完全涵盖系统的多重运行任务。例如, 除了电网消纳光伏电量这一主要目标外, 水光互补系统可能还需同时满足其他任务需求, 如供水、灌溉、防洪和生态流量等, 而模型中未能全面考虑或准确模拟这些需求, 可能导致调度策略在实际运行中出现偏差, 从而影响整体系统的运行效率与可靠性。

未来, 随着水库和水光互补调度在计算机科学和人工智能领域的深入研究, 探索基于强化学习的预训练模型在更长时间尺度和高度随机性调度问题中的应用将成为重要趋势。预训练强化学习模型能够在大规模、复杂场景下, 利用历史数据和模拟环境训练通用的策略模板, 在训练完成后能够直接实时生成调度方案, 从而显著降低训练时间和计算资源消耗; 此外结合物理机理的预训练深度强化学习模型具有应对动态变化与不确定性问题的独特优势。一方面, 物理机理建模可以为强化学习算法提供更加精准的约束条件,

使其优化结果更贴近实际运行要求; 另一方面, 预训练模型通过引入多样化场景的模拟与量化自然资源的不确定性, 可有效训练出具有高适应性和强鲁棒性的调度策略。此外, 这种方法能够在面临极端条件或未曾见过的运行环境时, 依托预训练经验实现快速调整, 保证系统的稳定性与优化性能。

这一方向为水资源和可再生能源领域解决复杂多尺度问题提供了新的方法与理论支持。例如, 预训练强化学习模型可以在涵盖丰水期、枯水期等多水文条件的广泛场景中进行学习, 并将其调度经验迁移到特定梯级水库或区域水光互补系统的实际应用中。在大规模梯级调度、跨区域多能互补系统优化等问题中, 预训练模型可以显著提高效率和精度, 减少对实时计算的依赖。

5 结论

提出基于深度强化学习的水光互补优化调度模型, 以澜沧江上游大型水光互补基地为算例, 以下结论:

(1) DDPG算法通过融合“需调”与“可调”概念, 有效提升了光伏消纳能力, 并显著加速了优化调度过程, 验证了有效性与求解的快速性。

(2) 设置1 000万kW、952.5万kW、862万kW三种梯级电站装机方案进行对比计算, 得出862万kW方案光伏资源消纳能力最高, 在来水情况为平水年时就已经能较为完全利用水电系统的互补能力。并且对于同一方案而言, 汛期(6—10月)径流来水情况偏丰, 水电量过大会挤占通道, 从而导致光伏消纳率不如年内来水情况较枯的情况。

(3) 基于深度强化学习算法的水光互补优化调度模型, 在不同水文条件下均能快速生成高效的调度方案, 为清洁能源基地的容量配置、运行优化提供了科学的决策支持。

参考文献(References):

- [1] 鲜于虎成, 黄显峰, 李旭, 等. 基于水-光-荷模块分层协同搜索的风险调度策略[J]. 电网技术, 2023, 47(3): 908-918.
XIANYU H C, HUANG X F, LI X, et al. Risk scheduling strategy based on hydro-pv-load module hierarchical cooperative searching [J]. Power System Technology, 2023, 47(3): 908-918.
- [2] 闻昕, 孙圆亮, 谭乔凤, 等. 考虑预测不确定性的风-光-水多能互补系统调度风险和效益分析[J]. 工程科学与技术, 2020, 52(3): 32-41.
WEN X, SUN Y L, TAN Q F, et al. Risk and benefit analysis of hydro-wind-solar multi-energy system considering the one-day ahead output forecast uncertainty [J]. Advanced Engineering Sciences, 2020, 52(3): 32-41.
- [3] 申建建, 王月, 程春田, 等. 水风光互补系统灵活性需求量化及协调优化模型[J]. 水利学报, 2022, 53(11): 1291-1303.
SHEN J J, WANG Y, CHENG C T, et al. Flexibility demand quantification and optimal operation model of water-wind-solar complementary system[J]. Journal of Hydraulic Engineering, 2022, 53(11): 1291-1303.
- [4] HUANG K D, LIU P, MING B, et al. Economic operation of a wind-solar-hydro complementary system considering risks of output shortage, power curtailment and spilled water[J]. Applied Energy, 2021, 290: 116805.
- [5] 吴丰昌. 我国水体污染控制与治理成效、科技支撑与展望[J]. 水利发展研究, 2023, 23(12): 1-8.
WU F C. Effectiveness, scientific and technological support, and prospects for water pollution control and management in China[J]. Water Resources Development Research, 2023, 23(12): 1-8.
- [6] 李国英. 为以中国式现代化全面推进强国建设、民族复兴伟业提供有力的水安全保障: 在 2024 年全国水利工作会议上的讲话[J]. 水利发展研究, 2024, 24(1): 1-10.
LI G Y. Improved water security for China's efforts to build itself into a stronger country and rejuvenate the Chinese nation on all fronts by pursuing Chinese modernization: Speech at the 2024 National Water Conservancy Work Conference [J]. Water Resources Development Research, 2024, 24(1): 1-10.
- [7] 李国英. 深入贯彻落实党的二十大精神扎实推动新阶段水利高质量发展: 在 2023 年全国水利工作会议上的讲话[J]. 水利发展研究, 2023, 23(1): 1-11.
LI G Y. Thoroughly implement the spirit of the 20th National Congress of the Communist Party of China and solidly promote the high-quality development of water conservancy in the new stage: Speech at the National Water Conservancy Work Conference in 2023 [J]. Water Resources Development Research, 2023, 23(1): 1-11.
- [8] YANG Z K, LIU P, CHENG L, et al. Sizing utility-scale photovoltaic power generation for integration into a hydropower plant considering the effects of climate change: A case study in the Longyangxia of China [J]. Energy, 2021, 236: 121519.
- [9] TAN Q, WEN X, SUN Y, et al. Evaluation of the risk and benefit of the complementary operation of the large wind-photovoltaic-hydropower system considering forecast uncertainty [J]. Applied energy, 2021, 285: 116442.
- [10] WEN X, SUN Y, TAN Q, et al. Optimizing the sizes of wind and photovoltaic plants complementarily operating with cascade hydropower stations: Balancing risk and benefit [J]. Applied Energy, 2022, 306: 117968.
- [11] 周永斐, 梅亚东, 谢凡仪, 等. 基于主从博弈的风光蓄网短期优化调度[J]. 水力发电学报, 2021, 40(12): 52-64.
ZHOU Y F, MEI Y D, XIE F Y, et al. Short-term optimal scheduling of wind-photovoltaic-pumped storage grid based on Stackelberg game [J]. Journal of Hydropower Engineering, 2021, 40(12): 52-64.
- [12] 朱燕梅, 黄炜斌, 陈仕军, 等. 水光互补日内优化运行策略[J]. 工程科学与技术, 2021, 53(3): 142-149.
ZHU Y M, HUANG W B, CHEN S J, et al. Intra-day optimization operation strategy of hydro-pv hybrid system [J]. Advanced Engineering Sciences, 2021, 53(3): 142-149.
- [13] 朱燕梅, 陈仕军, 马光文, 等. 计及发电量和出力波动的水光互补短期调度[J]. 电工技术学报, 2020, 35(13): 2769-2779.
ZHU Y M, CHEN S J, MA G W, et al. Short-term complementary operation of hydro-photovoltaic integrated system considering power generation and output fluctuation [J]. Transactions of China Electrotechnical Society, 2020, 35(13): 2769-2779.
- [14] 黄显峰, 格桑央拉, 吴志远, 等. 水光互补能源基地的多时间尺度优化调度[J]. 水力发电, 2022, 48(1): 106-111.
HUANG X F, KELSANG Y L, WU Z Y, et al. Multi-time scale optimization scheduling of hydro-photovoltaic complementary energy base [J]. Water Power, 2022, 48(1): 106-111.
- [15] 鲜于虎成, 黄显峰, 张艳青, 等. 考虑短期波动与弃电风险的水光互补系统中长期双层嵌套优化调度 [J]. 电工技术学报, 2023, 38(21): 5781-5793.
XIANYU H C, HUANG X F, ZHANG Y Q, et al. Medium-and long-term double-layer nested optimal scheduling of hydro-pv complementary system considering short-term power fluctuation and curtailment risk [J]. Transactions of China Electrotechnical Society, 2023, 38(21): 5781-5793.
- [16] 梁艺缤, 畅建霞, 郭爱军, 等. 考虑水光协调性与经济性的水电站优化调度[J]. 水力发电学报, 2022, 41(6): 22-33.
LIANG Y B, CHANG J X, GUO A J, et al. Oimal operation of hydropower stations considering hydro-photovoltaic coordination and economy [J]. Journal of Hydroelectric Engineering, 2022, 41(6): 22-33.
- [17] GONG Y, LIU P, MING B, et al. Identifying the effect of forecast uncertainties on hybrid power system operation: A case study of Longyangxia hydro-photovoltaic plant in China [J]. Renewable

- Energy, 2021, 178: 1303-1321.
- [18] GONG Y, LIU P, MING B, et al. Identifying the functional form of operating rules for hydro-photovoltaic power hybrid power systems[J]. Energy, 2022, 243: 123027.
- [19] 宋柯. 基于场景特征聚类的水光互补系统优化调度[J]. 水电能源科学, 2021, 39(6): 202-205.
- SONG K. Optimal scheduling of hydro-photovoltaic complementary generation system based on scene feature clustering [J]. Water Resources and Power, 2021, 39(6): 202-205.
- [20] 罗彬, 陈永灿, 刘昭伟, 等. 梯级水光互补系统最大化可消纳电量期望短期优化调度模型 [J]. 电力系统自动化, 2023, 47(10): 66-75.
- LUO B, CHEN Y C, LIU Z W, et al. Short-term optimal scheduling model for maximizing expected absorbable power of a cascade hydro-solar complementary system [J]. Automation of Electric Power Systems, 2023, 47(10): 66-75.
- [21] LI H, LIU P. Long-term complementary operation of a large-scale hydro-photovoltaic hybrid power plant using explicit stochastic optimization[J]. Applied Energy, 2019, 238: 863-875.
- [22] OPAN M, UNLU M, OZKALE C, et al. Optimal energy production from wind and hydroelectric power plants[J]. Energy Sources, 2019, 41(13/18): 2219-2232.
- [23] LIU W, ZHU F, CHEN J, et al. Multi-objective optimization scheduling of wind-photovoltaic-hydropower systems considering riverine ecosystem [J]. Energy Conversion & Management, 2019, 196: 32-43.
- [24] 冯欢, 苏华英, 刘臣亮, 等. 基于动态规划与灰狼算法的构皮滩水库防洪优化调度研究[J]. 人民珠江, 2024, 45(11): 62-71.
- FENG H, SU H Y, LIU C L, et al. optimal flood control scheduling of Goupitan reservoir based on dynamic programming and grey wolf algorithm[J]. Pearl River, 2024, 45(11): 62-71.
- [25] 王峰, 王建平, 王亦宁, 等. 改进的显随机动态规划在梯级水库优化调度中的应用[J]. 中国农村水利水电, 2023(8): 7-12.
- WANG F, WANG J P, WANG Y N, et al. Application of the improved explicit stochastic dynamic programming model in the optimal operation of cascade reservoirs [J]. China Rural Water and Hydropower, 2023(8): 7-12.
- [26] 黄显峰, 王宁, 刘志佳, 等. 基于改进 NSGA-II 算法的梯级水库多目标优化调度[J]. 水利水电科技进展, 2024, 44(4): 51-58.
- HUANG X F, WANG N, LIU Z J, et al. Multi-objective optimal operation of cascade reservoirs based on improved nsga-ii algorithm [J]. Advances in Science and Technology of Water Resources, 2024, 44(4): 51-58.
- [27] 曾茂森, 陈帝伊, 许贝贝, 等. 改进粒子群算法的厂内机组避振及负荷分配研究 [J]. 中国农村水利水电, 2023(12): 258-265.
- ZENG M S, CHEN D Y, XU B B, et al. Study on vibration avoidance and load distribution of in-house units based on improved particle swarm optimization algorithm [J]. China Rural Water and Hydropower, 2023(12): 258-265.
- [28] 曾茂森. 水(蓄)风光火的容量配置优化及水电站厂内避振负荷分配[D]. 咸阳: 西北农林科技大学, 2023.
- ZENG M S. Optimization of Capacity Allocation of Hydroelectric Power(Storage)-Wind-Photoelectricity-Thermal Power and Distribution of Anti-vibration Load in Hydropower Plant [D]. Xianyang: Northwest A&F University, 2023.
- [29] 李飞, 李咸善, 李振兴, 等. 基于梯级水电调节的多能联合发电系统短期优化调度 [J]. 电力系统保护与控制, 2022, 50(15): 11-20.
- LI F, LI X S, LI Z X, et al. Short-term optimal scheduling of multi-energy combined generation systems based on the regulation of cascade hydropower stations [J]. Power System Protection and Control, 2022, 50(15): 11-20.
- [30] 宋志红, 刘渊, 江生金, 等. 引江济淮工程(河南段)多目标水量优化调度 [J]. 南水北调与水利科技(中英文), 2024, 22(1): 39-47.
- SONG Z H, LIU Y, JIANG S J, et al. Multi-objective water optimal dispatching of the Yangtze-to-Huaihe River Water Diversion Project (Henan section) [J]. South-to-North Water Transfers and Water Science & Technology, 2024, 22(1): 39-47.
- [31] 陈子豪, 李莹莹, 董国涛. 黄藏寺水库水光互补日内优化调度研究 [J]. 人民黄河, 2023, 45(10): 151-157.
- CHEN Z H, LI Y Y, DONG G T. Research on optimal operation of Huangzangsi reservoir ' s hydro-pv complementation in days [J]. Yellow River, 2023, 45(10): 151-157.
- [32] ARULKUMARAN K, DEISENROTH M P, BRUNDAGE M, et al. Deep reinforcement learning: A brief survey [J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.
- [33] KAMRUZZAMAN M, DUAN J J, SHI D, et al. A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources [J]. IEEE Transactions on Power Systems, 2021, 36(6): 5525-5536.
- [34] HU D E, YE Z H, GAO Y Q, et al. Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization [J]. IEEE Transactions on Smart Grid, 2022, 13(6): 4873-4886.
- [35] LI Z, BAI L, TIAN W, et al. Online control of the raw water system of a high-sediment river based on deep reinforcement learning [J]. Water, 2023, 15(6): 1131.
- [36] WU R, WANG R, HAO J, et al. Multi objective multihydropower reservoir operation optimization with transformer-based deep reinforcement learning [J]. Journal of Hydrology, 2024: 130904.
- [37] 胡鹤轩, 钱泽宇, 胡强, 等. 离散四水库问题基准下基于 n 步 Q-learning 的水库群优化调度 [J]. 中国水利水电科学研究院学报(中英文), 2023, 21(2): 138-147.

- HU H X, QIAN Z Y, HU Q, et al. Optimal scheduling of multi-reservoir system based on n-step Q-learning under discrete four-reservoir problem benchmark[J]. *Journal of China Institute of Water Resources and Hydropower Research*, 2023, 21(2): 138-147.
- [38] 李文武, 周佳妮, 裴本林, 等. 梯级水库深度强化学习长期随机优化调度研究[J]. *水力发电学报*, 2023, 42(11): 21-32.
- LI W W, ZHOU J N, PEI B L, et al. Study on long-term stochastic optimal operation of cascade reservoirs by deep reinforcement learning [J]. *Journal of Hydropower Engineering*, 2023, 42(11): 21-32.
- [39] BUI V H, HUSSAIN A, KIM H M. Double deep Q-learning-based distributed operation of battery energy storage system considering uncertainties[J]. *IEEE Transactions on Smart Grid*, 2019, 11(1): 457-469.
- [40] ZHANG Y, WANG X, WANG J, et al. Deep reinforcement learning based volt-var optimization in smart distribution systems [J]. *IEEE Transactions on Smart Grid*, 2020, 12(1): 361-371.
- [41] 伍鑫, 陈植欣, 温庆博, 等. 基于强化学习的非常规水资源优化配置模型[J]. *水力发电学报*, 2021, 40(7): 23-31.
- WU X, CHEN Z X, WEN Q B, et al. Optimal allocation model of unconventional water resources based on reinforcement learning [J]. *Journal of Hydropower Engineering*, 2021, 40(7): 23-31.
- [42] 徐刚, 舒远丽, 任玉峰, 等. 基于深度学习的三峡水库实时防洪调度模型[J]. *水力发电学报*, 2022, 41(3): 60-69.
- XU G, SHU Y L, REN Y F, et al. Real-time flood control operation model of Three Gorges Reservoir based on deep learning[J]. *Journal of Hydropower Engineering*, 2022, 41(3): 60-69.
- [43] LILICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB/OL]. (2020-03-20) [2024-09-22]. <https://arxiv.org/abs/1509.02971v2>.
- [44] 杨挺, 赵黎媛, 刘亚闯, 等. 基于深度强化学习的综合能源系统动态经济调度[J]. *电力系统自动化*, 2021, 45(5): 39-47.
- YANG T, ZHAO L Y, LIU Y C, et al. Dynamic economic dispatch for integrated energy system based on deep reinforcement learning[J]. *Automation of Electric Power Systems*, 2021, 45(5): 39-47.
- [45] 李涛, 胡维昊, 李坚, 等. 基于深度强化学习算法的光伏-抽蓄互补系统智能调度[J]. *电工技术学报*, 2020, 35(13): 2757-2768.
- LI T, HU W H, LI J, et al. Intelligent economic dispatch for pv-phs integrated system: a deep reinforcement learning-based approach[J]. *Transactions of China Electrotechnical Society*, 2020, 35(13): 2757-2768.
- [46] 丁士东, 曾平良, 邢浩, 等. 一种风光水一体化发电系统中长期多目标优化运行方法[J]. *电力科学与工程*, 2019, 35(11): 17-25.
- DING S D, ZENG P L, XING H, et al. A medium and long term multi-objective optimal operation method for integrated wind/PV/hydro power generation system [J]. *Electric Power Science and Engineering*, 2019, 35(11): 17-25.
- [47] 马靖宇, 刘继春. 电力市场下计及负荷匹配和检修损失的水光互补系统中长期发电计划制定方法[J]. *桂林理工大学学报*, 2021, 41(3): 644-653.
- MA J Y, LIU J C. Medium and long term generation planning method in hydropower and photovoltaic complementary system considering load matching and maintenance loss in power market[J]. *Journal of Guilin University of Technology*, 2021, 41(3): 644-653.

(责任编辑 王海锋)