

基于多策略深度强化学习的水风光互补调峰调度

冯仲恺^{1,2,3}, 罗 涛^{1,2,3}, 牛文静⁴, 李建兵⁵, 申建建⁶, 武新宇⁶, 廖胜利⁶

(1. 河海大学水灾害防御全国重点实验室; 2. 河海大学水文水资源学院;
3. 河海大学洪涝灾害风险预警与防控应急管理部重点实验室; 4. 长江水利委员会水文局;
5. 中国电建集团成都勘测设计研究院有限公司; 6. 大连理工大学水电与水信息研究所)

摘要:为提升水风光互补系统应对风光出力不确定性的能力,以典型场景划分与策略适配为基础,提出了一种融合场景自适应判别与多策略深度强化学习的水风光互补系统调峰调度方法。该方法在离线训练阶段采用主成分分析与 K -means 聚类算法,在季节尺度对历史风光出力日曲线形态特征进行聚类分析,识别并构建不同出力规律的典型场景风光数据集,并基于该数据集在统一目标函数与约束条件下,针对各场景独立构建仿真环境,训练场景专属深度强化学习智能体,进而构建策略知识库;在线调度阶段通过设计 BiLSTM-Transformer 判别器,以当前季节特征与未来一日标准化风光出力预测序列为输入,输出场景隶属概率分布及标签,而后匹配并调用策略知识库中对应场景的智能体来生成调度指令,通过可行域投影处理确保满足约束条件,形成场景辨识-策略匹配-滚动调度的动态机制。工程验证结果表明,该方法在满足全部工程约束的前提下,可显著降低负荷波动标准差与峰谷差,决策响应时间达毫秒级;场景判别准确率达 97.54%,且具有较强的跨场景适应性与鲁棒性。

关键词:水风光互补;多策略优化;深度强化学习;调峰调度

中图分类号:TV213.4 文献标志码:A 文章编号:1004-6933(2026)05-0009-13

Peak-shaving operation for hydro-wind-solar complementary based on multi-strategy deep reinforcement learning//Feng Zhongkai^{1,2,3}, Luo Tao^{1,2,3}, Niu Wenjing⁴, Li Jianbing⁵, Shen Jianjian⁶, Wu Xinyu⁶, Liao Shengli⁶
(1. State Key Laboratory of Water Disaster Prevention, Hohai University; 2. College of Hydrology and Water Resources, Hohai University; 3. Key Laboratory of Flood Disaster Risk Warning, Prevention and Mitigation, Ministry of Emergency Management, Hohai University; 4. Bureau of Hydrology, Changjiang Water Resources Commission; 5. POWERCHINA Chengdu Engineering Corporation Limited; 6. Institute of Hydropower & Hydroinformatics, Dalian University of Technology)

Abstract: To enhance the ability of hydro-wind-solar complementary systems to cope with the uncertainty of wind and solar power output, a peak-shaving scheduling method integrating scenario adaptive discrimination and multi-strategy deep reinforcement learning for hydro-wind-solar complementary systems is proposed, based on typical scenario division and strategy adaptation. In the offline training phase, principal component analysis and the K -means clustering algorithm are used to cluster the morphological characteristics of daily historical wind and solar power output curves at the seasonal scale, identifying and constructing typical scenario datasets with different output patterns. Based on these datasets and under unified objective functions and constraints, simulation environments are independently built for each scenario to train scenario-specific deep reinforcement learning agents, thereby constructing a strategy knowledge base. In the online scheduling phase, a BiLSTM-Transformer discriminator is designed, which takes current seasonal features and standardized wind and solar power forecast sequences for the next day as input, outputs the scenario membership probability distribution and labels, then matches and calls the corresponding agent from the strategy knowledge base to generate scheduling instructions. Feasible region projection is applied to ensure constraint satisfaction, forming a dynamic mechanism of “scenario identification-strategy matching-rolling scheduling”. Case verification results show that, under the premise of satisfying all engineering constraints, the proposed method significantly reduces the standard deviation of load fluctuations and the peak-valley difference, with decision response time at the millisecond level. The scenario discrimination accuracy reaches 97.54%, and the proposed method exhibits strong cross-scenario adaptability and robustness.

Key words: hydro-wind-solar complementary; multi-strategy optimization; deep reinforcement learning; peak-shaving operation

基金项目:国家自然科学基金项目(52441901,52379009);江苏省自然科学基金优秀青年基金项目(BK20240189);北京江河水利发展基金会——水利青年科技英才项目(JHYC202310)

作者简介:冯仲恺(1988—),男,教授,博士,主要从事水文水资源研究。E-mail:myfellow@163.com

水风光互补电力系统作为支撑新型电力系统构建与“双碳”目标实现的关键路径,在提升可再生能源就地消纳能力、增强电网运行灵活性及保障能源安全方面具有重要作用^[1-5]。然而,实际运行中须协同满足电力负荷跟踪、生态流量保障、安全水位维持及出力爬坡限制等多重约束,形成复杂的多维约束体系^[6-7];风电、光伏出力受气象条件随机驱动,呈现显著间歇性与强不确定性,导致调峰调度问题兼具高维特征、强变量耦合、非凸目标函数及非线性约束等复杂属性,对调度决策的精准性与时效性构成严峻挑战^[8-12]。传统方法中,动态规划、线性规划及非线性规划在中小规模调度问题中具有一定适用性,但易因模型简化引入误差,或因维度灾难(尤其在风光出力场景下)与实时性不足难以满足工程需求^[13-15];群体智能算法如粒子群优化、遗传算法等,依托群体协作机制开展全局寻优,但其收敛稳定性对参数设置高度敏感,且在多约束场景下易出现可行解质量下降及搜索停滞现象^[16-19]。因此,亟须耦合场景驱动与多策略协同优化理论,构建兼顾可行域约束满足性、高效寻优能力与在线响应速度的新型调度方法,应对水风光互补系统调度挑战,提升清洁能源利用效率。

深度强化学习(deep reinforcement learning, DRL)通过端到端学习机制构建策略映射,能够从数据中自主学习复杂决策关系,为应对水风光互补调峰中的高维非线性难题提供了新方向^[20-21]。DRL在训练阶段利用历史数据构建仿真环境,学习负荷、风光出力与水动力响应间的耦合规律;在应用阶段,仅需单次前向推理即可输出调度指令,避免了传统方法各时段重复计算的高昂成本。DRL与水风光互补调度的适配性主要体现在三方面:①可直接在连续状态/动作空间输出连续指令,避免离散化带来的精度损失^[22];②可通过奖励函数同时编码多个目标与约束,将削峰填谷等调度需求统一为可学习的信号^[23-24];③其强大的表征与泛化能力能够适应风光出力的多场景、多模态特征,为解决风光不确定性带来的工况适配问题提供了新路径^[25]。在多能互补调峰场景中,近端策略优化(proximal policy optimization, PPO)通过裁剪式近端更新机制,保障训练稳定性,得到广泛应用^[26];群体相对策略优化(group relative policy optimization, GRPO)通过相对回报估计替代显式价值网络,可进一步降低训练复杂度^[27]。在GRPO基础上,本文面向连续调度问题构建近似群体相对策略优化(approximate group relative policy optimization, A-GRPO),将相对优势估计近似为固定滚动时间窗内蒙特卡罗回报的批内标

准化,并采用基于参考策略的Kullback-Leibler(KL)散度约束限制策略更新幅度,以提高训练稳定性。尽管如此,DRL直接应用于水风光调峰仍面临三方面挑战:约束密集交织(涉及库水位、下泄流量、出力爬坡率及出力边界等)、滚动运行失配(实际工况偏离训练场景分布)、多时间尺度耦合导致的奖励稀疏性。因此,DRL需与可行域投影技术及约束化奖励设计协同优化,方可满足工程要求。

风电与光伏出力具有季节性强且随机的特性,水风光系统呈多场景非平稳特征,单一策略跨场景易退化^[28-30];同时,梯级水库耦合性强、约束复杂,调度需在抑制剩余负荷波动的同时满足运行边界要求^[31]。针对上述难题,本文提出融合场景判别与多策略深度强化学习的水风光互补调峰方法:离线阶段依托历史风光出力序列构建典型场景库,在各场景下独立训练调度策略,形成场景专属策略集合;在线阶段利用场景判别器对滚动预测风光出力序列进行场景辨识,从策略集合中匹配最佳策略生成调度决策,构建“场景辨识—策略匹配—滚动调度”的动态响应机制。

1 水风光调峰调度模型

1.1 目标函数

风电、光伏出力受风速、太阳辐射等气象因子直接驱动,呈现显著的间歇性、波动性与随机性。梯级水电虽具快速响应、灵活调节能力,但其出力受流域来水情势、水库防洪兴利调度规则及机组工况等多重约束。在高比例可再生能源并网背景下,风光出力的强随机性易引发系统净负荷大幅波动,加剧调峰压力,威胁系统安全稳定运行。梯级水库群作为水风光互补系统核心调节单元,可通过精细化蓄泄调控与水风光出力协同优化,平抑风光出力波动,实现净负荷削峰填谷。基于此,本文以剩余负荷标准差最小为目标,依托梯级水库群蓄泄与水风光出力协同优化,平滑负荷过程、降低峰谷差与波动。目标函数公式为:

$$\min F = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(P_{r,t} - \frac{1}{T} \sum_{t=1}^T P_{r,t} \right)^2} \quad (1)$$

其中

$$P_{r,t} = P_{L,t} - P_{w,t} - P_{v,t} - \sum_{n=1}^N P_{h,nt}$$

$$P_{h,nt} = \kappa_n H_{nt} Q_{e,nt}$$

式中: F 为调度期内系统剩余负荷的标准差; T 为调度时段总数; t 为时段序号; N 为水库总数; n 为水库编号; $P_{L,t}$ 为 t 时段负荷; $P_{w,t}$ 、 $P_{v,t}$ 分别为 t 时段风电和光伏预测出力; $P_{r,t}$ 为 t 时段剩余负荷; κ_n 为水库 n 的出力系数; $P_{h,nt}$ 、 $Q_{e,nt}$ 、 H_{nt} 分别为水库 n 在时段 t 的水电出力、发电流量和发电水头。

1.2 约束条件

a. 水量平衡约束:

$$V_{n,t+1} = V_{nt} + (Q_{u,nt} - Q_{d,nt}) \Delta t \quad (2)$$

其中

$$Q_{u,nt} = \sum_{i \in \Omega_n} Q_{d,it} + Q_{1,nt}$$

$$Q_{d,nt} = Q_{e,nt} + Q_{s,nt}$$

式中: V_{nt} 、 $V_{n,t+1}$ 分别为水库 n 在时段 t 的初始库容和末库容; $Q_{u,nt}$ 、 $Q_{d,nt}$ 、 $Q_{1,nt}$ 、 $Q_{e,nt}$ 、 $Q_{s,nt}$ 分别为水库 n 在时段 t 的总入库流量、总下泄流量、区间入流、发电流量和弃水流量; Ω_n 为水库 n 的所有直接上游水库集合; Δt 为调度时段长度。

b. 坝前水位约束:

$$Z_{u,\min,n} \leq Z_{u,nt} \leq Z_{u,\max,n} \quad (3)$$

式中: $Z_{u,nt}$ 、 $Z_{u,\min,n}$ 、 $Z_{u,\max,n}$ 分别为水库 n 在时段 t 的时段初坝前水位及其最低、最高水位限制。

c. 发电流量约束:

$$Q_{e,\min,n} \leq Q_{e,nt} \leq Q_{e,\max,n} \quad (4)$$

式中: $Q_{e,\min,n}$ 、 $Q_{e,\max,n}$ 分别为水库 n 的最小、最大发电流量。

d. 下泄流量约束:

$$Q_{d,\min,n} \leq Q_{d,nt} \leq Q_{d,\max,n} \quad (5)$$

式中: $Q_{d,\min,n}$ 、 $Q_{d,\max,n}$ 分别为水库 n 的最小、最大总下泄流量。

e. 非线性曲线约束:

$$\begin{cases} Z_{u,nt} = f_n(V_{nt}) \\ Z_{d,nt} = g_n(Q_{d,nt}) \end{cases} \quad (6)$$

式中: f_n 为水库 n 的库容-坝前水位转换函数, $f_n(V_{nt})$ 为库容 V_{nt} 对应的坝前水位; g_n 为水库 n 的

下泄流量-坝下水位转换函数, $g_n(Q_{d,nt})$ 为下泄流量 $Q_{d,nt}$ 对应的坝下水位。

2 基于多策略深度强化学习的水风光调峰调度

2.1 总体框架

针对风光出力显著的季节性波动与强随机不确定性,以及单一调度策略难以适配多变风光出力场景导致调峰效果受限的问题,本文构建了融合场景判别与多场景分工况深度强化学习的水风光互补调峰调度方法,如图 1 所示。该方法以场景驱动的分层决策框架为基础,通过离线场景建模与在线动态匹配提升系统应对风光出力不确定性的能力。具体而言,首先基于历史风电与光伏出力数据,在季节尺度上对日尺度出力曲线形态特征进行聚类分析,识别出若干代表不同风光出力规律的典型场景及其出现概率。随后,基于统一的目标函数(最小化剩余负荷标准差)与运行约束(如水位、出力),并针对各典型场景独立构建仿真环境,训练对应的智能体,使其掌握该场景下水库群调度策略,从而形成场景专属的深度强化学习策略模型;在此基础上,设计场景判别模块,以当前季节信息和未来一日标准化风光出力预测序列为输入,快速在线辨识最可能的场景。在线运行阶段,判别模块依据最新风光出力预测结果确定主导场景,调用与之匹配的预训练场景专属智能体策略,生成各水库的出力设定与下泄流量指令;为确保指令满足工程实际约束,计算结果需经可行域投影处理,以保障方案在水位、流量及电站出力允许运

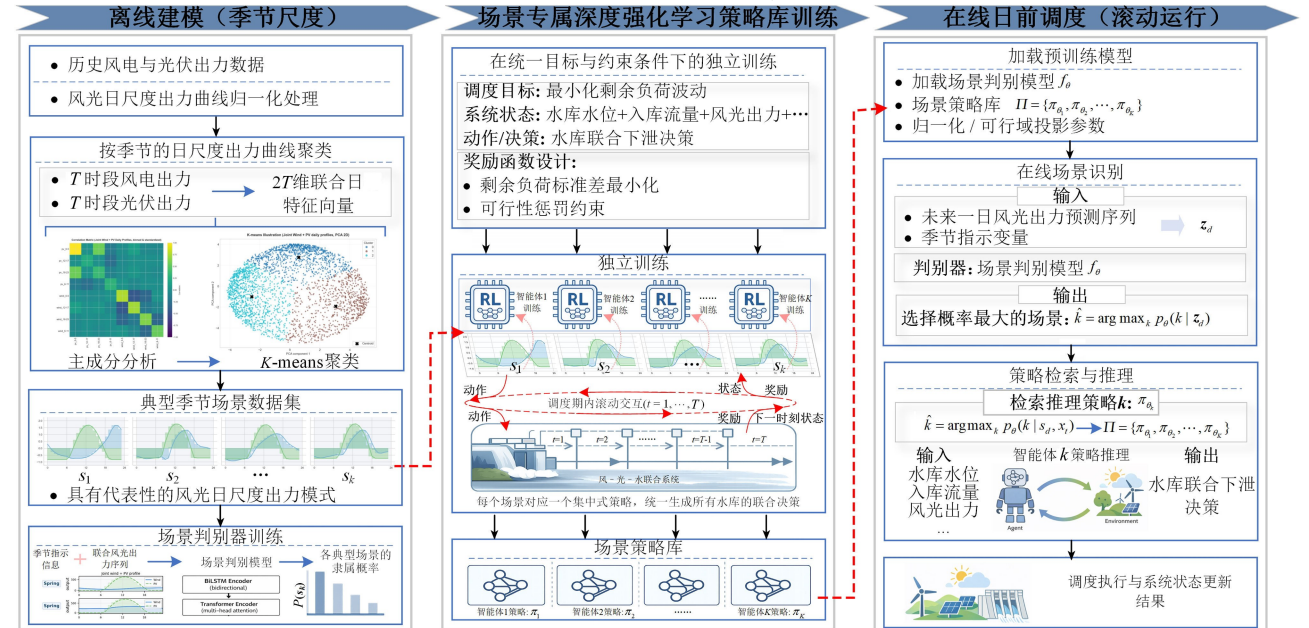


图 1 场景驱动的多策略水风光联合调度框架

Fig. 1 Scenario-driven multi-strategy framework for coordinated hydro-wind-solar dispatch

行范围内。该框架通过场景判别实现动态策略匹配,利用多个专属智能体分别优化不同场景下的调度策略。它结合了深度强化学习处理复杂时序关系的能力与显式场景划分的优势,提升了风光出力不确定条件下调峰决策的准确性、稳定性和抗干扰能力。

2.2 离线训练

2.2.1 训练场景划分

为刻画风光出力的不确定性与多模态特征,本文基于历史风电与光伏出力序列开展场景构建与划分。输入为按日组织的风光联合出力序列,对风电与光伏序列分别进行标准化处理以消除量纲差异,并拼接形成 $2T$ 维特征向量 $\mathbf{x}_d = (P_{w,d,0}, P_{w,d,1}, \dots, P_{w,d,T-1}, P_{v,d,0}, P_{v,d,1}, \dots, P_{v,d,T-1})$, 处理阶段先对 $\mathbf{x}_d$ 进行主成分分析(PCA)降维,并根据累计方差贡献率达到设定阈值(如 90%)自适应确定主成分数,将样本映射至低维空间;随后在低维空间采用 K -means 聚类,并通过 Calinski-Harabasz 指数确定最优聚类数 K 。输出为典型场景集合,对应的场景数据子集 $\{C_k\}_{k=1}^K$ (用于训练场景专属策略 π_{θ_k}),为场景适配策略库提供支撑。

2.2.2 强化学习训练要素设计

在模型求解层面,相较于传统数学规划方法在应对风光出力强随机性时易遭遇“维度灾难”与重复迭代导致的计算耗时问题,本文采用深度强化学习(DRL)实现高维非线性调度模型的近似求解。DRL 将复杂优化问题映射为有限时域调度区间的马尔可夫决策过程(Markov decision process, MDP):①通过奖励函数将调度目标统一转化为智能体可感知的反馈信号;②针对水位等复杂约束,设计动作可行域投影算子,在智能体输出动作进入物理环境前进行强制性截断与可行域映射,将约束满足过程前置,确保决策的物理可行性。由此,DRL 形成了“离线仿真训练+在线单次前向推理”的计算模式,将海量计算负担转移至离线阶段,在线阶段基于实时状态即可快速生成调度指令,实现毫秒级实时决策。上述 MDP 主要由状态、动作、转移与回报四要素构成,具体设计如下:

a. 状态向量。第 t 时段的状态向量记为 $\mathbf{s}_t$,由时间、库群坝前水位、入库流量、系统负荷、风电与光伏预测出力以及上一时段系统剩余负荷等信息构成,用于表征水风光联合系统的当前运行状态。为消除不同状态量之间的量纲和数值尺度差异,各状态量经归一化处理输入策略网络。

b. 动作向量。第 t 时段的动作向量记为 $\mathbf{a}_t$,由各水库发电流量以及弃水流量构成,用于表示离散时段智能体向各电站发出的调度指令,是智能体作

用于系统的核心输出。

c. 动作可行域。由发电流量、总下泄流量、生态流量、水电出力及出力爬坡等约束共同确定。策略网络输出的归一化动作经反归一化后,通过可行域投影修正为满足上述物理约束的调度指令。

d. 状态转移。系统状态由当前状态、调度动作和预测信息共同驱动。执行 $\mathbf{a}_t$ 后,下一时段的系统负荷、入库流量及风光出力由调度计划与预测结果确定;水库坝前水位和系统剩余负荷依据水量平衡及出力关系更新,形成下一时段的状态向量 $\mathbf{s}_{t+1}$ 。

e. 奖励函数。为将剩余负荷标准差最小化目标转化为可用于策略学习的时段反馈信号,本文第 t 时段的奖励函数 r_t 由调峰目标项和约束惩罚项组成。①调峰目标项。由各时段剩余负荷相对调度窗口平均值的偏离程度、相邻时段剩余负荷变幅评价调峰效果。首、末时段采用单侧差值,中间时段综合前、后相邻时段差值。两类代价经加权和归一化后取负值作为调峰奖励,以引导智能体降低剩余负荷波动和变化速率。②约束惩罚项。针对坝前水位越界、水电出力越界、生态下泄不足和期末水位偏离目标水位等情况设置惩罚。各类违反量经无量纲化处理,根据约束重要程度赋予相应权重和非线性惩罚系数;期末水位惩罚仅在调度窗口最后一个时段计算。③奖励合成。时段奖励 r_t 由调峰奖励扣除约束惩罚得到。剩余负荷越平稳且约束违反程度越小,最终奖励越高;约束违反程度增大时,惩罚值相应增加,从而引导智能体学习兼顾调峰效果与工程可行性的调度策略。

f. 策略映射。调度策略记为 π_{θ} ,其中 θ 为策略参数。该策略以当前状态 $\mathbf{s}_t$ 为输入,生成相应的动作 $\mathbf{a}_t$ 。在离线训练过程中,策略参数根据智能体与仿真环境交互获得的奖励反馈持续更新;在实际调用时,策略输出经动作转换和可行域投影后形成调度指令。

g. 状态动作轨迹。在长度为 T 的调度期内,策略与仿真环境逐时交互,依次记录各时段的系统状态、调度动作、反馈奖励和下一时段状态,形成完整的调度轨迹 τ 。该轨迹集中反映策略在整个调度期内的决策过程,并作为策略性能评价和参数更新的训练样本。

2.2.3 多场景策略库训练框架

2.2.3.1 多策略训练组织与智能体策略集合构建

基于典型风光联合出力场景划分结果,本文针对不同风光出力波动特征构建多策略强化学习框架。考虑不同运行场景下系统调节需求存在显著差异,为每类典型场景配置独立的强化学习智能体,通

过离线训练获得适用于对应状态的调度策略集合。具体而言,针对不同的典型风光出力场景,分别建立对应的仿真训练环境;而后以系统运行状态信息作为输入,以发电流量和弃水流量作为动作输出,通过奖励反馈持续优化策略参数,使智能体逐步学习兼顾复杂约束和调峰需求的调度规律。

在实际运行阶段,根据当前及未来时段风光出力特征辨识所属运行场景,并调用对应场景下训练完成的智能体策略生成调度方案。随后结合工程约束条件对策略输出进行可行性修正,形成满足实际运行要求的水风光联合调度指令,实现多变风光出力场景下的自适应优化调控。

2.2.3.2 基准对比算法

采用 OpenAI 于 2017 年提出的 PPO 算法作为基准对比算法,旨在缓解深度强化学习算法易出现的训练失稳、样本利用效率偏低等问题。PPO 基于策略梯度理论构建 Actor-Critic 框架^[26],Actor 网络输出动作策略,Critic 网络评估状态价值函数、估算期望累积回报并构造优势函数,以降低策略更新方差。PPO 与本文方法均属于策略优化类强化学习算法,可在相同的状态向量、动作空间、奖励函数和工程约束下进行对比。

2.2.3.3 A-GRPO 算法

针对水风光互补调度状态演化快、约束复杂及奖励稀疏的特性,传统 PPO 算法因依赖显式价值网络(Critic),易引发训练失稳、样本利用效率偏低问题。为克服该缺陷,本文基于 PPO-Clip 框架,提出一种轻量化的 A-GRPO 算法,并将其作为策略库的核心算法^[27]。训练过程中,智能体通过与仿真环境交互,在固定长度的滚动时间窗口内采集调度轨迹,并依据各时段奖励计算蒙特卡罗回报。随后,对同一批次内的回报进行标准化处理,以其相对差异评估各调度动作的优劣,从而在不设置价值网络的条件下形成策略更新所需的优势信息,降低不同场景和调度时段之间奖励尺度差异的影响。

策略更新以 PPO 的裁剪代理目标为基础,并引入参考策略、KL 散度约束和熵正则项。参考策略用于约束当前策略偏离更新前策略的程度,KL 散度约束用于限制当前策略与参考策略的分布差异,熵正则项用于维持策略探索能力。同时,根据训练过程中监测到的近似 KL 散度动态调节学习率、裁剪阈值和熵系数,以减小策略突变并提高训练收敛的稳定性。

2.3 在线决策

2.3.1 场景判别器构建

为实现水风光互补系统在线工况辨识与智能体

调度策略集合匹配,本文基于风光聚类结果构建场景判别器。首先,将聚类得到的场景标签作为监督信号,构造训练样本。第 d 天特征向量为 $\mathbf{z}_d = (s_d, P_{w,d,0}, \dots, P_{w,d,23}, P_{v,d,0}, \dots, P_{v,d,23})$,其中季节信息 s_d 采用 0~1 指示编码并入特征向量, $P_{w,d,0}, \dots, P_{w,d,23}, P_{v,d,0}, \dots, P_{v,d,23}$ 为该日 T 个时段风电与光伏的标准化出力序列,场景序号 $k_d \in \{1, 2, \dots, K\}$ 为分类标签。判别器采用 BiLSTM+Transformer 结构 $f_\theta(\cdot)$ ^[32]:BiLSTM 对联合风光出力序列进行双向编码以提取时序依赖,Transformer 通过多头自注意力捕捉长程相关与风光耦合特征,实现季节特征与全局时序表征的融合。模型经全连接层与 softmax 输出场景条件概率 $p_\theta(k | \mathbf{z}_d)$,并采用交叉熵损失进行离线训练。在线调度阶段,将滚动预测得到的风光日出力序列及当前季节信息输入判别器,获得场景标签 $\hat{k} = \operatorname{argmax}_k p_\theta(k | \mathbf{z}_d)$,并据此调用策略库中对应智能体执行调峰决策,实现“场景辨识-策略匹配-水风光联合调度”的动态联动。

2.3.2 在线调度流程与策略匹配机制

在线调度阶段,以离线构建的场景判别器与场景专属策略库 $\Pi = \{\pi_{\theta_1}, \pi_{\theta_2}, \dots, \pi_{\theta_K}\}$ 为基础,通过滚动预测,形成面向多场景自适应响应的机制,具体流程如下:

a. 离线模型初始化。在线调度启动时,载入场景判别器参数与策略库中各场景策略网络的参数 $\{\theta_1, \theta_2, \dots, \theta_K\}$;同步读取训练阶段一致的状态标准化参数、动作边界约束及可行域投影参数,确保在线推理与离线训练的数据分布与约束环境一致。

b. 状态感知与输入表征。在单个调度周期内,采集系统状态 $\mathbf{s}_t$ 并进行归一化处理。同时,结合当期季节特征与未来一日风光出力预测序列,构造场景判别器的输入向量 $\mathbf{z}_d$ 。

c. 场景辨识与策略匹配。将预处理后的风光出力预测序列及季节特征输入场景判别器,确定当前时段对应的场景标签,记为 $\hat{k} = \operatorname{argmax}_k p_\theta(k | \mathbf{z}_d)$;以场景标签 $\hat{k}$ 为索引,在策略库 Π 中匹配对应策略智能体 π_{θ_k} ,并同步加载该场景关联的标准化与投影参数。

d. 策略推理与动作生成。将归一化处理后的系统状态 $\mathbf{s}_t$ 输入已匹配的场景策略网络,输出相应的归一化动作。

e. 动作映射与可行校正。对策略网络输出的归一化动作进行反归一化和可行域投影,得到可行调度动作 $\mathbf{a}_t$,确保方案满足发电流量边界、生态下泄要求、水电出力限制及出力爬坡限制等工程约束。

f. 环境交互与滚动迭代。将物理可执行调度

量 a_i 施加于调度环境,驱动状态更新并进入下一时刻;伴随滚动预测的周期性更新,重复执行步骤 b~e,实现在线调度的连续动态响应。

g. 增量学习与参数更新。基于在线运行积累的状态-动作-回报数据,定期对策略网络进行微调;当风光出力特性出现显著变化时,利用新数据对场景判别器再训练,增强场景辨识与策略匹配的适应性。

3 工程算例验证

3.1 工程背景

黄河上游龙羊峡、拉西瓦、李家峡梯级水库群具备跨区域水资源优化配置、清洁能源供给与区域生态平衡维护等多重功能,是保障西北干旱半干旱区供水安全、支撑清洁能源跨区域外送及维系区域生态稳定的重要基石。本文在上述梯级水库群基础上,接入 400 MW 风电场及龙羊峡 850 MW 光伏电站,构建水风光互补系统。该系统既通过风光出力时空互补特性提升可再生能源供电可靠性,又依托梯级水库群灵活调节能力,为电网提供持续、稳定、优质的清洁调峰电源。

3.2 风光场景生成

风光出力数据取自公开平台 renewables. ninja,

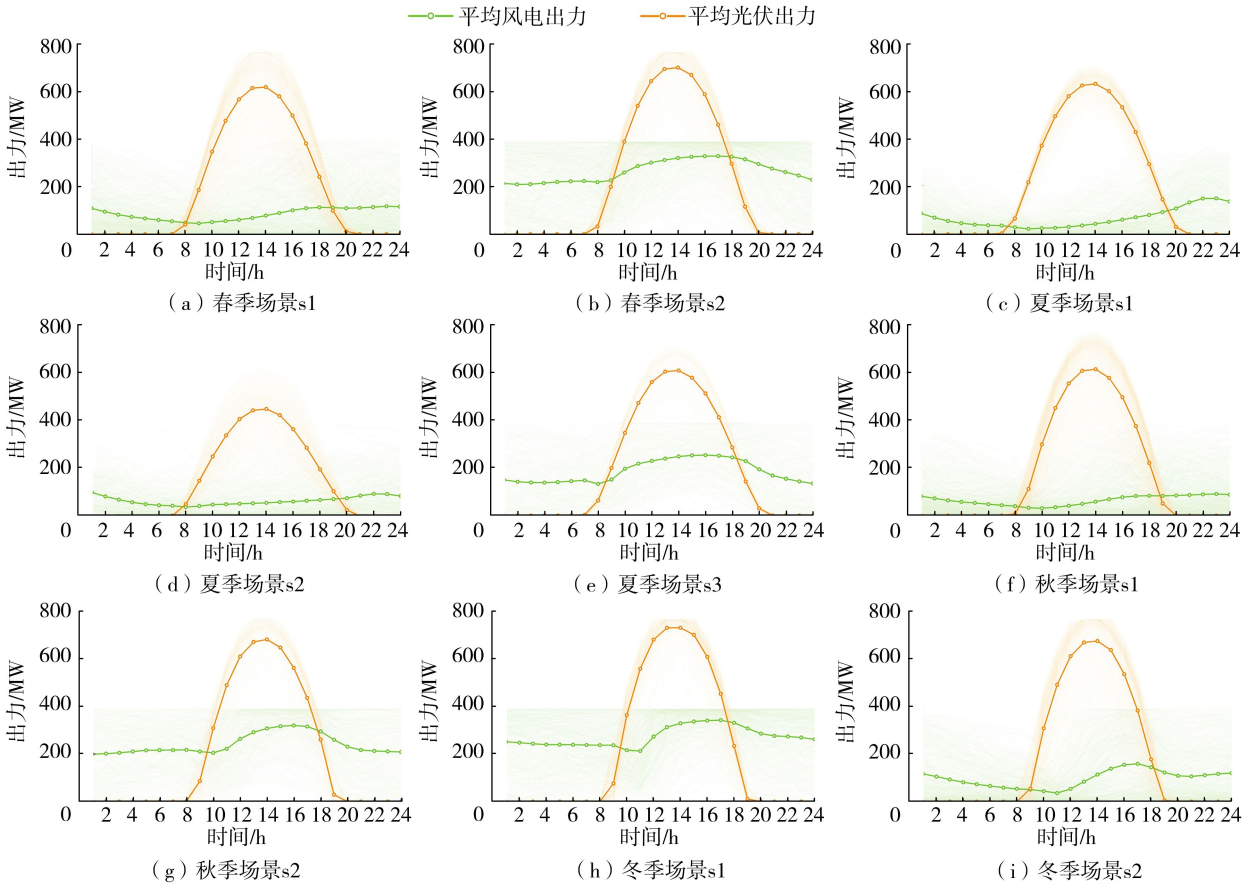


图2 风-光联合出力划分场景集

Fig. 2 Partitioned scenario set for joint wind-solar power output

本文提取 2016—2023 年逐小时风光出力序列,将其作为场景生成的基础数据源。将风电与光伏逐时出力在日尺度上重构为 48 维联合特征向量,按季节划分为春、夏、秋、冬 4 类样本。对各季节样本分别进行 PCA 降维与 K-means 聚类。聚类结果如图 2 所示,春季、秋季和冬季分别划分为 2 类场景,夏季划分为 3 类场景,共形成 9 类季节场景。由图 2 可见,不同季节典型场景集的风光出力特征差异显著。其中,在春季,场景 s1 风电、光伏的日均出力分别为 85 MW 与 195 MW,场景 s2 则分别增大至 266 MW 与 223 MW;在秋季,场景 s1 风电出力全天处于低位且波动平缓、光伏出力呈单峰分布,场景 s2 风电出力显著抬升,且午间光伏出力峰值进一步提升,光伏出力与风电出力同向隆起。通过上述处理,将高维日出力序列映射至低维空间并完成聚类,形成典型场景样本子集,为后续模型训练提供统一的场景划分标准,确保策略库覆盖多季节、多模态的风光出力特征。

3.3 基于 BiLSTM 的场景识别

基于场景划分结果,构建 BiLSTM-Transformer 混合模型作为场景判别器。实验采用 Adam 优化算法(学习率设为 2×10^{-4} ,权重衰减系数设为 1×10^{-4})。模型训练 60 轮,批处理大小设为 32,通过分

层随机抽样划分 15% 样本作为验证集。网络结构由单层双向 LSTM (隐层单元数 64) 与双层 Transformer 编码器 (模型维度 128, 多头注意力头数 4) 构成。图 3(a) 显示训练收敛平稳, 第 43 轮达到最优验证精度 98.27%。图 3(b) 混淆矩阵定量展示了验证集分类性能, 总体平均识别准确率达 97.32%, 且跨季节场景误判率普遍趋近于零。图 3(c) 显示各类别精确率与召回率普遍高于 95%。综合结果表明, 所构建的判别器在风光联合出力场景分类任务中能够稳定辨识风光出力典型场景。

3.4 超参数敏感性分析

为降低在全场景开展大规模超参数网格搜索的计算开销, 根据图 2 典型场景, 在各季节选取一条代表性风光联合出力序列构建训练环境, 对 A-GRPO 与 PPO 关键超参数进行敏感性分析。考虑折扣因子 γ 、学习率 α_π 与网络规模 N_h 对训练稳定性及调峰性能影响显著, 本文围绕这三类参数设计网格组合; 对每组参数在各季节代表环境中迭代训练 1500 次, 以剩余负荷标准差 (standard deviation, STD) 的最小值作为评价指标, 并绘制参数敏感性热力图 (图 4)。结果表明, 两类算法均对上述参数表现出明显敏感性, 但敏感模式存在差异。以夏季序列为例, 异常值 (STD 大于 150 MW) 对应的组合不具备稳定性; 在有效区间内, A-GRPO 的 STD 主要分布于 75.9~104.6 MW, 最优为 75.9 MW ($\gamma=0.9$, $\alpha_\pi=0.0002$, $N_h=640$); PPO 分布于 76.4~121.8 MW, 最优为 76.4 MW, 但在学习率与网络规模组合不当时更易出现极端值。综合各季节代表序列, A-GRPO 在风光出力波动更显著的场景中普遍获得更小的剩余负荷标准差, 体现出更好的调峰适应性。

3.5 训练收敛过程

为验证多策略强化学习模型的训练有效性, 本文选取不同季节的典型风光出力场景, 分别对 A-

GRPO 与 PPO 算法开展离线训练, 并比较训练过程中累计奖励的变化及策略稳定性。结果表明, 2 种算法均能通过与水风光联合调度环境持续交互, 逐步学习调峰策略, 训练后期总体趋于稳定, 说明所设置的状态向量、动作空间、奖励函数和约束处理方式能够支持策略学习。A-GRPO 在多数典型场景中的收敛过程更为平稳, 训练后期波动相对较小, 对不同风光出力特征表现出较好的适应性。该算法利用批次回报构建相对优势, 并对策略更新幅度加以约束, 有助于减弱回报尺度差异和策略突变对训练过程的影响。训练完成后, 保留各典型场景下表现较优的策略, 构建场景专属策略库, 并为在线阶段的场景辨识、策略匹配与调度决策提供基础。

3.6 不同典型场景日调峰效果对比

表 1 列出了春、夏、秋、冬各 2 个典型场景日的调峰效果与计算时间对比结果。结果表明, 相较于原始负荷, 各方法均能显著降低剩余负荷波动并提高负荷率。其中, CSA 算法调峰质量最优; 其在春、秋、冬三季的剩余负荷标准差均不高于 0.43 MW, 系统负荷率普遍不低于 99.85%, 峰谷差均控制在 1.47 MW 以内; 但其计算时间维持在 18~20.5 s, 计算成本相对较高。传统 POA 算法则存在明显不足: 调峰效果波动幅度较大, 夏季和秋季剩余负荷标准差 (最高达 205.18 MW)、峰谷差 (最高达 626.20 MW) 显著高于其他方法; 冬季负荷率最低仅 92.76%, 亦偏低。同时, POA 算法计算时间显著长于其他方法 (最高约 183 s), 制约其在实时场景中的应用。相比之下, A-GRPO 与 PPO 算法在计算效率上展现出显著优势, 计算时间均不超过 0.06 s, 可实现毫秒级决策。二者调峰效果亦较为显著, 将系统负荷率从原始水平的 69%~79% 提升至 96% 以上, 有效平抑波动并减小峰谷差。总体而言, CSA 算法调峰质量最优, 但计算成本较高; A-GRPO 与 PPO 算法在维持

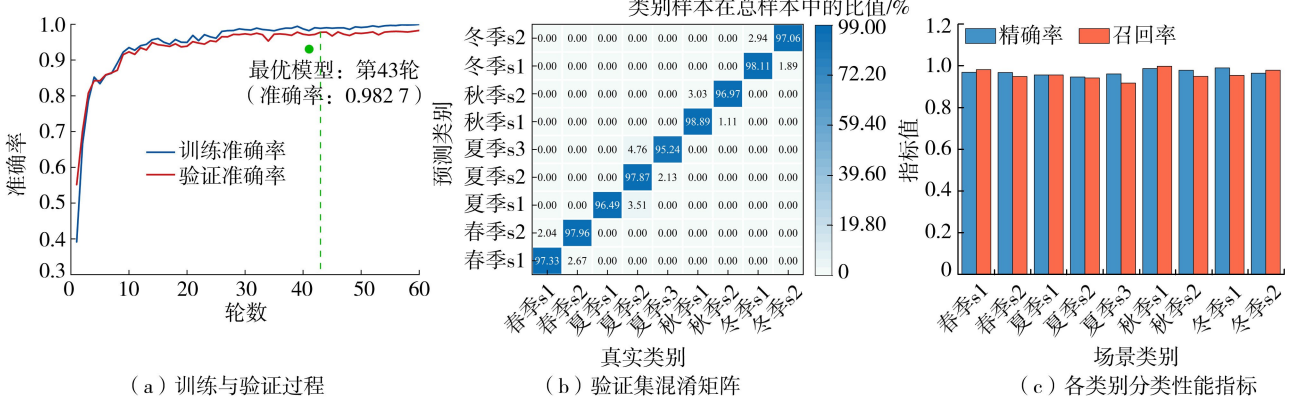
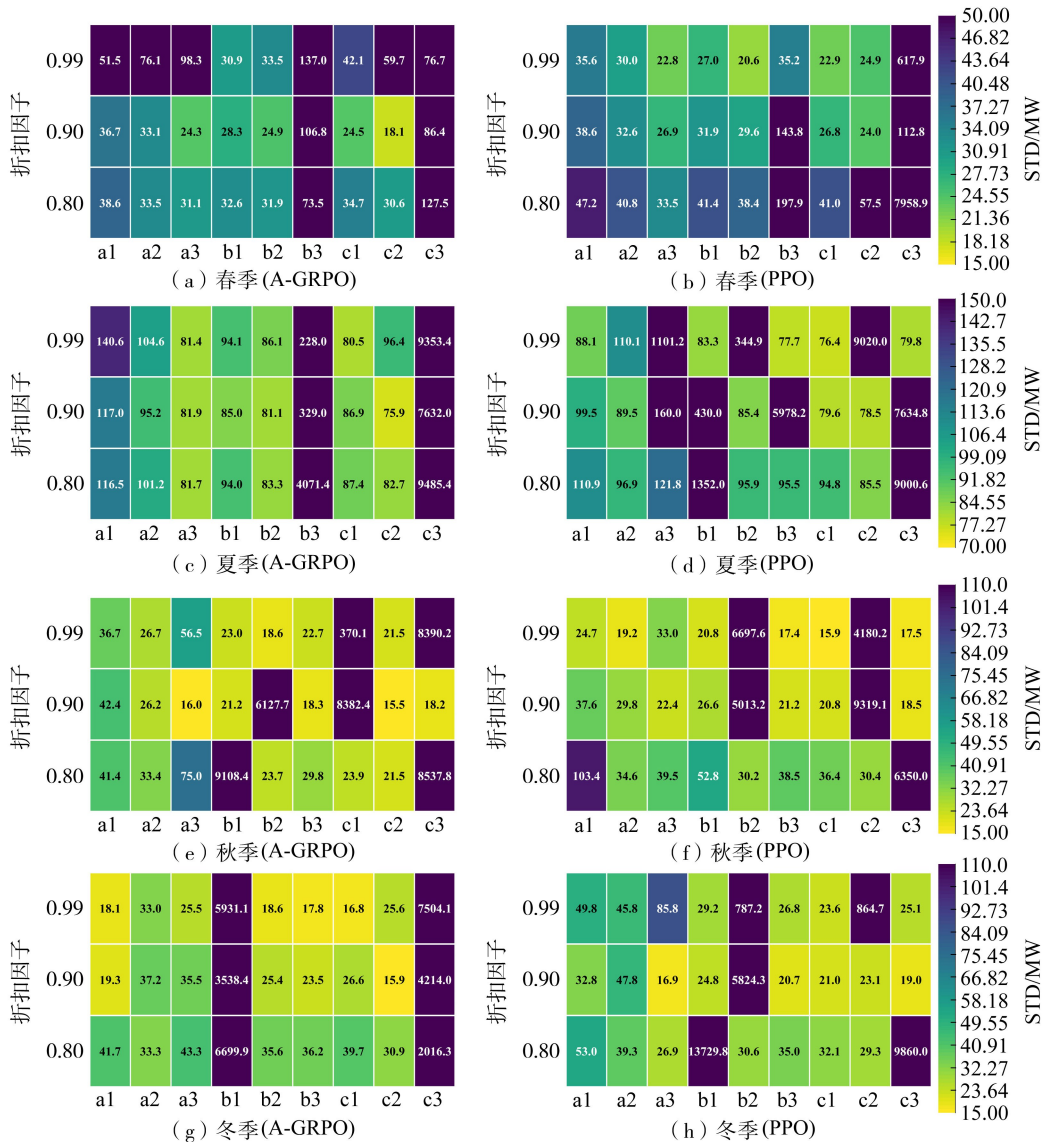


图 3 场景识别性能结果
Fig. 3 Scenario identification performance results



注:横轴表示参数组合。其中,a、b、c 分别表示每层隐藏单元数为 64、320 和 640;1、2、3 分别表示学习率为 0.000 1、0.000 2 和 0.002。字母与数字的组合表示相应的参数配置,如 b2 表示每层隐藏单元数为 320、学习率为 0.000 2。STD 为剩余负荷标准差。

图 4 参数敏感性分析结果
Fig. 4 Parameter sensitivity analysis results

毫秒级决策时效的同时,实现显著波动抑制与负荷率提升,更适宜作为在线滚动调峰调度的核心策略。综合跨季节、跨场景测试结果,A-GRPO 算法相较 PPO 算法在调峰效果(更小剩余负荷标准差、更低峰谷差、更高负荷率)上稳健性更优。

春季天然来水匮乏,水库维持高水头运行(保障运行经济性)与频繁削峰填谷(提供调节灵活性)之间的矛盾更为凸显。因此,选取春季典型场景 s1,对水库水位与水风光联合出力过程进行深入剖析。图 5 为龙羊峡、拉西瓦与李家峡三水库的日内水位变化过程,图 6 对比了不同算法下风光—梯级水电联合出力过程。分析结果显示,A-GRPO 与 PPO 的水位曲线平滑连续,呈缓慢消落与动态回升,表明其倾向于采用连续调节模式实现削峰填谷;相比之下,POA 水位曲线呈显著锯齿状波动,CSA 介于两者之

间,表明其易诱发频繁放水和蓄水切换,增大日内调节幅度。出力过程图进一步验证上述差异:A-GRPO 与 PPO 的梯级出力能紧密跟踪剩余负荷变化,且出力均衡平滑,有效抑制了波动。而 POA 与 CSA 的出力则在部分时段出现局部尖峰与陡变,对应水位快速涨落,表明其日内调节波动更明显。综上,A-GRPO 和 PPO 较 POA 算法具有更优的剩余负荷跟踪精度与更稳定的协同调度特性。

3.7 多场景条件下的自适应调峰调度效果分析

为检验所提方法的适应性,选取 2024 年风光出力数据作为独立测试集开展验证。测试采用基于 2016—2023 年逐小时出力序列训练得到的 BiLSTM-Transformer 场景判别器,对 2024 年逐日滚动预测的风光联合出力序列进行工况辨识,依据辨识标签从

表 1 典型场景下不同算法调峰性能与计算时间对比

Table 1 Comparison of peak-shaving performance and computation time among different algorithms under typical scenarios

指标	算法	春季 s1	春季 s2	夏季 s1	夏季 s2	秋季 s1	秋季 s2	冬季 s1	冬季 s2
剩余负荷 标准差/MW	CSA	0.13	0.26	40.05	86.37	0.38	0.43	0.36	0.28
	A-GRPO	3.30	3.72	55.00	94.28	2.26	2.41	2.53	2.28
	POA	11.73	8.42	172.11	205.18	102.11	95.97	8.55	8.40
	PPO	4.35	7.90	55.03	94.41	5.33	4.05	6.86	8.16
	原始	761.58	761.58	1292.95	1292.95	1001.59	1001.59	775.53	775.53
峰谷差/MW	CSA	0.52	1.05	176.37	289.49	1.32	1.47	1.27	1.12
	A-GRPO	11.46	17.02	190.80	325.24	8.68	9.84	10.87	9.28
	POA	35.48	38.58	535.58	626.20	244.63	459.51	39.05	36.75
	PPO	19.83	27.15	198.08	293.64	22.59	14.05	22.92	23.66
	原始	2200.00	2200.00	3700.00	3700.00	3500.00	3500.00	2000.00	2000.00
负荷率/%	CSA	99.99	99.98	98.52	98.05	99.98	99.98	99.85	99.93
	A-GRPO	99.79	99.69	98.16	97.44	99.84	99.89	99.19	99.53
	POA	99.39	99.22	97.71	96.99	98.38	98.61	92.76	96.51
	PPO	99.68	99.36	97.88	97.57	99.67	99.78	96.55	96.99
	原始	78.46	78.46	77.73	77.73	69.80	69.80	75.93	75.93
计算时间/s	CSA	19.528	19.447	20.499	19.286	18.550	19.136	20.419	19.746
	A-GRPO	0.053	0.054	0.053	0.015	0.017	0.017	0.016	0.017
	POA	176.100	182.425	144.762	131.028	182.823	163.541	85.968	113.213
	PPO	0.038	0.033	0.020	0.022	0.016	0.016	0.016	0.015

注:受表格横向版面限制,本表仅列出各季节 s1、s2 场景的结果。

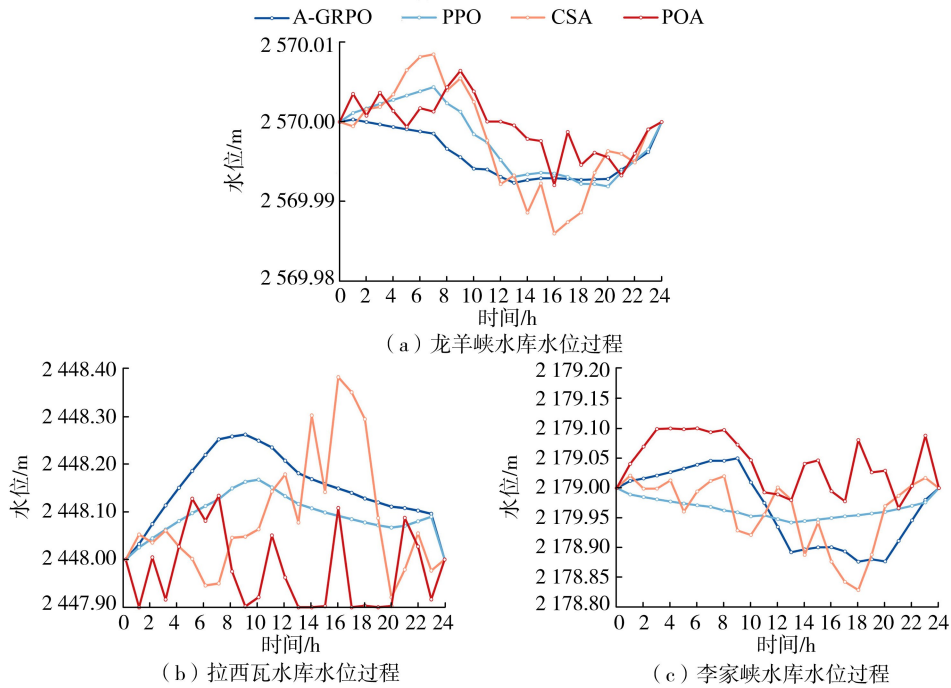


图 5 春季场景 s1 水库水位过程

Fig. 5 Reservoir water level hydrograph for Spring scenario s1

离线构建的策略库中检索并调用匹配的 A-GRPO 与 PPO 场景策略,统计剩余负荷标准差、负荷率、峰谷差及单步计算时延等指标,全过程满足工程边界条件。图 7 表明,场景判别器对 2024 年风光联合出力序列的工况辨识效果良好:全年 366 个日样本中有 357 个判别正确,总体准确率为 97.54%;按季节统计,春、夏、秋、冬场景准确率分别为 97.8%、95.7%、97.8%、98.9%,表明模型在各季节表现均

衡,误判次数少且分散,整体泛化能力较好。图 8~10 展示了 A-GRPO 与 PPO 的调度性能对比,表 2 给出了 2024 年独立测试中不同场景下剩余负荷指标的平均值及相应计算时间,与表 1 的典型场景日试验结果分别统计。总体上,A-GRPO 在所有场景组均优于 PPO,其剩余负荷标准差更小、负荷率更高、峰谷差更低,且保持毫秒级决策。以春季场景 s1 为例,A-GRPO 将剩余负荷标准差由 PPO

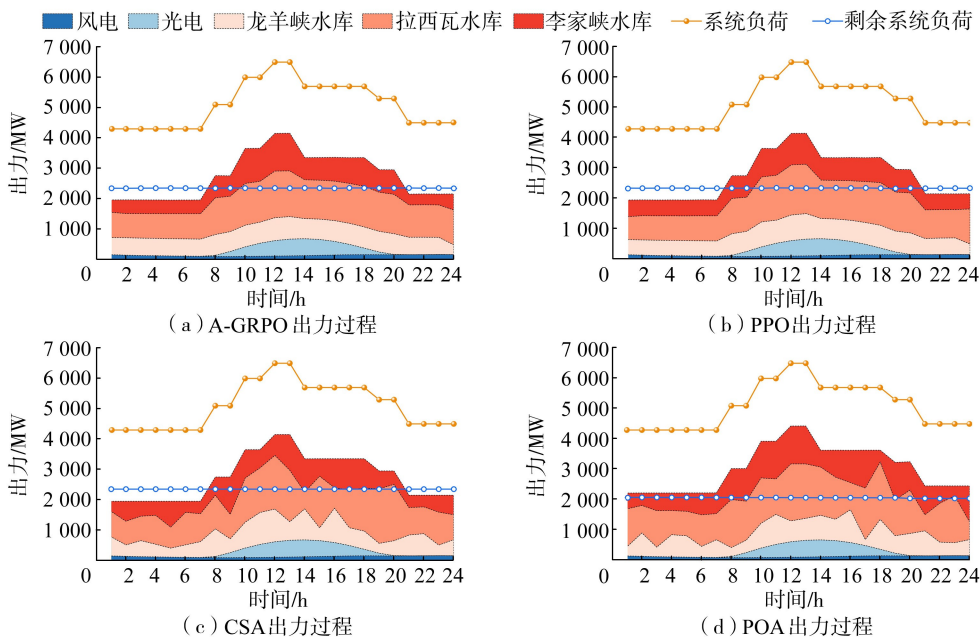


图 6 春季场景 s1 风光与梯级水电联合出力过程

Fig. 6 Joint output profiles of wind, solar and cascade hydropower for spring scenario s1

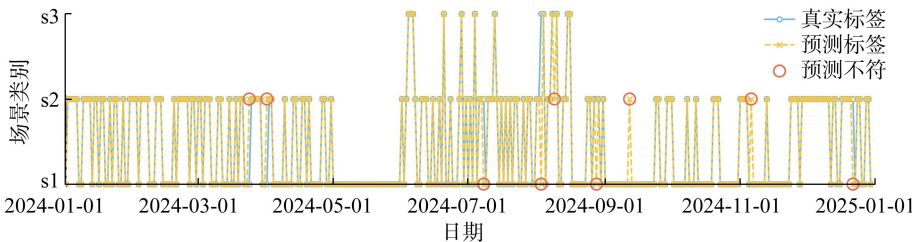


图 7 场景判别器测试效果

Fig. 7 Test performance of the scenario discriminator

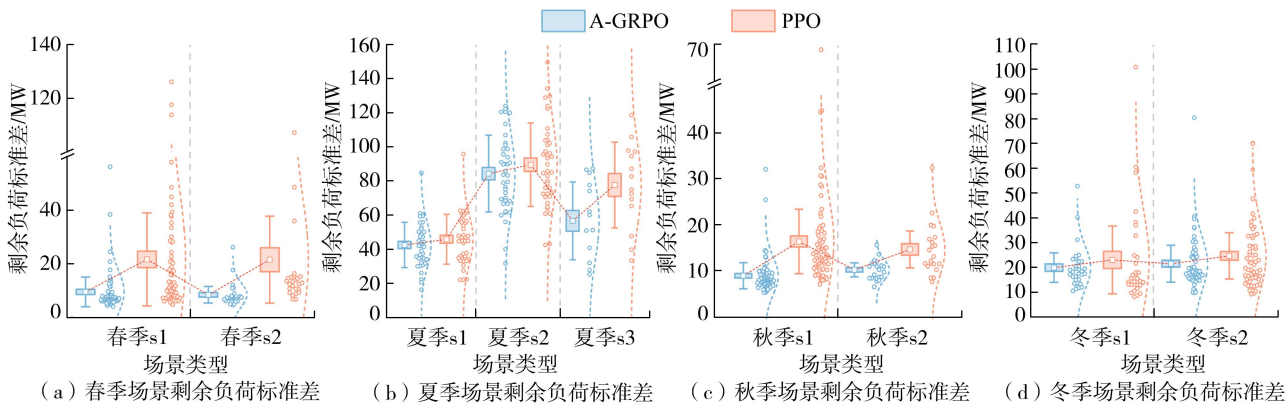


图 8 A-GRPO 与 PPO 剩余负荷标准差对比

Fig. 8 Comparison of residual load standard deviation between A-GRPO and PPO

的 21.69 MW 降至 9.54 MW,峰谷差由 91.47 MW 降至 36.89 MW,负荷率由 97.34% 提升至 98.77%;在风光出力波动更强的夏季场景 s3 中,相比 PPO,A-GRPO 将剩余负荷标准差由 77.71 MW 降至 56.66 MW,峰谷差由 319.97 MW 降至 215.02 MW,系统负荷率由 93.83% 提升至 96.35%。实时性方面,在 2024 年独立测试中,表 2 所列两种强化学习

策略的计算时间为 24.15 ~ 46.15 ms,均处于毫秒级,满足在线滚动调度需求;A-GRPO 与 PPO 时延差异小,但 A-GRPO 在调峰质量指标上表现更稳健。综上,本文方法在不确定风光出力条件下具备良好鲁棒性,其中 A-GRPO 相较 PPO 更适合作为调峰策略,为高比例可再生能源并网下的梯级水库动态调度提供技术支撑。

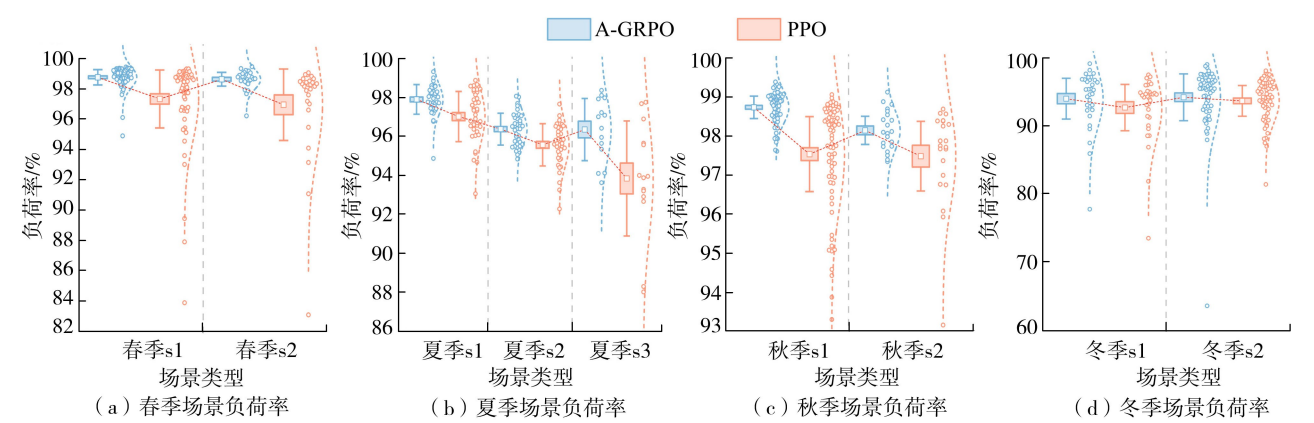


图 9 A-GRPO 与 PPO 负荷率对比
Fig. 9 Comparison of load rate between A-GRPO and PPO

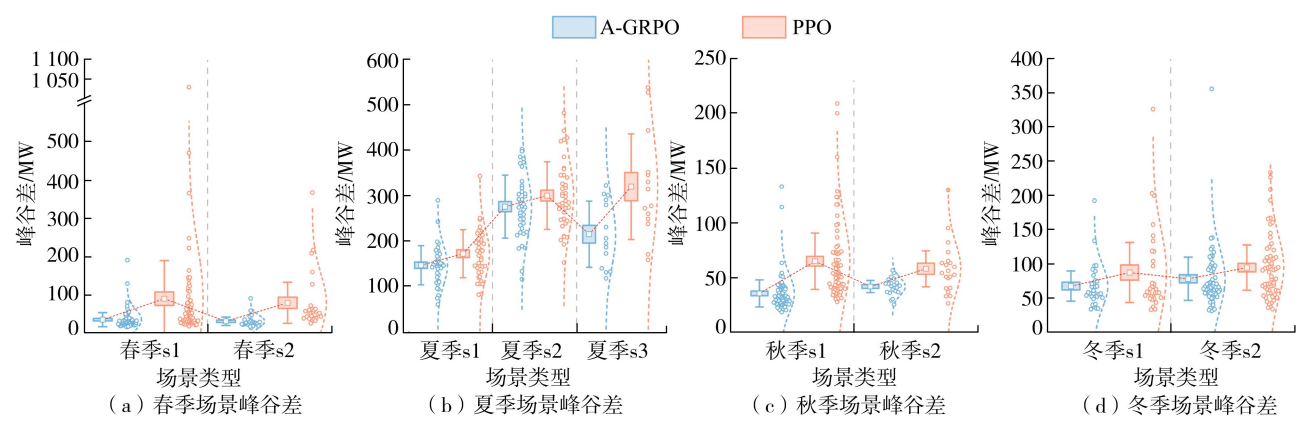


图 10 A-GRPO 与 PPO 峰谷差对比
Fig. 10 Comparison of peak-valley difference between A-GRPO and PPO

表 2 平均剩余负荷指标

季节	场景类别	算法	剩余负荷 标准差/MW	负荷 率/%	峰谷 差/MW	计算时 间/ms
春季	s1	A-GRPO	9.54	98.77	36.89	26.43
		PPO	21.69	97.34	91.47	28.29
	s2	A-GRPO	8.45	98.64	32.66	34.46
		PPO	21.58	96.96	80.65	41.12
夏季	s1	A-GRPO	42.51	97.90	146.60	30.55
		PPO	45.86	97.02	172.09	33.66
	s2	A-GRPO	84.38	96.37	275.91	32.24
		PPO	89.61	95.56	300.27	31.41
	s3	A-GRPO	56.66	96.35	215.02	36.24
		PPO	77.71	93.83	319.97	31.50
秋季	s1	A-GRPO	8.92	98.74	36.00	24.84
		PPO	16.37	97.55	65.30	24.15
	s2	A-GRPO	10.20	98.15	42.29	34.49
		PPO	14.65	97.49	58.42	46.15
冬季	s1	A-GRPO	19.92	94.02	67.40	34.57
		PPO	23.02	92.72	87.27	30.45
	s2	A-GRPO	21.51	94.23	78.09	24.75
		PPO	24.67	93.70	94.48	29.23

4 结 语

针对风光出力随机波动导致单一调度策略跨场景适应性不足,以及梯级水库群多重工程约束增加

在线决策难度的问题,本文提出了融合场景辨识与多策略深度强化学习的水风光联合调峰方法。该方法采用 PCA 降维与 K-means 聚类划分典型风光出力场景,利用 BiLSTM-Transformer 模型完成在线场景辨识,并针对各典型场景分别训练 A-GRPO 调度策略,形成“场景辨识-策略匹配-滚动调度”的在线决策机制。实例验证结果表明,所提方法能够有效辨识不同风光出力场景,满足水位、流量、出力及爬坡等工程约束,显著降低系统剩余负荷波动,改善削峰填谷效果;同时具有较好的训练稳定性和在线决策效率,增强了对多变风光出力场景的适应能力,为水风光协同调峰提供技术参考。

本文研究尚存在以下局限:一是场景库覆盖范围对低概率极端风光出力场景的泛化能力有待验证;二是多策略模型协同架构对计算资源需求较高。后续研究拟从以下方向深化:引入在线增量学习机制,实现策略库动态进化,以适应长期气候变化趋势;构建日内-周-旬多时间尺度协同优化框架,定量评估方法的适用性与鲁棒性边界,为高比例可再生能源并网调峰的工程实践提供理论支撑。

参考文献:

[1] 申建建,王月,程春田,等.水风光多能互补发电调度问题研究现状及展望[J].中国电机工程学报,2022,42(11):3871-3885. (Shen Jianjian, Wang Yue, Cheng Chuntian, et al. Research status and prospect of generation scheduling for hydropower-wind-solar energy complementary system [J]. Proceedings of the CSEE, 2022,42(11):3871-3885. (in Chinese))

[2] 张俊涛,程春田,于申,等.水电支撑新型电力系统灵活性研究进展、挑战与展望[J].中国电机工程学报,2024,44(10):3862-3885. (Zhang Juntao, Cheng Chuntian, Yu Shen, et al. Progress, challenges and prospects of research on hydropower supporting the flexibility of new power systems [J]. Proceedings of the CSEE, 2024,44(10):3862-3885. (in Chinese))

[3] 尹丽君,许昌,黄显峰,等.水光互补基地水库调度图绘制及应用策略[J].河海大学学报(自然科学版),2023,51(5):65-69. (Yin Lijun, Xu Chang, Huang Xianfeng, et al. Drawing and application strategy of dispatching diagram for water-light complementary base reservoir [J]. Journal of Hohai University (Natural Sciences), 2023,51(5):65-69. (in Chinese))

[4] 王立平,马实一,郭旭,等.风电-光伏发电-抽水蓄能联合优化调度方法[J].水利经济,2024,42(6):58-63. (Wang Liping, Ma Shiyi, Guo Xu, et al. Wind power-photovoltaic power generation-pumped storage joint optimal scheduling method [J]. Journal of Economics of Water Resources, 2024,42(6):58-63. (in Chinese))

[5] 黄显峰,黄晗,鲜于虎成,等.源-荷匹配机制下的水光互补系统中长期与短期嵌套优化调度模型[J].水利水电科技进展,2025,45(2):38-45. (Huang Xianfeng, Huang Han, Xianyu Hucheng, et al. Mid-long term and short term nested optimal scheduling model of hydro-photovoltaic complementary system considering source-load matching mechanism [J]. Advances in Science and Technology of Water Resources, 2025,45(2):38-45. (in Chinese))

[6] 冯仲恺,蒋林佚,牛文静,等.黄河流域九省区水资源-能源耦合系统研究(Ⅱ):演化规律[J].水资源保护,2025,41(6):52-65. (Feng Zhongkai, Jiang Linyi, Niu Wenjing, et al. Water-energy coupling system in nine provinces and autonomous regions of the Yellow River Basin (Ⅱ): evolution law [J]. Water Resources Protection, 2025,41(6):52-65. (in Chinese))

[7] 李继清,龙健,刘洋.基于水光电力系统时序生产模拟模型的水光优化配比研究[J].河海大学学报(自然科学版),2024,52(1):1-10. (Li Jiqing, Long Jian, Liu Yang. Study on optimal ratio of water and light based on time-series production simulation model of hydro-optical power system [J]. Journal of Hohai University (Natural Sciences), 2024,52(1):1-10. (in Chinese))

[8] 张验科,郭合斌,卢垚键,等.水风光多能互补系统多目标优化调度研究进展[J].水利水电快报,2025,46(10):101-109. (Zhang Yanke, Guo Hebin, Lu Yaojian, et al. Research progress on multi-objective optimal operation of hydro-wind-solar complementary system [J]. Express Water Resources & Hydropower Information, 2025, 46(10):101-109. (in Chinese))

[9] 刘欣雨,罗彬,陈永灿,等.考虑风光联合出力不确定性的混蓄-风光短期调度模型[J].水力发电学报,2025,44(5):44-60. (Liu Xinyu, Luo Bin, Chen Yongcan, et al. Short-term dispatch model for hybrid pumped storage-wind-solar power considering uncertainty of combined wind and solar outputs [J]. Journal of Hydroelectric Engineering, 2025,44(5):44-60. (in Chinese))

[10] 刘方,刘建行,唐成鹏,等.梯级“水蓄风光”互补系统增益分配调度策略[J].水利经济,2024,42(6):64-70. (Liu Fang, Liu Jianhang, Tang Chengpeng, et al. Gain allocation and dispatch strategy of cascade hydropower-pumping-storage-wind-photovoltaic multi-energy complementary system [J]. Journal of Economics of Water Resources, 2024,42(6):64-70. (in Chinese))

[11] 张验科,罗笙月,卢垚键,等.风-光-混蓄长周期互补发电优化调度研究[J].水利水电科技进展,2025,45(4):60-66. (Zhang Yanke, Luo Shengyue, Lu Yaojian, et al. Research on optimal operation for long-term complementary power generation of wind-photovoltaic-hybrid storage system [J]. Advances in Science and Technology of Water Resources, 2025,45(4):60-66. (in Chinese))

[12] 冯仲恺,蒋林佚,牛文静,等.黄河流域九省区水资源-能源耦合系统研究(Ⅲ):风险分析[J].水资源保护,2026,42(2):1-11. (Feng Zhongkai, Jiang Linyi, Niu Wenjing, et al. Water-energy coupling system in nine provinces and autonomous regions of the Yellow River Basin (Ⅲ): risk analysis [J]. Water Resources Protection, 2026,42(2):1-11. (in Chinese))

[13] 程雄,陈庆宁,李咸善,等.响应调峰需求的中小型流域梯级电站短期调度方法[J].水利学报,2023,54(2):159-171. (Cheng Xiong, Chen Qingning, Li Xianshan, et al. Short-term scheduling methods for peak load responding demand of cascade hydropower stations in small and medium-sized river basins [J]. Journal of Hydraulic Engineering, 2023,54(2):159-171. (in Chinese))

[14] 郭旭宁,秦韬,雷晓辉,等.水库群联合调度规则提取方法研究进展[J].水力发电学报,2016,35(1):19-27. (Guo Xuning, Qin Tao, Lei Xiaohui, et al. Advances in derivation method for multi-reservoir joint operation policy [J]. Journal of Hydroelectric Engineering, 2016,35(1):19-27. (in Chinese))

[15] 朱建全,朱文凯,刘海欣,等.近似动态规划在电力系统优化运行中的应用综述[J].电力系统自动化,2024,48(22):1-21. (Zhu Jianquan, Zhu Wenkai, Liu Haixin, et al. Overview on application of approximate dynamic

- programming in optimal operation of power system [J]. Automation of Electric Power Systems, 2024, 48 (22) : 1-21. (in Chinese))
- [16] 张梦婷,张军,何承烽. 元启发式算法研究综述[J]. 计算机工程与应用, 2026, 62 (2) : 40-53. (Zhang Mengting, Zhang Jun, He Chengfeng. Review of metaheuristic algorithms research [J]. Computer Engineering and Applications, 2026, 62 (2) : 40-53. (in Chinese))
- [17] 黄生志,闫俊,郭怿,等. 面向关中地区水网韧性提升的水库群优化调配研究[J]. 水资源保护, 2025, 41 (5) : 258-267. (Huang Shengzhi, Yan Jun, Guo Yi, et al. Optimization and allocation of reservoir groups for enhancing the resilience of water network in Guanzhong area [J]. Water Resources Protection, 2025, 41 (5) : 258-267. (in Chinese))
- [18] 闻昕,黄抒艺,谭乔凤,等. 江苏省南水北调多工程多目标联合优化调度方法[J]. 水资源保护, 2023, 39 (5) : 118-124. (Wen Xin, Huang Shuyi, Tan Qiaofeng, et al. A multi-objective joint optimal operation method for the South-to-North Water Diversion Project in Jiangsu Province [J]. Water Resources Protection, 2023, 39 (5) : 118-124. (in Chinese))
- [19] 钟加星,董增川,孟金玉,等. 基于机器学习的黄河上游梯级水库群多目标优化调度规则提取[J]. 河海大学学报(自然科学版), 2024, 52 (6) : 30-37. (Zhong Jiaxing, Dong Zengchuan, Meng Jinyu, et al. Extracting multi-objective optimal operation rules of cascade reservoirs in upper reaches of Yellow River based on machine learning [J]. Journal of Hohai University (Natural Sciences), 2024, 52 (6) : 30-37. (in Chinese))
- [20] 高冠中,杨胜春,郭晓蕊,等. 深度强化学习在含分布式柔性资源的电网优化调度中的应用研究综述[J]. 中国电机工程学报, 2024, 44 (16) : 6385-6404. (Gao Guanzhong, Yang Shengchun, Guo Xiaorui, et al. A review of research on the application of deep reinforcement learning in optimization dispatch of power grids with distributed flexible resources [J]. Proceedings of the CSEE, 2024, 44 (16) : 6385-6404. (in Chinese))
- [21] Mnih Volodymyr, Kavukcuoglu Koray, Silver David, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518 (7540) : 529-533.
- [22] Zhang Aonan, Liu Pan, Cheng Qian, et al. Deriving reservoir operating rules of spillway gates based on deep reinforcement learning [J]. Journal of Hydrology, 2025, 662:133858.
- [23] Wu Rixin, Wang Ran, Hao Jie, et al. Multiobjective multihydropower reservoir operation optimization with transformer-based deep reinforcement learning [J]. Journal of Hydrology, 2024, 632 (4) : 130904.
- [24] Ge Yuanyu, Xie Jun, Chang Jiaqi, et al. A multi-objective deep reinforcement learning method for intelligent scheduling of wind-solar-hydro-battery complementary generation systems [J]. International Journal of Electrical Power & Energy Systems, 2025, 167:110635.
- [25] Ge Yuanyu, Xie Jun, Feng Shuo, et al. Multi-task deep reinforcement learning with scenario clustering for real-time scheduling of wind-solar-hydro complementary generation systems [J]. IET Renewable Power Generation, 2025, 19 (1) : e70070.
- [26] Lu Peng, Lan Hanqing, Yuan Qiwei, et al. A bi-level solution strategy based on distributed proximal policy optimization for transmission and distribution network dispatch with EVs and variable energy [J]. Applied Energy, 2025, 384:125405.
- [27] Guo Daya, Yang Dejian, Zhang Haowei, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning [J]. Nature, 2025, 645 (8081) : 633-638.
- [28] Li Mingquan, Ma Qingyuan, Shan Rui, et al. Renewable energy quality trilemma and coincident wind and solar droughts [J]. Communications Earth & Environment, 2024, 5 (1) : 661.
- [29] Shen Qin, Mo Li, Liu Zixuan, et al. A multi-objective optimization method of sustainable wind-photovoltaic-hydro systems considering source-grid coordination [J]. Sustainability, 2024, 16 (1) : 61.
- [30] 苏华英,王珍妮,张俨,等. 考虑混蓄电站季调节特性的水风光蓄互补调度研究[J]. 河海大学学报(自然科学版), 2026, 54 (4) : 96-105. (Su Huaying, Wang Zhenni, Zhang Yan, et al. Study on scheduling of hydro-wind-photovoltaic-storage complementary system considering seasonal regulation characteristics of hybrid pumped storage power stations [J]. Journal of Hohai University (Natural Sciences), 2026, 54 (4) : 96-105. (in Chinese))
- [31] 吴文惠,张双虎,张忠波,等. 梯级水库集中调度发电效益考核评价研究:以乌江梯级水库为例[J]. 水力发电学报, 2015, 34 (10) : 60-69. (Wu Wenhui, Zhang Shuanghu, Zhang Zhongbo, et al. Assessment of power-generating benefits from cascade reservoirs integrated operation: a case study of Wu River cascade reservoirs [J]. Journal of Hydroelectric Engineering, 2015, 34 (10) : 60-69. (in Chinese))
- [32] 冯仲恺,罗涛,关铁生,等. 水库群发电调度多智能体深度强化学习方法 [J]. 水利学报, 2026, 57 (7) : 985-1003. (Feng Zhongkai, Luo Tao, Guan Tiesheng, et al. Multi-agent deep reinforcement learning method for cascade reservoir operation for power generation [J]. Journal of Hydraulic Engineering, 2026, 57 (7) : 985-1003. (in Chinese))

(收稿日期:2026 - 01 - 04 编辑:彭桃英)