

基于多智能体最优折中强化学习的多主体含氢 综合能源系统优化调度

彭春华, 钟沂辰, 孙惠娟, 张大权

(华东交通大学电气与自动化工程学院, 江西省 南昌市 330013)

Optimal Scheduling of Multi-entity Hydrogen Integrated Energy Systems Based on Multi-agent Optimal Compromise Reinforcement Learning

PENG Chunhua, ZHONG Yichen, SUN Huijuan, ZHANG Daquan

(School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, Jiangxi Province, China)

ABSTRACT: To fully account for the dynamic coupling of heterogeneous energy and multi-entity interactive collaboration in hydrogen-integrated energy systems, this study proposes a multi-entity collaborative optimization scheduling method for multi-energy carriers based on multi-agent optimal compromise reinforcement learning. First, an optimal scheduling model for multi-entity hydrogen-integrated energy systems is established to minimize operational costs. Subsequently, a multi-agent model is constructed with individual agents as entities, and the optimization scheduling model is solved under a multi-agent reinforcement learning (MARL) framework. Addressing the issues of insufficient collaboration, action conflicts, and low optimization efficiency in traditional multi-agent proximal policy optimization algorithms caused by independent states and actions, this method enhances state-space collaboration by incorporating adjacent agents' action information. Furthermore, a multi-scheme evaluation mechanism based on the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is developed to screen optimal action combinations through compromise solutions, thereby forming a multi-agent optimal compromise reinforcement learning algorithm that improves agent collaboration and solution efficiency. Case study results verify the superior solution accuracy and optimization performance of the proposed model and algorithm.

KEY WORDS: integrated energy systems; multi-agent reinforcement learning; optimal compromise; proximal policy optimization; optimal operation

摘要: 为充分考虑含氢综合能源系统中异质能源动态耦合和多主体交互协同, 实现多能源载体多主体协同优化运行, 提出了一种基于多智能体最优折中强化学习的多主体含氢综合能源系统优化调度方法。首先, 以系统运行成本最低为目标建立多主体含氢综合能源系统优化调度模型, 然后以各主体为对象构建多智能体模型, 并在多智能体强化学习框架下对优化调度模型进行求解。针对传统多智能体近端策略优化算法因状态和动作相互独立而导致协作不足、动作冲突及优化效率低的问题, 该文通过融入相邻智能体动作信息增强状态空间协作, 并基于逼近理想解排序法构建多方案评估机制, 筛选最优折中解的动作组合, 形成多智能体最优折中强化学习算法, 提升智能体间的协作和求解效率。算例仿真结果验证了所构建模型及算法在求解精度和优化性能上的优势。

关键词: 综合能源系统; 多智能体强化学习; 最优折中; 近端策略优化; 优化调度

DOI: 10.13335/j.1000-3673.pst.2025.0692

0 引言

2024 年国家发展改革委、国家能源局联合印发的《关于加强电网调峰储能和智能化调度能力建设的指导意见》明确提出“探索多能源品种和源网荷储协同调度机制”以及“明确源网荷各侧调节资源和风光储联合单元、负荷聚合商、虚拟电厂等主体的独立市场地位”的策略, 为多主体综合能源系统协同互动的未来能源利用模式指明了发展方向^[1]。而包含了含氢综合能源服务商(hydrogen integrated energy provider, HIEP)、电负荷聚合商(electrical load aggregators, ELA)、热负荷聚合商(thermal load aggregators, TLA)和氢负荷聚合商(hydrogen load aggregators, HLA)等多主体交互的含氢综合能源系统(hydrogen integrated energy systems, HIES)作为一

基金项目: 国家自然科学基金项目(52267007, 52567008); 江西省自然科学基金项目(20242BAB26070)。

Project Supported by the National Natural Science Foundation of China (52267007, 52567008); Jiangxi Provincial Natural Science Foundation (20242BAB26070).

种集成多种能源类型和运营主体的新型能源供应模式，因其具备的多能互补、高效转换和灵活调度能力，被视为推动能源绿色转型的关键路径之一[2-3]。

当前已有一些针对 HIES 的研究。文献[4]构建氢燃料电池热电协同控制模型，量化修正光伏出力预测曲线并优化多设备动态匹配策略，有效降低运行成本并提升系统可靠性；文献[5]计及氢能流构建了电热冷氢多能流综合能源系统，提高风光装机比例和利用率，提高系统经济性和稳定性；文献[6]建立了含多种氢能利用设备的氢能综合利用单元，研究其与多能系统运行的调度策略，解决高比例可再生能源科学消纳与能源系统灵活稳定运行问题。上述研究大多聚焦于单主体场景，且只将氢能作为耦合转化能源而不是作为独立决策主体参与能源优化运行。当 HIES 的优化调度在多主体场景和多种异质能源载体并存的情况下面临不同主体间的利益冲突、信息不对称以及能源供需的不确定性等挑战，使系统的协同优化成为一个典型的多目标、多约束的复杂决策问题[7]。

现有多主体综合能源系统研究大多通过博弈论构建主体间的利益分配框架。文献[8]提出基于合作博弈与 Shapley 值法的多主体协同调度策略，电气-热多能耦合按运营主体构建合作联盟，协同交互降低运行成本；文献[9]针对共享储能背景下的综合能源微网优化问题，提出基于 Stackelberg 博弈的混合求解策略，构建能保障均衡解唯一性的双层决策模型；文献[10]考虑碳约束建立电网公司、天然气公司与能源站的多主体收益模型，采用双层 Nash-Stackelberg 博弈满足了各主体利益的最大化。以上研究通过多主体博弈模型解决了利益分配冲突，同时实现了系统的经济运行，但主要针对传统能源场景，并使用粒子群算法、启发式算法和混合整数规划等传统优化算法。引入氢能后，设备耦合机理更复杂，系统运行不确定性显著增加，因此需构建能精准描述电热氢等能源动态耦合和多主体交互的模型，并使用更适应环境变化、能协调多方主体利益的优化算法。

相比传统优化算法，强化学习在复杂系统中处理非线性、不确定性和复杂决策方面的表现更优[11-12]。随着强化学习要处理的问题日益复杂，其应用已从单智能体强化学习转向多智能体强化学习，单智能体强化学习如文献[13]改进分布式近端策略优化算法，将高维非线性约束热电联产系统经

济调度问题建模为马尔可夫决策过程；文献[14]构建电-氢混合制氢综合利用运行模式，采用基于乐观行动-评判的深度强化学习方法实现能源系统的高效调度与降碳增效。多智能体强化学习更适用于多主体综合能源系统优化运行场景，如文献[15]将系统按利益主体划分为多智能体，采用多智能体深度确定性策略梯度算法，通过局部信息独立计算与全局奖励协同机制提高收敛稳定性；文献[16]将多智能体深度确定性策略梯度算法(multi-agent deep deterministic policy gradient, MADDPG)扩展到多主体虚拟电厂场景中用于解决效益分配问题；文献[17]将每个分布式储能视为独立智能体，采用多智能体深度确定性策略梯度算法求解高比例波动性能源接入配电网引发的电压调节问题；文献[18]提出基于多智能体近端策略优化(multi-agent proximal policy optimization, MAPPO)的多园区综合能源系统实时经济调度方法，降低了系统日运行成本。上述研究应用的各类算法实现了多主体场景下的调度优化，方法大多针对高维连续动作构建策略网络，其局限主要体现在梯度方差大和超参数敏感等问题。MAPPO 的概率比率剪切机制显著提升了收敛稳定性，但在分散执行阶段缺乏全局优化仍可能导致资源分配僵化，MADDPG 则因确定性策略噪声易陷局部最优，训练振荡明显。

鉴于此，本文拟通过构建含氢综合能源服务商智能体(HIEP-Agent)、电负荷聚合商智能体(ELA-Agent)、热负荷聚合商智能体(TLA-Agent)和氢负荷聚合商智能体(HLA-Agent)4 个智能体组成的多智能体强化学习模型，实现智能体间更高效的互动和协调。相比依赖高维连续动作空间的 MADDPG 等算法，MAPPO 的概率比率裁剪机制可限制策略更新幅度，解决梯度不稳定和收敛慢的问题；但 MAPPO 在策略更新时仍依赖独立状态、独立动作和单一策略网络，缺乏全局优化，易引发动作冲突，能源分配不够灵活，从而影响优化效率。针对上述不足，本文提出了一种将 MAPPO 算法与基于逼近理想解排序法(technique for order preference by similarity to ideal solution, TOPSIS)[19]的最优折中机制相结合，形成的多智能体最优折中强化学习算法(multi-agent optimal compromise proximal policy optimization, MAOCPPPO)用于求解 IES 多主体优化调度问题。该算法创新性体现在：改进了 MAPPO 的多轮次更新机制，通过并行多排列方案扩大搜索

空间；通过将相邻智能体动作信息纳入当前状态空间以增强协作，将逼近理想解排序法应用于筛选折中解并建立多方案评估机制，获取最优调度方案。该框架在训练过程中通过并行多排列方案和逼近理想解排序方法实现信息交互与多样化探索，有效平衡智能体间竞争合作关系，提升决策质量。

1 多主体含氢综合能源系统

本文研究的多主体含氢综合能源系统模型如图1所示。该综合能源系统由上级电网、天然气网、HIEP、ELA、TLA和HLA等主体构成。HIEP内主要设备有热电联产机组(combined heat and power, CHP)、电锅炉设备(electric boiler, EB)、电制氢设备(electrolyzer, EL)、燃料电池(hydrogen fuel cell, HFC)等^[20]，能够实现电、热、氢的转换与供应；负荷聚合商主要设备有电储、热储、氢储等设备。

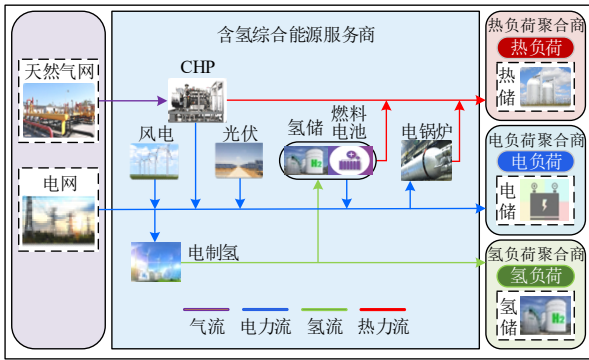


图1 多主体含氢综合能源系统模型图

Fig. 1 Multi-agent hydrogen integrated energy system model schematic

HIEP 可以与上级电网交互进行购电和售电。售电价低于 HIES 内部交易电价，而购电价高于向 ELA 售电的价格，因此 HIEP 优先满足各 LA 需求。ELA 管理 HIES 中分散的中小电负荷，配备电储；TLA 管理 HIES 热负荷，配备热储；HLA 管理 HIES 氢负荷，配备氢储。负荷分为重要固定负荷、可削减负荷和可平移负荷。各 LA 通过需求响应(demand response, DR)利用储能和弹性负荷优化用能^[21]，降低成本。

2 多主体含氢综合能源系统优化调度模型

2.1 优化调度目标函数

HIES 优化运行的目标是在满足各项约束条件的前提下，通过调整运行策略，实现成本最低，本文综合考虑 HIES 中各项成本，构建如下函数：

$$\begin{cases} C^{\text{om}} = \sum_{n=1}^N \sum_{t=1}^T c_n^{\text{om}} P_t^n \\ C^{\text{buy,e}} = \sum_{t=1}^T \mu_t^e P_t^e c_t^e \\ C^{\text{buy,g}} = c_g \sum_{t=1}^T \left(\frac{P_t^{\text{chp,e}}}{\mu_g \eta_{\text{chp,e}}} + \frac{P_t^{\text{chp,t}}}{\mu_g \eta_{\text{chp,t}}} \right) \\ C^{\text{waste,e}} = \sum_{t=1}^T (\tau_e P_t^{\text{re,e}} + \eta_e P_t^{\text{d,e}}) \\ C^{\text{waste,t}} = \sum_{t=1}^T (\tau_t P_t^{\text{re,t}} + \eta_t P_t^{\text{d,t}}) \\ C^{\text{waste,h}} = \sum_{t=1}^T (\tau_h P_t^{\text{re,h}} + \eta_h P_t^{\text{d,h}}) \end{cases} \quad (1)$$

$$\min F = \min(C^{\text{om}} + C^{\text{buy,e}} + C^{\text{buy,g}} + C^{\text{waste,e}} + C^{\text{waste,t}} + C^{\text{waste,h}}) \quad (2)$$

式中： C^{om} 为机组运行成本； c_n^{om} 和 P_t^n 分别为 t 时刻各设备的运维成本和出力值； $C^{\text{buy,e}}$ 表示与上级电网交互成本； $\mu_t^e \geq 0$ 表示购电； $\mu_t^e < 0$ 表示售电； P_t^e 为与电网交互功率； c_t^e 为与电网交互电价； $C^{\text{buy,g}}$ 为购气成本； c_g 为天然气单价； μ_g 为天然气热值； $P_t^{\text{chp,e}}$ 、 $\eta_{\text{chp,e}}$ 分别为 CHP 机组发电功率和发电效率； $P_t^{\text{chp,t}}$ 、 $\eta_{\text{chp,t}}$ 分别为 CHP 机组发热功率和发热效率； $C^{\text{waste,e}}$ 、 $C^{\text{waste,t}}$ 、 $C^{\text{waste,h}}$ 分别为电负荷、热负荷、氢负荷的削减补偿及平移成本； τ_e 、 τ_t 、 τ_h 分别为电负荷、热负荷、氢负荷的削减补偿成本系数； $P_t^{\text{re,e}}$ 、 $P_t^{\text{re,t}}$ 、 $P_t^{\text{re,h}}$ 分别为削减的电负荷、热负荷、氢负荷总量； η_e 、 η_t 、 η_h 分别为电负荷、热负荷、氢负荷的平移成本系数； $P_t^{\text{d,e}}$ 、 $P_t^{\text{d,t}}$ 、 $P_t^{\text{d,h}}$ 分别为平移的电负荷、热负荷、氢负荷总量。

2.2 约束条件

HIES 运行的约束条件包括 CHP 机组约束、EB 设备约束、EL 设备约束、HFC 设备约束、可削减负荷约束、可平移负荷约束、电功率平衡约束、热功率平衡约束、氢功率平衡约束、储能设备约束，其详细表达式见附录 A。

3 多智能体最优折中强化学习算法设计

针对前述多主体含氢综合能源系统模型，本文构建了由 HIEP-Agent、ELA-Agent、TLA-Agent 和 HLA-Agent 这 4 个智能体组成的多智能体强化学习模型。

多智能体强化学习算法需要将问题转换成多智能体马尔可夫决策过程(multi-agent markov decision process, MMDP)，需要考虑多个智能体之间的相互作用，主要包含 9 个元素 ($S, A, N, R,$

P, γ, π, o, Π), 其中: S 为状态空间; A 为动作空间; N 为智能体数量; R 为奖励; P 为转移概率; γ 为折扣因子; π 为智能体策略; o 为观测空间; Π 为联合智能体策略。

3.1 状态空间设计

多主体含氢综合能源系统环境通过数学模型描述系统中能量转换的动态过程, 用于支持智能体之间的交互。状态空间是智能体能够获取的环境特征, 用于描述其感知到的环境信息。

环境提供给 HIEP-Agent 的信息为预计总电负荷、预计总热负荷、预计总氢负荷、风机和光伏出力、HFC 热出力、CHP 热出力、HIES 内部各主体交易电价、上级电网购售电价、可削减电负荷、可削减热负荷、可削减氢负荷、可平移电负荷、可平移热负荷、可平移氢负荷和 HIEP 内氢储能荷电状态。

ELA-Agent 可获取的环境信息为 HIES 内部各主体交易电价、风机和光伏出力、EL 消耗电功率、电锅炉消耗电功率、CHP 电出力、HFC 电出力和预计总电负荷和电储能荷电状态。

TLA-Agent 的环境信息为 HIES 内部各主体交易热价、HFC 热出力、CHP 热出力、电锅炉热出力、预计总热负荷和热储能荷电状态。

HLA-Agent 的环境信息为 HIES 内部各主体交易氢价、HFC 消耗的氢功率、电制氢的氢出力、预计总氢负荷和 HLA 内氢储能荷电状态。

基于上述信息, 各智能体的状态空间定义式见附录 B。

3.2 动作空间设计

基于决策变量间的耦合特性, 各智能体动作空间可定义如下: HIEP-Agent 动作简化为 CHP 机组热出力、燃料电池 HFC 热出力和 HIEP 内氢储能出力; ELA 通过需求响应削减电负荷并调节储能充放电功率, 其动作空间由电负荷削减量和电储能功率构成; TLA 通过热负荷调控与储热系统配合, 动作空间包含热负荷削减量和热储能功率; HLA 则通过氢负荷削减与氢储能调节, 形成氢负荷削减量和氢储能功率组成的动作空间。各智能体的动作空间定义式见附录 C。

3.3 奖励函数设计

对于 HIEP-Agent, 其优化目标是为最小化系统运行成本, 目标函数公式即奖励函数, 各 LA 智能体优化目标是最小用能成本, 包括负荷削减补偿成本 c_t^{dis} 、负荷平移成本 c_t^{d} 、储能运维成本 c_t^{om} 和 HIES 内部各主体交易购能成本 c_t^{buy} :

$$r_t^{\text{HIEP}} = -\phi F + \varepsilon \quad (3)$$

$$r_t^{\text{load,e}} = -\phi(c_t^{\text{dis,e}} + c_t^{\text{d,e}} + c_t^{\text{om,e}} + c_t^{\text{buy,e}}) + \varepsilon \quad (4)$$

$$r_t^{\text{load,t}} = -\phi(c_t^{\text{dis,t}} + c_t^{\text{d,t}} + c_t^{\text{om,t}} + c_t^{\text{buy,t}}) + \varepsilon \quad (5)$$

$$r_t^{\text{load,h}} = -\phi(c_t^{\text{dis,h}} + c_t^{\text{d,h}} + c_t^{\text{om,h}} + c_t^{\text{buy,h}}) + \varepsilon \quad (6)$$

式中: r_t^{HIEP} 为 HIEP-Agent 奖励; $r_t^{\text{load,e}}$ 、 $r_t^{\text{load,t}}$ 、 $r_t^{\text{load,h}}$ 分别为 ELA-Agent、TLA-Agent、HLA-Agent 奖励; ϕ 为成本放缩比; ε 为使奖励函数回归正值的辅助值。

3.4 多智能体最优折中强化学习算法

3.4.1 MAPPO 算法基础

多智能体强化学习(multi-agent reinforcement learning, MARL)算法让多个智能体在复杂环境中通过自身经验和互动, 独立或合作地学习和优化策略。多智能体近端策略优化 MAPPO 算法是多智能体强化学习算法的一种, 是近端策略优化算法 PPO 在多智能体环境中的扩展, 确保每个智能体的策略更新具有稳定性的同时允许策略之间有一定程度的变化, 通过裁剪机制和多智能体特定的损失函数, 维持概率比率的稳定性和策略的多样性。

MAPPO 算法的网络更新基于 Actor-Critic 网络, 智能体使用 Actor 网络进行策略的输出, Critic 网络则估计状态-行动对的价值。Actor 网络的参数通过梯度上升进行更新, 以最大化裁剪后的目标函数, Actor 网络的更新目标函数:

$$\begin{cases} L_{\text{actor}}^i = L_{\text{clip}}^i + L_{\text{entropy}}^i \\ L_{\text{clip}}^i = E_t[\min(r_t^i A_t, \text{clip}(r_t^i, 1 - \varepsilon, 1 + \varepsilon) A_t)] \\ L_{\text{entropy}}^i = \eta \cdot E_t[H(\pi_{\theta^i})] \end{cases} \quad (7)$$

式中: r_t^i 为新旧策略概率比, 剪切机制限制策略更新幅度, 当新旧策略差异过大时, 用边界值替代实际概率比; A_t 为广义优势估计值, 通过 Critic 网络计算得到; ε 为剪切范围, 限制策略更新幅度; η 为熵正则系数; π_{θ^i} 是第 i 个智能体的策略网络参数; $H(\pi_{\theta^i})$ 为策略熵, 用于鼓励探索。

Critic 网络通过最小化值函数预测值与实际回报之间的均方误差进行参数更新:

$$L_{\text{critic}} = E_t[(V_{\phi}(s_t) - \hat{V}_t^{\text{target}})^2] \quad (8)$$

式中: $V_{\phi}(s_t)$ 为 Critic 网络对全局状态 s_t 的价值预测; $\hat{V}_t^{\text{target}}$ 是通过轨迹数据计算的目标价值。

3.4.2 最优折中机制

MAPPO 的多智能体决策方法存在 3 方面局限性: 1) MAPPO 依赖策略梯度迭代优化机制, 需通过多轮次策略更新与环境交互逐步收敛至稳态策略, 其在单步决策周期内无法找到当前最优动作序

列。2) 单一策略网络生成固定顺序的动作序列会固化能源分配模式,无法根据动态环境状态生成异构调度方案,导致资源供需匹配灵活性不足。3) 在线训练机制限制了策略多样性,复杂系统需通过多样化动作序列探索更优解空间。

为解决上述问题,本文提出多智能体最优折中机制,旨在通过智能体间的有序动作生成和依赖关系,得到最优折中解用于各主体奖励与系统整体效益的提升。最优折中机制主要通过结合有序决策生成机制、多方案并行运行与奖励累加机制和 TOPSIS 方法实现。

1) 有序决策生成机制。有序决策生成机制的设计灵感来源于意图共享(intention sharing, IS)的概念。Kim 等人提出了一种意图共享方案,使智能体不仅能够共享当前的局部观测,还能传递其在未来一段时间内的意图信息^[22]。在每一步决策中,各智能体基于自身的局部观测和前一步接收到的数据,预测后续的动作和观测序列,并通过注意力机制对这些序列进行编码,提取关键特征。编码后的信息随后广播至其他所有智能体。然而,智能体在共享意图后可能调整自身动作,导致实际行为偏离先前意图,从而对其他智能体产生误导。为解决这一问题,Liu 等人提出了一种层次化前馈(layered feed-forward, LFF)方法,通过动态构建智能体之间的层次结构来优化决策过程。在该方法中,高层次智能体优先完成决策,并将动作信息传递给低层次智能体,低层次智能体接收意图动作信息,并据此进行决策。智能体严格按照层次顺序逐级执行决策,从而实现单向意图传递,有效避免了误导现象的发生^[23]。

受上述研究的启发,本文设计有序决策生成机制,智能体按照设定的处理顺序依次生成动作。当前智能体的动作生成依赖于其自身的状态和前一个智能体的动作,前一个动作被动态添加到当前智能体的观测数据中,形成增强观测向量。通过这种方式,后续智能体能够结合前序智能体的决策信息通过神经网络调整自身动作,从而更好地适应全局任务目标。增强观测的数学表达如下:

$$O_t^i = \text{concat}(s_t^i, a_t^j) \quad (9)$$

式中: O_t^i 为第 i 个智能体的增强观测; concat 为拼接两部分信息的操作; s_t^i 为其当前状态; a_t^j 为前一个智能体的动作。

以 HIEP-Agent、ELA-Agent、TLA-Agent 和 HLA-Agent 的有序决策生成动作顺序为例,在 t 时刻, HIEP 状态观测空间不变, ELA-Agent 增强观测空间:

$$O_t^{\text{ELA}} = \text{concat}(s_t^{\text{ELA}}, a_t^{\text{HIEP}}) \quad (10)$$

式中: O_t^{ELA} 为 ELA-Agent 的增强观测; s_t^{ELA} 为 ELA-Agent 的状态空间; a_t^{HIEP} 为 HIEP-Agent 的动作。

ELA-Agent 增强观测空间作为输入,使用 Actor 网络生成动作, TLA-Agent 和 HLA-Agent 也按照有序决策生成机制生成动作。HIEP-Agent、ELA-Agent、TLA-Agent 和 HLA-Agent 这 4 个智能体按照排列组合,共有 24 种动作生成顺序。

2) 多方案并行运行与奖励累加机制。有序决策生成机制带来了 $n!$ 种智能体动作生成顺序(其中 n 为智能体数量),每种顺序对应一个完整的动作生成方案。为全面评估不同方案的优化性能,算法对所有生成的方案并行运行,具体是首先每种方案独立输入环境,与环境交互获得调度周期内的累积奖励和状态转移序列,其次所有方案均以并行方式执行,互不干扰。这种设计避免了单一方案可能导致的局部最优问题,同时加速了方案评估的效率。

对于给定调度周期,方案的累积奖励计算如下:

$$R_{\text{total}}^k = \sum_{t=1}^T r_t^k \quad (11)$$

式中: R_{total}^k 为方案 k 的累积奖励; T 为调度周期的时间步数。

3) 逼近理想解排序法。获得 24 个方案的累积奖励后,每个方案有 4 个智能体奖励,在综合考虑 4 个奖励的情况下利用逼近理想解排序法 TOPSIS 方法获得 24 个方案中的最优折中方案。TOPSIS 是一种多目标决策方法,通过比较每个方案与正理想解和负理想解的相对距离,选取最接近理想解的方案作为最终的最优折中解。TOPSIS 评价流程示意图见附录 D 图 D1。

3.4.3 多智能体最优折中强化学习算法

本文在此将多智能体强化学习 MAPPO 算法与基于 TOPSIS 的最优折中机制相结合,形成新型的多智能体最优折中强化学习算法 MAOCPPO。上文详细阐述了多智能体最优折中强化学习算法的相关机制和原理,MAOCPPO 框架图见附录 D 图 D4,为进一步清晰地展示多智能体最优折中强化学习的实现过程,阐述算法更新流程如下:

1) 初始化。初始化多智能体的策略网络、价值网络参数和环境状态,设定 TOPSIS 的权重向量及调度周期等超参数。

2) 动作生成与方案构建。通过有序决策生成机制,对 $n!$ 种智能体动作生成顺序依次进行动作采样,并构建对应的调度方案。

3) 环境交互与奖励累加。将所有方案输入环境, 进行完整调度周期的运行, 并累积其奖励值和状态信息。

4) TOPSIS 评价与折中解选择。基于 TOPSIS 方法对所有方案的累积奖励进行评价, 选出最优折中解, 该解对应的动作序列即为当前优化周期的最终调度方案。

5) 网络参数更新。根据最优折中解的奖励信号, 使用 MAPPO 算法更新智能体的策略网络和价值网络参数。

6) 时间步推进。若未达到设定的训练回合数, 则进入下一回合, 返回步骤 2)。

7) 训练终止。当达到最大训练回合数或算法收敛条件时, 训练终止, 并输出最终优化策略。

多智能体最优折中强化学习的主要流程见附录 D 表 D1。

4 算例分析

4.1 算例说明

以图 1 所建立的多主体含氢综合能源系统为例, 重点分析系统在降本增效方面的性能。通过采用 MAOCPPO 算法进行优化调度, 实现多主体含氢综合能源系统的高效运行。为验证所提方法的有效性与优势, 对不同优化算法进行对比分析, 综合评估算法的应用效果。HIES 内的设备参数见附录 D 表 D2; HIES 运行成本参数见附录 D 表 D3。HIES 从上级电网购电范围为 $[0, 35\text{MW}]$; HIES 与上级电网和天然气网交易分时价格见附录 D 图 D2。天然气热值为 $9.7\text{kW}\cdot\text{h}/\text{m}^3$, 氢气热值为 $40\text{kW}\cdot\text{h}/\text{kg}$, HIES 内部氢气交易价格为 25 元/kg, HIES 内部热能交易价格为 0.6 元/(kW·h)。算例中所用的设备参数及运行成本参数参考自文献[14]。

4.2 仿真结果分析

4.2.1 奖励曲线分析

算例仿真硬件为 AMD Ryzen 7 5800H 处理器。本文算法和训练基于 Pytorch 平台进行搭建, 采用 Python3.8 语言。图 2 是 MAOCPPO 算法和 MAPPO 算法在 HIES 中的训练过程奖励值曲线对比。在训练过程中定义单次训练成本奖励为 24h 的累计值, 奖励值越低代表系统运行成本越低。奖励值随训练次数的变化趋势表明, MAOCPPO 算法相较于 MAPPO 算法, 其在训练初期表现出更快的收敛速度, 并最终稳定获得更低的成本。MAOCPPO 算法具备探索更为多样化的动作序列的能力, 能够发现更高效的动作策略, 从而有效避免陷入局部最优

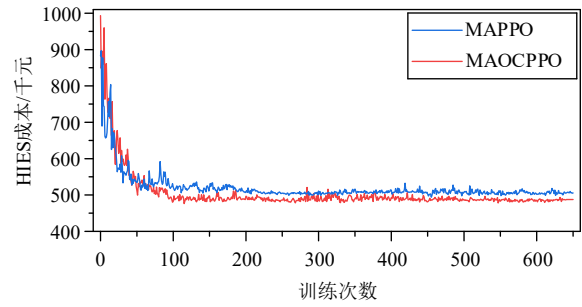


图 2 训练过程 HIES 成本变化曲线对比

Fig. 2 Comparison of HIES cost change curves during the training process

解, 提升系统的整体效益。

如图 3 所示, 随着 MAOCPPO 训练的推进, 网络参数逐步朝着各智能体奖励最大化及成本最小化的方向更新。在约 200 次训练后, 各 Agent 的成本值趋于平稳。ELA-Agent 的成本曲线在收敛稳定后仍呈现出一定范围的波动, 这主要是因为电价具有分时波动特性而导致 ELA 的成本结构不够稳定, 且该智能体可执行的动作相对有限, 导致其策略难以完全平滑电价变化带来的成本变化。相比之下, TLA-Agent 和 HLA-Agent 的成本值在收敛稳定后

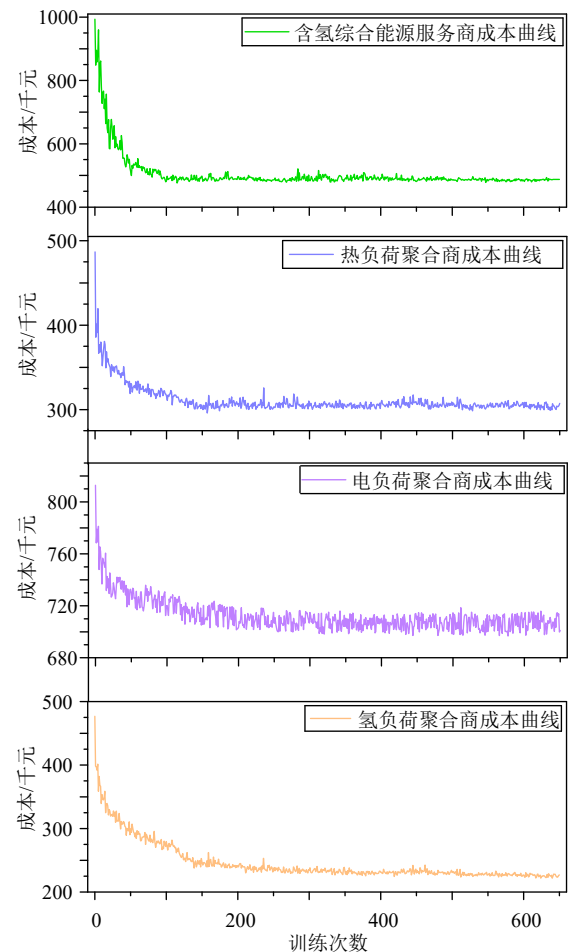


图 3 MAOCPPO 训练过程成本变化曲线

Fig. 3 MAOCPPO cost change curves during training

波动较小，其原因是系统内热价和氢价为固定值，成本结构相对稳定。此外，HIEP-Agent 涉及多种设备类型和系统层级的优化策略，其成本波动也较小，进一步表明 MAOCPPO 算法在复杂系统中的稳定性和优化效果。

4.2.2 优化调度结果分析

为进一步验证本文 MAOCPPO 算法的在线决策能力及多主体含氢综合能源系统模型的合理性，将训练好的算法用于多主体含氢综合能源系统优化调度运行。HLA 各项成本如表 1 所示。通过氢储能的灵活调节和氢负荷需求响应，HLA 的购氢成本从优化前的 25.175 万元降低至 22.755 万元，减少了 2.42 万元。而氢负荷削减补偿成本与氢储能运维成本的总和约 5.83 万元，远低于购买成本的降幅，可见氢储能和氢负荷需求响应显著降低了 HLA 的整体成本。

表 1 HLA 成本			
Table 1 Hydrogen load aggregators cost table			
成本项	成本/万元	成本项	成本/万元
优化前购氢成本	25.175	氢负荷削减补偿及平移成本	0.467
优化后购氢成本	22.755	氢储能运维成本	0.116

从图 4 的 HLA 负荷曲线与储能出力优化结果可以看出，氢储能的出力策略与电价的峰平谷时间段紧密相关。在 06:00—21:00 的峰平价时段，氢储能装置主要释放氢气以满足需求；而在 22:00—05:00 的谷价时段，则将低价电力转化成氢气储存备用。这种策略有效降低制氢成本，同时提高系统的经济性。电价的峰平时间段能保持高输出的原因在于氢储能初始容量设置较高，使得系统能够充分利用储氢资源，充分挖掘 HLA-Agent 的潜力，深度参与调度运行。

结合图 5 中 HIEP 氢能出力优化结果可以看出，在峰价时段 07:00—11:00 和 16:00—21:00，系统以

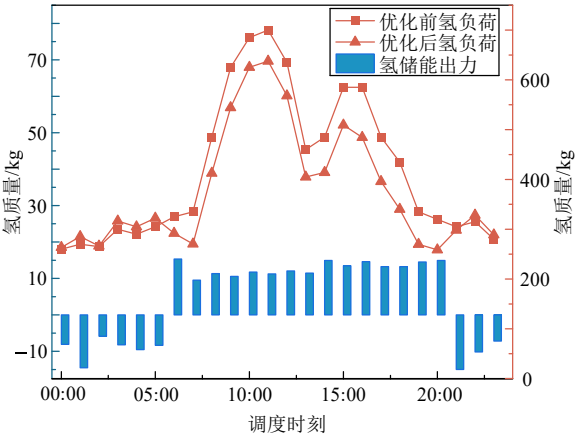


图 4 HLA 负荷曲线与储能出力优化结果
Fig. 4 Hydrogen load aggregators curve and energy storage output optimization results

降低运行成本为核心目标，通过大幅削减氢负荷减少对高价电网电力的依赖。此时氢能需求较高，但 EL 的电力成本激增，因此主动削减非必要氢负荷，同时降低 HFC 的耗氢量以控制消耗。氢储能装置在此阶段持续释放储存的氢气，补充系统缺口，缓解高价制氢压力。峰价时段，虽然风光总和出力较高，但是一方面因为电价处于峰值且系统需要电量较多，系统仍需要通过削减氢负荷和调整 HFC 出力来降低成本。在电价平缓时段，系统兼顾经济性与稳定性，氢负荷削减量适度调整，既避免过度依赖外部电力，又确保基础需求得到满足。EL 制氢量根据实时负荷灵活波动，结合平价电价优化生产成本。氢储能装置继续释放能量，弥补可再生能源的出力波动。夜间电价低谷时段 22:00—05:00，系统以最大化制氢和储能为核心策略。氢负荷削减量降至最低，充分满足需求；氢储能装置在此阶段持续充能，利用低价电力制氢气储存备用。此外，氢负荷高峰时间段与峰平谷价格时间段存在部分重合。在这些重合时段，系统也通过灵活调节氢储能和氢负荷需求响应，有效平衡高需求与高成本之间的矛盾。

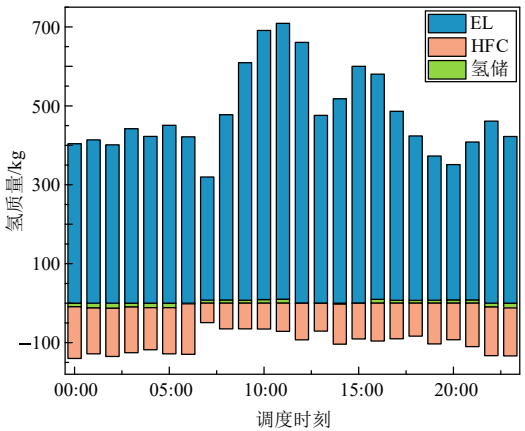


图 5 HIEP 氢能出力优化结果
Fig. 5 Hydrogen integrated energy providers hydrogen energy output optimization results

TLA 成本表如表 2 所示，通过热储能的灵活调节和热负荷需求响应，优化后购热成本从优化前购热成本的 35.094 万元降低至 30.120 万元，降幅约为 14.2%。而热负荷削减补偿成本与热储能运维成本的总和约 3.117 万元，仅为购买成本降幅的一半左右，热储能和热负荷需求响应也降低了 TLA 的整体成本。

表 2 TLA 成本表			
Table 2 Thermal load aggregators cost table			
成本项	成本/万元	成本项	成本/万元
优化前购热成本	35.094	热负荷削减补偿及平移成本	3.006
优化后购热成本	30.120	热储能运维成本	0.111

图 6 中 TLA 负荷曲线与储能出力优化结果和图 4 有相似之处,但热储能的出力策略不仅与电价的峰平谷时间段紧密相关,还与天然气价相关。由于天然气价与电价的峰平谷时间段一致,热储能装置同样在 06:00—21:00 的峰平价时段主要释放热能以满足需求,而在 22:00—05:00 的谷价时段,则将低价电力和低价天然气转化成热能储存备用。这种策略使得优化后热负荷高于优化前热负荷,一方面因为优化前热负荷比较平缓,另一方面因为 TLA 在低价时段从 HIEP 购买热能并储存,这部分热能被计入负荷。电价和天然气价的峰平时间段能够保持高输出的原因在于热储能初始容量设置较高,与氢储能的初始容量设置原因相同。

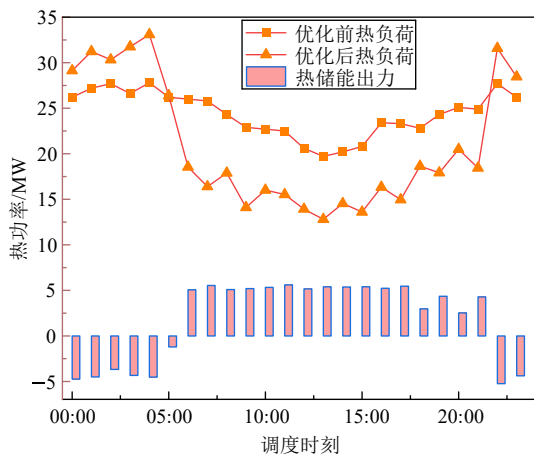


图 6 TLA 负荷曲线与储能出力优化结果

Fig. 6 Thermal load aggregators curve and energy storage output optimization results

结合图 6 和图 7,在电价与气价双高时段,系统以降低成本为核心,显著增加热负荷削减量,减少对高价能源的依赖。此时主要由 CHP 出力满足热能需求, HFC 热出力主要由氢负荷决定, EB 仅作为辅助热源少量启用,避免增加成本;热储能出力补充高峰时段的热能缺口,缓解高价能源的供能

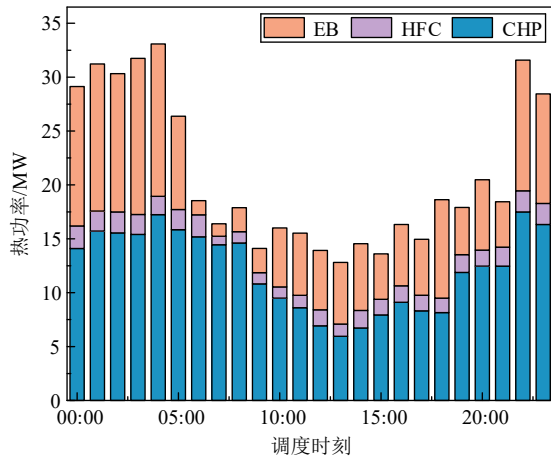


图 7 HIEP 热能出力优化结果

Fig. 7 Hydrogen integrated energy providers heat energy output optimization results

压力。电价与气价适中时段与光伏高峰出力时间段重叠, CHP 出力根据风光出力的波动动态优化, HFC 热出力逐步回升,与 CHP 形成互补,增强系统的调节能力;热储能保持放能模式,平滑可再生能源出力的波动,确保供需平衡。在电价与气价谷时段,系统以储热为重点,热负荷削减量降至最低,最大化满足基础需求。HFC 热出力维持稳定, EB 成为主力热源,充分利用低价电能大规模制热;热储切换至充能模式,为峰价时段储备资源。

ELA 成本如表 3 所示,通过电储能的调节和电负荷需求响应,优化后购电成本从优化前的 89.323 万元降低至 69.479 万元,降幅约为 22%。而电负荷削减补偿成本与电储能运维成本的总和约 5.94 万元,仅为购买成本降幅的约 30%,电储能和电负荷需求响应也降低了 ELA 的整体成本。

表 3 ELA 成本			
Table 3 Electric load aggregators cost table			
成本项	成本/万元	成本项	成本/万元
优化前购电成本	89.323	电负荷削减补偿及平移成本	4.117
优化后购电成本	69.479	电储能运维成本	1.823

ELA 负荷曲线与储能出力优化结果如图 8 所示。HIEP 电出力优化结果如图 9 所示。优化后电负荷高于优化前电负荷的原因在于 ELA 在谷价时段从电网购买低价电能并储存,这部分电能被计入负荷。峰平时间段能够保持高输出的原因在于电储能的初始容量较高,这与热储能和氢储能的初始容量设置逻辑一致。

图 9 中 HFC 设备、CHP 机组、EL 设备、EB 设备的出力逻辑与氢能、热能相似,但 HIEP 电能具有独特优势,可与上级电网进行交互,电储能不仅能补充高峰时段供电缺口,还能通过向电网售电优化收益,缓解高价能源的压力。

HIEP 与电网交互结果如图 10 所示。在谷价时

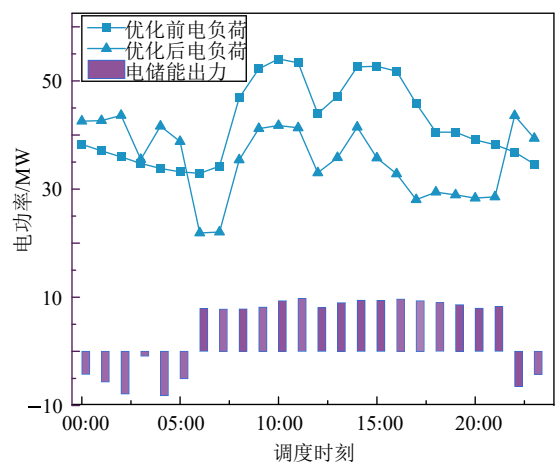


图 8 ELA 负荷曲线与储能出力优化结果

Fig. 8 Electric load aggregators curve and energy storage output optimization results

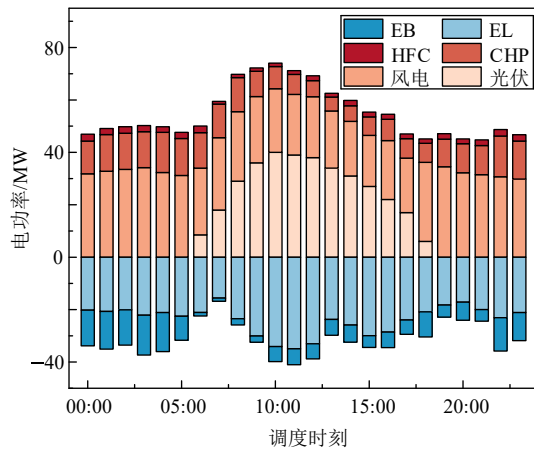


图 9 HIEP 电能出力优化结果
Fig. 9 Hydrogen integrated energy providers electric energy output optimization results

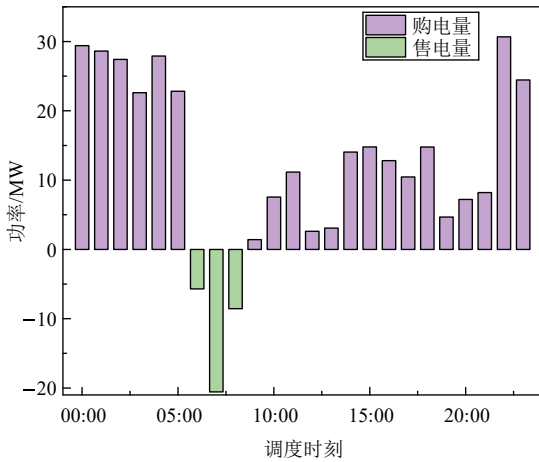


图 10 HIEP 与电网交互优化结果
Fig. 10 Hydrogen integrated energy provider grid interaction optimization results

段大量购电，一方面为储能充电，另一方面为高价时段向电网售电赚取差价做准备。在 06:00—08:00，将谷时电卖在峰平价时间段。其他峰平价时段的购电是为了维持电功率平衡，购电量相对较少。这种策略优化了能源利用效率，降低了系统的综合成本。

4.3 方法对比

4.3.1 算法对比

为验证本文算法对多主体含氢综合能源系统调度运行的优化效果，分别采用典型的群智能优化算法-粒子群优化(particle swarm optimization, PSO)及常用的多智能体强化学习算法 MADDPG、MAPPO 作为对比算法，对本文模型进行优化求解。各优化算法对比结果如表 4 所示。

由表 4 所示可知，PSO 优化算法在处理复杂调度问题时，由于存在易陷入局部最优问题且计算复杂度较高，其优化表现较差，优化成本和所需时间都最高。MADDPG 和 MAPPO 算法在优化时间和成本上都有所改进，MADDPG 由于依赖高维连续动作空间，面临梯度不稳定和收敛速度慢的问题，成本相较于 MAPPO 提高了约 1.5%。多智能体最优

表 4 不同优化方法对比		
Table 4 Comparison of different optimization methods		
优化方法	成本/万元	优化时间/s
PSO	53.6241	727
MADDPG	50.9716	1
MAPPO	50.2135	1
MAOCPPO	48.3864	1

折中强化学习算法 MAOCPPO 对 MAPPO 进行改进，调度成本较 MAPPO 降低了约 1.8 万元，调度成本下降约 3.64%，同时保持了高效的收敛速度。上述算法对比结果验证了 MAOCPPO 算法在多主体含氢综合能源系统调度中的优越性，能够更有效地降低调度运行成本并提高优化效率。

4.3.2 场景变化结果对比

为分析可再生能源随机波动性对算法鲁棒性的影响，基于对风电/光伏的预测场景数据及其随机分布特性，经聚类分析得到 5 种风电/光伏出力场景，出力曲线见附录 D 图 D3，并进一步采用 MAOCPPO 算法得到各不同场景下的系统运行成本，见表 5。

表 5 不同场景成本对比		
Table 5 Cost comparison across different scenarios		
场景	成本/万元	成本波动幅度/%
基准场景	48.3864	0
场景 2	48.9632	+1.19
场景 3	48.0974	-0.60
场景 4	49.0973	+1.47
场景 5	49.2865	+1.86

由表 5 可知，当可再生能源出力在一定范围内发生随机波动时，采用本文方法所得的多主体综合能源系统优化运行成本，其相对于基准场景的波动幅度可控制在 2%以内。可见，本文所提方法对于可再生能源随机波动具有较好的鲁棒性。

5 结论

本文深入研究了基于多智能体强化学习的多主体综合能源系统优化运行问题。算例结果表明：

1) 针对多主体含氢综合能源系统，构建了电网、天然气网与氢能网络的多能源协同运行框架，明确了综合能源服务商与负荷聚合商的功能分工，显著提升了系统经济性。

2) 多主体含氢综合能源系统能够实现电、热、氢的灵活转换，通过配备电储、热储和氢储设备，进一步增强了系统的灵活性和对波动的适应能力，提升了能源载体间的互补性和系统运行的可靠性。

3) 基于本文提出的多智能体最优折中强化学习算法的优化调度方法，通过有序决策生成和 TOPSIS 评价，实现了多主体间的高效协作和更低的运行成本，所提 MAOCPPO 算法在多主体含氢综合能源系统的整体运行成本和计算效率上均优于对比的典型算法，有效优化了能源调度与供应。

附录见本刊网络版(<http://www.dwjs.com.cn/CN/1000-3673/current.shtml>)。

参考文献

- [1] 国家发展改革委, 国家能源局. 关于加强电网调峰储能和智能化调度能力建设的指导意见[EB/OL]. (2024-01-27)[2025-05-18]. https://www.gov.cn/zhengce/zhengceku/202402/content_6934708.htm.
- [2] 张红, 袁铁江, 谭捷, 等. 面向统一能源系统的氢能规划框架[J]. 中国电机工程学报, 2022, 42(1): 83-94.
ZHANG Hong, YUAN Tiejia, TAN Jie, et al. Hydrogen energy system planning framework for unified energy system[J]. Proceedings of the CSEE, 2022, 42(1): 83-94(in Chinese).
- [3] 袁铁江, 曹继雷. 计及风电-负荷不确定性的风氢低碳能源系统容量优化配置[J]. 高电压技术, 2022, 48(6): 2037-2044.
YUAN Tiejia, CAO Jilei. Capacity optimization allocation of wind hydrogen low-carbon energy system considering wind power-load uncertainty[J]. High Voltage Engineering, 2022, 48(6): 2037-2044(in Chinese).
- [4] 韩子娇, 李正文, 张文达, 等. 计及光伏出力不确定性的氢能综合能源系统经济运行策略[J]. 电力自动化设备, 2021, 41(10): 99-106.
HAN Zijiao, LI Zhengwen, ZHANG Wenda, et al. Economic operation strategy of hydrogen integrated energy system considering uncertainty of photovoltaic output power[J]. Electric Power Automation Equipment, 2021, 41(10): 99-106(in Chinese).
- [5] 李健强, 余光正, 汤波, 等. 考虑风光利用率和含氢能流的多能流综合能源系统规划[J]. 电力系统保护与控制, 2021, 49(14): 11-20.
LI Jianqiang, YU Guangzheng, TANG Bo, et al. Multi-energy flow integrated energy system planning considering wind and solar utilization and containing hydrogen energy flow[J]. Power System Protection and Control, 2021, 49(14): 11-20(in Chinese).
- [6] 王文烨, 姜飞, 张新鹤, 等. 含规模氢能综合利用的高比例风光多能源系统低碳灵活调度[J]. 电网技术, 2024, 48(1): 197-211.
WANG Wenyue, JIANG Fei, ZHANG Xinhe, et al. Low-carbon and flexible dispatch of high-proportion wind-photovoltaic multi-energy system with comprehensive utilization of large-scale hydrogen energy[J]. Power System Technology, 2024, 48(1): 197-211(in Chinese).
- [7] WANG Yongli, MA Yuze, SONG Fuhao, et al. Economic and efficient multi-objective operation optimization of integrated energy system considering electro-thermal demand response[J]. Energy, 2020, 205: 118022.
- [8] 祝荣, 任永峰, 孟庆天, 等. 基于合作博弈的综合能源系统电-热-气协同优化运行策略[J]. 太阳能学报, 2022, 43(4): 20-29.
ZHU Rong, RENG Yongfeng, MENG Qingtian, et al. Electricity-heat-gas cooperative optimal operation strategy of integrated energy system based on cooperative game[J]. Acta Energetica Sinica, 2022, 43(4): 20-29(in Chinese).
- [9] 帅轩越, 马志程, 王秀丽, 等. 基于主从博弈理论的共享储能与综合能源微网优化运行研究[J]. 电网技术, 2023, 47(2): 679-687.
SHUAI Xuanyue, MA Zhicheng, WANG Xiuli, et al. Optimal operation of shared energy storage and integrated energy microgrid based on leader-follower game theory[J]. Power System Technology, 2023, 47(2): 679-687(in Chinese).
- [10] 曹逸滔, 王丹, 贾宏杰, 等. 考虑多能碳流约束的区域综合能源系统双层博弈扩展规划[J]. 电力系统自动化, 2023, 47(7): 12-22.
CAO Yitao, WANG Dan, JIA Hongjie, et al. Bilevel nash-stackelberg game expansion planning of regional integrated energy system considering multi-energy carbon flow constraints[J]. Automation of Electric Power Systems, 2023, 47(7): 12-22(in Chinese).
- [11] 刘俊峰, 陈剑龙, 王晓生, 等. 基于深度强化学习的微能源网能量管理与优化策略研究[J]. 电网技术, 2020, 44(10): 3794-3803.
LIU Junfeng, CHEN Jianlong, WANG Xiaosheng, et al. Energy management and optimization of multi-energy grid based on deep reinforcement learning[J]. Power System Technology, 2020, 44(10): 3794-3803(in Chinese).
- [12] NAKABI T A, TOIVANEN P. Deep reinforcement learning for energy management in a microgrid with flexible demand[J]. Sustainable Energy, Grids and Networks, 2021, 25: 100413.
- [13] ZHOU Suyang, HU Zijian, GU Wei, et al. Combined heat and power system intelligent economic dispatch: a deep reinforcement learning approach[J]. International Journal of Electrical Power & Energy Systems, 2020, 120: 106016.
- [14] 孙惠娟, 段伟男, 陈俐, 等. 基于乐观行动-评判深度强化学习的含氢综合能源系统低碳经济调度[J]. 电网技术, 2024, 48(5): 1873-1883.
SUN Huijuan, DUAN Weinan, CHEN Li, et al. Low-carbon economic scheduling of hydrogen integrated energy system based on optimistic actor-critic deep reinforcement learning[J]. Power System Technology, 2024, 48(5): 1873-1883(in Chinese).
- [15] 董雷, 刘雨, 乔骥, 等. 基于多智能体深度强化学习的电热联合系统优化运行[J]. 电网技术, 2021, 45(12): 4729-4737.
DONG Lei, LIU Yu, QIAO Ji, et al. Optimal dispatch of combined heat and power system based on multi-agent deep reinforcement learning[J]. Power System Technology, 2021, 45(12): 4729-4737(in Chinese).
- [16] 张继行, 张一, 王旭, 等. 基于多代理强化学习的多新型市场主体虚拟电厂博弈竞价及效益分配策略[J]. 电网技术, 2024, 48(5): 1980-1991.
ZHANG Jihang, ZHANG Yi, WANG Xu, et al. Game bidding and benefit allocation strategies for virtual power plants with multiple new market entities based on multi-agent reinforcement learning[J]. Power System Technology, 2024, 48(5): 1980-1991(in Chinese).
- [17] XIANG Yue, LU Yu, LIU Junyong. Deep reinforcement learning based topology-aware voltage regulation of distribution networks with distributed energy storage[J]. Applied Energy, 2023, 332: 120510.
- [18] 王丙文, 付明, 黄堃. 基于强化学习的多园区综合能源系统经济调度[J]. 电测与仪表, 2024, 61(9): 32-39.
WANG Bingwen, FU Ming, HUANG Kun. Economic dispatch of multi-area integrated energy system based on reinforcement learning[J]. Electrical Measurement & Instrumentation, 2024, 61(9): 32-39(in Chinese).
- [19] 邱威, 贺静波, 谭真, 等. 一种用于含直流馈入的受端电网故障筛选与排序方法[J]. 电网技术, 2020, 44(9): 3516-3524.
QIU Wei, HE Jingbo, TAN Zhen, et al. Contingency screening and ranking for receiving-end power grid with infeed HVDC[J]. Power System Technology, 2020, 44(9): 3516-3524(in Chinese).
- [20] 王世豪, 吕家君, 李更丰, 等. 考虑氢混燃机动态混氢特性的电气互联系统优化调度[J]. 电网技术, 2024, 48(5): 1896-1906.
WANG Shihao, LÜ Jiajun, LI Gengfeng, et al. Integrated electrical and natural gas system optimal scheduling considering dynamic hydrogen mixing characteristics of hydrogen mixed gas turbines[J]. Power System Technology, 2024, 48(5): 1896-1906(in Chinese).
- [21] 董萌苇, 朱勤婷, 徐成司, 等. 多主体参与的区域综合能源系统集中-分布式需求响应机制[J]. 电网技术, 2024, 48(6): 2336-2345.
DONG Mengwei, ZHU Mengting, XU Chengsi, et al. Centralized-distributed integrated demand response mechanism for regional integrated energy system considering multiple agents[J]. Power System Technology, 2024, 48(6): 2336-2345(in Chinese).
- [22] KIM W, PARK J, SUNG Y. Communication in multi-agent reinforcement learning: intention sharing[C]//Proceedings of the 9th International Conference on Learning Representations (ICLR 2021). Virtual Event, Austria: OpenReview. net, 2021. Available at: <https://openreview.net/forum?id=qpsl2dR9twy>.
- [23] LIU Zeyang, WAN Lipeng, SUI Xue, et al. Multi-agent intention sharing via leader-follower forest[EB/OL]. [2025-05-20]. <https://doi.org/10.48550/arXiv.2112.01078>.



彭春华

在线出版日期: 2025-09-01。

收稿日期: 2025-05-21。

作者简介:

彭春华(1973), 男, 通信作者, 博士, 教授, 博士生导师, 研究方向为电力系统优化运行、微电网优化运行与控制, E-mail: chinapch@163.com;

钟沂辰(2000), 男, 硕士研究生, 研究方向为综合能源系统优化调度, E-mail: 1745491048@qq.com;

孙惠娟(1982), 女, 教授, 研究方向为智能电网优化调度与配电网规划, E-mail: sunhuijuan1982@163.com;

张大权(1998), 男, 硕士研究生, 研究方向为综合能源系统优化调度, E-mail: 1183749349@qq.com。

(责任编辑 徐梅)

附录 A

1) CHP 机组约束:

$$\begin{cases} P_t^{\text{chp,e}} = \eta^{\text{chp}} H_t^{\text{chp}} \\ P_t^{\text{chp,t}} = \lambda^{\text{chp}} P_t^{\text{chp,e}} \\ P_{\text{chp}}^{\text{emin}} \leq P_t^{\text{chp,e}} \leq P_{\text{chp}}^{\text{emax}} \\ P_{\text{chp}}^{\text{tmin}} \leq P_t^{\text{chp,t}} \leq P_{\text{chp}}^{\text{tmax}} \\ -l_{\text{chp}}^{\text{down}} \leq P_{t+1}^{\text{chp,e}} - P_t^{\text{chp,e}} \leq l_{\text{chp}}^{\text{up}} \end{cases} \quad (\text{A1})$$

式中: $P_t^{\text{chp,e}}$ 、 $P_t^{\text{chp,t}}$ 和 H_t^{chp} 分别为 t 时段 CHP 的电出力、热出力和消耗的天然气功率; η^{chp} 为 CHP 的发电效率; λ^{chp} 为 CHP 的热电比; $P_{\text{chp}}^{\text{emin}}$ 和 $P_{\text{chp}}^{\text{emax}}$ 分别表示 CHP 的电出力下限和上限; $P_{\text{chp}}^{\text{tmin}}$ 和 $P_{\text{chp}}^{\text{tmax}}$ 分别表示 CHP 的热出力下限和上限; $l_{\text{chp}}^{\text{down}}$ 和 $l_{\text{chp}}^{\text{up}}$ 分别表示 CHP 的滑坡率和爬坡率。

2) 电锅炉 EB 设备约束:

$$\begin{cases} T_t^{\text{eb}} = \lambda^{\text{eb}} P_t^{\text{eb}} \\ T_{\text{eb}}^{\text{min}} \leq T_t^{\text{eb}} \leq T_{\text{eb}}^{\text{max}} \\ -l_{\text{eb}}^{\text{down}} \leq T_{t+1}^{\text{eb}} - T_t^{\text{eb}} \leq l_{\text{eb}}^{\text{up}} \end{cases} \quad (\text{A2})$$

式中: P_t^{eb} 和 T_t^{eb} 分别为 t 时段 EB 消耗的电功率和热出力; λ^{eb} 为 t 时段 EB 的能源转换效率; $T_{\text{eb}}^{\text{min}}$ 和 $T_{\text{eb}}^{\text{max}}$ 分别表示 EB 的热出力下限和上限; $l_{\text{eb}}^{\text{down}}$ 和 $l_{\text{eb}}^{\text{up}}$ 分别表示 EB 的滑坡率和爬坡率。

3) 电制氢 EL 设备约束:

$$\begin{cases} P_t^{\text{el,h}} = \lambda^{\text{el}} P_t^{\text{el}} \\ P_{\text{el}}^{\text{min}} \leq P_t^{\text{el,h}} \leq P_{\text{el}}^{\text{max}} \\ -l_{\text{el}}^{\text{down}} \leq P_{t+1}^{\text{el,h}} - P_t^{\text{el,h}} \leq l_{\text{el}}^{\text{up}} \end{cases} \quad (\text{A3})$$

式中: $P_t^{\text{el,h}}$ 和 P_t^{el} 分别为 t 时段 EL 氢出力和消耗的电功率; λ^{el} 为 EL 的能源转换效率; $P_{\text{el}}^{\text{min}}$ 和 $P_{\text{el}}^{\text{max}}$ 分别表示 EL 的氢出力下限和上限; $l_{\text{el}}^{\text{down}}$ 和 $l_{\text{el}}^{\text{up}}$ 分别表示 EL 的滑坡率和爬坡率。

4) 燃料电池 HFC 设备约束:

$$\begin{cases} P_t^{\text{hfc,e}} = P_t^{\text{hfc,h}} \lambda^{\text{hfc,e}} \\ T_t^{\text{hfc}} = P_t^{\text{hfc,h}} \lambda^{\text{hfc,t}} \\ \lambda^{\text{hfc,e}} + \lambda^{\text{hfc,t}} = 0.95 \end{cases} \quad (\text{A4})$$

式中: $P_t^{\text{hfc,h}}$ 、 $P_t^{\text{hfc,e}}$ 和 T_t^{hfc} 分别为 t 时段 HFC 消耗的氢功率、电出力和热出力; $\lambda^{\text{hfc,e}}$ 和 $\lambda^{\text{hfc,t}}$ 为 HFC 的电效率和热效率。

5) 可削减负荷约束:

一个优化周期内, 每个时段的可削减负荷量应不超过该时刻最大的可削减负荷量。

$$\begin{cases} 0 \leq P_t^{\text{re,e}} \leq P_t^{\text{re,emax}} \\ 0 \leq P_t^{\text{re,t}} \leq P_t^{\text{re,tmax}} \\ 0 \leq P_t^{\text{re,h}} \leq P_t^{\text{re,hmax}} \end{cases} \quad (\text{A5})$$

式中: $P_t^{\text{re,e}}$ 、 $P_t^{\text{re,t}}$ 和 $P_t^{\text{re,h}}$ 分别为 t 时段削减的电负荷总量、削减的热负荷总量和削减的氢负荷总量; $P_t^{\text{re,emax}}$ 、 $P_t^{\text{re,tmax}}$ 和 $P_t^{\text{re,hmax}}$ 分别表示 t 时段最大的可削减电负荷量、最大的可削减热负荷量和最大的可削减氢负荷量。

6) 可平移负荷约束:

一个优化周期内, 每个时段的可平移负荷量应不超过该时刻最大的可平移负荷量。且每一种负荷的平移功率之和为 0。

$$\begin{cases} 0 \leq P_t^{\text{d,e}} \leq P_t^{\text{d,emax}} \\ 0 \leq P_t^{\text{d,t}} \leq P_t^{\text{d,tmax}} \\ 0 \leq P_t^{\text{d,h}} \leq P_t^{\text{d,hmax}} \\ \sum_{t=1}^T P_t^{\text{d,e}} = 0 \\ \sum_{t=1}^T P_t^{\text{d,t}} = 0 \\ \sum_{t=1}^T P_t^{\text{d,h}} = 0 \end{cases} \quad (\text{A6})$$

式中: $P_t^{\text{d,e}}$ 、 $P_t^{\text{d,t}}$ 和 $P_t^{\text{d,h}}$ 分别为 t 时段平移的电负荷、热负荷、氢负荷总量; $P_t^{\text{d,emax}}$ 、 $P_t^{\text{d,tmax}}$ 和 $P_t^{\text{d,hmax}}$ 分别表示 t 时段最大的可平移电负荷量、最大的可平移热负荷量和最大的可平移氢负荷量。

7) 电功率平衡约束:

$$P_t^{\text{grid}} + P_t^{\text{new}} + P_t^{\text{hiep,e}} = P_t^{\text{eload}} + P_t^{\text{d,e}} - P_t^{\text{re,e}} \quad (\text{A7})$$

$$P_t^{\text{new}} = P_t^{\text{pv}} + P_t^{\text{wt}} \quad (\text{A8})$$

$$P_t^{\text{hiep,e}} = P_t^{\text{chp,e}} + P_t^{\text{hfc,e}} + P_t^{\text{es}} + P_t^{\text{el}} + P_t^{\text{eb}} \quad (\text{A9})$$

式中: P_t^{grid} 表示 t 时段与上级电网交互电量; P_t^{new} 表示 t 时段新能源电出力总和; P_t^{pv} 和 P_t^{wt} 为 t 时段光伏和风电出力; $P_t^{\text{hiep,e}}$ 表示 t 时段 HIEP 内设备电出力; P_t^{es} 表示电储电出力; P_t^{eload} 表示 t 时段总电负荷功率。

8) 热功率平衡约束:

$$P_t^{\text{chp,t}} + T_t^{\text{eb}} + T_t^{\text{hfc}} + P_t^{\text{ts}} = P_t^{\text{tload}} + P_t^{\text{d,t}} - P_t^{\text{re,t}} \quad (\text{A10})$$

式中: P_t^{ts} 表示 t 时段热储热出力; P_t^{tload} 表示 t 时段总热负荷功率。

9) 氢功率平衡约束:

$$P_t^{\text{el,h}} + P_t^{\text{hfc,h}} + P_t^{\text{hs}} + P_t^{\text{phs}} = P_t^{\text{hload}} + P_t^{\text{d,h}} - P_t^{\text{re,h}} \quad (\text{A11})$$

式中: P_t^{hs} 表示 t 时段氢储氢出力; P_t^{phs} 表示 t 时段 HIEP 氢储出力; P_t^{hload} 表示 t 时段总氢负荷功率。

10) 储能约束:

电储约束包括电储出力约束和电储荷电状态约束:

$$\begin{cases} P_t^{\text{es},\min} \leq P_t^{\text{es}} \leq P_t^{\text{es},\max} \\ S_t^{\text{es},\min} \leq S_t^{\text{es}} \leq S_t^{\text{es},\max} \end{cases} \quad (\text{A12})$$

式中: P_t^{es} 和 S_t^{es} 分别表示 t 时段电储出力和电储荷电状态; $P_t^{\text{es},\min}$ 和 $P_t^{\text{es},\max}$ 为电储出力的上下限; $S_t^{\text{es},\min}$ 与 $S_t^{\text{es},\max}$ 为电储荷电状态的上下限。

热储约束包括热储出力约束和热储容量约束:

$$\begin{cases} P_t^{\text{ts},\min} \leq P_t^{\text{ts}} \leq P_t^{\text{ts},\max} \\ S_t^{\text{ts},\min} \leq S_t^{\text{ts}} \leq S_t^{\text{ts},\max} \end{cases} \quad (\text{A13})$$

式中: P_t^{ts} 和 S_t^{ts} 分别表示 t 时段热储出力和热储容量; $P_t^{\text{ts},\min}$ 和 $P_t^{\text{ts},\max}$ 为热储出力的上下限; $S_t^{\text{ts},\min}$ 与 $S_t^{\text{ts},\max}$ 为热储容量的上下限。

氢储约束包括氢储出力约束和氢储容量约束:

$$\begin{cases} P_t^{\text{hs},\min} \leq P_t^{\text{hs}} \leq P_t^{\text{hs},\max} \\ S_t^{\text{hs},\min} \leq S_t^{\text{hs}} \leq S_t^{\text{hs},\max} \\ P_t^{\text{phs},\min} \leq P_t^{\text{phs}} \leq P_t^{\text{phs},\max} \\ S_t^{\text{phs},\min} \leq S_t^{\text{phs}} \leq S_t^{\text{phs},\max} \end{cases} \quad (\text{A14})$$

式中: P_t^{hs} 和 S_t^{hs} 分别表示 t 时段 HLA 氢储出力和氢储容量; $P_t^{\text{hs},\min}$ 和 $P_t^{\text{hs},\max}$ 为 HLA 氢储出力的上下限; $S_t^{\text{hs},\min}$ 与 $S_t^{\text{hs},\max}$ 为 HLA 氢储容量的上下限。 P_t^{phs} 和 S_t^{phs} 分别表示 t 时段 HIEP 氢储出力和氢储容量; $P_t^{\text{phs},\min}$ 和 $P_t^{\text{phs},\max}$ 为 HIEP 氢储出力的上下限; $S_t^{\text{phs},\min}$ 与 $S_t^{\text{phs},\max}$ 为 HIEP 氢储容量的上下限。

附录 B

$$\begin{cases} S_t^{\text{HIEP}} = [c_t^{\text{IEP}}, P_t^{\text{IEP}}, S_t^{\text{phs}}] \\ c_t^{\text{IEP}} = [c_t^{\text{IESe}}, c_t^{\text{buye}}, c_t^{\text{selc}}] \\ P_t^{\text{IEP}} = [P_t^{\text{eload}}, P_t^{\text{tload}}, P_t^{\text{hload}}, P_t^{\text{wt}}, P_t^{\text{pv}}, \\ T_t^{\text{hfc}}, P_t^{\text{chp},t}, P_t^{\text{re},e}, P_t^{\text{re},t}, P_t^{\text{re},h}, \\ P_t^{\text{d},e}, P_t^{\text{d},t}, P_t^{\text{d},h}] \end{cases} \quad (\text{B1})$$

$$S_t^{\text{ELA}} = [c_t^{\text{IESe}}, c_t^{\text{IESe}}, P_t^{\text{wt}}, P_t^{\text{pv}}, P_t^{\text{el}}, P_t^{\text{eb}}, P_t^{\text{chp},e}, P_t^{\text{hfc},e}, P_t^{\text{tload}}, S_t^{\text{es}}] \quad (\text{B2})$$

$$S_t^{\text{TLA}} = [c_t^{\text{IESt}}, P_t^{\text{hfc},t}, T_t^{\text{chp}}, T_t^{\text{eb}}, P_t^{\text{tload}}, S_t^{\text{ts}}] \quad (\text{B3})$$

$$S_t^{\text{HLA}} = [c_t^{\text{IESh}}, P_t^{\text{hfc},h}, P_t^{\text{el},h}, P_t^{\text{hload}}, S_t^{\text{hs}}] \quad (\text{B4})$$

式中: S_t^{HIEP} 、 S_t^{ELA} 、 S_t^{TLA} 和 S_t^{HLA} 分别表示 HIEP-Agent、ELA-Agent、TLA-Agent 和 HLA-Agent 的状态空间合集。

附录 C

$$\begin{cases} A_{\text{HIEP}} = \{T_{\text{CHP}}(t), T_{\text{HFC}}(t), H_{\text{ES}}^p(t)\} \\ A_{\text{ELA}} = \{L_{\text{ec}}(t), D_{\text{ec}}(t), P_{\text{ES}}(t)\} \\ A_{\text{TLA}} = \{L_{\text{tc}}(t), D_{\text{tc}}(t), T_{\text{ES}}(t)\} \\ A_{\text{HLA}} = \{L_{\text{hc}}(t), D_{\text{hc}}(t), H_{\text{ES}}(t)\} \end{cases} \quad (\text{C1})$$

式中: A_{HIEP} 、 A_{ELA} 、 A_{TLA} 和 A_{HLA} 分别表示含 HIEP、ELA、TLA 和 HLA 的动作空间; $T_{\text{CHP}}(t)$ 和 $T_{\text{HFC}}(t)$ 为 t 时段 CHP 机组热出力和燃料电池 HFC 热出力; $H_{\text{ES}}^p(t)$ 为 HIEP 内氢储能出力; $L_{\text{ec}}(t)$ 、 $D_{\text{ec}}(t)$ 和 $P_{\text{ES}}(t)$ 为 t 时段需求响应削减的电负荷、平移的电负荷和电储能出力; $L_{\text{tc}}(t)$ 、 $D_{\text{tc}}(t)$ 和 $T_{\text{ES}}(t)$ 为 t 时段需求响应削减的热负荷、平移的热负荷和热储能出力; $L_{\text{hc}}(t)$ 、 $D_{\text{hc}}(t)$ 和 $H_{\text{ES}}(t)$ 为 t 时段需求响应削减的氢负荷、平移的氢负荷和氢储能出力。

附录 D

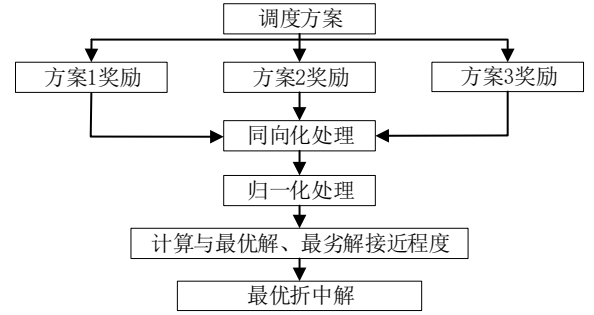


图 D1 TOPSIS 评价流程示意图

Fig. D1 Schematic diagram of TOPSIS evaluation process

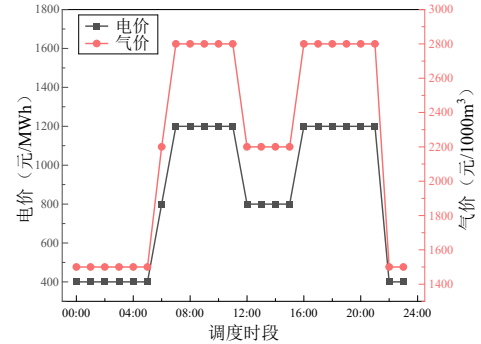


图 D2 分时价格曲线

Fig. D2 Time-share Curve

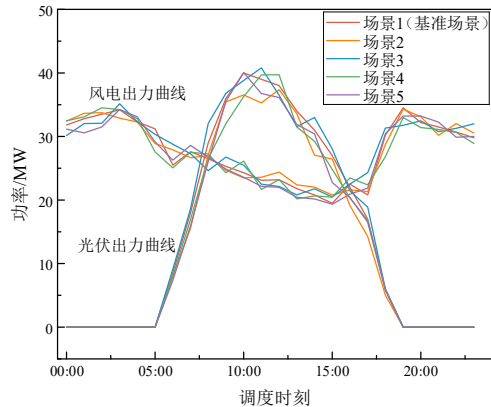


图 D3 风电和光伏出力曲线

Fig. D3 WT&PV generation curve

表 D1 MAOCPPO 训练过程
Table D1 MAOCPPO training process

输入: 智能体数量 n 、排列方式 $P = \{p_1, \dots, p_{n!}\}$ 、策略网络、值函数网络、超参数等

1 初始化策略网络参数 $\theta_1, \dots, \theta_n$ 和值函数网络参数 $\phi_1, \dots, \phi_n$ 、每个智能体初始化优化器、每个排列 $p \in P$ 初始化经验缓冲区 D^p

2 while $p \leq p_{\max}$ do

//并行收集所有排列的经验轨迹

for 每种排列 $p \in P$ do(并行执行)

重置环境, 得到初始状态 S_0^p 。

初始化周期计数器 $t=0$

while $t \leq T$ do

//按排列顺序生成动作

联合动作 $a^p = []$

for 智能体 i 按排列 p 的顺序 do

观察输入 $o_i^p = \text{concat}(s_t^p, a_{\text{previous}}^p)$ 包含前面智能体的动作

生成动作 $a_i^p \sim \pi_i(o_i^p)$

将 a_i^p 添加到 a^p

执行联合动作 a^p 得到下一状态 s_{t+1}^p 奖励

$r^p = (r_1^p, r_2^p, \dots, r_n^p)$

存储 $(s_t^p, a^p, r^p, s_{t+1}^p)$ 到 D^p

$t \leftarrow t+1$

//累积奖励并应用 TOPSIS 选择最优 D

计算每个 D^p 的累积奖励 R^p

使用 TOPSIS 选择得分最高的 D^{best}

//使用最优 D 更新策略和值函数

for 每个智能体 i do

从 D^{best} 中提取轨迹数据

计算优势和目标回报(同原始 MAPPO)

更新策略和值函数(同原始 MAPPO 的更新步骤)

end while

表 D2 HIES 设备参数

Table D2 Equipment parameters in HIES

参数	数值	参数	数值
μ^{chp}	0.4	$\lambda^{\text{ts,c}}$	0.9
λ^{chp}	1.125	$\lambda^{\text{ts,f}}$	0.95
λ^{eb}	0.95	$\lambda^{\text{hs,c}}$	0.9
λ^{el}	0.8	$\lambda^{\text{hs,f}}$	0.95
$\lambda^{\text{hfc,e}}$	0.5	S^{min}	0.1
$\lambda^{\text{hfc,t}}$	0.4	S^{max}	0.9
$\lambda^{\text{es,c}}$	0.9	I^{up}	0.6
$\lambda^{\text{es,f}}$	0.95	I^{down}	0.5

表 D3 HIES 运行成本参数

Table D3 Operating cost parameters in HIES

设备	成本参数	设备	成本参数
CHP	0.028(元/kW·h)	EL	0.013(元/kW·h)
HFC	0.015(元/kW·h)	EB	0.012(元/kW·h)
电、热储能	0.01(元/kW·h)	风电	0.06(元/kW·h)
氢储能	2.5(元/kg)	光伏	0.03(元/kW·h)
电负荷削减	0.6(元/kW·h)	电负荷平移	0.2(元/kW·h)
热负荷削减	0.7(元/kW·h)	热负荷平移	0.25(元/kW·h)
氢负荷削减	5(元/kg)	氢负荷平移	2(元/kg)

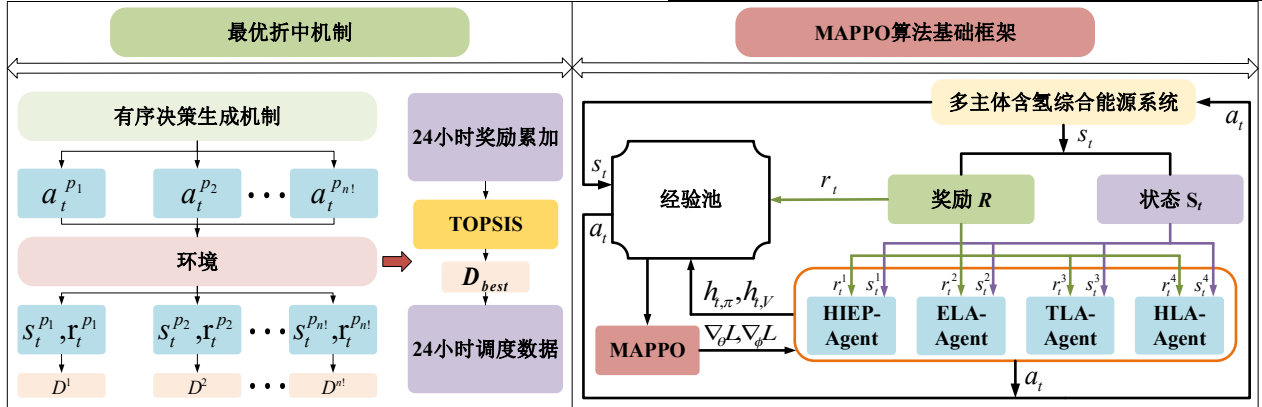


图 D4 MAOCPPO 强化学习框架

Fig. D4 MAOCPPO reinforcement learning framework