

DOI: 10.12158/j.2096-3203.2024.06.009

基于 PPO 算法的 CIES 低碳优化调度方法

陈凡, 吴凌霄, 王曼, 吕千云, 张小莲
(南京工程学院电力工程学院, 江苏 南京 211167)

摘要: 阶梯式碳交易机制以及优化调度模型求解算法是进行园区综合能源系统 (community integrated energy system, CIES) 优化调度的重要因素, 现有文献对这两个因素的考虑不够全面。为此, 文中在考虑阶梯式碳交易机制的基础上, 提出采用近端策略优化 (proximal policy optimization, PPO) 算法求解 CIES 低碳优化调度问题。该方法基于低碳优化调度模型搭建强化学习交互环境, 利用设备状态参数及运行参数定义智能体的状态、动作空间及奖励函数, 再通过离线训练获取可生成最优策略的智能体。算例分析结果表明, 采用 PPO 算法得到的 CIES 低碳优化调度方法能够充分发挥阶梯式碳交易机制减少碳排放量和提高能源利用率方面的优势。

关键词: 园区综合能源系统 (CIES); 优化调度; 近端策略优化 (PPO) 算法; 阶梯式碳交易机制; 惩罚系数; 碳排放
中图分类号: TM73 **文献标志码:** A **文章编号:** 2096-3203 (2024) 06-0088-12

0 引言

综合能源系统 (integrated energy system, IES) 是减少碳排放量、实现双碳目标的重要途径^[1-6]。园区综合能源系统 (community integrated energy system, CIES) 作为 IES 的基本单元, 将新能源、储能以及传统能源等融合互补, 为提高能源利用率和降低碳排放量提供了一种新的解决方案^[7-8]。

近年来, 高比例新能源的接入加剧了 CIES 的不确定性, 给系统的优化调度带来了挑战。文献[9-10]通过预测获取源荷数据, 将含有不确定性新能源出力数据的模型求解问题转变为确定性问题, 然后采用 CPLEX 求解器进行求解, 但受预测模型精度影响, 求解效果不佳。为解决上述问题, 文献[11]通过提升预测精度来改善优化效果; 而文献[12]则从算法入手, 采用粒子群优化 (particle swarm optimization, PSO) 算法对低碳调度模型进行求解, 通过启发式算法的可优化性、鲁棒性和适用性来获得更优的调度策略; 文献[13]在文献[12]的基础上进一步提出改进 PSO 算法求解优化调度模型。然而, 上述优化方法依赖预测模型获取风光荷数据, 风光荷预测误差将会影响 CIES 优化结果。

以马尔可夫决策理论为基础的强化学习方法可在线获取真实风光荷数据并进行优化, 避免了由预测模型误差产生的影响^[14-15]。文献[16]采用 Q 学习算法求解 CIES 低碳优化模型, 解决了预测模型误差的问题。相比数学规划方法和 PSO 算法的固定策略, Q 学习算法能根据真实源荷信息实时生成

自适应策略, 但面对高维连续问题时会出现维数灾难, 其离散化的状态空间也可能引起信息损失, 影响环境建模, 从而导致训练学习效果减弱。

为解决上述强化学习方法状态空间离散化的问题, 部分学者提出将深度神经网络与强化学习相结合, 将深度神经网络作为函数逼近器, 直接在连续空间中学习到一个策略或价值函数, 进而从原始输入中直接提取特征。文献[17-18]采用深度 Q 网络 (deep Q network, DQN) 算法求解电力系统运行优化问题, 通过深度神经网络避免了状态空间离散化, 进一步提升了优化调度效果。然而, DQN 无法处理连续动作, 必须将动作离散化才能进行交互训练, 导致动作空间中的部分动作无法被选择, 从而无法寻得最优策略。此外, DQN 使用经验回放提高样本效率, 但在部分可观察的环境下处理效果可能较差。为解决上述问题, 文献[19]采用深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 算法求解 IES 优化调度问题。该算法采用确定性策略梯度方法, 在给定的状态下直接输出确定的动作, 因此可直接适应连续的动作空间, 更贴近电热设备连续调节的动态特性, 且可适用于如部分可观测马尔可夫决策过程 (partially observable Markov decision processes, POMDPs) 这类环境, 但也可能会陷入局部最优, 难以找到全局最优策略。为此, 有学者提出使用随机策略的优势动作评论 (advantage actor critic, A2C) 算法^[20]求解优化调度问题, 但由于样本估计的方差和超参选择的敏感等原因, 无法确保训练优化模型时策略的单调提升, 容易出现不稳定的策略更新^[21]。近端策略优化 (proximal policy optimization, PPO) 算法引入重要性采样技术和动态

收稿日期: 2024-04-11; 修回日期: 2024-07-05

基金项目: 国家自然科学基金资助项目 (51977103)

步长剪切机制,降低了步长敏感程度,确保了模型训练时策略性能的单调提升;同时降低了策略梯度算法中的方差,从而提高了学习的稳定性,可以有效处理强不确定性下的 CIES 低碳优化调度问题^[22-23]。

阶梯式碳交易机制有利于减少碳排放量、提高能源效率^[9,24-25]。文献[26-28]研究阶梯式碳交易机制对 IES 碳排放总量的影响,但均采用传统优化方法进行模型求解,风光荷预测误差会使其无法获得最优解。文献[22,29]采用 PPO 算法求解 CIES 优化调度问题,但其优化模型未考虑碳交易成本或采用传统碳交易机制,不利于降低系统碳排放量。

综上所述,现有文献在进行 CIES 优化调度研究时,未能综合考虑阶梯式碳交易机制以及优化模型求解算法对调度结果的影响。因此,文中在考虑阶梯式碳交易机制的基础上,采用 PPO 算法对 CIES 低碳优化调度问题进行求解。以某园区的 IES 为例进行分析,算例结果验证了采用阶梯式碳交易机制的必要性和采用 PPO 算法求解 CIES 低碳优化调度问题的先进性。

1 CIES 低碳优化调度模型

1.1 优化调度框架

文中建立的 CIES 低碳优化调度模型主要包含可再生能源系统、传统机组、储能系统以及负荷。其中,可再生能源系统包括风电机组和光伏机组;传统机组包括热电联产(combined heat and power, CHP)机组和燃气锅炉(gas boiler, GB),CHP 机组由燃气轮机燃烧天然气,同时输出电能和热能,不仅可以满足电负荷的需求,还可以帮助 GB 应对热负荷的变化,而 GB 燃烧天然气产生热能,用于满足热负荷需求;储能系统包括电储能和热储能,通过储能系统的灵活运行可实现削峰填谷,满足电热平衡需求,提高可再生能源利用率和降低碳排放量。

1.1.1 目标函数

文中综合考虑 CIES 的购能成本 $f_{\text{buy}}^{\text{price}}$ 、碳交易成本 $f_{\text{CO}_2}^{\text{price}}$ 、弃风光成本 $f_{\text{DG,cut}}^{\text{price}}$ 和储能运行成本 $f_{\text{S}}^{\text{price}}$,构建总成本 F 最小的低碳经济调度目标为:

$$F = \min(f_{\text{buy}}^{\text{price}} + f_{\text{CO}_2}^{\text{price}} + f_{\text{DG,cut}}^{\text{price}} + f_{\text{S}}^{\text{price}}) \quad (1)$$

(1) 购能成本 $f_{\text{buy}}^{\text{price}}$ 。

$$P_{\text{buy,g}}(t) = P_{\text{CHP,g}}(t) + P_{\text{GB,g}}(t) \quad (2)$$

$$f_{\text{buy}}^{\text{price}} = \sum_{t=1}^T \alpha_t P_{\text{buy,e}}(t) + \sum_{t=1}^T \beta_t P_{\text{buy,g}}(t) \quad (3)$$

式中: $P_{\text{CHP,g}}(t)$ 、 $P_{\text{GB,g}}(t)$ 分别为 t 时刻 CHP 机组和 GB 的购气量; $P_{\text{buy,e}}(t)$ 、 $P_{\text{buy,g}}(t)$ 分别为 t 时刻

的购电功率和购气量; α_t 、 β_t 分别为 t 时刻的电价、气价; T 为调度周期,一般取 24 h。

(2) 弃风光成本 $f_{\text{DG,cut}}^{\text{price}}$ 。

$$f_{\text{WTG,cut}}^{\text{price}} = \delta_{\text{WTG}} \sum_{t=1}^T P_{\text{WTG,cut}}(t) \quad (4)$$

$$f_{\text{PV,cut}}^{\text{price}} = \delta_{\text{PV}} \sum_{t=1}^T P_{\text{PV,cut}}(t) \quad (5)$$

$$f_{\text{DG,cut}}^{\text{price}} = f_{\text{WTG,cut}}^{\text{price}} + f_{\text{PV,cut}}^{\text{price}} \quad (6)$$

式中: $f_{\text{WTG,cut}}^{\text{price}}$ 、 $f_{\text{PV,cut}}^{\text{price}}$ 分别为弃风、弃光费用; δ_{WTG} 、 δ_{PV} 分别为单位弃风、弃光惩罚系数; $P_{\text{WTG,cut}}(t)$ 、 $P_{\text{PV,cut}}(t)$ 分别为 t 时刻弃风、弃光功率。

(3) 储能运行成本 $f_{\text{S}}^{\text{price}}$ 。

$$f_{\text{S}}^{\text{price}} = \sum_{t=1}^T \sum_{n=1}^2 \gamma_n |P_{\text{ES,n}}(t)| \quad (7)$$

式中: n 为储能类型, $n \in \{1,2\}$,1和2分别表示电储能和热储能; γ_n 为第 n 种储能装置单位功率维护成本; $P_{\text{ES,n}}(t)$ 为 t 时刻第 n 种储能装置的充/放电功率,其中 $P_{\text{ES,n}}(t) > 0$ 表示储能装置为充电状态, $P_{\text{ES,n}}(t) \leq 0$ 表示储能装置为放电状态。

1.1.2 约束条件

(1) 风电出力约束。

$$0 \leq P_{\text{WTG}}(t) \leq P_{\text{WTG}}^{\text{max}}(t) \quad (8)$$

式中: $P_{\text{WTG}}(t)$ 为 t 时刻风电机组输出功率; $P_{\text{WTG}}^{\text{max}}(t)$ 为 t 时刻风电机组所能提供的最大输出功率。

(2) 光伏出力约束。

$$0 \leq P_{\text{PV}}(t) \leq P_{\text{PV}}^{\text{max}}(t) \quad (9)$$

式中: $P_{\text{PV}}(t)$ 为 t 时刻光伏机组输出功率; $P_{\text{PV}}^{\text{max}}(t)$ 为 t 时刻光伏机组所能提供的最大输出功率。

(3) 储能运行约束。电热储能模型相似^[5],可以统一表示为:

$$\begin{cases} -P_{\text{ES,n}}^{\text{max}} \leq P_{\text{ES,n}}(t) \leq P_{\text{ES,n}}^{\text{max}} \\ S_n^{\text{min}} \leq S_n(t) \leq S_n^{\text{max}} \\ S_n(1) = S_n(T) \\ P_{\text{ES,n}}^{\text{cha}}(t) = P_{\text{ES,n}}(t) & P_{\text{ES,n}}(t) > 0 \\ P_{\text{ES,n}}^{\text{dis}}(t) = P_{\text{ES,n}}(t) & P_{\text{ES,n}}(t) \leq 0 \\ S_n(t) = S_n(t-1) + P_{\text{ES,n}}^{\text{cha}}(t)\theta_{\text{ES,n}}^{\text{cha}} & P_{\text{ES,n}}(t) > 0 \\ S_n(t) = S_n(t-1) + P_{\text{ES,n}}^{\text{dis}}(t)/\theta_{\text{ES,n}}^{\text{dis}} & P_{\text{ES,n}}(t) \leq 0 \end{cases} \quad (10)$$

式中: $P_{\text{ES,n}}^{\text{max}}$ 为第 n 种储能装置单次充/放电最大功率; $P_{\text{ES,n}}^{\text{cha}}(t)$ 、 $P_{\text{ES,n}}^{\text{dis}}(t)$ 分别为 t 时刻第 n 种储能装置的充、放电功率; $\theta_{\text{ES,n}}^{\text{cha}}$ 、 $\theta_{\text{ES,n}}^{\text{dis}}$ 分别为第 n 种储能装置的充、放电效率; $S_n(t)$ 为 t 时刻第 n 种储能装置的储能量; S_n^{max} 、 S_n^{min} 分别为第 n 种储能装置储能量的上、下限。其中, $S_n(1) = S_n(T)$ 约束储能装置在

一个调度周期的始末储能状态一致,为下一日的调度做准备。

(4) CHP 机组运行约束。

$$\begin{cases} P_{\text{CHP},e}(t) = \theta_{\text{CHP},e} P_{\text{CHP},g}(t) \\ P_{\text{CHP},h}(t) = \theta_{\text{CHP},h} P_{\text{CHP},g}(t) \\ P_{\text{CHP},g}^{\min} \leq P_{\text{CHP},g}(t) \leq P_{\text{CHP},g}^{\max} \\ \Delta P_{\text{CHP},g}^{\min} \leq P_{\text{CHP},g}(t+1) - P_{\text{CHP},g}(t) \leq \Delta P_{\text{CHP},g}^{\max} \end{cases} \quad (11)$$

式中: $P_{\text{CHP},e}(t)$ 、 $P_{\text{CHP},h}(t)$ 分别为 t 时刻 CHP 机组输出的电、热功率; $\theta_{\text{CHP},e}$ 、 $\theta_{\text{CHP},h}$ 分别为 CHP 转换为电、热能的效率; $P_{\text{CHP},g}^{\max}$ 、 $P_{\text{CHP},g}^{\min}$ 分别为输入 CHP 机组的天然气功率上、下限; $\Delta P_{\text{CHP},g}^{\max}$ 、 $\Delta P_{\text{CHP},g}^{\min}$ 分别为 CHP 机组的爬坡上、下限。

(5) 电功率平衡约束^[30]。由于风光出力具有较大的随机性和波动性,为减轻主网调节压力^[9],文中不考虑 CIES 向上级电网售电。

$$\begin{cases} P_{\text{buy},e}(t) = P_{\text{Load},e}(t) + P_{\text{ES},1}(t) - \\ P_{\text{WTC}}(t) - P_{\text{PV}}(t) - P_{\text{CHP},e}(t) \\ 0 \leq P_{\text{buy},e}(t) \leq P_{\text{buy},e}^{\max} \end{cases} \quad (12)$$

式中: $P_{\text{Load},e}(t)$ 为 t 时刻电负荷; $P_{\text{ES},1}(t)$ 为 t 时刻输入电储能的功率; $P_{\text{buy},e}^{\max}$ 为所有时刻购电限值。

(6) 热功率平衡约束。

$$\begin{cases} P_{\text{GB},h}(t) = P_{\text{Load},h}(t) + P_{\text{ES},2}(t) - P_{\text{CHP},h}(t) \\ P_{\text{GB},g}^{\min} \leq P_{\text{GB},g}(t) \leq P_{\text{GB},g}^{\max} \\ P_{\text{GB},h}(t) = \theta_{\text{GB}} P_{\text{GB},g}(t) \\ \Delta P_{\text{GB},g}^{\min} \leq P_{\text{GB},g}(t+1) - P_{\text{GB},g}(t) \leq \Delta P_{\text{GB},g}^{\max} \end{cases} \quad (13)$$

式中: $P_{\text{GB},h}(t)$ 为 t 时刻 GB 输出的热功率; $P_{\text{Load},h}(t)$ 为 t 时刻热负荷; $P_{\text{ES},2}(t)$ 为 t 时刻输入热储能的功率; $P_{\text{GB},g}^{\max}$ 、 $P_{\text{GB},g}^{\min}$ 分别为输入 GB 的天然气功率上、下限; θ_{GB} 为 GB 的能量转换效率; $\Delta P_{\text{GB},g}^{\max}$ 、 $\Delta P_{\text{GB},g}^{\min}$ 分别为 GB 的爬坡上、下限。

1.2 阶梯式碳交易模型

采用碳交易机制的目的在于通过设立合理的碳排放权,使得生产商能够积极参与市场上的碳排放权交易,从而实现对整体碳排放量的有效掌控。监管部门首先会为每个碳排放源分发碳排放权配额,生产商则依据自身配额进行合理的生产和排放活动。当实际碳排放量低于分配的配额时,生产商有权将多余的碳排放配额投入碳交易市场进行交易;反之,则需要购买碳排放权配额。阶梯式碳交易机制模型主要涵盖碳排放权配额模型、实际碳排放模型以及阶梯式碳排放交易模型^[9]。

(1) 碳排放权配额模型。CIES 中的碳排放源

主要有 3 类:上级购电、GB 和 CHP 机组。目前国内主要采用的配额方法为无偿配额^[9],为方便计算,可认为上级购电均来源于燃煤机组发电。

$$\begin{cases} E_{\text{CIES}} = E_{\text{buy},e} + E_{\text{CHP},g} + E_{\text{GB},g} \\ E_{\text{buy},e} = \chi_{\text{buy}} \sum_{t=1}^T P_{\text{buy},e}(t) \\ E_{\text{CHP},g} = \chi_{\text{CHP}} \sum_{t=1}^T P_{\text{CHP},g}(t) \\ E_{\text{GB},g} = \chi_{\text{GB}} \sum_{t=1}^T P_{\text{GB},g}(t) \end{cases} \quad (14)$$

式中: E_{CIES} 、 $E_{\text{buy},e}$ 、 $E_{\text{CHP},g}$ 、 $E_{\text{GB},g}$ 分别为 CIES、上级购电、CHP 机组、GB 的碳排放权配额; χ_{buy} 、 χ_{CHP} 、 χ_{GB} 分别为燃煤机组单位电力消耗和 CHP 机组、GB 单位天然气消耗的碳排放权配额。

(2) 实际碳排放模型。

$$\begin{cases} E_{\text{CIES},a} = E_{\text{buy},e,a} + E_{\text{CHP},g,a} + E_{\text{GB},g,a} \\ E_{\text{buy},e,a} = \chi_{\text{buy},a} \sum_{t=1}^T P_{\text{buy},e}(t) \\ E_{\text{CHP},g,a} = \chi_{\text{CHP},a} \sum_{t=1}^T P_{\text{CHP},g}(t) \\ E_{\text{GB},g,a} = \chi_{\text{GB},a} \sum_{t=1}^T P_{\text{GB},g}(t) \end{cases} \quad (15)$$

式中: $E_{\text{CIES},a}$ 、 $E_{\text{buy},e,a}$ 、 $E_{\text{CHP},g,a}$ 、 $E_{\text{GB},g,a}$ 分别为 CIES、上级购电、CHP 机组、GB 的实际碳排放量; $\chi_{\text{buy},a}$ 、 $\chi_{\text{CHP},a}$ 、 $\chi_{\text{GB},a}$ 分别为燃煤机组单位电力消耗和 CHP 机组、GB 单位天然气消耗的实际碳排放。

(3) 阶梯式碳排放交易模型。通过计算 CIES 碳排放权配额及实际碳排放量可求得实际参与碳交易市场的碳排放权交易额。

$$E_{\text{CIES},\text{tr}} = E_{\text{CIES},a} - E_{\text{CIES}} \quad (16)$$

式中: $E_{\text{CIES},\text{tr}}$ 为 CIES 的碳排放权交易额。

阶梯式定价机制划分了多个购买区间,需要购买的碳排放权配额越多,相应区间的购价越高。阶梯式碳交易成本 $f_{\text{CO}_2}^{\text{price}}$ 为:

$$f_{\text{CO}_2}^{\text{price}} = \begin{cases} \lambda E_{\text{CIES},\text{tr}} & E_{\text{CIES},\text{tr}} \leq l \\ \lambda(1+a)(E_{\text{CIES},\text{tr}} - l) + \lambda l & l < E_{\text{CIES},\text{tr}} \leq 2l \\ \lambda(1+2a)(E_{\text{CIES},\text{tr}} - 2l) + \lambda(2+a)l & 2l < E_{\text{CIES},\text{tr}} \leq 3l \\ \lambda(1+3a)(E_{\text{CIES},\text{tr}} - 3l) + \lambda(3+3a)l & 3l < E_{\text{CIES},\text{tr}} \leq 4l \\ \lambda(1+4a)(E_{\text{CIES},\text{tr}} - 4l) + \lambda(4+6a)l & E_{\text{CIES},\text{tr}} > 4l \end{cases} \quad (17)$$

式中: λ 为碳交易基价; l 为碳排放量区间长度; a 为价格增长率。

2 基于PPO算法的CIES低碳优化调度框架

PPO算法是一种强化学习方法,其本质是通过基于深度神经网络生成的智能体与环境不断交互来学习和探索,以最大化回报或实现特定目标的马尔可夫决策过程。其中,用于训练的交互环境是由动作空间、状态空间、状态转移概率和奖励函数构成的元组^[29]。图1为基于CIES低碳优化调度模型的强化学习环境与智能体的交互过程。图1中,交互环境以CIES各电热设备为基础进行构建,在训练学习时为智能体提供设备运行状态及动作反馈奖励,使智能体动作概率分布不断更新,引导其学习最佳策略从而获得CIES最优调度方案。

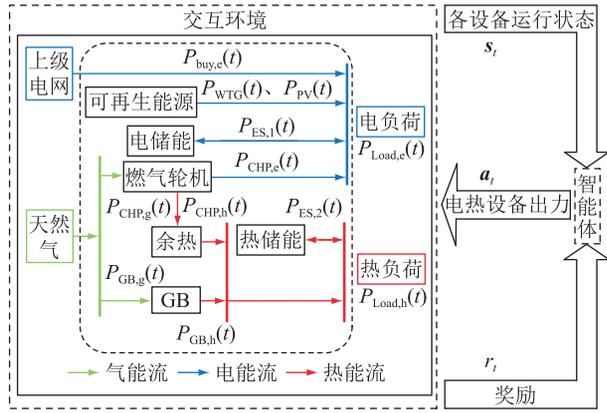


图1 CIES低碳优化调度框架

Fig.1 CIES low-carbon optimized scheduling framework

2.1 动作空间描述

动作空间设计为优化调度问题的决策变量,即图1环境模型中各设备出力。

智能体动作 a_t 为各电热设备出力:

$$a_t = [P_{\text{CHP,g}}(t) \ P_{\text{ES,1}}(t) \ P_{\text{ES,2}}(t) \ P_{\text{WTG}}(t) \ P_{\text{PV}}(t)] \quad (18)$$

其中,各设备存在出力约束,如式(8)一式(13)所示。动作 a_t 可由神经网络输出的向量乘以 CHP 机组、电储能、热储能以及风光出力的范围得到,其中神经网络生成的向量的元素范围为 $0 \sim 1$ 。

2.2 状态空间描述

智能体根据观察到的系统状态做出动作,文中设计的状态向量 s_t 为各设备运行状态及电、热负荷大小,包括 CIES 在 t 时刻由风速确定的风电最大输出功率、由辐照强度确定的光伏最大输出功率;同时为保证燃气轮机和 GB 输出功率满足爬坡功率约束,将其 $t-1$ 时刻的输出功率作为 t 时刻状态空间的一部分:

$$s_t = [P_{\text{WTG}}^{\text{max}}(t) \ P_{\text{PV}}^{\text{max}}(t) \ P_{\text{Load,e}}(t) \ P_{\text{Load,h}}(t) \ S_1(t) \ S_2(t) \ P_{\text{CHP,g}}(t-1) \ P_{\text{GB,g}}(t-1)] \quad (19)$$

2.3 状态转移概率

状态转移概率 G 描述了将动作 a_t 应用到环境状态 s_t 后,环境从当前状态 s_t 转移至下一状态 s_{t+1} 的概率:

$$G = P\{s_{t+1} = s' \mid s_t = s, a_t = a\} \quad (20)$$

式中: $P\{\cdot\}$ 为状态转移概率的集合; s, a 为当前状态的; s' 为下一状态的值。

图2为状态转移过程,即动作 a_t 输入到环境后,环境状态 s_t 遵循状态转移概率转移至下一环境状态 s_{t+1} 的过程。图2中,将状态空间中 $S_1(t)$ 、 $S_2(t)$ 、 $P_{\text{CHP,g}}(t-1)$ 和 $P_{\text{GB,g}}(t-1)$ 遵循各模型的约束条件进行计算,状态转移概率为1;而 $P_{\text{WTG}}^{\text{max}}(t)$ 、 $P_{\text{PV}}^{\text{max}}(t)$ 、 $P_{\text{Load,e}}(t)$ 和 $P_{\text{Load,h}}(t)$ 具有不确定性,文中采用风光荷深度神经网络预测模型,通过样本数据学习环境的状态转移概率,从而确定环境的下一状态 s_{t+1} 。其中环境的状态转移概率无须直接计算,由神经网络的参数间接反映。

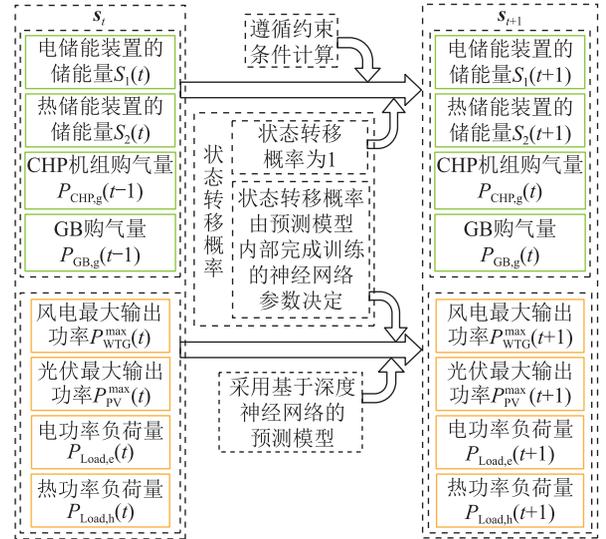


图2 状态转移过程示意

Fig.2 Illustration of state transition process

2.4 奖励函数构建

构建合理的奖励函数能够有效引导智能体寻得最优调度策略。智能体以整个回合内总奖励最大为目标,在神经网络训练中更新网络参数,而CIES优化调度目标为日运行成本最低。因此,文中设计奖励函数 r_t 为:

$$r_t = - \left(\alpha_t P_{\text{buy,e}}(t) + \beta_t P_{\text{buy,g}}(t) + \delta_{\text{WTG}} P_{\text{WTG,cut}}(t) + \delta_{\text{PV}} P_{\text{PV,cut}}(t) + \sum_{n=1}^2 \gamma_n |P_{\text{ES,n}}(t)| + C(t) \right) \quad (21)$$

式中: $C(t)$ 为 t 时刻违反约束的惩罚成本。

$$C(t) = \eta(P_h(t) + P_{buy}(t)) \quad (22)$$

$$P_h(t) = P_{Load,h}(t) + P_{ES,2}(t) - P_{CHP,h}(t) - P_{GB,h}(t) \quad (23)$$

$$P_{buy}(t) = f(0, P_{buy,e}^{max}, P_{buy,e}(t)) \quad (24)$$

式中: η 为计算惩罚成本的越限惩罚系数; $P_h(t)$ 、 $P_{buy}(t)$ 分别为 t 时刻热功率不平衡量、购电越限量, 其中购电越限量也反映了电功率不平衡的情况; $f(x, y, z)$ 为度量变量 z 超出范围 $[x, y]$ 的程度函数。

$$f(x, y, z) = (|x - z| + |y - z| - |y - x|) / 2 \quad (25)$$

显然, $z \in [x, y]$ 时, $f(x, y, z) = 0$, 而 $z \notin [x, y]$ 时, $f(x, y, z)$ 表示越限程度, z 超出范围 $[x, y]$ 越多, 越限越严重。

3 基于 PPO 算法的 CIES 低碳优化调度方法

文中提出了一种基于 PPO 算法的 CIES 低碳优化调度方法。PPO 算法基于动作-评价(actor-critic, AC)架构, 同时融合重要性采样和动态步长剪切机制, 利用深度神经网络构建策略网络 Actor 与价值网络 Critic, 策略网络生成动作概率分布, 价值网络评估动作价值。重要性采样技术将同策略(on-policy)转换为异策略(off-policy), 通过设置新、旧策略网络进行参数更新与样本获取, 实现对同批样本的重复使用, 可以有效利用先前的经验优化策略, 从而加快 PPO 算法的收敛速度。动态步长剪切机制根据新、旧策略网络之间的差异动态调整并更新步长, 从而限制新、旧策略的距离, 降低模型的步长敏感度, 保障模型训练的稳定性^[29]。文中基于 PPO 算法构建的 CIES 低碳优化调度模型的训练与应用机制如图 3 所示。PPO 算法优化求解过程涵盖离线训练和在线应用 2 个阶段, 离线训练阶段通过含风光荷预测模型的 CIES 环境的交互训练, 实现策略网络和价值网络的参数更新; 在线应用阶段则基于离线训练所获得的策略网络, 依据实际的 CIES 环境状态生成最优调度策略。

离线训练中利用风光荷预测模型生成风光荷数据, 从而获取不同风光荷下的 CIES 系统运行场景, 同时通过策略网络与含风光荷预测模型的 CIES 环境的不断交互获得大量训练样本以便策略网络和价值网络学习训练。策略网络和价值网络的输入都是由上一时刻传统机组的调度策略、储能状态和本时刻负荷及风光最大输出功率构成的系统状态向量 s_t 。off-policy 机制的新、旧策略网络 Actor2

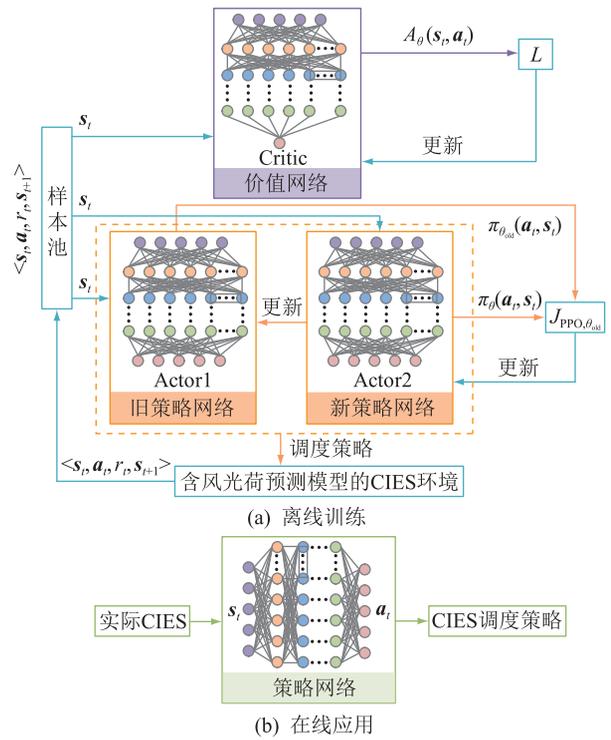


图 3 PPO 训练及应用过程

Fig.3 PPO training and application process

和 Actor1 输出为动作概率分布 π_{θ} 和 $\pi_{\theta_{old}}$, 前者被用于生成新的最优调度策略, 以便与 Actor1 的策略进行比较, 从而更新策略参数; 后者被用于生成智能体与 CIES 环境的交互数据以收集经验数据。其中, θ 为新策略网络 Actor2 的参数; θ_{old} 为旧策略网络 Actor1 的参数。采用价值网络输出的优势函数 $A_{\theta}(s_t, a_t)$ 评价当前状态 s_t 下采取动作 a_t 相对于所有可能动作的平均效果的提升, 从而引导策略网络选择更优的动作。文中设计的策略网络和价值网络采用全连接层结构, 具有处理 CIES 低碳优化调度问题中任务与资源之间强非线性关系的优势。

PPO 利用强化学习方法的自我学习机制以及价值网络逐步改善策略网络的决策性能, 探索学习不同风光荷场景下的最优调度策略。策略网络通过不断交互学习最终掌握在不同风光荷场景下的最优策略。在实际的 IES 环境中, 训练完成的智能体策略网络能够自适应地根据所面临的各种运行场景, 在线给出实时的决策策略。因此, 基于 PPO 算法的 CIES 低碳优化调度决策网络的训练包括样本产生与提取、价值网络训练和策略网络训练三部分。

首先, 在状态 s_t 驱动下, 基于旧策略网络 Actor1 得到动作概率分布, 智能体以此为基础进行采样, 获得动作 a_t , 即 t 时刻输入 CHP 机组的天然气功率、风光实际输出功率及电热储能充放功率。然后, 将生成的调度策略即 a_t 传递给 CIES 交互环境,

基于当前时刻的经济调度成本和惩罚成本计算奖励函数 r_t 。同时,生成下一时刻的状态 s_{t+1} ,由此得到训练样本序列 $\langle s_t, \mathbf{a}_t, r_t, s_{t+1} \rangle$,并将其存储至样本池,以便策略网络和价值网络学习训练。

价值网络训练中,首先,从样本池中提取样本序列 $\langle s_t, \mathbf{a}_t, r_t, s_{t+1} \rangle$,将状态 s_t 输入到价值网络,通过计算优势函数 $A_{\theta}(s_t, \mathbf{a}_t)$ 评价当前状态 s_t 下的价值大小并输出。然后,通过式(26)构造价值网络损失函数 L ,对价值网络进行梯度更新^[31]。

$$L = \text{mean}(\text{square}(A_{\theta}(s_t, \mathbf{a}_t))) \quad (26)$$

式中: $\text{mean}(\cdot)$ 为均值函数; $\text{square}(\cdot)$ 为平方函数。

策略网络训练与价值网络相同,需要先从样本池中提取样本序列 $\langle s_t, \mathbf{a}_t, r_t, s_{t+1} \rangle$,并输入到策略网络 Actor1 和 Actor2,得到相应的动作概率分布 π_{θ} 和 $\pi_{\theta_{\text{old}}}$ 。为使旧策略网络逼近最优策略,寻找使奖励期望值最大的网络参数。PPO 算法利用策略梯度提升算法,策略梯度计算如下:

$$\nabla R(\pi_{\theta}) = E_{(s_t, \mathbf{a}_t) \sim \pi_{\theta}}(A_{\theta}(s_t, \mathbf{a}_t) \nabla \ln \pi_{\theta}(\mathbf{a}_t | s_t)) \quad (27)$$

式中: $\nabla R(\pi_{\theta})$ 为奖励期望的梯度; $E_{(s_t, \mathbf{a}_t) \sim \pi_{\theta}}(\cdot)$ 为在状态 s_t 、动作 \mathbf{a}_t 、动作概率分布 π_{θ} 下计算期望的函数; $\pi_{\theta}(\mathbf{a}_t | s_t)$ 为在状态 s_t 下、动作 \mathbf{a}_t 的动作概率分布。

基于重要性采样技术原理可将式(27)变形为:

$$\nabla R(\pi_{\theta}) = E_{(s_t, \mathbf{a}_t) \sim \pi_{\theta_{\text{old}}}} \left(\frac{\pi_{\theta}}{\pi_{\theta_{\text{old}}}} A_{\theta_{\text{old}}}(s_t, \mathbf{a}_t) \nabla \ln \pi_{\theta} \right) \quad (28)$$

通过复合函数及对数函数对式(28)求导可得:

$$\nabla \pi_{\theta} = \pi_{\theta} \nabla \ln \pi_{\theta} \quad (29)$$

将式(29)代入式(28)得:

$$\nabla R(\pi_{\theta}) = E_{(s_t, \mathbf{a}_t) \sim \pi_{\theta_{\text{old}}}} \left(\frac{\nabla \pi_{\theta}}{\pi_{\theta_{\text{old}}}} A_{\theta_{\text{old}}}(s_t, \mathbf{a}_t) \right) \quad (30)$$

根据策略梯度反推得到目标函数 $J_{\theta_{\text{old}}}$ 为:

$$J_{\theta_{\text{old}}} = E_{(s_t, \mathbf{a}_t) \sim \pi_{\theta_{\text{old}}}} \left(\frac{\pi_{\theta}}{\pi_{\theta_{\text{old}}}} A_{\theta_{\text{old}}}(s_t, \mathbf{a}_t) \right) \quad (31)$$

在策略网络训练中对 $J_{\theta_{\text{old}}}$ 进行梯度提升训练获得使奖励期望值最大的策略网络参数 θ 。

在进行了一定批次的更新后,将新策略网络的参数 θ 同步更新至旧策略网络。

此外,为保证策略更新的稳定提升,新、旧策略网络必须具有相似性,即输入同样的样本序列,两者输出的动作概率分布不能相差过大。PPO 算法为了实现上述目标引入剪切机制,以此限制网络更新幅度,确保新策略在旧策略的近端产生^[32],其具

体公式为:

$$J_{\text{PPO}, \theta_{\text{old}}} = \sum_{(s_t, \mathbf{a}_t)} \min \left(\frac{\pi_{\theta}}{\pi_{\theta_{\text{old}}}} A_{\theta_{\text{old}}}(s_t, \mathbf{a}_t), \text{clip} \left(\frac{\pi_{\theta}}{\pi_{\theta_{\text{old}}}} A_{\theta_{\text{old}}}(s_t, \mathbf{a}_t), 1 - \varepsilon, 1 + \varepsilon \right) \right) \quad (32)$$

式中: $J_{\text{PPO}, \theta_{\text{old}}}$ 为引入剪切机制后的目标函数; ε 为控制剪切区间的超参数; $\text{clip}(\cdot)$ 为剪切函数,用来确保新、旧策略网络的采样概率比总处于区间 $[1 - \varepsilon, 1 + \varepsilon]$ 之内,从而防止新、旧策略差异过大导致算法训练失稳。

PPO 算法与含风光荷预测模型的 CIES 环境进行交互,批量生成样本,并采用梯度提升方法进行网络的批量训练,直至达到最大训练周期,使奖励函数收敛。经过训练的智能体可以在线应用,以应对不同的风光荷场景,实时生成最优调度策略。

4 算例分析

4.1 算例介绍

文中以国内某园区为例,以 24 h 为一个周期进行优化调度。从上级电网购电的分时电价见表 1^[9];天然气价格为 0.45 元/(kW·h)^[33],各设备运行参数见表 2^[9]。系统当前参与的阶梯式碳交易为每日结算,碳交易基价 λ 为 250 元/t,区间长度 l 为 2 t,价格增长率 a 为 25%^[9]。系统碳排放参数见表 3。算法的策略网络和价值网络采用全连接层结构,其学习率为 0.001,折扣因子为 0.99。文中模型算法采用 Python 3.9 编写,计算机环境为 CPU Intel(R) Core(TM) i7-12700H 2.30 GHz、GPU NVIDIA GeForce RTX 3060,算法迭代循环为 32 000 回合。

表 1 分时电价

时段	电价/[元·(kW·h) ⁻¹]
00:00-06:00, 22:00-23:00	0.38
07:00-10:00, 14:00-17:00	0.68
11:00-13:00, 18:00-21:00	1.20

4.2 碳交易机制分析

为验证阶梯式碳交易机制对碳排放的减少效果,文中设置 3 种运行场景进行对比分析。

场景 1:考虑阶梯式碳交易机制,不计及碳交易成本,目标函数仅考虑系统运行成本和弃风光成本的传统经济调度场景。

场景 2:考虑传统碳交易机制,目标函数计及固定碳交易价格的碳排放成本、系统运行成本及弃风光成本的低碳优化调度场景。

表2 设备运行参数

Table 2 Equipment operating parameters

设备名称	参数	数值
CHP 机组	容量/kW	600
	爬坡约束/%	20
	产电效率/%	92
GB	容量/kW	800
	爬坡约束/%	20
	产热效率/%	95
电储能	容量/kW	450
	容量上限约束/%	90
	容量下限约束/%	10
	输入、输出损耗系数	0.95
热储能	容量/kW	500
	容量上限约束/%	10
	容量下限约束/%	90
	输入、输出损耗系数	0.95

表3 碳交易机制参数

Table 3 Carbon trading mechanism parameters

设备名称	实际碳排放	碳排放配额
电网	1.080 kg/(kW·h)	0.728 kg/(kW·h)
CHP 机组	0.102×10^{-3} kg/kJ	0.068×10^{-3} kg/kJ
GB	0.102×10^{-3} kg/kJ	0.062×10^{-3} kg/kJ

场景3:考虑阶梯式碳交易机制,目标函数考虑碳排放成本、系统运行成本及弃风光成本的低碳优化调度场景。

3种运行场景下的系统总运行成本、碳排放总量以及风光消纳率如表4所示。由表4可知,优化目标考虑碳排放成本时,无论是采用传统碳交易机制还是阶梯式碳交易机制,碳排放总量都要小于不考虑碳交易成本的场景,分别降低了107.8 kg和248.9 kg。采用阶梯式碳交易机制的场景3能最大程度降低碳排放总量,经济成本也比采用传统碳交易机制的场景2大幅减少。可见,文中采用阶梯式碳交易机制能最大程度约束碳排,在保证系统运行经济性的同时大大降低碳排放量。

表4 不同场景下的系统调度结果

Table 4 Results of system scheduling under different scenarios

指标	场景1	场景2	场景3
系统总成本/元	15 795.2	15 670.7	15 283.6
碳排放总量/kg	2 555.9	2 448.1	2 307.0
风电消纳率/%	100	100	100
光伏消纳率/%	100	100	100

4.3 CIES 优化调度算法性能比较

图4展示了文中PPO算法和A2C算法离线训

练过程中奖励函数的变化情况。由图4可知,训练至32 000回合时奖励值已完全稳定,这说明2种算法均收敛,智能体已掌握使得系统运行成本最小的最优调度策略。此外,与A2C算法相比,PPO算法尽管收敛速度慢一些,但在训练过程中策略更新更为稳定。

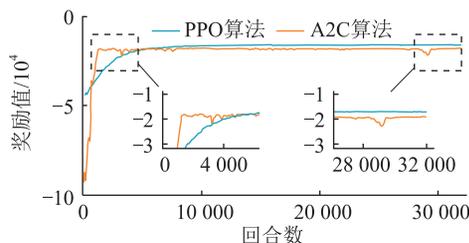


图4 训练过程中奖励值变化曲线

Fig.4 Curves of reward value variation during training

为验证PPO算法能够发挥阶梯式碳交易机制减排作用的优势,将文中算法与强化学习方法中基于AC机制的A2C算法、启发式算法中的PSO算法和基于CPLEX求解器的传统数学规划方法进行对比,CPLEX求解器采用分支定界(branch and bound, BB)法求解混合整数线性规划问题。其中,PSO算法和BB算法属于传统优化算法,对于不确定性采用先预测后优化的处理方案;A2C算法和文中算法属于强化学习方法,对于不确定性采用先离线训练后在线应用、实时获取真实数据的处理方案。采用不同算法对园区某日系统运行成本进行优化,其获得的碳交易成本、运行总成本以及碳排放总量等结果如表5所示。

表5 不同算法的求解结果比较

Table 5 Comparison of results obtained by different algorithms

指标	PSO 算法	BB 算法	A2C 算法	PPO 算法
CHP 机组购电量/(kW·h)	14 400.0	14 400.0	14 400.0	14 309.7
GB 购电量/(kW·h)	11 433.7	11 035.2	11 040.6	11 084.4
购电量/(kW·h)	11 514.6	8 674.7	4 847.4	3 912.2
碳交易成本/元	1 500.0	1 120.2	699.2	595.9
运行总成本/元	22 841.2	19 896.4	15 941.6	15 283.6
碳排放总量/kg	5 000.0	3 984.5	2 637.5	2 307.0

由表5可知,PPO算法所得的平均碳交易成本比另外三者中最优的A2C算法降低14.8%,运行总成本比三者中最低的A2C算法还要减少4.1%。BB算法与PSO算法受预测模型误差限制无法找到最优解;A2C算法属于强化学习方法,因在线获取真实风光数据而获得较优解;而文中PPO算法在AC

结构算法的基础上引入了重要性采样技术和动态步长剪切机制,保证模型训练时策略性能单调提升,可以更有效地处理强不确定性的 CIES 低碳优化调度问题,因此在寻优效果上更为优秀。

采用 4 种算法得到的一个调度周期内的风电和光伏消纳率分别如图 5 和图 6 所示。

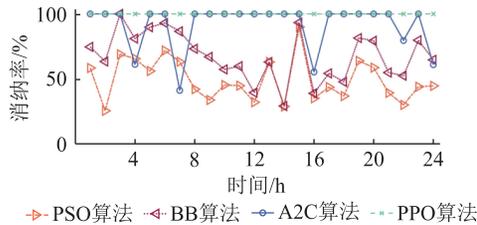


图 5 不同算法得到的风电消纳率比较

Fig.5 Comparison of wind power utilization rates obtained by different algorithms

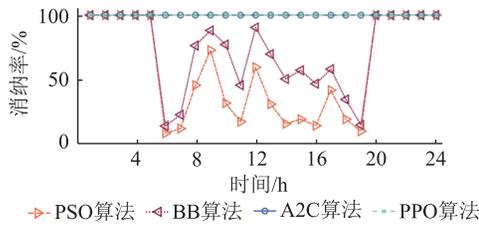


图 6 不同算法得到的光伏消纳率比较

Fig.6 Comparison of photovoltaic utilization rates obtained by different algorithms

由图 5 和图 6 可知,整个调度周期内,PPO 算法和 A2C 算法由于在线获取真实风光数据,可充分利用风光资源,从而获得较高的风光消纳率。然而,由于 A2C 算法在策略更新时的不稳定性,其对某些状态过度探索,从而影响了智能体的参数调整,阻碍了参数优化过程,这不仅削弱了优化调度的效果,而且限制了储能设备在削峰填谷时充分消纳风电和光伏的潜力,因此在 4 时、7 时、16 时、22 时和 24 时风电消纳率不高。尽管 BB 算法的风光消纳率超过 PSO 算法,某些时刻可达 80%甚至近 100%,但比起文中算法和 A2C 算法仍有大量弃风弃光。这是因为不论是 BB 算法还是 PSO 算法,都采用先预测后优化的方案,受预测精度影响无法获得准确的风光数据,因而浪费部分风光资源,造成风光消纳率降低及碳排放量升高。而 PSO 算法风光消纳率最低是因为该算法容易陷入局部最优。

4.4 CIES 优化调度算法的策略比较

为进一步说明在考虑阶梯式碳交易机制下文中算法所得策略较其他三者的优势,图 7 和图 8 分别展示了电功率和热功率的优化调度结果。

由图 7(a)一(d)可知,文中算法和 A2C 算法所得策略实现电功率平衡主要基于风光出力和 CHP

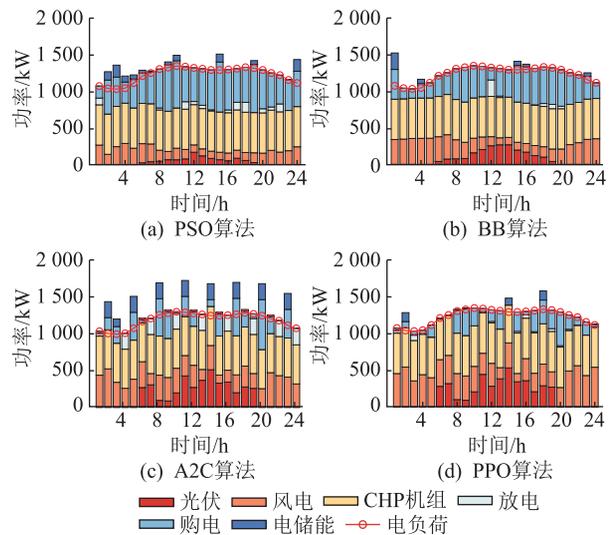


图 7 电功率优化调度结果

Fig.7 Results of electric power optimization scheduling

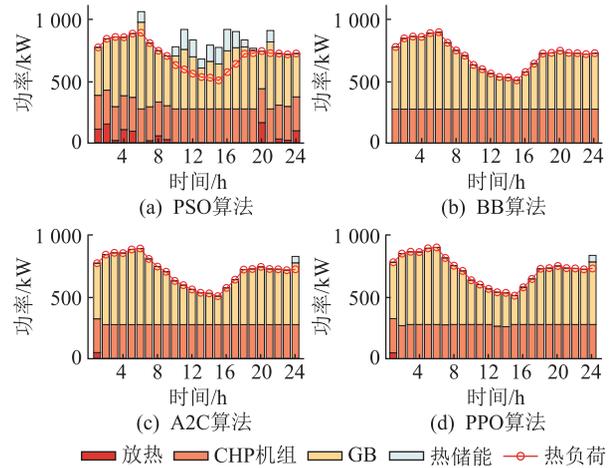


图 8 热功率优化调度结果

Fig.8 Results of thermal power optimization scheduling

机组出力,购电功率占比少;而 PSO 算法和 BB 算法满足电负荷需求的调度策略主要为 CHP 机组出力和购电相互配合,风光出力较少,遇到功率缺额则简单使用购电功率弥补,整体购电功率远远大于文中算法和 A2C 算法。观察上述策略侧重调度手段的不同,可以发现当 CHP 机组出力无法满足电负荷需求时,充分利用风光出力可以大幅降低购电功率。由此可见,强化学习方法利用在线获取真实风光数据的优势可以减小预测模型误差带来的影响,使得风光消纳水平大幅提高。

进一步观察图 7(c)和(d)可以发现,文中算法所得策略不会对电储能频繁进行大功率充放,而是精确使用电储能进行削峰填谷;而 A2C 算法所得策略多次使用电储能进行大功率充放,其原因为 A2C 算法易陷入局部最优而无法获得最佳的电储调度策略。多次使用电储能装置进行大功率充放则会

导致电储运行成本增加,因此 A2C 算法所得策略的运行成本大于 PPO 算法,这与表 5 中的数据一致。

对比图 8(a)—(d) 可知,文中算法、A2C 算法、PSO 算法以及 BB 算法均实现了热功率平衡。观察 4 种算法的调度策略可以发现,文中算法、A2C 算法以及 BB 算法皆使用 GB 出力弥补热功率缺口,而 PSO 算法除了使用 GB 出力弥补热功率缺口,还须通过多次热储能装置调度才能实现热储始末状态一致。由于满足热负荷需求的调度手段只有热储能、GB 和 CHP 机组,同时天然气价格在各个时间段也相同,因此在保证热储能装置始末状态一致的前提下只需要使用 CHP 机组和 GB 满足热负荷需求,无须多次调用储能,频繁调用热储能只会大幅提高热储运行成本。综上所述,文中算法、A2C 算法以及 BB 算法所得策略通过降低热储运行成本获得比 PSO 算法所得策略更低的运行成本。

上述对于调度策略的分析进一步验证了表 5 中数据的合理性。由表 5 可知,文中算法在碳交易成本上分别低于 A2C 算法、BB 算法和 PSO 算法 14.8%、46.8% 及 60.3%,这是因为在调度策略上文中算法优先消纳风光资源并精确使用电储能削峰填谷,避免使用购电弥补;对于运行总成本,A2C 算法优化策略所得成本十分接近文中算法,其他两者的成本则远远大于文中算法,这是因为文中算法和 A2C 算法所得策略皆能高比例消纳风光,较其他两者节省了弃风光成本和原本可以用风光代替的传统机组出力成本,而 A2C 算法没有寻得削峰填谷的最优电储调度策略,因此其系统运行总成本略高于文中算法。

4.5 惩罚系数敏感性分析

为研究越限惩罚系数 η 对文中算法的影响,详细分析不同越限惩罚系数的影响效果。图 9 展示了不同惩罚系数下运行总成本和碳排放总量的变化情况。

由图 9 可知,当 η 值介于 1~9 范围内时,该惩罚系数对智能体的训练和学习具有良好的指导效果。当 η 值等于 5 时,系统运行总成本和碳排放总量均最低;当 η 值从 5 增大或减小时,运行总成本和碳排放总量均呈增加趋势,且变化曲线呈先平缓后急剧上升的趋势。这主要是因为惩罚系数过小时,智能体会忽视惩罚函数的作用,导致其无法充分学习如何在满足约束的同时优化行为策略,难以达到预期的任务目标;而惩罚系数过大时,智能体则会过分集中于减少惩罚值而掩盖奖励目标的变化,导致其在探索环境时过于谨慎,减少尝试新行为从而

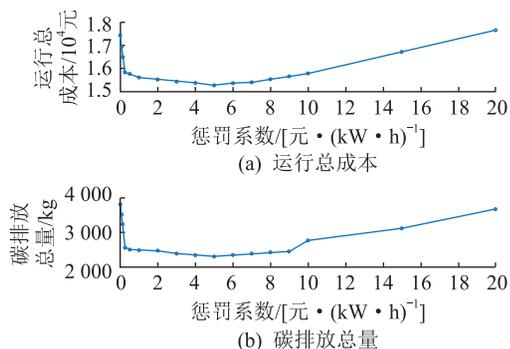


图 9 运行总成本和碳排放总量随惩罚系数的变化
Fig.9 Changes of total operating costs and total carbon emissions with penalty coefficients

陷入局部最优。

4.6 储能装置对优化调度结果的影响性分析

为说明储能对 CIES 优化调度的影响,表 6 给出了储能设备投入前后的系统运行总成本、碳排放总量、风光消纳率等指标。

表 6 储能对 CIES 优化调度的影响分析
Table 6 Impact analysis of energy storage on CIES optimization scheduling

指标	无储能	有储能
运行总成本/元	15 720.6	15 283.6
碳排放总量/kg	2 490.6	2 307.0
风电消纳率/%	97.9	100
光伏消纳率/%	100	100

由表 6 可知,安装储能装置之后系统的运行总成本降低了 437.0 元,碳排放总量减少了 183.6 kg,风电的消纳水平也得到了提高。无储能条件下,光伏实现了完全消纳而风电的消纳率却只有 97.9%,这是由于风电功率在 24 h 内均不为 0,光伏在夜间低负荷需求时段输出为 0,因此只能弃风以满足供需平衡。

5 结论

为充分发挥阶梯式碳交易机制的作用和提高 CIES 低碳优化调度模型的求解性能,文中采用 PPO 算法对计及阶梯式碳交易机制的 CIES 低碳优化调度问题进行了研究。通过算例分析,得出了以下结论:

- (1) 引入阶梯式碳交易机制能在保持 CIES 系统运行经济性的同时有效减少碳排放量。
- (2) 同传统优化求解算法相比,PPO 算法求得到的调度策略能够进一步降低系统碳排放量和运行成本,并提高系统的风光利用率。
- (3) 越限惩罚系数会对 PPO 算法的求解结果

产生影响,越限惩罚系数选择不当会导致求得的结果与全局最优解之间存在较大差异。

文中仅通过敏感性分析研究了越限惩罚系数对PPO算法求解性能的影响,未能给出准确的越限惩罚系数求解方法,未来可对PPO算法中越限惩罚系数的确定方法开展进一步研究。

致 谢

本文得到江苏省研究生科研与实践创新计划项目(SJCX23_1205)资助,谨此致谢!

参考文献:

- [1] CHENG Y H, ZHANG N, WANG Y, et al. Modeling carbon emission flow in multiple energy systems[J]. IEEE Transactions on Smart Grid, 2019, 10(4): 3562-3574.
- [2] 李晖, 刘栋, 姚丹阳. 面向碳达峰碳中和目标的我国电力系统发展研判[J]. 中国电机工程学报, 2021, 41(18): 6245-6259.
- LI Hui, LIU Dong, YAO Danyang. Analysis and reflection on the development of power system towards the goal of carbon emission peak and carbon neutrality[J]. Proceedings of the CSEE, 2021, 41(18): 6245-6259.
- [3] 张沈习, 王丹阳, 程浩忠, 等. 双碳目标下低碳综合能源系统规划关键技术及挑战[J]. 电力系统自动化, 2022, 46(8): 189-207.
- ZHANG Shenxi, WANG Danyang, CHENG Haozhong, et al. Key technologies and challenges of low-carbon integrated energy system planning for carbon emission peak and carbon neutrality[J]. Automation of Electric Power Systems, 2022, 46(8): 189-207.
- [4] DOU Z L, ZHANG C Y, WANG W, et al. Review on key technologies and typical applications of multi-station integrated energy systems[J]. Global Energy Interconnection, 2022, 5(3): 309-327.
- [5] 王一清, 沙倩, 刘秋林. 区域综合能源系统接入的配电网扩展规划研究[J]. 电力科学与技术学报, 2023, 38(6): 198-205.
- WANG Yiqing, SHA Qian, LIU Qiulin. Research on distribution network expansion planning for regional integrated energy system access[J]. Journal of Electric Power Science and Technology, 2023, 38(6): 198-205.
- [6] 贲保记, 张恩硕, 张国, 等. 考虑综合需求响应与“双碳”机制的综合能源系统优化运行[J]. 电力系统保护与控制, 2022, 50(22): 11-19.
- YUN Baoji, ZHANG Enshuo, ZHANG Guo, et al. Optimal operation of an integrated energy system considering integrated demand response and a "dual carbon" mechanism[J]. Power System Protection and Control, 2022, 50(22): 11-19.
- [7] QIN C, YAN Q Y, HE G. Integrated energy systems planning with electricity, heat and gas using particle swarm optimization[J]. Energy, 2019, 188: 116044.
- [8] 梁泽琪, 周云, 冯冬涵, 等. 考虑电碳绿证市场耦合的园区综合能源系统日前优化调度[J]. 电力建设, 2023, 44(12): 43-53.
- LIANG Zeqi, ZHOU Yun, FENG Donghan, et al. Day-ahead optimal scheduling of park-integrated energy system considering electricity-carbon-green certificate market[J]. Electric Power Construction, 2023, 44(12): 43-53.
- [9] 陈锦鹏, 胡志坚, 陈颖光, 等. 考虑阶梯式碳交易机制与电制氢的综合能源系统热电优化[J]. 电力自动化设备, 2021, 41(9): 48-55.
- CHEN Jinpeng, HU Zhijian, CHEN Yingguang, et al. Thermoelectric optimization of integrated energy system considering ladder-type carbon trading mechanism and electric hydrogen production[J]. Electric Power Automation Equipment, 2021, 41(9): 48-55.
- [10] 程杉, 陈诺, 徐建宇, 等. 考虑综合需求响应的楼宇综合能源系统能量管理优化[J]. 电力工程技术, 2023, 42(2): 40-47, 57.
- CHENG Shan, CHEN Nuo, XU Jianyu, et al. Optimal energy management of residential integrated energy system with consideration of integrated demand response[J]. Electric Power Engineering Technology, 2023, 42(2): 40-47, 57.
- [11] 马志侠, 张林鎰, 邱朝洁, 等. 基于CEEMD-SSA-LSTM的园区综合能源系统两阶段优化调度[J]. 高电压技术, 2023, 49(4): 1430-1440.
- MA Zhixia, ZHANG Linxuan, QIU Chaojie, et al. Two-stage optimal scheduling of the park integrated energy system based on CEEMD-SSA-LSTM[J]. High Voltage Engineering, 2023, 49(4): 1430-1440.
- [12] 陈曦, 袁梦玲, 王松, 等. 考虑碳交易影响风电消纳的综合能源系统优化运行[J]. 重庆理工大学学报(自然科学), 2022, 36(1): 268-276.
- CHEN Xi, YUAN Mengling, WANG Song, et al. Optimal scheduling of integrated energy system considering impact of carbon trading on wind power consumption[J]. Journal of Chongqing University of Technology (Natural Science), 2022, 36(1): 268-276.
- [13] 李玲, 曹锦业, Nikita Tomin, 等. 计及电动汽车接入的区域综合能源系统双层日前协调优化调度[J]. 电力建设, 2023, 44(5): 23-33.
- LI Ling, CAO Jinye, TOMIN N, et al. Bi-level coordinated day-ahead optimal dispatch of regional integrated energy system considering the integrations of electric vehicles[J]. Electric Power Construction, 2023, 44(5): 23-33.
- [14] 陈明昊, 孙毅, 谢志远. 基于双层深度强化学习的园区综合能源系统多时间尺度优化管理[J]. 电工技术学报, 2023, 38(7): 1864-1881.
- CHEN Minghao, SUN Yi, XIE Zhiyuan. The multi-time-scale management optimization method for park integrated energy system based on the bi-layer deep reinforcement learning[J]. Transactions of China Electrotechnical Society, 2023, 38(7): 1864-1881.
- [15] DONG L, WEI J, LIN H, et al. Distributed optimization of elec-

- tricity-gas-heat integrated energy system with multi-agent deep reinforcement learning [J]. *Global Energy Interconnection*, 2022, 5(6):604-617.
- [16] 郑洁云, 宋倩芸, 吴桂联, 等. 基于 Q 学习的区域综合能源系统低碳运行策略[J]. *电力科学与技术学报*, 2022, 37(2):106-115, 128.
- ZHENG Jieyun, SONG Qianyun, WU Guilian, et al. Low-carbon operation strategy of regional integrated energy system based on the Q learning algorithm [J]. *Journal of Electric Power Science and Technology*, 2022, 37(2):106-115, 128.
- [17] 孙立钧, 顾雪平, 刘彤, 等. 一种基于深度强化学习算法的电网有功安全校正方法[J]. *电力系统保护与控制*, 2022, 50(10):114-122.
- SUN Lijun, GU Xueping, LIU Tong, et al. A deep reinforcement learning algorithm-based active safety correction method for power grids[J]. *Power System Protection and Control*, 2022, 50(10):114-122.
- [18] 孙广明, 陈良亮, 王瑞升, 等. 基于深度强化学习的充光储能电站调度策略[J]. *电力工程技术*, 2021, 40(5):17-24.
- SUN Guangming, CHEN Liangliang, WANG Ruisheng, et al. A deep reinforcement learning-based scheduling strategy of photovoltaic-storage-charging integrated energy stations[J]. *Electric Power Engineering Technology*, 2021, 40(5):17-24.
- [19] 彭寒梅, 胡磊, 谭貌, 等. 基于深度强化学习的电-气区域综合能源系统安全校正控制决策方法[J]. *电网技术*, 2024, 48(3):1019-1030.
- PENG Hanmei, HU Lei, TAN Mao, et al. Decision-making method for electricity-gas region integrated energy system security correction based on deep reinforcement learning [J]. *Power System Technology*, 2024, 48(3):1019-1030.
- [20] ALIBABAEI K, GASPARD P D, ASSUNÇÃO E, et al. Comparison of on-policy deep reinforcement learning A2C with off-policy DQN in irrigation optimization: a case study at a site in Portugal [J]. *Computers*, 2022, 11(7):104.
- [21] SUTTON R S, BARTO A G. Reinforcement learning: an introduction [M]. 2nd ed. London: The MIT Press, 2014.
- [22] 王桂兰, 张海晓, 刘宏, 等. 基于近端策略优化算法含碳捕集的综合能源系统低碳经济调度[J]. *计算机应用研究*, 2024, 41(5):1508-1514.
- WANG Guilian, ZHANG Haixiao, LIU Hong, et al. Low carbon economic scheduling of integrated energy systems based on proximal policy optimization algorithm with carbon capture [J]. *Application Research of Computers*, 2024, 41(5):1508-1514.
- [23] 雷嘉明, 姜爱华, 吴新飞, 等. 计及源荷不确定性的综合能源系统近端策略优化调度[J]. *电力科学与技术学报*, 2023, 38(5):1-11.
- LEI Jiaming, JIANG Aihua, WU Xinfei, et al. Proximal policy optimization dispatch of integrated energy system considering source-load uncertainty [J]. *Journal of Electric Power Science and Technology*, 2023, 38(5):1-11.
- [24] 王守相, 郑婉婷, 赵倩宇, 等. 基于碳-绿证互认和电热柔性负荷的含氢多能系统源荷低碳经济调度方法[J/OL]. *高技术技术*, 1-12 [2024-05-23]. <https://doi.org/10.13336/j.1003-6520.hve.20232057>.
- WANG Shouxiang, ZHENG Wanting, ZHAO Qianyu, et al. Source-load low-carbon economic dispatch method for hydrogen multi-energy system based on mutual recognition of carbon-green certificates and electric and thermal flexible loads [J/OL]. *High Voltage Engineering*: 1-12 [2024-05-23]. <https://doi.org/10.13336/j.1003-6520.hve.20232057>.
- [25] 原希尧, 王关涛, 朱若源, 等. 碳-绿色证书交易机制下考虑回收 P2G 余热和需求响应的 PIES 优化调度[J]. *电力建设*, 2023, 44(3):25-35.
- YUAN Xiyao, WANG Guantao, ZHU Ruoyuan, et al. Optimal scheduling of park integrated energy system with P2G waste heat recovery and demand response under carbon-green certificate trading mechanism [J]. *Electric Power Construction*, 2023, 44(3):25-35.
- [26] 骆钊, 王菁慧, 王华, 等. 考虑碳捕集和电转气的综合能源系统优化调度[J]. *电力自动化设备*, 2023, 43(12):127-134.
- LUO Zhao, WANG Jinghui, WANG Hua, et al. Optimal scheduling of integrated energy system considering carbon capture and power-to-gas [J]. *Electric Power Automation Equipment*, 2023, 43(12):127-134.
- [27] 王利猛, 刘雪梦, 李扬, 等. 阶梯式碳交易机制下考虑需求响应的综合能源系统低碳优化调度[J]. *电力建设*, 2024, 45(2):102-114.
- WANG Limeng, LIU Xueming, LI Yang, et al. Low-carbon optimal dispatch of integrated energy system considering demand response under the tiered carbon trading mechanism [J]. *Electric Power Construction*, 2024, 45(2):102-114.
- [28] 黄冬梅, 王干帅, 孙锦中, 等. 碳交易机制下园区综合能源系统优化调度[J]. *电力系统及其自动化学报*, 2024, 36(5):90-98.
- HUANG Dongmei, WANG Ganshuai, SUN Jinzhong, et al. Optimal scheduling of park-level integrated energy system under carbon trading mechanism [J]. *Proceedings of the CSU-EPSC*, 2024, 36(5):90-98.
- [29] 杨志学, 任洲洋, 孙志媛, 等. 基于近端策略优化算法的新能源电力系统安全约束经济调度方法[J]. *电网技术*, 2023, 47(3):988-998.
- YANG Zhixue, REN Zhouyang, SUN Zhiyuan, et al. Security-constrained economic dispatch of renewable energy integrated power systems based on proximal policy optimization algorithm [J]. *Power System Technology*, 2023, 47(3):988-998.
- [30] 刘必晶. 基于深度强化学习的综合能源系统优化调度[J]. *现代电力*, 2024, 41(4):710-717.
- LIU Bijing. Optimal dispatch of integrated energy system based on deep reinforcement learning [J]. *Modern Electric Power*, 2024, 41(4):710-717.
- [31] 李奇, 刘鑫, 孟翔, 等. 基于近端策略优化算法的燃料电池混合动力系统综合价值损耗最小能量管理方法[J]. *中国电机工程学报*, 2024, 44(12):4788-4799.

- LI Qi, LIU Xin, MENG Xiang, et al. Comprehensive value depletion minimization energy management method for fuel cell hybrid systems based on proximal policy optimization algorithm [J]. Proceedings of the CSEE, 2024, 44(12):4788-4799.
- [32] LIU C, WANG Y J, WANG L, et al. Load-adaptive real-time energy management strategy for battery/ultracapacitor hybrid energy storage system using dynamic programming optimization [J]. Journal of Power Sources, 2019, 438:227024.
- [33] 张晓辉, 刘小琰, 钟嘉庆. 考虑奖惩阶梯型碳交易和电-热转移负荷不确定性的综合能源系统规划[J]. 中国电机工程学报, 2020, 40(19):6132-6142.
- ZHANG Xiaohui, LIU Xiaoyan, ZHONG Jiaqing. Integrated energy system planning considering a reward and punishment lad-

der-type carbon trading and electric-thermal transfer load uncertainty[J]. Proceedings of the CSEE, 2020, 40(19):6132-6142.

作者简介:



陈凡

陈凡(1981),女,博士,教授,研究方向为电力系统规划与运行分析(E-mail:fanchen_nj@163.com);

吴凌霄(2000),男,硕士在读,研究方向为电力系统运行与分析;

王曼(1977),女,硕士,副教授,研究方向为电力系统运行与规划。

A low-carbon optimization scheduling method of CIES based on PPO algorithm

CHEN Fan, WU Lingxiao, WANG Man, LYU Ganyun, ZHANG Xiaolian

(School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

Abstract: The tiered carbon trading mechanism and optimization scheduling model solving algorithm are pivotal for the community integrated energy system (CIES). CIES plays a crucial role in optimizing scheduling, yet existing literature often does not fully consider these two factors. To address this gap, the adoption of the proximal policy optimization (PPO) algorithm is proposed, which incorporates a ladder-type carbon trading mechanism to solve the low-carbon optimization scheduling problem of CIES. This method constructs a reinforcement learning interactive environment based on a low-carbon optimization scheduling model. The intelligent agent's state, action space, and reward function are defined using device status and operating parameters. An intelligent agent capable of generating the optimal policy is obtained through offline training. Case study analysis results demonstrate that the low-carbon optimization scheduling scheme for CIES achieved through the PPO algorithm, effectively leverages the advantages of the tiered carbon trading mechanism, significantly reducing carbon emissions and improving energy utilization efficiency.

Keywords: community integrated energy system (CIES); optimize scheduling; proximal policy optimization (PPO) algorithm; ladder-type carbon trading mechanism; penalty coefficient; carbon emission

(编辑 陆海霞)