

热力与综合能源系统

文章编号:1001-2060(2026)02-0140-13

基于双延迟深度确定性策略梯度算法的燃煤机组耦合碳捕集系统优化调度研究

钱烽雷¹, 张子腾¹, 王思腾², 吴 哮²

(1. 中国能源建设集团江苏省电力设计院有限公司, 江苏南京 211102; 2. 东南大学 能源与环境学院, 江苏南京 210096)

摘要: 配备碳捕集系统是实现燃煤机组低碳化改造的重要途径, 但燃煤机组耦合碳捕集系统存在复杂的电、热、碳耦合关系, 需要通过优化调度合理分配系统出力, 以保障其经济、低碳、灵活运行。由于电、碳负荷的强随机波动和机组宽工况运行的非线性特征, 基于线性模型的常规优化调度方法难以取得满意效果。为此, 提出一种基于双延迟深度确定性策略梯度算法(TD3)的数据驱动燃煤机组-碳捕集系统智能调度方法, 采用不确定负荷指令分解策略提高智能体在不确定场景中的寻优效果。设计计及机组煤耗成本、运维成本和负荷偏差惩罚的优化调度目标函数, 考虑设备运行、能量平衡和日均碳捕集率在内的调度约束, 构建数学规划调度框架, 进而将此框架转化为马尔科夫决策过程, 再通过智能体与仿真模型的离线交互训练, 自适应学习不确定环境下的最优调度策略。仿真结果表明: 在 1 000 MW 超超临界热电联产机组耦合单乙醇胺吸收碳捕集系统中, 相比基于线性模型的混合整数线性规划(MILP)算法和常规不确定 TD3 算法, 所提方法的日运行总成本分别降低了 0.66% 和 0.52%。

关键词: 燃煤发电; 碳捕集; 灵活运行; 优化调度; 深度强化学习

中图分类号: TK019 文献标识码: A DOI: 10.16146/j.cnki.rndlgec.2026.02.016

[引用本文格式] 钱烽雷, 张子腾, 王思腾, 吴 哮. 基于双延迟深度确定性策略梯度算法的燃煤机组耦合碳捕集系统优化调度研究 [J]. 热能动力工程, 2026, 41(2): 140-152. QIAN Fenglei, ZHANG Ziteng, WANG Siteng, WU Xiao. Research on the optimal scheduling of coal-fired power units integrated with carbon capture system based on twin delayed deep deterministic policy gradient algorithm [J]. Journal of Engineering for Thermal Energy and Power, 2026, 41(2): 140-152.

Research on the Optimal Scheduling of Coal-fired Power Units Integrated with Carbon Capture System based on Twin Delayed Deep Deterministic Policy Gradient Algorithm

QIAN Fenglei¹, ZHANG Ziteng¹, WANG Siteng², WU Xiao²

(1. China Energy Engineering Group Jiangsu Power Design Institute Co., Ltd., Nanjing, China, Post Code: 211102;
2. School of Energy and Environment, Southeast University, Nanjing, China, Post Code: 210096)

Abstract: Equipping with carbon capture system for coal-fired power units (CFPU) is an important means of achieving low-carbon retrofits. CFPU integrated with post-combustion carbon capture systems (CFPU-PCC) contain complex electrical, thermal and carbon coupling relationships, thus requiring optimal scheduling to reasonably allocate system output to ensure economical, low-carbon and flexible opera-

收稿日期: 2025-07-17; 修订日期: 2025-09-23

基金项目: 国家自然科学基金项目(52376002); 中国能源建设集团江苏省电力设计院有限公司科技项目(32-JK-2024-010); 国家重点研发计划(2022YFB4100404)

Fund-supported Project: National Natural Science Foundation of China (52376002); Science and Technology Project of Jiangsu Power Design Institute Co., Ltd. of China Energy Engineering Group (32-JK-2024-010); National Key Research and Development Program of China (2022YFB4100404)

作者简介: 钱烽雷(1980-), 男, 中国能源建设集团江苏省电力设计院有限公司高级工程师。

通信作者: 吴 哮(1986-), 男, 东南大学教授。

tion. For the strong random fluctuations in electricity and carbon loads, and the nonlinear characteristics of wide operating conditions for power units, the conventional linear model-based scheduling methods are ineffective. To address this, a data-driven intelligent scheduling method was proposed for CFPU-PCC based on twin delayed deep deterministic policy gradient algorithm (TD3), an uncertainty load command decomposition strategy was adopted to improve optimization capability of agents under uncertain situations. An optimized scheduling objective function consisting of coal consumption cost of power units, operation and maintenance cost, and load deviation penalty was designed. A mathematical programming scheduling frame was constructed, including scheduling constraints of equipment operation, energy balance and daily average carbon capture rate. And this conventional optimal scheduling frame was converted into a Markov decision process. Through offline interacting training between the agent and simulation model, the optimal scheduling strategy under uncertain conditions was adaptively learned. Simulation results indicate that in a 1 000 MW ultra-supercritical combined heat and power (CHP) unit coupled with a monoethanolamine absorption carbon capture system, the proposed method reduces total daily operating cost by 0.66% and 0.52% respectively, compared to mixed-integer linear programming (MILP) algorithm based on linear models and conventional uncertain TD3 algorithm.

Key words: coal-fired power generation; carbon capture; flexible operation; optimal scheduling; deep reinforcement learning

符号说明

CFPU—燃煤发电机组	
HP—汽轮机高压缸	
IP—汽轮机中压缸	
LP—汽轮机低压缸	
PCC—燃烧后碳捕集系统	
P —电功率, MW	
η —CO ₂ 捕集率, %	
m —质量流量, kg/s	
st —抽汽质量流量, kg/s	
H —热功率, MW	
k —公式中的转换系数	
F —总成本, 元	
C —单位成本, 元	
Δt —优化间隔, h	
U —CO ₂ 储罐储量, t	
L —负荷指令	
r —奖励, 元	
α —一步约束越界惩罚值	
θ —评估网络参数	
ϕ —策略网络参数	
θ' —目标评估网络参数	
ϕ' —目标策略网络参数	
a —动作	
\tilde{a} —目标策略平滑动作	
s —状态	
s' —新状态	
$\pi(\cdot)$ —策略函数	
ε —探索噪声	
σ —探索噪声标准差	
$\tilde{\sigma}$ —目标策略平滑噪声标准差	
c —裁剪限值	
γ —折扣因子	
y —目标动作价值	

Q —动作价值函数	
J —策略网络性能指标	
∇ —梯度算子	
τ —软更新系数	
下标:	
t —优化时刻	
coal—燃煤机组给煤	
elec—供电	
cap—CO ₂ 捕集	
heat—供热	
comp—压缩机	
all—全部	
imb—出力不平衡	
om—运行维护	
max—最大值	
min—最小值	
set—设置限值	
vio—越限	
ope—运行	
上标:	
pre—预测	
bias—偏差	
step—步	
epi—回合	

引言

2024年,全国煤电发电量占比为54.8%,是电力行业碳排放的主要来源,约占全国CO₂排放总量的三分之一^[1]。发改委《煤电低碳化改造建设行动方案》明确指出,碳捕集利用与封存(CCUS)是煤电低碳改造的重要技术方向。其中,基于化学吸收的

燃烧后碳捕集(PCC)技术工艺成熟,已在国能锦界、国能泰州^[2]、加拿大边界大坝^[3]、美国佩特拉诺瓦^[4]等项目示范应用,具备规模化、商业化水平。

耦合化学吸收燃烧后碳捕集系统对燃煤发电机组运行提出了诸多挑战。碳捕集过程需要从汽轮机中抽取大量蒸汽为吸收剂再生供能,机组出力将降低 20%~30%^[5]。同时,为适应机组变负荷运行下的烟气变化,捕集系统也需频繁调整抽汽量,对机组的稳定、灵活运行产生不利影响。因此,对燃煤机组耦合碳捕集耦合系统进行优化调度,协调系统的电、热、碳出力,是保证机组经济、灵活、低碳运行的关键。

目前,燃煤机组耦合碳捕集系统优化调度的相关研究已有较多。Guo 等人^[6]提出一种考虑深度调峰的光伏-燃煤碳捕集联合发电系统低碳、经济调度方法,通过在优化指标中引入调峰成本及补偿收益提高系统的经济性能和光电消纳水平。Zantye 等人^[7]提出一种包含可再生能源、燃煤碳捕集和电锅炉的综合能源系统,通过构造两阶段配置-运行优化策略,验证了系统具有在不同地区长时间运行的优势。Li 等人^[8]考虑到风力发电和用电需求的不确定性,构建了包含燃煤-碳捕集机组电力系统的日前随机优化调度模型,协同碳捕集及需求响应的灵活性,降低了系统运行成本,促进了风电消纳。Wu 等人^[9]采用核密度估计法拟合年度碳价的非参数概率分布模型,采用鲁棒优化进行燃煤碳捕集电厂的电、碳协同调度,提高了系统在变动碳市场中的降碳效果。

上述研究均采用简化的线性模型表征机组燃料消耗-发电-碳排放间的关联,难以反映机组变工况运行下的非线性特性,在优化过程中也未能充分表征机组整体电、热、碳出力间的耦合约束。为此,Zantye 等人^[10]提出一种煤电碳捕集系统两阶段随机规划调度模型,通过代数模型自学习方法(Automatic Learning of Algebraic Models for Optimization, ALAMO)^[11]拟合不同调度阶段的系统非线性模型,并通过 BARON 求解器求解,验证了电价波动对系统盈利的增益效果。Chen 等人^[12]采用深度置信网络(DBN)构建了燃煤机组耦合碳捕集系统的非线性替代模型,提出了考虑可再生能源消纳、碳减排等性能指标的电站碳捕集系统优化调度方法并采用粒子群算法求解,结果表明,碳捕集系统灵活运行对发

电机组保供、调峰具有显著的支撑作用。

综上所述,目前有关燃煤机组耦合碳捕集系统优化调度的研究大多基于反映系统核心变量间关联的模型,根据负荷预测信息开展系统电碳出力的逐时优化分配研究仍面临两方面主要问题。一方面,调度优化品质严重依赖于模型性能,线性模型优化求解效率良好,但难以反映机组宽负荷运行中的非线性耦合特性,而非线性模型存在难以求解、易陷入局部最优的问题;另一方面,调度结果可靠性严重依赖预测数据的质量或对其概率分布的掌握,而此类信息在实际运行中常常难以准确提供^[13]。

近年来,人工智能算法的发展为复杂优化问题的求解提供了崭新的解决思路。深度强化学习(Deep Reinforcement Learning, DRL)融合深度神经网络与强化学习技术,通过多层特征提取增强环境感知能力,构建状态、动作的高维映射关系,可显著提升在复杂非线性、不确定性系统调度任务中的决策精度与泛化能力。Li 等人^[14]采用 SARSA 算法,在不同的碳配额条件下开展燃煤机组耦合碳捕集系统的投标-运行过程联合优化,可实现碳捕集率时变的灵活运行,辅助决策者提高发电机组运行的折现累计利润。刘倩等人^[15]利用深度 Q 网络(Deep Q-Network, DQN)算法提升了求解园区综合能源系统调度问题的源荷波动适应性和计算效率。上述研究采用强化学习中的价值学习算法,该算法在包含高维、连续决策变量的复杂问题中计算能力不足,且易忽略状态与动作间的关联信息^[16]。Li 等人^[17]采用深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)算法求解风-光-火-储电力系统的低碳、经济指标优化问题,验证了所提方法的电-碳协同优化优势。尽管 DDPG 实现了决策变量空间的连续化,但易出现 Q 值过拟合问题,导致策略学习失准^[18]。

上述研究从电力系统宏观视角开展基于 DRL 的调度优化研究,但仍缺乏面向燃煤-碳捕集耦合系统,且兼顾机组变工况非线性、电热碳耦合约束及负荷不确定性的智能调度方法。为此,本文提出一种基于双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3)算法的 DRL 燃煤机组耦合碳捕集系统调度方法:首先构建数据驱动的系统替代模型,作为智能体实验训练环境;进而构建基于数学规划算法的系统调度优化框架,将其

转化为马尔科夫决策过程,并在状态中引入实际与预测负荷间的偏差信息,增强对不确定性的适应能力;最后利用TD3的双评估网络、延迟更新及动作平滑机制离线训练智能体,获得自适应调度策略。案例分析验证了本文所提方法的优越性。

1 系统稳态模型构建

1.1 系统运行流程

本文以1 000 MW超超临界燃煤热电联产机组耦合基于质量分数为30%单乙醇胺(MEA)溶剂的全烟气燃烧后碳捕集系统(Coal-Fired Power Units integrated with Post-Combustion Carbon Capture Systems, CFPU-PCC)为研究对象,系统结构如图1所示。

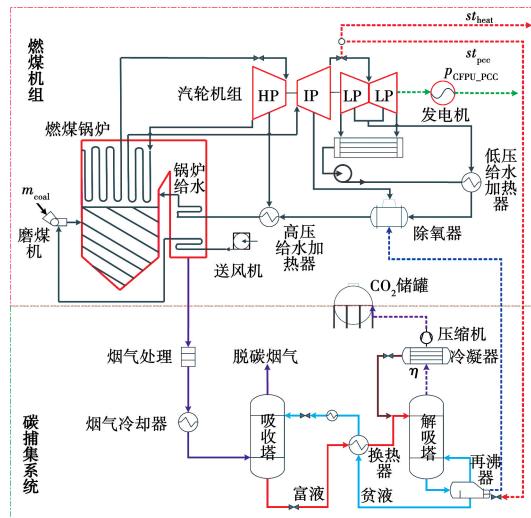


图1 燃煤机组耦合碳捕集系统结构图

Fig. 1 Structural diagram of CFPU-PCC

系统主要运行流程为:燃煤锅炉排烟经脱硫、脱硝、除尘、冷却等工艺处理后送入碳捕集系统吸收塔底,在塔内与自塔顶流下的吸收溶液逆向接触,实现CO₂脱除,净化气从塔顶排出。富含CO₂的溶液经换热器加热后流入解吸塔顶,被供入再沸器的汽轮机组中压缸(IP)出口抽汽加热,分离出CO₂后从塔底流出,参与下一轮吸收-解吸循环。抽汽经再沸器利用后被冷凝,送入除氧器,与低压给水加热器给水混合后流入高压给水加热器,混合为锅炉给水。解吸塔顶流出的CO₂经压缩、冷却等工艺后送至储罐储存,供下游用户使用。系统典型工况运行参数如表1所示。

表1 典型工况下系统主要运行参数

Tab. 1 Main operating parameters of the system under typical working conditions

参数	数值
给煤量/(kg·s ⁻¹)	105.5
输出电功率/MW	852.3
供热抽汽量/(kg·s ⁻¹)	50.3
给水流量/(kg·s ⁻¹)	715.8
主汽阀开度/%	94.3
主蒸汽压力/MPa	26.5
烟气CO ₂ 质量分数/%	21.5
碳捕集抽汽量/(kg·s ⁻¹)	105.8
分离器比焓/(kJ·kg ⁻¹)	2 687.4
再沸器抽汽比焓/(kJ·kg ⁻¹)	1 580
再沸器抽汽流量/(kg·s ⁻¹)	193.6
再沸器温度/K	386
贫液流量/(kg·s ⁻¹)	213
碳捕集率/%	80

1.2 系统替代模型构建

前期研究已基于gCCS[®]仿真软件^[19]搭建了上述系统的机理模型。考虑到模型中含有大量偏微分方程及中间变量,计算复杂,不利于深度强化学习框架下智能体的高效学习,本文采用DBN^[20]构建表征核心变量间稳态关联的替代模型f_{CFPU-PCC},作为智能体学习交互的环境,如式(1)所示:

$$(P_t, \eta_t) = f_{\text{CFPU-PCC}}(m_{\text{coal},t}, st_{\text{cap},t}, st_{\text{heat},t}) \quad (1)$$

DBN模型的输出验证结果如图2所示。

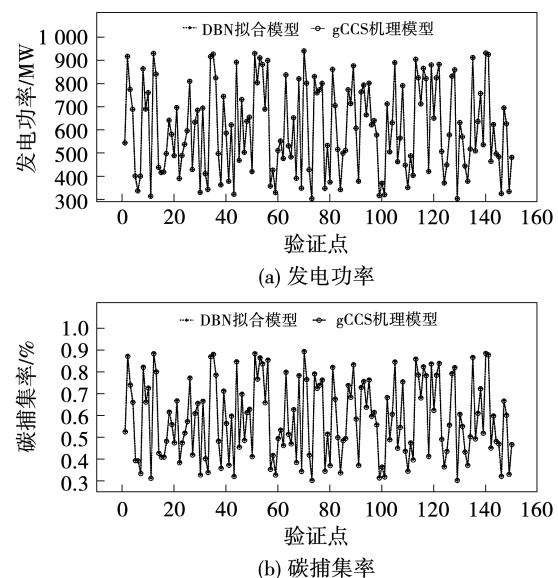


图2 模型拟合精度验证结果

Fig. 2 Model fitting accuracy validation results

由图 2 可知, DBN 模型能在宽工况运行范围内很好地跟踪 gCCS 机理模型的输出, 从而为训练的可靠性提供有效支撑。

采用均方误差 (MSE) 和决定系数 (R^2) 对比了 DBN、BP 神经网络和线性拟合 3 种建模方法的拟合精度, 结果如表 2 所示。

表 2 不同建模方法拟合精度指标对比

Tab. 2 Comparison of fitting accuracy indicators of different modeling methods

建模方法	输出电功率		碳捕集率	
	R^2	MSE	R^2	MSE
DBN	0.998 1	5.81	0.997 4	0.000 2
BP 神经网格	0.990 1	21.47	0.989 1	0.001 5
线性拟合	0.969 2	98.64	0.09 70	0.002 5

由表 2 可知, DBN 模型各输出的 MSE 最小、 R^2 最大, 说明其对机理模型的非线性特征捕捉能力与拟合精度最优, 证明了该建模方法的有效性。对比表 2 中的 3 种模型与 gCCS 机理模型的计算效率, 结果如表 3 所示。由表 3 可知, DBN 模型的单步计算耗时为 0.09 s, 远低于机理模型, 可为智能体提供高效、可靠的学习环境。

表 3 不同建模方法计算效率对比

Tab. 3 Comparison of computational efficiency of different modeling methods

建模方法	单步计算耗时/s
DBN	0.09
BP 神经网络	0.06
线性拟合	0.05
gCCS 机理	0.50

除采用 DBN 模型来反映 CFPU-PCC 核心变量的关联外, 补充模型(2)~(4), 以量化各优化时刻下的机组供热功率、 CO_2 捕集量及 CO_2 压缩电耗:

$$H_t = k_{\text{heat}} \cdot m_{\text{heat},t} \quad (2)$$

$$m_{\text{cap},t} = k_{\text{cap}} \cdot m_{\text{coal},t} \cdot \eta_t \quad (3)$$

$$P_{\text{comp},t} = k_{\text{comp}} \cdot m_{\text{cap},t} \quad (4)$$

2 基于 TD3 的燃煤机组 - 碳捕集系统智能优化调度

以第 1 节构建的 DBN 替代模型作为智能体交

互对象, 将基于 TD3 算法构建燃煤机组耦合碳捕集系统经济、低碳智能优化调度框架。

2.1 目标函数

智能优化调度框架以基于目标函数 - 约束条件的常规优化框架为基础, 故首先进行该常规优化框架的构建。

假设燃煤机组 - 碳捕集系统运行需满足外界的电、热、碳负荷需求, 本文优化目标函数如式(5)所示:

$$\min F_{\text{all}} = F_{\text{coal}} + F_{\text{om}} + F_{\text{imb}} \quad (5)$$

其中, 燃料成本可按式(6)计算得到:

$$F_{\text{coal}} = C_{\text{coal}} \sum_{t=1}^{24} 3.6 m_{\text{coal},t} \Delta t \quad (6)$$

系统设备运行维护成本可按式(7)计算得到:

$$F_{\text{om}} = \sum_{t=1}^{24} (C_{\text{elec}} P_t + C_{\text{heat}} H_t + 3.6 C_{\text{cap}} m_{\text{cap},t}) \quad (7)$$

当系统的电、热、碳供应量高于负荷需求时, 对超出量进行切除; 当低于需求时, 须施加相应的负荷不平衡惩罚, 不平衡惩罚可按式(8)计算得到:

$$F_{\text{imb}} = \sum_{t=1}^{24} (C_{\text{elec}} P_{\text{imb},t} + C_{\text{heat}} H_{\text{imb},t} + 3.6 C_{\text{cap}} m_{\text{cap},t}^{\text{imb}}) \quad (8)$$

2.2 约束条件

优化调度中需考虑系统的能量平衡约束和设备运行约束。由于需从汽轮机中抽取过热蒸汽用于供热及碳捕集吸收剂再生, 系统发电量将会显著下降。考虑机组的电热耦合关系和汽轮机的安全运行限制, 将如图 3 所示的系统电热可行区域^[21]作为运行约束在优化调度中予以考虑。

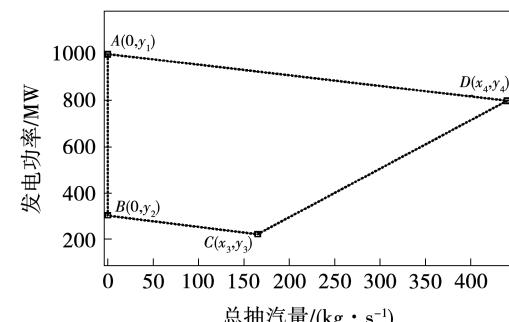


图 3 燃煤机组耦合碳捕集系统电热可行区域^[21]

Fig. 3 Feasible electrical-thermal operating range of CFPU-PCC^[21]

在图 3 中, 纵坐标为燃煤机组的发电功率, 横坐

标为机组非供电总抽汽量,由碳捕集抽汽和供热抽汽两部分组成。点 A,B 分别代表机组在无抽汽情况下汽轮机最大连续额定负荷(Turbine Maximum Continuous Rating, TMCR) 和 30% 汽轮机热耗验收(Turbine Heating Acceptance, THA) 工况下的发电量;点 C,D 分别代表系统在抽汽量最大时,30% THA 和 TMCR 工况下对应的发电量。上述 4 点围合所得区域形成的约束如式(9)所示:

$$\begin{cases} \frac{y_2 - y_3}{x_3} (st_{\text{cap},t} + st_{\text{heat},t}) + y_2 - P_t \leq 0; \\ \frac{y_3 - y_4}{x_3 - x_4} (st_{\text{cap},t} + st_{\text{heat},t}) + y_3 - \frac{y_3 - y_4}{x_3 - x_4} x_3 - P \geq 0; \\ \frac{y_1 - y_4}{x_4} (st_{\text{cap},t} + st_{\text{heat},t}) + y_1 - P_t \geq 0 \end{cases} \quad (9)$$

燃煤机组-碳捕集系统的设备运行约束如式(10)~式(15)所示:

$$m_{\text{coal,min}} \leq m_{\text{coal},t} \leq m_{\text{coal,max}} \quad (10)$$

$$st_{\text{cap,min}} \leq st_{\text{cap},t} \leq st_{\text{cap,max}} \quad (11)$$

$$st_{\text{heat,min}} \leq st_{\text{heat},t} \leq st_{\text{heat,max}} \quad (12)$$

$$U_{\min} \leq U_t \leq U_{\max} \quad (13)$$

$$U_0 = U_{24} \quad (14)$$

$$\eta_{\min} \leq \eta_t \leq \eta_{\max} \quad (15)$$

需指出的是, U_0 和 U_{24} 分别为优化调度周期开始、结束时刻的 CO_2 储罐储量, 为保证调度结果的可重复性, 强制要求其相等。

系统电、热、碳供需平衡约束如式(16)~式(18)所示:

$$P_t + P_{\text{imb},t} - P_{\text{comp},t} = L_{\text{elec},t} \quad (16)$$

$$H_t + H_{\text{imb},t} = L_{\text{heat},t} \quad (17)$$

$$m_{\text{cap},t} + m_{\text{cap,imb},t} = L_{\text{cap},t} \quad (18)$$

为满足系统低碳运行要求, 同时扩大优化空间, 设计如式(19)所示的日平均碳捕集率约束:

$$\eta_{\text{set}} \leq \frac{\sum_{t=1}^{24} m_{\text{cap},t}}{\sum_{t=1}^{24} (k_{\text{cap}} \cdot m_{\text{coal},t})} \quad (19)$$

2.3 TD3 智能优化调度框架

2.3.1 交互元素设计

针对以式(15)所示为目标函数, 以式(9)~式(19)为约束条件的 CFPU-PCC 系统数学规划调度框架, 需将其转化为马尔科夫决策过程(MDP), 以

实现基于 DRL 算法的优化调度问题求解。MDP 包含的智能体与环境的交互元素包括状态空间 S、动作空间 A、奖励函数 R 和状态转移概率 P^[22]。其中, S 包含智能体感知的环境信息, 反映系统运行状态; A 包含可供智能体执行的调度动作, 直接影响优化目标达成; R 用于量化环境对动作的反馈评估, 内嵌目标函数、约束条件, 建立从算法到优化目标的桥梁; P 表征环境在动作执行后转移到下一状态的概率, 反映不确定性因素, 由环境内在特性决定。

本文设计如下动作、状态和奖励函数实现燃煤机组-碳捕集系统的优化调度:

(1) 状态空间 S。状态的选取需全面反映系统运行状况, 且应与决策变量有直接关联。通常将负荷指令纳入 S, 而由于各优化时刻下负荷指令本身数值变化范围大、分布存在不确定因素, 将导致 S 具有高维、变化剧烈特征, 因而会对智能体特征提取造成不利影响, 寻优能力降低。鉴于此, 本文将存在较强不确定性的电、碳负荷指令拆分为“预测基准”和“修正偏差”两个维度, 前者为负荷指令预测值, 反映其变化的总体趋势; 后者为日内实际指令与日前预测值间的偏差, 表征实际负荷指令的不确定特征, 以此在训练过程中引导智能体将复杂状态分解, 提升不确定因素下的调度鲁棒性。因此, 设计如式(20)所示的状态空间 S:

$$S = [t, L_{\text{elec},t}^{\text{pre}}, L_{\text{elec},t}^{\text{bias}}, L_{\text{cap},t}^{\text{pre}}, L_{\text{cap},t}^{\text{bias}}, L_{\text{heat},t}, U_t] \quad (20)$$

(2) 动作空间 A。选取 CFPU-PCC 系统优化调度的决策变量作为智能体动作, 包括 t 时刻给煤量、碳捕集抽汽流量及供热抽汽流量, 组成如式(21)所示的动作空间 A:

$$A = [m_{\text{coal},t}, st_{\text{cap},t}, st_{\text{heat},t}] \quad (21)$$

(3) 奖励函数 R。综合考虑调度目标和约束条件, 构造如式(22)所示的奖励函数:

$$r_t = -k_{\text{ope}} F_{\text{all},t} - k_{\text{vio}}^{\text{step}} r_{\text{vio},t}^{\text{step}} - k_{\text{vio}}^{\text{epi}} r_{\text{vio},t}^{\text{epi}} \quad (22)$$

需指出的是, k_{ope} 为系统运行性能的权重系数, $k_{\text{vio}}^{\text{step}}$ 和 $k_{\text{vio}}^{\text{epi}}$ 分别为每步(1 h)及每回合(24 h)约束越界的惩罚系数。

在每个优化时刻, 若智能体动作执行结果违反约束式(9)~式(18), 向奖励函数中添加如式(23)所示的步约束越界惩罚:

$$r_{vio,t}^{step} = \alpha \quad (23)$$

其中, α 的取值根据智能体训练收敛情况进行调试。若智能体动作执行结果未违反式(9)~式(18)中的任一约束, 则有 $r_{vio,t}^{step} = 0$ 。

在每个优化周期末时刻, 若智能体动作执行结果违反日平均碳捕集率约束式(19), 向奖励函数中添加如式(24)所示的回合约束越界惩罚:

$$r_{vio,t}^{epi} = \eta_{set} - \frac{\sum_{t=1}^{24} m_{cap,t}}{\sum_{t=1}^{24} (k_{CO_2} \cdot m_{coal,t})} \quad (24)$$

若智能体动作执行结果未违反约束式(19), 则有 $r_{vio,t}^{epi} = 0$ 。

2.3.2 TD3 智能体训练流程

通过大量回合的决策-状态感知-奖励-反馈的闭环迭代训练, DRL 算法将得到解决上述复杂

MDP 问题的最优策略, 以实现对目标调度优化问题的求解, 本文采用 TD3 算法执行该训练过程。TD3 是一种基于演员-评论家(Actor-Critic)框架构建的用以解决连续动作空间中复杂决策问题的无模型深度强化学习算法。其中, 演员负责学习参数化的策略函数, 评论家通过学习状态价值函数预测特定策略下从当前状态开始可获得的预期总回报, 即预测未来累计奖励^[23]。该算法以深度确定性策略梯度算法(DDPG)为基础, 做出 3 点主要改进:(1) 使用 2 套截断的评估网络(Critic)估算 Q 值(动作价值), 取较小者作为更新目标, 以抑制 Q 值过估计;(2) 延迟策略网络(Actor)参数的更新频率, 确保 Critic 训练充分、稳定;(3) 在目标动作中加入噪声, 平滑 Q 函数沿不同动作的变化, 稳定 Critic 的训练波动^[24]。算法训练原理如图 4 所示。

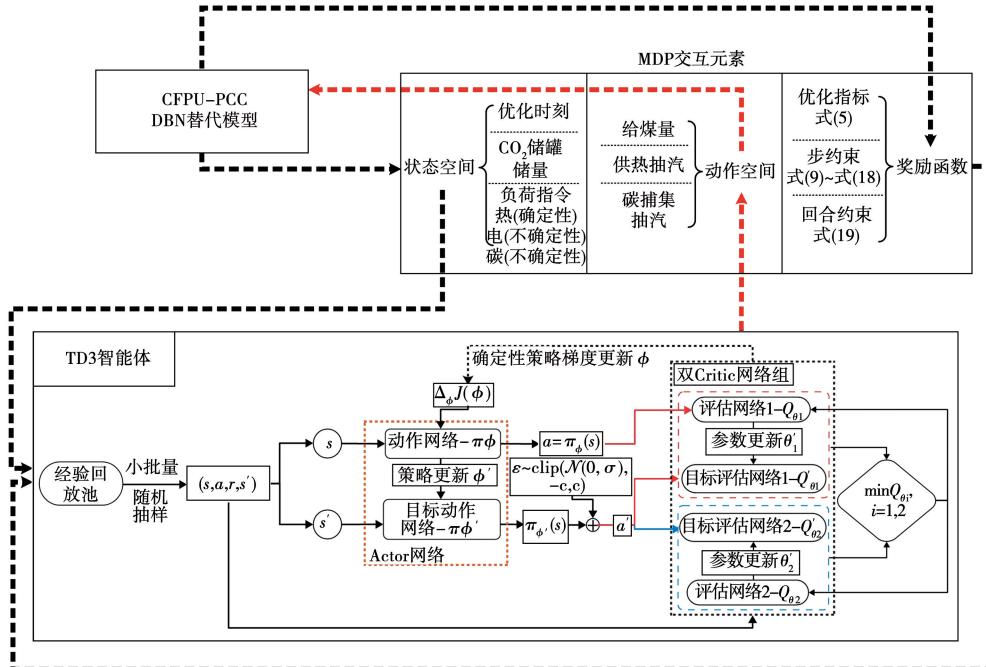


图 4 TD3 算法训练原理图

Fig. 4 Training schematic diagram of TD3 algorithm

算法的训练流程^[25]为:

- (1) 随机初始化 2 个评估网络的参数(θ_1, θ_2)和策略网络的参数(ϕ);
- (2) 初始化上述网络对应的目标网络的参数($\theta'_1, \theta'_2, \phi'$);
- (3) 初始化经验回放池 B;

(4) 执行从 $t=1$ 到 $t=T$ (回合长度, 即调度周期)为周期的循环训练:

- ①智能体根据初始(或当前)状态 s 选择动作 a 并加入探索噪声 $a \sim \pi(s) + \varepsilon$, 其中 $\varepsilon \sim N(0, \sigma)$, 进而得到奖励 r , 反馈至环境, 得到新状态 s' ;
- ②储存上述过程中的元素 (s, a, r, s') 至 B;

③从B中对过程元素 (s, a, r, s') 进行小批量随机采样,计算目标动作,并加入裁剪噪声:

$$\tilde{a} \leftarrow \pi_{\phi}(s) + \varepsilon, \varepsilon \sim \text{clip}(N(0, \tilde{\sigma}), -c, c) \quad (25)$$

计算目标值:

$$y \leftarrow r + \gamma \min_{i=1,2} Q_{\theta_i}(s', \tilde{a}) \quad (26)$$

按梯度下降原理更新评估网络参数:

$$\theta_i \leftarrow \min_{i=1,2} N^{-1} \sum (y - Q_{\theta_i}(s, a))^2 \quad (27)$$

当 t 可被 d 整除时(延迟 d 步执行策略更新),根据确定性策略梯度(DPG)更新策略网络参数 ϕ :

$$\nabla_{\phi} J(\phi) = N^{-1} \sum \nabla_a Q_{\theta_1}(s, a) |_{a=\pi_{\phi}(s)} \nabla_{\phi} \pi_{\phi}(s) \quad (28)$$

更新各目标网络参数:

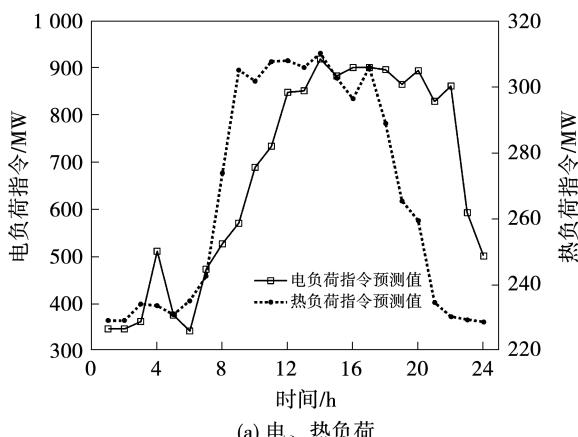
$$\theta'_i \leftarrow \tau \theta_i + (1 - \tau) \theta'_i \quad (29)$$

$$\phi' \leftarrow \tau \phi + (1 - \tau) \phi' \quad (30)$$

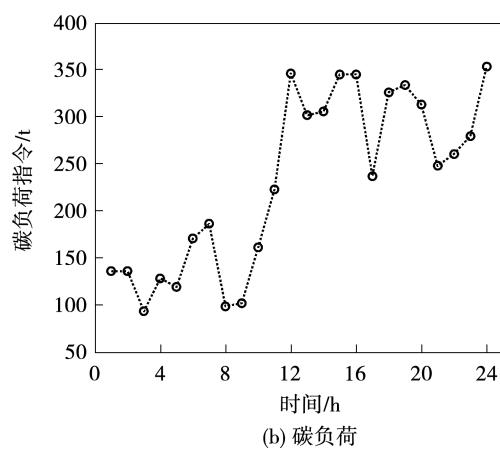
依据上述训练流程,在MDP框架下,TD3智能体通过感知当前时刻CFPU-PCC系统状态,执行动作,获得即时奖励,并使系统转移至新状态。之后将每步交互得到的[状态,动作,奖励,新状态]元组存储于经验回放池,在后续训练中通过小批量随机采样利用过去经验,以消除经验数据关联,进而通过并行策略网络/评估网络训练持续提升智能体对系统优化调度问题的决策能力。

3 算例分析

本节通过算例对比与分析,验证所提调度方法的有效性和优越性。算例分析中,作为调度优化计算的边界条件确定性场景下(算例1)的电、热、碳负荷预测值如图5所示。



(a) 电、热负荷

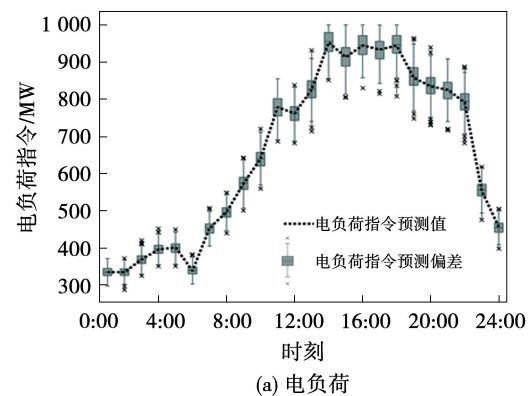


(b) 碳负荷

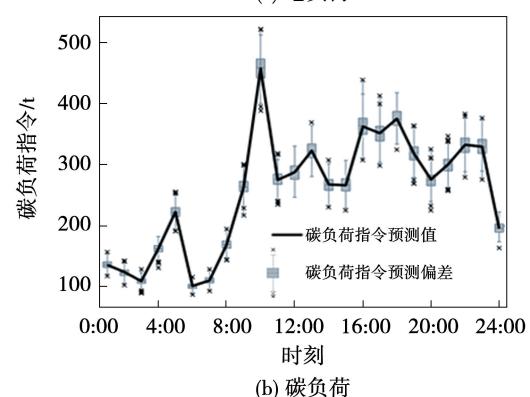
图5 确定性调度预测负荷指令

Fig. 5 Load forecast commands of deterministic scheduling

不确定性场景下的(算例2,3),考虑来自天气、市场、用户等方面短期随机性影响,假设电、碳负荷指令存在不确定性(热负荷以季节为周期变化,随机性较小,假设仍为确定性指令),其预测值及偏差(用每时刻下的箱型图表示)如图6所示。



(a) 电负荷



(b) 碳负荷

图6 不确定性调度预测负荷指令及偏差

Fig. 6 Load forecast commands and deviations of uncertain scheduling

此外,调度目标、约束条件中涉及的参数如表4所示。TD3训练采用的超参数如表5所示。表4

中,煤炭价格参考自文献[26],为2021年国内动力煤市场价;考虑电、热供应为煤电机组运行的主要任务,均需满足用户负荷需求,故设置相同的惩罚权重。

表4 调度计算所用参数

Tab. 4 Parameters used in optimal scheduling calculation

名称	参数
煤炭成本 ^[25] $C_{\text{coal}}/(\text{元}\cdot\text{t}^{-1})$	1 500
机组发电单位运维成本 $C_{\text{elec}}/(\text{元}\cdot\text{MW}^{-1})$	40
机组供热单位运维成本 $C_{\text{heat}}/(\text{元}\cdot\text{MW}^{-1})$	40
碳捕集单位运维成本 $C_{\text{cap}}/(\text{元}\cdot\text{t}^{-1})$	9.62
供热不足单位惩罚成本 $C_{\text{heat}}^{\text{imb}}/(\text{元}\cdot\text{MW}^{-1})$	1 500
碳捕集不足单位惩罚成本 $C_{\text{cap}}^{\text{imb}}/(\text{元}\cdot\text{t}^{-1})$	60
发电不足单位惩罚成本 $C_{\text{elec}}^{\text{imb}}/(\text{元}\cdot\text{MW}^{-1})$	1 500
日平均碳捕集率最低限值 $\eta_{\text{set}}/\%$	60
CO_2 储罐储存量下限 U_{\min}	50
CO_2 储罐储存量上限 U_{\max}	2 800
系统运行性能权重系数 k_{ope}	0.000 1
步约束越限惩罚系数 $k_{\text{vio}}^{\text{step}}$	1
步约束越限惩罚值 α	10
回合约束越限惩罚系数 $k_{\text{vio}}^{\text{epi}}$	10
碳捕集率上限 η_{\max}	0.3
碳捕集率下限 η_{\min}	0.9

表5 TD3 算法超参数设置

Tab. 5 Hyperparameter setting for TD3 algorithm

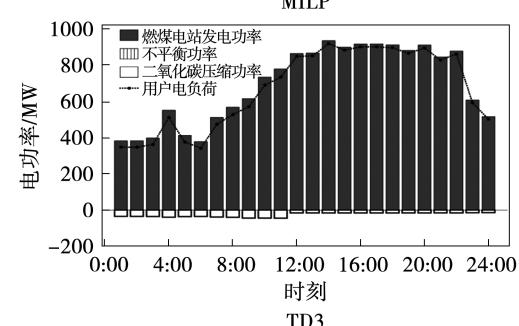
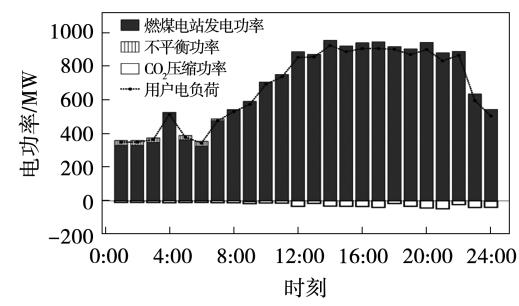
参数	数值
策略网络学习率	0.001
评估网络学习率	0.000 1
折扣因子 γ	0.995
经验回放池容量	2 000 000
经验回放池采样批量大小	512
目标平滑因子	0.005
目标更新频率	10

3.1 验证 TD3 算法处理系统非线性特性的优势

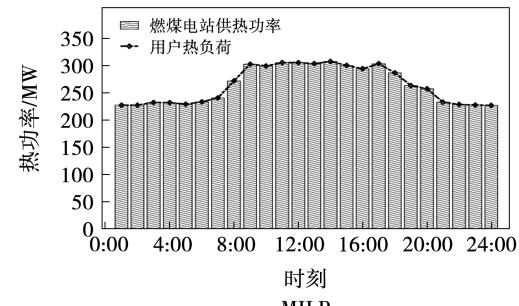
为验证 TD3 算法在处理 CFPU-PCC 变工况运行非线性特性上的优势设置算例 1,将本文所提 TD3 算法与基于线性模型的混合整数线性规划(MILP)算法进行对比,在确定性场景下求解 CFPU-PCC 系统的优化调度问题。MILP 调度的目标函数、

约束设置见式(5)~式(19)。

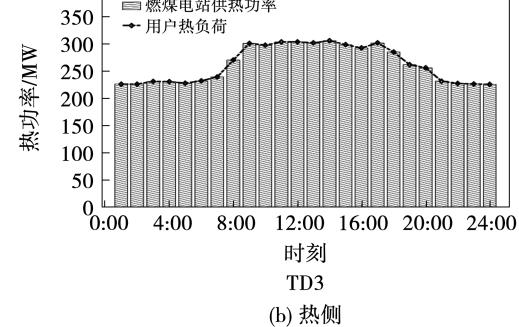
两种调度方法下,CFPU-PCC 系统的电、热、碳出力及碳储罐情况如图 7 所示。



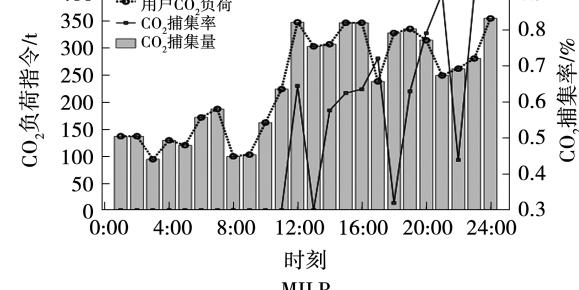
(a) 电侧



MILP



TD3



MILP

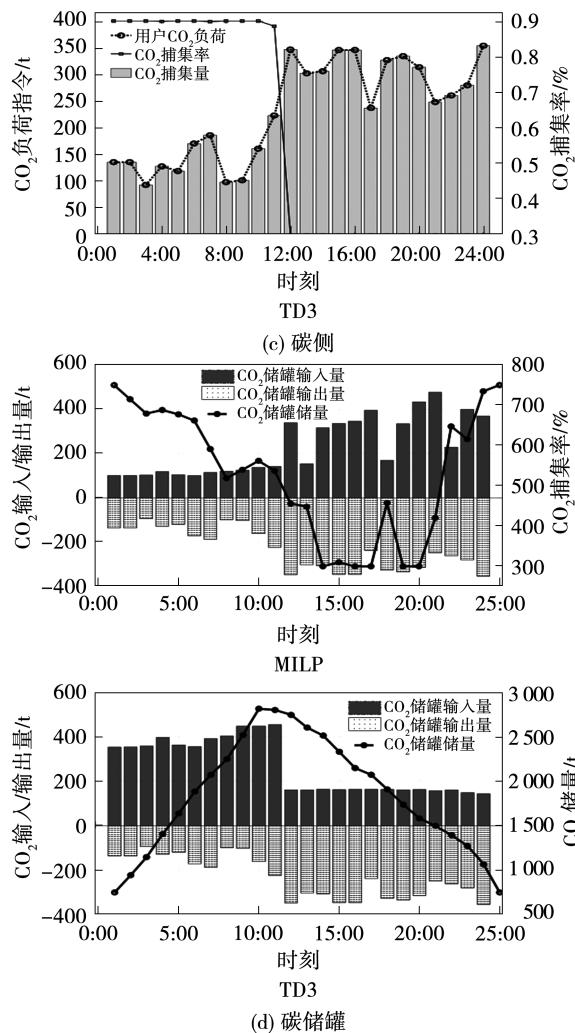


图7 算例1中燃煤机组耦合碳捕集系统调度结果

Fig. 7 Scheduling results of CFPU-PCC in Example I

由图7(a)可知,在电侧,MILP方法在1:00~7:00时段出现明显的电功率不平衡现象,引起额外惩罚。原因在于,在电负荷低谷时段,发电机组运行于低负荷工况,发电效率较低,而MILP基于额定工况的线性模型开展优化,无法充分反映机组宽负荷区间运行的非线性特性,过高估计了机组的发电功率,导致机组发电量无法满足负荷需求。而TD3通过智能体与非线性DBN代理模型交互,能准确识别机组宽工况范围内的运行状态,结合调度指令有效调整机组出力,调度周期内不平衡功率几乎为零。由图7(b)和7(c)可知,热侧和碳侧并未出现不平衡现象,这主要是因为负荷指令波动较小,且CO₂储罐对碳侧供需不平衡有缓冲作用。结合图7(c)和7(d)可知,TD3调度方法使碳捕集系统在电负荷低

谷时以高捕集率运行,在储罐中积累大量CO₂,而在电负荷高峰时段降低捕集率,通过此种运行模式实现碳捕集任务时移,在满足式(19)的前提下充分利用储罐的灵活性,以保障峰时电侧出力的充足与稳定。而在MILP调度方法下,在电负荷低谷时碳捕集率最低,依靠释放储罐中的CO₂满足用户需求,导致电负荷高峰时,CO₂存量无法满足碳负荷需求,捕集率被动提高,影响了电侧顶峰出力。

TD3和MILP方法下系统的经济性指标对比如表6所示。结合上述分析,TD3方法通过DBN模型交互实现了对发电机组宽工况运行状态的识别,相比MILP方法电功率不平衡惩罚降低了98.78%。尽管发电量增加导致机组燃料和维护成本有所增加,TD3调度方法下燃煤机组耦合碳捕集系统的日运行总成本仍较MILP方法降低了0.66%。

表6 算例1经济性指标对比

Tab. 6 Comparison of economic indicators in Example I

费用明细/万元	确定性 TD3	确定性 MILP
燃料成本	1 114.80	1 077.20
设备运维成本	72.71	70.41
出力不平衡惩罚	0.59	48.38
日运行总成本	1 188.10	1 195.99

综上所述,TD3算法通过智能体与DBN代理模型的交互学习,可在确定性场景中深入挖掘系统电、热、碳耦合非线性特性并形成宽工况运行下的自适应策略,在满足热侧需求的同时充分发挥碳侧灵活性,提高电侧负荷跟踪性能,实现系统协同运行优化。需要注意的是,尽管MILP可以获得线性模型下的全局最优解,但由于线性模型难以反映机组变工况运行的非线性特征,调度结果存在精度不足问题。

3.2 验证改进TD3算法处理不确定性的优势

为验证本文提出的不确定负荷指令分解策略在处理负荷不确定性上的优势设置算例2,将常规TD3算法与本文所提改进TD3算法进行对比。将图6所示的电、碳负荷预测不确定场景作为智能体离线训练场景,选取图5和图6中折线代表的电、碳负荷数据分别作为验证场景中的实际和预测负荷指

令,以检验智能体对 CFPU-PCC 系统负荷不确定性的处理效果。

两种方法下系统电功率日不平衡分布对比如图 8 所示。由图 8 可知,常规 TD3 方法下,机组在 8:00~19:00 时段供电不足,导致较多不平衡惩罚。该现象是由于在电负荷上升及高峰期的不确定性较大,而常规 TD3 方法在状态中仅考虑电负荷指令的实际值,难以充分学习掌握各时刻指令的不确定性特征,预测效果不佳。改进 TD3 方法通过将实际电负荷指令分解为预测指令与预测偏差之和,可深入学习各时刻指令的预测-实际偏差特征,使发电机组出力在负荷指令变化趋势的基础上对预测偏差进行修正,有效降低了供电不平衡惩罚。

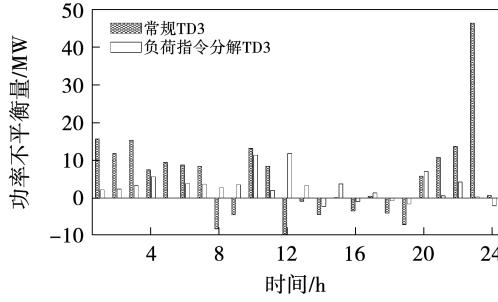


图 8 算例 2 两方法电功率不平衡情况对比

Fig. 8 Comparison of electrical power imbalances between the two methods in Example II

算例 2 经济指标对比如表 7 所示。由表 7 可知,改进 TD3 调度的不平衡惩罚相比常规 TD3 降低了 37.2%, 日运行总成本降低了 0.52%, 验证了所提策略改进 TD3 在 CFPU-PCC 系统负荷不确定场景调度问题求解中的优势。

表 7 算例 2 经济性指标对比

Tab. 7 Comparison of economic indicators in Example II

费用明细/万元	改进不确定 TD3	常规不确定 TD3
燃料成本	1 079.3	1 082.70
设备运维成本	70.4	70.67
出力不平衡惩罚	3.9	6.21
日运行总成本	1 153.6	1 159.58

3.3 验证 TD3 相比其他 DRL 算法处理不确定调度问题的优势

算例 3 应用 TD3、DDPG 和 SAC 3 种 DRL 算法

求解 3.2 节中的 CFPU-PCC 系统不确定优化调度问题,以验证 TD3 算法处理该问题的优势。为保证对比公平,3 种算法共有的训练超参数、网络深度和训练环境设为一致,且均采用本文所提不确定负荷指令分解策略。需指出的是,由于非线性优化无法保证全局最优,尽管 3 种算法采用相同的参数设置,取得的调度结果仍存在差异。

表 8 总结了各方法的经济性指标。由表 8 可知,本文所提策略改进 TD3 调度下,系统各项成本均最低,其不平衡惩罚相比改进 SAC 方法降低了 87.25%, 相比改进 DDPG 方法降低了 90.23%, 总成本相比上述算法分别降低了 4.43% 和 9.12%。这说明通过引入双评估网络、延迟策略更新、动作平滑机制,TD3 算法在处理 CFPU-PCC 系统调度问题时,相对其他 DRL 算法能探索得到更优的调度策略,获得更好的运行经济性和可靠性。

表 8 算例 3 经济性指标对比

Tab. 8 Comparison of economic indicators in Example III

费用明细/万元	改进 TD3	改进 SAC	改进 DDPG
燃料成本	1 079.3	1 104.2	1 154.4
设备运维成本	70.4	72.3	75.1
出力不平衡惩罚	3.9	30.6	39.9
日运行总成本	1 153.6	1 207.1	1 269.4

4 结 论

本文提出 1 种基于 TD3 算法的燃煤机组耦合碳捕集系统深度强化学习调度框架,结合负荷需求对机组电、热、碳指令进行优化分配,实现系统整体的经济、低碳、灵活运行。主要结论如下:

(1) TD3 调度方法的数据驱动特点,使其可以有效处理机组宽工况运行下的非线性问题。相比常规基于线性模型开展的 MILP 调度方法,功率不平衡惩罚可降低 98.78%,从而使得系统日运行总成本降低 0.66%;

(2) 所提复杂负荷状态分解策略有效提高了负荷指令不确定条件下智能体的学习能力,相比常规 TD3 调度,功率不平衡惩罚降低 37.2%,系统日运行总成本降低 0.52%;

(3) 相比 SAC、DDPG 等其他深度强化学习方

法,TD3 算法能够探索得到更经济、可靠的调度策略,使系统总运行成本分别降低 4.43% 和 9.12%。

参考文献:

- [1] 樊启祥,蒋敏华,许世森,等. 碳中和目标下新一代超高参数煤电机组技术体系 [J]. 中国电机工程学报, 2024, 44(18): 7167–7178.
FAN Qixiang, JIANG Minghua, XU Shisen, et al. Technical systems of advanced ultra-supercritical coal-fired power units under the carbon neutralization target [J]. Proceedings of the CSEE, 2024, 44(18): 7167–7178.
- [2] 孙泽远. 煤电低碳化改造背景下的 CCUS 产业发展机遇与挑战 [J]. 能源, 2024(10): 74–80.
SUN Zeyuan. Development opportunities and challenges for the CCUS industry in the context of low-carbon retrofits for coal-fired power plants [J]. Energy, 2024(10): 74–80.
- [3] 刘吉臻,杜忠明,王庆华,等. 新型电力系统下新一代煤电发展路径 [J]. 新型电力系统, 2024, 2(4): 357–370.
LIU Jizhen, DU Zhongming, WANG Qinghua, et al. Development paths for next-generation coal-fired power generation under new electric power systems [J]. New Type Power Systems, 2024, 2(4): 357–370.
- [4] 陈阿小,冉真真,马双忱,等. 火电厂碳捕集技术:现状、应用与发展 [J]. 洁净煤技术, 2024, 30(S1): 52–64.
CHEN Axiao, RAN Zhenzhen, MA Shuangchen, et al. Carbon capture technology in thermal power plants: Present situation, application and development [J]. Clean Coal Technology, 2024, 30(S1): 52–64.
- [5] 袁鑫,刘骏,陈衡,等. 碳捕集技术应用对燃煤机组调峰能力的影响 [J]. 发电技术, 2024, 45(3): 373–381.
YUAN Xin, LIU Jun, CHEN Heng, et al. Effect of carbon capture technology application on peak shaving capacity of coal-fired units [J]. Power Generation Technology, 2024, 45(3): 373–381.
- [6] GUO X, LIAO Y, LI G, et al. Low-carbon economic dispatch of photovoltaic-carbon capture power plant considering deep peak regulation [J]. Journal of Cleaner Production, 2023, 420: 138418.
- [7] ZANTYE M S, ARORA A, HASAN M M F. Renewable-integrated flexible carbon capture: A synergistic path forward to clean energy future [J]. Energy & Environmental Science, 2021, 14(7): 3986–4008.
- [8] LI X, ZHANG R, BAI L, et al. Stochastic low-carbon scheduling with carbon capture power plants and coupon-based demand response [J]. Applied Energy, 2018, 210: 1219–1228.
- [9] WU S, QI X, LI X, et al. Collaborative robust dispatch of electricity and carbon under carbon allowance trading market [J]. Global Energy Interconnection, 2024, 7(4): 391–401.
- [10] ZANTYE M S, ARORA A, HASAN M M F. Operational power plant scheduling with flexible carbon capture: A multistage stochastic optimization approach [J]. Computers & Chemical Engineering, 2019, 130: 106544.
- [11] WILSON Z T, SAHINIDIS N V. The ALAMO approach to machine learning [J]. Computers & Chemical Engineering, 2017, 106: 785–795.
- [12] CHEN X, WU X, LEE K Y. The mutual benefits of renewables and carbon capture: Achieved by an artificial intelligent scheduling strategy [J]. Energy Conversion and Management, 2021, 233: 113856.
- [13] 崔杨,郭福音,仲悟之,等. 多重不确定性环境下的综合能源系统区间多目标优化调度 [J]. 电网技术, 2022, 46(8): 2964–2975.
CUI Yang, GUO Fuyin, ZHONG Wuzhi, et al. Interval multi-objective optimal dispatch of integrated energy system under multiple uncertainty environment [J]. Power System Technology, 2022, 46(8): 2964–2975.
- [14] LI Z, DING Z, WANG M. Optimal bidding and operation of a power plant with solvent-based carbon capture under a CO₂ allowance market: A solution with a reinforcement learning-based Sarsa temporal-difference algorithm [J]. Engineering, 2017, 3(2): 257–265.
- [15] 刘倩,王馨,朱刘柱,等. 基于深度强化学习的园区级综合能源系统双层规划策略 [J/OL]. 武汉大学学报(工学版), 1–13[2025–07–13]. <https://link.cnki.net/urlid/42.1675.T.20240326.1452.002>.
LIU Qian, WANG Xin, ZHU Liuzhu, et al. A two-layer optimal planning strategy for park-level integrated energy system based on deep reinforcement learning [J/OL]. Engineering Journal of Wuhan University, 1–13[2025–07–13]. <https://link.cnki.net/urlid/42.1675.T.20240326.1452.002>.
- [16] 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.
- [17] LI J, HE X, LI W, et al. Low-carbon optimal learning scheduling of the power system based on carbon capture system and carbon emission flow theory [J]. Electric Power Systems Research, 2023, 218: 109215.

- [18] LIANG T, CHAI L, TAN J, et al. Dynamic optimization of an integrated energy system with carbon capture and power-to-gas interconnection: A deep reinforcement learning-based scheduling strategy [J]. *Applied Energy*, 2024, 367: 123390.
- [19] CHEN X, SHI Z, ZHANG Z, et al. Dynamic modeling and comprehensive analysis of an ultra-supercritical coal-fired power plant integrated with post-combustion carbon capture system and molten salt heat storage [J]. *Energy*, 2024, 308: 132961.
- [20] LI F, ZHANG J, SHANG C, et al. Modelling of a post-combustion CO₂ capture process using deep belief network [J]. *Applied Thermal Engineering*, 2018, 130: 997–1003.
- [21] CHEN X, WU X. The roles of carbon capture, utilization and storage in the transition to a low-carbon energy system using a stochastic optimal scheduling approach [J]. *Journal of Cleaner Production*, 2022, 366: 132860.
- [22] 张超. 基于深度强化学习的虚拟电厂实时优化调度研究 [D]. 北京: 华北电力大学, 2024.
ZHANG Chao. Research on real-time optimization scheduling of virtual power plants based on deep reinforcement learning [D]. Beijing: North China Electric Power University, 2024.
- [23] 史文龙. 数据与知识增强的多智能体深度强化学习电网实时调度 [D]. 北京: 中国电力科学研究院, 2024.
SHI Wenlong. Data and knowledge-enhanced multi-agent deep reinforcement learning for real-time power grid scheduling [D]. Beijing: China Electric Power Research Institute, 2024.
- [24] 钟皓俊, 王振雷. 基于双经验回放池 TD3 算法的 PID 参数优化 [J/OL]. 控制理论与应用, 1–9 [2025–01–18]. <http://kns.cnki.net/kcms/detail/44.1240.TP.20241024.1423.002.html>.
- [25] FUJIMOTO S, VAN HOOF H, MEGER D. Addressing function approximation error in actor-critic methods [J/OL]. CoRR, 2018: abs/1802.09477. <http://arxiv.org/abs/1802.09477>.
- [26] 陈昊. 政策干预下的动力煤现货价格预测方法研究 [D]. 北京: 北京邮电大学, 2023.
CHEN Hao. Research on forecasting method for thermal coal price under policy intervention [D]. Beijing: Beijing University of Posts and Telecommunications, 2023.

(刘颖 编辑)