

DOI: 10.19783/j.cnki.pspc.240303

# 基于行为克隆 TD3 强化学习的低碳园区柔性资源优化策略

舒展<sup>1</sup>, 孙旻<sup>1</sup>, 吴越<sup>1</sup>, 万子镜<sup>1</sup>, 段伟男<sup>2</sup>, 彭春华<sup>2</sup>

(1. 国网江西省电力有限公司电力科学研究院, 江西 南昌 330096;  
2. 华东交通大学电气与自动化工程学院, 江西 南昌 330013)

**摘要:** 园区作为我国产业聚集地, 是我国二氧化碳排放的重要来源, 优先在园区实现碳中和是助力我国达成“双碳”目标的重要举措。通过对园区综合能源系统供能侧加入电解槽-掺氢燃气轮机碳捕集进行低碳化改造, 同时考虑含储能侧、供能侧、用能侧多类型柔性资源构建低碳园区综合能源系统。为对该园区综合能源系统中各类柔性资源进行高效的在线低碳经济优化调度, 提出采用考虑行为克隆的 TD3 强化学习算法对低碳园区综合能源系统进行离线训练和在线优化求解。最后, 通过算例仿真验证了所提优化策略的优越性。

**关键词:** 园区综合能源系统; 多类型柔性资源; 强化学习; 行为克隆; 低碳经济调度

## Flexible resource optimization strategy for low-carbon parks based on behavioral cloning TD3 reinforcement learning

SHU Zhan<sup>1</sup>, SUN Min<sup>1</sup>, WU Yue<sup>1</sup>, WAN Zijing<sup>1</sup>, DUAN Weinan<sup>2</sup>, PENG Chunhua<sup>2</sup>

(1. State Grid Jiangxi Electric Power Research Institute, Nanchang 330096, China; 2. School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** Industrial parks in China are significant contributors to the country's carbon dioxide emissions. Prioritizing the achievement of carbon neutrality in parks is a crucial in helping China reach its 'dual-carbon' goal. This paper presents the construction of a low-carbon park integrated energy system. The system incorporates electrolyzers and hydrogen-blended gas turbines with carbon capture technology into the energy supply side, and considers various flexible resources on the storage, supply, and consumption sides. To efficiently optimize the low-carbon economic dispatch of various flexible resources in this integrated energy system, a TD3 reinforcement learning algorithm considering behavioral cloning is proposed for offline training and online optimization. Finally, the superiority of the proposed optimization strategy is verified through simulation examples.

This work is supported by the Science and Technology Project of the Headquarters of State Grid Corporation of China (No. 5400-202325227A-1-1-ZN).

**Key words:** park integrated energy system; multiple flexible resources; reinforcement learning; behavioral cloning; low-carbon economic dispatch

## 0 引言

园区作为我国产业及能耗聚集地, 是我国的碳排放主体, 据统计省级以上园区碳排放占全国碳排放的 1/3 以上<sup>[1]</sup>。因此, 为助力我国“双碳”目标的实现, 构建清洁低碳园区能源系统迫在眉睫。综合能源系统(integrated energy system, IES)作为一种能

实现低碳排放、高清洁能源消纳率的能源系统, 可为建立低碳园区提供支撑<sup>[2-4]</sup>, 助力我国“双碳”目标的实现。

目前已有一些国内外学者对低碳园区 IES 调度展开研究。文献[5-6]分析了碳捕集电厂的工作机理与数学模型, 并采用碳捕集燃气电厂捕获排放的 CO<sub>2</sub>, 有效地降低了系统整体碳排放; 文献[7-8]采用碳捕集电厂与电转气(power-to-gas, P2G)联合运行, 将碳捕集捕获的 CO<sub>2</sub> 供给 P2G, 降低了运行成本, 并提高了可再生能源消纳率。以上研究大多考

基金项目: 国家电网公司总部科技项目资助(5400-202325227A-1-1-ZN)

虑 P2G-燃气机组-碳捕集协同运行,以降低园区 IES 碳排放,并消纳园区内弃风弃光。但现有的 P2G-燃气机组-碳捕集协同运行过程中通常忽略了电转氢和掺氢燃气轮机的利用,而实际上电转氢相较于甲烷化具有更高的能源转化率,且已有研究表明掺氢燃气机组较普通燃气机组的碳排放更低<sup>[9-10]</sup>。此外,在园区 IES 调度中利用柔性资源也是降低园区碳排放的可靠途径。文献[11]通过在园区内考虑参数自适应阶梯型需求响应,有效提高了园区综合能源系统的经济性和低碳性;文献[12-13]通过在源侧考虑碳捕集措施和荷侧考虑需求响应措施,提高碳捕集灵活性,进一步降低了系统碳排放;文献[14]在可再生能源大量并网的背景下,考虑了电动汽车和空调的柔性资源潜能,从而建立了考虑响应意愿的电动汽车和空调集群需求响应模型,有效降低了系统的弃风光率。上述文献考虑了用能侧或供能侧柔性资源,但未同时考虑用能侧、供能侧、储能侧多种柔性资源在园区 IES 降碳潜能。综上分析,本文拟通过对园区 IES 供能侧加入电解槽-掺氢燃气轮机-碳捕集进行低碳化改造,同时考虑含储能侧、供能侧、用能侧多类型柔性资源构建低碳园区 IES。而如何充分考虑各类柔性资源对低碳园区 IES 进行高效的在线优化调度则是实现其低碳经济运行的关键。

近年来,强化学习作为能快速处理序列决策问题的高效方法,已在电力系统运行决策中被广泛应用。采用强化学习的方法一般可通过历史数据进行离线训练,并对训练好的模型进行在线优化,如文献[15-16]采用深度 Q 学习算法实现了电热 IES 的优化调度和孤岛微电网的负荷频率控制问题。但上述方法只能处理离散动作空间,若应用于更为复杂的系统中易出现维度爆炸问题,将导致训练结果的精度显著下降。因此,越来越多的学者提出将具有连续性决策能力的强化学习算法应用于 IES 调度中。如文献[17-18]通过采用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法实现 IES 的动态经济调度;文献[19]采用异步优势行动者-评论家算法对含大量可再生能源的 IES 系统实现了在线经济优化调度;文献[20]采用近端策略优化算法对含混合时间尺度调度的综合能源系统进行了离线训练和在线优化。然而,上述算法存在收敛速度和稳定性都较差的问题,随着综合能源系统快速发展其模型日趋复杂,已难以满足要求。因此,有学者提出将具有双 Q 函数和目标策略平滑的孪生延迟 DDPG(twin delayed deep deterministic policy gradient, TD3)算法应用于综合能源系统低碳经济调度中<sup>[21]</sup>。然而,现有 TD3 算法未能有效解决训练过程中的分

布偏移问题,若直接应用于本文的低碳园区 IES 多类型柔性资源优化调度中,易导致训练精度和稳定性下降。

综上分析,为实现低碳园区 IES 的低碳经济在线优化调度,本文提出了一种考虑多种柔性资源的低碳园区 IES 强化学习调度方法。通过在供能侧考虑电转氢和含碳捕集的掺氢燃气机组协同运行,并综合考虑多种柔性资源,含蓄电池、电动汽车(electric vehicle, EV)、电/热/冷柔性负荷,以充分调动园区 IES 的降碳潜能;同时,鉴于行为克隆 TD3 算法在策略更新中考虑加权行为克隆损失项可有效解决 TD3 算法存在的分布偏移问题,且通过对状态量归一化处理可提高训练的稳定性<sup>[22]</sup>,本文提出采用行为克隆 TD3 算法实现低碳园区柔性资源优化及在线实时决策。

### 1 低碳园区综合能源系统运行架构

图 1 为低碳园区 IES 运行架构。其中能源供给侧包含上级电网、天然气源和风光,能源转化设备包含:EV 充电站、电锅炉(electric boiler, EB)、电解槽(electrolyzer, EL)、电制冷(electric chillers, EC)、碳捕集(carbon capture, CC)等耗电设备;掺氢燃气轮机(hydrogen-doped gas turbines, HGT)、掺氢燃气锅炉(hydrogen-doped gas boilers, HGB)等耗气设备;耗热设备吸收式制冷机(absorption chillers, AC);储能设备蓄电池。在园区负荷侧包含电负荷、热负荷和冷负荷。

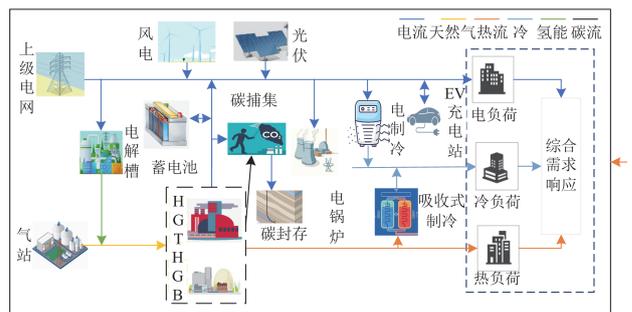


图 1 低碳园区综合能源系统运行架构

Fig. 1 Low-carbon park IES operation architecture

#### 1.1 风光-电解槽-掺氢燃气-碳捕集电厂子系统

本文将碳捕集设备、掺氢燃气机组、电制氢设备聚合为风-光-电解槽-掺氢燃气-碳捕集电厂子系统,如图 2 所示。其中 EL 将系统风光出力转换为氢能,并直接通过管道掺入供气管道供给掺氢燃气机组,从而降低机组碳排放。

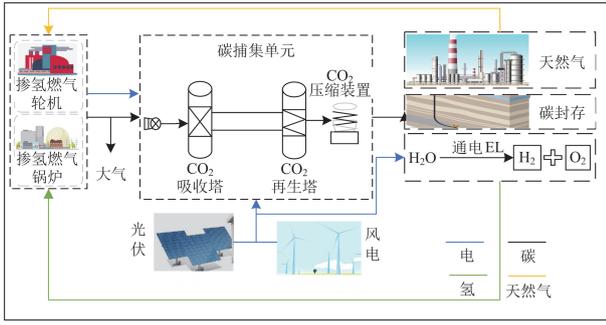


图 2 风-光-电解槽-掺氢燃气-碳捕集电厂子系统

Fig. 2 Wind-solar-EL-HG-CC subsystem

### 1.1.1 子系统能耗和出力

在该子系统中 EL 装置消耗系统电能, 为系统提供氢能, 而 HGT、HGB 装置主要消耗天然气或掺氢天然气, 为系统提供电/热功率。其中, 碳捕集装置能耗主要由 HGT 和风光提供。子系统的能耗和出力数学模型如式(1)所示。

$$\begin{cases} P_{\text{hgt}}^t = P_{\text{hgt,g}}^t + P_{\text{hgt,h}}^t \\ P_{\text{hgb}}^t = P_{\text{hgb,g}}^t + P_{\text{hgb,h}}^t \\ P_{\text{cap}}^t = P_{\text{A}}^t + P_{\text{D}}^t \\ P_{\text{hgt,c}}^t = \eta_{\text{hgt,c}} P_{\text{hgt}}^t \\ P_{\text{el}}^t = P_{\text{wvh}}^t \\ P_{\text{el,h}}^t = \eta_{\text{el}} P_{\text{el}}^t \\ H_{\text{hgt}}^t = \eta_{\text{hgt,h}} P_{\text{hgt}}^t \\ H_{\text{hgb}}^t = \eta_{\text{hgb}} P_{\text{hgb}}^t \end{cases} \quad (1)$$

式中:  $P_{\text{hgt}}^t$ 、 $P_{\text{hgb}}^t$  分别为 HGT 和 HGB  $t$  时段总能耗;  $P_{\text{hgt,g}}^t$ 、 $P_{\text{hgb,g}}^t$  分别为 HGT 和 HGB  $t$  时段消耗的天然气;  $P_{\text{hgt,h}}^t$ 、 $P_{\text{hgb,h}}^t$  分别为 HGT 和 HGB  $t$  时段消耗的氢气;  $P_{\text{cap}}^t$  为碳捕集设备  $t$  时段总能耗;  $P_{\text{A}}^t$ 、 $P_{\text{D}}^t$  分别为碳捕集设备  $t$  时段运行能耗、固定能耗;  $P_{\text{hgt,c}}^t$ 、 $H_{\text{hgt}}^t$  分别为 HGT 机组  $t$  时段电出力、热出力;  $H_{\text{hgb}}^t$  为 HGB 机组  $t$  时段热出力;  $P_{\text{el}}^t$  为 EL 装置  $t$  时段电能耗;  $P_{\text{el,h}}^t$  为 EL 装置  $t$  时段提供的氢能;  $P_{\text{wvh}}^t$  为风光  $t$  时段提供给 EL 的电功率;  $\eta_{\text{el}}$  为 EL 装置制氢效率;  $\eta_{\text{hgt,c}}$ 、 $\eta_{\text{hgt,h}}$  分别为 HGT 机组的电效率、热效率;  $\eta_{\text{hgb}}$  为 HGB 机组热效率。

该子系统中一部分风光和 HGT 机组功率供给碳捕集设备, 另外一部分功率输入主系统中, EL 机组消纳部分风光, 其数学模型如式(2)所示。

$$\begin{cases} P_{\text{wvc}}^t + P_{\text{hgt,cc}}^t = P_{\text{cap}}^t \\ P_{\text{A}}^t = \omega_{\text{c}} M_{\text{CC}}^t \\ P_{\text{wvc}}^t + P_{\text{wvh}}^t + P_{\text{wvl}}^t = P_{\text{wv}}^t \\ P_{\text{hgt,l}}^t + P_{\text{hgt,cc}}^t = P_{\text{hgt,e}}^t \end{cases} \quad (2)$$

式中:  $P_{\text{wvc}}^t$  为风光提供的碳捕集能耗;  $P_{\text{hgt,cc}}^t$  为 HGT 机组  $t$  时段提供的碳捕集能耗;  $M_{\text{CC}}^t$  为  $t$  时段系统处理的  $\text{CO}_2$  总量;  $\omega_{\text{c}}$  为碳捕集设备处理  $\text{CO}_2$  单位能耗;  $P_{\text{wvl}}^t$  为  $t$  时段风光上网功率;  $P_{\text{wv}}^t$  为  $t$  时段风光总出力;  $P_{\text{hgt,l}}^t$  为  $t$  时段 HGT 机组上网功率。

### 1.1.2 子系统碳排放和燃气掺氢

本系统中碳排放主要是上级电网购电涉及的虚拟碳排放和掺氢燃气机组消耗天然气的碳排放, 虚拟碳排放主要通过阶梯碳交易机制约束, 而掺氢燃气机组装置则可通过碳捕集装置直接捕获, 其具体捕获  $\text{CO}_2$  量  $M_{\text{CC}}^t$  的数学模型如式(3)所示。

$$\begin{cases} M_{\text{CC}}^t = \beta \delta^t (M_{\text{hgt}}^t + M_{\text{hgb}}^t) \\ M_{\text{hgt}}^t = \mu_{\text{hgt}} P_{\text{hgt,g}}^t \\ M_{\text{hgb}}^t = \mu_{\text{hgb}} P_{\text{hgb,g}}^t \end{cases} \quad (3)$$

式中:  $M_{\text{hgt}}^t$ 、 $M_{\text{hgb}}^t$  分别为 HGT 和 HGB  $t$  时段的碳排放;  $\beta$  为碳捕效率;  $\delta^t$  为烟气分流比;  $\mu_{\text{hgt}}$ 、 $\mu_{\text{hgb}}$  分别为 HGT 和 HGB 的碳排放系数。

此外, 通过采用 HGT 和 HGB 减少高碳能源的使用, 达到碳减排效果。选取掺氢比例  $\varepsilon^t$  不高于 20% 这一基数<sup>[23]</sup>, 其掺氢环节的数学模型如式(4)所示。

$$\begin{cases} V_{\text{gh}}^t = V_{\text{g}}^t (1 - \varepsilon^t) \\ P_{\text{gh}}^t = V_{\text{g}}^t \varepsilon^t \end{cases} \quad (4)$$

式中:  $V_{\text{g}}^t$ 、 $V_{\text{gh}}^t$  分别为  $t$  时段掺氢前后天然气消耗量;  $P_{\text{gh}}^t$  为  $t$  时段掺氢消耗的氢功率。

## 1.2 热/冷/储能运行单元模型

本文其他运行单元还包括制冷单元、制热单元以及储能单元。其中制冷单元包含电制冷和吸收式制冷设备; 制热单元包含电锅炉; 储能单元为蓄电池, 其能提高系统用能的灵活性, 因此该装置作为一种重要的柔性资源参与本园区的优化调度中。上述设备的数学模型见文献[24]。

## 1.3 阶梯碳交易模型

本文通过阶梯碳交易模型进一步约束园区碳排放量, 其具体阶梯碳交易模型见文献[25]。

## 2 EV 慢充负荷和综合需求响应

为进一步实现该园区的节能减排和调度灵活性目标,本文考虑了多类型柔性资源,包含储能、电动汽车和用户响应负荷。

### 2.1 电价引导 EV 有序充电

#### 2.1.1 EV 充电数学模型

美国交通部公布的美国家庭出行调查显示, EV 到达园区充电站的时间可近似用正态分布表达,其到达时间的概率密度函数  $f_s(t)$  如式(5)所示。

$$f_s(t) = \frac{1}{\sigma_{ev,s} \sqrt{2\pi}} \exp \left[ -\frac{(t_{ev} - \mu_{ev,s})^2}{2\sigma_{ev,s}^2} \right] \quad (5)$$

式中:  $t_{ev}$  为 EV 到达充电站的时间;  $\sigma_{ev,s}$  为 EV 到站时间拟合函数的标准差;  $\mu_{ev,s}$  为 EV 到站时间拟合函数的均值。

为拟合 EV 在园区的充电负荷,除 EV 到站时间外,还与 EV 每日的行驶里程和充电时长有关。EV 的日行驶里程也服从正态分布,其概率密度函数如式(6)所示。

$$f_d(x) = \frac{1}{x\sigma_{ev,d} \sqrt{2\pi}} \exp \left[ -\frac{(\ln x - \mu_{ev,d})^2}{2\sigma_{ev,d}^2} \right] \quad (6)$$

式中:  $x$  为 EV 日行驶里程;  $\sigma_{ev,d}$  和  $\mu_{ev,d}$  分别为 EV 日行驶里程的标准差和均值。

对于 EV 的充电时长,主要取决于 EV 充电前后的荷电状态,如式(7)所示。

$$T_{ev} = \frac{(S_{ev2} - S_{ev1})K}{P_{ev,z}\mu_{ev}} \quad (7)$$

式中:  $T_{ev}$  为 EV 的充电时长;  $S_{ev1}$ 、 $S_{ev2}$  分别为 EV 充电前后的荷电状态;  $K$  表示 EV 的电池容量;  $P_{ev,z}$  为充电桩功率;  $\mu_{ev}$  为充电效率。

其中充电后的实际荷电状态,与 EV 的行驶里程和初始荷电状态有关,如式(8)所示。

$$S_{ev2} = S_{ev1} + \frac{x E_{100}}{100K} \quad (8)$$

式中,  $E_{100}$  为 EV 的百公里耗电量。

根据式(5)~式(8),可采用蒙特卡洛法模拟出园区内无序充电状态下的充电负荷。

#### 2.1.2 分时电价引导 EV 有序充电

由于无序充电下的 EV 充电负荷较为集中,不利于系统的经济运行。因此,将 EV 纳入工业园区后,可基于分时电价引导 EV 充电负荷充电。EV 电池约束与 EV 有序充电功率约束数学模型如式(9)所示。

$$\left\{ \begin{array}{l} 0 \leq P_{EV}^{ch,t} \leq P_{EV}^{ch,max} \\ 0 \leq P_{EV}^{dc,t} \leq P_{EV}^{dc,max} \\ P_{EV}^t = P_{EV}^{dc,t} - P_{EV}^{ch,t} \\ P_{EV}^{ch,t} = \sum_{n=1}^N P_{EV,n}^{ch,t} \\ P_{EV}^{dc,t} = \sum_{n=1}^N P_{EV,n}^{dc,t} \\ R_{EV,n}^t = R_{EV,n}^{t-1} + P_{EV,n}^{ch,t} \eta_{EV}^{ch} - P_{EV,n}^{dc,t} / \eta_{EV}^{dc} \\ P_{EV,n}^{ch,t} \times P_{EV,n}^{dc,t} = 0 \\ R_{EV,n}^{min} \leq R_{EV,n}^t \leq R_{EV,n}^{max} \\ L + M \leq N_{max} \end{array} \right. \quad (9)$$

式中:  $P_{EV}^{ch,t}$ 、 $P_{EV}^{dc,t}$  分别为园区内 EV 在  $t$  时段的总充、放电功率;  $P_{EV,n}^{ch,t}$  和  $P_{EV,n}^{dc,t}$  分别为第  $n$  辆 EV 在  $t$  时段的充、放电功率;  $P_{EV}^t$  为园区内 EV 充电站在  $t$  时段的总充放电功率(大于 0 为放电,小于 0 为充电);  $P_{EV}^{ch,max}$ 、 $P_{EV}^{dc,max}$  分别为园区内 EV 充电站的最大允许充、放电功率;  $R_{EV,n}^t$ 、 $R_{EV,n}^{t-1}$  分别为园区内第  $n$  辆 EV  $t$  时段和  $t-1$  时段电池的荷电状态;  $R_{EV,n}^{max}$ 、 $R_{EV,n}^{min}$  分别为第  $n$  量 EV 电池容量上下限;  $\eta_{EV}^{ch}$ 、 $\eta_{EV}^{dc}$  分别为 EV 的充、放电效率;  $N$  为 EV 数量;  $N_{max}$  为园区内充电桩数量。

在满足充电需求前提下, EV 可作为一种柔性资源,在系统电负荷高峰时或电价高峰时释放一部分 EV 储能电池中的能量,缓解系统用能压力。

### 2.2 电热冷负荷综合需求响应

#### 2.2.1 价格型需求响应

本文中参与价格型需求响应的负荷主要为电负荷,电负荷需求响应模型如式(10)所示。

$$\left\{ \begin{array}{l} L_c^{DR} = L_c^{DR,0} + \Delta L_c^{DR} \\ \Delta L_c^{DR} = \sum_t \Delta l_{t,e} \\ \left[ \frac{\Delta l_{1,e}}{l_{1,e}^0}, \frac{\Delta l_{2,e}}{l_{2,e}^0}, \dots, \frac{\Delta l_{t,e}}{l_{t,e}^0} \right]^T = \mathbf{B}_p \left[ \frac{\Delta p_{1,e}}{p_{1,e}^0}, \frac{\Delta p_{2,e}}{p_{2,e}^0}, \dots, \frac{\Delta p_{t,e}}{p_{t,e}^0} \right]^T \end{array} \right. \quad (10)$$

式中:  $L_c^{DR,0}$ 、 $L_c^{DR}$  分别为需求响应前后的调度周期内的总负荷;  $T$  表示总调度周期;  $\Delta L_c^{DR}$  为总的可转移负荷变化量;  $\Delta l_{t,e}$ 、 $\Delta p_{t,e}$  分别为  $t$  时段可转移负荷变化量和电价变化量;  $l_{t,e}^0$  和  $p_{t,e}^0$  分别表示需求响应实施前第  $t$  时段的负荷量和电力价格;  $\mathbf{B}_p$  为电力价格弹性系数矩阵。

#### 2.2.2 激励型需求响应

激励型需求响应主要是 IES 管理商在系统负荷

高峰时期对用户采取补贴激励政策, 鼓励用户在此时段对部分负荷削减或者中断。

由于供热系统中, 供给用户热能主要采用回水管道, 其供给用户热负荷具有延迟性, 因此在考虑供热负荷时应当考虑供热系统的热惯性。当供热的实际温度在一定范围内即可满足用户的热需求, 因此在保证用户一定舒适度的前提下, 考虑用户热惯性后, 其供暖负荷模型如式(11)所示。

$$\mu_H P_H' \leq L_H' \leq P_H' \lambda_H \quad (11)$$

式中:  $L_H'$  为  $t$  时段系统供热负荷;  $P_H'$  为  $t$  时段系统热负荷需求;  $\mu_H$ 、 $\lambda_H$  为供热系统的惯性系数。

此外, 人体对于供暖温度的感知具有模糊性。为量化人体对于温度感知的情况, 采用预测热指标(predicted mean vote, PMV)来描述<sup>[21]</sup>, 其指标计算公式如式(12)所示。

$$M_{pmv} = 2.43 - \frac{3.76(T_s - T_n)}{\psi(R_f + 0.1)} \quad (12)$$

式中:  $M_{pmv}$  为预测热指标;  $\psi$  为人体能量代谢率, 取 0.2;  $R_f$  为服装热阻, 一般取 1clo;  $T_s$  为舒适状态下人体皮肤平均温度, 一般取 24 °C;  $T_n$  为室内温度。

由于日间用户活动较为频繁, 对温度变化的敏感度高于夜间, 对于舒适度的要求也较高, 本文 PMV 指标可设置如式(13)所示。

$$\begin{cases} |M_{pmv}| \leq 1, t \in [01:00-07:00] \cup [20:00-24:00] \\ |M_{pmv}| \leq 0.5, t \in [08:00-19:00] \end{cases} \quad (13)$$

根据式(12)即可求出考虑 PMV 的室内温度最大值  $T_{n,max}$  和最小值  $T_{n,min}$ 。由此得出系统考虑热惯性和 PMV 情况下的系统供热功率, 其中标准供热温度  $T_b$  取 24 °C。

$$\begin{aligned} \max(\mu_H P_H', (1 - \frac{T_b - T_{n,min}}{T_b}) P_H') &\leq L_H' \leq \\ \min(P_H' \lambda_H, (1 - \frac{T_b - T_{n,max}}{T_b}) P_H') & \end{aligned} \quad (14)$$

其中冷负荷也具有一定的供冷延迟性, 因此冷负荷供冷模型如式(15)所示。

$$\mu_C P_C' \leq L_C' \leq P_C' \lambda_C \quad (15)$$

式中:  $L_C'$  为  $t$  时段系统供冷负荷;  $P_C'$  为  $t$  时段系统冷负荷需求;  $\mu_C$ 、 $\lambda_C$  为供冷系统的惯性系数。

从上述公式中可知, 冷热负荷具有可调控的属性, 且该部分需求响应量不影响用户的正常使用, 无需补偿。为进一步增加园区的柔性资源, 本文考虑存在影响用户体验的可中断冷热负荷, 并对该部

分负荷采取补偿措施。其中电负荷的可中断负荷也会影响用户体验, 也应对其进行补偿。

$$0 \leq L_{z,id}' \leq L_{z,idm}' \quad (16)$$

式中:  $z \in \{e, H, C\}$ , e、H、C 分别表示电、热、冷;

$L_{z,id}'$  表示中断负荷量;  $L_{z,idm}'$  为最大中断负荷量。

冷/热/电中断负荷阶梯式价格补贴如式(17)所示。

$$\begin{cases} f_{IDR,z}' = \lambda_{id}^z L_{z,id}' \\ \lambda_{id}^z = \begin{cases} \lambda_{id,1}^z & 0 \leq L_{z,id}' \leq L_{z,id,1}' \\ \lambda_{id,2}^z & L_{z,id,1}' \leq L_{z,id}' \leq L_{z,idm}' \end{cases} \end{cases} \quad (17)$$

式中:  $f_{IDR,z}'$  为  $t$  时段可中断负荷补贴成本;  $\lambda_{id}^z$  为  $t$  时段补贴价格;  $\lambda_{id,1}^z$ 、 $\lambda_{id,2}^z$  为两种阶梯补贴价格;  $L_{z,id,1}'$  为负荷划分边界。

### 3 目标函数与约束条件

#### 3.1 目标函数

本文综合考虑园区运行成本和 CO<sub>2</sub> 相关成本, 构建低碳经济调度目标函数如式(18)所示。

$$\min F = \min \sum_{t=1}^T F^t = \min \sum_{t=1}^T (C_{op}^t + C_{co_2}^t) \quad (18)$$

$$\begin{cases} C_{op}^t = C_{om}^t + C_{waste}^t + C_{buy}^t + C_{id}^t \\ C_{co_2}^t = f_{carbon}^t + c_{seq} M_{CC}^t + C_{cap}^t \\ C_{om}^t = \sum_{n=1}^N c_{om}^n P_n^t \\ C_{waste}^t = \mu^{waste} P_{waste}^t \\ C_{buy}^t = (c_{buy,c}^t P_{grid}^t + c_{buy,g}^t P_{gas}^t) \\ C_{id}^t = f_{IDR,e}^t + f_{IDR,H}^t + f_{IDR,C}^t \\ C_{cap}^t = C^{FL} \frac{(1+r)^{N^Z} r}{365 \times 24 [(1+r)^{N^Z} - 1]} \end{cases} \quad (19)$$

式中:  $F^t$  为  $t$  时段系统总成本;  $C_{op}^t$ 、 $C_{co_2}^t$  分别为  $t$  时段系统运行成本和 CO<sub>2</sub> 相关成本;  $C_{om}^t$  为  $t$  时段园区内除碳捕集外所有设备总运行成本;  $C_{waste}^t$  为  $t$  时段弃风光惩罚成本;  $C_{buy}^t$  为  $t$  时段购电购气成本;  $C_{id}^t$  为  $t$  时段电/热/冷负荷削减补偿成本;  $c_{seq}$  为 CO<sub>2</sub> 封存成本;  $f_{carbon}^t$  为阶梯碳交易成本;  $C_{cap}^t$  为  $t$  时段碳捕集折旧成本;  $c_{om}^n$ 、 $P_n^t$  分别为  $t$  时段第  $n$  台设备的运行维护成本和出力;  $\mu^{waste}$  为弃风弃光惩罚系数;  $P_{waste}^t$  为  $t$  时段弃风光量;  $c_{buy,c}^t$ 、 $c_{buy,g}^t$  分别为  $t$  时段购电、购气价格;  $P_{grid}^t$  为  $t$  时段低碳园区购电量;  $P_{gas}^t$

为  $t$  时段低碳园区购气量； $C^{FL}$ 、 $N^Z$  分别为碳捕集设备总成本和折旧年限； $r$  为碳捕集电厂项目贴现率。

### 3.2 约束条件

#### 1) 常规机组约束

其中 HGT、HGB、EL、EB 机组约束见文献[9]，AC、EC、储能设备和风光出力约束见文献[24]。

#### 2) 柔性资源约束

本文的柔性资源主要指 EV 充电和冷热电负荷的需求响应量，其要满足式(9)一式(10)、式(14)一式(16)。

#### 3) 功率平衡约束

园区整体功率平衡公式如式(20)所示。

$$\begin{cases} P_{cap}^t + P_{cl}^t + P_{ec}^t + P_{cb}^t + P_e^t + L_{e,DR}^t = \\ P_{hgt,e}^t + P_{wv}^t + P_{grid}^t + L_{e,id}^t + P_{EV}^t \\ H_{ac}^t + L_H^t = H_{hgt}^t + H_{eb}^t + H_{hgb}^t + L_{H,id}^t \\ L_C = C_{ac}^t + C_{ec}^t + L_{C,id}^t \end{cases} \quad (20)$$

式中： $P_e^t$  表示储电充放动作状态量，大于 0 表示存储电能，反之为释放电能； $H_{eb}^t$  和  $P_{cb}^t$  分别为 EB 在  $t$  时段的热出力和电消耗； $C_{ac}^t$  和  $H_{ac}^t$  分别为 AC 机组在  $t$  时段的冷出力和消耗的热功率； $C_{ec}^t$  和  $P_{ec}^t$  分别为 EC 设备在  $t$  时段的冷出力和消耗的电功率； $L_{e,DR}^t$  为  $t$  时段价格型需求响应后的电负荷。

## 4 面向低碳园区 IES 的强化学习模型

强化学习模型中智能体与环境的交互可采用马尔可夫决策过程描述，该过程中主要信息包含状态空间  $S$ 、动作空间  $A$ 、奖励值  $R$  和折扣因子  $\gamma$ 。而其中的状态空间、动作空间与园区 IES 模型中的决策变量和一些已知量相关。

### 4.1 动作空间设计

动作空间表示动作变量在每个时间步长变化量的集合，动作变量包括：风电光伏、HGT 机组电出力、电锅炉、掺氢比、烟气分流比、AC 机组出力、蓄电池动作、电/冷/热削减负荷、考虑热惯性后的冷热负荷、EV 充放电电量。 $t$  时段动作空间集合表示为

$$A^t = \left[ \begin{array}{c} P_w^t, P_v^t, P_{hgt,e}^t, P_{cb}^t, \varepsilon^t, C_{ac}^t, \delta^t, C_{ac}^t, \\ P_e^t, L_{e,id}^t, L_{H,id}^t, L_{C,id}^t, L_C, L_H, P_{EV}^t \end{array} \right] \quad (21)$$

式中， $A^t$  表示为  $t$  时段动作空间集合。在该动作空间中的动作变量，应当满足 3.2 节中的各类约束条件。

### 4.2 状态空间设计

状态空间一般考虑会对决策产生影响的因素。

本文中状态空间包括下一时段的电负荷、热负荷、冷负荷、风光预测值、电价、气价、购电量以及上一时段的动作量，即

$$S^t = [L_{e,DR}^t, P_H^t, P_C^t, f_w^t, f_v^t, c_{buy,e}^t, c_{buy,g}^t, P_{grid}^t, A^t] \quad (22)$$

式中： $S^t$  表示为  $t$  时段状态空间集合； $f_w^t$ 、 $f_v^t$  分别为  $t$  时段风、光预测值。

### 4.3 奖励函数设计

强化学习通过智能体不断地学习更新自身策略，从而达到最大化期望奖励值。因此，奖励函数的设置应当与调度模型中的目标函数相关。奖励函数  $r^t$  如式(23)所示。

$$r^t = -gF^t + \kappa \quad (23)$$

式中： $g$  为系统总运行成本比重系数； $\kappa$  为修正参数，使得奖励函数趋向正值。

### 4.4 考虑行为克隆 TD3 强化学习

#### 1) 截断 Double Q-Learning 与目标策略平滑

在 TD3 算法中存在两个 Q 函数神经网络和两个目标 Q 函数神经网络用于跟踪主神经网络，提高算法训练的稳定性，这两个网络具有相同的结构，但其初始权重值不同。此外，还存在一个用来输出动作的主策略神经网络和一个用于跟踪主网络的目标策略神经网络。

在强化学习中，Q 函数值不是真实值而是通过 Q 函数神经网络拟合。因此，为避免高估 Q 值对智能体最优动作策略的更新产生影响，TD3 算法通过构建两个 Q 函数神经网络，选取两个网络估计拟合 Q 值的最小值，为策略更新的依据。

此外，Q 函数估计中对于值空间中的窄峰估计可能存在过拟合的问题，TD3 算法通过将目标动作周围的一小块区域的值进行模糊拟合，即在每个动作中加入截断的正态分布噪声作为正则化，以平滑 Q 值计算，其 Q 函数的贝尔曼公式如式(24)所示。

$$\begin{cases} Q(s^t, a^t) = \min(Q_{LB}^{(1,2)}(s^t, a^t)) \\ Q_{LB}^{(1,2)}(s^t, a^t) = r(s^t, a^t) + \gamma \left( \min \left( \begin{array}{c} \check{Q}_{LB}^1(s^{t+1}, a^{t+1}), \\ \check{Q}_{LB}^2(s^{t+1}, a^{t+1}) \end{array} \right) \right) \\ a^{t+1} = \pi_\theta(s^{t+1}) + \xi \end{cases} \quad (24)$$

式中： $s^t$ 、 $a^t$  分别为  $t$  时段状态量和动作量； $Q(s^t, a^t)$  为在状态对  $(s^t, a^t)$  下的 Q 值； $Q_{LB}^{(1,2)}(s^t, a^t)$  表示两个神经网络估计的 Q 值； $\check{Q}_{LB}^1(s^{t+1}, a^{t+1})$  与  $\check{Q}_{LB}^2(s^{t+1}, a^{t+1})$  分别为两个目标神经网络在状态对  $(s^{t+1}, a^{t+1})$  下计算的 Q 值； $\gamma$  为折扣因子，表示智能体在学习过程中对未来奖励的重视程度； $r(s^t, a^t)$

为在状态  $s^t$  下采取动作  $a^t$  所获得的奖励值； $\pi_\theta(s^{t+1})$  为策略神经网络在参数  $\theta$  下得到的策略； $\xi$  为加入的噪声，从正态分布中取样得到。

### 2) 考虑行为克隆的策略

在 TD3 中其策略  $\pi$  通过确定性策略梯度更新，如式(25)所示。

$$\pi = \arg \max_{\pi} \mathbb{E}_{(s^t, a^t) \sim D} [Q(s^t, \pi(s^t))] \quad (25)$$

式中： $\pi$  为 TD3 算法当前优化策略，其表示要最大化期望奖励值； $\mathbb{E}$  表示期望； $D$  为经验池，即离线数据集； $\pi(s^t)$  为在状态  $s^t$  下智能体的策略。

为了使得策略神经网络中输出动作尽可能接近数据集中的动作分布，避免分布偏移问题，考虑行为克隆 TD3 强化学习算法在 TD3 算法的策略更新基础上，加入了一个加权行为克隆损失项  $(\pi(s^t) - a^t)^2$ ，如式(26)所示。

$$\pi = \arg \max_{\pi} \mathbb{E}_{(s^t, a^t) \sim D} [\lambda Q(s^t, \pi(s^t)) - (\pi(s^t) - a^t)^2] \quad (26)$$

式中， $\lambda$  为超参数，控制了行为克隆损失项占比，若  $\lambda$  过大，算法可能会出现过拟合情况， $\lambda$  过小则可能出现欠拟合和分布偏移的问题。

此外为进一步提高算法的稳定性，避免分布偏移问题，对状态量进行归一化处理，如式(27)所示。

$$s_* = \frac{s - \mu}{\sigma + \varepsilon} \quad (27)$$

式中： $s_*$  为归一化后的状态量； $s$  为原始状态量； $\mu$ 、 $\sigma$  分别为状态数据集中的均值和标准差； $\varepsilon$  为归一化常数，本文取  $\varepsilon = 10^{-3}$ 。

### 3) 智能体网络参数更新

本文中的参数主要包括：主 Q 函数神经网络参数  $\omega_1$ 、 $\omega_2$ ；目标 Q 函数神经网络参数  $\omega_1^*$ 、 $\omega_2^*$ ；策略神经网络参数  $\theta$ ；目标策略神经网络参数  $\theta^*$ 。

两个 Q 函数神经网络参数  $\omega_1$ 、 $\omega_2$  可通过最小化贝尔曼残差  $J_Q(\omega_i)_{i \in \{1,2\}}$  方式更新，如式(28)所示。

$$J_Q(\omega_i)_{i \in \{1,2\}} = \mathbb{E}_{(s^t, a^t, s^{t+1}) \sim D, a^t \sim \pi_\theta} \left[ \frac{1}{2} (Q_{LB}^i(s_*^t, a^t) - r(s_*^t, a^t) - \gamma Q(s_*^t, a^t))^2 \right] \quad (28)$$

式中： $\omega_i$  为第  $i$  个 Q 函数神经网络参数； $Q_{LB}^i$  为第  $i$  个 Q 函数神经网络估计的 Q 值； $\pi_\theta$  为策略神经网络在参数  $\theta$  下得到的策略； $r(s_*^t, a^t)$  为状态动作对  $(s_*^t, a^t)$  下的奖励值； $s_*^t$ 、 $s_*^{t+1}$  分别为归一化后  $t$  时段和  $t+1$  时段的状态量。

$\omega_i^*$  为目标 Q 函数神经网络的参数，通过式(29)更新。

$$\omega_{i, i \in \{1,2\}}^* = \tau \omega_i + (1 - \tau) \omega_i^* \quad (29)$$

式中， $\tau$  为软更新系数。通过梯度下降法最小化函数  $J_\pi(\theta)$  实现策略神经网络参数  $\theta$  的更新，如式(30)所示。

$$J_\pi(\theta) = \mathbb{E}_{(s^t, a^t) \sim D} [\lambda Q(s_*^t, \pi_\theta(s_*^t)) - (\pi_\theta(s_*^t) - a^t)^2] \quad (30)$$

式中： $J_\pi(\theta)$  表示智能体寻找一个策略神经网络参数  $\theta$  下的最优策略  $\pi$ ，达到最大化期望奖励值； $Q(s_*^t, \pi_\theta(s_*^t))$  为 Q 函数神经网络在状态动作对  $(s_*^t, \pi_\theta(s_*^t))$  下的 Q 值。

目标策略神经网络参数  $\theta^*$  更新方式如式(31)所示。

$$\theta^* = \tau \theta + (1 - \tau) \theta^* \quad (31)$$

本文算法离线训练具体更新流程如下所述。

步骤 1: 初始化  $\omega_1$ 、 $\omega_2$ 、 $\omega_1^*$ 、 $\omega_2^*$ 、 $\theta$ 、 $\theta^*$ ；

步骤 2: 从状态空间初始化状态量  $s_0$ ，令时段  $t = 0$ ，训练次数  $k = 0$ ，经验池容量大于预训练次数；

步骤 3: 由策略噪声<sup>[22]</sup>来采样动作量  $a^t$ ；

步骤 4: 通过给出的动作量传入环境中更新状态量和奖励值，并将得到的  $(s^t, a^t, s^{t+1}, r^t)$  存储到经验池中，其中  $r^t$  为  $t$  时段的奖励；

步骤 5: 若经验样本量小于设定最大容量，则转至步骤 3，否则从经验池中随机采样  $(s^t, a^t, s^{t+1}, r^t)$ ；

步骤 6: 将状态量  $s^{t+1}$  输入策略神经网络得到  $a^{t+1}$ ；

步骤 7: 对状态量进行归一化处理得到  $(s_*^t, s_*^{t+1})$ ；

步骤 8: 输入  $(s_*^t, a^t, s_*^{t+1}, a^{t+1}, r^t)$  至智能体进行网络参数更新，更新两个 Q 函数神经网络参数  $\omega_1$ 、 $\omega_2$ ，再更新两个目标 Q 函数神经网络参数  $\omega_1^*$ 、 $\omega_2^*$  和策略神经网络参数  $\theta$ 、目标策略神经网络参数  $\theta^*$ ；

步骤 9: 若  $t$  未达到 23 时段调度周期，则进入下一个时间断面，令  $t = t + 1$ ，转至步骤 3；

步骤 10: 若已训练至最大训练次数  $K$ ，则训练终止，否则回到步骤 2，进入下一轮训练。

## 4.5 基于行为克隆 TD3 的优化调度流程

本文低碳园区 IES 强化学习优化调度采用“离线训练，在线调度”的模式。采用考虑行为克隆的 TD3 强化学习算法对低碳园区综合能源系统模型进行在线优化调度，具体强化学习优化调度流程如图 3 所示。其主要流程如下所述。

1) 将历史数据输入强化学习环境，初始化强化学习各网络参数；

2) 进行离线调度训练，从园区 IES 获取此时段的状态量，并采用式(27)进行归一化；

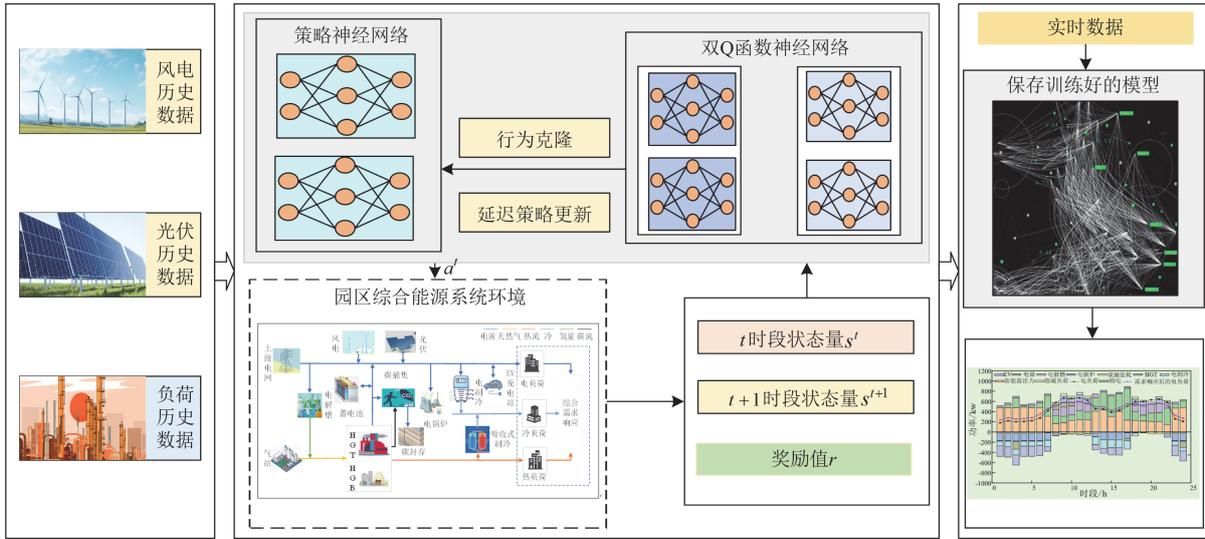


图3 低碳园区 IES 强化学习优化调度

Fig. 3 Low carbon park IES reinforcement learning optimal dispatch

- 3) 将归一化后的状态及其他量输入智能体中;
- 4) 智能体根据所输入的状态给出动作策略;
- 5) IES 根据动作, 更新下一时段的状态;
- 6) 当达到设定的训练次数, 保存模型; 将实时数据输入保存好的训练模型中, 该模型给出实时调度策略。

## 5 仿真算例分析

### 5.1 算例设置

以图 1 所示的低碳园区综合能源系统为例, 对各类柔性资源降碳、降成本的效果进行分析。园区综合能源系统中各单元参数如附录表 A1—表 A3 所示, 其中氢能和碳捕集相关设备参数参考文献[24]; 其余参数见文献[24, 26]。上级电网购电分时电价见表 1, 气站购气价格 0.512 元/kWh; 碳交易基价为 0.27 元/kg, 区间长度为 120 kg, 碳价增加率为 25%。其中低碳园区中的 EV 数量设置为 500 辆, 园区内总充电桩充电功率上限为 200 kW, 其通过蒙特卡洛法模拟的园区无序充电日负荷见图 4。典型风/光/负荷预测曲线如图 5 所示。冷热需求响应各参数见文献[27]。

表 1 分时电价

Table 1 Time of use power price

时段	时刻	电价/(元/kWh)
峰时段	08:00—12:00,	0.90
	18:00—22:00	
平时段	07:00,	0.55
	13:00—17:00	
谷时段	01:00—06:00,	0.295
	23:00—24:00	

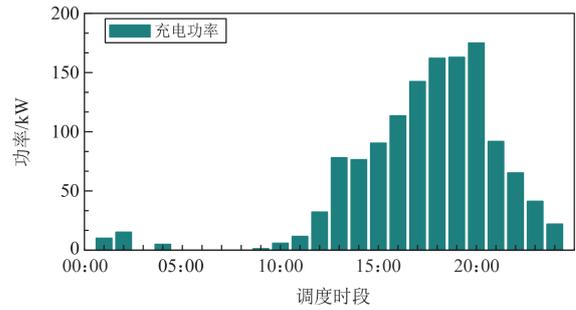


图 4 电动汽车无序充电日负荷

Fig. 4 Daily load of disorderly charging of electric vehicles

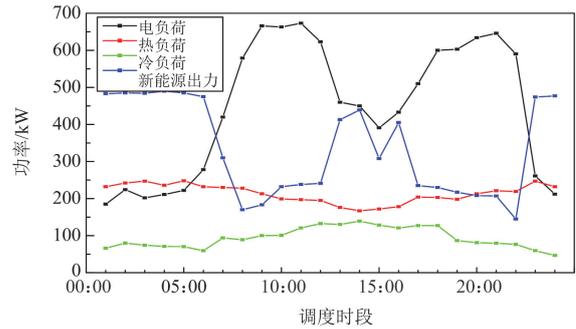


图 5 风/光/负荷预测曲线

Fig. 5 Forecast curves of wind power, PV and load

### 5.2 训练收敛性分析

算法中奖励折扣因子取 0.9, 神经元个数为 256, 噪声为 0.2, 延迟策略更新频率为 2。

由图 6 可知, 本文强化学习智能体在 1600 迭代次数后达到收敛, 最终奖励值稳定在 52 左右。在 500~600 次迭代中智能体奖励值有明显上升, 在

600~700 次迭代中又有明显下降, 这是因为在训练初期, 智能体寻优幅度较大后, 而后双 Q 网络的限制寻优空间变得较为保守, 出现奖励值下降。在 700~1600 次迭代中智能体奖励值逐步缓慢达到稳定。在稳定之后, 智能体的振荡幅度明显减少。

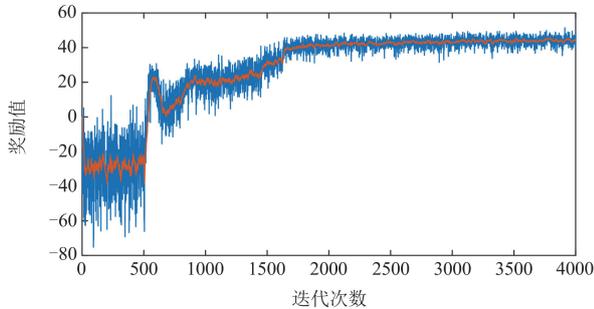


图 6 强化学习算法训练过程奖励值

Fig. 6 Reinforcement learning algorithm training process reward value

### 5.3 在线优化结果分析

通过训练好的强化学习模型测试如图 5 所示的风/光/负荷预测场景, 各设备的最终优化出力结果如图 7—图 9 所示。

图 7 为电负荷优化调度结果。由图 7 可知, 23:00—24:00 和 01:00—06:00 时段为风电出力高峰期, 且电负荷和电价处于低谷期, 系统电负荷供应主要由购电和新能源满足; 同时在该时段满足柔性资源 EV 和蓄电池的充电需求, 电锅炉和电制冷满功率输出。由图 10 可知, 在该时段, 烟气分流比和掺氢比都是处于最高位, 以减少系统碳排放。在 08:00—12:00 和 18:00—22:00 时段新能源出力较低且电负荷和电价处于高峰期, 此时 HGT 和购电满足系统的用电需求; 系统柔性资源储能和 EV 对系统放电, 以减轻系统用电压力; 同时需求响应对负荷进行一定程度的削减, 减少用电需求; 由于

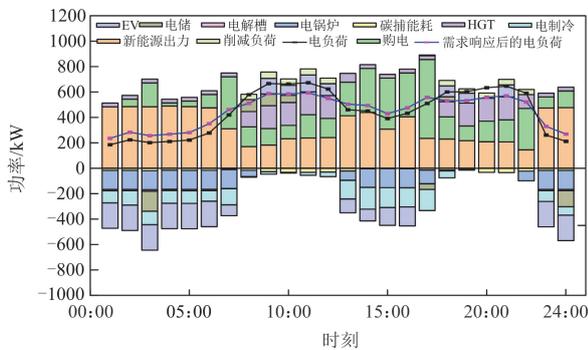


图 7 电负荷优化调度结果

Fig. 7 Optimal scheduling results of electricity load

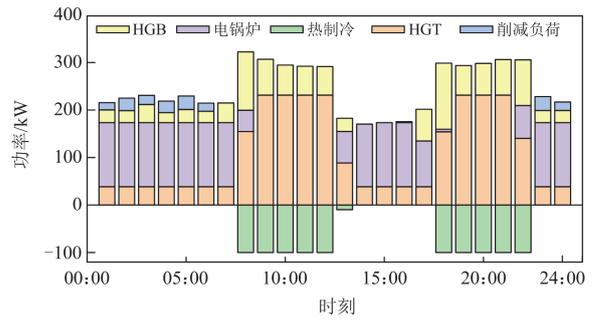


图 8 热负荷优化调度结果

Fig. 8 Optimal scheduling results of heat load

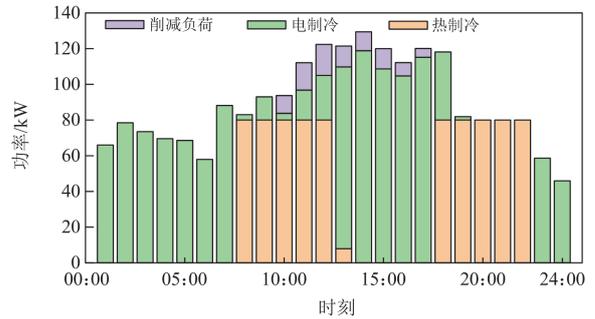
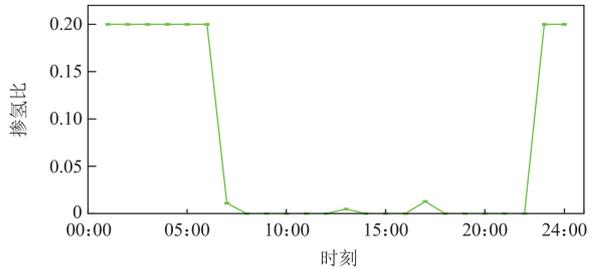
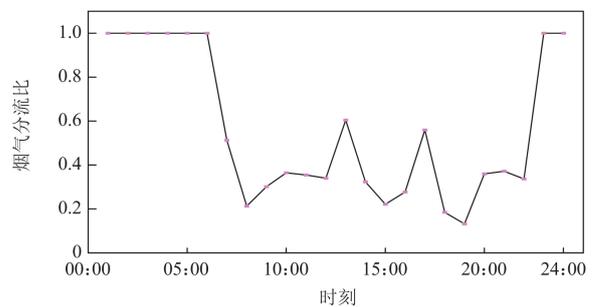


图 9 冷负荷优化调度结果

Fig. 9 Optimal scheduling results of cooling load



(a) 掺氢比



(b) 烟气分流比

图 10 掺氢与烟气分流比

Fig. 10 Hydrogen blending to flue gas split ratio

系统用电需求和购电成本高, 系统已无多余电能供给碳捕集设备和 EL 设备, 烟气分流和掺氢比急剧

下降(如图 10 所示), 导致系统碳排放量升高。

图 8 为热负荷优化结果。在 23:00—24:00 和 01:00—06:00 时段由于热负荷处于高峰时段, 在该时段采取了强制削减措施, 降低此时系统热负荷的需求, 并对用户进行补偿。在 08:00—12:00 和 18:00—22:00 时段热负荷需求较低, 且此时 HGT 动作为系统供热供电, EB 由于购电费用太高停止供电, 不足部分由 HGB 供给。

图 9 为系统冷负荷优化结果。由图 9 可知, 在电价高峰期, 冷负荷主要由热制冷提供, 这是由于此时系统 HGT 达到最高出力阶段且热负荷处于低谷期, 系统有多余热能提供给 AC 装置供冷; 除在电价高峰期, 其冷负荷主要由电制冷提供。在冷负荷高峰段, 本调度对负荷用户也采取了强制削减措施, 以保证系统供冷。

除激励需求响应的强制削减措施外, 由热冷惯性和感知模糊需求响应措施对处于高峰时段的热冷负荷用户也进行一定程度的削减(见图 11), 此部分负荷削减不会影响用户的正常使用, 因此无需额外补偿费用。

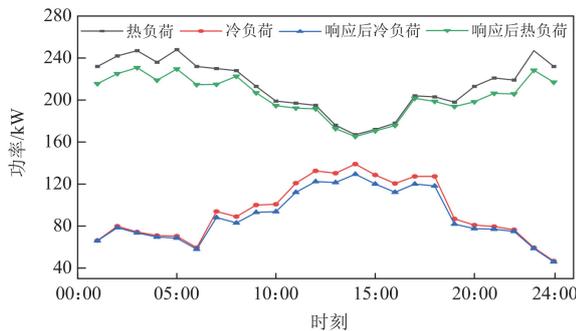


图 11 热冷负荷需求响应

Fig. 11 Demand response for heating and cooling loads

## 5.4 方法对比分析

### 5.4.1 模型对比

场景 1: 不考虑系统柔性资源参与调度, 即不含 EV 有序充电和冷热电综合需求响应;

场景 2: 仅考虑单一柔性资源, 即考虑综合需求响应但不考虑 EV 有序充电;

场景 3: 仅考虑单一柔性资源, 考虑 EV 有序充电不考虑综合需求响应;

场景 4: 考虑多类型需求响应资源。

以上场景均采用基于行为克隆的 TD3 算法进行离线训练, 用同一历史数据集进行训练, 再采用图 5 调度日数据对训练好的模型进行测试。各场景调度结果如表 2 所示。

表 2 优化调度结果对比

场景	成本/元	碳排放/kg	购电成本/元	购气成本/元
场景 1	10924.11	6918.60	5178.00	3973.25
场景 2	9987.19	6086.96	4174.38	3725.49
场景 3	10205.95	5948.21	4365.52	4055.59
场景 4	8898.67	5667.13	3009.72	3844.77

相较于场景 1, 场景 2 的调度成本和碳排放都有所降低。这是由于场景 2 中的综合需求响应使得冷热电负荷的峰谷差减少, 从而减少了系统的购气购电成本; 场景 3 由于实施了 EV 的有序充电, 避免在系统用电高峰期和电价高峰期集中充电, 且在高峰时放电, 从而减少了系统的购电需求; 对于场景 4 采取了 EV 有序充电和综合需求响应双措施, 该场景通过多类型柔性资源调控, 大幅降低了系统的峰谷差, 并在降低系统调度成本之外, 系统有多余电能供给碳捕集装置, 使得系统碳捕能力上升, 大幅降低了系统的碳排放。

### 5.4.2 算法对比

为验证本文算法对低碳园区综合能源系统低碳经济调度的优化效果, 分别采用了一种常用的群智能优化算法(粒子群优化)和常用的深度强化学习算法(TD3)作为对比算法, 同样对本文模型进行优化求解。各优化算法对比结果和各强化学习算法的训练收敛图如表 3 和图 12 所示。

表 3 不同优化方法对比

优化方法	成本/元	优化时间/s
粒子群优化	9305.52	195
TD3 <sup>[17]</sup>	9124.51	1
考虑行为克隆 TD3	8898.67	1

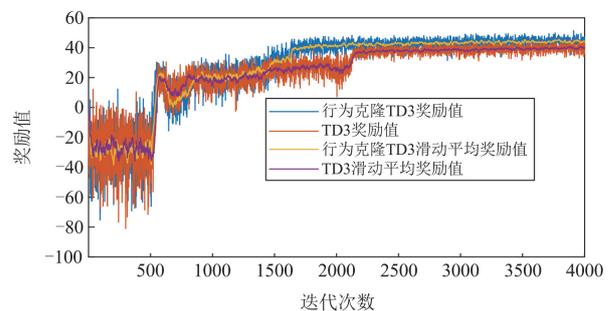


图 12 强化学习算法训练收敛图对比

Fig. 12 Comparison of training convergence plots for reinforcement learning algorithms

由表 3 可知, 由于传统群智能优化算法易陷入局部最优, 相比于本文强化学习算法其最终的成本

高出 4.6%，且其寻优速度为 195 s，远比强化学习算法所需的求解时间长，难以满足在线优化调度的要求；而 TD3 算法和本文考虑行为克隆 TD3 算法的优化时间和优化结果都具有明显优势，可实现在线优化调度的快速求解；此外，考虑行为克隆 TD3 算法相比 TD3 算法在最终的优化结果上更好，且在离线训练过程中，TD3 算法的收敛速度和最终收敛的奖励值都低于行为克隆 TD3 算法(如图 12 所示)，因此本文算法对于实现低碳园区 IES 经济调度更具优势。

## 6 结论

本文考虑了园区综合能源系统内的各类柔性资源，并结合基于行为克隆 TD3 算法实现了含多类型柔性资源的园区综合能源系统低碳经济调度。首先建立考虑储、供、用能侧各类柔性资源的园区综合能源系统低碳经济调度模型，同时通过该模型建立强化学习环境，并采用基于行为克隆 TD3 算法对其进行优化求解，得到以下结论。

通过调度储能、柔性负荷和引导 EV 有序充电可有效降低园区综合能源系统的碳排放和调度成本。此外相较于群智能算法和常用的 TD3 强化学习算法，本文所采用的基于行为克隆 TD3 算法在优化性能方面具有明显的优势，可有效提升园区综合能源系统的低碳性和经济性。

鉴于本文研究在储能侧仅考虑了电储能柔性资源，且未考虑含储液式的碳捕集装置，后续还可综合考虑热、冷、空气储能柔性资源和含储液式碳捕集装置进一步研究。

## 附录 A

表 A1 各设备运行参数

Table A1 Operating parameters of each equipment

设备名称	出力上限/kW	出力下限/kW	爬坡率/(kW/h)	滑坡率/(kW/h)
HGT	0	180	100	120
HGB	0	180	—	—
AC	0	100	100	100
EB	0	150	90	100
电储	30	300	200	80

表 A2 各设备运行效率

Table A2 Operating efficiency of each equipment

设备名称	效率	设备名称	效率
HGT	0.35/0.45	HGB	0.7
AC	0.8	EB	0.9
电储	0.95/0.9	EL	0.9
碳捕集	0.9	EC	1.44
EV	0.95/0.9	—	—

表 A2 中，HGT 的 0.35/0.45 分别表示气转电能效率和气转热能效率；电储的 0.95/0.9 分别表示电储能的充、放电效率；EV0.95/0.9 分别为 EV 的充、放电效率。

表 A3 运行成本

Table A3 Operating cost

设备	成本	设备	成本
HGT	0.026 元/kWh	HGB	0.03 元/kWh
AC	0.013 元/kWh	EB	0.013 元/kWh
电储	0.02 元/kWh	EL	0.01 元/kWh
碳封存	0.03 元/kg	削减电负荷	0.6 元/kWh
风电	0.05 元/kWh	削减热负荷	0.7 元/kWh
光伏	0.039 元/kWh	削减冷负荷	0.75 元/kWh
EC	0.013 元/kWh	碳捕集折旧	309.75 元/天

## 参考文献

- [1] 陈艳波, 张宁, 李嘉祺, 等. 零碳园区研究综述及展望[J/OL]. 中国电机工程学报: 1-21[2024-05-24]. <https://doi.org/10.13334/j.0258-8013.pcsee.231293>.  
CHEN Yanbo, ZHANG Ning, LI Jiaqi, et al. Review and prospect of zero carbon park research[J]. Proceedings of the CSEE: 1-21[2024-05-24]. <https://doi.org/10.13334/j.0258-8013.pcsee.231293>.
- [2] 赵海祺, 董树锋, 南斌, 等. 考虑电炉短流程的钢铁工业园区综合能源系统低碳运行模式分析[J]. 电力建设, 2023, 44(6): 23-32.  
ZHAO Haiqi, DONG Shufeng, NAN Bin, et al. Analysis of low carbon operating models in steel industrial communityintegrated energy system considering electric arc furnace short process[J]. Electric Power Construction, 2023, 44(6): 23-32.
- [3] 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.
- [4] 王扬, 路菲, 李骥, 等. 基于综合优化目标的低碳园区能源系统规划配置[J]. 中国电力, 2024, 57(4): 14-24.  
WANG Yang, LU Fei, LI Ji, et al. Multi-energy system planning and configuration study for low-carbon parks based on comprehensive optimization objectives[J]. Electric Power, 2024, 57(4): 14-24.
- [5] 陈启鑫, 康重庆, 夏清. 碳捕集电厂的运行机制研究与调峰效益分析[J]. 中国电机工程学报, 2010, 30(7): 22-28.  
CHEN Qixin, KANG Chongqing, XIA Qing. Operation

- mechanism and peak-load shaving effects of carbon-capture power plant[J]. Proceedings of the CSEE, 2010, 30(7): 22-28.
- [6] 寇洋, 武家辉, 张华, 等. 考虑碳捕集与 CVaR 的电力系统低碳经济调度[J]. 电力系统保护与控制, 2023, 51(11): 131-140.  
KOU Yang, WU Jiahui, ZHANG Hua, et al. Low carbon economic dispatch for a power system considering carbon capture and CVaR[J]. Power System Protection and Control, 2023, 51(11): 131-140.
- [7] MA Yiming, WANG Haixin, HONG Feng, et al. Modeling and optimization of combined heat and power with power-to-gas and carbon capture system in integrated energy system[J]. Energy, 2021, 236: 121392.
- [8] 孙惠娟, 刘昀, 彭春华, 等. 计及电转气协同的含碳捕集与垃圾焚烧虚拟电厂优化调度[J]. 电网技术, 2021, 45(9): 3534-3544.  
SUN Huijuan, LIU Yun, PENG Chunhua, et al. Optimization scheduling of virtual power plant with carbon capture and waste incineration considering power-to-gas coordination[J]. Power System Technology, 2021, 45(9): 3534-3544.
- [9] 陈登勇, 刘方, 刘帅. 基于阶梯碳交易的含 P2G-CCS 耦合和燃气掺氢的虚拟电厂优化调度[J]. 电网技术, 2022, 46(6): 2042-2054.  
CHEN Dengyong, LIU Fang, LIU Shuai. Optimization of virtual power plant scheduling coupling with P2G-CCS and doped with gas hydrogen based on stepped carbon trading[J]. Power System Technology, 2022, 46(6): 2042-2054.
- [10] 刘妍, 胡志坚, 陈锦鹏, 等. 含碳捕集电厂与氢能多元利用的综合能源系统低碳经济调度[J]. 电力系统自动化, 2024, 48(1): 31-40.  
LIU Yan, HU Zhijian, CHEN Jinpeng, et al. Low-carbon economic dispatch of integrated energy system considering carbon capture power plant and multi-utilization of hydrogen energy[J]. Automation of Electric Power Systems, 2024, 48(1): 31-40.
- [11] 周伟, 孙永辉, 王建喜, 等. 计及自适应阶梯型需求响应激励机制的园区综合能源系统优化调度[J]. 电网技术, 2023, 47(10): 4210-4221.  
ZHOU Wei, SUN Yonghui, WANG Jianxi, et al. Optimal dispatch of park-level integrated energy system considering adaptive stepped demand response incentive mechanism[J]. Power System Technology, 2023, 47(10): 4210-4221.
- [12] 毕锐, 王孝淦, 袁华凯, 等. 考虑供需双侧响应和碳交易的氢能综合能源系统鲁棒调度[J]. 电力系统保护与控制, 2023, 51(12): 122-132.  
BI Rui, WANG Xiaogan, YUAN Huakai, et al. Robust dispatch of a hydrogen integrated energy system considering double side response and carbon trading mechanism[J]. Power System Protection and Control, 2023, 51(12): 122-132.
- [13] 崔杨, 邓贵波, 赵钰婷, 等. 考虑源荷低碳特性互补的含风电电力系统经济调度[J]. 中国电机工程学报, 2021, 41(14): 4799-4815.  
CUI Yang, DENG Guibo, ZHAO Yuting, et al. Economic dispatch of power system with wind power considering the complementarity of low-carbon characteristics of source side and load side[J]. Proceedings of the CSEE, 2021, 41(14): 4799-4815.
- [14] 胡志勇, 郭雪丽, 王爽, 等. 考虑响应意愿的电动汽车群-空调集群需求响应策略研究[J]. 电力系统保护与控制, 2023, 51(15): 109-119.  
HU Zhiyong, GUO Xueli, WANG Shuang, et al. Demand response strategy for electric vehicles and air conditioners considering response willingness[J]. Power System Protection and Control, 2023, 51(15): 109-119.
- [15] 王新迎, 赵琦, 赵黎媛, 等. 基于深度 Q 学习的电热综合能源系统能量管理[J]. 电力建设, 2021, 42(3): 10-18.  
WANG Xinying, ZHAO Qi, ZHAO Liyuan, et al. Energy management approach for integrated electricity-heat energy system based on deep Q-learning network[J]. Electric Power Construction, 2021, 42(3): 10-18.
- [16] 范培潇, 杨军, 肖金星, 等. 基于深度 Q 学习的含电动汽车孤岛微电网负荷频率控制策略[J]. 电力建设, 2022, 43(4): 91-99.  
FAN Peixiao, YANG Jun, XIAO Jinxing, et al. Load frequency control strategy based on deep Q learning for island microgrid with electric vehicles[J]. Electric Power Construction, 2022, 43(4): 91-99.
- [17] 杨挺, 赵黎媛, 刘亚闯, 等. 基于深度强化学习的综合能源系统动态经济调度[J]. 电力系统自动化, 2021, 45(5): 39-47.  
YANG Ting, ZHAO Liyuan, LIU Yachuang, et al. Dynamic economic dispatch for integrated energy system based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2021, 45(5): 39-47.
- [18] 王珣, 王钟, 沈海华, 等. 基于三维 Kmeans-DDPG 的多能园区优化调度[J]. 电工技术, 2023(6): 111-117, 120.  
WANG Xun, WANG Zhong, SHEN Haihua, et al. Multi-functional campus optimization scheduling based on 3D Kmeans-DDPG[J]. Electric Engineering, 2023(6): 111-117, 120.
- [19] ZHOU Xiang, WANG Jiye, WANG Xinying, et al. Optimal dispatch of integrated energy system based on deep

- reinforcement learning[J]. *Energy Reports*, 2023, 9: 373-378.
- [20] 彭春华, 陈思畏, 徐佳璐, 等. 综合能源系统混合时间尺度多目标强化学习低碳经济调度[J]. *电网技术*, 2022, 46(12): 4914-4925.  
PENG Chunhua, CHEN Siwei, XU Jialu, et al. Low carbon economic scheduling for integrated energy systems with mixed timescale & multi-objective reinforcement learning[J]. *Power System Technology*, 2022, 46(12): 4914-4925.
- [21] 邱革非, 何虹辉, 刘铠铭, 等. 基于改进 TD3 算法的综合能源系统低碳经济调度[J]. *电力科学与工程*, 2023, 39(10): 52-62.  
QIU Gefei, HE Honghui, LIU Kaiming, et al. Low-carbon economic dispatch of integrated energy system based on improved TD3 algorithm[J]. *Electric Power Science and Engineering*, 2023, 39(10): 52-62.
- [22] FUJIMOTO S, GU S S. A minimalist approach to offline reinforcement learning[C] // *Advances in Neural Information Processing Systems*, NeurIPS, 2021: 10874-10885.
- [23] 邱玥, 周苏洋, 顾伟, 等. “碳达峰、碳中和”目标下混氢天然气技术应用前景分析[J]. *中国电机工程学报*, 2022, 42(4): 1301-1321.  
QIU Yue, ZHOU Suyang, GU Wei, et al. Application prospect analysis of hydrogen enriched compressed natural gas technologies under the target of carbon emission peak and carbon neutrality[J]. *Proceedings of the CSEE*, 2022, 42(4): 1301-1321.
- [24] 程静, 谭智钢, 岳雷. 考虑负荷综合需求响应的 CCHP-SESS 双层优化配置[J]. *电网技术*, 2023, 47(3): 918-929.  
CHENG Jing, TAN Zhigang, YUE Lei. CCHP-SESS bi-layer optimal configuration considering comprehensive load demand response[J]. *Power System Technology*, 2023, 47(3): 918-929.
- [25] 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.
- [26] 彭春华, 郑聪, 陈婧, 等. 基于置信间隙决策的综合能源系统鲁棒优化调度[J]. *中国电机工程学报*, 2021, 41(16): 5593-5604.  
PENG Chunhua, ZHENG Cong, CHEN Jing, et al. Robust optimal dispatching of integrated energy system based on confidence gap decision[J]. *Proceedings of the CSEE*, 2021, 41(16): 5593-5604.
- [27] 马一鸣, 周夕然, 董鹤楠, 等. 考虑电转气与冷热负荷惯性的综合能源系统优化调度[J]. *电网与清洁能源*, 2021, 37(8): 118-127, 138.  
MA Yiming, ZHOU Xiran, DONG Henan, et al. Optimal dispatch of integrated energy system considering power-to-gas and load inertia[J]. *Power System and Clean Energy*, 2021, 37(8): 118-127, 138.

收稿日期: 2024-03-16; 修回日期: 2024-05-26

作者简介:

舒展(1977—), 男, 硕士, 高级工程师, 研究方向为电力系统运行与控制; E-mail: shuzhan2003@126.com

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

(编辑 魏小丽)