

DOI: 10.19666/j.rlfed.202208176

基于深度强化学习的热电联产机组 多工况自适应控制

叶婧, 蔡霞, 张磊, 杨楠, 李振华
(三峡大学电气与新能源学院, 湖北 宜昌 443002)

[摘要] 针对热电联产机组受大范围出力工况运行波动性影响, 导致汽轮机-锅炉协调控制系统发电负荷-抽汽流量-机前压力整体控制品质较差, 提出一种基于多智能体深度确定性策略梯度(MA-DDPG)的多工况自适应控制方法。首先, 根据机组非线性动态机理建立计及关键状态参数变化的多个工况子模型, 利用积分型函数切换机制获取最优工况子模型; 其次, 提出一种多智能体同步运行机制, 以快速响应、稳定供热及安全运行为协调控制目标来设计奖励函数, 通过训练智能体与环境进行互动, 不断在线调整多回路增益, 实现机组多工况自适应控制。仿真结果表明, 与传统控制方法相比, 所提方法在保证供热平稳性的同时, 有效提升了机组大范围出力工况发电负荷响应速率。

[关键词] 热电联产机组; MA-DDPG; 多工况自适应控制; 供热平稳性; 负荷响应速率

[引用本文格式] 叶婧, 蔡霞, 张磊, 等. 基于深度强化学习的热电联产机组多工况自适应控制[J]. 热力发电, 2023, 52(4): 104-112. YE Jing, CAI Xia, ZHANG Lei, et al. Multi-condition adaptive control of combined heat and power unit based on deep reinforcement learning[J]. Thermal Power Generation, 2023, 52(4): 104-112.

Multi-condition adaptive control of combined heat and power unit based on deep reinforcement learning

YE Jing, CAI Xia, ZHANG Lei, YANG Nan, LI Zhenhua

(College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China)

Abstract: Combined heat and power (CHP) units are affected by the output fluctuation of large-scale operating conditions, resulting in poor overall control quality of power generation load-extraction steam flow-throttle pressure in the turbine-boiler coordinated control system. To solve this problem, a multi-condition adaptive control method based on multi-agent deep deterministic policy gradient (MA-DDPG) is proposed. Firstly, according to the nonlinear dynamic mechanism of the unit, multiple operating condition sub-models considering the changes of state parameters are established, and the optimal operating condition sub-model is obtained by the integral function switching mechanism. For the multi-loop complex control requirements of the coordinated control system, a multi-agent synchronous operation mechanism is proposed, and the reward function is designed with the coordinated control objectives of rapid response, stable heating and safe operation. Finally, by training the agent to interact with the environment, the multi-loop gains are continuously adjusted online to achieve multi-condition adaptive control. Simulation results show that, compared with the conventional control method, the proposed method can effectively improve the load response rate under wide range operating conditions, and ensure the stability of heating.

Key words: CHP unit; MA-DDPG; multi-condition adaptive control; heating stability; load response rate

随着热电联产灵活性改造不断进行, “三北地区”大规模装机^[1]的热电联产 (combined heat and

power, CHP) 机组在较大出力区间内参与电网灵活性快速调节^[2-3]。但是, CHP 机组出力区间大时, “供

收稿日期: 2022-08-25 网络首发日期: 2022-11-25

基金项目: 国家自然科学基金项目 (52007103)

Supported by: National Natural Science Foundation of China (52007103)

第一作者简介: 叶婧 (1986), 女, 博士, 讲师, 主要研究方向为大规模新能源接入后电力系统优化运行与控制, yejing2000310@163.com。

通信作者简介: 张磊 (1986), 男, 博士, 副教授, 主要研究方向为大规模新能源接入电力系统的优化调度, leizhang3188@163.com。

电”与“供热”运行状况复杂多变，机组热力系统关键状态参数波动较大，导致被控对象延迟，非线性特性出现显著差异^[4]，机组响应过程变得复杂甚至出现不可控现象。因此，亟需对 CHP 机组出力区间的优化运行与控制展开研究，从而为 CHP 机组参与电网灵活性调节提供支撑。

目前，计及抽汽供热快速调节的变负荷优化控制^[5-7]能够在保证供热质量基础上，提升 CHP 机组额定工况负荷响应速率。但是，这些控制方法^[5-7]大多围绕机组特定出力工况展开，很少关注机组特定工况模型受大范围出力波动性影响，以及易忽略热力系统状态参数变化^[4]的重要信息，导致整体控制品质变差。近年来，多模型控制被广泛应用于解决非线性、变工况、参数不确定性等问题。文献[8]采用一种基于间隙度量的多模型控制方法，有效克服了模型失配；文献[9-10]利用多模型切换控制以削弱变工况对系统控制的影响，提高被控对象工况适应性。文献[8-10]为解决 CHP 机组大范围出力过程中被控对象特性差异问题提供了新思路。但是，CHP 机组传统协调控制系统的多条控制回路往往采用固定 PID 控制器^[5-7]，无法适应多模型控制的工况切换过程，机组发电负荷-抽汽流量-机前压力控制品质难以得到有效保证。文献[11-12]将深度强化学习^[13-14]入控制系统自适应设计环节，在保证系统控制可靠性的基础上，显著提升了其面对复杂环境的适应能力。

基于此，本文提出一种基于多智能体深度确定性策略梯度（multi-agent deep deterministic policy gradient, MA-DDPG）的多工况自适应控制方法。首先，根据机组较大出力波动及其状态参数非线性变化特征，构造多个工况子模型；然后，利用多智能体同步运行机制与多目标奖励函数的协调配合，在反复迭代训练过程中自适应调整子控制系统的多回路增益；最后，仿真验证所提控制方法的有效性。

1 CHP 机组工况建模及其控制

1.1 CHP 机组多工况模型与控制系统

CHP 机组动态响应过程是一个包含了机械能转换、热交换和两相流动的复杂过程。典型抽汽式 CHP 机组非线性动态机理^[5]可表述如下。

正压直吹式制粉系统惯性描述：

$$T_b \dot{r}_B(t) = -r_B(t) + u_B(t)(t - \tau) \quad (1)$$

锅炉能量平衡方程：

$$C_b \dot{p}_d(t) = -K_3 p_T(t) u_T(t) + K_1 r_B(t) \quad (2)$$

过热器机前压力压差特性描述：

$$p_T(t) = p_d(t) - K_2 (K_1 r_B(t))^{1.5} \quad (3)$$

汽轮机能量平衡方程：

$$T_t \dot{P}_e(t) = -P_e(t) + K_5 p_z(t) u_H(t) + K_4 K_3 p_T(t) u_T(t) \quad (4)$$

中间抽汽压力动态关系：

$$C_h \dot{p}_z(t) = -K_6 q_b(t)(96 p_z(t) - t_c(t) + 103) + (1 - K_4) K_3 p_T(t) u_T(t) - K_5 p_z(t) u_H(t) \quad (5)$$

抽汽流量计算：

$$q_m(t) = K_7 K_6 q_b(t)(96 p_z(t) - t_c(t) + 103) \quad (6)$$

式中： T_b 为制粉惯性时间，s； τ 为制粉延迟时间，s； C_b 为锅炉蓄热系数，MJ/MPa； $u_B(t)$ 为给煤量，t/h； $u_H(t)$ 为蝶阀开度，%； $u_T(t)$ 为高压调节阀开度，%； $p_T(t)$ 为机前压力，MPa； $P_e(t)$ 为发电负荷，MW； $q_m(t)$ 为抽汽流量，t/h； q_b 为循环水流量，t/h； t_c 为回水温度，℃； T_t 为汽轮机惯性时间，s； C_h 为热网加热器蓄热系数，MJ/℃； $r_B(t)$ 为锅炉燃烧率，t/h； p_d 为汽包压力，MPa； p_z 为供热抽汽压力，MPa； K_1 为燃料量增益； K_2 为压差拟合系数； K_3 为汽轮机增益； K_4 为高中压缸占汽轮机做功比例； K_5 为低压缸蒸汽做功增益； K_6 为热网循环水比热容； K_7 为抽汽热量折合蒸汽流量系数。

文献[15]对式(1)一式(6)非线性动态模型中各项状态参数的变化情况进行了详细分析，其中受机组热耗率验收（turbine heat acceptance, THA）工况变化影响较大同时又对整体控制品质影响极大的参数包括锅炉蓄热系数 C_b 、燃料量增益 K_1 、制粉惯性时间 T_b 、制粉延迟时间 τ 。其中，锅炉蓄热系数 C_b 表示单位压力变化下锅炉存储或释放的有效热量，受压力变化影响较大。机前压力 p_T 降低时，锅炉蓄热系数 C_b 显著增加。

燃料量增益 K_1 与锅炉效率、汽轮机热耗率及主蒸汽流量呈单值函数关系：

$$K_{1,i} = K_{1,\text{ref}} \eta_B \eta_{T,\text{ref}} / (\eta_{B,\text{ref}} \eta_T) \quad (7)$$

式中： $K_{1,\text{ref}}$ 为100%THA工况的燃料量增益； η_B 和 $\eta_{B,\text{ref}}$ 分别为实际工况和100%THA工况的锅炉效率； η_T 和 $\eta_{T,\text{ref}}$ 分别为实际工况和100%THA工况的汽轮机热耗率。

锅炉制粉系统作为 CHP 机组重要组成部分，其制粉惯性时间 T_b 主要受给煤量、煤质变化因素影响。给煤量 u_B 越小，制粉惯性时间 T_b 越小。而制粉延迟时间 τ 和给煤量 u_B 呈近似反比关系：

$$\tau = \tau_0 u_{B0} / u_B \quad (8)$$

式中： τ_0 和 u_{B0} 分别为 100%THA 工况下的制粉延迟时间和给煤量。

因此，在机组工况模型的划分中考虑上述关键状态参数的变化情况，用以精确描述被控对象不同出力状态下的非线性动态过程^[15-17]：

$$F_i = \begin{cases} T_{b,i} \dot{r}_B(t) = -r_B(t) + u_{B,i}(t - \tau_i) \\ C_{b,i} \dot{p}_d(t) = -K_3 p_T(t) u_{T,i} + K_{1,i} r_B(t) \\ p_T(t) = p_d(t) - K_2 (K_{1,i} r_B(t))^{1.5} \\ T_t \dot{p}_e(t) = -P_e(t) + K_5 p_z(t) u_{H,i} + \\ \quad K_4 K_3 p_T(t) u_{T,i} \\ C_h \dot{p}_z(t) = (1 - K_4) K_3 p_T(t) u_{T,i} - \\ \quad K_5 p_z(t) u_{H,i} - \\ \quad K_6 q_{b,i} (96 p_z(t) - t_{c,i} + 103) \\ q_m(t) = K_7 K_6 q_{b,i} (96 p_z(t) - t_{c,i} + 103) \end{cases} \quad (9)$$

式中： i 为不同工况编号； F_i 为第 i 个工况模型； $T_{b,i}$ 、 τ_i 、 $C_{b,i}$ 、 $K_{1,i}$ 分别为工况模型 F_i 的制粉惯性时间、制粉延迟时间、锅炉蓄热系数及燃料量增益； $u_{H,i}(t)$ 、 $u_{T,i}(t)$ 、 $u_{B,i}(t)$ 、 $q_{b,i}(t)$ 分别为工况模型 F_i 的蝶阀开度、高压调节阀开度、给煤量及循环水流量的稳态值。

CHP 机组协调控制系统将机前压力偏差 $\Delta p_T(t)$ 送入第 1 条 u_T 控制回路，发电负荷偏差 $\Delta P_e(t)$ 送入第 2 条 u_H 控制回路，而第 3 条 u_B 控制回路输入由 $\Delta P_e(t)$ 与抽汽流量扰动量 $\Delta q_m(t)$ 的等效发电负荷偏差构成^[5]。因此，机组工况模型 F_i 的控制系统为：

$$Y_i = [\Delta y_{1,i}(t) \quad \Delta y_{2,i}(t) \quad \Delta y_{3,i}(t)]$$

$$\Delta y_{j,i}(t) = y_{j,0} + k_{p,j} \Delta x_{j,i}(t) + \quad (10)$$

$$k_{1,j} \int_0^t \Delta x_{j,i}(t) dt + k_{d,j} \dot{\Delta x}_{j,i}(t), \quad j=1,2,3$$

式中： Y_i 为工况模型 F_i 的控制回路矩阵； $\Delta y_{j,i}(t)$ 、 $y_{j,0}$ 、 $\Delta x_{j,i}(t)$ 、 $k_{p,j}$ 、 $k_{1,j}$ 、 $k_{d,j}$ 分别为 F_i 的控制系统 Y_i 的第 j 条控制回路控制器的输出增量、输出初始值、输入偏差及 3 个控制增益，这些控制增益决定了协调控制系统调节性能。

1.2 CHP 机组多工况模型控制

多工况模型控制结构如图 1 所示。其中，根据积分型函数切换机制选取最优工况子模型及其控制系统的过程如下。

1) 根据式(9)构造 n 个工况子模型，逼近被控对象整个区间特性：

$$\Omega = \{F_i | i=1,2,\dots,n\} \quad (11)$$

式中： Ω 为模型元素 F_i 的集合。

2) 根据式(10)建立工况模型 F_i 的控制系统：

$$C = \{Y_i | i=1,2,\dots,n\} \quad (12)$$

式中： C 为子控制系统 Y_i 的集合。

3) 利用积分型函数切换指标评估各工况子模型与当前被控对象匹配程度，切换至最优工况模型 F_0 ：

$$J = \arg \min_{i \in \Omega} J_i (i=1,2,\dots,n)$$

$$J_i(t) = \sum_{j=1}^3 (\alpha |\Delta z_{j,i}(t)| + \beta \int_0^t e^{-\gamma(t-\tau)} \Delta z_{j,i}^2(\tau) d\tau) \quad (13)$$

式中： J 为切换机制； $\Delta z_{j,i}(t)$ 为第 i 个工况模型第 j 条控制输入 $z_{j,i}(t)$ 与实际被控对象模型第 j 条控制回路输入 $z_{j,real}(t)$ 的误差； α 和 β 分别为当前瞬态误差和记忆长度内误差的权重； γ 为遗忘因子，保证 $J_i(t)$ 的收敛性； $\alpha > 0$ ， $\beta > 0$ ， $\gamma > 0$ 。

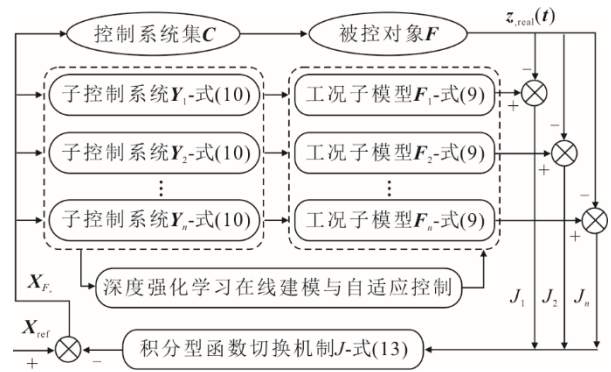


图 1 多工况模型控制结构
Fig.1 Control structure of the multi-condition model

最后，将最优工况模型 F_0 的输出与机组参考指令 X_{ref} 作比较，将该偏差矩阵 $X_{F_0} = [\Delta x_{1, F_0}(t) \quad \Delta x_{2, F_0}(t) \quad \Delta x_{3, F_0}(t)]$ 送入控制系统 Y_i ，深度强化学习不断调整 Y_i 的多控制回路增益，从而为 F_0 匹配最优子控制系统 Y_0 ；同时在复杂工况环境中学习最优策略，快速建立新的最优工况子模型-子控制系统偶对。

2 基于 MA-DDPG 的多工况自适应控制设计

基于 MA-DDPG 多工况自适应控制设计流程如图 2 所示。智能体 Actor 网络 $\mu(s|\theta^a)$ 根据当前状态 s 决策动作 a 并返回奖励值 r ，Critic 网络 $Q(s,a|\theta^Q)$ 对当前动作策略进行估计，在与环境交互过程中智能体不断评估 Q 值并修正策略^[18-19]，以实现期望最优决策，从而获取当前工况模型最优子控制系统。

2.1 DDPG 多智能体同步运行机制的提出

为满足 CHP 机组子控制系统 3 条控制回路的

负荷快速响应及供热稳定与运行安全复杂控制要求, 设置 DDPG 多智能体同步运行机制如图 3 所示。DDPG 并行、同步执行多个 Agent, 同时给出 3 个控制器动作并作用于同一工况模型 F_i , 利用多智能体^[20-21]模式分享模型与积累经验。

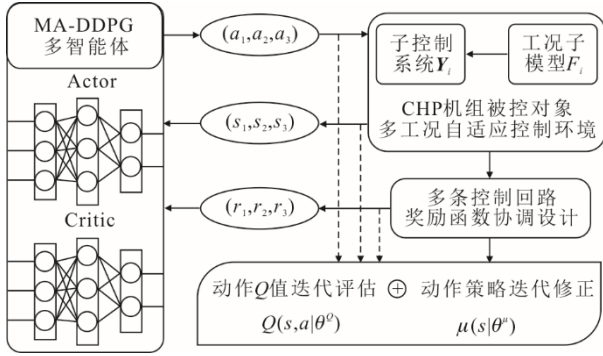


图 2 MA-DDPG 多工况自适应控制设计流程
Fig.2 Design process of the MA-DDPG-based multi-condition adaptive control

在任意优化时刻, 3 个 Agent 共同探索工况子模型环境, 借助经验回放池 M 在训练周期 T 内的 N 个随机样本快速更新各自网络参数和动作策略, 通过最小化损失函数 $L(Q)$ 和确定性策略梯度 (deterministic policy gradient, DPG) 训练更新 Q 、 μ 网络参数。因此, 每个 Agent 梯度更新公式:

$$\begin{cases} \min_{\theta^Q} L(Q) = \frac{1}{N} \cdot \sum_l (Y_t^l - Q(s_t^l, a_t^l | \theta^Q))^2 \\ Y_t^l = r_t^l + \omega Q(s_t, \mu'(s_t | \theta^\mu) | \theta^Q) \Big|_{s=s_t^l, a=\mu(s_t^l | \theta^\mu)} \end{cases} \quad (14)$$

$$\Delta_{\theta^\mu} \mu \approx \frac{1}{N} \cdot \sum_l \nabla_a Q(s, a | \theta^Q) \nabla_{\theta^\mu} \mu(s | \theta^\mu) \Big|_{s=s_t^l, a=\mu(s_t^l | \theta^\mu)} \quad (15)$$

式中: l 为回放池的随机样本编号, $\{a_t^l, s_t^l, r_t^l, s_{t+1}^l | t=1, 2, \dots\} \sim M, l \in N$; Y_t^l 为目标值函数; ω 为 Q 值迭代折扣因子; θ^Q 、 θ^Q 、 θ^μ 、 θ^μ 为每个 Agent 的 Critic、Actor 的当前 Q 、 μ 网络参数和目标 Q' 、 μ' 网络参数。

进一步, 更新 Agent 的 Critic、Actor 的网络参数 θ^Q 、 θ^Q 和 θ^μ 、 θ^μ :

$$\theta_{k+1}^Q = \theta_k^Q + \varphi \nabla_{\theta^Q} Y(\theta^Q) \quad (16)$$

$$\theta_{k+1}^{Q'} = \sigma \theta_{k+g}^Q + (1 - \sigma) \theta_{k+g}^Q \quad (17)$$

$$\theta_{k+1}^\mu = \theta_k^\mu + \varphi \nabla_{\theta^\mu} Y(\theta^\mu) \quad (18)$$

$$r_{2,i}(t) = \begin{cases} \xi_1, & u_H(t) \in [0, 100], |\Delta x_{2,i}(t)| \leq 0.476 |\Delta x_{2,i}| \\ \xi_1 - 1, & u_H(t) \notin [0, 100], |\Delta x_{2,i}(t)| \leq 0.476 |\Delta x_{2,i}| \\ -0.05 \Delta x_{2,i}(t), & u_H(t) \in [0, 100], |\Delta x_{2,i}(t)| > 0.476 |\Delta x_{2,i}| \\ -0.05 \Delta x_{2,i}(t) - 1, & u_H(t) \notin [0, 100], |\Delta x_{2,i}(t)| > 0.476 |\Delta x_{2,i}| \end{cases} \quad (22)$$

$$\theta_{k+1}^{\mu'} = \sigma \theta_{k+g}^\mu + (1 - \sigma) \theta_{k+g}^\mu \quad (19)$$

式中: θ_{k+1}^Q 、 $\theta_{k+1}^{Q'}$ 和 θ_{k+1}^μ 、 $\theta_{k+1}^{\mu'}$ 分别为第 $k+1$ 次迭代的 Q 、 μ 和 Q' 、 μ' 网络参数; χ 和 φ 分别为 Critic、Actor 当前网络学习权重值; σ 、 g 分别为 Critic、Actor 目标网络的更新权重及延迟步数。

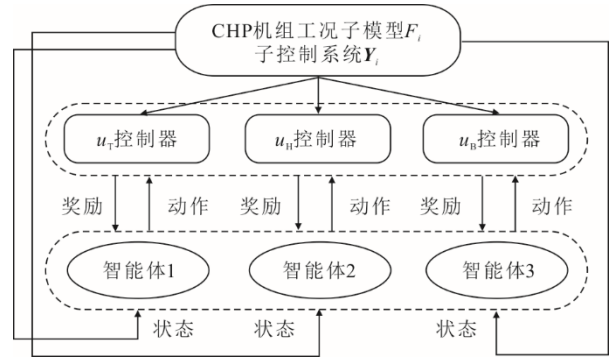


图 3 DDPG 多智能体同步运行示意
Fig.3 Schematic diagram of synchronous operation of DDPG-based multi-agent

这种同步运行机制使得每个 Agent 仅需处理较少决策数据, 克服了训练过程中单智能体对于子控制系统 3 个控制器训练周期长、收敛困难的问题, 进而提高 Agent 学习效率和运行速度。

2.2 并行多智能体的奖励函数设计

本文所述多目标控制问题等价于最大化工况模型 F_i 对应控制系统 3 个控制器最优响应策略下的期望奖励值:

$$\max E[R_{j,i} | \pi_{j,i}], j = 1, 2, 3 \quad (20)$$

式中: $R_{j,i} = \sum_{t=1}^T r_{j,i}(t)$; $\pi_{j,i}$ 表示第 j 个控制器在 T 内作用于工况模型 F_i 的总奖惩值及最优响应策略。

由于机组子控制系统 Y_i 中各被控变量之间存在耦合关系, 奖励函数设计要在保证 3 个 Agent 决策相对独立的同时, 协调各 Agent 之间的学习动作, 实现每个 Agent 决策最优与全局最优:

$$r_{1,i}(t) = \begin{cases} \xi_0, & |u_{T,i} - u_T(t)| < 20, |\Delta x_{1,i}(t)| \leq 0.35 \\ \xi_0 - 1, & |u_{T,i} - u_T(t)| \geq 20, |\Delta x_{1,i}(t)| \leq 0.35 \\ -2\Delta x_{1,i}(t), & |u_{T,i} - u_T(t)| < 20, |\Delta x_{1,i}(t)| > 0.35 \\ -2\Delta x_{1,i}(t) - 1, & |u_{T,i} - u_T(t)| \geq 20, |\Delta x_{1,i}(t)| > 0.35 \end{cases} \quad (21)$$

$$r_{3,i}(t) = \begin{cases} \xi_2 + \xi_3, & |u_{B,i} - u_B(t)| \leq 30, |\Delta x_{2,i}(t)| \leq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| \leq 1 \\ \xi_2 + \xi_3 - 1, & |u_{B,i} - u_B(t)| > 30, |\Delta x_{2,i}(t)| \leq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| \leq 1 \\ -0.05\Delta x_{2,i}(t) + \xi_3, & |u_{B,i} - u_B(t)| \leq 30, |\Delta x_{2,i}(t)| \geq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| \leq 1 \\ -0.05\Delta x_{2,i}(t) + \xi_3 - 1, & |u_{B,i} - u_B(t)| > 30, |\Delta x_{2,i}(t)| \geq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| \leq 1 \\ \xi_2 - 0.05\Delta x_{3,i}(t), & |u_{B,i} - u_B(t)| \leq 30, |\Delta x_{2,i}(t)| \leq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| > 1 \\ \xi_2 - 0.05\Delta x_{3,i}(t) - 1, & |u_{B,i} - u_B(t)| > 30, |\Delta x_{2,i}(t)| \leq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| > 1 \\ -0.05(\Delta x_{2,i}(t) + \Delta x_{3,i}(t)), & |u_{B,i} - u_B(t)| \leq 30, |\Delta x_{2,i}(t)| \geq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| > 1 \\ -0.05(\Delta x_{2,i}(t) + \Delta x_{3,i}(t)) - 1, & |u_{B,i} - u_B(t)| > 30, |\Delta x_{2,i}(t)| \geq 0.524\Delta x_{2,\text{ref}}, |\Delta x_{3,i}(t)| > 1 \end{cases} \quad (23)$$

式中： $r_{1,i}(t)$ 、 $r_{2,i}(t)$ 、 $r_{3,i}(t)$ 分别为3个Agent对应控制器被控变量的响应奖惩； $\Delta x_{2,\text{ref}}$ 为 u_T 控制器的参考输入偏差； ξ_0 、 ξ_1 、 ξ_2 、 ξ_3 均通过试验获得， $\xi_0=2$ ， $\xi_1=5$ ， $\xi_2=3$ ， $\xi_3=15$ 。

3 系统仿真结果与分析

以330 MW抽汽式CHP机组为例，在MATLAB 2020b/Simulink中搭建模型进行仿真分析，验证所提基于MA-DDPG的CHP机组多工况自适应控制方法的有效性。该热电联产机组不同工况模型的关键状态参数变化情况^[4-5,15]见表1。

表1 CHP机组典型出力工况关键参数变化情况
Tab.1 Changes of key parameters of CHP unit under typical output conditions

关键状态参数	不同热耗率验收工况(THA)工作点			
	100%	75%	50%	30%
P_e/MW	275	208	130	88
$q_m/(\text{t}\cdot\text{h}^{-1})$	400	300	222	105
p_T/MPa	16.67	14.61	9.96	7.36
$u_T/\%$	82	71	67	58
$u_H/\%$	76	56	33	28
$u_B/(\text{t}\cdot\text{h}^{-1})$	251	200	162	128
$q_b/(\text{t}\cdot\text{h}^{-1})$	2 500	1 875	1 388	656
τ 相对值/%	100	125	155	196
T_b 相对值/%	100	91	80	66
K_i 相对值/%	100.0	98.5	93.7	88.5
C_b 相对值/%	100	118	141	165

3.1 多工况模型性能分析

在大范围出力工况运行区间进行负荷扰动实验，对比验证所提多工况模型描述方法的有效性。CHP机组响应达到稳态后，1 800 s施加+20 MW负荷扰动，基于机组多工况模型与传统模型的热电输出误差响应性能指标见表2。机组控制系统3个传统固定控制器控制增益为 $[(12.0, 0.100, 0, 0), (25.0, 0.000, 4, 1.2), (44.3, 0.002, 8, 18.5)]$ 。

表2 不同模型描述下热电输出误差响应指标
Tab.2 Thermal-electrical output error response indexes of different models

模型描述	工况 THA/%	上升时间/s	响应时间/s	稳态误差	
		ΔP_e	ΔP_e	ΔP_e	Δq_m
多工况模型	100	257.53	272.23	0.155	0.530
	75	335.09	499.82	0.947	28.391
	50	348.41	499.82	1.143	36.775
	30	397.40	499.82	1.106	19.311
传统模型	100	257.53	272.23	0.155	0.530
	75	226.86	272.23	0.287	-0.654
	50	222.01	275.17	0.203	-1.120
	30	283.46	324.02	0.269	-0.077

由表2可知，相较于传统模型，利用多工况模型刻画大范围负荷工况运行时，CHP机组发电负荷上升时间平均由244.11 s延长至360.30 s，抽汽流量响应缓慢、稳态误差达36 t/h左右，机组响应过程波动趋势复杂、过渡时间不一、跟踪效果差异显著。可见，基于理想状态参数的传统模型忽略了大范围负荷工况运行时系统状态参数非线性变化，无法准确有效地描述机组整个区间特性。

3.2 不同算法收敛性能分析

在多工况模型描述方法基础上，对比本文MA-DDPG与其他算法的收敛性能。以训练75%THA工况子模型对应子控制系统发电负荷控制回路Agent为例，MA-DDPG、DDPG与DQN不同算法在线学习训练过程及其性能比较如图4和表3所示。

从图4可知，训练开始时Agent的累计奖励均为负值，学习效果欠佳。但是，与基准方法DDPG、DQN以及GA算法相比，MA-DDPG在多个Agent相互学习、独立决策训练过程中，奖励值不断增加，迭代75次便基本收敛，证实了深度强化学习的收敛速度优势，这是由于在线训练环节得到离线知识快速指导，避免了大量迭代工作。此外，与其他算

法相比,表3所提 MA-DDPG 的累计奖励值有效提升了 18.76%、38.28%、12.55%。可见,多个 Agent 协同合作相较于单个 Agent 连续和离散动作算法,具有更快的收敛速度和更好的训练效果。

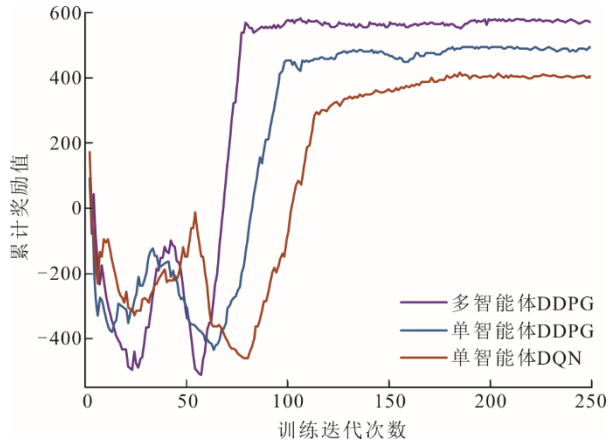


图4 不同算法训练过程累计奖励曲线
Fig.4 Cumulative reward curves of different algorithms during training process

表3 不同算法性能比较
Tab.3 Performance comparison of different algorithms

不同算法	MA-DDPG	DDPG	DQN	GA
迭代次数	75	90	110	189
累计奖励值	572.2	481.8	413.8	508.4

3.3 大范围工况阶跃扰动实验分析

在大范围工况进行阶跃扰动实验,对比所提 MA-DDPG 自适应控制方案与传统控制方案下机组动态响应性能。工况子模型-子控制系统偶对 MA-DDPG 自适应结果见表4。

表4 典型工况模型-子控制系统偶对自适应结果
Tab.4 The adaptive results of typical working condition model-sub-control system pair

工况 THA/%	子控制系统			
	控制增益	u_T 控制器	u_H 控制器	u_B 控制器
100	比例	12.1	25.3	46.7
	积分	0.112 0	0.000 4	0.002 8
	微分	0	1.5	18.0
75	比例	13.3	27.6	50.3
	积分	0.130 3	0.000 4	0.003 1
	微分	0	1.8	20.6
50	比例	15.6	28.2	72.8
	积分	0.181 2	0.000 4	0.005 6
	微分	0	2.1	27.3
30	比例	19.2	31.5	61.9
	积分	0.232 1	0.000 4	0.004 3
	微分	0	2.0	21.5

CHP 机组响应达到稳态后,1 800 s 施加 +20 MW 负荷扰动,传统控制方案与 MA-DDPG 自适应控制方案下的误差响应曲线簇如图5所示,响应指标见表5。

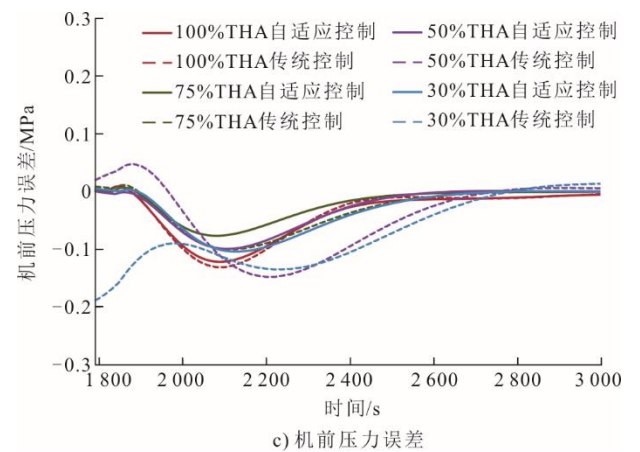
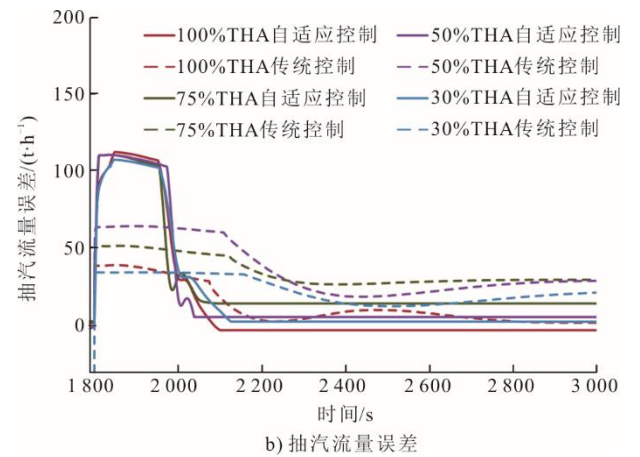
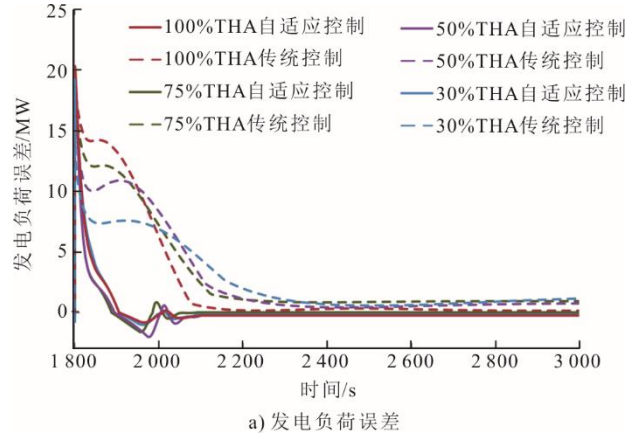


图5 CHP 机组不同工况的阶跃扰动误差响应曲线
Fig.5 Step disturbance response curves of CHP unit under different conditions

由图5及表5可知,与传统控制方案相比,自适应控制方案使机组发电负荷上升时间由 253~345 s 大幅缩短至 58~79 s,响应时间平均减少了 52.98%;

抽汽流量误差短时波动后迅速稳至 0 附近；机前压力峰值平均缩减了 25.07%，有效避免安全失稳现象。

可见，在工况子模型-子控制系统偶对 MA-DDPG 自适应作用下，CHP 机组在大范围区间迅速准确响应需求而不影响供热及运行安全。

表 5 CHP 机组不同工况的阶跃扰动响应指标对比
Tab.5 The step disturbance response indexes of CHP unit under different conditions

控制方案	工况 THA/%	上升时间/s	响应时间/s	超调量/%			稳态误差		
		ΔP_e	ΔP_e	Δp_T	ΔP_e	Δq_m	Δp_T	ΔP_e	Δq_m
自适应控制	100	78.980	161.940	1.174	5.002	-28.834	0	-0.126	-4.602
	75	58.497	170.631	0.766	8.002	-33.920	0	0.032	12.892
	50	64.220	192.732	0.994	10.043	-48.687	0	-0.239	4.052
	30	78.867	93.213	1.042	4.004	-101.990	0	-0.230	1.030
传统控制	100	252.763	269.412	1.271	0	-9.426	-0.004	0.151	0.335
	75	289.261	312.323	1.003	0	-8.189	0.000	0.945	28.398
	50	308.340	345.261	1.540	0	-18.435	0.006	0.751	27.795
	30	345.262	388.460	1.492	0	-15.551	0.014	1.163	20.033

3.4 低负荷工况连续扰动实验分析

在 30%~50% THA 低负荷工况区间进行连续扰动实验，对比所提 MA-DDPG 自适应控制方案与传统控制方案下机组响应性能。CHP 机组发电负荷在 83~160 MW 内波动，9.96 MPa 定压运行，抽汽流量保持 220 t/h。传统控制方案与 MA-DDPG 自适应控制方案下误差响应曲线及响应指标如图 6 和表 6 所示。

由图 6 及表 6 可知，相较于传统控制方案下发电负荷响应缓慢、抽汽流量和机前压力的长时、大幅波动失稳趋势，本文所提自适应控制方案使得机组发电负荷上升时间平均减小了 79.54%，响应时间由 221~416 s 大幅缩至 72~162 s，且抽汽流量及机前压力响应过程过渡极为迅速平稳。

综上所述，由表 5、表 6 及图 5、图 6 分析可见，面对整个区间大范围出力要求，基于 MA-DDPG 的多工况自适应控制下 CHP 机组发电负荷响应快速性、供热稳定性和运行安全性等指标均优于传统控制。

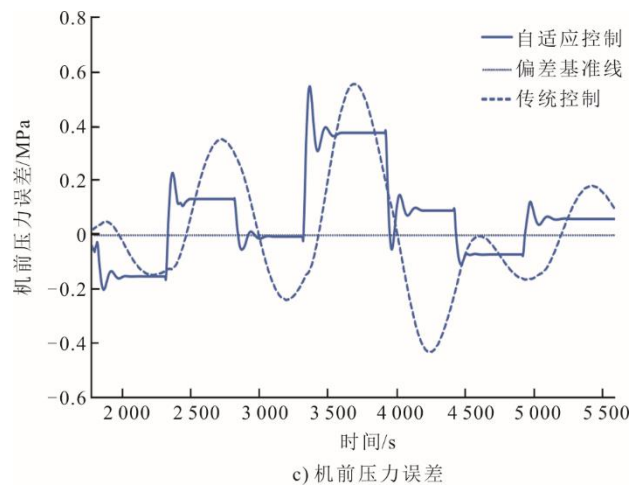
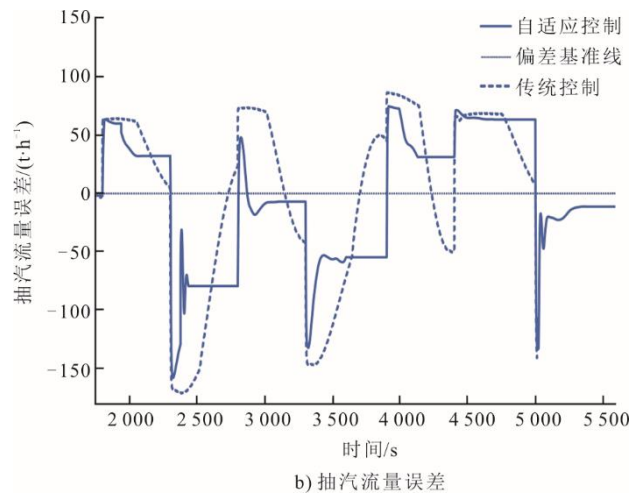
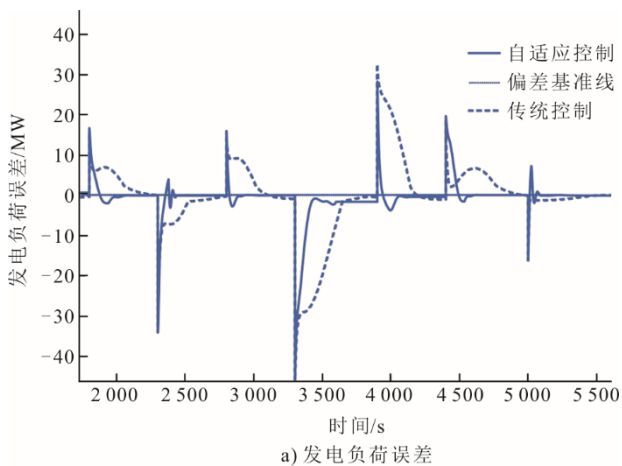


图 6 CHP 机组低负荷工况连续扰动误差响应曲线
Fig.6 Continuous disturbance error response curves of CHP unit under low load conditions

表6 CHP 机组低负荷工况的连续扰动响应指标对比
Tab.6 The continuous disturbance response indexes of CHP unit under low load conditions

控制	连续扰动 负荷指令	上升时间/s		超调量/%			稳态误差		
		ΔP_e	ΔP_e	Δp_T	ΔP_e	Δq_m	Δp_T	ΔP_e	Δq_m
自适应控制	(1 800 s,+16 MW)	64.02	162.27	0.285	10.99	-16.437	-0.153	0	21.845
	(2 300 s,-34 MW)	40.06	97.05	-0.877	10.82	91.301	0.132	-0.02	-59.620
	(2 800 s,+16 MW)	18.15	71.97	0.472	12.09	-23.968	-0.007	0.02	-7.286
	(3 300 s,-45 MW)	92.61	103.47	-1.784	0	27.993	0.376	-1.58	-48.954
	(3 900 s,+33 MW)	27.46	112.76	1.449	10.12	-22.736	0.089	-0.34	30.838
	(4 400 s,+20 MW)	81.51	142.30	1.604	6.858	-5.092	-0.072	-0.09	52.939
	(5 000 s,-16 MW)	10.16	72.76	-0.006	45.49	52.334	0.059	-0.01	-11.543
传统控制	(1 800 s,+16 MW)	271.08	338.74	0.104	0	-27.278	-0.137	0.07	4.222
	(2 300 s,-34 MW)	175.00	220.57	-0.367	0	97.903	0.317	-0.20	23.313
	(2 800 s,+16 MW)	257.89	318.42	0.513	0	-43.658	-0.188	-0.84	-42.593
	(3 300 s,-45 MW)	287.44	331.97	-3.256	0	108.170	0.239	-0.43	44.756
	(3 900 s,+33 MW)	240.16	270.58	1.712	0	-49.824	-0.257	-0.99	-49.296
	(4 400 s,+20 MW)	390.21	415.92	1.582	0	-28.424	-0.155	-0.05	6.872
	(5 000 s,-16 MW)	10.81	280.53	-0.008	0	61.041	0.103	0	25.262

4 结 论

针对 CHP 机组受大范围工况出力波动性影响以及易忽略其中状态参数变化重要信息,本文提出了一种基于 MA-DDPG 的多工况自适应控制方法。面对 CHP 机组大范围出力工况复杂控制需求,所提控制方法具备良好的工况自适应能力。

1) 通过多模型控制理论构建了多工况模型,有效刻画了机组整个区间特性。经仿真验证传统模型无法有效描述机组整个区间特性,多工况模型实现了对系统运行状态、参数变化等因素有效估计。

2) 设计了基于 DDPG 的多智能体同步运行机制,并协同多目标奖励函数,在模型状态参数不确定及出力波动性影响下自适应调整多回路增益。仿真表明,本文控制方法在保证供热质量及运行安全基础上,有效提升了 CHP 机组发电负荷响应速率。

[参 考 文 献]

- [1] 赵紫原. 热电联产如何“一石三鸟”[N/OL]. 中国能源报, (2022-03-09)[2022-06-09]. http://www.cnenergynews.cn/dianli/2022/03/09/detail_20220309119037.html. ZHAO Ziyuan. How can combined heat and power unit “kill three birds with one stone”[N/OL]. China Energy News, (2022-03-09)[2022-06-09]. http://www.Cnenergynews.cn/dianli/2022/03/09/detail_20220309119037.html.
- [2] 帅永, 赵斌, 蒋东方, 等. 中国燃煤高效清洁发电技术现状与展望[J]. 热力发电, 2022, 51(1): 1-10. SHUAI Yong, ZHAO Bin, JIANG Dongfang, et al. Status and prospect of coal-fired high efficiency and clean power generation technology in China[J]. Thermal Power Generation, 2022, 51(1): 1-10.
- [3] 潘尔生, 田雪沁, 徐彤, 等. 火电灵活性改造的现状、

关键问题与发展前景[J]. 电力建设, 2020, 41(9): 58-68. PAN Ersheng, TIAN Xueqin, XU Tong, et al. Status, critical problems and prospects of flexibility retrofit of thermal power in China[J]. Electric Power Construction, 2020, 41(9): 58-68.

- [4] 李玲, 刘鑫屏. 新能源大规模并网条件下火电机组深度调峰控制策略优化[J]. 中国电力, 2020, 53(1): 155-160. LI Ling, LIU Xinping. Control strategy optimization for thermal power unit adapted to deep peak shaving for large-scale new energy source integration[J]. Electric Power, 2020, 53(1): 155-160.
- [5] 王玮, 孙阳, 刘吉臻, 等. 适应电网快速调频的热电联产机组新型变负荷控制策略[J]. 电力系统自动化, 2018, 42(21): 63-69. WANG Wei, SUN Yang, LIU Jizhen, et al. Load-change control strategy for combined heat and power units adapted to rapid frequency regulation of power grid[J]. Automation of Electric Power Systems, 2018, 42(21): 63-69.
- [6] WANG W, LIU J Z, ZENG D L, et al. Modeling and flexible load control of combined heat and power units[J]. Applied Thermal Engineering, 2020, 166: 114624.
- [7] HANG Y H, CHEN Q, YE J, et al. Study on dynamic process characteristics of CHP unit with variable load based on working point linearization modeling[J]. Frontiers in Energy Research, 2022, 9: 815272.
- [8] 马增辉, 徐慧仪. 基于间隙度量的主汽温多模型 Smith 预估控制[J]. 热能动力工程, 2020, 35(9): 148-153. MA Zenghui, XU Huiyi. Multiple model smith predictor control of main steam temperature based on gap metric[J]. Journal of Engineering for Thermal Energy and Power, 2020, 35(9): 148-153.
- [9] 关成龙, 潘维加, 李鹏辉. 基于多模型的超临界机组过热汽温改进模糊控制[J]. 电力科学与技术学报, 2019, 34(1): 129-135. GUAN Chenglong, PAN Weijia, LI Penghui. Improved fuzzy control for superheated steam temperature of supercritical unit based on multi-model[J]. Journal of Electric Power Science and Technology, 2019, 34(1): 129-135.

- [10] 周超, 谢七月, 左毅, 等. 基于多模型切换的锅炉主蒸汽温度预测控制[J]. 电力科学与技术学报, 2020, 35(4): 154-160.
ZHOU Chao, XIE Qiyue, ZUO Yi, et al. Multi-model switching based predictive control for main steam temperature in boiler[J]. Journal of Electric Power Science and Technology, 2020, 35(4): 154-160.
- [11] 严家政, 专祥涛. 基于强化学习的参数自整定及优化算法[J]. 智能系统学报, 2022, 17(2): 341-347.
YAN Jiazheng, ZHUAN Xiangtao. Parameter self-tuning and optimization algorithm based on reinforcement learning[J]. CAAI Transactions on Intelligent Systems, 2022, 17(2): 341-347.
- [12] 刘勇, 李祥, 蒋沛阳, 等. 基于DDPG-PID的机器人动态目标跟踪与避障控制研究[J]. 南京航空航天大学学报, 2022, 54(1): 41-50.
LIU Yong, LI Xiang, JIANG Peiyang, et al. Research on robot dynamic target tracking and obstacle avoidance control based on DDPG-PID[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2022, 54(1): 41-50.
- [13] ARULKUMARAN K, DEISENROTH M P, BRUNDAGE M, et al. A brief survey of deep reinforcement learning[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.
- [14] ZHAO X Y, DING S F, AN Y X, et al. Applications of asynchronous deep reinforcement learning based on dynamic updating weights[J]. Applied Intelligence, 2019, 49(2): 581-591.
- [15] 刘鑫屏. 热力发电过程建模与状态参数检测研究[D]. 北京: 华北电力大学, 2010: 43-44.
LIU Xinping. Research on modeling and the state parameter detection of thermal power generation process[D]. Beijing: North China Electric Power University, 2010: 43-44.
- [16] 董泽, 尹二新. 基于状态跟踪的非线性工业系统全工况建模[J]. 系统仿真学报, 2018, 30(3): 846-856.
DONG Ze, YIN Erxin. Modeling of nonlinear industrial system at all operating conditions based on state tracking[J]. Journal of System Simulation, 2018, 30(3): 846-856.
- [17] 赵帅, 顾煜炯, 李欣, 等. 供热机组多元工况下能效状态基准研究[J]. 热力发电, 2022, 51(4): 88-95.
ZHAO Shuai, GU Yujiong, LI Xin, et al. Research on energy efficiency state benchmark of heating units under multiple operating conditions[J]. Thermal Power Generation, 2022, 51(4): 88-95.
- [18] 邱高, 刘友波, 许立雄, 等. 基于深度确定性策略梯度的电网断面极限传输能力动态趋优控制[J]. 中国电机工程学报, 2021, 41(15): 5128-5139.
QIU Gao, LIU Youbo, XU Lixiong, et al. A deep deterministic policy gradient based-dynamic optimizing control for power system total transfer capability[J]. Proceedings of the CSEE, 2021, 41(15): 5128-5139.
- [19] 谢启跃, 应雨龙. 基于深度确定性策略梯度的电-气综合能源微网优化调度研究[J]. 热力发电, 2022, 51(3): 141-147.
XIE Qiyue, YING Yulong. Research on optimal dispatch of electricity-gas integrated energy microgrid based on deep deterministic policy gradient[J]. Thermal Power Generation, 2022, 51(3): 141-147.
- [20] 郑宇, 董朝阳, 赵俊华, 等. 考虑多智能体协同的电动汽车充放电分散式实时优化策略[J]. 电力建设, 2017, 38(6): 15-20.
ZHENG Yu, DONG Zhaoyang, ZHAO Junhua, et al. Distributed real-time optimization strategy of electric vehicle charge/discharge in cooperation with multi-agent system[J]. Electric Power Construction, 2017, 38(6): 15-20.
- [21] 苏永新, 吴泽旋, 谭貌, 等. 基于深度强化学习的家庭综合需求响应在线优化[J]. 中国电机工程学报, 2021, 41(16): 5581-5593.
SU Yongxin, WU Zexuan, TAN Mao, et al. Online optimization for home integrated demand response based on deep reinforcement learning[J]. Proceedings of the CSEE, 2021, 41(16): 5581-5593.

(责任编辑 杜亚勤)