

DOI: 10.19783/j.cnki.pspc.230458

基于 DDPG 算法的微网负载端接口变换器自抗扰控制

周雪松¹, 张心茹¹, 赵滢宇¹, 王博¹, 赵明², 问虎龙³

(1. 天津市新能源电力变换传输与智能控制重点实验室(天津理工大学), 天津 300384; 2. 承德电智尚节能科技有限公司, 河北 承德 067000; 3. 天津瑞能电气有限公司, 天津 300385)

摘要: 直流微电网是新能源综合利用的重要形式, 但其中的分布式接口往往存在着强随机性扰动, 这给直流变换器的稳压控制带来了诸多问题。为了尽可能地抑制控制器参数固定时这种不确定性特征引起的不利影响, 提出了一种利用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法整定线性自抗扰控制器参数的方法。依靠引入了智能算法的自抗扰微电网控制系统, 实现了控制器参数的自适应调整, 从而实现了微电网接口变换器的稳定运行。通过仿真对比了各类典型工况下, DDPG-LADRC 与传统线性自抗扰控制器(linear active disturbance rejection control, LADRC)、双闭环比例-积分控制器的性能差异, 验证了所提控制策略的有效性。而参数扰动下的鲁棒性分析结果结合多项指标下的系统整体性分析, 充分体现了控制器参数的智能化调整所带来的多工况自适应性增益的优越性, 具备较强的工程价值。

关键词: 微电网; DC-DC 变换器; 线性自抗扰控制; 深度强化学习; DDPG 算法; 抗扰性

Active disturbance rejection control of a microgrid load-side interface converter based on a DDPG algorithm

ZHOU Xuesong¹, ZHANG Xinru¹, ZHAO Hanyu¹, WANG Bo¹, ZHAO Ming², WEN Hulong³

(1. Tianjin Key Laboratory of New Energy Power Conversion, Transmission and Intelligent Control (Tianjin University of Technology), Tianjin 300384, China; 2. Chengde Electric Zhishang Energy Saving Technology Co., Ltd., Chengde 067000, China; 3. Tianjin Reineng Electric Co., Ltd., Tianjin 300385, China)

Abstract: The DC microgrid is an important form of comprehensive utilization of new energy, but the distributed interface often has strong random disturbance, which brings many problems to the voltage stabilization control of a DC converter. To suppress the adverse effects caused by this uncertainty feature when the controller parameters are fixed as much as possible, this paper proposes a method for setting linear ADRC controller parameters using a deep deterministic policy gradient algorithm (DDPG), and relies on the automatic rejection microgrid control system introduced by introducing intelligent algorithms to realize the adaptive adjustment of controller parameters, so as to realize the stable operation of the microgrid interface converter. Simulation curves are used to compare the performance differences between DDPG-LADRC and traditional LADRC and the double-loop PI controller proposed in this paper in various typical working conditions. The effectiveness of the control strategy proposed in this paper is verified. The robustness analysis results under parameter perturbation combined with the overall analysis of the system under multiple indicators fully reflect the superiority of multi-condition adaptive gain brought by the intelligent adjustment of controller parameters, and have strong engineering value.

This work is supported by the National Natural Science Foundation of China (No. 51877152).

Key words: microgrid; DC-DC converters; linear active disturbance rejection control; deep reinforcement learning; deep deterministic policy gradient (DDPG) algorithm; disturbing resistance

0 引言

随着能源结构的调整, 以光伏、风机为代表的

新能源形式在工业现代化发展过程中发挥着更加关键的作用^[1]。直流微网将分布式电源、DC-DC 变换器和各种负载进行整合, 解决了配电系统中大规模可再生能源的利用问题, 从而得到广泛应用^[2-3]。但由于融合了分布式能源和负载, 常表现出随机性、间断性的特点, 给直流母线和输出侧接口的电压造成较大影响, 进而影响微网中接口变换器运行的稳

基金项目: 国家自然科学基金项目资助(51877152); 天津市科技特派员项目资助(22YDTPJC00340); 天津市研究生科研创新实践项目资助(2022SKY180)

定性^[4-6]。因此,对接口变换器施加具有强鲁棒性与抗扰性的稳压控制策略显得尤为重要。

由于微网负载端接口变换器非线性、时变的结构特点,导致所建立的模型存在不精确性。PI 控制往往受到滞后效应和抗扰性不足的制约^[7],难以实现较好的控制效果。文献[8]将自抗扰控制器(active disturbance rejection control, ADRC)用于控制 DC-DC 变换器,但传统 ADRC 由于非线性的扩张状态观测器和误差反馈控制率均含有非线性函数,因此调参问题较为复杂。文献[9]对传统 ADRC 进行了线性化处理,提出了带宽法的思想,减少了需要整定的参数数量,但仍需要手动整定。上述控制方法均需要依靠设计人员对系统的认知以及经验,且设计出的参数难以适应复杂多变的微网系统。因此,文献[10-12]用遗传算法、粒子群算法等启发式算法整定了线性自抗扰控制器的参数。但这些算法在处理比较复杂的问题时常常陷入局部最优。

针对上述问题,本文将深度强化学习(deep reinforcement learning, DRL)与 LADRC 相结合,具有以下特点:

1) 通过 DDPG 智能体与环境交互学习 LADRC 参数整定策略,从而实现参数的智能化整定;

2) 将深度强化学习算法和经典 LADRC 相结合,充分利用 DRL 的自学习能力,提高系统稳定性和抗扰性。

考虑微网系统的耦合性以及 ADRC 控制过程的马尔可夫性,结合 DRL 在学习优化时不忽略参数间耦合关系^[13]以及其寻找最优策略的马尔可夫决策过程,将 DRL 引入到直流微网中接口变换器的自抗扰控制调参过程中。由于参数通常分布在连续空间中,而 DDPG 算法^[14]为行动者-评论家结构算法,通常用于解决连续动作空间问题^[15-17]。因此,本文选择了 DDPG 作为 DRL 算法对 LADRC 参数进行整定。通过 3 种控制器对同参数接口变换器控制过程的对比,验证了所提 DDPG-LADRC 在不同工况下的有效性。

1 被控对象与常规线性自抗扰控制策略

1.1 直流微网结构与被控对象建模

直流微网结构如图 1 所示,其主要由分布式电源(光伏发电和风力发电)、混合储能装置、负载以及各种接口变换器组成。本文主要针对直流母线与低压负载之间的降压变换器进行了研究。

图 1 主要由开关管 S、二极管 Q、电感 L、输出电容 C_{out} 以及负载 R_{load} 等组成,其中 v_g 和 v_o 分别为输入、输出电压。

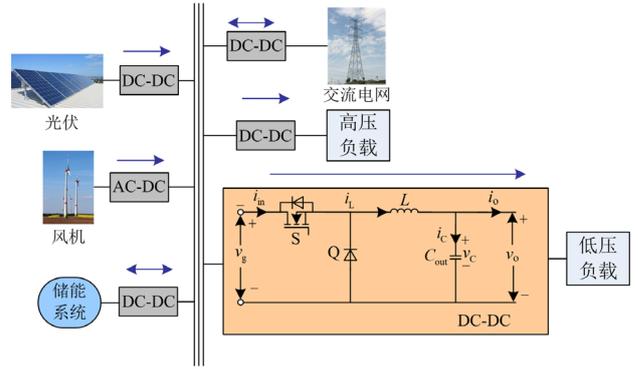


图 1 直流微网结构图

Fig. 1 Structure diagram of DC microgrid

通过开关器件的导通和关断对直流电的压降进行控制。针对一个周期 T_s 内开关管的导通与关断,电路的工作状态被分为 $[0, DT_s)$ 和 $[DT_s, T_s)$ 两部分,其中 D 为占空比, $i_c(t)$ 为电容电流, $i_L(t)$ 为电感电流, $v_L(t)$ 为电感电压。为了便于分析,忽略开关管导通阻抗和寄生电容。

分别对两个工作状态下的 $v_L(t)$ 和 $i_c(t)$ 建模,采用状态空间平均法^[18]可得

$$\begin{bmatrix} \dot{i}_L(t) \\ \dot{v}_o(t) \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C_{out}} & -\frac{1}{R_{load}C_{out}} \end{bmatrix} \begin{bmatrix} i_L(t) \\ v_o(t) \end{bmatrix} + \begin{bmatrix} D \\ \frac{D}{L} \\ 0 \end{bmatrix} v_g(t) \quad (1)$$

由式(1)可得

$$\begin{cases} L \frac{di_L(t)}{dt} = Dv_g(t) - v_o(t) \\ C_{out} \frac{dv_o(t)}{dt} = i_L(t) - \frac{v_o(t)}{R_{load}} \end{cases} \quad (2)$$

对式(2)进行拉普拉斯变换并利用小信号分析法^[18-19]可得 DC-DC 变换器控制量到输出电压之间的传递函数 $G_{vd}(s)$ 。

$$G_{vd}(s) = \frac{\hat{v}_o(s)}{\hat{D}(s)} = \frac{v_g}{s^2 LC_{out} + \frac{L}{R_{load}}s + 1} \quad (3)$$

1.2 传统线性自抗扰控制器

线性自抗扰控制器主要由线性扩张状态观测器(linear extended state observer, LESO)、线性状态误差反馈控制率(linear state error feedback control, LSEFC)两部分组成。其中,LESO 通过检测系统内部输入、输出信号来即时估计和干预系统中的总扰动;LSEFC 用于控制系统误差^[20-21]。

针对式(3)所示的二阶被控对象,选用图 2 所示二阶 LADRC 进行控制。

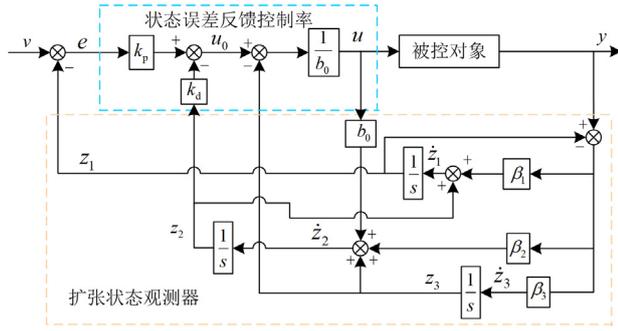


图 2 LADRC 结构示意图

Fig. 2 Schematic diagram of the LADRC structure

含扰动的二阶系统可以表示成

$$\ddot{y} = g(y, \dot{y}, w, t) + bu \quad (4)$$

式中： u 为系统控制量； w 为扰动量； y 为系统输出量； g 表示被控系统模型变量之间的耦合关系； b 为控制量增益，通常情况下， b 用 b_0 估计^[20]。

设 $f(y, \dot{y}, w, t) = g(y, \dot{y}, w, t) + (b - b_0)u$ 为系统广义扰动，则式(4)可改写为

$$\ddot{y} = f + b_0u \quad (5)$$

因总扰动 f 可导，且 $\dot{f} = h$ ，将 $y(t)$ 、 $\dot{y}(t)$ 及 $f(t)$ 分别用状态变量 x_1 、 x_2 、 x_3 来表示，即 $x_1 = y$ 、 $x_2 = \dot{y}$ 、 $x_3 = f$ ，得到状态方程(6)。

$$\begin{cases} \dot{x} = Ax + Bu + Eh \\ y = Cx \end{cases} \quad (6)$$

其中

$$\begin{cases} A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\ B = [0 \quad b_0 \quad 0]^T \\ E = [0 \quad 0 \quad 1] \\ C = [1 \quad 0 \quad 0] \end{cases} \quad (7)$$

将扩张状态观测器设计成

$$\begin{cases} \dot{z} = Az + Bu + L(x_1 - z_1) \\ \hat{y} = Cz \end{cases} \quad (8)$$

式中： z 为状态变量的估计值； L 为观测器增益矩阵 $L = [\beta_1, \beta_2, \beta_3]^T$ ，选择合适的 L 可实现对系统中各变量的实时跟踪。

设计 LSEFC 为

$$u_0 = k_p(v - z_1) - k_d z_2 \quad (9)$$

式中： k_p 、 k_d 为控制器增益系数； u_0 为控制器输出。

系统最终控制量为

$$u = \frac{u_0 - z_3}{b_0} \quad (10)$$

将式(10)代入式(5)可知，系统会被简化为双积分器串联标准型结构。

$$\ddot{y} = (f - z_3) + u_0 \approx u_0 \quad (11)$$

采用带宽法思想^[10]将 LESO 的参数配置在极点处，可得

$$\begin{cases} \beta_1 = 3\omega_c \\ \beta_2 = 3\omega_c^2 \\ \beta_3 = \omega_c^3 \end{cases} \quad (12)$$

类似地，可以得到控制器增益如式(13)所示。

$$\begin{cases} k_p = \omega_c^2 \\ k_d = 2\omega_c \end{cases} \quad (13)$$

因此，对 LADRC 的参数配置可以看作是对控制器带宽 ω_c 和观测器带宽 ω_o 的选取问题。

对式(3)进行拉普拉斯反变换可得

$$\ddot{y} = -a_0 y(t) - a_1 \dot{y}(t) + w(t) + b_0 u(t) + (b - b_0)u(t) \quad (14)$$

$$\begin{cases} a_0 = \frac{1}{LC_{out}} \\ a_1 = \frac{1}{C_{out} R_{load}} \\ b_0 = \frac{v_g}{LC_{out}} \end{cases} \quad (15)$$

根据式(6)和式(12)可得

$$\begin{cases} z_1 = \frac{3\omega_o s^2 + 3\omega_o^2 s + \omega_o^3}{(s + \omega_o)^3} y + \frac{b_0 s}{(s + \omega_o)^3} u \\ z_2 = \frac{(3\omega_o s + \omega_o^3)s}{(s + \omega_o)^3} y + \frac{b_0 (s + 3\omega_o)s}{(s + \omega_o)^3} u \\ z_3 = \frac{\omega_o^3 s^2}{(s + \omega_o)^3} y - \frac{b_0 \omega_o^3}{(s + \omega_o)^3} u \end{cases} \quad (16)$$

根据式(9)、式(10)和式(13)可得

$$u = \frac{1}{b_0} (\omega_c^2 (v - z_1) - 2\omega_c z_2 - z_3) \quad (17)$$

将式(17)代入式(16)可得

$$u = \left[\frac{(s + \omega_o)^3 / b_0 \cdot ((s + \omega_o)^3 + 2\omega_c s^2 + (\omega_c^2 + 6\omega_c \omega_o)s - \omega_o^3)}{\omega_c^2 v - \left((3\omega_c^2 \omega_o + 6\omega_c \omega_o^2 + \omega_o^3)s^2 + (3\omega_c^2 \omega_o^2 + 2\omega_c \omega_o^3)s + \omega_c^2 \omega_o^2 \right) \cdot y / (s + \omega_o)^3} \right] \quad (18)$$

根据式(18)可以将系统结构表示为图 3。

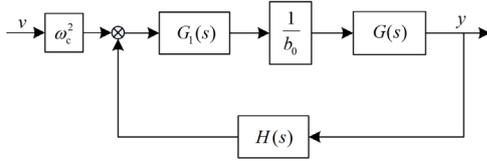


图3 系统结构图

Fig. 3 System structure diagram

$$G_1(s) = \frac{(s + \omega_0)^3}{(s + \omega_0)^3 + 2\omega_c s^2 + (\omega_c^2 + 6\omega_c \omega_0)s - \omega_0^3} \quad (19)$$

$$H(s) = \left[\frac{(3\omega_c^2 \omega_0 + 6\omega_c \omega_0^2 + \omega_0^3)s^2 + (3\omega_c^2 \omega_0^2 + 2\omega_c \omega_0^3)s + \omega_c^2 \omega_0^2}{(s + \omega_0)^3} \right] \quad (20)$$

求得系统的闭环传递函数 $G_c(s)$ 为

$$G_c(s) = \frac{\omega_c^2 G_1(s) G(s) / b_0}{1 + G_1(s) G(s) H(s) / b_0} \quad (21)$$

其中, $G(s)$ 为 1.1 节所求变换器的传递函数。由于 ω_0 与 ω_c 之间存在 k 倍的关系, $k \in (2, 10)^{[19]}$, b_0 已知, 根据赫尔维茨判据可以求得系统稳定的 ω_0 的范围。从而使参数的取值范围缩小, 加快深度强化学习算法训练的收敛速度。

2 DDPG-LADRC 设计

2.1 LADRC 控制系统的深度强化学习框架

智能体在环境中学习, 根据环境的状态执行动作并根据环境的奖惩指导更好的动作。其原理图如图 4 所示。

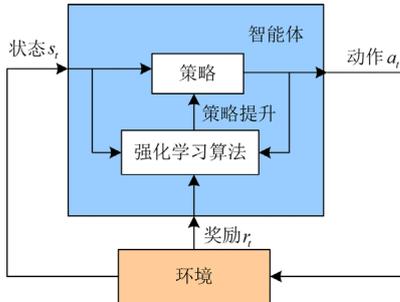


图4 强化学习原理图

Fig. 4 Reinforcement learning schematic

所提控制策略 DDPG-LADRC 将整个变换器控制系统视为环境, 以系统的控制效果为奖励评价标准。DDPG-LADRC 智能体基于当前环境确定动作, 然后环境根据输出值给出新的状态, 并计算奖励值。DDPG-LADRC 智能体做出判断, 优化更新下一个动作并与环境交互, 直到奖励收敛。这一过程的本质是马尔可夫决策过程(Markov decision process, MDP)^[22], 主要包括状态空间 S 、状态转移概率 P 、动作空间 A 、折扣因子 γ 和奖励函数 R 五个元素,

记为 $\{S, A, P, \gamma, R\}$ 。

分析可得 ADRC 的控制过程具有马尔可夫性, 可看作一类序列决策问题。由此将 DDPG 算法集成到 DC-DC 变换器的 LADRC 设计中以建立 DRL 模型, 从而实现较好控制效果。结合 LADRC 和 MDP 的 DDPG-LADRC 控制逻辑框图如图 5 所示。

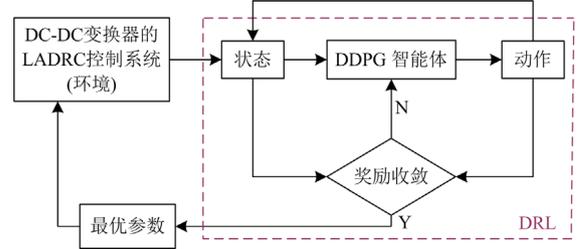


图5 DDPG-LADRC 逻辑结构

Fig. 5 DDPG-LADRC logic structure

根据接口变换器的结构模型定义系统的状态空间 S 、动作空间 A 以及奖励函数 R 。

1) 状态空间选取 (e, y)

(1) 误差 e 包括以下三部分: 变换器的输出电压 y 与观测值 z_1 之间的误差 e_1 ; 观测值 z_1 与输入的参考电压 v 之间的误差 e_2 ; 变换器输出电压 y 与输入参考电压 v 之间误差的微分 \dot{e}_3 。

(2) 变换器的输出电压 y 。

2) 奖励函数 R

为了减少最终误差, 根据变换器的最终输出电压 y 与期望电压设定值 v 之间误差 e_3 , 将奖励函数设计为分段函数的形式, 从而训练智能体学习到较好的控制策略。其中, 奖励函数表示为

$$R = \begin{cases} 200 & e_3 < 0.5 \\ 20 & 0.5 \leq e_3 < 2 \\ -10 & 2 \leq e_3 < 10 \\ -300 & e_3 \geq 10 \end{cases} \quad (22)$$

2.2 深度强化学习和 DDPG 算法

在 t 时刻, 从状态 s 到动作 a 的映射称为策略 π 。

$$a_t = \pi(s_t) \quad (23)$$

通过重复 MDP 可以根据策略 π 产生的动作连续获得新的状态和奖励值 r , 累积奖励的计算公式为

$$G_t = \sum_{t=1}^T \gamma^t r^t \quad (24)$$

状态价值函数的贝尔曼方程表示为

$$V_\pi(s) = E_\pi \left[\sum_{k=0}^{\infty} G_t \mid s = s_t \right] = E_\pi [r_{t+1} + \gamma V_\pi(s_{t+1}) \mid s = s_t] \quad (25)$$

考虑动作对值函数的影响, 状态动作价值函数

的贝尔曼方程表示为

$$Q_{\pi}(s, a) = E_{\pi}[r_{t+1} + \gamma Q_{\pi}(s_{t+1}, a_{t+1}) | s = s_t, a = a_t] \quad (26)$$

最优贝尔曼方程表示为

$$Q^*(s, a) = r_{s_t}^a + \gamma Q_{\pi^*}(s_{t+1}, a_{t+1}) \quad (27)$$

最优策略 π^* 通过最大化累积奖励以及其对应的最优贝尔曼方程得到

$$\pi(s) \rightarrow a_t = \arg \max Q_{\pi^*}(s_t, a_t) \quad (28)$$

DDPG 算法包括价值网络(Critic)、策略网络(Actor)和经验回放池, Actor 和 Critic 各包含一个当前网络和目标网络, 当前网络用于训练, 目标网络用于更新参数。经验回放池使网络可以学习历史经验数据。从而增强智能体学习的稳定性, 使算法收敛更加可靠。

根据式(29)计算损失函数, y_t 为目标 Q 值。

$$\begin{cases} y_t = r_t + \gamma Q'(s_{t+1}, \pi'(s_{t+1} | \theta^{\pi'})) | \theta^{Q'} \\ L(\theta^Q) = E(y_t - Q(s_t, a_t | \theta^Q))^2 \end{cases} \quad (29)$$

根据梯度更新法则, 通过对损失函数求梯度更新当前价值网络^[23]:

$$\begin{cases} \theta_k^Q = \theta_{k-1}^Q - \mu_Q \nabla_{\theta^Q} L(\theta_{k-1}^Q) \\ \nabla_{\theta^Q} L(\theta_{k-1}^Q) = E(2(y_t - Q(s, a | \theta_{k-1}^Q)) \cdot \nabla_{\theta^Q} Q(s, a | \theta_{k-1}^Q) |_{s=s_t, a=a_t}) \\ \nabla_{\theta^Q} Q(s, a | \theta_{k-1}^Q) |_{s=s_t, a=a_t} \end{cases} \quad (30)$$

策略网络使用价值网络输出的 Q 函数作为损失函数。通过对 Q 函数求策略梯度, 得到更新公式(31)^[24]。

$$\begin{cases} \theta_k^{\pi} = \theta_{k-1}^{\pi} - \mu_{\pi} \nabla_{\theta^{\pi}} L(\theta_{k-1}^{\pi}) \\ \nabla_{\theta^{\pi}} L(\theta_{k-1}^{\pi}) = \nabla_a Q(s, a | \theta_{k-1}^{\pi}) |_{s=s_t, a=\pi(s_t)} \cdot \nabla_{\theta^{\pi}} \pi(s | \theta_{k-1}^{\pi}) |_{s=s_t} \end{cases} \quad (31)$$

目标网络采用式(32)所示的软更新方式。

$$\begin{cases} \theta_k^{Q'} = \tau \theta_{k-1}^{Q'} + (1 - \tau) \theta_{k-1}^{Q'} \\ \theta_k^{\pi'} = \tau \theta_{k-1}^{\pi'} + (1 - \tau) \theta_{k-1}^{\pi'} \end{cases} \quad (32)$$

上述分析中各个参数所代表的含义如表 1 所示。

2.3 DDPG 算法训练模型

DDPG 算法模型结构如图 6 所示。由图 6 可知, DDPG 智能体将与 LADRC 控制系统交互获得的样本数据 $\{a_t, s_t, r_t, s_{t+1}\}$ 存放至经验池中, 在学习策略时, 随机从经验池中采样 m 个样本数据, 通过不断迭代更新网络梯度值实现算法的优化。

本文未选择图像作为输入, 因此采用了全连接网络构建 DDPG 网络。通过以下三部分保证过程收敛。

Actor: 根据当前环境状态和噪声选取动作值, 输入为 2.1 节设定的四维状态变量 $S = \{e_1, e_2, \dot{e}_3, y\}$,

表 1 DDPG 算法参数及含义

Table 1 DDPG algorithm parameters and meaning

算法参数	含义
$Q(s, a \theta^Q)$	当前价值网络在 t 时刻输出的 Q 值
$Q'(s_{t+1}, \pi'(s_{t+1} \theta^{\pi'})) \theta^{Q'}$	目标网络的输入 Q 值
$\pi'(s_{t+1} \theta^{\pi'})$	目标策略网络输出的动作变量
$\theta_k^Q, \theta_k^{\pi}$	第 k 轮学习时当前网络的参数
μ_Q	价值网络的学习率
$\nabla_{\theta^Q} L(\theta_{k-1}^Q)$	损失函数 $L(\theta_{k-1}^Q)$ 对参数 θ_{k-1}^Q 的梯度
μ_{π}	策略网络的学习率
$\nabla_{\theta^{\pi}} \pi$	策略梯度
$\theta_k^{Q'}, \theta_k^{\pi'}$	第 k 次学习时目标网络参数
τ	软更新系数

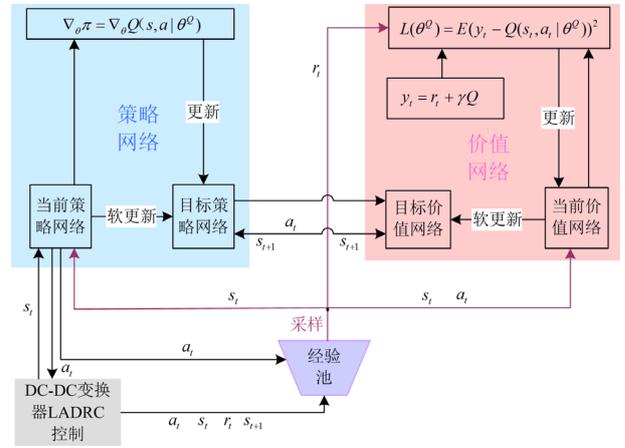


图 6 DDPG 算法模型结构

Fig. 6 DDPG algorithm model structure

输出为一维动作变量。

Critic: 用于评估当前策略网络输出动作值的好坏。其将状态 s_t 和动作 a_t 作为输入, 输出为状态 s_t 下的状态动作价值函数 $Q(s_t, a_t)$ 。

噪声: DDPG 算法是确定性策略, 为确定性策略加入噪声, 如式(33)所示, 使得智能体更加有效地探索环境, 防止其陷入局部最优, 通过确定性策略梯度使 Critic 收敛并更新网络参数。

$$a_t = \mu_{\theta}(s_t) + N_{\text{oise}} \quad (33)$$

基于 DDPG 算法的接口变换器 LADRC 参数整定流程如图 7 所示。

3 模型训练

3.1 控制策略设计

使用 DRL 作为 LADRC 参数自动调节机制, 通过 DDPG 算法使网络收敛。根据 1.2 节的分析, 通过训练观测器带宽 ω_o , 可得 LESO 的参数。因此,

在训练时保持 LSEFC 参数不变。为加快训练的收敛速度, 通过李纳德-威帕特稳定判据对式(21)进行计算, 并通过多次仿真调试将 ω_0 的范围从 $[0, +\infty)$ 缩减到 $(1000, +\infty)$ 。同时, 在训练时使直流母线侧电压和负载产生一定程度跳变, 以模拟系统复杂的干扰环境, 增强所提方法的抗扰性。本文设计的控制策略如图 8 所示。

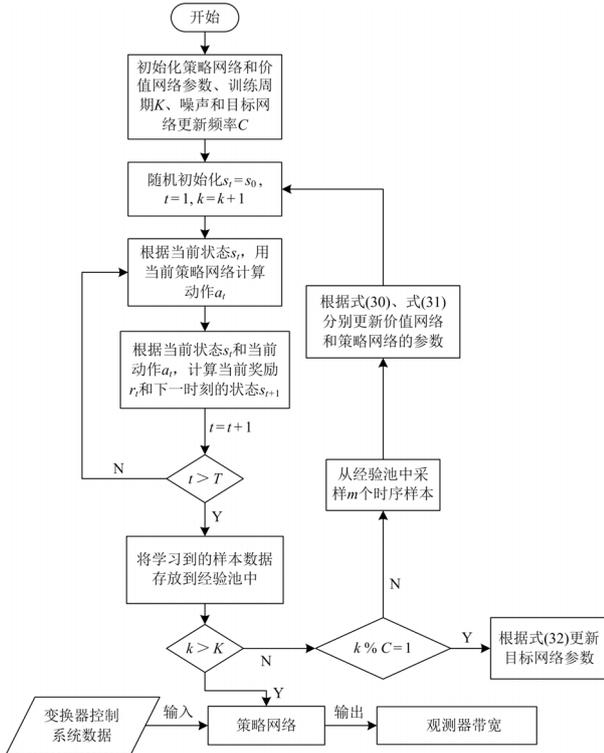


图 7 基于 DDPG 算法的参数整定流程图

Fig. 7 Flow chart of parameter tuning based on DDPG

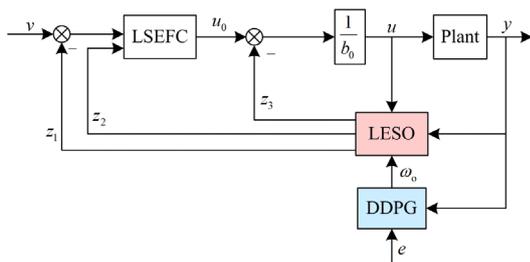


图 8 DDPG-LADRC 控制策略

Fig. 8 Control strategy for DDPG-LADRC

3.2 算法训练过程

在进行训练时, 对 DDPG 算法中学习率 α 和折扣因子 γ 等参数的设置也至关重要。 α 控制网络模型的学习进度, 过大会使网络在最优值附近徘徊不能收敛, 过小会导致网络收敛缓慢^[15,24]。 γ 表示未来奖励对当前的重要程度, 取值范围为 $(0,1)$, 应在

算法能够收敛的前提下选择较大的折扣因子^[25]。经验池用来提高数据利用效率, 超过设定值, 程序会自动删除旧的数据^[26-27]。

经过多次调试, 最终选定的 DDPG 参数如表 2 所示。

表 2 DDPG 算法参数

Table 2 Parameters of DDPG algorithm

超参数	值
演员网络学习率	0.00001
评论家网络学习率	0.0001
小批量采样样本数	128
折扣因子	0.97
经验池大小	1 000 000
噪声方差	0.4
噪声衰减系数	0.000 01
每个回合持续的最大时间步长	100

平均奖励代表智能体在一段时间内获得奖励的平均值, 一般用其判断智能体训练是否收敛^[15], 训练后的奖励随训练轮次变化曲线如图 9 所示。

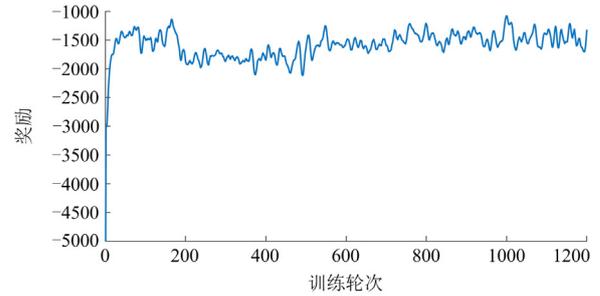


图 9 训练奖励变化曲线

Fig. 9 Training reward change curve

4 仿真分析

本文依据试错寻优整定出双闭环 PI 控制器和传统 LADRC 的参数, 并从跟踪性、直流母线侧电压突变、负载突变和鲁棒性角度对双闭环 PI 控制器、LADRC 和 DDPG-LADRC 的控制效果进行仿真对比分析。

4.1 仿真参数

DC-DC 变换器参数如表 3 所示。

双闭环 PI 控制器和传统 LADRC 的参数为: $\{K_{Pi} = 0.1, K_{Ii} = 55, K_{Pv} = 0.55, K_{Iv} = 900, k_p = 39690000, k_d = 12600, \beta_1 = 6 \times 10^4, \beta_2 = 12 \times 10^8, \beta_3 = 8 \times 10^{12}\}$ 。

4.2 仿真结果与分析

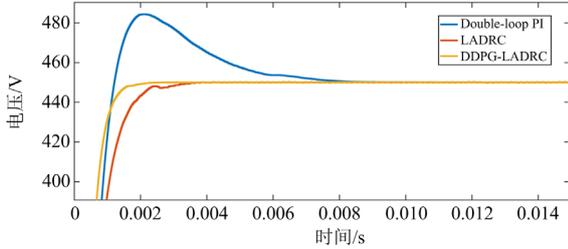
1) 跟踪性能对比

图 10 的仿真结果主要性能指标见表 4 和表 5。

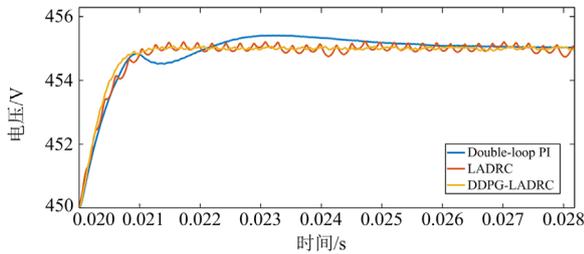
表 3 DC-DC 变换器参数

Table 3 DC-DC converter simulation parameters

参数	数值	参数	数值
参考电压/V	450	负载电阻/ Ω	5
开关频率/kHz	100	电容/ μF	300
输入侧电压/V	550	电感/ μH	120



(a) 对起始阶段的450 V启动信号的跟踪



(b) 对稳态下的+5 V电压调整信号跟踪

图 10 跟踪性能对比

Fig. 10 Traceability comparison

表 4 起始阶段跟踪效果

Table 4 Tracking performance at the beginning stage

性能指标	PI	LADRC	DDPG-LADRC
$\sigma\%$	9.04	0.044	0
调节时间/ms	13	4.039	2.814

表 5 电压调整信号跟踪效果

Table 5 Voltage adjustment signal tracking effect

性能指标	PI	LADRC	DDPG-LADRC
$\sigma\%$	0.088	0.044	0
调节时间/ms	9	2.16	2.01

表中, σ 表示输出电压最大波动量, 为正时, 即超调量。由于加入了 DDPG 算法, DDPG-LADRC 参数可以实时调整, 因此, 在调节过程中相比于双闭环 PI 控制器和传统 LADRC, 其在跟踪输入信号时无超调, 且调节时间最短。同时, 在对+5 V 电压调整信号进行跟踪时, 相比于双闭环 PI 控制器和传统 LADRC, DDPG-LADRC 在超调抑制以及跟踪性方面均具有明显优越性。

2) 抗扰性分析

工况一: 直流母线侧电压突变

考虑直流母线侧电压波动对系统的影响, 在 0.05 s 处分别加入直流母线侧电压突增 20%和突减 10%的扰动。低压侧输出电压动态变化波形如图 11 和图 12 所示, 不同扰动下系统的电压最大波动量和调节时间分别如表 6 和表 7 所示。

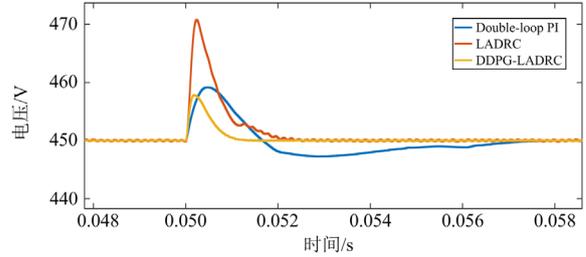


图 11 电压突增 20%

Fig. 11 Voltage surge by 20%

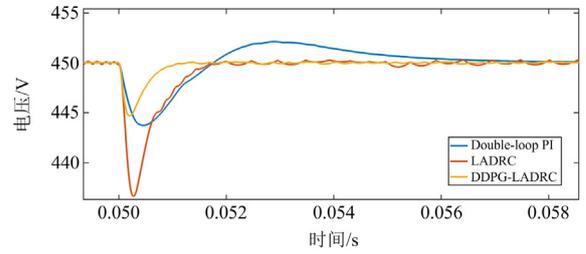


图 12 电压突减 10%

Fig. 12 Voltage sag by 10%

表 6 电压突增情况下的性能指标

Table 6 Performance indicators under voltage surge

性能指标	PI	LADRC	DDPG-LADRC
$\sigma\%$	2.09	4.6	1.77
调节时间/ms	8.9	2.3	1.3

表 7 电压突减情况下的性能指标

Table 7 Performance indicators under voltage sag

性能指标	PI	LADRC	DDPG-LADRC
$\sigma\%$	-1.36	-2.97	-1.15
调节时间/ms	9.4	2.1	1.2

由表 6 和表 7 可知, 针对直流母线电压波动的扰动抑制, 在受到直流母线电压突增时, 与双闭环 PI 控制器和传统 LADRC 相比, DDPG-LADRC 分别优化 0.32%和 2.83%的超调量。此外, 在 DDPG-LADRC 控制下调节时间也明显优于另外两种控制方式。在受到直流母线电压突减时, DDPG-LADRC 控制下的电压最大波动量小于双闭环 PI 控制器和传统 LADRC 的, 且可以使系统更快的恢复到稳定状态。

工况二: 负载突变

考虑负载突变对系统的干扰, 在低压负载波动

工况下, 分析输出电压动态波形, 在 0.05 s 处施加突增 20% 和突减 20% 的扰动。负载波动下输出侧电压波形如图 13 和图 14 所示。

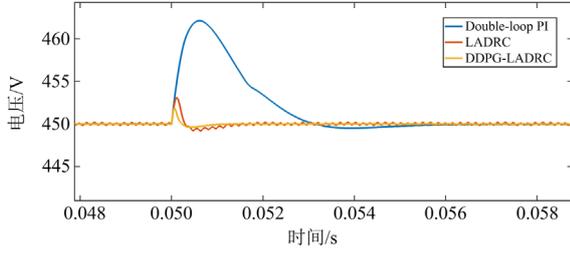


图 13 负载突增 20%

Fig. 13 Load surge by 20%

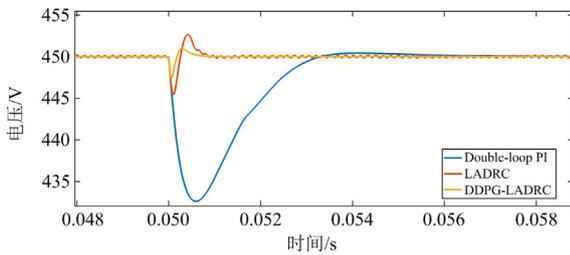


图 14 负载突减 20%

Fig. 14 Load sag by 20%

由表 8、表 9 可知, 针对负载突增 20% 的扰动, DDPG-LADRC 相比于双闭环 PI 控制器和传统 LADRC 分别优化 2.29% 与 0.29% 的超调量。针对负载突减 20% 的扰动, DDPG-LADRC 的电压最大波动量相比于双闭环 PI 控制器和 LADRC 分别减小了 3.27% 和 0.45%。且在两种工况下, DDPG-LADRC 的调节时间均最短。

表 8 负载突增情况下性能指标

Table 8 Performance indicators under load surge

性能指标	PI	LADRC	DDPG-LADRC
$\sigma\%$	2.69	0.69	0.40
调节时间/ms	6.5	1.4	1.0

表 9 负载突减情况下性能指标

Table 9 Performance indicators under load sag

性能指标	PI	LADRC	DDPG-LADRC
$\sigma\%$	-3.84	-1.02	-0.57
调节时间/ms	1.0	0.8	0.5

3) 鲁棒性分析

保持控制器的参数不变, 令被控系统内部的电容和电感值在 $\pm 20\%$ 内随机波动, 对 3 种控制器各进行 20 次重复实验。选取平方误差积分(integration of square error, ISE)以及 ISE 的分布(distribution of ISE, dISE)作为性能指标。其中, dISE 表示每个 ISE 指标与该组 ISE 指标平均值之间的误差值。性能指标

点分布越靠近原点, 说明系统具有更好的跟踪性和抗扰性; 分布越密集说明系统具有更好的鲁棒性, 绘制系统的散点图如图 15 所示。

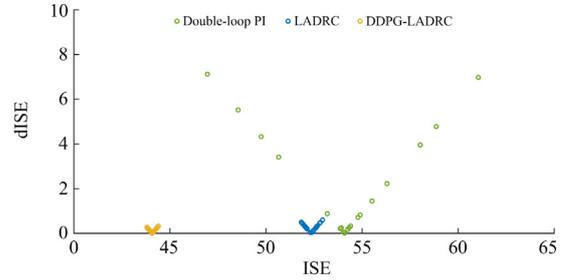


图 15 鲁棒性对比散点图

Fig. 15 Robustness comparison scatter chart

由图 15 可知, 在内部参数偏离标称值的情况下, DC-DC 变换器在 3 种控制器的作用下均可在较小数量级的恢复过程后实现对目标信号的稳定跟踪, 且 DDPG-LADRC 相比于 LADRC 和双闭环 PI 控制器的 ISE 性能指标分布更靠近原点, 说明其具有更好的跟踪性和抗扰性。同时, DDPG-LADRC 的性能指标点分布更加集中, 说明其鲁棒性也明显优于另外两种控制器。

4) 系统整体性能分析

误差积分型指标能够反映系统动态响应的综合性能。因此, 选取了绝对误差积分(integration of absolute error, IAE)和平方误差积分(ISE)两个指标分别对 3 种控制器进行对比分析。

绝对误差积分 J_{IAE} 为

$$J_{IAE} = \int_0^{\infty} |e(t)| dt \quad (34)$$

3 种控制器在不同工况下的 IAE 值如图 16 所示。

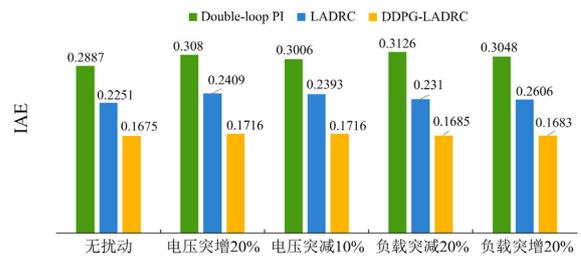


图 16 不同工况下的 IAE 值

Fig. 16 IAE values under different working conditions

平方误差积分 J_{ISE} 为

$$J_{ISE} = \int_0^{\infty} [e(t)]^2 dt \quad (35)$$

3 种控制器在不同工况下的 ISE 值如图 17 所示。由图 16、图 17 可知, 在各种工况下, DDPG-LADRC 整体性能均优于双闭环 PI 控制器和传统 LADRC; 当发生电压突变和负载突变时, IAE 值和 ISE 值均

会增大；在无扰动时，DDPG-LADRC 的 IAE 值分别降低了 25.59%和 41.98%，ISE 值分别降低了 15.76%和 18.29%。

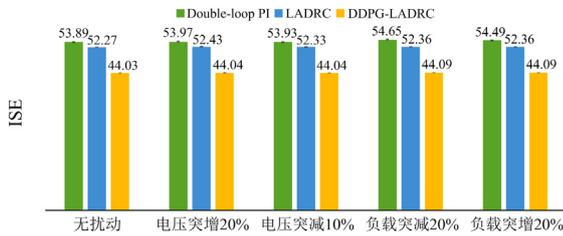


图 17 不同工况下的 ISE 值

Fig. 17 ISE values under different working conditions

综上，根据上述 3 种控制器在不同工况下的对比仿真分析结果可知，本文提出的 DDPG-LADRC 在跟踪性、抗扰性以及鲁棒性等方面均优于双闭环 PI 控制器和传统 LADRC。

5 结论

针对直流微网中负载端接口变换器非线性、时变的结构特点以及由于工况不确定性而导致控制效果不佳的问题，本文利用自抗扰控制器鲁棒性强的特点，在深度强化学习的框架下，提出了一种通过 DDPG 算法实时整定 LADRC 参数的策略。并在数字仿真平台上搭建相应模型，将双闭环 PI 控制器、传统 LADRC 以及本文提出的 DDPG-LADRC 的控制效果进行了对比分析。实验表明，本文提出的 DDPG-LADRC 在跟踪性、抗扰性以及鲁棒性等方面均有较显著的优势，最后通过 IAE 和 ISE 两个指标对系统的整体性能进行分析比较，指出了本文所提出的参数整定策略的可行性以及有效性。显然，相比于双闭环 PI 控制器和传统 LADRC，本文提出的 DDPG-LADRC 更适用于多变量、强耦合、含有较大随机性以及未知扰动的微电网系统。

参考文献

- [1] 高强, 刘畅, 金道杰, 等. 考虑综合需求响应的园区综合能源系统优化配置[J]. 高压电器, 2021, 57(8): 159-168.
GAO Qiang, LIU Chang, JIN Daojie, et al. Optimal configuration of park-level integrated energy system considering integrated demand response[J]. High Voltage Apparatus, 2021, 57(8): 159-168.
- [2] 程林, 万宇翔, 齐宁, 等. 含多种分布式资源的配电系统运行可靠性研究评述及展望[J]. 电力系统自动化, 2021, 45(22): 191-207.
CHENG Lin, WAN Yuxiang, QI Ning, et al. Review and prospect of research on operation reliability of power distribution and consumption system considering various distributed energy resources[J]. Automation of Electric Power Systems, 2021, 45(22): 191-207.
- [3] 蒋谦, 黄志豪, 程启明, 等. 基于 MMC 的 PET 中间隔离级 DC-DC 变换器的新型模型预测控制策略[J]. 电力建设, 2022, 43(4): 49-57.
JIANG Qian, HUANG Zhihao, CHENG Qiming, et al. Novel model predictive control strategy for mmc-based dc-dc converter in the intermediate isolation stage of PET[J]. Electric Power Construction, 2022, 43(4): 49-57.
- [4] 杨晓峰, 郑琼林, 林智钦, 等. 用于直流电网的大容量 DC/DC 变换器研究综述[J]. 电网技术, 2016, 40(3): 670-677.
YANG Xiaofeng, ZHENG Qionglin, LIN Zhiqin, et al. Survey of high-power DC/DC converter for HVDC grid application[J]. Power System Technology, 2016, 40(3): 670-677.
- [5] 胡鹏飞, 朱乃璇, 江道灼, 等. 柔性互联智能配电网关键技术研究进展与展望[J]. 电力系统自动化, 2021, 45(8): 2-12.
HU Pengfei, ZHU Naixuan, JIANG Daozhuo, et al. Research progress and prospects of key technologies of flexible interconnected smart distribution network[J]. Automation of Electric Power Systems, 2021, 45(8): 2-12.
- [6] 张宇涵, 杜贵平, 雷雁雄, 等. 直流微网混合储能系统控制策略现状及展望[J]. 电力系统保护与控制, 2021, 49(3): 177-187.
ZHANG Yuhuan, DU Guiping, LEI Yanxiong, et al. Current status and prospects of control strategy for a DC microgrid hybrid energy storage system[J]. Power System Protection and Control, 2021, 49(3): 177-187.
- [7] 李虹, 尚佳宁, 陈姚, 等. 基于 fal 函数的非线性 PI 控制器在 DC-DC 变换器中的应用[J]. 电工技术学报, 2014, 29(增刊 1): 326-331.
LI Hong, SHANG Jianing, CHEN Yao, et al. The applications of nonlinear PI controller based on the fal function in the DC-DC converter[J]. Transactions of China Electrotechnical Society, 2014, 29(S1): 326-331.
- [8] 李乔, 吴捷. 自抗扰控制及其在 DC-DC 变换器中的应用[J]. 电工技术学报, 2005, 20(1): 83-88.
LI Qiao, WU Jie. Auto-disturbance rejection control and its application to DC-DC converter[J]. Transactions of China Electrotechnical Society, 2005, 20(1): 83-88.
- [9] GAO Z. Scaling and bandwidth-parameterization based controller tuning[C] // ACC, 2003: 4989-4996.
- [10] 孙雨萌, 张旭秀. 改进遗传算法整定自抗扰控制器参数及应用[J]. 自动化与仪表, 2020, 35(3): 13-17, 45.
SUN Yumeng, ZHANG Xuxiu. Parameters optimization and application of auto-disturbance-rejection-controller based on improved genetic algorithm[J]. Automation & Instrumentation, 2020, 35(3): 13-17, 45.
- [11] 陈宗祥, 张武林, 陈克难, 等. LLC 谐振变换器自抗扰

- 控制研究[J]. 电机与控制学报, 2022, 26(8): 130-138.
CHEN Zongxiang, ZHANG Wulin, CHEN Kenan, et al. Research on active disturbance rejection control of LLC resonant converter[J]. Electric Machines and Control, 2022, 26(8): 130-138.
- [12] 魏立新, 赵默林, 范锐, 等. 基于改进鲨鱼优化算法的自抗扰控制参数整定[J]. 控制与决策, 2019, 34(4): 816-820.
WEI Lixin, ZHAO Molin, FAN Rui, et al. Parameter tuning of active disturbance rejection control based on ameliorated shark smell optimization algorithm[J]. Control and Decision, 2019, 34 (4): 816-820.
- [13] WANG Y, FANG S, HU J. Active disturbance rejection control based on deep reinforcement learning of PMSM for more electric aircraft[J]. IEEE Transactions on Power Electronics, 2022, 38(1): 406-416.
- [14] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv:1509.02971, 2015.
- [15] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2018, 41(1): 1-27.
LIU Quan, ZHAI Jianwei, ZHANG Zongchang, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27.
- [16] CAO D, HU W, ZHAO J, et al. Reinforcement learning and its applications in modern power and energy systems: a review[J]. Journal of Modern Power Systems and Clean Energy, 2020, 8(6): 1029-1042.
- [17] 张延宇, 饶新朋, 周书奎, 等. 基于深度强化学习的电动汽车充电调度算法研究进展[J]. 电力系统保护与控制, 2022, 50(16): 179-187.
ZHANG Yanyu, RAO Xinpeng, ZHOU Shukui, et al. Research progress of electric vehicle charging scheduling algorithms based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(16): 179-187.
- [18] 王瑶. 基于状态空间平均模型的电压控制 SIDO Buck 变换器稳定性分析[J]. 中国电机工程学报, 2018, 38(6): 1810-1817, 1918.
WANG Yao. Stability analysis for voltage controlled single-inductor dual-output buck converter based on state space average model[J]. Proceedings of the CSEE, 2018, 38(6): 1810-1817, 1918.
- [19] 陶珑, 王萍, 王议锋, 等. 微电网负载端接口变换器的自抗扰稳压控制[J]. 电工技术学报, 2022, 37(8): 2076-2085.
TAO Long, WANG Ping, WANG Yifeng, et al. Voltage stabilization strategy for load-side interface converter of microgrid combined with active disturbance rejection control[J]. Transactions of China Electrotechnical Society, 2022, 37(8): 2076-2085.
- [20] 韩京清. 自抗扰控制器及其应用[J]. 控制与决策, 1998, 13(1): 19-23.
HAN Jingqing. Auto-disturbances-rejection controller and it's applications[J]. Control and Decision, 1998, 13(1): 19-23.
- [21] 高志强. 自抗扰控制思想探究[J]. 控制理论与应用, 2013, 30(12): 1498-1510.
GAO Zhiqiang. On the foundation of active disturbance rejection control[J]. Control Theory & Applications, 2013, 30(12): 1498-1510.
- [22] JIA S, SHEN L, XUE H. Continuous-time Markov decision process with average reward: using reinforcement learning method[C] // 2015 34th Chinese Control Conference (CCC). IEEE, 2015: 3097-3100.
- [23] LIANG Y, GUO C, DING Z, et al. Agent-based modeling in electricity market using deep deterministic policy gradient algorithm[J]. IEEE Transactions on Power Systems, 2020, 35(6): 4180-4192.
- [24] SEHGAL A, LA H, LOUIS S, et al. Deep reinforcement learning using genetic algorithm for parameter optimization[C] // 2019 Third IEEE International Conference on Robotic Computing (IRC), IEEE, 2019: 596-601.
- [25] 尚立, 蔡硕, 崔俊彬, 等. 基于软件定义网络的电网边缘计算资源分配[J]. 电力系统保护与控制, 2021, 49(20): 136-143.
SHANG Li, CAI Shuo, CUI Junbin, et al. SDN-based MEC resource allocation of a power grid[J]. Power System Protection and Control, 2021, 49(20): 136-143.
- [26] 陈希亮, 曹雷, 李晨溪, 等. 基于重抽样优选缓存经验回放机制的深度强化学习方法[J]. 控制与决策, 2018, 33(4): 600-606.
CHEN Xiliang, CAO Lei, LI Chenxi, et al. Deep reinforcement learning via good choice resampling experience replay memory[J]. Control and Policy, 2018, 33(4): 600-606.
- [27] 杨思明, 单征, 丁煜, 等. 深度强化学习研究综述[J]. 计算机工程, 2021, 47(12): 19-29.
YANG Siming, SHAN Zheng, DING Yu, et al. Survey of research on deep reinforcement learning[J]. Computer Engineering, 2021, 47(12): 19-29.

收稿日期: 2023-04-23; 修回日期: 2023-07-25

作者简介:

周雪松(1964—), 男, 博士, 教授, 研究方向为新能源发电、电力系统控制; E-mail: sjteam2023@163.com

张心茹(2001—), 女, 硕士研究生, 研究方向为智能电网、电力系统保护; E-mail: 1658328066@qq.com

赵洽宇(1989—), 女, 通信作者, 博士, 副教授, 研究方向为新能源发电、磁特性测量。E-mail: zhaohanyu@email.tjut.edu.cn

(编辑 姜新丽)