

规模化新能源接入电网下的多区域协同算法

吴应双¹, 刘蔚², 刘明顺¹, 张野², 王寅¹, 唐王倩云²

(1. 贵州电网公司电力调度控制中心, 贵阳 550000; 2. 直流输电技术全国重点实验室(南方电网科学研究院), 广州 510663)

摘要: 可再生能源的规模化接入电网具有间歇性和强随机特性, 容易引起显著的频率波动。针对多区域互联电网, 现有的基于神经网络的多智能体协同算法在面对强随机扰动时, 常因梯度爆炸和梯度消失问题导致控制性能并不理想。为此, 提出了一种自适应权重残差近端策略优化自动发电控制方法。该方法将智能体策略神经网络的各层之间通过残差连接, 并为残差连接添加一个可训练的自适应权重, 以取得最优的残差比例, 缓解神经网络梯度爆炸和梯度消失问题, 从而使得深层网络的训练过程更加稳定和高效, 进而更快地获得强随机扰动下的多区域协同最优解, 以消除由于强随机扰动所引起的频率波动。通过在 IEEE 两区域和三区域负荷频率控制系统上的仿真测试验证了所提算法的有效性。且相较于多种强化学习算法, 所提算法具有更快的收敛性和更强的稳定性, 更高的频率稳定性及控制性能。

关键词: 多区域互联电网; 残差连接; 自适应权重; 近端策略优化

Multi-Area Cooperative Algorithm Under Large-Scale New Energy Access to Power Grid

WU Yingshuang¹, LIU Wei², LIU Mingshun¹, ZHANG Ye², WANG Yin¹, TANG Wangqianyun²

(1. Power Dispatching and Control Center of Guizhou Power Grid Co., Ltd., Guiyang 550000, China; 2. State Key Laboratory of HVDC, Electric Power Research Institute, CSG, Guangzhou 510663, China)

Abstract: The large-scale access of renewable energy to the power grid has intermittent and strong random characteristics, which is easy to cause significant frequency fluctuations. For multi-area power grids, the existing multi-agent cooperative algorithms based on neural networks often have unsatisfactory control performance due to gradient explosion and gradient disappearance in the face of strong random disturbances. To this end, this paper proposes an adaptive weight residual proximal policy optimization method for automatic power generation control. This method connects the layers of the agent policy neural network through residual connections, and adds a trainable adaptive weight to the residual connection to achieve the optimal residual ratio and alleviate the problem of gradient explosion and gradient disappearance of the neural network, so that the training process of the deep network is more stable and efficient, and then the multi-area cooperative optimal solution under strong random disturbance is obtained faster to eliminate the frequency fluctuation caused by strong random disturbance. The effectiveness of the proposed algorithm is verified by simulation tests on IEEE two-area and three-area load frequency control systems. Compared with various reinforcement learning algorithms, the proposed algorithm has faster convergence and stronger stability, higher frequency stability and control performance.

Key words: multi-area power grids; residual connection; adaptive weight; proximal policy optimization

0 引言

随着全球能源结构持续优化和可再生能源渗透率

不断提升, 部分地区电网呈现出由传统集中式向区域协同互联方向演进的趋势, 这对电网的运行控制技术提出了更高要求。作为电力系统二次调频的自动发电控制(automatic generation control, AGC)^[1]也将从中式向多区域互联结构转变。然而, 具有间歇性、强随机特性的规模化源、荷并网会引起较强的频率波动^[2], 进而对电网控制性能带来影响。为此, 学者们试图从多区域互联 AGC 的角度进行探索, 以寻找一

基金项目: 国家自然科学基金青年科学基金资助项目(52207110); 中国南方电网有限责任公司科技项目(GZKJXM20222217)。

Foundation item: Supported by the Youth Science Foundation Project of National Natural Science Foundation (52207110); the Science and Technology Project of China Southern Power Grid Co., Ltd. (GZKJXM20222217).

种有效的协同控制方法来应对集中式电网所面临的强随机扰动^[3-6]问题。

现今主流的AGC控制方法分为模型驱动和数据驱动两种。在基于模型的方法中,文献[7]基于多事件触发机制提出分布式风力自动发电控制方法,充分利用了各个子系统的事件信息,提高了分布式发电网络的资源利用效率。文献[8]基于最优里程分布和分布式自适应拍卖算法开发了一种AGC功率分配方法,该方法改善了因大规模新能源并网而引起的随机波动问题,提高了电网对负荷的响应能力。文献[9]采用增强Lyapunov-Krasovskii函数对考虑电动汽车(electric vehicle, EV)并网通信时延的负荷频率控制系统(load frequency control, LFC)进行稳定性分析,结合线性矩阵不等式对系统的稳定性的判定更加精准,EV的需求响应能力得到了充分挖掘。文献[10]基于模型预测控制,重点讨论了EV并网技术的优势和不足,结合风力发电技术验证了EV参与LFC的可行性,提高了新能源的利用率。然而上述控制策略能改善随机扰动对电网的不利影响的前提是首先需要获得详细的系统模型或部分参数,这在如今可持续能源规模化并网的新型电力系统中难以实现并且工作量巨大。在这一背景下,学者们开始对数据驱动方法例如回归分析、分类算法、推荐系统和强化学习^[11]等进行研究,而强化学习因具有在线优化和持续的学习能力受到广泛关注。

自Watkins提出Q-learning^[12]以来,强化学习便不断发展。文献[13]率先将Q学习算法应用于AGC的负荷频率控制,并且在CPS标准下取得了较为理想的控制效果。为实现电力系统中大规模光伏整列的构建排布优化,文献[14]利用改进群双Q学习算法与储氢系统联动有效降低了电力调节成本,最高降低调节功率达89.22%。但Q学习及其衍生算法除了Q值过估计问题之外,还存在状态-动作值离散化的局限性,这导致工程师在设计控制器之前需要评估电网的负载特性,依据负载特性设置相应大小的状态-动作矩阵,这难以面对高维度状态-动作^[15]任务,显然限制了经典Q学习算法的应用范围,亟需开发一种泛化能力^[16]更强的多智能体协同方法。

在过去10年中,基于神经网络(neural network, NN)的深度强化学习迎来了快速发展期,在电力系统中的应用也越来越受到重视。在近5年对深度强化学习的研究中,文献[17]基于双重深度Q网络(double

deep Q-network, DDQN)和演员-评论家(action discovery, AD)策略提出了DDQN-AD深度强化学习算法,通过NN解决了状态值离散的诸多问题,并在多区域互联电网中验证了该算法的有效性。文献[18]基于DDQN和多激励演员-评论家(actor-critical, AC)策略提出了协同学习演员-评论家(collaborative learning actor-critic, CLAC)深度强化学习算法,通过引入NN解决了电力系统泛在物联网带来的强随机干扰问题,并且其经验共享机制提高了系统整体的控制性能。文献[19]将改进深度Q网络(double deep Q-network, DQN)应用于智能电压控制研究,以取代经典的多时间尺度无功功率优化和电压调节方案,相较于经典的比例-积分控制,所提算法降低电压偏差最高达到68.95%,并且具有持续交互能力,可对智能体参数进行持续优化。然而上述提及的无论强化学习亦或深度强化学习,智能体目标动作完全来源于预设的离散动作集合,依然存在动作空间维数灾难^[20]所导致的高维环境适应能力不强和调节精度不高等问题。虽然学者们针对动作“维数灾(提出的改进动作选择策略会提升算法在AGC系统中的控制性能,但依然无法处理连续控制^[21]问题,从根本上影响了算法更广泛的应用。

2017年,由NN近似离散动作集合的AC^[19]深度强化学习架构被提出,并开发了较为先进的近端策略优化^[22](proximal policy optimization, PPO)算法,其直接对策略网络梯度进行剪裁,有效防止了策略在每次更新中发生的剧烈变化,从而降低了策略梯度估计的方差,也让训练过程更加的平稳,并减少了计算复杂度。正是因为这些特点,在对稳定性要求高但计算资源相对不足的电力系统中,PPO算法的应用成为了可能。

因此,本文面向大规模新能源接入的新型电力系统的AGC频率控制问题,提出了一种自适应权重残差近端策略优化(adaptive weight residual proximal policy optimization, Ares-PPO)自动发电控制方法。所提方法为智能体策略网络各层之间新增了自适应权重残差连接(adaptive weight residual connection, AWR-Connection),通过引入可训练的权重参数可以动态调整残差连接的比例,从而防止NN中的梯度爆炸和梯度消失问题,进而可以在AGC系统中获得更优的频率控制性能。在IEEE标准两区域和三区域负荷频率控制(load frequency control, LFC)模型中进行仿真测试,与多种强化学习方法相比,所提方法具有更快的收敛性和更强的稳定性。

1 多区域互联电力系统频率控制模型

1.1 多区域互联频率控制模型

对于具有 i 个区域的多区域互联频率控制系统模型, 在频率调节过程中, 机组的调节输出功率 $\Delta P_{T,i}$ 和负荷功率增量 $\Delta P_{D,i}$ 的差值^[9,19]可由式(1)计算得出。

$$\Delta P_{T,i} - \Delta P_{D,i} = \frac{d}{dt} W_i + \Delta P'_{D,i} + \Delta P_{tie,i} \quad (1)$$

式中: dW/dt 为发电机发电功率增量; $\Delta P'_{D,i}$ 为负载频率调节功率; $\Delta P_{tie,i}$ 为第 i 个区域的联络线交换功率。通过联络线从区域 1 流向区域 2 的联络线功率可由如下公式得到。

$$L[\Delta P_{tie,12}] = \frac{2\pi}{l} T_{12} (L[\Delta f_1] - L[\Delta f_2]) \quad (2)$$

式中: $L[\cdot]$ 表示拉普拉斯变换; l 为拉普拉斯变换变量; Δf_1 和 Δf_2 分别为区域 1 和区域 2 的频率偏差; T_{12} 为区域联络线时间常数。

本文所采用的模型不具体设置调度中心, 而是在每个区域中单独设置一个智能体进行独立控制运行, 各自维持本区域的区域控制误差 (area control error, ACE) 指标稳定。那么所采用的 ACE 计算方式为:

$$C_{ace,i} = \Delta P_{tie,i} + B_i \Delta f_i \quad (3)$$

$$B_i = \frac{1}{R_i} + D_i \quad (4)$$

式中: $C_{ace,i}$ 为区域 i 的 ACE 值; B_i 为区域 i 的偏差系数; R_i 为区域 i 的一次调频系数; D_i 为区域 i 的阻尼系数。

2 Ares-PPO

2.1 PPO

在 PPO 中, 策略以目标函数期望的最大值进行策略参数更新, 具体如下。

$$\theta_{\zeta+1} = \arg \max_{\theta} E_{s,a \sim \pi_{\theta_{\zeta}}} [L(s,a,\theta_{\zeta},\theta)] \quad (5)$$

式中: s 为当前系统状态; a 为当前选择的动作; θ_{ζ} 为策略 ζ 的参数; θ 为当前策略的参数; $E(\cdot)$ 为期望函数; $L(\cdot)$ 为 PPO 的目标函数, 表示相邻两个策略之间的偏差, 具体表示如下:

$$L(s,a,\theta_{\zeta},\theta) = \min \left(\frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{\zeta}}(a_t|s_t)} A^{\pi_{\theta_{\zeta}}}(s_t, a_t), \right. \\ \left. clip \left(\frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{\zeta}}(a_t|s_t)}, 1 - \varepsilon, 1 + \varepsilon \right) A^{\pi_{\theta_{\zeta}}}(s_t, a_t) \right) \quad (6)$$

式中: $clip(\cdot)$ 为策略梯度截断函数, 目的是将策略

变化量限制在一个合理的范围内, 防止梯度爆炸; ε 为梯度截断参数; s_t 为智能体在 t 轮的状态; a_t 为在 t 轮选择的动作; π_{θ} 为当前策略函数, 表示在给定状态 s_t 下选择动作 a_t 的概率分布; $\pi_{\theta_{\zeta}}$ 为基于旧的策略网络参数 θ_{ζ} 的策略函数, 表示之前策略的概率分布; $A^{\pi_{\theta_{\zeta}}}(s_t, a_t)$ 为选择动作 a_t 时优势函数, 相对于某个基准的动作的价值。

2.2 自适应权重残差连接

PPO 的策略网络通常是由多层不同个数的神经元组合而成层级结构, 通常用随机梯度下降法以反向传播的方式对每一层神经元进行权值更新。当网络层数增加时, 梯度在反向传播过程中可能会逐渐减小或增大, 导致前几层的权重无法有效更新, 造成梯度消失或梯度爆炸, 导致智能体无法正常学习和运行, 进而造成频率失控等严重问题。

而计算机领域通常采用基本残差连接 (如图 1 所示) 来缓解上述两个问题。如式(7)所示, 假设第 i 层 NN 有一个输入 x_i 经过变换函数 $F_i(\cdot)$ 后输出为 y , 即基本残差连接可以表示为:

$$y = F_i(x_i) + x_i \quad (7)$$

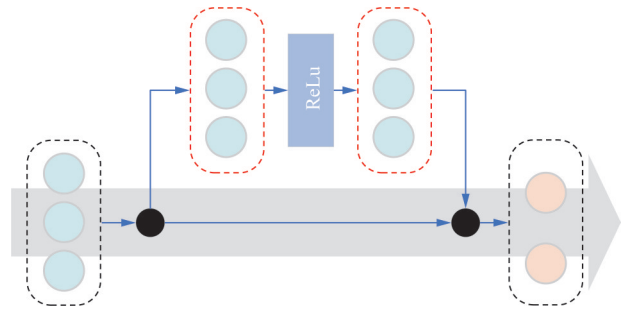


图 1 残差连接示意图

Fig. 1 Sketch diagram of Residual connection

输入 x_i 和变换后的输出 $F_i(x_i)$ 的权重是固定的, 在整个训练过程中无法调整各层之间的残差比例, 这种固定的连接方式在所有层中都是相同的, 进而导致模型的表达能力在某些情况下可能受到限制。因此, 每层 NN 考虑一个可训练的权重参数 α_i , 使得网络能够根据训练数据自动调整残差的贡献, 即式(7)可变换为如下形式。

$$y = \alpha_i F_i(x_i) + (1 - \alpha_i) x_i \quad (8)$$

在策略网络的反向传播假设我们有一个损失函数 L , 通过链式法则对输入 x 求导。

考虑一个 3 层残差 NN, 对 y 关于 x 的导数为:

$$\frac{\partial y}{\partial x} = \alpha \frac{\partial F(x)}{\partial x} + (1 - \alpha) \quad (9)$$

对 y 关于 $F(x)$ 的导数为:

$$\frac{\partial y}{\partial F(x)} = \alpha \quad (10)$$

根据链式法则梯度传播公式可写为:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial x} = \frac{\partial L}{\partial y} \cdot (\alpha \frac{\partial F(x)}{\partial x} + (1 - \alpha)) \quad (11)$$

假设有一个三层网络，每一层都采用AWR-Connection，输入为 x_0 ，经过第一层的变换 $F_1(x_0)$ ，由式(8)可知输出为：

$$x_1 = \alpha_1 F_1(x_0) + (1 - \alpha_1)x_0 \quad (12)$$

将第一层输出作为第二层输入，以此类推，第二层和第三层输出分别为：

$$x_2 = \alpha_2 F_2(x_1) + (1 - \alpha_2)x_1 \quad (13)$$

$$x_3 = \alpha_3 F_3(x_2) + (1 - \alpha_3)x_2 \quad (14)$$

在反向传播时，对输入 x_0 求导：

$$\frac{\partial L}{\partial x_0} = \frac{\partial L}{\partial x_3} \cdot \frac{\partial x_3}{\partial x_2} \cdot \frac{\partial x_2}{\partial x_1} \cdot \frac{\partial x_1}{\partial x_0} \quad (15)$$

其中:

$$\frac{\partial x_3}{\partial x_2} = \alpha_3 \frac{\partial F_3(x_2)}{\partial x_2} + (1 - \alpha_3) \quad (16)$$

$$\frac{\partial x_2}{\partial x_1} = \alpha_2 \frac{\partial F_2(x_1)}{\partial x_1} + (1 - \alpha_2) \quad (17)$$

$$\frac{\partial x_1}{\partial x_0} = \alpha_1 \frac{\partial F_1(x_0)}{\partial x_0} + (1 - \alpha_1) \quad (18)$$

将式(12)~(18)代入总的梯度式(9)中可得:

$$\begin{aligned} \frac{\partial L}{\partial x_0} = & \frac{\partial L}{\partial x_3} \cdot (\alpha_3 \frac{\partial F_3(x_2)}{\partial x_2} + (1 - \alpha_3)) \cdot (\alpha_2 \frac{\partial F_2(x_1)}{\partial x_1} + \\ & (1 - \alpha_2)) \cdot (\alpha_1 \frac{\partial F_1(x_0)}{\partial x_0} + (1 - \alpha_1)) \end{aligned} \quad (19)$$

由式(19)可知, AWR-Connection中的 $1-\alpha_i$ 项保证了即使 $\alpha_i \times \partial F_i(x_i)/\partial x_i$ 非常小, 仍有一部分梯度 $1-\alpha_i$ 直接传递, 这减少了梯度消失的风险。通过 α_i 控制 $F_i(x_i)$ 的梯度贡献, 使得 $F_i(x_i)$ 的梯度贡献受到限制, 从而减少了梯度爆炸的风险。

初始训练阶段，网络可能需要更多依赖于直接输入 x ，此时 α 值较小，使得 $1-\alpha$ 较大，从而更多地保留原始输入信息。随着训练的进行，网络逐渐学习到有效的特征，此时 α 值可以逐渐增大，使得 $F_i(x_i)$ 的贡献增加，网络能够更充分地利用变换后的特征。在训练的收敛阶段，网络可能需要对 α_i 进行微调，以找到最优的残差比例，使得最终的特征

表示最为有效。

2.3 Ares-PPO

Ares-PPO(如图2所示)是基于AC架构改进得到的无模型在线策略深度强化学习算法,图中 S_t 为 S_i 下一状态。无模型强化学习体系,能够解决部分可观测的复杂非线性电力系统建模难题。基于AC架构的Ares-PPO能够解决强化学习维数灾难问题,无需复杂繁琐的划分状态-动作矩阵;而AWR-Connection策略网络提升模型训练的速度和稳定性,能在AGC系统中获得更稳定快速的频率调节能力。Ares-PPO的策略网络可简单表示如下。

$$X_1 = \text{Activate}(\text{Linear}(X)) \quad (20)$$

$$X_2 = \text{Activate}(\text{Linear}(X_1)) \quad (21)$$

$$X_3 = \alpha \times X + (1 - \alpha) \times X, \quad (22)$$

$$X_{\text{mean}} = \tanh(\text{meanLayer}(X_3)) \quad (23)$$

$$a = \text{Sample}(\text{Normal}(X_{\text{mean}})) \quad (24)$$

式中： $Linear(\cdot)$ 为一个线性变换层，它通过对输入数据进行加权求和并加上偏置，得到网络的中间输出； $Activate(\cdot)$ 为激活函数； $meanLayer(\cdot)$ 为自定义层，用来计算输入数据的均值； $Normal(\cdot)$ 为将数据处理为正态分布； $sample(\cdot)$ 为对分布进行采样得到一个动作； X 为系统输入； X_1 、 X_2 、 X_3 和 X_{mean} 分别为策略网络输入层、隐藏层、自适应残差连接和输出层输出； a 为通过输出动作的概率分布采样得到的动作值。

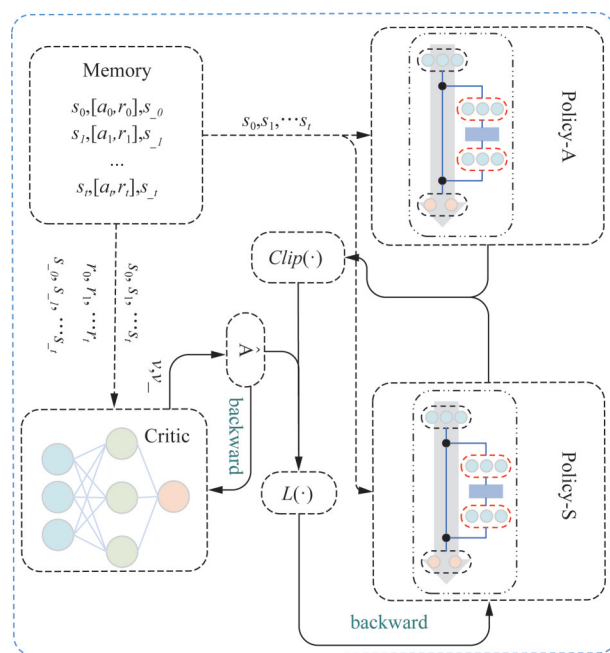


图2 Ares-PPO框架

Fig. 2 Ares-PPO framework

3 基于 Ares-PPO 算法的 AGC 系统设计

基于 Ares-PPO 算法的 AGC 控制系统设计主要包括了控制性能标准^[26] (control performance standards, CPS)、状态设置、奖励函数、参数设置和 Ares-PPO 的 AGC 控制流程。

3.1 控制性能标准

CPS 是北美电力可靠性委员会提出用于评价 AGC 系统供电质量的国际标准, 注重中长期 AGC 性能指标, 具体由 CPS1 指标和 CPS2 指标构成, 国内一般采用 10 min 考核周期, 详细 CPS 考核标准如下。

1) 若 CPS1 值 $\geq 200\%$, 且 CPS2 值为任意值, CPS 指标合格。

2) 若 $100\% \leq \text{CPS1 值} < 200\%$, 且 CPS2 值 $\geq 90\%$, CPS 指标合格。

3) 若 CPS1 值 $< 100\%$, 则 CPS 指标不合格。

3.2 状态设置

Ares-PPO 合理的状态-动作设计对 AGC 的性能起着至关重要的作用。由于 Ares-PPO 不需要再对状态和动作进行划分, 所以在 AGC 系统的一个区域中, 我们将 CPS 中的 CPS1 指标 S_{CPS1} 、区域控制误差 S_{ACE} 和区域频率偏差 $S_{\Delta f}$ 的组合向量作为智能体状态输入, 具体如下。

$$S = [S_{\text{CPS1}}, S_{\text{ACE}}, S_{\Delta f}] \quad (25)$$

在区域 i 中, 每一轮控制周期的功率差值 ΔP_i 设置为智能体的动作值, AGC 控制周期设置为 4 s, 二次调频时延为 20 s。

3.3 奖励函数

奖励函数作为强化学习的重要组成部分, 直接影响着 Ares-PPO 的控制性能。本文主要追求 AGC 系统对于 CPS 指标的长期效益, 除此之外, S_{ACE} 和 $S_{\Delta f}$ 也是衡量电力系统稳定的重要指标, 所以在本文中综合如上 3 个指标作为系统的奖励函数。为了数据尺度一致性, 本文对 CPS1 指标 S_{CPS1} 映射到 0 附近为 $(S_{\text{CPS1}} - 200)$, 且对整个三元函数取负数, 奖励函数 R_t 具体如下。

$$R_t = -\left\{ [S_{\text{ACE}}(t)]^2 + [S_{\text{CPS1}}(t) - 200]^2 + [S_{\Delta f}(t)]^2 \right\} \quad (26)$$

式中: $S_{\text{ACE}}(t)$ 为第 t 次迭代 ACE 的数值; $S_{\text{CPS1}}(t)$ 为第 t 次迭代 CPS1 值; $S_{\Delta f}(t)$ 为第 t 次迭代 Δf 值。

3.4 参数设置

对多区域互联电力系统控制器设置合理的参数是保证控制性能的前提, 参数具体设置如表 1 所示。

表 1 控制器参数对照表

Tab. 1 Controller parameters comparison table

参数	设置值
截断参数 ε	0.2
学习率 α	0.0001 (Ares-PPO/Res-PPO/PPO) 0.01 (DQN/Ares-DQN)
记忆池 D	128
抽样容量 D_M	32
折扣因子 γ	0.99

1) 截断参数 ε 限制了新旧策略更新的尺度, 在第一批更新过程中, 新旧策略的更新幅度被限制在了 $[1-\varepsilon, 1+\varepsilon]$ 之内, 防止因梯度爆炸导致算法不收敛的问题。在文献[21]中建议 ε 的取值范围为 0.1~0.3, 无具体设置办法, 本文在实验的基础上取 $\varepsilon = 0.2$ 。

2) 学习率 α 决定了智能体学习的步长, 本质上决定了智能体的更新幅度。 α 过大会导致智能体更新幅度过大, 有可能错过权值的最优值; α 过小可能会导致智能体收敛于局部最优点, 降低模型的有效容量。Ares-PPO 及其他对比算法具体的学习率如表 1 所示。

3) 记忆池 D 是存放历史信息的缓冲池, 合适的记忆池容量可以打乱学习顺序, 避免顺序学习面临的联系影响风险, 最终防止了 NN 的连续震动和发散, 还影响着算法的收敛速度。综合考虑, 本文具体取记忆池容量 $D=128$ 。

4) 抽样容量 D_M 是从记忆池 D 中抽取的用于训练 NN 的样本数量。 D_M 打乱了训练的样本顺序, 防止 NN 出现振荡和崩溃。综合考虑, 本文具体取 $D_M=32$ 。

5) 折扣因子 γ ($0 < \gamma < 1$), 权衡立即奖励和后期奖励之间的重要性, 考虑 γ 取值应接近 1。

3.5 基于 Ares-PPO 的 AGC 流程

基于 Ares-PPO 的 AGC 的主要逻辑和流程如表 2 所示。表中描述了一个基于 Ares-PPO 算法流程, 通过使用动态 S_{CPS1} 、 S_{ACE} 和 $S_{\Delta f}$ 数据作为输入, 以输出总功率指令 ΔP_{Σ} 。算法的主要步骤包括选择动

作、计算奖励值和总功率指令、存储数据到记忆池、计算累积奖励值、计算优势函数、更新策略网络参数以及复制参数至目标策略网络,以不断优化控制过程并最终获得所需的总功率指令。

表2 Ares-PPO的AGC控制流程
Tab. 2 AGC control process of Ares-PPO

输入:动态 S_{CPS1} 、 S_{ACE} 和 $S_{\Delta f}$ 数据
输出:总功率指令 ΔP_{Σ}
初始化:神经网络Critic、Policy-A和Policy-S;经验池;初始状态 s_0 ; 设置参数 ε 、 α 、 D 、 D_M 、 γ 。
Repeat:
1:依据Policy-A执行动作 a ,根据式(21)得到下一状态 s_{-} ,并依据式(22)计算奖励值 r ,依据动作 a 输出总功率指令 ΔP_{Σ} ;
2:将包含了当前状态 s 、动作 a 、奖励 r 和下一状态 s_{-} 的四维数组 (s, a, r, s_{-}) 存入记忆池;
3:如果记忆池连续存入 D 条数据则运行步骤4,否则返回步骤1;
4:在记忆池中随机采样 D_M 条数据;
5:用折扣因子 γ 计算每个样本的累积奖励值 R ;
6:用Critic网络的估值和累积奖励 R ,计算优势函数 $\hat{A}$;
7:根据当前状态 s 、动作 a 和优势函数 $\hat{A}$,计算得到动作网络损失loss,并反向传播loss;
8:将Policy-S的参数复制给Policy-A,完成更新;
9:返回步骤1。
End

4 算例分析

选取离散动作集合的深度强化学习DQN和Res-DQN,连续动作的PPO和Res-PPO作为对比算法与Ares-PPO进行对比分析。搭建了两区域LFC模型和基于西南电网的三区域LFC模型。

为了凸显各个算法在不同状态-动作集上的表现,本文将DQN和Res-DQN的动作区间设置为 $[-30, 30]$ MW,步长为2 MW;Ares-PPO、Res-PPO和PPO的动作范围为连续区间 $[-30, 30]$ MW。为保证实验结果的真实性和严谨的实验过程,5种算法中涉及到的相同参数和NN都保持一致的情况下进行仿真实验。

对于仿真模型,标准频率为50 Hz,频率偏差控制目标设置为0.2 Hz,ACE的合格目标设置为正负2 MW。预学习阶段采用周期性负荷扰动和特定阶跃负荷扰动,在多区域仿真中着重采用非周期随机负荷扰动来评估控制器的性能。

4.1 IEEE标准两区域LFC模型

IEEE标准两区域LFC模型是电力系统中经典的频率控制模型之一,用于研究和设计电力系统的频率调节和功率平衡。如图3所示,该模型分为两个区域,每个区域都有特定且不必相同的发电机和负荷需求,同时两区域之间通过联络线进行电力交换。

图3中, T_g 为火电机调速器时延常数; T_t 为火电机时间常数; T_p 为频响函数时间常数; K_p 为频响函数系数; P_{tie} 为联络线交换功率; P_f 为火电机输出功率; P_L 为负荷功率; T_{12} 为联络线时间常数,详细参数如图3所示。

4.1.1 预学习

强化学习需要足够的环境信息逐渐调整自身参数,达到优化控制能力的目的。在正式运行之前需

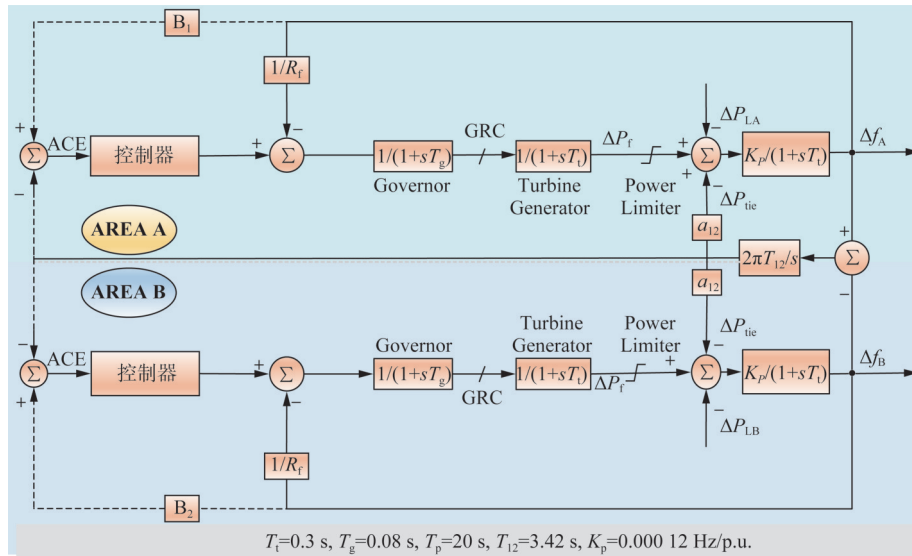


图3 IEEE两区域LFC模型

Fig. 3 IEEE two-area LFC model

要建立电网等效数学模型,让智能体能获取足够多的数字信息训练数据,以投入到实际电网中。

因此,在预学习阶段选用峰值为 $[-50, 50]$ MW,周期为 $1\ 200\text{ s}$ 的正弦负荷作为预学习蓝本,预学习负荷跟踪曲线如图4所示。

图4的负荷扰动跟踪曲线展示了Ares-PPO、Res-PPO、PPO、Ares-DQN和DQN共5种不同算法在两个区域下的负荷响应情况。Ares-PPO和PPO显示出更好的稳定性,负荷波动较小且较为平稳,波动幅度在 $\pm 60\text{ MW}$ 之间。Res-PPO和Ares-DQN在初期有剧烈波动,波动幅度分别在 $\pm 300\text{ MW}$ 和 $\pm 400\text{ MW}$ 之间,但随后逐渐趋于稳定。相比之下,DQN初期波动非常剧烈,尤其是Area B,波动幅度最大可达 $\pm 1\ 800\text{ MW}$,并且在后期会出现负荷扰动跟踪失效,表现最不稳定。综上,Ares-PPO在收敛性、稳定性和波动幅度方面表现最佳,而DQN存在较大波动风险,Res-PPO和Ares-DQN虽然初

期波动较大,但最终能趋于稳定。

4.1.2 收敛性分析

依据式(22),智能体在学习过程中所能得到的最大奖励为0,即奖励越接近0说明智能体的学习完备性越高。图5比较了Ares-PPO、Res-PPO、PPO、Ares-DQN和DQN这5种算法在Area A和Area B中的表现。Ares-PPO在0至 $5\ 000\text{ s}$ 期间,奖励值有较大波动,最低达到 $-1\ 500$ 左右。但从 $5\ 000\text{ s}$ 开始,奖励值趋于稳定,波动减小,平均奖励值约为 -500 ,波动范围在 ± 500 内。Res-PPO初期奖励值波动较大,最低达到 $-1\ 500$ 左右。在 $5\ 000\sim 15\ 000\text{ s}$ 期间,奖励值逐渐稳定,但仍有一定波动,最低奖励值约为 $-1\ 000$,平均奖励值约为 -500 。PPO奖励值波动频繁且幅度较大,最低奖励值达到 $-2\ 000$ 。尽管在 $15\ 000\text{ s}$ 后波动有所减小,但奖励值仍然不稳定,平均奖励值约为 $-1\ 000$ 。Ares-DQN在初期奖励值波动剧烈,最低达到 $-2\ 000$ 左右,但

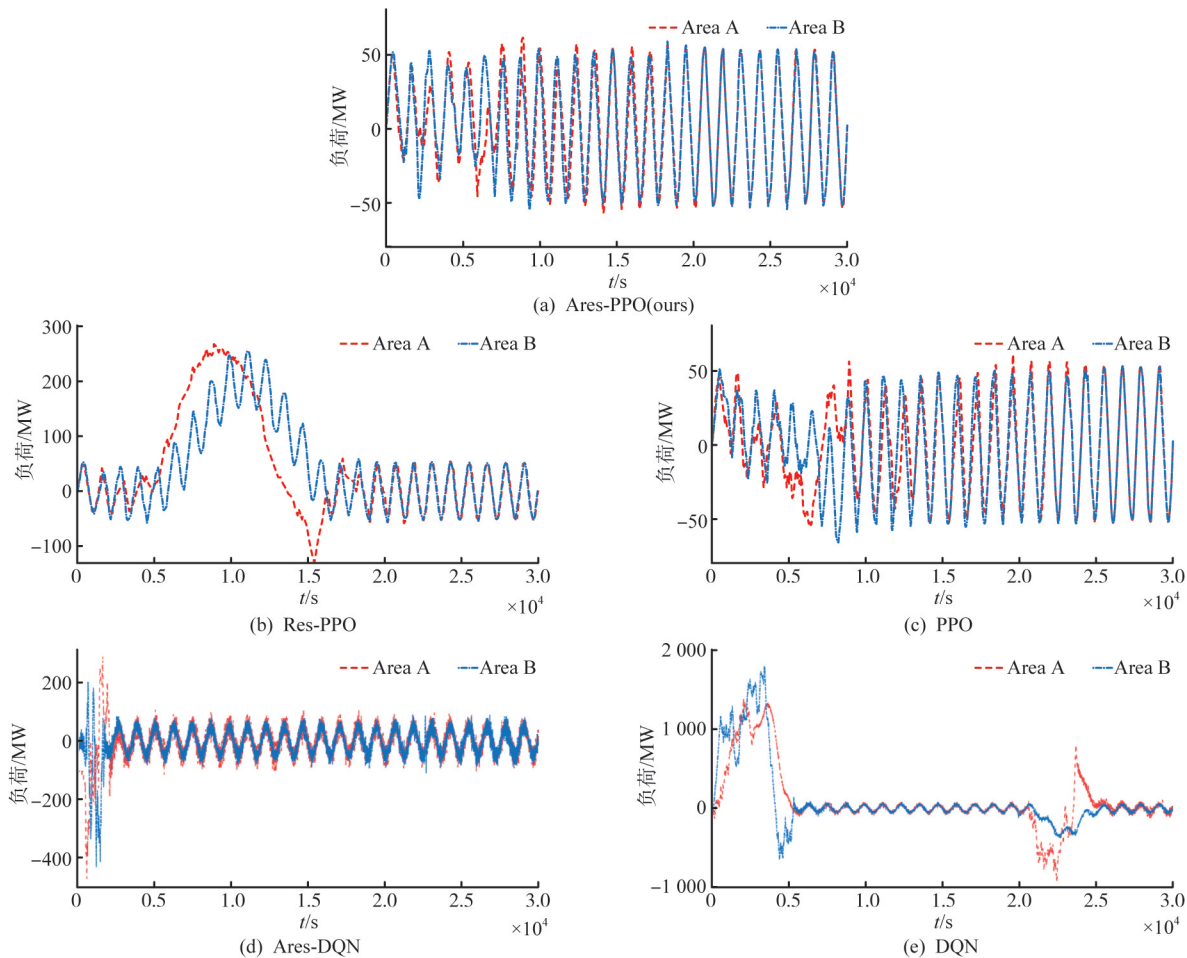


图4 预学习负荷跟踪曲线

Fig. 4 Pre-learning load tracking curves

在1 000 s后逐渐趋于稳定。平均奖励值约为-500,波动范围在 $\pm 1\ 000$ 内。DQN的奖励值波动最大,初期频繁出现奖励值为-2 000的情况。整个过程波动幅度较大且频率较高,平均奖励值约为-1 000,波动范围广泛。

综上所述,Ares-PPO算法的表现最佳,尽管初期有较大波动,但很快趋于稳定,且波动范围较小。Res-PPO算法次之,虽然波动略大,但总体稳定性较好。PPO和Ares-DQN算法的波动较为剧烈,稳定性较差。DQN算法的奖励值波动最大,表现最不稳定。

因此,可总结出如表3中所反应各对比算法的性能差异。

4.1.3 阶跃负荷扰动

在充分预学习后,为了对比5种算法在负荷明显突增的情况下的性能,我们利用幅值为1 000 MW的阶跃负荷扰动来验证控制算法的快速性、稳定性、准确性。图6为连续运行100幕后选取的最优负荷控制曲线及其指标。

图6(a)展示了不同算法在负荷分配中的动态响应情况。图中包括Ares-PPO、Res-PPO、PPO、Ares-DQN和DQN这5种算法的负荷变化曲线。从

表3 预学习阶段性能对比

Tab. 3 Performance comparison of the pre-learning stage

算法	区域	负荷波动幅度/MW	负荷稳定性	收敛时间/s	奖励稳定性	平均奖励值	奖励波动范围
Ares-PPO	Area A/B	± 60	最佳,平稳	5 000	较高	-500	± 500
Res-PPO	Area A/B	± 300	稳定性较好	24 000	中等	-500	± 500
PPO	Area A/B	± 60	稳定性较差	25 000	中等	-1000	$\pm 1\ 000$
Ares-DQN	Area A/B	± 400	初期波动大,后期趋稳	3 000	较高	-500	$\pm 1\ 000$
DQN	Area A/B	$\pm 1\ 800$	最差,不稳定	30 000+	最差	-1 000	$\pm 2\ 000$

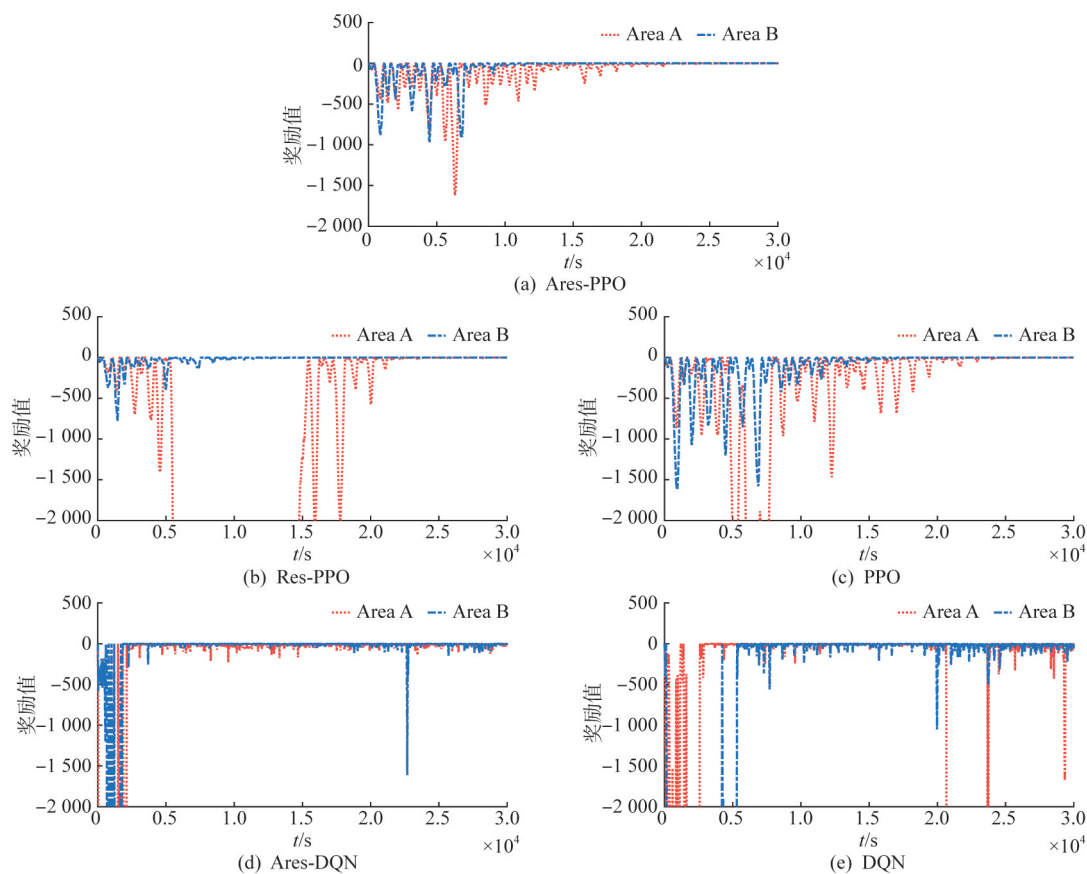


图5 预学习奖励函数曲线

Fig. 5 Pre-learning reward function curves

结果来看, Ares-PPO 算法表现最佳, 其负荷稳定在约 1 000 MW, 并且波动最小。具体来看, Ares-PPO 在约 2 000 s 时达到稳定状态, 负荷在 1 000 MW 上下波动约 10 MW。相比之下, Res-PPO 和 PPO 算法在约 2 000 s 后也能稳定在 1 000 MW 左右, 但波动略大, 约在 20 MW 左右。Ares-DQN 算法在达到稳定状态后负荷波动较大, 约在 1 000 MW 上下 100 MW 范围内。DQN 算法则表现最差, 其负荷在达到稳定状态后波动最为剧烈, 尤其是在 2 000~4 000 s 之间, 负荷显著下降, 最低降至约 600 MW, 随后逐渐回升, 但波动仍然明显。

图 6(b) 展示了不同算法在负荷扰动下的频率波动情况, 对比了 DQN、Ares-DQN、PPO、Res-PPO 和 Ares-PPO 这 5 种算法在 Area A 和 Area B 中的表现。从结果来看, Ares-PPO 算法在负荷扰动后的频率恢复最快且波动最小。在 Area A 中, Ares-PPO 算法的频率在扰动后的最低点约为 49.93 Hz, 随后迅速恢复至 50 Hz, 并保持稳定, 波动在 ± 0.01 Hz 内。PPO 和 Res-PPO 算法在扰动后频率最低点均在 49.92 Hz 左右, 恢复至 50 Hz 后波动在 ± 0.02 Hz 内。Ares-DQN 算法表现稍差, 频率最低点为 49.90 Hz, 恢复较慢且波动范围在 ± 0.03 Hz 内。DQN 算法表现最差, 频率最低点为 49.88 Hz, 恢复缓慢且波动范围较大。在 Area B 中, Ares-PPO 算法的频率最低点约为 49.91 Hz, 恢复至 50 Hz 后波动较小, 保持在 ± 0.015 Hz 内。PPO 和 Res-PPO 算法在频率最低点为 49.90 Hz 左右, 恢复至 50 Hz 后波动在 ± 0.02 Hz 内。Ares-DQN 算法表现出明显的波动, 频率最低点为 49.87 Hz, 恢复缓慢且波动范围在 ± 0.035 Hz 内。DQN 算法最低点降至 49.85 Hz, 恢复最慢, 波动范围最大。

图 6(c) 展示了不同算法在负荷扰动下的 10 min ACE 均值 (IACEI) 变化情况。在 Area A 中, Ares-PPO 和 Res-PPO 算法的 ACE 表现依旧最好, 最低点约为 -60 MW, 并能迅速恢复到 0 并保持稳定。PPO 算法的 IACEI 在扰动后最低达到约 -50 MW, 恢复速度较快并能稳定在 0。Ares-DQN 算法的 IACEI 最低值为约 -55 MW, 恢复速度较慢但最终也能恢复到 0。DQN 算法表现最差, IACEI 最低值约为 -60 MW, 恢复最慢且波动较大。

图 6(d) 展示了不同算法在负荷扰动下的 10 min CPS1 均值 (ICPS1I) 变化情况。在 Area A 中,

Ares-PPO 和 Res-PPO 算法的 ICPS1I 在负荷扰动后的最低点约为 185, 在扰动发生后迅速恢复到 200 并保持稳定。PPO 算法的 ICPS1I 在扰动后最低降至约 190, 恢复速度稍慢但也最终恢复至 200。Ares-DQN 算法的 ICPS1I 最低点约为 187, 恢复较慢但最终也能恢复到 200。DQN 算法表现最差, ICPS1I 在扰动后最低降至约 185, 恢复最慢且波动较大。在 Area B 中, Ares-PPO 和 Res-PPO 算法的 ICPS1I 表现依旧最好, 最低点约为 185, 并能迅速恢复到 200 并保持稳定。PPO 算法的 ICPS1I 在扰动后最低达到约 190, 恢复速度较快并能稳定在 200。Ares-DQN 算法的 ICPS1I 最低值为 187, 恢复速度较慢但最终也能恢复到 200。DQN 算法表现最差, ICPS1I 最低值约为 185, 恢复最慢且波动较大。

4.2 西南电网三区域 LFC 模型

为进一步分析 Ares-PPO 在更为复杂工况下控制性能, 以西南电网为基础, 搭建了具有风电、光伏等可再生能源并网的改进三区域 LFC 模型 (如图 7 所示)。需要说明的是, 本文的目的是验证所设计的算法在负荷复杂变化情况下的控制性能, 因此将新能源并网看作负荷, 并未对其内部进行建模。同样分别采用 Ares-PPO、Res-PPO、PPO、Ares-DQN 和 DQN 共 5 种算法进行对比分析, 一幕的仿真时间为 10 000 s, 总共仿真 50 幕。

4.2.1 随机负荷扰动

图 8 为三区域 LFC 模型中的负荷跟踪曲线。由图 8 中可知, Ares-PPO 在各个区域的负荷曲线与实际负荷曲线几乎重合, 显示出极高的跟踪能力。在 5 000 s 和 6 000 s 的负荷剧烈变化中, Ares-PPO 能迅速回到稳定状态, 恢复时间短, 波动幅度较小, 且跟踪误差最小, 显示出优异的跟踪精度。Res-PPO 总体表现良好, 但在负荷变化剧烈时存在明显波动, 恢复时间较 Ares-PPO 长, 波动较大, 跟踪误差比 Ares-PPO 略大。PPO 表现中等, 特别是蓝色曲线在 3 000~8 000 s 处未能完全跟踪实际负荷, 存在明显偏差, 恢复速度较慢, 波动幅度较大, 跟踪误差较大。Ares-DQN 总体表现良好, 但在 5 000 s 和 7 000 s 的负荷变化中存在一定偏离, 恢复时间较长, 波动幅度较小。但跟踪误差较大, 精度较低。DQN 表现最差, 多个时间段出现较大波动, 曲线偏离实际负荷显著, 恢复时间最长, 波动幅度最大, 跟踪误差最大。

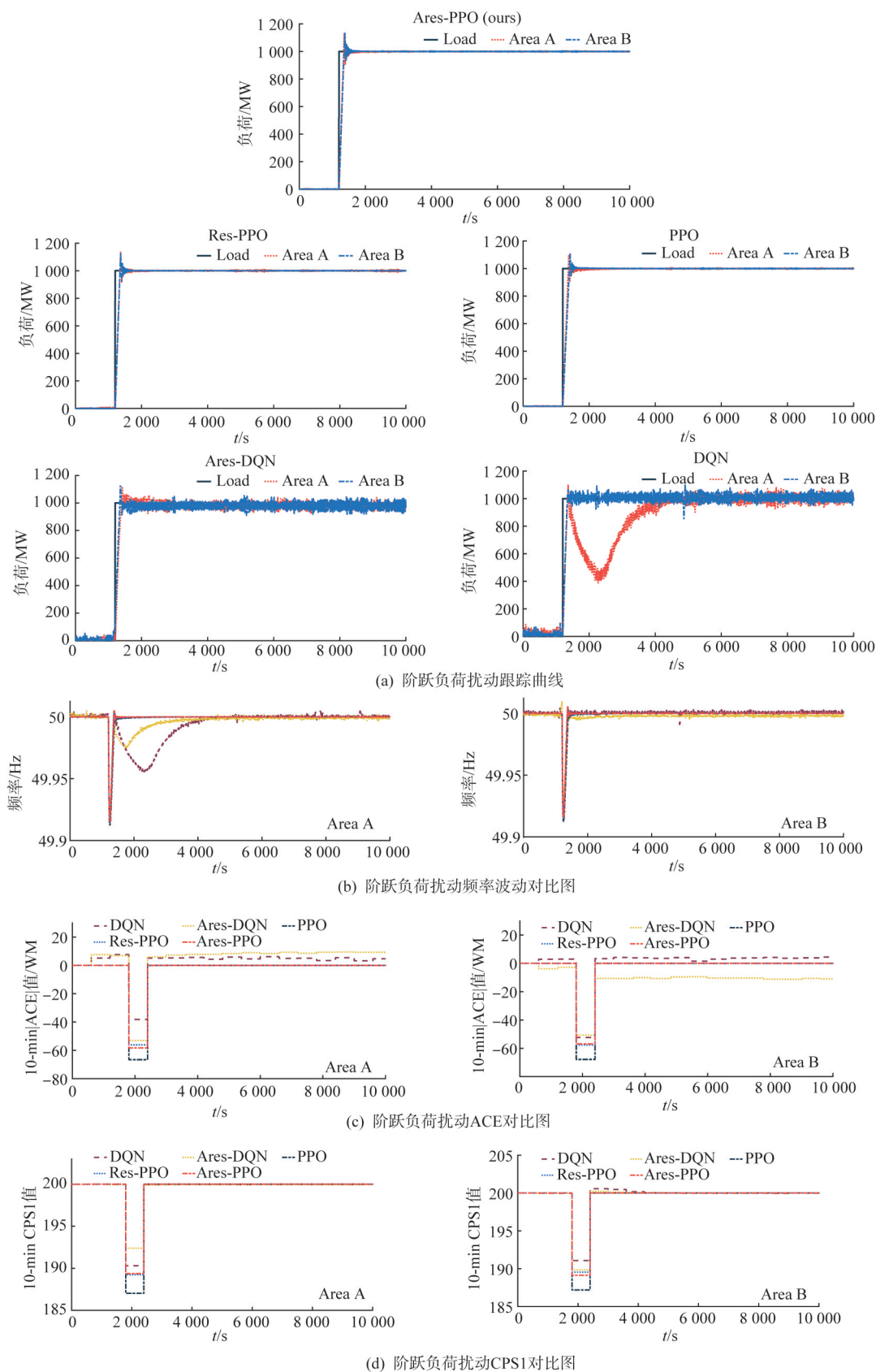


图6 阶跃负荷下AGC性能指标

Fig. 6 AGC performance indexes under step load

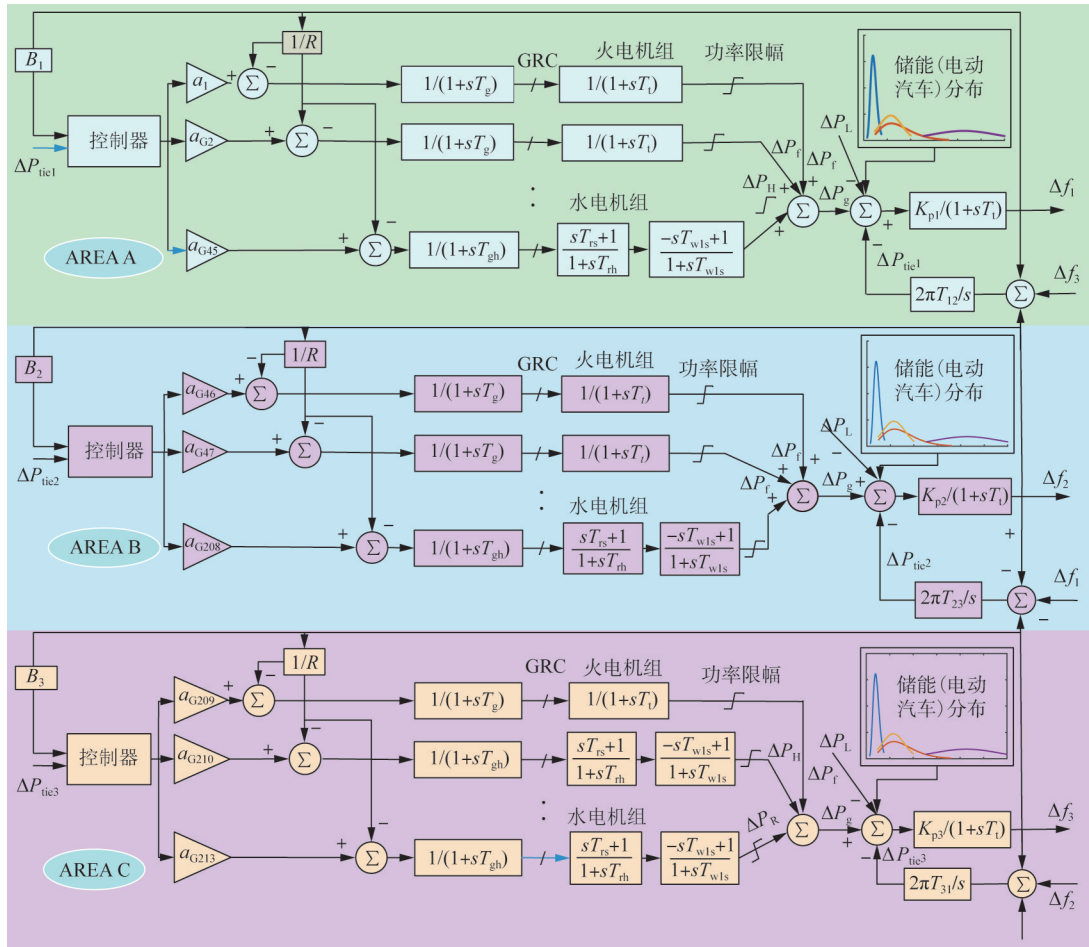


图7 改进的三区域LFC模型

Fig. 7 Improved three-area LFC model

在量化数据方面,通过表4分析可见,Ares-PPO算法在负荷跟踪效果上普遍表现最好。例如在Area A,Ares-PPO相比最差的Ares-DQN,绝对误差积分(integral of absolute error, IAE)减少了约74.1%,平方误差积分(integral of square error, ISE)减少了约90.1%,瞬时误差标准差(standard deviation of instantaneous error, SDIE)降低了约66.7%,瞬时误差方差(variance of instantaneous error, VIE)减少了约88.9%。在Area B,Ares-PPO相对于Ares-DQN,IAE减少了约81.0%,ISE值减少了约92.3%,SDIE值降低了约72.2%,VIE值减少了约92.3%。在Area C,Ares-PPO相比Ares-DQN,IAE减少了约92.7%,ISE值减少了约87.2%,SDIE值降低了约88.4%,VIE值减少了约98.7%。这些数据表明,Ares-PPO在各个区域中均显著提升了负荷跟踪的精度和稳定性。

表4展示了不同算法在随机负荷扰动下的三区

域LFC性能指标。Ares-PPO在所有区域中的表现均为最佳。以Area A为例,Ares-PPO的频率偏差 $|\Delta f|$ 为0.000 6 Hz,比PPO的0.002 2 Hz降低了72.7%,比Res-PPO的0.000 9 Hz降低了33.3%,比Ares-DQN的0.003 0 Hz降低了80%,比DQN的0.013 3 Hz降低了95.5%。在IAEI方面,Ares-PPO的值为2.094 4 MW,比PPO的8.781 2 MW降低了76.1%,比Res-PPO的2.716 6 MW降低了22.9%,比Ares-DQN的5.215 5 MW降低了59.8%,比DQN的66.366 5 MW降低了96.8%。在CPS2方面,Ares-PPO的值为98.83%,比PPO的92.54%提高了6.8%,比Res-PPO的97.35%提高了1.5%,比Ares-DQN的99.25%略低0.4%,但远高于DQN的69.88%。综合来看,Ares-PPO在控制精度和稳定性方面显著优于其他算法,特别是在频率偏差和IAEI的表现上尤为突出。

通过以上3个仿真实验可以证明,在连续系统

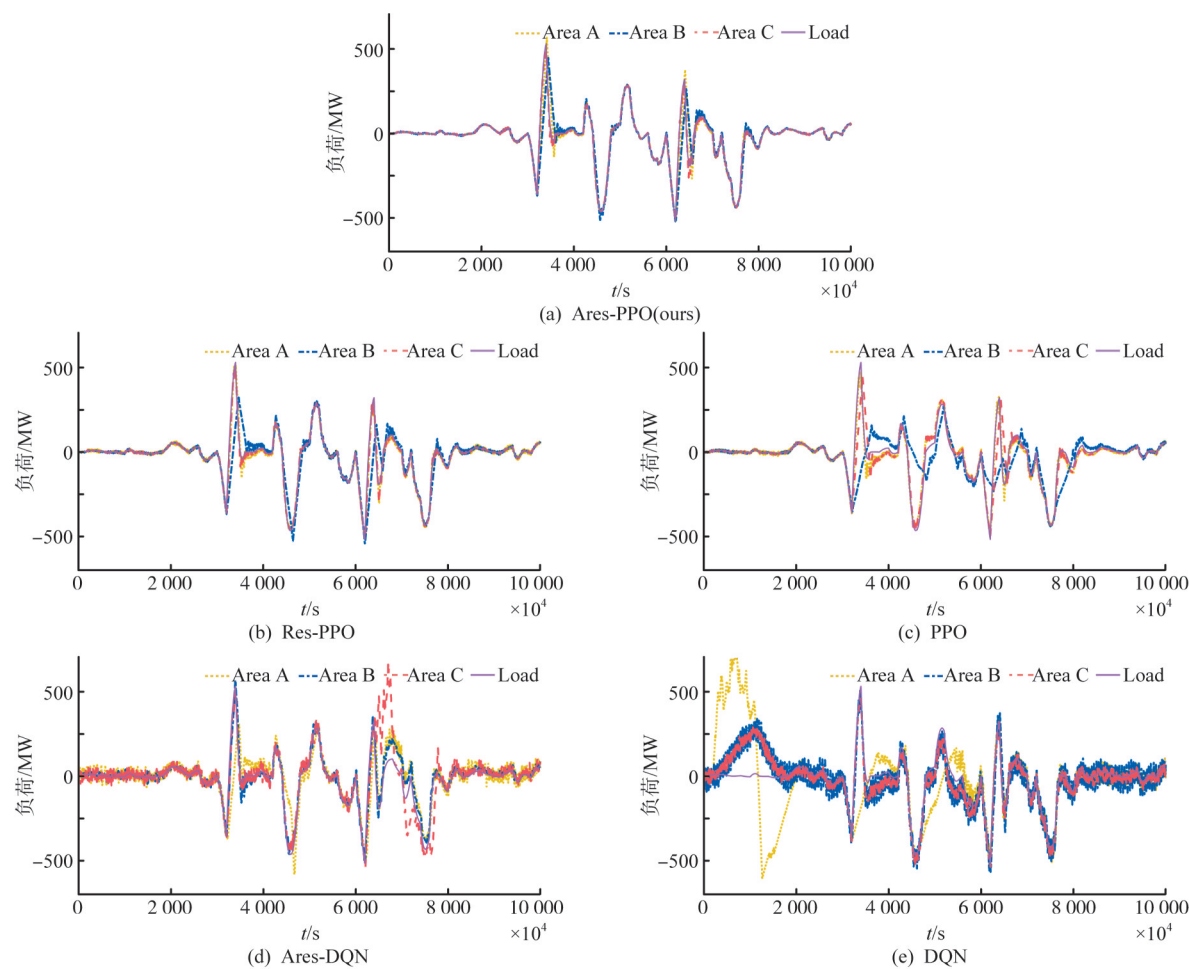


图8 随机负荷跟踪曲线
Fig. 8 Random load tracking curves

表 4 负荷跟踪效果评价指标
Tab. 4 Evaluation indexes of load tracking effect

区域	算法	绝对误差积分(IAE值)	平方误差的积分(ISE值)	瞬时误差标准差(STD值)	瞬时误差方差(VAR值)
Area A	Ares-PPO	83 645. 106 0	2 816 713. 901	16. 633 605	276. 676 802
	Res-PPO	112 507. 515 0	4 244 174. 586	20. 236 283	409. 507 131
	PPO	164 667. 471 0	6 203 761. 855	24. 581 626	604. 256 334
	Ares-DQN	322 828. 034 6	28 571 033. 200	50. 003 962	2 500. 396 231
	DQN	256 847. 525 3	10 229 997. 510	31. 057 692	964. 580 262
Area B	Ares-PPO	97 514. 581 47	5 917 490. 179	24. 206 540	585. 956 564
	Res-PPO	167 097. 577 9	19 619 746. 260	44. 207 988	1 954. 346 182
	PPO	209 177. 651 6	38 900 215. 430	61. 775 972	3 816. 270 701
	Ares-DQN	514 125. 677 0	77 273 334. 400	86. 955 595	7 561. 275 532
	DQN	354 942. 009 9	26 736 855. 200	42. 957 202	1 845. 321 227
Area C	Ares-PPO	143 067. 534 7	17 660 112. 680	41. 973 136	1 761. 744 126
	Res-PPO	314 956. 835 1	44 104 085. 060	66. 182 274	4 380. 093 429
	PPO	1 472 261. 317 0	340 982 366. 100	118. 592 401	14 064. 157 590
	Ares-DQN	1 953 761. 460 0	1 378 313 027. 000	362. 257 907	131 230. 791 200
	DQN	1 181 186. 351 0	425 675 057. 500	206. 307 581	42 562. 817 870

表 5 随机负荷扰动下的三区域 LFC 性能

Tab. 5 Three-area LFC performance under random load disturbance

区域	算法	$\Delta f/\text{Hz}$	$\Delta ACE/\text{MW}$	CPS1/%	CPS2/%	CPS3/%
Area A	Ares-PPO	0.000 6	2.094 4	199.970 7	98.83	100
	Res-PPO	0.000 9	2.716 6	199.955 5	97.35	100
	PPO	0.002 2	8.781 2	199.407 8	92.54	100
	Ares-DQN	0.003 0	5.215 5	199.950 7	99.25	100
	DQN	0.013 3	66.366 5	191.377 2	69.88	91.34
Area B	Ares-PPO	0.002 3	8.308 5	199.523 8	93.45	100
	Res-PPO	0.003 3	12.560 8	199.078 6	90.05	100
	PPO	0.003 6	14.140 5	198.910 4	87.90	100
	Ares-DQN	0.008 2	36.516 8	195.133 2	86.12	93.99
	DQN	0.005 6	13.120 6	199.869 9	97.03	100
Area C	Ares-PPO	0.000 4	0.839 7	199.996 8	100	100
	Res-PPO	0.000 7	1.947 6	199.970 8	98.44	100
	PPO	0.002 3	9.201 4	199.390 5	91.86	100
	Ares-DQN	0.002 8	3.317 9	199.975 9	99.53	100
	DQN	0.005 0	2.332 4	199.994 6	100	100

中,强化学习算法的状态-动作表达形式对控制效果具有显著的影响力。此外,从有规律的预学习阶段,到应对阶跃负荷变化,再到三区域 LFC 系统中的随机负荷场景,Ares-PPO 在各类复杂环境下均展现出卓越的性能和适应性。这种跨场景的鲁棒表现,表明所提算法能够灵活调整策略以应对动态变化,充分验证了其在不同应用场景下的可靠性与有效性。

5 结论

针对可再生能源并网引起的电力系统频率随机波动问题,本文提出了能够缓解策略网络梯度爆炸或梯度消失的多智能体协同算法 Ares-PPO。Ares-PPO 通过 AWR-Connection 策略网络,从理论上规避了策略网络梯度急剧变化而带来的频率失稳问题。通过在两区域和三区域负荷频率控制系统进行仿真,验证了所提算法能够实现了对多区域电网稳定的频率控制,且相较于 DQN、PPO 及其变体,所提算法收敛速度更快,控制精度更高并且具有更优的控制性能。

Ares-PPO 可扩展至光伏和风电占比高的跨区域电网频率控制场景。然而,其实际应用面临高维状态空间在线学习的计算复杂性,以及在硬件系统中

实现实时控制的算力限制。

参考文献

- [1] ALOMOUSH M I. Load frequency control and automatic generation control using fractional-order controllers [J]. Electrical Engineering, 2010, 91(7): 357–368.
- [2] 席磊,刘治洪,李彦营.基于拉格朗日松弛强化学习算法的自动发电控制[J].中国电机工程学报,2023,43(4):1359–1369.
XI Lei, LIU Zhihong, LI Yanying. Automatic generation control based on lagrangian relaxation reinforcement learning algorithm [J]. Proceedings of the CSEE, 2023, 43(4): 1359–1369.
- [3] LI Jiawen, YU Tao, ZHANG Xiaoshun. Coordinated automatic generation control of interconnected power system with imitation guided exploration multi-agent deep reinforcement learning [J]. International Journal of Electrical Power & Energy Systems, 2022 (136): 1–22.
- [4] XIAO Hao, PEI Wei, WU Lei, et al. A novel deep learning based probabilistic power flow method for multi-microgrids distribution system with incomplete network information [J]. Applied Energy, 2023(335): 1–21.
- [5] XIAO Hao, PU Xiaowei, PEI Wei, et al. A novel energy management method for networked multi-energy microgrids based on improved DQN [J]. IEEE Transactions on Smart Grid, 2023, 14(6): 4912–4926.
- [6] LI Jiawen, YU Tao, ZHANG Xisoshun. Coordinated load frequency control of multi-area integrated energy system using multi-agent deep reinforcement learning [J]. Applied Energy, 2022(306): 1–12.
- [7] HUO Zhihong, XU Chang. Distributed cooperative automatic generation control and multi-event triggered mechanisms co-design for networked wind-integrated power systems [J]. Renewable Energy, 2022(193): 41–56.
- [8] ZHANG Xiaoshun, TAN Tian, ZHOU Bin, et al. Adaptive distributed auction-based algorithm for optimal mileage based AGC dispatch with high participation of renewable energy [J]. International Journal of Electrical Power & Energy Systems, 2021(124): 1–12.
- [9] BABAEI F, SAFARI A, FARROKHIFAR M, et al. Participation of electric vehicles in a delay-dependent stability analysis of LFC considering demand response control [J]. Electronics, 2022, 11(22): 1–19.
- [10] KHAN M, SUN H. Complete provision of MPC-based LFC by electric vehicles with inertial and droop support from DFIG-based wind farm [J]. IEEE Transactions on Power Delivery, 2022, 37(2): 716–726.
- [11] 席磊,全悦,刘治洪,等.基于自适应强化探索悲观 Q 的多智能体协同 AGC 算法[J].高电压技术,2023,49(6):2286–2298.
XI Lei, QUAN Yue, LIU Zhihong, et al. Multi-agent collaborative agc algorithm based on self-adaptive reinforcement-exploration maxmin Q [J]. High Voltage Engineering, 2023, 49(6): 2286–2298.
- [12] WATKINS CJCH, DAYAN P. Q-learning [J]. Machine Learning, 1992, 8(3): 279–292.
- [13] YU Tao, ZHOU Bin. A novel self-tuning CPS controller based on Q-learning method [C]// Proceedings of the Power &

- Energy Society General Meeting-conversion & Delivery of Electrical Energy in the Century, July 20 - 24, 2008, Pittsburgh, USA. New York: IEEE, 2008: 1 - 5.
- [14] ZHANG Xiaoshun, MENG Die, CAI Jiahui, et al. A swarm based double Q-learning for optimal PV array reconfiguration with a coordinated control of hydrogen energy storage system [J]. *Energy*, 2023(266): 1 - 15.
- [15] ZHAO Wenmeng, ZENG Tuo, LIU Zhihong, et al. Automatic generation control in a distributed power grid based on multi-step reinforcement learning [J]. *Protection and Control of Modern Power Systems*, 2024(4): 1 - 12.
- [16] ZHANG Xiaoshun, YU Tao, PAN Zhenning, et al. Lifelong learning for complementary generation control of interconnected power grids with high-penetration renewables and EVs [J]. *IEEE Transactions on Power Systems*, 2018, 33(4): 4097 - 4110.
- [17] XI Lei, YU Lu, XU Yanchun, et al. A novel multi-agent DDQN-AD method-based distributed strategy for automatic generation control of integrated energy systems [J]. *IEEE Transactions on Sustainable Energy*, 2020, 11(4): 2417 - 2426.
- [18] XIE Lihui, WU Junnan, LI Yanying, et al. Automatic generation control strategy for integrated energy system based on ubiquitous power Internet of Things [J]. *IEEE Internet of Things Journal*, 2022, 10(9): 1 - 10.
- [19] YIN Linfei, HE Xiaoyu. Artificial emotionab deep Q learning for real-time smart voltage control of cyber-physical social power systems [J]. *Energy*, 2023(273): 1 - 15.
- [20] ZHANG Xiaoshun, LI Qing, YU Tao. Consensus transfer q-learning for decentralized generation command dispatch based on virtual generation tribe [J]. *IEEE Transactions on Smart Grid*, 2018, 9(3): 2152 - 2165.
- [21] YAN Ziming, XU Yan. Data-driven load frequency control for stochastic power systems: a deep reinforcement learning method with continuous action search [J]. *IEEE Transactions on Power Systems*, 2019, 34(2): 1653 - 1656.
- [22] LIU Zhihong, XI Lei, QUAN Yue, et al. Brain-inspired collaborative power dynamic dispatch for the digital twin smart generation system [J]. *IEEE Transactions on Industry Applications*, 2025, 61(1): 1420 - 1430.
- [23] BUŞONIU Lucian, DE Bruin Tim, TOLIĆ Domagoj, et al. Reinforcement learning for control: performance, stability, and deep approximators [J]. *Annual Reviews in Control*, 2018(46): 8 - 28.
- [24] CHEN Ting, KORNBLITH Simon, SWERSKY Kevin, et al. Big self-supervised models are strong semi-supervised learners [J]. *arXiv.2006.10029*.
- [25] BELLMAN Richard. A Markovian decision process [J]. *Journal of Mathematics and Mechanics*, 1957: 679 - 684.
- [26] JALEELI Nasser, VANSLYCK L S. NERC's new control performance standards [J]. *IEEE Transactions on Power Systems*, 1999, 14(3): 1092 - 1099.

收稿日期: 2024-11-04; 网络首发日期: 2025-07-30

作者简介:

吴应双(1988), 男, 工程师, 博士, 从事电力系统分析、运行与控制、新能源并网控制工作, 871340366@qq.com;

刘蔚(1974), 男, 通信作者, 高级工程师, 博士, 从事交直流混合大电网仿真与计算分析工作, liuwe3@csg.cn;

刘明顺(1986), 男, 高级工程师, 硕士, 从事电网分析、运行与控制工作, lms2800@qq.com。

(上接第106页 Continued from Page 106)

- [25] DEBNATH U K, AHMAD I, HABIBI D. Quantifying economic benefits of second life batteries of gridable vehicles in the smart grid [J]. *International Journal of Electrical Power & Energy Systems*, 2014, 63(12): 577 - 587.
- [26] 李盈, 赵伟, 赵立宁, 等. 基于电池储能SOC的调峰、调频混合控制策略[J]. *高压电器*, 2023, 59(7): 48 - 55, 74.
LI Ying, ZHAO Wei, ZHAO Lining, et al. mixed control strategy of peak regulation and frequency modulation based on SOC of battery energy storage [J]. *High Voltage Engineering*, 2023, 59(7): 48 - 55, 74.
- [27] 颜宁, 张骞, 李相俊, 等. 基于荷电状态一致性的电池组双层动态均衡方法. *高电压技术*. 2024, 50(6): 2410 - 2423
YAN Ning, ZHANG Qian, LI Xiangjun, et al. A two-layer dynamic equalization method for battery modules based on state of charge [J]. *High Voltage Engineering*, 2024, 50(6): 2410 - 2423.
- [28] 张博钊, 荷斌, 徐燕璋. 循环使用与储存条件对石墨/LiCoO2 电池寿命的影响分析[J]. *电气工程学报*, 2022, 17(2): 38 - 48.
ZHANG Bozhao, GOU Bin, XU Yanzhang. Effect analysis of recycling and storage conditions on graphite/LiCoO2 battery life [J]. *Journal of Electrical Engineering*, 2022, 17(2): 38 - 48.
- [29] HASAN M, HEDAYAT S, SHAHRAM J. Power-to-gas utilization in optimal sizing of hybrid power, water, and hydrogen microgrids with energy and gas storage [J]. *Journal of Energy Storage*, 2022(45): 103745.1 - 103745.11.
- [30] FAN G Y, LIU Z J, LIU X, et al. Two-layer collaborative optimization for a renewable energy system combining electricity storage, hydrogen storage, and heat storage [J]. *Energy*, 2022, 259(15): 1 - 19.
- [31] ZHOU J L, WU Y N, ZHONG Z M, et al. Modeling and configuration optimization of the natural gas-wind-photovoltaic-hydrogen integrated energy system: a novel deviation satisfaction strategy [J]. *Energy Conversion and Management*, 2021(243): 114340.

收稿日期: 2025-04-22; 网络首发日期: 2025-08-22

作者简介:

王宁(1981), 男, 通信作者, 副教授, 博士, 研究方向为电氢系统运行优化控制和微电网容量优化配置, wnbull@ysu.edu.cn;

耿伊薇(1999), 女, 助理工程师, 硕士, 研究方向为综合能源配置, gyw202304@163.com;

王迪生(1984), 男, 工程师, 硕士, 研究方向为综合能源运行控制, disin-2000@163.com。

征 稿 启 事

《南方电网技术》是由中国南方电网公司主管、南方电网科学研究院有限责任公司主办的电力技术类科技期刊,向国内外公开发行人。刊登电力系统的科研、规划、设计、工程、生产运行、设备和系统维护等方面的成果、经验和动态,尤其是科研创新方面的论文。设有特约专稿、高压直流输电、高电压与绝缘、系统分析与运行、控制与保护、调度与通信、输变电技术、电力技术经济、智能电网、分布式电源与微电网、低碳电力等栏目。热烈欢迎国内外的电力研究人员、工程技术人员、大专院校师生踊跃投稿。

1 文稿要求

1) 文稿应未在其他刊物上发表过,译文稿件须附上原作者的书面许可证明。内容要有科学性、创新性和实用性;论点明确、数据可靠、说理严谨、数学推导简明;语言流畅、文字简练、层次分明、重点突出。

2) 学术论文请按 GB 7713—1987《科学技术报告、学位论文和学术论文的编写格式》书写,篇幅以版面不超过 9 000 字(包括图表所占篇幅)为宜。同时请写明如下 a、b、c、d、e 五项:

a) 作者署名(不宜超过 7 位)、作者单位、地址、电话、email 信箱和邮政编码;

b) 论文题目和作者单位英译文,作者姓名的汉语拼音;

c) 作者的简历,包括姓名、出生年、性别、籍贯、职称、学位、当前从事的研究或工作等项;

d) 350 字左右的中文摘要,3~8 个关键词,以及摘要和关键词相应的英文文本。摘要采用第三人称写法,内容包括论文的目的、方法、结果和结论;

e) 若为国家基金资助项目或部、省级重大科研攻关项目,需提供项目名称及编号。

2 书写格式

1) 要求文稿正文字号为 5 号,汉字字体为宋体,英数字体为 Times New Roman,希腊字体为 Symbol。文稿须采用可编辑的电子文档形式。

2) 计量单位采用国家法定计量单位和符号,不能用“大气压”、“kgf/cm²”、“卡”、“ppm”等已废除的计量单位。

3) 文中和公式中容易混淆的字符请注明文种(希文、英文、罗马字等)、大小写、上下标及上下标字母含义,表示矢量及矩阵的字符用黑斜体表示或特别注明。

4) 文稿标题中不宜用缩略词(化学符号和公知公用者除外);摘要和正文中的缩略词在第一次出现时必须写出全称,后加括号附缩略词。

3 表格、插图及参考文献

1) 表格尽量采用“三线表”。表格的上方写表序和表名。表名应有自明性且中文、英文表名并列。表注放在表底,缩入 1 个汉字以“注:”起头排版。

2) 插图的下放应有图序和图名。图名应有自明性且中英文图名并列。工程图、电气图和函数图采用 AutoCAD、

Adobe Illustrator 或 Corel DRAW 软件绘制,工程图和电气图要符合国家标准规定,函数图要标明曲线序号及其注释,坐标轴上要有标值、标目,注明物理量和单位;数码照片图具备 200 万总像素以上。

3) 图片和表格中的文字、变量、单位和数字要齐全、清晰。

4) 参考文献应选用公开发表的资料,按引用的先后次序列表于文后,以[1]、[2]…标识序号,且与正文中的指示序号一致。按《信息与文献 参考文献著录规则,GB/T 7714—2015》和《中国学术期刊(光盘版)检索与评价数据规范,CAJ-CD B/T 1—2006 修订版试行稿》的要求著录文后参考文献。中文参考文献后要并列其英文译文(如示例[6]所示)。

文献类型及其标识为:普通图书[M];会议论文[C];报纸文章[N];期刊文章[J];学位论文[D];报告[R];标准[S];专利[P];汇编[G];档案[B];古籍[O];参考工具[K];其他未说明的文献类型,例如可公开的政府行政部门编号文件、行业或大公司的技术规范或工作手册[Z]。网上期刊[J/OL];网上电子公告[EB/OL]。电子文献尚需在载体标记后加上发表或更新日期(加圆括号)、引用日期(加方括号)和电子文献网址。

4 投稿

1) 请登录《南方电网技术》编辑部网站 <https://nfdwjs.csg.cn>,选择“作者登录”,登录或注册登录后选择投稿并按提示操作,登录方面出现问题请致电 020-36625646 咨询。

2) 投稿时务必附上联系电话(手机为佳)。

3) 编辑部不保证留存稿件,请作者自留底稿。稿件评审进程和结论可在《南方电网技术》编辑部网站上即时查询。作者也可致电 020-36625642、36625643 或发送电子邮件到 nfdwjs@csg.cn 向编辑部了解审稿情况。

4) 文章一经发表,编辑部便取得作者的自动授权,可将文章收录入本刊网站,及与本刊有合作关系的数据库。

参考文献著录格式示例:

- [1] 刘国钧,陈绍业,王凤翥. 图书馆目录[M]. 北京:高等教育出版社,1957:15-18.
- [2] 钟文发. 非线性规划在可燃毒物配置中的应用[C]//赵玮. 运筹学的理论与应用——中国运筹学会第五届大会论文集. 西安:西安电子科技大学出版社,1996:468-471.
- [3] 冯西桥. 核反应堆压力管道与压力容器的 LBB 分析[R]. 北京:清华大学核能技术设计研究院,1997.
- [4] 中国国家标准化管理委员会. 信息与文献 参考文献著录规则:GB/T 7714—2015[S]. 北京:中国标准出版社,2015.
- [5] 邓一刚. 全智能节电器:200610171314.3[P]. 2006-12-13.
- [6] 陈祥. 无线技术在现代电网管理中的应用与展望[J]. 南方电网技术,2009,3(6):85-89.
CHEN Xiang. Development of wireless technology in modern power grid management and its perspective [J]. Southern Power System Technology, 2009,3(6):85-89.
- [7] 王明亮. 关于中国学术期刊标准化数据库系统工程的进展[EB/OL]. (1998-08-16) [1998-10-04]. <http://www.cajcd.edu.cn/pub/wml.txt/980810-2.html>.

《南方电网技术》编辑部