

文章编号: 2095-2163(2023)10-0108-08

中图分类号: TM734

文献标志码: A

基于 MATSAC-LSTM 的综合能源系统自动发电控制算法研究

李昊, 荣娜, 邓棋宸

(贵州大学 电气工程学院, 贵阳 550025)

摘要: 为提高综合能源系统自动发电控制(Automatic Generation Control, AGC)的控制性能和算法收敛速度, 本文提出了一种基于多智能体迁移柔性行动器-批判器与长短时记忆网络(Multi-Agent Transfer Soft Actor-Critic with Long-Short Term Memory, MATSAC-LSTM)的AGC控制法。首先, 用LSTM网络将采集的区域控制误差等环境状态量进行时序特征提取, 并作为MATSAC算法的输入, 使智能体能结合历史信息进行快速的有功功率分配决策; 其次, 采用集中训练分散执行框架, 将一个智能体观察的环境状态量以及其他智能体的动作信息作为相应智能体Critic网络的输入, 以便训练时能够让多智能体之间共享信息; 最后, 通过迁移学习将旧任务训练的Critic和Actor网络模型参数转移到新任务相应模型参数中, 以提高智能体的训练效率。算例分析在一个修改的IEEE标准两区域负荷频率控制系统模型和一个五区域综合能源系统模型展开, 仿真结果表明, 与比例积分微分、Q学习、双延迟深度确定性策略梯度、基于动态策略的赢或快速学习爬坡策略、柔性行动器-批判器等传统算法相比, 本文所提MATSAC-LSTM算法提高了AGC控制性能标准和算法收敛速度, 降低了系统的区域控制误差和频率偏差。

关键词: 综合能源系统; 迁移学习; 柔性行动器-批判器; 集中训练分散执行

Research on automatic generation control algorithm of integrated energy system based on MATSAC-LSTM

LI Hao, RONG Na, DENG Qichen

(School of Electrical Engineering, Guizhou University, Guiyang 550025, China)

[Abstract] To improve the control performance of automatic generation control (AGC) in integrated energy system and the convergence speed of the algorithm, this paper proposes an AGC algorithm based on multi-agent transfer soft actor-critic with long-short term memory (MATSAC-LSTM). Firstly, the environmental state variables such as area control error are extracted with temporal features using LSTM networks and used as input to the MATSAC algorithm so that agents can combine the historical information to make fast active power allocation decisions. Furthermore, the centralized training decentralized execution framework is used to allow the environmental state variables observed by one agent and the actions of other agents to be used as input to the corresponding agent critic network, which enables information sharing among multiple agents during training. Finally, the parameters of the critic and actor network models trained on the old task are transferred to the corresponding model parameters of the new task through transfer learning to improve the training efficiency of the agents. The case studies are carried out in a modified IEEE standard two-area load frequency control system model and a five-area integrated energy system model. The simulation results show that the MATSAC-LSTM algorithm effectively improves the control performance standard and convergence speed of the system and reduces area control error and frequency deviation of the system compared with the conventional algorithms such as proportion integration differentiation and Q learning.

[Key words] integrated energy system; transfer learning; soft actor-critic; centralized training with decentralized execution

0 引言

随着全球能源转型, 新能源大规模分散式接入电网, 致使电力系统随机性、间歇性不断增强, 传统

自动发电控制(AGC)策略已经不满足电网需求^[1-2]。因此, 从AGC策略的角度, 寻找一种在综合能源系统背景下有效提高系统安全稳定运行的控制策略具有重要意义^[3-4]。

基金项目: 贵州省科学技术基金(2021277); 贵州省优秀青年科技人才项目资助(20215645)。

作者简介: 李昊(1998-), 男, 硕士研究生, 主要研究方向: 综合能源与自动发电控制; 荣娜(1979-), 女, 博士, 讲师, 硕士生导师, 主要研究方向: 电力系统与综合能源、电力电子装备与系统。

通讯作者: 荣娜 Email: nrong@gzu.edu.cn

收稿日期: 2022-09-13

AGC 控制策略实际上是根据电网实时运行工况在线计算出最优决策^[5]。目前, AGC 控制策略可以分为传统 AGC 控制策略和智能 AGC 动态优化策略。传统 AGC 控制策略有模糊控制、自适应控制、鲁棒控制、比例积分微分 (Proportional Integral Derivative, PID) 控制等方法。文献[6]使用自适应乌鸦搜索算法提出了一个最优模糊 PID 控制器, 将其应用于非线性两区域和三区域再热系统的 AGC, 减少频率偏差。随着大规模新能源接入电网, 数据维度变大, 传统 PID 控制方法控制难度急剧增大。伴随人工智能的进步与发展, 有关 AGC 的智能算法也相继应用, 试图解决传统控制方法暴露的问题。

由于强化学习具有实时性和自适应性, 智能体可以与环境交互收集信息, 不断试错和探索, 从而可以在综合能源系统获得最优控制策略。文献[7]将 Q 学习算法应用于 AGC, 依靠 Q 值函数和控制性能标准 (Control Performance Standards, CPS) 控制动作形成闭环反馈来形成最优控制策略, 增强了算法的适应性和控制性能; 文献[8]提出多经验池概率回放的双延迟深度确定性策略梯度 (Multiple Experience pool experience replay Twin Delayed deep deterministic policy gradient, ME-TD3) 算法, 采用不同概率从不同经验池采样, 提高最优策略的质量; 文献[9]引入卷积神经网络 (Convolution Neural Network, CNN), 解决了传统强化学习算法中维数灾难的问题, 但缺乏考虑历史状态变化, 造成局部最优的问题; 文献[10]提出一种基于动态策略的赢或快速学习爬坡策略 (Policy Dynamics based Win or Learn Fast Policy Hill-Climbing, PDWoLF-PHC) 算法, 通过改变学习率在各种复杂的电力系统环境中得到最优策略, 解决了新能源和分布式能源接入电

网时产生的强随机扰动问题, 但由于采样复杂度较高, 算法收敛速度较慢。

为解决复杂综合能源系统背景下 AGC 算法收敛速度慢、控制性能差的问题, 本文提出了一种基于多智能体迁移柔性行动器-批判器与长短时记忆网络 (Multi-Agent Transfer Soft Actor-Critic with Long-Short Term Memory, MATSAC-LSTM) 的算法, 该算法具有更强鲁棒性, 最大熵政策可以使智能体具有更强的探索能力, 能够得到全局最优解, 融合迁移学习使得算法的收敛速度变快。在一个修改的 IEEE 标准两区域负荷频率控制系统模型和一个五区域综合能源系统进行了算例分析。结果表明, 该方法能有效提高系统的控制性能指标和收敛速度, 降低了系统的区域控制误差和频率偏差。

1 MATSAC-LSTM 算法

传统强化学习算法在 AGC 应用上收敛速度慢, 控制性能差, 因此本文在柔性行动器-批判器 (Soft Actor-Critic, SAC) 算法的基础上, 用 LSTM 网络将采集的区域控制误差等环境状态量进行时序特征提取, 作为 MATSAC 算法的输入, 多智能体框架使得智能体之间信息共享, 并通过迁移学习来解决收敛速度慢的问题, 进而提出 MATSAC-LSTM 算法来提高收敛速度和控制性能。

本文提出 MATSAC-LSTM 算法架构如图 1 所示。本文所研究的 AGC 系统是动态随机环境, MATSAC-LSTM 算法根据系统所处的状态计算出相应的奖励值, 将当前系统环境的状态量作为 MATSAC-LSTM 算法的输入, 智能体在每个控制周期中, 给出最优功率发电指令。

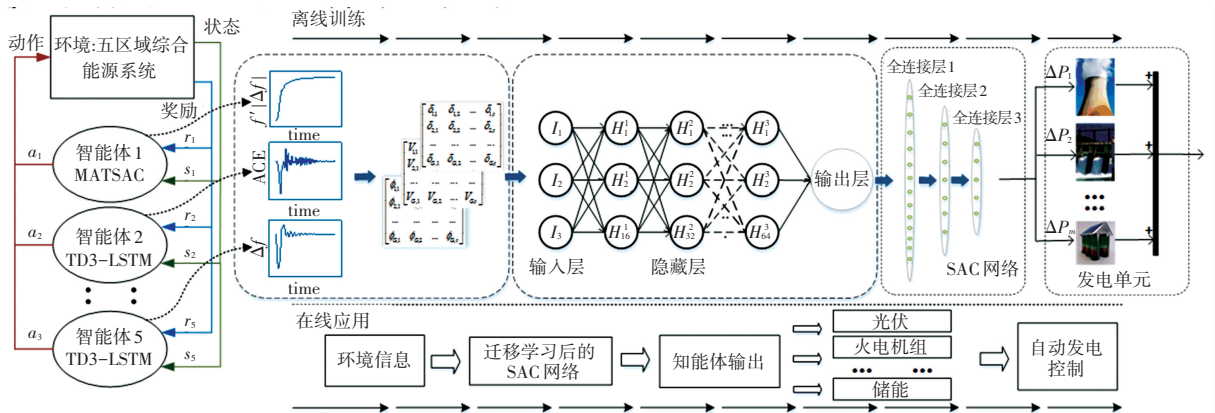


图 1 MATSAC-LSTM 算法架构

Fig. 1 Architecture of MATSAC-LSTM algorithm

1.1 柔性行动器-批判器算法

深度强化学习具有处理高维连续状态-动作空间的特点,而柔性行动器-批判器(SAC)算法是最好异策略深度强化学习算法之一,相较于双延迟深度确定性策略梯度(Twin Delayed deep deterministic policy gradient, TD3)算法和近端策略优化(Proximal Policy Optimization, PPO)算法,该算法可以使政策的熵值和预期收益最大化,从而使样本学习的效率得到提高,SAC算法框架如图2所示。

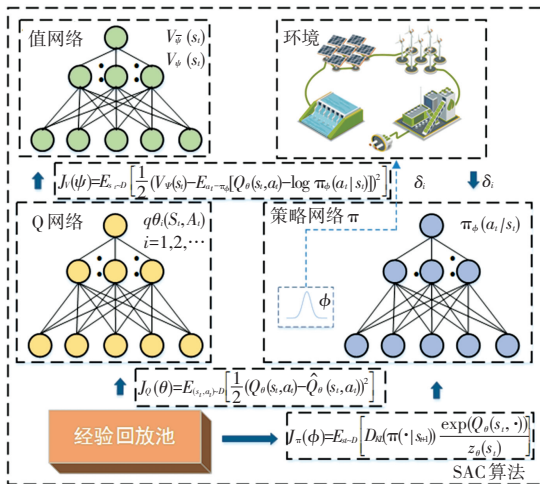


图2 SAC算法框架

Fig. 2 Framework of SAC algorithm

在SAC算法中,概率策略的熵可以描述为式(1):

$$H(\pi(\cdot | s_t)) = - \sum_a \pi(a | s_t) \ln \pi(a | s_t) \quad (1)$$

其中, s_t 代表当前智能体的状态, a 代表当前智能体的动作。

在强化学习算法框架中的最大熵值函数,可以描述为式(2):

$$Q_h^{\pi}(s_t, a_t) = E_{\tau \sim \pi} \left[\sum_{t=0}^T r^t R_t + \alpha \sum_{t=1}^T r^t H(\pi(\cdot | s_t)) | s_0 = s, a_0 = a \right] \quad (2)$$

其中, E 代表数学期望; T 为智能体和环境交互的时间步数; a_t 代表智能体在时间 t 执行的动作。

在SAC算法中,有3个神经网络,分别是: V 网络, Q 网络和策略网络。

V 网络 $V_{\psi}(s_t)$ 由损失函数的均方误差更新,式(3):

$$J_V(\psi) = E_{s_t \sim D} \left[\frac{1}{2} V_{\psi}(s_t) - E_{a_t \sim \pi_{\phi}}(Q_{\theta}(s_t, a_t)) - \right.$$

$$\left. \log \pi_{\phi}(a_t | s_t) \right]^2 \quad (3)$$

其中, D 是训练样本的经验回放池, λ_V 是 V 网络的学习速率。

策略网络 $\pi_{\phi}(a_t | s_t)$ 可以由 KL 散度损失公式更新,式(4):

$$J_{\pi}(\phi) = E_{s_t \sim D} \left[D_{KL}(\pi(\cdot | s_{t+1}) || \frac{\exp(Q_{\theta}(s_t, \cdot))}{Z_{\theta}(s_t)}) \right]$$

$$\phi \leftarrow \phi - \lambda_{\pi} \nabla_{\psi} J_{\pi}(\phi) \quad (4)$$

其中, λ_{π} 代表 V 网络的学习速率。

Q 网络 $Q_{\theta}^{\pi}(s_t, a_t)$ 通过贝尔曼误差公式进行更新,式(5)和式(6):

$$J_Q(\theta) = E_{(s_t, a_t) \sim D} \left[\frac{1}{2} (Q_{\theta}(s_t, a_t) - \hat{Q}_{\theta}(s_t, a_t))^2 \right] \quad (5)$$

$$\hat{Q}_{\theta}(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1} \sim p} [V_{\psi}(s_{t+1})] \quad (6)$$

其中, $r(s_t, a_t)$ 代表智能体执行动作时获得的奖励, $V_{\psi}(s_t)$ 代表目标网络。

1.2 多智能体迁移柔性行动器-批判器算法

与SAC算法相比,多智能体迁移柔性行动器-批判器算法是将单智能体采用集中训练分散执行(Centralized Training with Decentralized Execution, CTDE)框架拓展到多智能体。在训练阶段,本文在每个区域设置一个智能体,智能体的Actor网络和目标Actor网络采集当前状态的环境信息和下一状态的环境信息,然后生成智能体的当前动作和目标动作。将一个智能体观察的环境状态量,以及其他智能体的动作信息作为相应智能体Critic网络的输入,Critic网络输出当前动作的 Q 值。此外,每个智能体都有一个自己的经验回放池,以提高MASAC算法的稳定性。在智能体收集足够经验回放池中的数据后,每个智能体从经验回放池中随机取样来训练自己。当经验回放池中有足够的数据时,智能体随机抽样得到的数据接近于独立的相同分布,设置经验回放池可以打破序列之间关联性,避免模型陷入局部最优。

1.3 基于LSTM网络的MASAC框架构建

长短时记忆网络(Long-Short Term Memory, LSTM)网络是在循环神经网络(Recurrent Neural Network, RNN)基础上改进的,每个LSTM都是一组捕获数据的单元,这些单元从一个模块连接到另一个模块,传输过去的的数据,并收集当前的数据。LSTM模型示意图如图3所示。

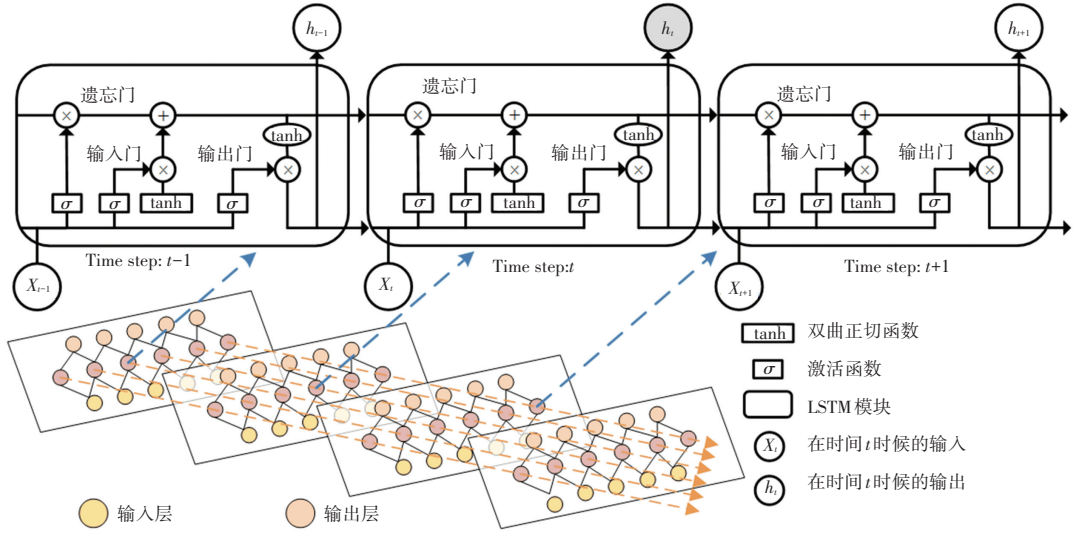


图 3 LSTM 模型示意图

Fig. 3 Schematic diagram of LSTM model

LSTM 网络包括 3 个门:输入门、输出门和遗忘门,帮助 LSTM 处理顺序数据。遗忘门 f_t 、输入门 i_t 、输出门 o_t 由式(7)~式(9)计算得出。

$$f_t = \sigma_g(\mathbf{W}_f x_t + \mathbf{U}_f h_{t-1} + \mathbf{b}_f) \quad (7)$$

$$i_t = \sigma_g(\mathbf{W}_i x_t + \mathbf{U}_i h_{t-1} + \mathbf{b}_i) \quad (8)$$

$$o_t = \sigma_g(\mathbf{W}_o x_t + \mathbf{U}_o h_{t-1} + \mathbf{b}_o) \quad (9)$$

其中, x_t, h_t 分别代表输入层和输出层; $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_o, \mathbf{W}_c$ 是隐藏层输入映射到3个门的权重矩阵; $\mathbf{U}_f, \mathbf{U}_i, \mathbf{U}_o, \mathbf{U}_c$ 是关于3个门与输入单元状态之间连接相关的权重矩阵; $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_o, \mathbf{b}_c$ 是偏差向量; σ_g 是门激活函数。

计算得到的3个门的输出后,由式(10)和式(11)更新输出:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (10)$$

$$h_t = o_t \tanh(C_t) \quad (11)$$

其中, C_t 代表长期记忆信息, h_t 代表短期记忆信息。

在模型训练时,LSTM网络将已经输出的记忆信息与电力系统的当前状态连接,传递给 Actor 网络和 Critic 网络作为输入,使得智能体输入的信息更加完整。

1.4 迁移学习

迁移学习涉及源领域和目标领域,从源领域学习到的知识可以转移到目标领域,源领域和目标领域相似,则知识转移的过程相对容易,从而不考虑在两者之间分布和适应性方面的差异。不同综合能源系统由于其机组不同、结构不同,所以其控制策略也有差异。因此,迁移学习可以与 SAC 算法相结合,从而使 SAC 算法在 AGC 中学习效率得到提高。在迁移学习中,通常采用两种策略,一种是传递学习模

型的参数;另一种是共享由以前训练过的智能体。本文采用第一种方法,即在源领域训练后的智能体向目标领域智能体进行参数传递。

2 基于 MATSAC-LSTM 的 AGC 设计

基于 MATSAC-LSTM 算法的自动发电控制是通过各个区域的智能体实时感知综合能源系统中的环境信息,智能体通过收集当前系统的环境信息状态量作为 MATSAC-LSTM 算法的输入,以奖励函数作为目标函数,计算出相应奖励值,算法在每个控制周期中,智能体输出最优动作作为实际电网调度端所有机组最优的总发电调节指令。

综合能源的控制性能可以通过区域控制偏差(Area Control Error, ACE)ACE 和 CPS1 来衡量。

ACE 计算,式(12):

$$ACE = -10B\Delta f + \Delta P_T \quad (12)$$

其中, Δf 代表电网实际频率与计划频率之差; ΔP_T 代表联络线上实际交换功率与计划交换功率之差; B 代表区域定义的频率偏差系数。

CPS1 指标,式(13):

$$CPS1 = \left(2 - \frac{1}{\varepsilon_1} \left(\frac{ACE}{-10B} \right)_{\min} \times \Delta f_{\min} \right) \times 100 \quad (13)$$

其中, ε_1 是互联电网对全年 1 min 频率平均偏差均方根的控制目标值, Δf_{\min} 代表相应变量在 1 min 内的平均值。

2.1 智能体设计

(1)状态空间:对于任何时刻 t , S_t 代表智能体在 t 时刻当前的状态,可以用式(14)表示:

$$S_t^i = \left[\int ACE_i dt, \Delta f_i, ACE_{i+1} \right] \quad (14)$$

其中, ACE_i 是第 i 区域的区域控制误差的瞬时值, Δf_i 是第 i 区域频率偏差的瞬时值。

(2) 动作空间: 对于任何时刻 t , a_t 包括所有智能体的动作, a_t^i 代表智能体在 t 时刻当前的动作, 可以用式(15)表示:

$$a_t^i = [\Delta P_{Gi}] \quad (15)$$

其中, ΔP_{Gi} 是第 i 区域实际电网调度端所有机组的总发电调节指令。

(3) 奖励函数: 将 ACE 作为目标函数, 能够使 CPS 指标保持高水平稳定且功率限制在小范围内波动。而系统控制性能也能由频率偏差的绝对值 $|\Delta f|$ 的大小直接反应。本文将 ACE 和 $|\Delta f|$ 作为算法的奖励函数, 并且对 ACE 和 $|\Delta f|$ 的量纲进行归一化处理, 奖励函数式(16):

$$R_t = -\alpha_1 |ACE(t)| - \alpha_2 100 |\Delta f(t)| \quad (16)$$

其中, $|ACE(t)|$ 是 t 时刻 ACE 的绝对值; $|\Delta f(t)|$ 是 t 时刻频率偏差的绝对值; α_1 和 α_2 是 $|ACE(t)|$ 和 $|\Delta f|$ 的权重, 且 $\alpha_1 = \alpha_2 = 0.5$ 。

2.2 神经网络结构

本文智能体模型的 Critic 网络和 Actor 网络的输入是 LSTM 网络的输出, 相较于其他神经网络, LSTM 神经网络能够更好的将历史信息和目前的状态信息结合起来, 生成最优策略。Actor 网络设置 3 个全连接层, ReLU 激活函数层, Tanh 激活函数层, 其结构示意图如图 4 所示。Actor 神经网络的隐藏神经元分别是: 128、64、1; Critic 网络的输入为环境状态和动作的合集, 也含 3 个全连接层, ReLU 激活函数层, Tanh 激活函数层, 最后输出最优策略, 3 个全连接层的神经元的个数分别是: 64、32、1。

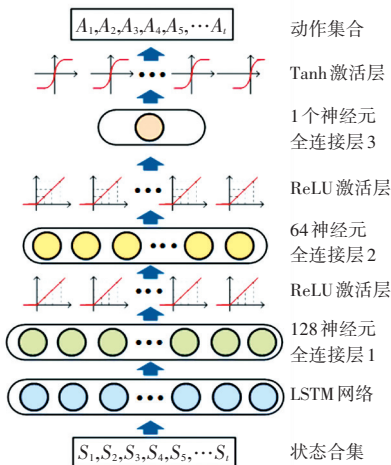


图 4 Actor 网络结构示意图

Fig. 4 Schematic diagram of Actor network structure

3 算例分析

为验证本文所提的 MATSAC-LSTM 算法的有效性和适用性, 在一个修改的 IEEE 标准两区域负荷频率控制系统模型上采用多智能体柔性行动器-批判器与长短时记忆网络算法验证 LSTM 网络和多智能体结合 SAC 算法在 AGC 中的控制效果, 并通过迁移学习在一个五区域综合能源系统模型上采用 MATSAC-LSTM 算法进行仿真实验。

模型中的一些参数的选择将影响算法在自动发电控制中的应用效果。这些参数按照以下原则选取:

(1) 折扣因子: 代表知识矩阵在更新过程中对过去奖励值的折扣。如果累计的奖励值对实际问题的求解影响大, 则折扣因子选择较大的值。对于自动发电控制而言, 目前环境所得到的奖励的瞬时值更为重要, 所以折扣因子选择较小的值。

(2) 学习率: 如果设置太小, 则算法收敛缓慢, 但容易找到全局最优解。设置越大, 算法的收敛速度越快, 但可能导致算法无法收敛。本文在引入迁移学习的方法后, 算法已经具有较好的先验知识, 所以学习率选择较小的值。

(3) 经验池长度: 如果经验池过大, 则会导致无用的经验加入到抽样过程中; 如果经验池过小, 则会导致算法不能通过经验池的抽样获得最优解。本文通过大量实验, 选取 1 000 000。

(4) dropout: dropout 如果过大则会影响算法的拟合能力; 如果过小, 则会导致数据样本不足而过拟合。本文在自动发电过程中, 会获得大量样本, 所以 dropout 选择较小的值。

(5) 批次大小: 在一定范围内, 批次越大, 引起训练震荡越小, 收敛精度越高。但如果过大, 则会增加训练时间, 且收敛精度也不会提高。本文进行了大量的实验, 选取批次大小为 512 最为有效。

通常, 在不同环境下综合能源系统的运行工况会实时变化, 需要设置不同的最优参数才能寻求到不同环境下自动发电控制的最优解, 但是这将耗费大量时间, 本文根据参数选取的原则和大量的实验仿真得到一组最优的参数见表 1。

3.1 一个修改的 IEEE 标准两区域负荷频率控制系统模型

本文在 IEEE 标准两区域负荷频率控制系统模型基础上融入电池储能、风电、光伏。

表 1 模型参数

Tab. 1 Parameters of model

参数	数值
折扣因子 (discount factor)	0.1
学习率 (actor learning rate)	0.001
经验池长度 (Experience Buffer Length)	1 000 000
dropout	0.01
L2 正则化因子 (L2 regularization factor)	0.000 2
批次大小 (batch size)	512
目标平滑因子 (target smooth factor)	0.001
输出层激活函数 (activation function for output layer)	Tanh
隐藏层激活函数 (activation function for input layer)	ReLU

3.1.1 智能体训练

MASAC-LSTM 算法分为离线训练和在线测试两个阶段。离线训练阶段,智能体的学习步长为 AGC 系统的控制周期,该标准算例中取 4 s,不断更新智能体的策略,使控制器的控制效果达到最优。最终对修改的 IEEE 标准两区域负荷频率控制系统模型的第一个区域施加一个周期 800 s,幅值 1 000 MW,时间 8 000 s 的正弦负荷扰动。

各种算法智能体学习过程如图 5 所示。与 PID、Q、TD3、PDW_oLF-PHC、SAC 算法相比,MASAC-LSTM 算法在 1 200 s 后,已经能够稳定地跟踪负载扰动变化,能够更准确的跟踪负荷扰动,收敛速度最快。由于 MATSAC-LSTM 算法可以通过 LSTM 网络将采集的区域控制误差等环境状态量进

行时序特征提取,并作为 MATSAC 算法的输入,使得智能体能够结合历史信息做出更优的决策,因此使得奖励值震荡次数减少,具有更好的动态性能。

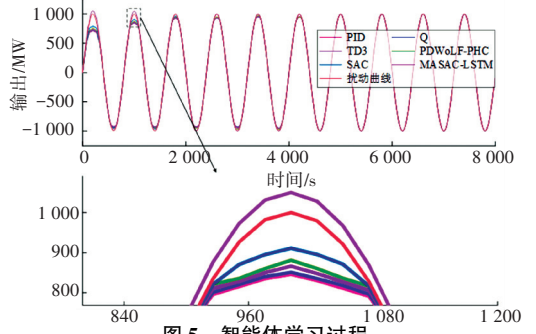


图 5 智能体学习过程

Fig. 5 The learning process of agents

3.1.2 阶跃扰动

为了评估 MASAC-LSTM 算法的可靠性和鲁棒性,引入振幅为 1 000 MW 阶跃扰动。基于 Q、PID、TD3、PDW_oLF-PHC、SAC、MASAC-LSTM 算法的在线测试结果如图 6 所示。可以看出,MASAC-LSTM 算法在两个区域的 ACE 和 Δf 的峰值明显小于其他 3 种算法,说明 MASAC-LSTM 算法有效地减小了 ACE 的偏差,在 220 s 内 MASAC-LSTM 算法可以使各区域达到稳定。区域联络线功率偏差 ACE 可以直接看出区域之间数据的共享程度以及多个区域之间协同控制的配合程度,MASAC-LSTM 算法采用了多智能体集中训练分散执行框架,使智能体之间信息共享,有效实现多个区域之间的最优协同控制。

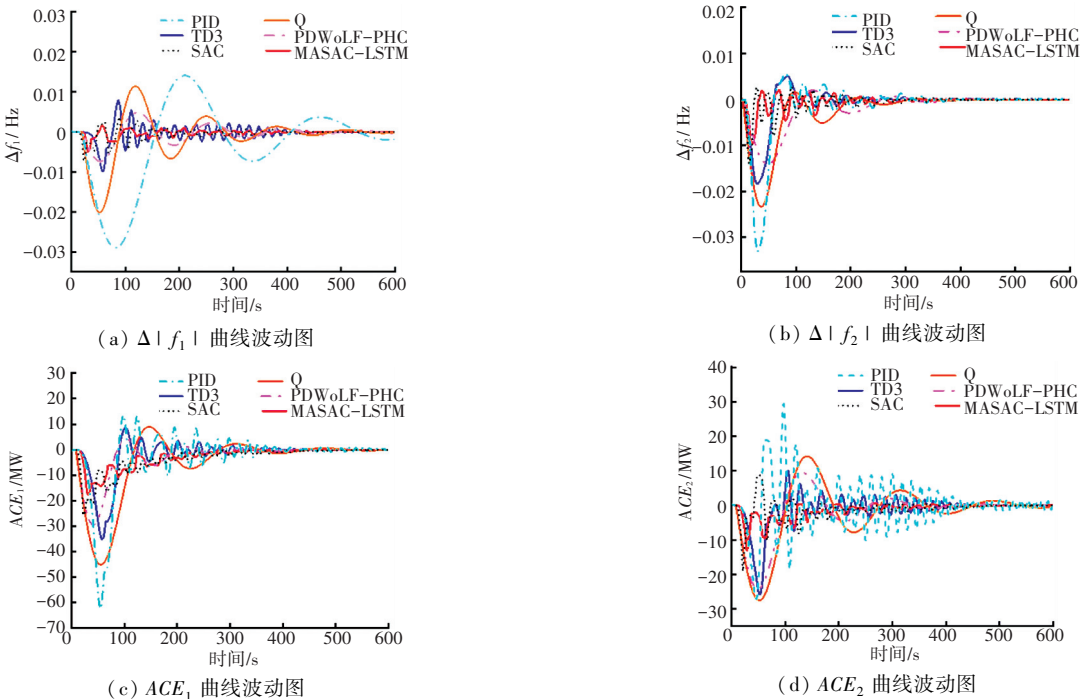


图 6 阶跃扰动曲线

Fig. 6 Curve of step perturbation

3.1.3 方波扰动

引入考核周期为 1 500 s,幅值在 800 MW 以内的方波负荷扰动,5 种算法的控制性能见表 2。可以看出,MASAC-LSTM 的 $|\Delta f|$ 的平均值降低 0.002 2~0.004 Hz, $|ACE|$ 的平均值降低 2.391~11.869 MW, CPS1 的平均值增加 0.841%~2.311%。

表 2 不同算法的控制性能

Tab. 2 Control performance of different algorithms

算法	$ \Delta f / \text{Hz}$	$ ACE / \text{MW}$	CPS1 / %
PID	0.008 3	28.123	196.541
Q	0.007 3	22.482	197.521
TD3	0.006 5	18.645	198.011
PDWoLF-PHC	0.005 9	17.985	198.211
SAC	0.005 7	17.525	198.456
MASAC-LSTM	0.004 3	16.254	198.852

3.2 一个五区域综合能源系统模型

随着电网新能源占比提高,为了考虑大量分布能源并入电网造成的影响,本文构建五区域综合能源模型。此模型在一个 IEEE 标准两区域负荷频率控制系统模型上加入了电池储能、风电、光伏、热电联产、柴油发电、核能水电等机组,每个区域设置独立的智能体,实现多智能体协同控制。

因为 MASAC-LSTM 算法需要智能体与环境长时间进行交互,收集数据从而获得最佳策略,本文对修改的 IEEE 标准两区域负荷频率控制系统模型的智能体向五区域综合能源系统模型的智能体传递参数,从而缩短智能体的训练时间。

第一组实验采用 MASAC-LSTM 算法在构建的五区域综合能源系统模型上训练 320 轮;第二组实验采用迁移学习的方法,将修改的 IEEE 标准两区域负荷频率控制系统模型的 Critic 网络和 Actor 网络的第一层全连接层参数转移到五区域综合能源系统模型参数中;第三组实验将修改的 IEEE 标准两区域负荷频率控制系统模型的训练模型 Critic 网络和 Actor 网络的第一层和第二层全连接层转移到五区域综合能源系统模型参数中;第二组实验和第三组实验的 Critic 网络和 Actor 网络的输出层都被重置,且将回放缓冲区清空,训练 320 轮。

迁移学习训练曲线如图 7 所示。由图 7 可见,第一组实验的起始奖励为 -26.7,第二组实验为 -23.2,第三组实验为 -21.2,说明 MATSAC-LSTM 算法使智能体能更好地收集经验;在 100 轮时,第三组实验奖励已经达到 -13,而第一组实验的奖励还在缓慢上升,MATSAC-LSTM 算法通过迁移学习可以使奖励增长率变大;第一组实验最后奖励稳定在

-12.4,第二组实验为 -9.1,第三组实验为 -7.9,说明通过迁移学习 MATSAC-LSTM 算法在求解质量方面得到提高。实验结果说明通过迁移学习将旧任务训练的 Critic 和 Actor 网络模型参数转移到新任务相应模型参数中,可以减少整个算法的训练时间,而对于转移模型参数的选择,在重置输入层和输出层的基础上,转移的神经网络参数越多,训练的收敛速度越快,且寻优结果也更好。

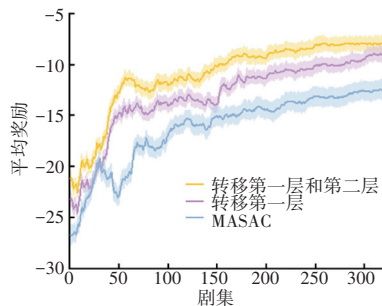


图 7 迁移学习训练

Fig. 7 Training diagram for transfer learning

考虑到综合能源系统的随机性和间歇性,在五区域综合能源系统中引入幅值为 1 000 MW,持续时间为 10 000 s 的随机扰动,扰动曲线如图 8 所示。

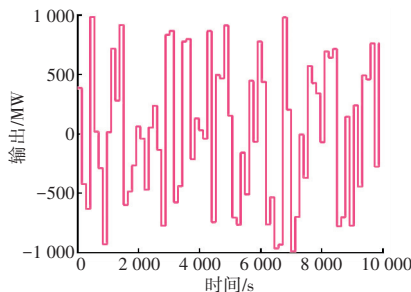


图 8 随机扰动曲线

Fig. 8 The curves of random disturbance

PID、Q、TD3、PDWoLF-PHC、SAC、MASAC-LSTM、MATSAC-LSTM 7 种算法在随机扰动下的控制性能如图 9 所示。相较于其他算法,本文所提算法的 $|\Delta f|$ 、 $|ACE|$ 、稳态误差、超调量有所降低,CPS1 有所提高。

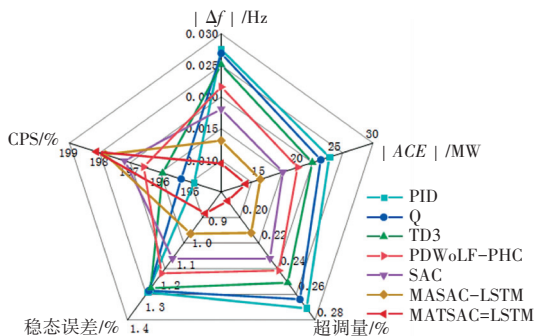


图 9 随机扰动下算法的控制性能

Fig. 9 Control performance of algorithms under random perturbation