



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



多微网系统端网协同分布式实时智能优化

王丹璐, 孙秋野, 苏涵光

引用本文:

王丹璐, 孙秋野, 苏涵光. 多微网系统端网协同分布式实时智能优化[J]. 控制与决策, 2024, 39(11): 3801–3809.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.1122>

您可能感兴趣的其他文章

Articles you may be interested in

求解约束优化问题的改进果蝇优化算法及其工程应用

Improved fruit fly optimization algorithm for solving constrained optimization problems and engineering applications

控制与决策. 2021, 36(2): 314–324 <https://doi.org/10.13195/j.kzyjc.2019.0557>

认知智能电网中基于能效优化的频谱分配策略

Spectrum allocation strategy based on energy efficiency optimization in cognitive smart grid

控制与决策. 2021, 36(8): 1901–1910 <https://doi.org/10.13195/j.kzyjc.2019.1448>

区间数可重入混合流水车间调度与预维护协同优化

Collaborative optimization of interval number reentrant hybrid flow shop scheduling and preventive maintenance

控制与决策. 2021, 36(11): 2599–2608 <https://doi.org/10.13195/j.kzyjc.2020.0973>

基于深度强化学习与迭代贪婪的流水车间调度优化

Scheduling optimization for flow-shop based on deep reinforcement learning and iterative greedy method

控制与决策. 2021, 36(11): 2609–2617 <https://doi.org/10.13195/j.kzyjc.2020.0608>

基于协同控制的串联超级电容电压均衡策略

Voltage equalization strategy for series-connected ultracapacitors based on cooperative control

控制与决策. 2021, 36(8): 1997–2001 <https://doi.org/10.13195/j.kzyjc.2019.1736>

多微网系统端网协同分布式实时智能优化

王丹璐, 孙秋野[†], 苏涵光

(东北大学 信息科学与工程学院, 沈阳 110819)

摘要: 多微网互联对于提升微网稳定性和促进可再生能源消纳具有重要作用, 对此, 考虑多微网系统网内多智能终端协同和网间功率互济需求, 提出一种基于多智能体强化学习的端-网实时协同优化算法. 该算法可自适应源荷不确定性和系统拓扑变化, 实时产生网间能量互济和网内智能终端协同优化策略. 首先建立端-网协同优化系统结构和优化模型; 其次建立基于多智能体马尔科夫决策过程的强化学习模型, 进而提出基于多智能体近端策略优化算法 (MAPPO) 的分布式协同优化算法; 然后考虑调节过程中功率平衡约束, 设计一种新的功率平衡反馈信号, 能够有效避免功率不平衡现象的出现; 最后针对 3 个典型场景进行仿真, 结果表明各微电网无需全局信息便可得到准确的近似全局最优解. 将所提出方法分别与状态完全观测和状态部分观测的分布式强化学习算法进行对比, 结果表明所提出的方法既能获得良好的协同优化效果, 又能满足实时优化对于算法效率的需求.

关键词: 端网协同; 多智能体; 近端策略优化; 能量互济; 多微网; 实时优化

中图分类号: TM73; TP18

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.1122

引用格式: 王丹璐, 孙秋野, 苏涵光. 多微网系统端网协同分布式实时智能优化 [J]. 控制与决策, 2024, 39(11): 3801-3809.

Collaborative distributed real-time intelligent optimization of multi-microgrid system

WANG Dan-lu, SUN Qiu-ye[†], SU Han-guang

(College of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract: Interconnection between multiple microgrids plays an important role on enhancing stability of microgrids and renewable energy utilization. Considering the needs of collaborative operation of intelligent terminals and power mutual assistance between microgrids, a real-time collaborative terminal-microgrid optimization algorithm based on multi-agent reinforcement learning is proposed. The proposed algorithm can flexibly adapt to the system uncertainties and topology changes, and optimizes energy interaction between microgrids and collaboration among intelligent terminals. Firstly, the structure of the collaborative terminal-microgrid optimization system is established, and the corresponding optimization model is formulated. Then, a multi-agent Markov decision process based reinforcement learning model is proposed. Furthermore, a distributed collaborative optimization algorithm based on the multi-agent proximal policy optimization (MAPPO) algorithm is proposed. Considering the power balance constraint, a novel power balance feedback is designed to effectively avoid the occurrence of power imbalance. Simulations are conducted under three typical scenarios, and the results show that approximate global optimal solutions can be obtained without global state observation. What is more, the proposed algorithm is compared with the distributed reinforcement learning algorithms with fully state observation and partial state observation, respectively. Results show that the proposed method can achieve good collaborative optimization results and meet the requirements on compute efficiency of real-time optimization.

Keywords: terminal-grid collaboration; multi-agent; proximal policy optimization; energy interaction; multi-microgrid; real-time optimization

0 引言

在“碳中和、碳达峰”^[1-3]的背景下, 分布式可再生能源 (distributed renewable generator, DRG) 因其清

洁低碳的特点而成为能源供给的主要来源之一. 以 DRG 为主体的能源系统得到了快速发展, 微网作为实现可再生能源高效利用的重要方式, 受到了国内外

收稿日期: 2023-08-07; 录用日期: 2023-12-21.

基金项目: 国家重点研发计划项目 (2018YFA0702200); 国家自然科学基金重点项目 (U20A20190); 国家自然科学基金面上项目 (62073065).

[†]通讯作者. E-mail: sunqiuye@ise.neu.edu.cn.

学者的关注^[4-5]。但单一的独立微电网存在电力不稳定、容量有限、鲁棒性差、运行成本高等缺点。因此,形成了多个独立微电网互联的多微网系统,例如,对于远洋海岛供电,架设与大陆电网相连的海底电缆。这种多微网系统具有造价高、建设难度大的缺陷。为了降低建设成本,远洋海岛之间往往采用多个独立微电网互联的形式。在此模式下,各个微网既可以独立运行,也可以相互协调支撑,网间存在更复杂的功率、信息交互。从已有的研究来看,对于多微网系统的能量管理,传统多以总运行成本最低为优化目标,采用集中式方法进行优化调度。文献[6]考虑微网独立运行的情况,针对包含柔性负荷的微网系统进行能量管理;文献[7]只考虑微电网与大电网之间的能量交互,而未研究独立微电网之间的互济和微网内能源终端的协同调节;文献[8]针对光伏发电和风力发电混合发电的微网系统,研究微网内部各能源终端之间的能源供给分配策略。上述研究以独立微网为研究主体,研究微网独立运行或互联运行的单一模态,缺乏对微网多模态运行的综合考虑。

集中式方法对系统通信能力的要求较高,且忽视了能源终端的计算和优化能力,因此为了减轻通信负担,充分发挥能源终端的计算能力,分布式方法逐渐成为研究热点。文献[9]基于多智能体一致性理论,研究多微网间的能量优化调度和能源终端出力的按比例分配,但在该研究中多微网工作在互联模态下,未考虑微网独立运行与互联运行的切换问题。文献[10]针对可再生能源高比例渗透的家庭能源网络,提出一种分布式双层优化模型,通过日前优化和日内调整兼顾用户舒适度和系统经济性,实现网内能源终端的协同优化。文献[11]提出一种改进的粒子群算法,对微网内的电池进行实时的出力行为优化。由于多微网系统具有的高维度、不确定性等特点,上述分布式算法难以有效地处理具有复杂状态空间的优化问题,且这些算法泛化学习能力不足、计算耗时长,难以满足当前源荷随机性强的多微网系统实时优化要求。

随着人工智能技术的发展,作为人工智能重要分支之一的强化学习也受到研究者的广泛关注。强化学习中的智能体在动态环境中反复探索和试错,最终学习到令累计奖励最大的策略,对于处理强不确定性环境下的优化问题具有天然优势^[12-13]。近年来,强化学习在微网优化调度中也有较为广泛的应用。对于结构简单的微网系统,往往采用动作和状态空间离散的强化学习方法。文献[14]在微电网孤岛运行的工况下,采用模糊方法将电池等能源终端

的出力离散化,在此基础上提出了一种多智能体 Q 学习算法以实现微网的经济运行。该文献研究的微网系统结构较为简单,离散处理后状态空间和动作空间结合形成的决策空间小,因此,采用查表式的 Q 学习就能达到较好的优化结果。随着系统复杂度增加, Q 表的大小将急剧增加,查表式的 Q 学习效率较低,无法有效求解复杂系统的优化问题。为了解决决策空间较大的优化问题,文献[15]将深度神经网络与 Q 学习结合,利用深度神经网络拟合 Q 表,通过更新神经网络参数实现 Q 值的更新。然而,微网系统的状态空间往往是连续的,因此,一些状态空间连续的强化学习方法被应用于微网优化运行,比较典型的方法包括执行-评判算法(actor-critic algorithm, AC)^[16]、深度确定性策略梯度算法(deep deterministic policy gradient, DDPG)^[17]和近端策略优化算法(proximal policy optimization, PPO)^[18]。文献[19]提出一种结合深度卷积网络的集中式AC算法;文献[20]针对孤岛运行的微电网考虑不确定性提出一种分布式AC算法,实现了网内能源终端的协同优化;文献[21]提出一种结合深度神经网络的集中式DDPG方法,解决了包含不确定性的微网能量管理问题;文献[22]提出一种分布式的DDPG算法,解决了微网中各个储能单元的日前优化问题。A2C和DDPG算法都需要事先在环境中对神经网络进行长时间的离线训练,当系统的拓扑发生变化时,需要重新对算法进行训练,因此,这两种算法难以适应用户的即插即用行为,难以满足多微网系统实时优化的要求。与上述两种算法相比,PPO算法在实时性上具有较好的表现。文献[23]提出一种集中式的PPO算法,以实现实时的微网能量管理。然而,针对多微网系统,集中式的PPO算法存在对系统的通信和计算能力要求较高的缺陷。

为了发挥多微网能量互补的优势,同时充分利用微网内能源终端的计算能力,本文研究多微网系统的端-网实时协同优化,综合考虑多微网系统的网间能量互济和网内能源终端互相协同,以经济成本最低为目标,提出基于多智能体近端策略优化(MAPPO)的实时端-网协同优化策略。本文首先建立多微网系统端-网协同优化结构和优化模型;其次建立基于多智能体马尔科夫决策过程的强化学习模型;然后提出基于MAPPO算法的分布式端-网协同优化算法;最后针对3个典型场景验证算法的有效性和准确性。本文的贡献如下:

1) 为了降低功率互济中微网间通信负担,本文设

计了功率互济智能体.功率互济智能体存储所在微网的功率缺额信息,当进行功率互济时,仅由功率互济智能体进行通信,就能得到接近全局最优解的优化结果,降低了互济过程中的通信负担;

2)可自适应分布式电源出力和负荷的不确定性,不依赖于对源荷的准确预测,根据源荷的随机变化实时给出优化调度结果,考虑优化过程中功率不平衡约束难以满足的问题,设计一种新的参数可调的功率平衡反馈,能够克服传统的线性反馈信号快速趋向于零的缺点,更好地避免功率不平衡现象的出现,保障算法后期对于功率平衡约束的实现;

3)与状态完全观测和状态部分观测的策略相比,所提出的策略兼具高效性和稳定性,能够实现多微网系统的网间能量互济和网内能源终端协同,大幅提升系统整体的经济性和鲁棒性.

1 系统结构和双微网功率互济优化建模

1.1 端-网协同优化系统结构

端-网协同优化系统主要实现两个功能,即网内多终端协同和网间功率互济.以两个独立的微电网为例,构建了如图1所示的端-网协同优化系统.各个微电网通过发电机和风力发电对柔性负荷和固定负荷供电.

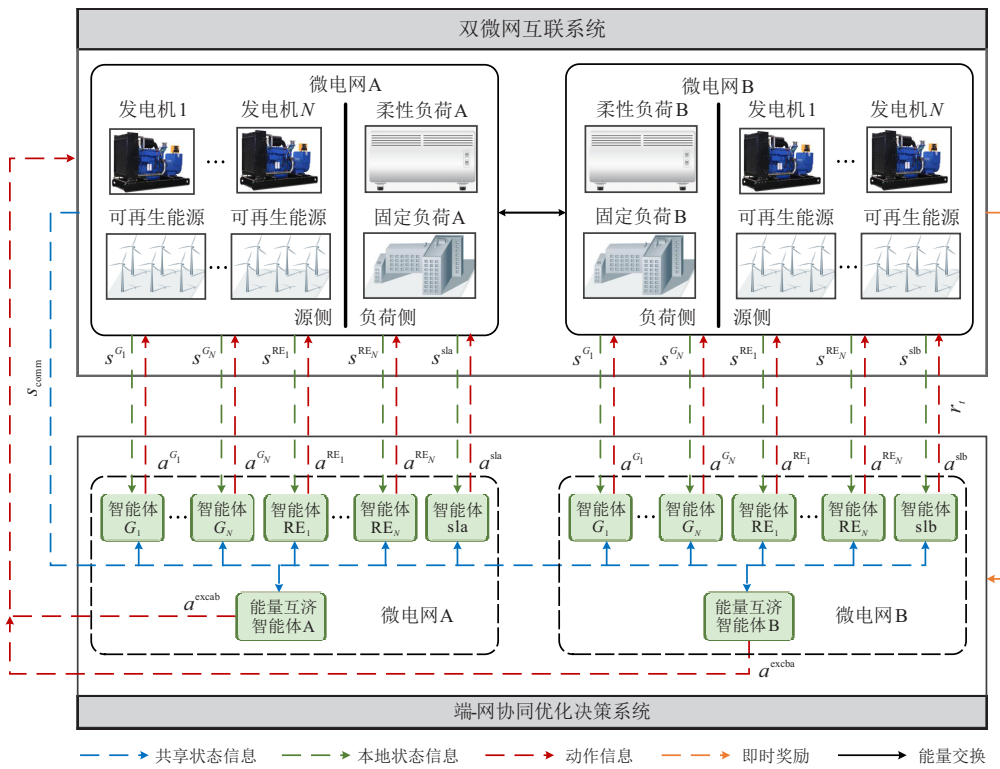


图1 端-网协同优化系统框图

1.2 端-网协同优化模型

1.2.1 目标函数

本系统以总运行成本最低为目标,进行实时端-网协同优化.成本函数为

$$F = \min \sum_{t=1}^N \sum_{\phi=WT,G} (\beta P_{\phi,t} + \gamma P_{\phi,t}^2) + \min \sum_{t=1}^N c_{pr,t} (s_a P_{exab,t} + s_b P_{exba,t}). \quad (1)$$

其中: $c_{pr,t}$ 是第 t 时刻的电价, $P_{exab,t}$ 是该时刻的微电网A向微电网B输送的功率值, $P_{exba,t}$ 是该时刻的微电网B向微电网A输送的功率值, $P_{\phi,t}$ 是风机和发电机的输出功率. s_a 和 s_b 分别是微电网A和微电网B的功率输送状态,当微电网存在功率盈余并向其他微

电网传输功率时, $s_i = 1$ ($i = a, b$);当微电网内发电设备供不应求,需要从其他微电网接收功率时, $s_i = 0$.电机和风机的发电成本用一个二次函数统一进行表示, β 和 γ 是对应的成本系数.

1.2.2 净功率

在研究多微电网互联的系统中,净功率对于判断微电网是否互联有着至关重要的作用. t 时段净功率是微网中的发电设备与用电负荷的差,表示如下:

$$\Delta P_a = \sum_{i=1}^{N_A} P_{G_i,t} - P_{nsla,t} - P_{sla,t}, \quad (2)$$

$$\Delta P_b = P_{WT,t} + \sum_{j=1}^{N_B} P_{G_j,t} - P_{nslb,t} - P_{slb,t}. \quad (3)$$

其中: $P_{G_i,t}$ 和 $P_{G_j,t}$ 分别代表微电网A和微电网B中发电机的出力, $P_{\text{nsla},t}$ 和 $P_{\text{nslb},t}$ 分别代表微电网A和微电网B中的固定负荷大小, $P_{\text{sla},t}$ 和 $P_{\text{sלב},t}$ 分别代表微电网A和微电网B中的柔性负荷大小, ΔP_a 和 ΔP_b 分别代表微电网A和微电网B在 t 时段的净功率. 当 $\Delta P > 0$ 时, 说明微电网此时处于超额状态, 系统内的发电设备不仅可以满足负荷的需求, 还有盈余. 需要注意的是, 当各个微电网均处于发电超额的状态且可再生能源发电量多于用户需求时, 将出现弃风现象. 当 $\Delta P < 0$ 时, 说明此时处于缺额状态, 需要通过调节增加系统内发电设备的发电量或通过功率互济来补足功率缺额.

1.2.3 约束条件

根据系统内设备供能和用能关系, 端-网协同优化的动态功率约束可以表示为

$$\begin{cases} \sum_{i=1}^{N_A} P_{G_i,t} - P_{\text{nsla},t} - P_{\text{sla},t} + \\ (s_a P_{\text{exab},t} + s_b P_{\text{exba},t}) \geq 0, \\ P_{\text{WT},t} + \sum_{j=1}^{N_B} P_{G_j,t} - P_{\text{nslb},t} - P_{\text{sלב},t} + \\ (s_a P_{\text{exab},t} + s_b P_{\text{exba},t}) \geq 0. \end{cases} \quad (4)$$

其中, 如果等号成立, 则代表此时微电网处于功率平衡状态.

设备上下限约束为

$$\begin{cases} P_{\text{WT}}^{\min} \leq P_{\text{WT},t} \leq P_{\text{WT}}^{\max}, \\ P_{G_i}^{\min} \leq P_{G_i,t} \leq P_{G_i}^{\max}, \\ P_{G_j}^{\min} \leq P_{G_j,t} \leq P_{G_j}^{\max}. \end{cases} \quad (5)$$

其中: P_{WT}^{\min} 和 P_{WT}^{\max} 分别是风机的功率上下限约束, $P_{G_i}^{\min}$ 、 $P_{G_j}^{\min}$ 和 $P_{G_i}^{\max}$ 、 $P_{G_j}^{\max}$ 分别是发电机的功率上下限约束.

2 基于MAPPO的端-网协同优化策略

2.1 多智能体马尔科夫决策过程

基于MAPPO的端-网协同优化可以用多智能体马尔科夫决策过程 (multi-agent Markov decision process, MAMDP) 来表示, 该过程通常可以概括为一个元组, 即 $\langle M, \{S^m\}, \{A^m\}, \mathcal{P}, \mathcal{R} \rangle$. 其中: M 是智能体的数量; S^m 是第 m 个智能体的状态集合, $s_t^m \in S^m$ 表示 t 时段该智能体的状态, 对所有智能体状态组成的联合状态有 $S^1 \times S^2 \times \dots \times S^m = \mathcal{S}$, 且 $s_t \in \mathcal{S}$; A^m 是第 m 个智能体的动作集合, $a_t^m \in A^m$ 表示 t 时段该智能体的动作, 对所有智能体动作组成的联合动作有

$A^1 \times A^2 \times \dots \times A^m = \mathcal{A}$, 且 $a_t \in \mathcal{A}$; \mathcal{P} 为状态转移概率, 满足 $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$; \mathcal{R} 是奖励函数, 表示在状态 s_t 下采用动作 a_t 时, 从系统环境中获得的即时反馈 r_t , 有 $r_t \in \mathcal{R}$. 各智能体根据观测到的状态, 采取动作 a_t 作用于环境, 使得环境状态由 s_t 转移至 $s(t+1)$, 并获得反馈 r_t , 如此循环交替, 通过与环境的不断互动最终形成策略 $\pi : \mathcal{S} \rightarrow \mathcal{A}$, 且 $a \sim \pi(a|s)$.

2.2 基于MAPPO的优化模型

根据MAMDP的基本原理, 将能量互济优化模型转化为基于MAPPO的优化模型, 该模型的状态空间、动作空间和即时奖励函数定义如下.

2.2.1 状态空间

状态空间是智能体所观测到的环境信息. 根据图1所示的端-网系统优化结构, 状态空间的定义可以分为两部分, 即公共的观测状态和各区域专用的观测状态. 公共的观测状态包括各微电网的净功率以及是否参与能量互济的状态信息, 定义如下:

$$s_{\text{comm},t} = (t, \Delta P_{a,t}, \Delta P_{b,t}, s_{\text{part},t}^a, s_{\text{part},t}^b). \quad (6)$$

其中: $\Delta P_{a,t}$ 和 $\Delta P_{b,t}$ 分别为 t 时段微电网A和微电网B的净功率; $s_{\text{part},t}^a$ 和 $s_{\text{part},t}^b$ 分别为能量互济的参与状态, 若微网参与互济, 则该状态为1, 否则为0.

微电网A各个发电机智能体的状态空间可以表示为 $s_t^{G_i} = (P_{G_i,t}, s_{\text{comm},t})$; 柔性负荷智能体的状态空间可以表示为 $s_t^{\text{sla}} = (P_{\text{sla},t}, s_{\text{comm},t})$; 能量互济智能体的状态空间可以表示为 $s_t^{\text{excab}} = (s_{\text{comm},t})$. 类似地, 微电网B中各个发电机智能体、柔性负荷智能体和能量互济智能体的状态空间可以分别表示为 $s_t^{G_j} = (P_{G_j,t}, s_{\text{comm},t})$, $s_t^{\text{WT}} = (P_{\text{WT},t}, s_{\text{comm},t})$ 和 $s_t^{\text{exba}} = (s_{\text{comm},t})$.

2.2.2 动作空间

动作空间为相关决策变量, 每个区域智能体的动作空间为其控制区域内可控设备的出力. 微电网A中各个发电机智能体的动作空间为 $a_t^{G_i} = (\Delta P_{G_i,t})$; 柔性负荷智能体的动作空间为 $a_t^{\text{sla}} = (\Delta P_{\text{sla},t})$; 能量互济智能体的动作空间为 $a_t^{\text{excab}} = (P_{\text{exab},t})$. 同样地, 微电网B中发电机智能体、柔性负荷智能体和能量互济智能体的动作空间可以分别表示为 $a_t^{G_j} = (\Delta P_{G_j,t})$, $a_t^{\text{sלב}} = (\Delta P_{\text{sלב},t})$ 和 $a_t^{\text{exba}} = (P_{\text{exba},t})$.

2.2.3 奖励函数

多智能体强化学习的训练目标为寻找最优策略使得累积回报期望值最大, 即

$$R_t = \sum_{k=t}^T \gamma^{k-t} r_k. \quad (7)$$

其中: T 为总优化时长, r_k 为第 k 个时刻智能体获得的即时反馈. 本文设计的及时奖励包括以下几个部分: 运行成本项、调节柔性负荷导致用户满意度下降的惩罚项、弃风惩罚项和功率不平衡惩罚项.

1) 运行成本奖励函数为

$$r_{\text{cost},t} = \Delta F. \quad (8)$$

2) 用户满意度惩罚函数为

$$\begin{cases} r_{\text{sla},t} = -\varphi_{\text{disa}}(\Delta P_{\text{sla},t})^2, \\ r_{\text{slb},t} = -\varphi_{\text{disb}}(\Delta P_{\text{slb},t})^2, \end{cases} \quad (9)$$

其中 φ_{disa} 和 φ_{disb} 分别为微电网 A 和微电网 B 中用户满意度惩罚系数. 该项用二次函数表示, 随着柔性负荷削减量的增加, 用户满意度将大幅度下降, 与此同时该惩罚函数带来的负反馈也大大增加. 通过设计该惩罚函数, 可以避免优化过程中对柔性负荷的过度削减, 保证用户的使用体验.

3) 弃风惩罚函数为

$$r_{\text{abd},t} = -\varphi_{\text{abd}}(P_{\text{WT},t} - P_{\text{WT},t}^{\text{max}})^2, \quad (10)$$

其中 φ_{abd} 为弃风惩罚系数. 随着弃风量的增加, 该项负反馈也相应增加, 从而使得优化算法避免弃风行为, 提高可再生能源的利用率.

4) 功率不平衡惩罚函数.

在实时的能量互济中, 需要保证各区域的功率平衡. 传统的功率不平衡惩罚函数往往呈线性, 这使得在优化进入后期时, 功率不平衡惩罚值快速趋向于 0, 与其他反馈信号相比, 该项对于智能体寻优的影响变得微乎其微, 从而导致功率不平衡现象时有发生. 因此针对实时调节中功率不平衡的问题, 本文设计了如下参数可调的惩罚函数:

$$\begin{cases} r_{\text{pba},t} = -\varphi_{\text{pba}}(1 + c_1|\Delta P_{\text{imba}}|) \cdot |\Delta P_{\text{imba}}|, \\ r_{\text{pbb},t} = -\varphi_{\text{pbb}}(1 + c_2|\Delta P_{\text{imbb}}|) \cdot |\Delta P_{\text{imbb}}|. \end{cases} \quad (11)$$

其中: φ_{pba} 和 φ_{pbb} 分别为微电网 A 和微电网 B 的功率不平衡惩罚系数; c_1 和 c_2 分别控制惩罚函数的变化趋势, 通过选取合理的参数值, 既能加快算法在初期的收敛速度, 避免算法在功率差额较大的区域持续探索, 又能在算法进入优化后期时, 可以放大功率不平衡这一因素对智能体寻优的影响, 引导智能体向消除功率不平衡的方向移动.

5) 总奖励函数.

根据上述奖励函数定义, 各微网的总奖励函数定义如下:

$$\begin{cases} r_{a,t} = r_{\text{cost},t} + r_{\text{sla},t} + r_{\text{pba},t}, \\ r_{b,t} = r_{\text{cost},t} + r_{\text{slb},t} + r_{\text{abd},t} + r_{\text{pbb},t}. \end{cases} \quad (12)$$

2.3 基于MAPPO的端-网协同算法

MAPPO 算法采用策略神经网络来近似优化策略, 用评判网络来评价优化策略的好坏. 第 i 个智能体的策略网络参数和评判网络参数分别表示为 θ_i^π 和 θ_i^V . 其中, 状态 s 下评判网络的值函数定义为

$$V(s) = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s\right], \quad (13)$$

γ^k 为折扣因子. 值函数表示状态 s 下可获取累积回报的期望值. 基于值函数, 可以得到第 i 个智能体的优势函数为

$$A_i(s_i, a_i) = E\left[\sum_{t=0}^{\infty} \gamma^t r_t | s_i = s_0, a_i = a_0\right] - E\left[\sum_{t=0}^{\infty} \gamma^t r_t | s_i = s_0, a_i = a\right]. \quad (14)$$

根据式(14)可知, 优势函数的第 1 项反映了在 s_0 状态下采取动作 a_0 得到的实际累计回报的期望值, 第 2 项反映了该状态下所有采集过的动作 a 的平均累计回报的期望值. 通过将这两项相减, 所得的结果便可用来衡量当前采用动作相对于历史动作的优劣.

优势函数可以近似地表示为

$$\hat{A}_{it} = -V_i(s_{iT}) + r_t + \gamma r_{t+1} + \dots + \gamma^{T-t-1} r_{T-1} + \gamma^{T-t} V_i(s_{iT}). \quad (15)$$

将未更新的策略网络与环境互动获得的采样数据(即迹 τ_i) 存储于经验池, 可以多次用于更新策略网络, 以提高数据的利用率, 表示为 $\tau_i = (s_{i1}, a_{i1}, r_1, s_{i2}, a_{i2}, r_2, \dots, s_{iT}, a_{iT}, r_T)$.

在训练过程中, 策略网络参数和评判网络参数每一步都需要进行更新, 然而这种更新方式使得迹 τ_i 非常难以获取. 为了解决该问题, 额外设置了一个参数为 $\theta_i^{\pi'}$ 的神经网络, 该网络每一回合进行一次更新以实现迹的存储.

2.3.1 策略网络更新规则

在策略网络的更新中, 需要限制新策略和旧策略的差距, 避免优化策略的突变. 采用如下替代函数对网络参数进行更新:

$$L_{i,t}^{\text{clip}}(\theta_i^\pi) = E[\min(\mu_t(\theta_i) \hat{A}'_{i,t}, \text{clip}(\mu_t(\theta_i), 1 - \varepsilon, 1 + \varepsilon)) \hat{A}'_{i,t}]. \quad (16)$$

$$\text{clip}(\mu_t(\theta_i), 1 - \varepsilon, 1 + \varepsilon) = \begin{cases} 1 - \varepsilon, & \mu_t(\theta_i) < 1 - \varepsilon; \\ \mu_t(\theta_i), & 1 - \varepsilon \leq \mu_t(\theta_i) \leq 1 + \varepsilon; \\ 1 + \varepsilon, & \mu_t(\theta_i) > 1 + \varepsilon. \end{cases} \quad (17)$$

其中: $L_{i,t}^{\text{clip}}(\theta_i^\pi)$ 是替代函数; $\mu_t(\theta_i)$ 为概率比, 反映了新旧策略的相似度; ε 是裁剪参数, 用来控制概率比的取值范围. 概率比的定义如下:

$$\mu_t(\theta_i) = \frac{\pi_{\theta_i^\pi}(a_{it}|s_{it})}{\pi_{\theta_i^{\pi'}}(a_{it}|s_{it})}. \quad (18)$$

在此基础上, 策略网络的参数更新策略为

$$\theta_i^\pi = \theta_i^\pi + \ell_i^\pi \nabla_{\theta_i^\pi} L_{i,t}^{\text{clip}}(\theta_i^\pi), \quad (19)$$

其中 ℓ_i^π 是动作网络的学习因子.

2.3.2 评判网络更新规则

评判网络参数的更新也采用梯度下降的方式, 规则如下:

$$L(\theta_i^V) = E[(r_t + \gamma V_i(s_{i,t+1}) - V_i(s_{it}))^2], \quad (20)$$

$$\theta_i^V = \theta_i^V + \ell_i^V \nabla_{\theta_i^V} L(\theta_i^V). \quad (21)$$

根据 MAPPO 算法的更新规则, 可以得到端-网协同优化的流程, 如图2所示.

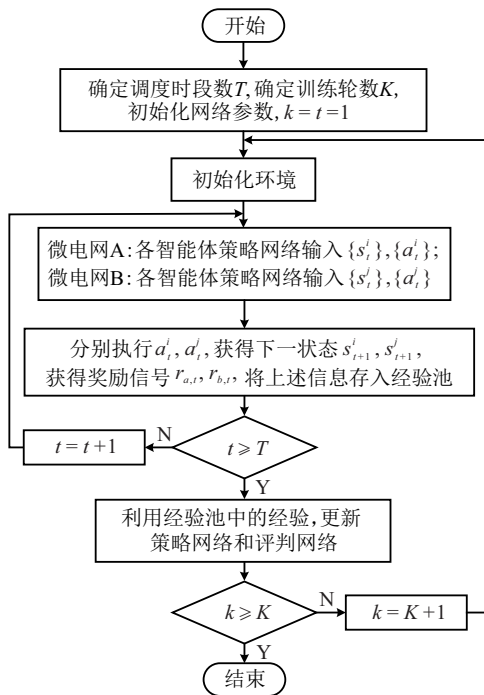


图2 端-网协同优化流程

3 仿真与分析

为验证所提出算法的有效性, 以海岛双微网系统为例进行仿真研究. 各个设备的成本参数以及功率上下限如表1所示.

表1 设备成本参数及功率上下限

设备名称	β	γ	下限/kW	上限/kW	
微电网A	发电机 G_1	53.296	0.765	45	200
	发电机 G_2	50.768	0.871	40	160
微电网B	发电机 G_3	45.340	1.018	35	150
	发电机 G_4	38.216	1.254	30	120
	风机	110	0	0	20~130

3.1 算法收敛性分析

算法的收敛过程如图3所示, 统计单个优化时段的所有区域智能体总奖励的平均值, 得到奖励曲线. 可以看出: 训练开始时, 智能体获得的全局奖励变化剧烈, 因为此时智能体对动作不断进行探索; 随着训练的进行, 智能体获得的全局奖励值收敛, 模型的表现效果越来越好, 所得的奖励曲线逐渐趋向平稳, 在第22个回合时, 算法达到收敛的稳定状态.

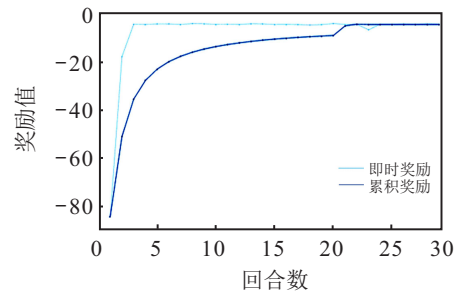


图3 MAPPO算法收敛曲线

3.2 多场景端-网协同优化仿真

3.2.1 场景1: 微网独立运行的网内终端协同优化

夜间风力发电机的发电量充足, 且用户负荷需求处于低谷, 风力发电能够满足微电网B的负荷需求, 此时微电网A的发电设备上限之和大于负荷需求. 因为微电网A和微电网B都能满足各自需求, 所以微电网A和微电网B独立运行, 无需进行能量互济, 只进行网内的能源终端协同优化. 优化结果如表2所示.

表2 场景1的优化结果 单位: kW

设备及负荷	调节区间	独立优化后	互济功率
微电网A	G_1	45~200	158
	G_2	40~160	122
	柔性负荷	80~130	120
	固定负荷		160
微电网B	G_3	35~150	123
	G_4	30~120	80
	风机	0~130	117
	柔性负荷	100~150	140
固定负荷		180	

从表2结果可以看出, 由于微电网内部存在柔性负荷, 优化算法通过适当调减柔性负荷需求, 在用户满意度和运行成本之间进行权衡, 便可实现微电网独立运行状态下的网内终端协同优化运行.

3.2.2 场景2: 风力发电不足情况下网间能量互济

日间风力发电量减少, 但用户需求相较于夜间大大增加, 该场景下微电网B的所有发电设备发电能力均达到上限, 柔性负荷可调节能力达到下限后, 仍不能满足用户需求, 此时需要通过能量互济, 从微电网

A 获取电能以实现功率平衡. 场景 2 的优化结果如表 3 所示.

		设备及负荷	调节区间	互济优化后	互济功率
微电网 A	G_1		45 ~ 200	176	输出 58
	G_2		40 ~ 160	141	
	柔性负荷		80 ~ 130	89	
	固定负荷		170		
微电网 B	G_3		35 ~ 150	135	获得 58
	G_4		30 ~ 120	113	
	风机		0 ~ 20	18	
	柔性负荷		100 ~ 150	114	
	固定负荷		210		

从表3的结果可以看出,通过能量互济,微电网 A 向微电网 B 传送 58 kW 功率,与此同时,微电网 B 内部各终端设备(发电机 G_3 、 G_4 和柔性负荷)互相协同以满足微电网 B 的用户需求.

3.2.3 场景3: 发电设备异常工况下网间能量互济

考虑发电机退出或故障等情况,假设微电网 A 中的发电机 G_2 退出,微电网 A 的其发电设备发电能力均达到上限,柔性负荷可调节能力达到下限后,仍不能满足用户需求,此时需要通过能量互济,从微电网

B 获取电能以实现功率平衡. 场景 3 的优化结果如表 4 所示.

		设备及负荷	调节区间	互济优化后	互济功率
微电网 A	G_1		45 ~ 200	190	获得 50
	G_2		—	—	
	柔性负荷		80 ~ 130	80	
	固定负荷		160		
微电网 B	G_3		35 ~ 150	148	输出 50
	G_4		30 ~ 120	104	
	风机		0 ~ 130	119	
	柔性负荷		100 ~ 150	141	
	固定负荷		180		

从表4的结果可以看出,通过能量互济,微电网 B 向微电网 A 传送 50 kW 功率,在此基础上,微电网 A 内部发电机 G_1 和柔性负荷共同调节,满足了微电网 A 的用户需求,并且微电网 B 的风能也得到了较充分的应用.

3.3 算法准确性验证

为验证所提出方法优化结果的准确性,将该方法与传统优化方法进行对比. 上述 3 个场景的结果对比如图 4 所示.

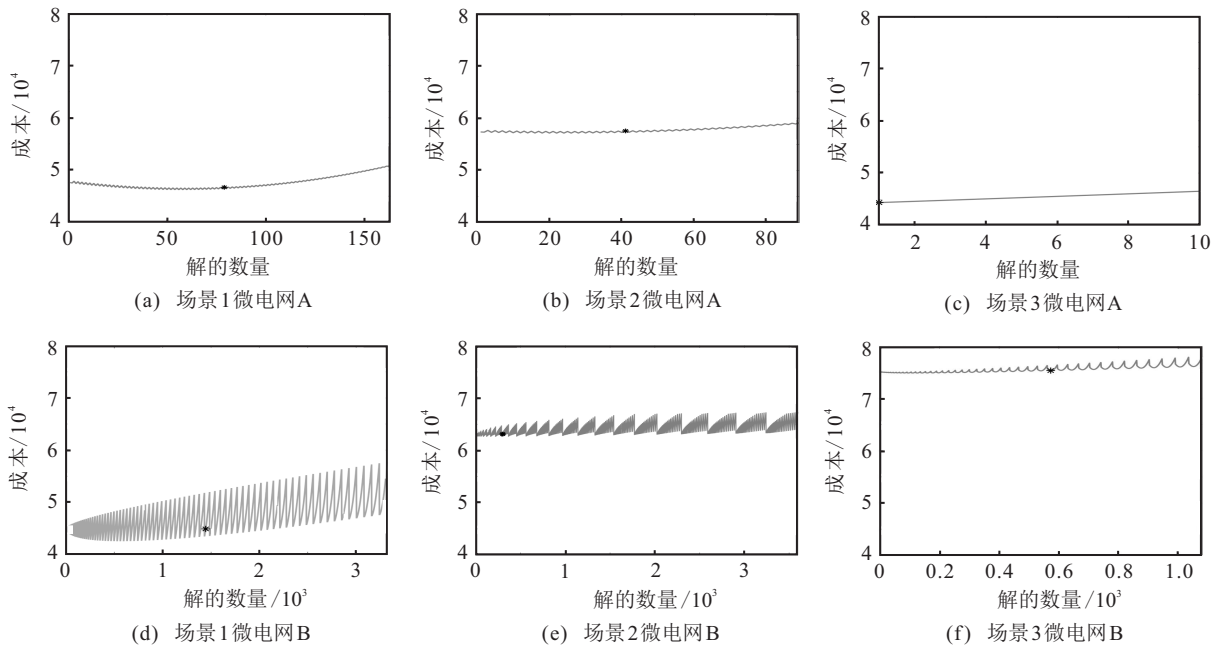


图 4 优化结果准确性验证

图 4 中: 灰色的范围为优化所有可行解对应的成本分布情况, 黑色的点对应本文提出的优化算法得到的结果在有可能解中所处的位置. 黑点分布的位置越低, 说明得到的优化策略对应的系统运行成本越低, 即越接近全局最优解. 从图 4 中可以看出本文算法得到的结果比较接近全局最优解, 对于场景 3, 微电网 A 的优化较为简单, 可以获得全局最优解. 本文所

得的优化结果与最优解的比值如表 5 所示.

表5 MAPPO 优化结果与最优解的比值

	场景 1	场景 2	场景 3
微电网 A	1.00094	1.0073	1
微电网 B	1.0554	1.0274	1.0450

由表5可知,由于MAPPO算法的寻优过程具有一定的随机性,得到的优化结果往往会在一定范围内波动.本文提出的算法与全局最优解的平均偏差为2.41%,说明本文算法所得的优化结果准确性较高.

3.4 对比分析

将本文提出的MAPPO算法与以下两种方法进行对比,即状态完全观测的MAPPO算法和状态部分观测的MAPPO算法.

3.4.1 方法1: 状态完全观测的MAPPO算法

对于状态完全观测的MAPPO算法,各个智能体的观测量相同,其状态空间可以表示为 $s = (P_{G_k,t}, s_{comm,t}, P_{sla,t}, P_{slb,t}, P_{WT,t}) (k = 1, 2, \dots, 5)$. 采用该算法进行优化的收敛曲线如图5所示.

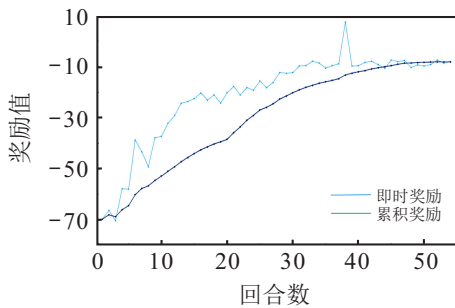


图5 状态完全观测MAPPO算法收敛曲线

由图5可知,状态完全观测的MAPPO算法大约经过50个回合的训练才可以收敛.从算法效率的角度来看,与如图3中本文所提出算法经过22回合收敛相比,状态完全观测的MAPPO算法效率较低,当系统对实时性的要求较高时,这种方法不宜使用.

3.4.2 方法2: 状态部分观测的MAPPO算法

对于状态部分观测的MAPPO算法,假设A、B两个微电网的智能体之间的通信拓扑如图6所示.根据图6所示的通信拓扑,可以得到部分观测MAPPO算法中各个智能体的状态空间,如表6所示.

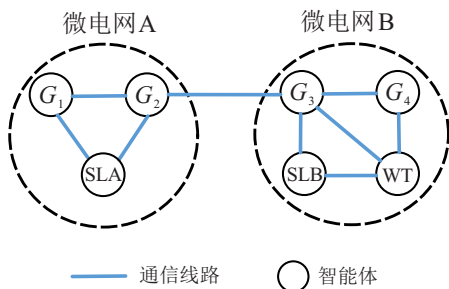
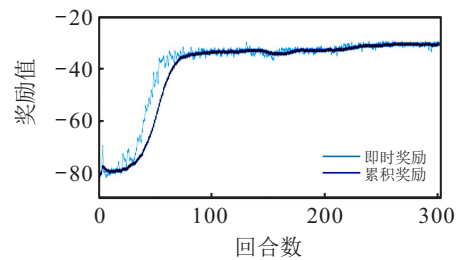


图6 双微电网系统通信拓扑

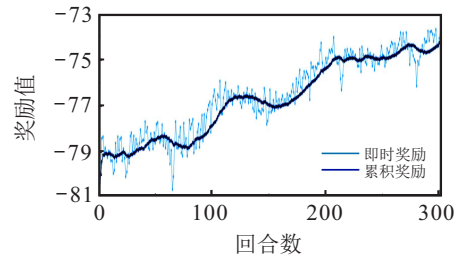
如表6所示,除智能体 G_2 和 G_3 外,其余智能体只能观测到所在微网的智能体状态信息,而无法获得相邻微网内智能体的状态信息.该算法的收敛情况如图7所示.

表6 部分观测MAPPO算法的状态空间

智能体状态空间	
微电网A	$s_t^{G_1} = (P_{G_1,t}, P_{G_2,t}, P_{sla,t}, s_{comm,t})$
	$s_t^{G_2} = (P_{G_1,t}, P_{G_2,t}, P_{G_3,t}, P_{sla,t}, s_{comm,t})$
	$s_t^{sla} = (P_{G_1,t}, P_{G_2,t}, P_{sla,t}, s_{comm,t})$
微电网B	$s_t^{G_3} = (P_{G_2,t}, P_{G_3,t}, P_{G_4,t}, P_{slb,t}, P_{WT,t}, s_{comm,t})$
	$s_t^{G_4} = (P_{G_3,t}, P_{G_4,t}, P_{WT,t}, s_{comm,t})$
	$s_t^{slb} = (P_{G_3,t}, P_{slb,t}, P_{WT,t}, s_{comm,t})$
	$s_t^{WT} = (P_{G_3,t}, P_{G_4,t}, P_{slb,t}, P_{WT,t}, s_{comm,t})$



(a) 微电网A



(b) 微电网B

图7 状态部分观测MAPPO算法收敛曲线

图7中,采用状态部分观测的MAPPO算法时,微电网A和微电网B中智能体是不收敛的.值得注意的是,尽管微电网A中的智能体经过约90个回合后,奖励值趋于某一值,但是微电网A中的智能体并非真正达到收敛状态.此时微电网A的总发电量与总负荷严重不平衡,而且供需两侧的功率差是大幅变化的,只是每个回合的平均奖励值趋于某一值而已.由此可见,虽然采用状态部分观测的策略能够在一定程度上保护各微网用户和设备发电量等隐私信息,但这种方法的收敛性无法得到保证.综上所述,与上述两种方法相比,本文通过设计能量互济智能体,具有良好的寻优效率,能够满足系统的实时性要求.

4 结论

为了充分发挥多微网系统的能源终端的独立运行能力和功率支撑优势,本文针对多微网系统研究了端-网协同优化问题.考虑多微网系统源荷不确定性和设备即插即用导致的拓扑变化,提出了基于MAPPO方法的实时优化方法.该方法具有在线训练和在线优化的特点,能够自适应分布式电源出力和负荷的不确定性,不依赖于对源荷的准确预测,根据源荷的随机变化实时给出优化调度结果.在此基础上

上,进一步考虑功率平衡约束,区别于传统的线性反馈信号,所设计的功率平衡反馈信号能够有效避免功率不平衡情况的出现,提高了可再生能源利用率。此外,设计了能源互济智能体用于交换功率盈缺信息,各微网仅基于本区域信息的收集和相邻区域之间的功率缺额信息,便能得到接近全局最优解的优化结果。将所提出的方法与状态完全观测和状态部分观测的MAPPO算法进行对比,结果显示,所提出的方法具有良好的寻优效率。

参考文献(References)

- [1] Zhou T R, Kang C Q, Chen X Y, et al. Evaluating low-carbon effects of demand response from smart distribution grid[C]. The 3rd IEEE PES Innovative Smart Grid Technologies Europe. Berlin, 2012: 1-6.
- [2] Cheng Y H, Zhang N, Lu Z X, et al. Planning multiple energy systems toward low-carbon society: A decentralized approach[J]. IEEE Transactions on Smart Grid, 2019, 10(5): 4859-4869.
- [3] Xin B A. Contribute wisdom and strength to achieve the goal of “peak carbon dioxide emissions, carbon neutrality” [J]. State Grid, 2021(3): 2-3.
- [4] Ufa R A, Malkova Y Y, Rudnik V E, et al. A review on distributed generation impacts on electric power system[J]. International Journal of Hydrogen Energy, 2022, 47(47): 20347-20361.
- [5] Ji Y, Wang J H. Online optimal scheduling of a microgrid based on deep reinforcement learning[J]. Control and Decision, 2022, 37(7): 1675-1684.
- [6] Igualada L, Corchero C, Cruz-Zambrano M, et al. Optimal energy management for a residential microgrid including a vehicle-to-grid system[J]. IEEE Transactions on Smart Grid, 2014, 5(4): 2163-2172.
- [7] Shen J S, Jiang C W, Liu Y Y, et al. A microgrid energy management system with demand response for providing grid peak shaving[J]. Electric Power Components and Systems, 2016, 44(8): 843-852.
- [8] Wang R, Ma D Z, Li M J, et al. Accurate current sharing and voltage regulation in hybrid wind/solar systems: An adaptive dynamic programming approach[J]. IEEE Transactions on Consumer Electronics, 2022, 68(3): 261-272.
- [9] Sun Q Y, Fan R Y, Li Y S, et al. A distributed double-consensus algorithm for residential we-energy[J]. IEEE Transactions on Industrial Informatics, 2019: 15(8): 4830-4842.
- [10] Mansouri S A, Ahmarinejad A, Nematbakhsh E, et al. A sustainable framework for multi-microgrids energy management in automated distribution network by considering smart homes and high penetration of renewable energy resources[J]. Energy, 2022, 245: 123228.
- [11] Hossain M A, Pota H R, Squartini S, et al. Modified PSO algorithm for real-time energy management in grid-connected microgrids[J]. Renewable Energy, 2019, 136: 746-757.
- [12] Kiran B R, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: A survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 4909-4926.
- [13] Kuznetsova E, Li Y F, Ruiz C, et al. Reinforcement learning for microgrid energy management[J]. Energy, 2013, 59: 133-146.
- [14] Kofinas P, Dounis A I, Vouros G A. Fuzzy Q-learning for multi-agent decentralized energy management in microgrids[J]. Applied Energy, 2018, 219: 53-67.
- [15] Liu Y K, Zhang D X, Hoay B G. Optimization strategy based on deep reinforcement learning for home energy management[J]. CSEE Journal of Power and Energy Systems, 2020, 6(3): 572-582.
- [16] Wang Z H, Zhang Y X, Huang Z Q, et al. Multi-agent collaboration based on RGMAAC algorithm under partial observability[J]. Control and Decision, 2023, 38(5): 1267-1277.
- [17] Zheng K C, Jia X L, Chi K K, et al. DDPG-based joint time and energy management in ambient backscatter-assisted hybrid underlay CRNs[J]. IEEE Transactions on Communications, 2023, 71(1): 441-456.
- [18] Soleimanzade M A, Kumar A, Sadrzadeh M. Novel data-driven energy management of a hybrid photovoltaic-reverse osmosis desalination system using deep reinforcement learning[J]. Applied Energy, 2022, 317: 119184.
- [19] Qin Z M, Liu D, Hua H C, et al. Privacy preserving load control of residential microgrid via deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2021, 12(5): 4079-4089.
- [20] Jendoubi I, Bouffard F. Multi-agent hierarchical reinforcement learning for energy management[J]. Applied Energy, 2023, 332: 120500.
- [21] Kaewdornhan N, Chatthaworn R. Model-free data-driven approach assisted deep reinforcement learning for optimal energy management in microgrid[J]. Energy Reports, 2023, 9: 850-858.
- [22] Harrold D J B, Cao J, Fan Z. Renewable energy integration and microgrid energy trading using multi-agent deep reinforcement learning[J]. Applied Energy, 2022, 318: 119151.
- [23] Guo C Y, Wang X, Zheng Y H, et al. Real-time optimal management of microgrid with uncertainties based on deep reinforcement learning[J]. Energy, 2022, 238: 121873.

作者简介

王丹璐(1994—),女,博士生,从事能源系统智能优化运行的研究,E-mail: danlu_w@163.com;

孙秋野(1977—),男,教授,博士生导师,从事能源互联网的建模与优化运行、分布式多能源综合互补优化等研究,E-mail: sunqiuye@ise.neu.edu.cn;

苏涵光(1991—),男,副教授,博士,从事综合能源系统优化控制的研究,E-mail: suhanguang@ise.neu.edu.cn.