

# 基于分布式联邦强化学习的多区域综合能源系统优化调度

朱新文,王家奇,李生炜,林文杰,吴祥,郭方洪

(浙江工业大学 信息工程学院,浙江 杭州 310023)

**摘要:**针对传统优化方法及集中式联邦强化学习在隐私保护和计算效率方面存在的局限性,提出一种基于分布式联邦强化学习的多区域综合能源系统优化调度方法。每个区域综合能源系统由单独智能体管理,各智能体通过双延迟确定性策略梯度算法优化本地Critic网络的参数,并与邻域智能体进行参数信息交互,无需额外的中央服务器即可高效管理能量调度。为了保证全局最优,参数交互的权重系数由双随机矩阵元素确定。算例分析结果表明,所提方法能在增强对综合能源系统隐私保护的同时,展现出良好的收敛性能,有效降低了运营成本。

**关键词:**综合能源系统;优化调度;分布式联邦强化学习;智能体;双延迟确定性策略梯度算法

**中图分类号:**TM73;TK01

**文献标志码:**A

**DOI:**10.16081/j.epae.202601006

## 0 引言

在全球能源需求增长和能源安全问题日益严峻的背景下,综合能源系统(integrated energy system, IES)因具有多能互补和能源利用效率高的特点,逐渐成为现代能源系统发展的重要方向<sup>[1]</sup>。IES通过整合多种能源类型,能够实现资源的高效配置,多区域IES互联则可实现资源共享和网间互补<sup>[2]</sup>。但随着多区域IES规模的扩大,其资源特性和需求波动带来了显著的调度复杂性,使优化管理面临诸多挑战<sup>[3-4]</sup>。

传统基于数学模型的数学优化方法(如非线性规划<sup>[5]</sup>、交替方向乘子法<sup>[6]</sup>、鲁棒优化<sup>[7]</sup>)在早期的IES调度问题中得到了广泛的应用。这类方法通常通过构建系统的物理模型和数学约束,以最小化成本为目标进行全局优化。然而,当应对多区域IES的复杂动态环境时,这些方法难以满足实时优化的要求<sup>[8-9]</sup>。

在此背景下,强化学习技术凭借其动态环境的适应性和高效性被认为是未来调度优化的核心趋势<sup>[10-11]</sup>。当采用强化学习算法解决多区域IES日内实时调度优化问题时,通过学习历史数据,并与实际运行环境进行实时交互,选择最优的能量分配和调度策略。基于强化学习算法,智能体能够动态地适应能源需求波动、可再生能源出力等不确定性因素<sup>[12]</sup>。文献[13]将氢耦合电-热IES的优化调度问题转化为马尔可夫决策过程,采用单智能体算法在连续状态-动作空间中学习调度策略。文献[14]提

出了一种基于多智能体深度Q网络的分布式优化调度方法,基于数据驱动的强化学习算法实现了园区IES的经济调度。文献[15]提出了一种基于目标值竞争机制的多智能体强化学习算法,解决了分布式非凸经济调度问题,实现了智能体之间的协调,并确保算法的收敛性。然而,多智能体强化学习算法虽然在一定程度上提升了系统的灵活性和分布性,但由于智能体之间的交互仍需要涉及各区域的部分状态信息和策略信息传输,仍存在数据隐私问题<sup>[16]</sup>。这是因为多智能体强化学习算法需要将各区域的敏感数据(如能耗数据、调度策略等详细信息)集中于1个经验池进行统一学习和优化,这种数据汇集方式虽然提升了系统的优化效果,但也不可避免地会产生数据隐私风险。

针对上述数据隐私问题,近年来联邦学习框架被应用于强化学习中,并已得到了广泛的研究,被称为联邦强化学习<sup>[17]</sup>(federated reinforcement learning, FRL)。FRL允许各区域智能体在本地环境中训练独立模型,各智能体仅需向中央服务器发送模型参数,而不传递非敏感数据,很好地弥补了强化学习在隐私安全方面的缺陷<sup>[18-19]</sup>。文献[20]构建了基于FRL的分布式微电网协同优化框架。针对基于分时电价的多住宅能量调度问题,文献[21]基于集中式联邦强化学习(centralized FRL, CFRL)方法实现了协作式分布式能量管理,有效降低了总能耗成本。但对于规模较大的分布式IES,如多个园区协作或区域型IES,随着节点数量增加,CFRL方法由于依赖中央服务器作为聚合节点,其通信和计算负担显著增加,容易导致通信效率降低<sup>[22]</sup>。此外,CFRL涉及数据集中化处理,不仅增加了隐私泄露风险,还使得系统面临更严重的单点故障问题,这在实际部署中可能成为关键瓶颈。

为了解决多区域IES调度中的数据隐私和实时

收稿日期:2024-11-19;修回日期:2025-04-24

在线出版日期:2026-01-29

基金项目:国家自然科学基金资助项目(62373328)

Project supported by the National Natural Science Foundation of China(62373328)

响应问题,本文提出一种分布式联邦强化学习(distributed FRL, Dis-FRL)方法。这种分布式框架使得各区域智能体无须依赖中央服务器,即可在本地环境中独立优化,从而避免了数据集中存储和共享。同时,该方法引入双随机矩阵机制,根据邻域特征调整模型参数的交互权重,并结合离线训练和在线学习的优势,不仅降低了计算复杂度,还能够快速适应系统的动态变化和不确定性,满足日内实时调度需求。

## 1 IES 优化调度模型

本文所考虑的 IES 包括柴油发电机、冷热电联供(combined cooling, heating and power, CCHP)机组、电锅炉、电制冷机、电储能(electric energy storage, EES)这5类关键运行设备,均由所在区域的能源管理系统(energy management system, EMS)进行管理。EMS 与主电网、天然气网络和光伏发电共同构成一个复杂的综合能源网络。单个区域 IES 的内部能量流向如图1所示。

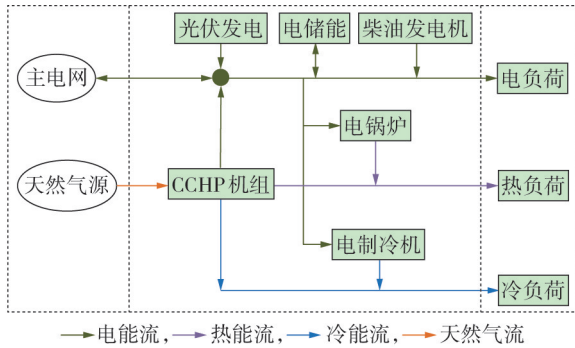


图1 单个区域 IES 的内部能量流向

Fig.1 Internal energy flow within a single regional IES

### 1.1 多区域 IES 的交易策略

为了增强多区域 IES 之间的互联性和协作性,本文引入一种基于合作博弈理论的策略<sup>[23]</sup>,在多区域 IES 间促进协作并最大化整体效益。各区域 IES 优先使用本地能源,当本地供能不足以满足需求时,优先与其他区域 IES 进行能源交换,然后与主电网进行交易,从而降低对主电网的依赖性。多区域 IES 的系统结构见附录 A 图 A1。受文献[24]的启发,本文提出多区域 IES 的交易策略。

$t$ 时刻第  $i$  个区域 IES 的目标交易功率  $p_{\text{target},i}^t$  为:

$$p_{\text{target},i}^t = p_{g,i}^t + \sum_{\substack{j=1 \\ j \neq i}}^N p_{\text{IES},i,j}^t \quad (1)$$

式中: $p_{\text{target},i}^t > 0$  表示第  $i$  个区域 IES 购电, $p_{\text{target},i}^t < 0$  表示第  $i$  个区域 IES 售电; $p_{g,i}^t$  为  $t$  时刻第  $i$  个区域 IES 与主电网的交易功率, $p_{g,i}^t > 0$  表示第  $i$  个区域 IES 从主电网购电, $p_{g,i}^t < 0$  表示第  $i$  个区域 IES 向主电网售电;

$p_{\text{IES},i,j}^t$  为  $t$  时刻第  $i$  个区域 IES 与第  $j$  个区域 IES 的交易功率; $N$  为区域 IES 的数量。

记多区域 IES 运行时所产生的额外收益为  $C_g^-$ ,该收益反映了多区域 IES 相较于各独立区域与主电网交互时所产生的累计交易成本的降低量,如式(2)所示。

$$C_g^- = \sum_{i=1}^N C(p_{\text{target},i}^t) - C\left(\sum_{i=1}^N p_{\text{target},i}^t\right) \quad (2)$$

式中: $C(\cdot)$  为与主电网的交易成本函数。

根据各区域 IES 的目标交易功率,将多区域 IES 分为购电 IES 和售电 IES,计算总购电功率  $p_b^t$  和总售电功率  $p_s^t$ ,分别如式(3)和式(4)所示。

$$p_b^t = \sum p_{\text{target},i}^t \quad p_{\text{target},i}^t > 0 \quad (3)$$

$$p_s^t = \sum p_{\text{target},i}^t \quad p_{\text{target},i}^t < 0 \quad (4)$$

IES 的内部交易电价  $\varepsilon_{\text{IES}}^t$  可表示为:

$$\varepsilon_{\text{IES}}^t = \begin{cases} \varepsilon_{\text{IES},b}^t & p_{\text{IES},i,j}^t \geq 0 \\ \varepsilon_{\text{IES},s}^t & p_{\text{IES},i,j}^t < 0 \end{cases} \quad (5)$$

$$\varepsilon_{\text{IES},s}^t = \frac{\varepsilon_{\text{grid},s}^t p_s^t - C_g^-/2}{p_s^t} \quad (6)$$

$$\varepsilon_{\text{IES},b}^t = \frac{\varepsilon_{\text{grid},b}^t p_b^t - C_g^-/2}{p_b^t} \quad (7)$$

$$\varepsilon_{\text{grid}}^t = \begin{cases} \varepsilon_{\text{grid},b}^t & p_{g,i}^t \geq 0 \\ \varepsilon_{\text{grid},s}^t & p_{g,i}^t < 0 \end{cases} \quad (8)$$

式中: $\varepsilon_{\text{IES},b}^t$ 、 $\varepsilon_{\text{IES},s}^t$  分别为  $t$  时刻 IES 之间的购、售电电价; $\varepsilon_{\text{grid}}^t$  为  $t$  时刻 IES 与主电网的交易电价; $\varepsilon_{\text{grid},b}^t$ 、 $\varepsilon_{\text{grid},s}^t$  分别为  $t$  时刻 IES 与主电网交易的购、售电电价。

系统中 CCHP 机组<sup>[25]</sup>、电锅炉、EES 等关键运行设备的模型见附录 B 式(B1)一(B10)。

### 1.2 目标函数

以 IES 的总运营成本最小化为优化目标,总运营成本主要考虑柴油发电机的发电成本、与主电网和其他 IES 的交易成本、从天然气网购买天然气的成本以及碳交易成本。

1) 在满足负荷需求和确保电能平衡的前提下,各区域 IES 中柴油发电机的发电成本可表示为:

$$f_{1,i}^t = a_i (p_{\text{DG},i}^t)^2 + b_i p_{\text{DG},i}^t + c_i \quad (9)$$

式中: $f_{1,i}^t$  为  $t$  时刻第  $i$  个区域 IES 中柴油发电机的发电成本; $a_i$ 、 $b_i$ 、 $c_i$  为第  $i$  个区域 IES 中柴油发电机的发电成本系数; $p_{\text{DG},i}^t$  为  $t$  时刻第  $i$  个区域 IES 中柴油发电机的输出功率。

2) 各区域 IES 与主电网和其他 IES 的交易成本可表示为:

$$f_{2,i}^t = \varepsilon_{\text{grid}}^t p_{g,i}^t + \sum_{\substack{j=1 \\ j \neq i}}^N \varepsilon_{\text{IES}}^t p_{\text{IES},i,j}^t \quad (10)$$

式中: $f_{2,i}^t$  为  $t$  时刻第  $i$  个区域 IES 与主电网和其他 IES 的交易成本。

3)各区域IES为了满足CCHP机组需求,从天然气网购买天然气的成本可表示为:

$$f_{3,i}^t = \varepsilon_{\text{gas}} \frac{p_{\text{cchp},i}^t}{\eta_{\text{cchp}}} \quad (11)$$

式中: $f_{3,i}^t$ 为 $t$ 时刻第 $i$ 个区域IES从天然气网购买天然气的成本; $\varepsilon_{\text{gas}}$ 为从天然气网购买天然气的价格; $p_{\text{cchp},i}^t$ 为 $t$ 时刻第 $i$ 个区域IES中CCHP机组输出的电功率; $\eta_{\text{cchp}}$ 为CCHP机组的发电效率。

4)为了进一步在调度模型中刻画低碳经济性,需要考虑碳排放交易政策对CCHP机组的影响。通过合理设计适用于IES的调度运行策略,能够有效提升系统的低碳经济运行水平<sup>[26]</sup>。由于CCHP机组在运行过程中消耗天然气,会产生二氧化碳,本文引入碳交易机制,并将其纳入优化目标。若实际碳排放量超出碳配额,则需要在碳交易市场购买额外的碳配额,从而产生额外的碳交易成本。碳交易成本可表示为:

$$f_{4,i}^t = \varepsilon_{\text{CO}_2} [E_{\text{cchp},i}^t - Q_{\text{CO}_2}^t]_+ \quad (12)$$

$$E_{\text{cchp},i}^t = \alpha_{\text{cchp}} \frac{p_{\text{cchp},i}^t}{\eta_{\text{cchp}}} \quad (13)$$

式中: $f_{4,i}^t$ 为 $t$ 时刻第 $i$ 个区域IES的碳交易成本; $\varepsilon_{\text{CO}_2}$ 为碳交易市场的价格; $E_{\text{cchp},i}^t$ 、 $Q_{\text{CO}_2}^t$ 分别为 $t$ 时刻第 $i$ 个区域IES中CCHP机组的碳排放量、碳配额; $[x]_+ = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$ ;  $\alpha_{\text{cchp}}$ 为CCHP机组的碳排放因子。

结合式(9)~(13),本文IES优化调度模型的目标函数可表示为:

$$\min \sum_{i=1}^T \sum_{i=1}^N (f_{1,i}^t + f_{2,i}^t + f_{3,i}^t + f_{4,i}^t) \quad (14)$$

式中: $T$ 为优化周期。

### 1.3 约束条件

1)与主电网和其他区域IES交易功率的上、下限约束。

$$\begin{cases} |p_{g,i}^t| \leq P_{\text{grid}}^{\text{max}} \\ |p_{\text{IES},i,j}^t| \leq P_{\text{IES}}^{\text{max}} \end{cases} \quad (15)$$

式中: $P_{\text{grid}}^{\text{max}}$ 、 $P_{\text{IES}}^{\text{max}}$ 分别为IES与主电网、其他区域IES的功率传输限值。

2)功率平衡约束。

$$p_{\text{target},i}^t + p_{\text{pv},i}^t + p_{\text{DG},i}^t + p_{\text{cchp},i}^t + p_{\text{ess},i}^t = p_{\text{load},i}^t + p_{\text{eb},i}^t + p_{\text{ec},i}^t \quad (16)$$

$$h_{\text{cchp},i}^t + h_{\text{eb},i}^t = h_{\text{load},i}^t \quad (17)$$

$$c_{\text{cchp},i}^t + c_{\text{ec},i}^t = c_{\text{load},i}^t \quad (18)$$

式中: $p_{\text{load},i}^t$ 、 $h_{\text{load},i}^t$ 、 $c_{\text{load},i}^t$ 分别为 $t$ 时刻第 $i$ 个区域IES的电、热、冷负荷; $p_{\text{pv},i}^t$ 为 $t$ 时刻第 $i$ 个区域IES中光伏的发电功率; $p_{\text{ess},i}^t$ 为 $t$ 时刻第 $i$ 个区域IES中ESS的充放电功率; $p_{\text{eb},i}^t$ 、 $p_{\text{ec},i}^t$ 分别为 $t$ 时刻第 $i$ 个区域IES中电锅炉、电制冷机的耗电功率; $h_{\text{cchp},i}^t$ 、 $h_{\text{eb},i}^t$ 分别为 $t$ 时刻第

个区域IES中CCHP机组、电锅炉输出的热功率; $c_{\text{cchp},i}^t$ 、 $c_{\text{ec},i}^t$ 分别为 $t$ 时刻第 $i$ 个区域IES中CCHP机组、电制冷机输出的冷功率。

3)IES内部交易功率约束。

$$\sum_{i=1}^N \sum_{j=1, j \neq i}^N p_{\text{IES},i,j}^t = 0 \quad (19)$$

式(19)表示IES内部的交易功率之和为0。

## 2 基于Dis-FRL的优化求解方法

式(14)与约束条件式(15)~(19)所构建的优化问题属于二次约束规划问题,可以采用现有的基于模型的方法和基于学习的方法进行求解。然而,这些方法在实时性、隐私保护以及更为严重的单点故障风险方面可能存在一些局限性。为了克服这些缺点,本章提出一种去中心化的FRL算法。

### 2.1 状态、动作、奖励的定义

为了方便利用强化学习算法求解第1章中构建的多区域IES优化调度模型,将各区域IES的EMS定义为一个智能体,将上述优化调度模型的数学表述转换为马尔可夫决策过程。

第 $i$ 个区域IES中智能体的状态集 $\Xi_i^t$ 包含 $t$ 时刻的负荷数据、光伏发电功率、EES的荷电状态(state of charge, SoC)、与主电网的交易电价以及时刻 $t$ ,如式(20)所示。

$$\Xi_i^t = \{ p_{\text{load},i}^t, h_{\text{load},i}^t, c_{\text{load},i}^t, p_{\text{pv},i}^t, S_{\text{soc},i}^t, \varepsilon_{\text{grid}}^t, t \} \quad (20)$$

式中: $S_{\text{soc},i}^t$ 为 $t$ 时刻第 $i$ 个区域IES中ESS的SoC。

选择优化模型中的变量 $p_{\text{DG},i}^t$ 、 $p_{\text{cchp},i}^t$ 、 $p_{\text{ess},i}^t$ 构成第 $i$ 个区域IES智能体的动作集 $A_i^t$ ,如式(21)所示。

$$A_i^t = \{ p_{\text{DG},i}^t, p_{\text{cchp},i}^t, p_{\text{ess},i}^t \} \quad (21)$$

一旦智能体的动作确定,则该区域IES的目标交易功率 $p_{\text{target},i}^t$ 可由式(16)获得。因此,在强化学习的训练过程中,功率平衡约束始终能得到满足。

设计智能体的奖励由两部分组成:一部分与式(14)所示目标函数有关,另一部分则是对智能体采取违反约束或不当行动的惩罚。因此,奖励 $R_i^t$ 可表示为:

$$R_i^t = R_{c,i}^t + R_{p,i}^t \quad (22)$$

$$R_{c,i}^t = -(f_{1,i}^t + f_{2,i}^t + f_{3,i}^t + f_{4,i}^t) \quad (23)$$

$$R_{p,i}^t = -\kappa \sum_{c \in C_i} (\max \{ 0, p_{c,\min}^t - p_c^t \} + \max \{ 0, p_c^t - p_{c,\max}^t \}) \quad (24)$$

式中: $R_{c,i}^t$ 与目标函数相关,表示为经济成本的负值; $R_{p,i}^t$ 为智能体违反约束时受到的惩罚; $\kappa$ 为惩罚权重; $C_i$ 为式(15)~(19)中的不等式约束集合; $p_c^t$ 为 $t$ 时刻不等式约束 $c$ 对应的变量; $p_{c,\max}^t$ 、 $p_{c,\min}^t$ 分别为 $t$ 时刻不等式约束 $c$ 对应变量的上、下限。

### 2.2 Dis-FRL框架

CFRL框架如图2(a)所示,其局限性在于:传统CFRL方法依赖中央服务器来实现多区域智能体模

型参数的聚合。首先,每个智能体在本地环境中进行训练,更新其本地模型参数  $\theta_i^t (i=1, 2, \dots, N)$  并将其上传至中央服务器;然后,中央服务器以一种典型的联邦平均聚合方式计算全局模型参数  $\theta_g^{t+1}$ [24],如式(25)所示,并将该全局模型参数分发至所有智能体,以确保各区域智能体的一致性。虽然 CFRL 能够在一定程度上减少数据隐私泄露风险,但由于每次聚合都依赖中央服务器,在实际应用中可能加剧对中央服务器的依赖性,从而面临单点故障的潜在挑战。

$$\theta_g^{t+1} = \frac{1}{N} \sum_{i=1}^N \theta_i^t \quad (25)$$

针对 CFRL 的局限性,本文提出了一种 Dis-FRL 方法,对多区域 IES 进行能量优化调度。Dis-FRL 框架如图 2(b)所示,每个区域由 1 个独立的智能体管理,智能体采取 4 个步骤进行操作:①上传数据集;②本地模型训练;③参数交换;④参数聚合。在 Dis-FRL 框架中,中央服务器的角色被完全移除,各智能体之间通过去中心化方式进行参数交换和协同

优化。这种设计不仅降低了对中央服务器的依赖性,还有效提升了系统的隐私保护能力。该方式依靠智能体之间的信息交互,不涉及敏感数据的直接传输,且每个智能体基于其邻域的信息更新自身模型,无须经过中央服务器的聚合。

虽然 Dis-FRL 无法从根本上完全避免单点故障的影响,但其去中心化架构使系统具有较高的灵活性。例如,当某节点发生故障时,可通过隔离失效节点,并重新设计分布式拓扑,使剩余节点通过更新智能体的交互关系,形成新的优化协同结构,从而维持系统整体功能的正常运行。这种灵活性使 Dis-FRL 在应对动态变化和节点失效时优于 CFRL 方法。

### 2.3 Dis-FRL 算法实施

在 Dis-FRL 框架下,每个区域智能体均通过双延迟确定性策略梯度(twin delayed deep deterministic policy gradient, TD3)算法在本地环境中进行训练,并与邻域智能体通过去中心化的参数交换实现协同优化。基于 Dis-FRL 的多区域 IES 优化调度实施过程见附录 B 图 B1。为了在保护数据隐私的同时增强调度性能,Dis-FRL 引入双随机矩阵的加权融合机制,以保证智能体之间参数交互的随机性和动态性。下面将介绍 Dis-FRL 的算法实施过程,包括智能体本地训练过程、分布式参数共享机制、双随机矩阵的权重设置。

#### 1) 智能体本地训练过程。

在 Dis-FRL 框架中,各智能体在本地环境中以 TD3 算法独立进行策略优化。具体而言,TD3 算法通过引入双 Critic 网络  $\{Q_1, Q_2\}$  来估计价值函数,并利用两者之间的最小值作为目标  $Q$  值的估计,如式(26)所示。这样可以有效避免单一 Critic 网络可能出现的高估  $Q$  值问题,从而提升价值估计的准确性。

$$y = r + \gamma \min \{Q_1(s', a'), Q_2(s', a')\} \quad (26)$$

式中: $y$  为目标  $Q$  值; $s', a'$  分别为下一时刻的状态、动作; $\gamma$  为折扣因子; $r$  为奖励值。

为了提高策略更新的稳定性,TD3 算法在每 2 次 Critic 网络更新后才对 Actor 网络进行 1 次更新。该机制能够使 Critic 网络更充分地学习当前策略的  $Q$  值估计,从而为策略优化提供更稳定的梯度方向,避免 Actor 网络在目标函数波动时频繁更新导致的不稳定性。在训练过程中,智能体引入了目标网络,对其采取软更新的方式以减缓策略优化中的波动。

Dis-FRL 框架通过强化学习的环境交互特性,动态适应不确定性条件。智能体在本地训练过程中通过学习历史数据,获取了需求波动的分布特征,在实际运行中,智能体能够直接根据当前环境状态(如实时能源需求、实时可再生能源出力等)进行策略优化。这种离线训练与在线学习相结合的方式,不仅显著降低了计算复杂度,还确保了策略优化的实

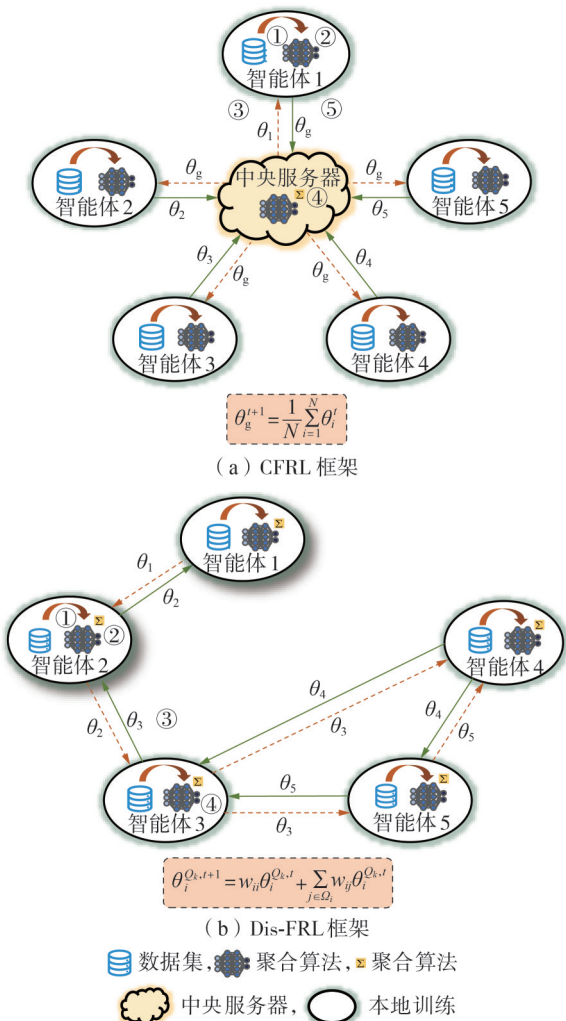


图 2 CFRL 和 Dis-FRL 的训练与聚合过程  
Fig.2 Training and aggregation process of CFRL and Dis-FRL

时性。

### 2) 分布式参数共享机制。

Dis-FRL通过去中心化的参数共享机制,使各区域智能体在不依赖中央服务器的情况下实现协同优化。该过程主要是受分布式无向图结构的启发,根据拓扑结构,每个智能体仅与其直接邻居进行交互,从而克服了传统集中式聚合方法的缺点。具体而言,每个智能体在本地训练完成后,将其Critic网络参数 $\theta_i^{k,t}$  ( $k=1, 2$ )与邻域智能体进行交换,从而在去中心化环境下实现全局策略的协同优化。智能体间的分布式参数共享方式为:

$$\theta_i^{k,t+1} = w_{ii} \theta_i^{k,t} + \sum_{j \in \Omega_i} w_{ij} \theta_j^{k,t} \quad k=1, 2 \quad (27)$$

式中: $\Omega_i$ 为与智能体 $i$ 相邻的智能体集合; $w_{ii}$ 为智能体 $i$ 自身的本地参数权重; $w_{ij}$ 为智能体 $i$ 对邻居智能体 $j$ 的参数权重。

Dis-FRL正是基于上述机制,当智能体之间进行参数交换时,不需要共享原始数据,从而避免了数据集中存储带来的隐私泄露风险。

### 3) 双随机矩阵的权重设置。

当分配参数时,双随机矩阵的元素通过分析智能体之间的连接性来定义,以确保信息交换的有效性和平衡。根据文献[27],得到参数共享的双随机矩阵元素,如式(28)所示。

$$w_{ij} = \begin{cases} \min \left\{ \frac{1}{n_i}, \frac{1}{n_j} \right\} & j \in \Omega_i \\ 1 - \sum_{j \in \Omega_i} \min \left\{ \frac{1}{n_i}, \frac{1}{n_j} \right\} & j = i \\ 0 & \text{其他} \end{cases} \quad (28)$$

式中: $n_i, n_j$ 分别为与智能体 $i, j$ 相邻智能体的数量。矩阵的行元素和列元素之和均为1,即 $\sum_{i=1}^N w_{ij} = \sum_{j=1}^N w_{ji} = 1$ 。

双随机矩阵的引入实现了智能体之间的动态权重分配,在保护隐私的同时,通过随机化权重分配降低了信息泄露的风险。Dis-FRL方法通过TD3算法的本地训练、双随机矩阵的加权参数交互,实现了无中央服务器的分布式优化。相较于传统集中式方法采取式(25)所示联邦平均的方式,Dis-FRL方法无须将其智能体的参数发送到中央服务器进行更新,而是根据式(27)所示方式与邻近智能体进行参数共享。且双随机矩阵的引入进一步增强了隐私保护能力,并确保各区域智能体在独立优化的同时实现全局协同优化。

## 3 算例分析

### 3.1 算例配置

#### 1) 神经网络及算法参数配置。

在本文算例中,每个智能体的Actor、Critic网络

均采用2层前馈神经网络架构。Actor网络通过2层全连接层对输入状态进行处理,每层包含256个神经元,激活函数为ReLU,输出层使用Tanh激活函数以限制动作空间的范围。Critic网络接收当前状态和动作信息,通过2层全连接层进行处理,每层包含256个神经元,并且使用ReLU函数激活输出值估计。此外,为了确保Critic网络的训练稳定性,采用LayerNorm对网络权重进行归一化处理。

TD3算法的关键超参数设置如下:设折扣因子 $\gamma=0.995$ ,惩罚权重 $\kappa=3 \times 10^5$ ,软更新因子 $\tau=10^{-3}$ ;策略延迟系数为2,即每2次Critic网络更新后执行1次Actor网络更新;小批量大小设置为48,经验重放缓冲区设置为100000,Critic网络和Actor网络的学习率分别为 $10^{-4}$ 、 $10^{-3}$ ,并采用Adam优化器。实验在Python 3.7.0环境下通过TensorFlow 1.15.0实现。

#### 2) IES参数配置。

为了验证本文所提Dis-FRL方法在多区域IES优化调度中的有效性,构建一个包含5个区域的多区域IES( $IES_1 - IES_5$ )仿真模型,每个区域均配置5种关键设备(柴油发电机、CCHP机组、电锅炉、电制冷机、EES)。负荷需求曲线和光伏发电出力曲线基于某城市能源管理数据获取,见附录C图C1。使用拉丁超立方体采样方法构建样本数据集,以确保数据样本的多样性<sup>[28]</sup>。调度周期 $T=24$  h,时间步长为1 h。系统采用分时定价机制,分时电价见附录C表C1。天然气价格为70 \$ / MW。IES中各设备的详细参数和其他系统参数设置分别见附录C表C2和表C3。

### 3.2 IES调度结果分析

在分时电价下 $IES_1$ 的调度结果如图3所示。该调度结果表明本文所提方法能在保证能源供需平衡的同时,有效应对电价波动,显著降低能源成本。

#### 1) 电功率调度结果。

由图3(a)可知,系统充分利用了电价波动来优化电力调度。在电价低谷时段(如01:00—07:00),EES优先充电以储存低价电能,从而能在后续高峰时段提供电能支撑;而在电价高峰时段(08:00—12:00、17:00—21:00),EES放电以减少从主电网的购电功率,降低购电成本。在非高峰时段,电锅炉和电制冷机按需启动,以满足冷、热负荷需求;在高峰时段,则通过CCHP机组发电供电,减少电锅炉、电制冷机的耗电功率,从而避免以高价购电。在光伏出力充足且电价较高时段,系统优先售电至主电网或其他区域IES,以提升收益。此外,各区域IES优先通过内部交易解决电力需求与供应的缺额问题,仅在本地资源不足或有剩余电力时才与主电网进行交易。整体调度策略体现了“谷充峰放”和“本地优先”的原则,有效实现了能量互补和经济性提升。

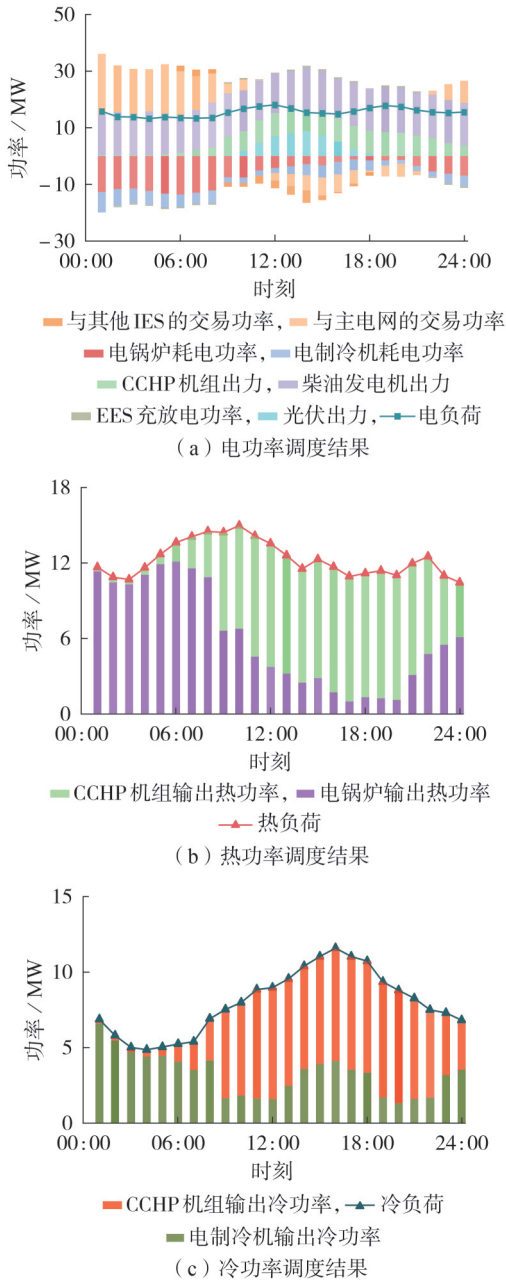


图3 基于Dis-FRL的IES<sub>1</sub>调度结果

Fig.3 Scheduling results of IES<sub>1</sub> based on Dis-FRL

2)冷、热功率调度结果。

由图3(b)可知,CCHP机组和电锅炉的不同运行策略可以满足热负荷需求。若天然气价格低于电价,则CCHP机组的供热功率增大;否则,电锅炉会增大热出力来满足热负荷需求。由图3(c)可知,CCHP机组和电制冷机的运行策略相似。所得优化调度策略有效降低了对主电网的依赖,不仅提升了系统的能源效率,还增强了经济性和运营灵活性。

为了验证Dis-FRL方法在更高时间分辨率下的适用性,将时间步长细化至15 min,此时的调度结果见附录C图C2。结果表明,Dis-FRL方法在更高时间分辨率调度场景下仍具备较强的优化性能和实时

性潜力。

3.3 不同场景对比分析

为了验证碳排放交易政策对IES调度的影响,在已有调度场景的基础上新增以下场景进行对比分析:①场景1,不考虑碳交易限制,优化目标不考虑碳交易成本;②场景2,碳排放量受限于固定碳配额,无碳交易机制;③场景3,允许通过碳交易市场弥补超额碳排放。

不同场景的运行结果如表1所示。由表可知:在无碳交易限制的场景1中,虽然运行成本最低,但碳排放量高达95.66 t,远超碳配额;场景2通过严格限制碳排放量,虽然满足了政策要求,但因高排放时段的CCHP机组出力受限,导致需要更多地依赖其他高成本能源供给,使运行成本显著增加至\$ 68 076;场景3引入了碳交易机制,通过购买碳配额来弥补超额碳排放,相较于场景2,总成本减少了7.2%。上述结果验证了碳交易机制在保持经济性与满足环境约束之间的优势,凸显了Dis-FRL方法在碳交易政策约束下的实际应用价值。

表1 不同场景的运行结果

Table 1 Operation results of different scenarios

场景	碳排放量 / t	运行成本 / \$	碳交易成本 / \$	总成本 / \$
1	95.66	61 284	0	61 284
2	83.91	68 076	0	68 076
3	92.75	62 888	910	63 798

3.4 不同方法的对比分析

本节将本文所提Dis-FRL方法与商业求解器Gurobi和2种现有的多智能体强化学习方法(多智能体双延迟确定性策略梯度(multi-agent twin delayed deep deterministic policy gradient,MATD3)算法<sup>[29]</sup>和CFRL方法<sup>[24]</sup>)进行比较,结果如表2所示。由表可知:Dis-FRL方法所得总成本为\$ 261 995,相较于MATD3算法和CFRL方法所得结果,分别减少了3.1%、1.5%左右;Gurobi求解器所得总成本为\$ 240 510,虽然总成本最小,但其平均计算时间较长,难以满足实时调度需求。而3种多智能体强化学习方法在离线训练阶段学习到的调度策略能够

表2 不同方法的结果比较

Table 2 Results comparison among different methods

方法	运行成本 / \$					总成本 / \$	平均计算时间 / s
	IES <sub>1</sub>	IES <sub>2</sub>	IES <sub>3</sub>	IES <sub>4</sub>	IES <sub>5</sub>		
Gurobi求解器	66544	52151	38812	33190	49813	240510	13.60
MATD3算法	74794	58616	43624	37305	55989	270328	0.75
CFRL方法	73592	57675	42923	36705	55089	265984	0.92
Dis-FRL方法	72488	56809	42280	36155	54263	261995	0.82

直接用于在线优化,从而显著缩短了在线计算时间。Dis-FRL方法的平均计算时间为0.82 s,远少于Gurobi求解器的13.60 s,这是因为Dis-FRL方法通过离线训练在系统运行特性方面建立了长期学习能力,相较于传统优化方法,其能够以较低的计算复杂度快速做出调度决策,尤其适用于高频次或实时响应的调度场景。

MATD3算法、CFRL方法、Dis-FRL方法的奖励收敛曲线如图4所示。由图可知:虽然CFRL方法的收敛速度略快于Dis-FRL方法,但Dis-FRL方法下的奖励值在收敛前期平滑增长,波动较小,表明其学习过程更加稳定;此外,Dis-FRL方法大约迭代750个回合后达到收敛,具有较快的收敛速度,而MATD3算法的收敛速度和稳定性均不及Dis-FRL方法,其大约迭代1000个回合后达到收敛,且在训练后期的奖励值波动较大。上述结果表明,Dis-FRL方法通过去中心化的双随机矩阵协同机制,有效协调各智能体的策略更新,从而提高了收敛过程的稳定性。

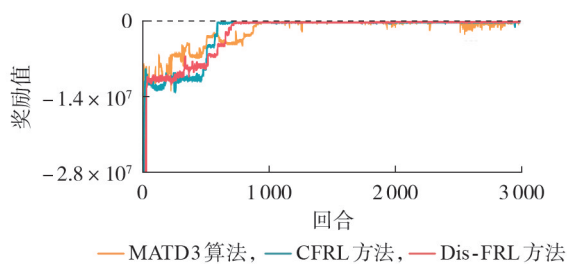


图4 不同方法的奖励值收敛曲线

Fig.4 Reward value convergence curves of different methods

为了验证双随机矩阵在智能体协同优化中的作用,本节设计了对比实验,分别采用静态非双随机矩阵和双随机矩阵进行参数交互,实验结果如图5所示。由图可知,双随机矩阵能确保智能体之间的信息流动更加均衡,减少了训练过程中的参数冲突,使策略更新过程更加高效。因此,当采用双随机矩阵进行参数交互时,大约迭代750个回合后达到收敛,而当采用静态非双随机矩阵进行参数交互时,在早

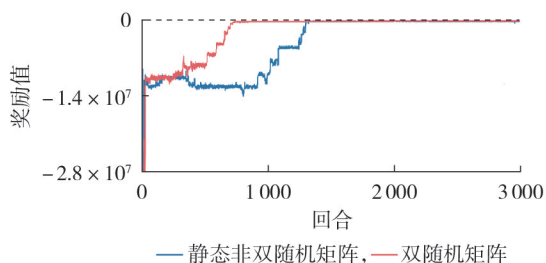


图5 双随机矩阵与静态非双随机矩阵下的奖励值收敛曲线

Fig.5 Reward value convergence curves under double random matrix and static non-double random matrix

期阶段参数交换不充分,导致奖励值的波动较大,大约需要迭代1300个回合才能达到类似的收敛效果。

不同优化方法在数据共享类型、训练方式和隐私性方面的差异如表3所示。从数据共享类型来看:Gurobi求解器需要对全局数据进行集中化处理,直接暴露了原始数据的敏感信息;MATD3算法通过经验回放共享历史数据,并采用集中式训练、分散式执行的架构,数据隐私性也较弱;CFRL方法采用分布式训练、分散式执行架构,通过模型参数共享实现了一定程度的隐私保护,但由于依赖中央服务器,其隐私性取决于服务器的安全性;相较而言,Dis-FRL方法完全摒弃了中央服务器,智能体之间仅通过参数交互完成优化,避免了原始数据的共享和集中存储,显著提升了隐私保护水平<sup>[30]</sup>。

表3 不同优化方法的隐私性比较

Table 3 Privacy comparison among different optimization methods

方法	数据共享类型	训练执行方式	隐私性
Gurobi求解器	全局数据	—	弱
MATD3算法	经验回放	集中式训练、分散式执行	弱
CFRL方法	网络模型参数	分布式训练、分散式执行,依赖中央服务器	中
Dis-FRL方法	网络模型参数	分布式训练、分散式执行,不依赖中央服务器	强

## 4 结论

本文提出了一种基于Dis-FRL的IES优化调度方法,以解决传统集中式方法中的数据隐私泄露和计算效率问题。Dis-FRL方法通过去中心化的智能体协同优化,结合双随机矩阵的权重分配,实现了无中央服务器的参数交换,在提升隐私保护的同时,有效降低了系统的计算成本。算例仿真结果表明,本文所提方法能使IES中的敏感数据得到保障,且能降低能源系统的经济成本,提高能源分配效率,同时具有较好的收敛性能。未来的研究将探索异构型智能体之间的协调机制,以进一步提升分布式联邦学习在更广泛应用场景中的性能和可靠性。

附录见本刊网络版(<http://www.epae.cn>)。

## 参考文献:

- [1] 黄子硕,何桂雄,闫华光,等. 园区级综合能源系统优化模型功能综述及展望[J]. 电力自动化设备,2020,40(1):10-18. HUANG Zishuo, HE Guixiong, YAN Huaguang, et al. Overview and prospect of optimization model function for community-scale integrated energy system [J]. Electric Power Automation Equipment, 2020,40(1):10-18.
- [2] 米阳,孔令飞,王育飞,等. 面向配电网的多综合能源系统容量规划和能源交易优化[J]. 电力自动化设备,2025,45(9):38-45,87.

- MI Yang, KONG Lingfei, WANG Yufei, et al. Capacity planning and energy trading optimization of multiple integrated energy systems oriented to distribution network[J]. *Electric Power Automation Equipment*, 2025, 45(9):38-45, 87.
- [3] 齐先军, 蒋中琦, 张晶晶, 等. 考虑碳捕集与综合需求响应互补的综合能源系统优化调度[J]. *电力自动化设备*, 2023, 43(7):133-141.
- QI Xianjun, JIANG Zhongqi, ZHANG Jingjing, et al. Optimal dispatching of integrated energy system considering complementation of carbon capture and integrated demand response[J]. *Electric Power Automation Equipment*, 2023, 43(7):133-141.
- [4] WANG Y L, WANG Y D, HUANG Y J, et al. Optimal scheduling of the regional integrated energy system considering economy and environment[J]. *IEEE Transactions on Sustainable Energy*, 2019, 10(4):1939-1949.
- [5] 申鸿帅, 吕家君, 李更丰, 等. 考虑热储灵活配置与热网水力热力特性的电热综合能源系统优化规划[J]. *电力自动化设备*, 2024, 44(7):214-222, 254.
- SHEN Hongshuai, LÜ Jiajun, LI Gengfeng, et al. Optimal planning of integrated electricity and heating system considering flexible configuration of thermal storage and hydraulic-thermal characteristics of thermal network[J]. *Electric Power Automation Equipment*, 2024, 44(7):214-222, 254.
- [6] 程义, 李更丰. 基于双层模仿学习的多园区综合能源系统分布式协同优化调度[J]. *电力系统自动化*, 2022, 46(24):16-25.
- CHENG Yi, LI Gengfeng. Distributed collaborative optimal dispatch of multi-park integrated energy system based on bilayer imitation learning[J]. *Automation of Electric Power Systems*, 2022, 46(24):16-25.
- [7] LU Z X, XU X Y, YAN Z, et al. Multistage robust optimization of routing and scheduling of mobile energy storage in coupled transportation and power distribution networks[J]. *IEEE Transactions on Transportation Electrification*, 2022, 8(2):2583-2594.
- [8] 乔骥, 王新迎, 张擎, 等. 基于柔性行动器-评判器深度强化学习的电-气综合能源系统优化调度[J]. *中国电机工程学报*, 2021, 41(3):819-833.
- QIAO Ji, WANG Xinying, ZHANG Qing, et al. Optimal dispatch of integrated electricity-gas system with soft actor-critic deep reinforcement learning[J]. *Proceedings of the CSEE*, 2021, 41(3):819-833.
- [9] 沈运帷, 徐凯, 林顺富, 等. 考虑广义储能参与的多园区综合能源系统低碳优化运行策略[J]. *电力自动化设备*, 2024, 44(11):41-51.
- SHEN Yunwei, XU Kai, LIN Shunfu, et al. Low-carbon optimal operation strategy of multi-park integrated energy system considering generalized energy storage participation[J]. *Electric Power Automation Equipment*, 2024, 44(11):41-51.
- [10] 杨挺, 赵黎媛, 刘亚闯, 等. 基于深度强化学习的综合能源系统动态经济调度[J]. *电力系统自动化*, 2021, 45(5):39-47.
- YANG Ting, ZHAO Liyuan, LIU Yachuang, et al. Dynamic economic dispatch for integrated energy system based on deep reinforcement learning[J]. *Automation of Electric Power Systems*, 2021, 45(5):39-47.
- [11] 赵鹏杰, 吴俊勇, 王斌, 等. 基于深度强化学习的微电网优化运行策略[J]. *电力自动化设备*, 2022, 42(11):9-16.
- ZHAO Pengjie, WU Junyong, WANG Yi, et al. Optimal operation strategy of microgrid based on deep reinforcement learning[J]. *Electric Power Automation Equipment*, 2022, 42(11):9-16.
- [12] 窦嘉铭, 王小君, 刘盟, 等. 面向综合能源系统调度多场景泛化提升的扩散策略学习[J]. *电力系统自动化*, 2026, 50(3):145-155.
- DOU Jiaming, WANG Xiaojun, LIU Zhao, et al. Diffusion-policy learning for multi-scenario generalized boosting of integrated energy system scheduling[J]. *Automation of Electric Power Systems*, 2026, 50(3):145-155.
- [13] 梁涛, 柴露露, 谭建鑫, 等. 基于深度强化学习算法的氢耦合电-热综合能源系统优化调度[J]. *电力自动化设备*, 2025, 45(1):59-66.
- LIANG Tao, CHAI Lulu, TAN Jianxin, et al. Optimal scheduling of hydrogen coupled electrothermal integrated energy system based on deep reinforcement learning algorithm[J]. *Electric Power Automation Equipment*, 2025, 45(1):59-66.
- [14] 张帆, 武东吴, 陈玉萍, 等. 多智能体深度强化学习的分布式园区综合能源系统经济调度策略[J]. *电力系统及其自动化学报*, 2022, 34(12):18-26.
- ZHANG Fan, WU Donghao, CHEN Yuping, et al. Economic scheduling strategy for integrated energy system in distributed parks based on multi-agent deep reinforcement learning[J]. *Proceedings of the CSU-EPSA*, 2022, 34(12):18-26.
- [15] DING L F, LIN Z Y, SHI X S, et al. Target-value-competition-based multi-agent deep reinforcement learning algorithm for distributed nonconvex economic dispatch[J]. *IEEE Transactions on Power Systems*, 2023, 38(1):204-217.
- [16] YE Y J, TANG Y, WANG H Y, et al. A scalable privacy-preserving multi-agent deep reinforcement learning approach for large-scale peer-to-peer transactive energy trading[J]. *IEEE Transactions on Smart Grid*, 2021, 12(6):5185-5200.
- [17] 杨文伟, 彭显刚, 全欢, 等. 基于联邦强化学习的主动配电网多主体博弈协同优化策略[J]. *电力系统自动化*, 2025, 49(13):70-82.
- YANG Wenwei, PENG Xiangang, QUAN Huan, et al. Multi-agent game collaborative optimization strategy for active distribution networks based on federated reinforcement learning[J]. *Automation of Electric Power Systems*, 2025, 49(13):70-82.
- [18] 陈明昊, 孙毅, 胡亚杰, 等. 基于纵向联邦强化学习的居民社区综合能源系统协同训练与优化管理方法[J]. *中国电机工程学报*, 2022, 42(15):5535-5550.
- CHEN Minghao, SUN Yi, HU Yajie, et al. The collaborative training and management-optimized method for residential integrated energy system based on vertical federated reinforcement learning[J]. *Proceedings of the CSEE*, 2022, 42(15):5535-5550.
- [19] 李锦辉, 吴毓峰, 余涛, 等. 数据孤岛下基于联邦学习的用户电价响应刻画及其应用[J]. *电力系统保护与控制*, 2024, 52(6):164-176.
- LI Jinhui, WU Yufeng, YU Tao, et al. Characterization of user price response behavior and its application based on federated learning considering a data island[J]. *Power System Protection and Control*, 2024, 52(6):164-176.
- [20] 蒲天骄, 杜帅, 李焱, 等. 面向隐私保护基于联邦强化学习的分布式电源协同优化策略[J]. *电力系统自动化*, 2023, 47(8):62-70.
- PU Tianjiao, DU Shuai, LI Ye, et al. Collaborative optimization strategy of distributed generators based on federated reinforcement learning for privacy preservation[J]. *Automation of Electric Power Systems*, 2023, 47(8):62-70.
- [21] LEE J H, PARK J Y, SIM H S, et al. Multi-residential energy scheduling under time-of-use and demand charge tariffs with federated reinforcement learning[J]. *IEEE Transactions on Smart Grid*, 2023, 14(6):4360-4372.
- [22] 孙晓荣, 杨爽爽, 潘学萍, 等. 交互能源机制下多区域综合能源系统能源共享协同调度[J]. *电力自动化设备*, 2023, 43(11):9-17.

- SUN Xiaorong, YANG Shuangshuang, PAN Xueping, et al. Energy sharing and cooperative scheduling of multi-regional integrated energy system under transactive energy mechanism [J]. *Electric Power Automation Equipment*, 2023, 43(11): 9-17.
- [23] 杨茂, 杜茜, 孙莉, 等. 基于双层博弈的多区域综合能源系统协同优化运行策略[J]. *电力自动化设备*, 2026, 46(2): 30-38.
- YANG Mao, DU Qian, SUN Li, et al. Collaborative optimization operation strategy of multi-regional integrated energy system based on dual-layer game [J]. *Electric Power Automation Equipment*, 2026, 46(2): 30-38.
- [24] 周毅斌, 肖浩, 裴玮, 等. 基于纵向联邦学习的微电网群协同优化运行与策略进化[J]. *电力系统自动化*, 2023, 47(11): 121-132.
- ZHOU Yibin, XIAO Hao, PEI Wei, et al. Collaborative optimization operation and strategy evolution of microgrid cluster based on vertical federated learning [J]. *Automation of Electric Power Systems*, 2023, 47(11): 121-132.
- [25] 李佳佳, 李兴朔, 周国文, 等. 基于火-储联合循环的冷热电三联供系统多运行模式热力性能分析[J]. *中国电机工程学报*, 2024, 44(9): 3598-3610.
- LI Jiajia, LI Xingshuo, ZHOU Guowen, et al. Thermo-dynamics analysis under different operation modes of trigenerative system based on CFPP-CAES combined cycle [J]. *Proceedings of the CSEE*, 2024, 44(9): 3598-3610.
- [26] 张靖, 李勇, 曹一家, 等. 计及需求侧管理与碳排放的综合能源系统光-储设备优化配置方法[J]. *电力自动化设备*, 2024, 44(7): 189-197.
- ZHANG Jing, LI Yong, CAO Yijia, et al. Optimal configuration approach of PV-ESS for integrated energy system considering demand-side management and carbon emission [J]. *Electric Power Automation Equipment*, 2024, 44(7): 189-197.
- [27] GUO F H, WEN C Y, LI G Q, et al. Distributed economic dispatch for a multi-area power system [C]//2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA). Auckland, New Zealand: IEEE, 2015: 620-625.
- [28] 徐询, 谢丽蓉, 叶林, 等. 基于非参数核密度估计的风电场有功功率双层优化模型[J]. *电力系统自动化*, 2022, 46(2): 43-55.
- XU Xun, XIE Lirong, YE Lin, et al. Bi-level optimization model of active power for wind farm based on nonparametric kernel density estimation [J]. *Automation of Electric Power Systems*, 2022, 46(2): 43-55.
- [29] CHEN T Y, BU S R, LIU X, et al. Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning [J]. *IEEE Transactions on Smart Grid*, 2022, 13(1): 715-727.
- [30] YU W R, LI Q X, LOPUHAÄ-ZWAKENBERG M, et al. Provable privacy advantages of decentralized federated learning via distributed optimization [J]. *IEEE Transactions on Information Forensics and Security*, 2025, 20: 822-838.

#### 作者简介:

朱新文(1998—),男,硕士研究生,主要研究方向为综合能源系统优化运行(E-mail: 211122030085@zjut.edu.cn);

郭方洪(1987—),男,副教授,博士研究生导师,通信作者,主要研究方向为微电网控制与优化(E-mail: fhguo@zjut.edu.cn)。

(编辑 陆丹)

## Distributed federated reinforcement learning-based optimal scheduling of multi-regional integrated energy system

ZHU Xinwen, WANG Jiaqi, LI Shengwei, LIN Wenjie, WU Xiang, GUO Fanghong  
(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract:** Aiming at the limitations of traditional optimization methods and centralized federated reinforcement learning in terms of privacy protection and computational efficiency, a distributed federated reinforcement learning-based optimal scheduling method of multi-regional integrated energy system is proposed. Each regional integrated energy system is managed by an individual agent. Each agent optimizes the parameters of its local Critic network using the twin delayed deep deterministic policy gradient algorithm and exchanges parameter information with the neighboring agents, thereby enabling efficient energy scheduling management without reliance on an additional centralized server. To ensure global optimization, the weight coefficients of parameter interaction are determined by the elements of a doubly stochastic matrix. The results of case study analysis show that the proposed method can enhance the privacy protection of the integrated energy system while demonstrating excellent convergence performance and effectively reducing operation costs.

**Key words:** integrated energy system; optimal scheduling; distributed federated reinforcement learning; agent; twin delayed deep deterministic policy gradient algorithm