文章编号：1000-3673（2026）01-0081-11 中图分类号：TM 721 文献标志码：A 学科代码：470·40

基于个性化联邦强化学习的多园区综合能源系统低碳经济调度

张东升[1]，黄媛[1]，许潇[1]，刘俊勇[1]，孙增杰[2]，樊茂森[3]
（1. 四川大学电气工程学院，四川省 成都市 610065；
2. 国网河北省电力有限公司，河北省 石家庄市 050000；
3. 国网河北省电力有限公司保定供电公司，河北省 保定市 071000）

Low-carbon Economic Dispatch of Multi-park Integrated Energy Systems Based on Personalized Federated Reinforcement Learning

ZHANG Dongsheng[1], HUANG Yuan[1], XU Xiao[1], LIU Junyong[1], SUN Zengjie[2], FAN Maosen[3]
(1. College of Electrical Engineering, Sichuan University, Chengdu 610065, Sichuan Province, China;
2. State Grid Hebei Electric Power Co., Ltd., Shijiazhuang 050000, Hebei Province, China;
3. Baoding Power Supply Company of State Grid Hebei Electric Power Co., Ltd., Baoding 071000, Hebei Province, China)

ABSTRACT: Collaborative optimization of multi-park integrated energy systems is essential for enhancing multi-energy coupling and complementarity, reducing carbon emissions, and lowering operational costs. However, this process faces significant challenges, including heterogeneous privacy protection among multiple agents and the need for real-time decision-making. To address these issues, this paper proposes a low-carbon economic dispatch model that balances personalized decision-making for each park with privacy protection. The model employs an offline training and online decision-making approach to meet real-time requirements. At the individual park level, an improved Deep Deterministic Policy Gradient (DDPG) reinforcement learning method is utilized to develop a dispatch model that comprehensively considers electric and thermal loads, time-of-use electricity pricing, and tiered carbon trading. Additionally, embedded vectors are incorporated into the DDPG neural network to enhance the agent's perception and adaptability to the operational characteristics of each park. To ensure data privacy and address the challenges of multi-agent heterogeneity, a personalized federated framework is established for collaborative training across multiple parks within the integrated energy system. In this framework, each park's DDPG model transmits only model parameters, excluding embedded vectors, to a central aggregator, thereby preventing the direct exchange of operational data between parks. Simulation experiments conducted across multiple scenarios and algorithms demonstrate that each park can achieve global optimization using only local model information, thereby facilitating system-wide low-carbon economic dispatch.

KEY WORDS: multi-park integrated energy systems; personalized federated learning; privacy preserving; multi-agent heterogeneity; tiered carbon trading

摘要：多园区综合能源系统协同优化在提升多能耦合互补、降低碳排放与运行成本的过程中，亟需应对多主体异构、隐私保护与实时决策等问题。为此，该文提出一套兼顾园区个性化决策与隐私保护的低碳经济调度模型，通过离线训练与在线决策的方式，满足实时决策需求。首先，在单园区层面，采用改进的深度确定性策略梯度(deep deterministic policy gradient，DDPG)强化学习方法，构建综合考虑电-热负荷、分时电价和阶梯式碳交易的调度模型，并在 DDPG 的神经网络中增加嵌入向量增强智能体感知和适应该园区运行特性的能力。其次，为满足数据隐私保护需求并应对多主体异构挑战，搭建多园区综合能源系统的个性化联邦框架进行协同训练，此时单园区 DDPG 模型仅向聚合中心传递除嵌入向量以外的模型参数，避免园区间直接交换运行数据。最后，基于多场景多算法进行仿真实验，结果表明各园区仅利用局部模型信息即可实现全局优化，实现了系统低碳经济调度。

关键词：多园区综合能源系统；个性化联邦；隐私保护；多主体异构；阶梯式碳交易

DOI：10.13335/j.1000-3673.pst.2024.2232

0 引言

为满足提高经济效益与降低碳排放的现实需求，能源领域逐渐构建起可统筹多种能源形态的综合能源系统，从而实现更为低碳的经济运行[1-3]。然

基金项目：国家自然科学基金项目(52207128)。
Project Supported by National Natural Science Foundation of China (52207128).

而，相比于单个园区综合能源系统(park integrated energy system，PIES)单独运行，同一配电区域内多个园区综合能源系统(multiple park integrated energy systems，MPIES)的互联运行能够进一步挖掘灵活性资源，通过能源和信息的交互，提高了各园区运行方式的互补性[4]，减少了园区对上级高碳排放能源网络的依赖。因此，对 MPIES 的调度研究逐渐成为助力实现“双碳”目标的重要手段。

目前，国内外学者针对 MPIES 的优化调度问题进行了广泛研究。文献[5]提出了一种基于可再生能源配额制和绿证交易机制的 MPIES 优化调度模型，采用集中优化求解，在保证系统经济性的同时降低了碳排放。然而，集中优化会显著增加通信成本并影响数据隐私。为此，文献[6]提出了分布式鲁棒的多能源枢纽优化调度模型，采用交替方向乘子(alternating direction method of multipliers，ADMM)算法进行分布式求解，在保护隐私的同时优化调度决策。文献[7]利用能源共享策略，协同区域 IES 中互联的多微能网进行优化调度，并采用 ADMM 算法实现能量分配的分布式优化。文献[8]将 MPIES 内部分设备整合为广义储能，结合碳捕集技术，在有限信息交互基础上，采用分布式算法实现了博弈均衡解。然而，传统的模型驱动方法已难以适应愈发复杂的 MPIES 优化调度需求：在建模层面，高精度物理建模存在困难，固定的调度方案无法灵活应对动态变化[9]；在场景层面，多种能源设备间的强耦合与不确定性因素叠加，使得算法鲁棒性受到严重限制[10]；在决策层面，实时决策场景对高性能计算与快速响应的迫切需求亦使现有方法难以胜任[11]。在此多重挑战下，传统模型驱动方法已然步入瓶颈期。

为有效应对上述挑战，深度强化学习(deep reinforcement learning，DRL)被广泛应用于 MPIES 优化调度领域。作为一种无模型[12]方法，DRL 利用历史数据自适应学习，应对负荷和新能源的不确定性[13-14]，实现离线训练、在线决策的复杂调度需求。文献[15]提出了一种基于多智能体双深度 Q 网络的多微能网能源管理模型，但其离散动作空间难以实现精确管理且面临“维数灾难”问题。文献[16]采用多智能体深度确定性策略梯度算法，在连续动作空间中实现了 MPIES 精细化调度。然而，全局 Critic 网络需要获取全部园区运行数据，与实际情况中多主体交互存在数据壁垒问题相矛盾。此外，随着多个不同利益主体园区在协同训练过程中不愿共享本地数据，数据孤岛和隐私安全的问题日益突出。在此背景下，联邦强化学习作为分布式训练框架，凭借避免共享原始数据即可实现多智能体协同训练的优势，逐步应用于多智能体场景。文献[17]将深度 Q 网络强化学习算法与联邦学习技术结合，在离散状态和动作空间中打破了数据壁垒的问题，实现了居民社区的能量优化管理模型，但同时带来高维度灾难问题。文献[18]将 PIES 划分为多个供能子系统，采用具有连续决策能力的 DDPG 算法结合联邦学习框架，实现了电热综合能源系统的优化调度，避免了离散空间的维数灾问题。然而上述研究均未考虑多智能体的异构性，致使全局策略难以适配个体需求。为此，文献[19]考虑到每栋建筑都高度个性化且不愿共享数据，采用分类联邦学习算法构建了不同类型的暖通空调系统容量评估模型。文献[20]则通过匹配学习使训练不足的智能体向成熟智能体学习，以适应不同能源枢纽需求模式。尽管现有研究对异构性进行了有益探索，但 MPIES 协同调度领域仍面临隐私保护与个性化调度需求的互斥难题。此外，现有优化调度研究多以运行成本最优为目标函数，系统运行中的环境成本并未得到充分考虑。随着碳交易市场政策机制的推广与完善[9]，为进一步促进多能互补与低碳调度目标的深度融合，文献[21]在 MPIES 中引入了碳交易机制，虽然在一定程度上降低了碳排放量，但其对碳排放量的控制能力依然有限，难以有效激励多园区协同调度环境下的节能减排。因此，本文考虑阶梯式碳交易机制，以期实现更高效协同的低碳经济调度。

综上所述，本文提出一种面向 MPIES 的个性化联邦强化学习(personalized federated reinforcement learning，Per-FedRL)低碳经济调度模型。首先，构建基于改进 DDPG 算法的单园区日前调度模型。该模型综合考虑电、热能源消耗、分时电价和阶梯式碳交易机制，通过在 DDPG 神经网络的输入端嵌入低维向量捕捉各园区特定的动作与状态空间特征，即能量转换和碳排放等运行特性，确保个性化调度需求得以满足。继而，为解决多园区模型聚合过程中的隐私安全和多主体异构问题，将单园区 DDPG 模型拓展至个性化联邦强化学习框架，该框架通过在本地维持嵌入向量独立更新，并在通信周期内聚合中心仅聚合平均除嵌入向量以外的模型参数，避免了直接交互各园区运行数据的情况，实现隐私保护与多主体差异化特征保留，从而有效避免模型同质化，提升整体系统的鲁棒性与泛化能力。最后，通过算例仿真验证了所构建的低碳经济调度模型的有效性。

1　多园区综合能源系统低碳经济调度模型

1.1　基于 P2P 能源交易的多园区综合能源系统

本文研究的 MPIES 由 n 个独立的 PIES、一个 P2P 能源交易市场、上级电网、天然气网和区域碳交易市场组成，结构如图 1 所示。

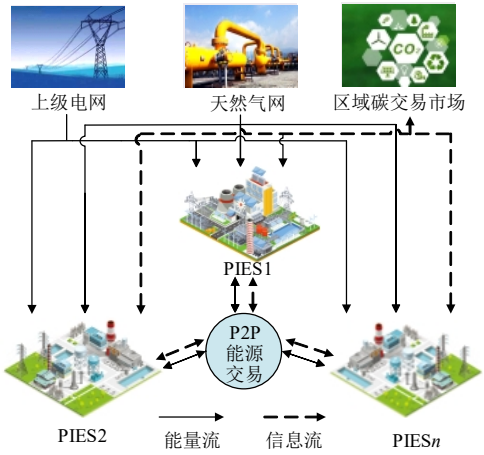

图 1　多园区综合能源系统结构

Fig. 1　Multi-park integrated energy system structure

在该结构下，PIES 可通过上级电网购售电能，并从气网购买天然气作为燃料来实现内部自治。同时，每个园区也可通过 P2P 能源交易实现园区间的能量交易，满足多园区的多能互补需求。此外，n 个园区还通过区域碳交易市场进行碳排放额度交易，激励各自实现低碳运行。为确保各园区的信息安全，避免内部运行数据泄露，园区间通过可信的 P2P 能源市场实时传递交易电量和内部交易电价，并通过可信的区域碳交易市场传递购售碳排放额度。每个园区配备智能电表和智能插头，为模型训练提供基础数据输入，通过协同训练保证各自内部能量平衡，实现最优低碳经济运行。

1.2　P2P 能源交易市场

P2P 能源交易市场基于中间市场利率(mid-market-rate，MMR)定价方案[22]。每个园区只需要向 P2P 市场提供自己的总售能量或总购能量数据，无需提供详细的运行数据，避免了隐私泄露风险。因此，无论对于 PIES 买家或是卖家，都能充分激励合作参与园区间能源交易。与相对较高的电网购买价格和电网销售价格相比，MMR 计划下的 PIES 可以分别受益于较低的本地购买价格和较高的本地销售价格。将每一个 PIES 的净需求/发电定义为 $P_{in}^{i,t}$，其值取决于该园区内的电力器件组合，数值为正代表净需求大于发电，为负代表净需求小于发电。将整个多园区的净需求/发电定义为 P_{mn}^{t}：

$$P_{mn}^{t}=\sum_{i\in N}P_{in}^{i,t} \tag{1}$$

将参与 P2P 交易的多园区净需求 P_{md}^{t} 和发电量 P_{mg}^{t} 定义为

$$\begin{cases}P_{md}^{t}=\sum\limits_{i\in\Omega^{d}}P_{in}^{i,t}\\ P_{mg}^{t}=\sum\limits_{i\in\Omega^{g}}P_{in}^{i,t}\end{cases} \tag{2}$$

式中：$\Omega^{d}=\{i\in N:P_{in}^{i,t}>0\}$ 和 $\Omega^{g}=\{i\in N:P_{in}^{i,t}<0\}$ 分别是 P2P 能源交易市场中作为买家和卖家的 PIES 集合。

MMR 定价方案将主电网分时电价 $\lambda_{\mathrm{MG}}^{t}$ 的平均值 $(\lambda_{\mathrm{MG+}}^{t}+\lambda_{\mathrm{MG-}}^{t})/2$ 作为本地购电价格 $\lambda_{\mathrm{local+}}^{t}$ 和本地售电价格 $\lambda_{\mathrm{local-}}^{t}$，表示为中间价格 $\lambda_{\mathrm{local}}^{t}$。由于多园区间的净需求 P_{md}^{t} 和发电量 P_{mg}^{t} 难以平衡，因此，不平衡的电量 P_{mn}^{t} 必须通过主电网以上网分时电价进行交易，并且按比例分配成本或收益，从而导致 $\lambda_{\mathrm{local-}}^{t}$ 和 $\lambda_{\mathrm{local-}}^{t}$ 进行相应的调整[23]：

$$\begin{cases}\lambda_{\mathrm{local+}}^{t}=\begin{cases}\lambda_{\mathrm{local}}^{t}, & \text{if } P_{mn}^{t}\le 0\\ (\lambda_{\mathrm{local}}^{t}\,|P_{mg}^{t}|+\lambda_{\mathrm{MG+}}^{t}P_{mn}^{t})/P_{md}^{t}, & \text{if } P_{mn}^{t}>0\end{cases}\\ \lambda_{\mathrm{local-}}^{t}=\begin{cases}\lambda_{\mathrm{local}}^{t}, & \text{if } P_{mn}^{t}\ge 0\\ (\lambda_{\mathrm{local}}^{t}P_{md}^{t}+\lambda_{\mathrm{MG-}}^{t}\,|P_{mn}^{t}|)/|P_{mg}^{t}|, & \text{if } P_{mn}^{t}<0\end{cases}\end{cases} \tag{3}$$

出清公式(3)表明，当园区之间供需完全匹配时，本地购售电价格均等于中间价格 $\lambda_{\mathrm{local}}^{t}$；当园区之间供不应求时，本地购电价格将达到上限，高于中间价格 $\lambda_{\mathrm{local}}^{t}$，而本地售电价格仍等于中间价；当园区之间供大于求时，本地售电价格将达到下限，低于中间价格 $\lambda_{\mathrm{local}}^{t}$，本地购电价格仍等于中间价。

1.3　基于阶梯式碳交易的 PIES 日前调度

考虑多主体异构性导致每个 PIES 的内部产能设备、用能特性等存在差异，因此，本节以设备种类最丰富的 PIES 进行介绍。同时，为激励每个 PIES 积极参与降碳行动，本文采用了一种借鉴分时电价理念的阶梯式碳交易市场，该市场在区域内进行交易，类似于上级电网。其通过在不同的碳排放水平区间设定阶梯式碳价，鼓励园区优化生产以减少碳排放。PIES 整体结构如图 2 所示。

PIES 主要包括以下 3 部分：1）产能设备主要包括热电联产机组(combined heat and power，CHP)、电热锅炉、燃气锅炉、太阳能板等设备；2）储能设备为蓄电池；3）用能负荷包括电、热等固定负荷和可控负荷。

1.3.1　目标函数

考虑 MPIES 的不同利益，以计及碳交易的最小化日运行成本为目标，包括购能成本 C_{E}^{t}、P2P 交易成本 C_{P}^{t}、阶梯式碳交易成本 $C_{\mathrm{CO_2}}^{t}$ 和储能设备等

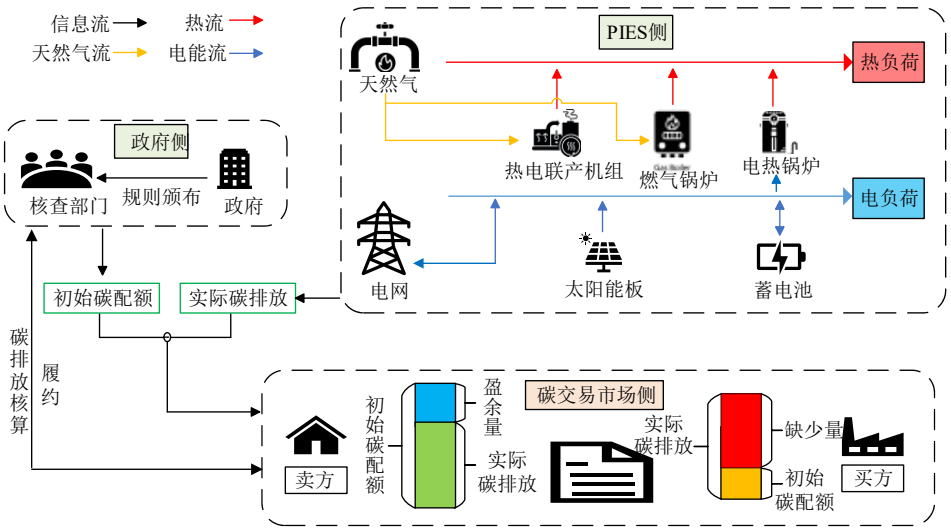

图 2　园区综合能源系统结构

Fig. 2　Comprehensive energy system structure of the park

运维成本 C_m^t，单 PIES 优化目标函数表示为

$$\min F=\sum_{t=1}^{T}(C_E^t+C_P^t+C_{\mathrm{CO_2}}^t+C_m^t) \tag{4}$$

购能成本 C_E^t 包含主网购电和购气成本，表示为

$$C_E^t=\sum_{t=1}^{T}(\lambda_{\mathrm{MG}}P_{\mathrm{MG}}^t+\lambda_{\mathrm{NG}}(\frac{P_{\mathrm{CHP}}^t}{\eta_{\mathrm{CHP}}H_{\mathrm{NG}}}+\frac{H_{\mathrm{GB}}^t}{\eta_{\mathrm{GB}}H_{\mathrm{NG}}})) \tag{5}$$

式中：λ_{MG}、λ_{NG} 为 t 时刻的主电网分时电价和天然气的单位热量价格，单位分别为元、元/(kW·h)。

P2P 交易成本表示为

$$C_p^t=\lambda_{\mathrm{local}}^t P_{\mathrm{trade}}^t \tag{6}$$

式中：P_{trade}^t 为 PIES 在 P2P 能源交易平台实际交易的电量。

为更好地反映碳排放量对总体运行成本的影响，进一步限制碳排放，本文将参与碳交易的实际碳排放交易量合理划分为多个区间，具体碳交易成本表示为

$$C_{\mathrm{CO_2}}=\begin{cases}\lambda E_{\mathrm{IES},t}, & E_{\mathrm{IES},t}\le l\\ \lambda(1+\alpha)(E_{\mathrm{IES},t}-l)+\lambda l, & l\le E_{\mathrm{IES},t}\le 2l\\ \lambda(1+2\alpha)(E_{\mathrm{IES},t}-2l)+\lambda(2+\alpha)l, & 2l\le E_{\mathrm{IES},t}\le 3l\\ \lambda(1+3\alpha)(E_{\mathrm{IES},t}-3l)+\lambda(3+3\alpha)l, & E_{\mathrm{IES},t}\ge 3l\end{cases} \tag{7}$$

式中：$E_{\mathrm{IES},t}$ 为实际参与碳市场的交易量，由公式(11)求得；λ 为基础碳价；l 为碳排放区间长度；α 为碳价增长率。

储能设备等运维和折旧成本表示为

$$C_m^t=\sum_{n\in\Gamma}\sum_{t=1}^{T}{}_n^t P_n^t \tag{8}$$

式中：Γ 为设备集合，$\Gamma=\{$CHP,电锅炉，燃气锅炉，蓄电池$\}$；ϖ_n^t 表示第 n 种设备的单位电能维护成本；P_n^t 表示第 n 种设备在 t 时刻的输出功率。

1.3.2　碳配额分配和实际碳排放量

对于初始碳配额的分配，由于电热锅炉通过电力驱动，因此，忽略其碳排放。所构建模型中排放源主要包括：主网购电(来源于燃煤机组发电)、燃气锅炉、CHP 机组[24]。采用无偿配额方式分配，具体计算公式如下：

$$\begin{cases}E_{\mathrm{IES}}=E_{\mathrm{buy}}+E_{\mathrm{CHP}}+E_{\mathrm{GB}}\\ E_{\mathrm{buy}}=\chi_e\sum_{t=1}^{T}P_{\mathrm{buy}}^t\\ E_{\mathrm{CHP}}=\chi_g\sum_{t=1}^{T}(P_{\mathrm{CHP},e}^t+P_{\mathrm{CHP},h}^t)\\ E_{\mathrm{GB}}=\chi_g\sum_{t=1}^{T}P_{\mathrm{GB},h}^t\end{cases} \tag{9}$$

式中：E_{buy}、E_{CHP}、E_{GB} 分别代表上级购电、CHP 机组、燃气锅炉初始碳配额量；χ_e、χ_c、χ_g 分别代表燃煤机组、CHP 机组和燃气锅炉单位有功碳配额系数；P_{buy}^t、$P_{\mathrm{CHP},e}^t$、$P_{\mathrm{CHP},h}^t$ 和 P_{GB}^t 分别代表 t 时段上级主网购电量、CHP 机组发电功率、CHP 机组输出热功率和燃气锅炉输出热功率，T 为 24h。

园区实际运行过程中产生的碳排放量[25]如下式所示：

$$\begin{cases}E_{\mathrm{IES},a}=E_{\mathrm{buy},a}+E_{\mathrm{ng},a}\\ E_{\mathrm{buy},a}=\sum_{t=1}^{T}(a_1+b_1\cdot P_{\mathrm{buy}}^t+c_1\cdot(P_{\mathrm{buy}}^t)^2)\\ E_{\mathrm{ng},a}=\sum_{t=1}^{T}(a_2+b_2\cdot P_{\mathrm{ng}}(t)+c_2\cdot(P_{\mathrm{ng}}^t)^2)\\ P_{\mathrm{ng}}^t=P_{\mathrm{CHP},e}^t+P_{\mathrm{CHP},h}^t+P_{\mathrm{GB},h}^t\end{cases} \tag{10}$$

式中：$E_{\text{buy},a}$、$E_{\text{ng},a}$分别为上级电网购电和燃气设备(CHP 和燃气锅炉)的实际碳排放量；P_{ng}^{t}为 t 时刻以燃气为原料的供能机组的实际等效输出功率；a_1、b_1、c_1和a_2、b_2、c_2分别为上级电网中燃煤机组和以燃气为原料的供能机组的碳排放计算系数。

通过上述公式(8)和(9)的计算之后，利用下式求得参与到碳市场的实际交易量：

$$E_{\text{IES},t}=E_{\text{IES},a}-E_{\text{IES}} \tag{11}$$

1.3.3　设备模型及约束条件

PIES 中包含有 CHP 机组、燃气锅炉、电热锅炉和蓄电池 4 类设备模型，通过协调运行满足园区运行需求。

CHP 机组作为核心设备，以天然气作为输入燃料，以较高的转换效率同时输出热能和电能，满足系统多能需求。其发电功率和输出热功率的计算公式如下：

$$\begin{cases}P_{\text{CHP}}^{t}=v_{\text{CHP}}^{t}H_{\text{NG}}\eta_{\text{CHP}}\\ H_{\text{CHP}}^{t}=\beta P_{\text{CHP}}\eta_{\text{HE}}\end{cases} \tag{12}$$

式中：P_{CHP}^{t}和H_{CHP}^{t}分别代表 CHP 机组在 t 时刻的发电功率和输出热功率，单位为 kW；v_{CHP}^{t}代表 CHP 机组在 t 时刻的天然气消耗量，单位为 m^3/h；H_{NG}为天然气的热值，单位为 $\text{kW}\cdot\text{h}/\text{m}^3$；$\eta_{\text{CHP}}$代表 CHP 机组的发电效率，为常数；$\beta$ 为 CHP 机组的输出热功率与输出电功率的比值，称为热电比；η_{HE}代表 CHP 机组余热回收效率，为常数。

在 CHP 机组运行的过程中，考虑其发电功率输出上下限约束为

$$P_{\text{CHP}}^{\min}\le P_{\text{CHP}}^{t}\le P_{\text{CHP}}^{\max} \tag{13}$$

对于园区中以天然气作为输入原料，单独提供热能的燃气锅炉，输出热能满足系统部分热能需求，其输出热功率和上下限约束为

$$\begin{cases}H_{\text{GB}}^{t}=v_{\text{GB}}^{t}H_{\text{NG}}\eta_{\text{GB}}\\ H_{\text{GB}}^{\min}\le H_{\text{GB}}^{t}\le H_{\text{GB}}^{\max}\end{cases} \tag{14}$$

式中：v_{GB}^{t}代表燃气锅炉在 t 时刻的天然气消耗量，单位为 m^3/h；H_{NG}为天然气的热值，单位为 $\text{kW}\cdot\text{h}/\text{m}^3$；$\eta_{\text{GB}}$为燃气锅炉的发热效率；$H_{\text{GB}}^{\min}$和$H_{\text{GB}}^{\max}$分别代表燃气锅炉输出热功率的最小值和最大值。

对于系统中以电能作为输入，单独提供热能的电热锅炉，其输出热功率和上下限约束为

$$\begin{cases}H_{\text{EB}}^{t}=P_{\text{EB}}^{t}\eta_{\text{EB}}\\ H_{\text{EB}}^{\min}\le H_{\text{EB}}^{t}\le H_{\text{EB}}^{\max}\end{cases} \tag{15}$$

式中：P_{EB}^{t}为电热锅炉在 t 时刻为产热而消耗的电功率，单位为 kW；η_{EB}为电热锅炉的发热效率，为常数；$H_{\text{EB}}^{\min}$和$H_{\text{EB}}^{\max}$分别代表电热锅炉输出热功率的最小值和最大值。

通过合理配置蓄电池，实现平抑可再生能源出力波动、保障供需平衡、在峰谷电价下实现峰谷套利的作用。因此选择典型蓄电池模型，满足以下公式：

$$\begin{cases}B_{\text{soc}}^{t}=B_{\text{soc}}^{t-1}-\dfrac{\eta_{\text{BA}}P_{\text{BA}}^{t}\Delta t}{Q_{e\max}}\\ \eta_{\text{BA}}=\begin{cases}\eta_{\text{dis}}, & P_{\text{BA}}\ge 0\\ 1/\eta_{\text{ch}}, & P_{\text{BA}}\le 0\end{cases}\end{cases} \tag{16}$$

式中：B_{soc}^{t}代表蓄电池在 t 时刻的储能状态；P_{BA}^{t}代表蓄电池在 t 时刻的运行功率，为正时代表处于放能状态，为负时代表处于充能状态；$Q_{e\max}$代表蓄电池的最大容量，单位为 $\text{kW}\cdot\text{h}$；η_{BA}代表蓄电池的充放电系数。同时考虑到实际运行，设置合理的约束条件保证蓄电池在运行过程中安全稳定，延长使用寿命，具体约束如下所示：

$$\begin{cases}B_{\text{soc}}^{\min}\le B_{\text{soc}}^{t}\le B_{\text{soc}}^{\max}\\ P_{\text{BA}}^{\min}\le P_{\text{BA}}^{t}\le P_{\text{BA}}^{\max}\end{cases} \tag{17}$$

此外，为了保证系统内部能源平衡，还需要考虑热电功率平衡约束，具体包括电功率平衡、天然气、外购电供应平衡：

$$\begin{cases}P_{\text{MG}}^{t}+P_{\text{pv}}^{t}+P_{\text{BA}}^{t}+P_{\text{CHP}}^{t}-P_{\text{EB}}^{t}+P_{\text{trade}}^{t}=P_{\text{load}}^{t}\\ H_{\text{CHP}}^{t}+H_{\text{GB}}^{t}+H_{\text{EB}}^{t}=H_{\text{load}}^{t}\\ v_{\text{NG}}^{t}=v_{\text{CHP}}^{t}+v_{\text{GB}}^{t}\end{cases} \tag{18}$$

式中：P_{MG}^{t}、P_{pv}^{t}、P_{load}^{t}和P_{trade}^{t}分别代表系统在 t 时刻与上级主网交易电量、光伏出力，电负荷和在 P2P 能源交易市场交易的电能，单位均为 kW，其中，P_{MG}^{t}为正时表示该园区从上级电网购买电能，为负时表示园区向上级电网售电；H_{load}^{t}代表 t 时刻的热负荷，单位为 kW；v_{NG}^{t}代表 t 时刻供应给系统的天然气总量，其中购售电量和天然气供应量还满足如下约束：

$$\begin{cases}P_{\text{MG}}^{\min}\le P_{\text{MG}}^{t}\le P_{\text{MG}}^{\max}\\ v_{\text{NG}}^{\min}\le v_{\text{NG}}^{t}\le v_{\text{NG}}^{\max}\end{cases} \tag{19}$$

2　基于 Per-FedRL 的 MPIES 低碳经济调度模型及求解

2.1　马尔可夫决策过程建模

深度强化学习方法通过智能体对动态环境的不断试错来学习最优策略，本文将智能体定义为每个 PIES。常见的 MDP 过程通常包括元组(S,A,P_{sa},R,γ)，其中 S 表征状态集合，A 表征动作

集合，$P_{sa}: S \times A \to P(S)$ 是状态转移概率的集合，R 是奖励函数，γ 表示折扣因子，用于在决策过程中平衡未来可能获得奖励与当前奖励之间的关系。在一个有限时间范围 T 内的 MDP 过程中，智能体在某个状态 $s_t \in S$ 时基于策略 μ 选择动作 $a_t \in A$ 执行，并获得奖励 r 。

1）状态空间：每个 PIES 智能体从环境中观测到的状态信息由全局观测状态和局部观测状态两类组成，在每个决策周期内，定义状态空间如下：

$$\begin{cases} s_i^t = [s_{i,l}^t, s_g] \\ s_{i,l}^t = [P_{i,\mathrm{pv}}^t, P_{i,\mathrm{load}}^t, H_{i,\mathrm{load}}^t, B_{i,\mathrm{soc}}^{t-1}] \\ s_g = [\lambda_{\mathrm{MG}+}^t, \lambda_{\mathrm{MG}-}^t, \lambda_{\mathrm{local}+}^{t-1}, \lambda_{\mathrm{local}-}^{t-1}, t] \end{cases} \tag{20}$$

式中：局部观测状态中的 $P_{i,\mathrm{pv}}^t$ 、$P_{i,\mathrm{load}}^t$ 、$H_{i,\mathrm{load}}^t$ 和 $B_{i,\mathrm{soc}}^{t-1}$ 分别代表第 i 个园区的光伏出力、电负荷、热负荷和蓄电池荷电状态；全局观测状态中的 $\lambda_{\mathrm{MG}+}^t$ 、$\lambda_{\mathrm{MG}-}^t$ 、$\lambda_{\mathrm{local}+}^t$ 、$\lambda_{\mathrm{local}-}^t$ 、t 分别代表上网购电电价、上网售电电价、上一时刻本地购电电价、上一时刻本地售电电价和调度时间。值得注意的是，每个智能体的全局观测空间都是相同的，而局部观测空间则在一定程度上反映多主体异构性。

2）动作空间：PIES 的日前调度目标是确定设备最佳出力情况。其中，$P_{i,\mathrm{CHP}}^t$ 确定后 $H_{i,\mathrm{CHP}}^t$ 可由公式(11)得到，$P_{i,\mathrm{EB}}^t$ 确定后 $H_{i,\mathrm{EB}}^t$ 可由公式(14)求得，进一步，$H_{i,\mathrm{GB}}^t$ 通过公式(17)中的热平衡等式约束求得。已知 $P_{i,\mathrm{BA}}^t$ 和 $P_{i,\mathrm{trade}}^t$ 后，$P_{i,\mathrm{MG}}^t$ 由公式(17)中的电平衡等式约束求得。因此，动作空间定义如下：

$$a_i^t = [P_{i,\mathrm{CHP}}^t, P_{i,\mathrm{EB}}^t, P_{i,\mathrm{BA}}^t] \tag{21}$$

式中：$P_{i,\mathrm{CHP}}^t$ 、$P_{i,\mathrm{EB}}^t$ 和 $P_{i,\mathrm{BA}}^t$ 分别代表第 i 个园区在 t 时刻的 CHP 机组输出电功率、电热锅炉消耗电功率和蓄电池运行功率。

3）奖励函数：作为指导策略 μ 更新的指标，能够引导智能体在约束范围内采取动作，使得运行成本最小化，同时考虑到蓄电池日常运行平衡是一个强时间耦合约束，智能体无法保证时刻满足该约束，故定义如下：

$$R_i^t(s_i^t, a_i^t) = -(\varepsilon_1 F + \varepsilon_2 F_p) \tag{22}$$

式中：ε_1 为控制成本的缩放系数；ε_2 为约束惩罚的缩放系数；F_p 为约束惩罚成本，表示为

$$F_p = -\sum(\Delta B_{i,\mathrm{soc}}^t) \tag{23}$$

式中：$\Delta B_{i,\mathrm{soc}}^t$ 为蓄电池储电/放电超额量，通过引入蓄电池惩罚项，进一步防止蓄电池的过充过放动作，延长使用寿命。

2.2 基于 Per-FedRL 的问题求解

在第 1.1 节，基于 P2P 能量交易的 MPIES 低碳经济调度模型通过每个 PIES 智能体不断调整内部设备的出力，并利用 P2P 能量交易市场实现效益最大化，但多智能体 MDP 环境中的环境不稳定和信息受限等问题导致传统的多智能体强化学习过度依赖智能体间的真实数据共享[26]，从而限制了其实际应用。因此，本文提出个性化联邦框架应对隐私泄露问题。具体而言，Per-FedRL 方法包括两个具体的实施细节：1）单园区采用基于改进的 DDPG 方法的 Actor-Critic 架构，Actor 网络输出连续动作，Critic 网络修正 Actor 网络的权重，嵌入向量感知本地运行特性。2）将个性化联邦框架与改进的 DDPG 方法结合，在保护隐私的同时实现分布式控制。其具体架构如图 3 所示。

图 3 中的算法，包括以下 3 个主要的内部步骤：

1）采用并行方式。各园区采用高斯嵌入初始化各自嵌入向量，并从环境中获取状态作为输入，Actor 网络输出一组动作并作用于 PIES，根据 1.3 节 PIES 模型计算内部能量流进行调度管理，将当前状态 s_i^t，动作 a_i^t，奖励 $R_i^t(s_i^t, a_i^t)$ 和下一时刻状态 s_i^{t+1} 保存到经验回放池中。

2）各 PIES 智能体从经验回放池中抽取一小批经验，训练并更新局部模型。

3）经过一段通信间隔后，聚合中心收集并聚合平均 n 个本地 DDPG 模型参数，得到全局模型参数 $\omega_g = (\theta_g, \phi_g)$ 并下发至各个园区更新。

更具体地说，Per-FedRL 训练过程包括单园区 DDPG 模型的训练和模型的全局聚合。n 个 PIES 智能体构建各自的数据集并训练自己的 Actor-Critic 网络模型，嵌入向量则在本地以梯度下降法不断更新。在每个通信周期内，聚合中心将每个园区除嵌入向量以外的模型参数 $\omega_i = (\theta_i, \phi_i)$ 进行传输和聚合，以估计所有 PIES 的全局模型 $\omega_g = (\theta_g, \phi_g)$，并将 ω_g 广播到所有 PIES 取代 ω_i。此后，本地 DDPG 模型不断训练并聚合，直到损失曲线或收益曲线收敛。

2.3 差异化感知的改进 DDPG 算法

对于单园区日前调度问题，采用改进的 DDPG 算法进行本地训练，如图 3 中步骤 2 所示部分。该算法基于 Actor-Critic 体系结构，包括两个在线网络和两个目标网络。为有效解决在联邦聚合平均过程中异构园区的 DDPG 模型生成全局单一策略的问题，对 DDPG 算法进行了如下改进：1）初始化阶段，增加高斯嵌入向量 $e_{\mathrm{actor},i}$ 和 $e_{\mathrm{critic},i}$ 作为本地 DDPG

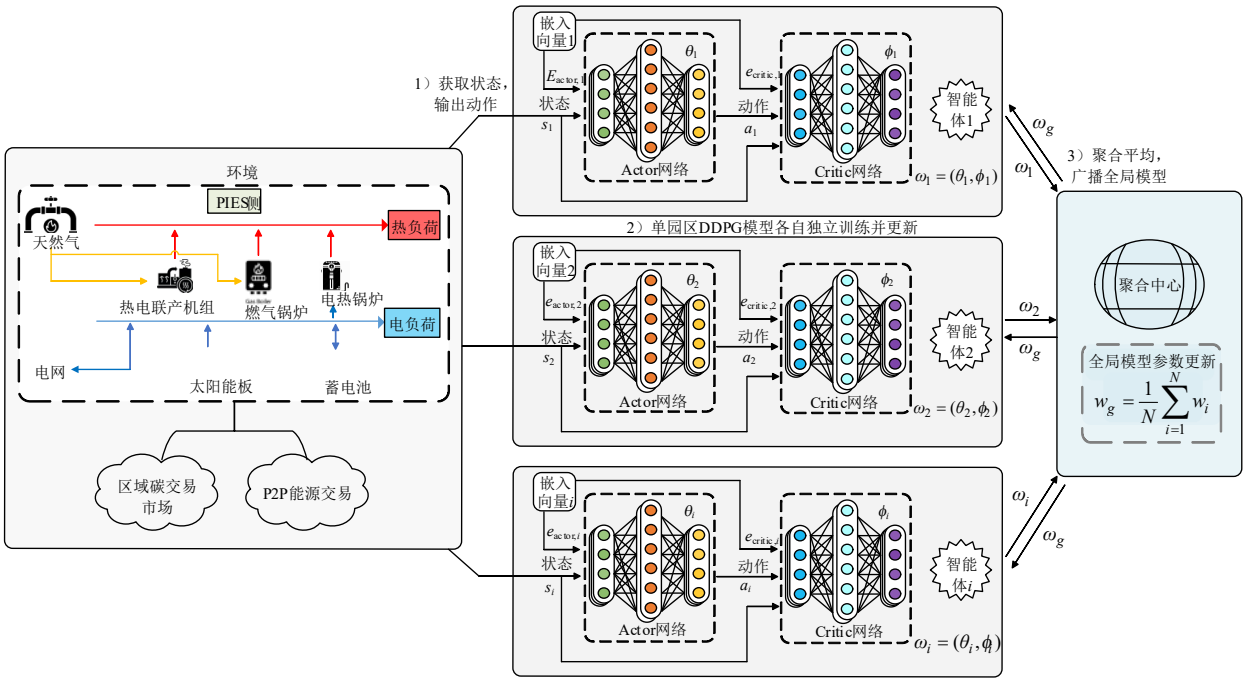

图 3　MPIES 的个性化联邦学习框架

Fig. 3　Federal learning framework for multi-park integrated energy systems

模型的特有参数。2）状态预处理阶段，嵌入向量根据不同的网络类型与状态空间拼接成为全新的状态空间：

$$\begin{cases} s_{\text{actor},i}=(e_{\text{actor},i};s_i) \\ s_{\text{critic},i}=(e_{\text{critic},i};s_i) \end{cases} \tag{24}$$

训练过程中，在线 Actor 网络 $\mu_{\theta_i}(s_{\text{actor},i})$ 以 θ_i 作为参数，将智能体 i 的状态空间 $s_{\text{actor},i}$ 作为输入值，输出动作 a_i。在线 Critic 网络 $Q_i(s_{\text{critic},i},a_i)$ 以 ϕ_i 作为参数，将智能体 i 的状态空间 $s_{\text{critic},i}$ 和动作 a_i 作为输入值，输出 Q 值的标量估计值来执行策略评估任务。

具体而言，采用经验回放池存储从环境中获得的经验元组 $e_i^t=(s_i^t,a_i^t,r_i^t,s_i^{t+1})$，抽取小批量 $(e_i^t)_{m=1}^M$ 进行训练，计算在线 Critic 网络的均方差分误差：

$$L(_i)=\mathrm{E}[(r_i+\gamma Q_i^{'}(s_{\text{critic},i}^{'},\mu_{\theta_i}^{'}(s_{\text{actor}}^{'}))-Q_i(s_{\text{critic},i},a_i)^2)\ ]=$$

$$\frac{1}{M}\sum_{m=1}^{M}[(y_m-Q(s_{\text{critic},i},a_m))^2] \tag{25}$$

式中：$Q_{\phi_i}^{'}(\cdot)$ 和 $\mu_{\theta_i}^{'}(\cdot)$ 分别代表目标 Critic 网络和目标 Actor 网络。随后采用确定性策略梯度理论求得在线 Actor 网络的策略梯度：

$$\nabla_{\theta_i}J(\mu_{\theta_i})=\frac{1}{M}\sum_{m=1}^{M}[\nabla_{\theta_i}\mu_{\theta_i}(s_{\text{actor},i})\cdot$$

$$\nabla_{a_m}Q_i(s_{\text{critic},i},a_j)|_{a_m=\mu_{\theta_i}(s_{\text{actor},i})}] \tag{26}$$

然后，按照以下公式对 4 个网络进行更新：

$$\theta_i\leftarrow\theta_i+\alpha_\theta\cdot\nabla_{\theta_i}J(\mu_{\theta_i}),\ \theta_i^{'}\leftarrow\tau\theta_i+(1-\tau)\theta_i^{'} \tag{27}$$

$$\phi_i\leftarrow\phi_i-\alpha_\phi\cdot\nabla_{\phi_i}L(\phi_i),\ \phi_i^{'}\leftarrow\tau\phi_i+(1-\tau)\phi_i^{'} \tag{28}$$

式中：α_θ、α_ϕ 代表在线 Actor 网络和在线 Critic 网络的学习率；τ 为目标网络软更新率。$e_{\text{actor},i}$ 和 $e_{\text{critic},i}$ 使用学习率为 α_e 的梯度下降法更新：

$$\begin{cases} e_{\text{actor},i}\leftarrow e_{\text{actor},i}-\alpha_e\nabla_{e_{\text{actor}}}J \\ e_{\text{critic},i}\leftarrow e_{\text{critic},i}-\alpha_e\nabla_{e_{\text{critic}}}L \end{cases} \tag{29}$$

3　算例仿真

为验证所提框架的有效性，采用图 1 所示的 3 个园区综合能源系统结构进行仿真实验。电负荷、热负荷来源于文献[27]，光伏发电数据来源于文献[28]。采用 200 天的负荷和光伏数据进行训练和测试。系统调度时长设为 24h，时间间隔为 1h。各园区设备模型参数见附录 A 表 A1，不同能源设备类型也在一定程度上体现了多主体异构型。设备运行参数见附录 A 表 A2 所示。

本文采用分时电价，详见附录 A 表 A3 所示。PIES 与上级主网购售电功率范围为[−500, 1000] kW 之间，其余参数见附录 A 表 A4 所示。天然气固定价格为 0.45 元 / (kW·h)，碳排放基础价格设定为 150 元/t[29]。

实验采用 Python3.7 和 PyTorch1.2.0 实现，设置本地 DDPG 模型每训练 100 次后进行聚合平均，每个实验取 10 次结果的均值。改进的 DDPG 算法

中，Actor 网络和嵌入向量的学习率为 0.0005，Critic 网络的学习率为 0.005。奖励函数折扣因子取 0.95，目标网络的软更新率为 0.001。Actor 和 Critic 网络均含有两层隐藏层，采用 ReLU 函数激活和 Adam 优化器更新网络权重。

3.1 个性化与普通联邦强化学习训练结果对比

首先验证所提 Per-FedRL 对 3 个园区模型的奖励函数收敛性，以及 Per-FedRL 方法对奖励函数收敛曲线的改善效果。主要分为以下两种场景：

1）园区设备见附录 A 表 A1 所示。

2）3 个相同设备的园区。

表 1 不同算法运行结果

Table 1 Results of different algorithms

方法	碳交易成本/元	总成本/元	碳排放量/t	训练时间/min
FedRL	8217.04	40134.29	318.827	54.5
Per-FedRL	7897.26	38459.14	315.778	26.2

如附录 A 图 A1 所示，场景 1 中应用 Per-FedRL 的各个园区模型奖励函数都在迭代训练过程中实现了奖励函数的收敛，显著优于 FedRL 方法，后者在相同的有限训练回合内并未收敛。而表 1 则通过从数据中随机抽取 10 天测试并延长训练回合后得到对比结果，其表明 Per-FedRL 方法训练效率更高，且相比于 FedRL 方法总成本降低了 4.17%，碳排放量降低 0.96%。Per-FedRL 通过嵌入向量有效学习并适应了园区间设备差异，使模型在训练迭代过程中逐渐学习到了更适合自身园区的调度策略，最终的奖励值更高且波动更小，表明 Per-FedRL 方法更强的稳定性。

如附录 A 图 A2 所示，场景 2 中使用 FedRL 与 Per-FedRL 方法最终得到的奖励值基本一致，但 FedRL 方法在训练 750 回合之后才逐渐收敛，收敛速度远低于 Per-FedRL 方法。即使设备一致性减小了模型训练的难度，Per-FedRL 方法仍体现出更高的灵活性和收敛效率。通过比较相同设备的园区和不同设备的园区奖励函数收敛曲线，反映了嵌入向量通过本地独立更新的方式，捕捉并学习到了不同园区的运行特性，提升了模型在多主体异构 MPIES 场景下的适应能力。

3.2 调度结果分析

利用历史数据对模型进行离线训练后，保存得到的网络模型用于 MPIES 的动态低碳经济调度。为验证所提方法的鲁棒性和泛化性，选择典型日(1 月 15 日)作为测试日进行调度管理，该天的特点是热需求较大，光伏发电量少。

通过在个性化联邦框架下协同训练 MPIES 模型，并对 PIES 使用改进的 DDPG 算法实现实时调度，最终得到各个园区的调度动作 $P^t_{i,\mathrm{CHP}}$、$P^t_{i,\mathrm{EB}}$ 和 $P^t_{i,\mathrm{BA}}$，以及由相应计算得到的 $H^t_{i,\mathrm{CHP}}$、$H^t_{i,\mathrm{EB}}$、$H^t_{i,\mathrm{GB}}$、$P^t_{i,\mathrm{trade}}$ 和 $P^t_{i,\mathrm{MG}}$。

图 4(a)显示，配备 CHP 机组的园区主要通过 CHP 机组发电，在电负荷较小且满足自身调度需求的同时，充分利用机组冗余容量参与 P2P 能源交易市场以获取额外收益。仅装配电热锅炉和燃气锅炉的园区 3 则主要通过上级电网和 P2P 能源市场满足电能需求，在 P2P 市场中有盈余电量时，优先选择从该市场购电，如 02:00—04:00、09:00—11:00，降低运行成本和碳排放。3 个园区的蓄电池主要在谷电价时段进行充电，并在电价高峰或平时段放电，同时在光伏出力较大时储存电能，发挥削峰填谷和消纳光伏的作用，如 00:00—06:00 时段的低谷电价充电，10:00—12:00 光伏出力较高时继续储能。图 4(b)显示，配备电热锅炉的园区主要在低谷电价以电热锅炉耗电制热为主，如 00:00—02:00。园区 1 和 2 在电价平时段或高峰时段，如 10:00—15:00，降低 CHP 机组出力并增加燃气锅炉购气制热，原因是燃气价格作为一个固定值，要低于平时或峰值电价。图 4(c)显示，园区 1 的碳配额始终低于实际排放，且在 15:00 左右碳排放降至最低，此时 CHP 机组出力最小，表明适度出力有助于降低碳排放成本。园区 2 在 15:00 左右通过碳市场获利，此时 CHP 机组出力最小，表明碳交易受 CHP 机组、燃气锅炉和上级电网的影响依次递减。园区 3 从上级电网购电较多且燃气锅炉出力较多时，增加碳配额购入，而在 P2P 能量市场交易时段，如 22:00—00:00，将产生盈余的碳配额在碳市场中出售。

3.3 不同场景对比分析

针对本文提出的 MPIES 低碳经济调度模型，本节设置 3 个场景，选取典型日进行对比分析阶梯式碳交易和 P2P 市场对 MPIES 运行的影响：

场景 1：在综合能源系统调度中考虑定价碳交易机制，优化目标考虑碳交易成本。

场景 2：在综合能源系统调度中考虑阶梯式碳交易机制，优化目标考虑碳交易成本。

场景 3：在综合能源系统调度中考虑阶梯式碳交易机制，同时加入多园区间能源交易的 P2P 市场，优化目标考虑碳交易成本。

基于表 2 的实验结果，场景 3 总成本为 38997.87 元，较场景 1 和场景 2 分别降低 6.57%和 4.78%；碳排放量为 311.761t，较场景 1 和场景 2 分别减少 2.95%和 1.66%。具体而言，场景 3 与场景 2 相比，

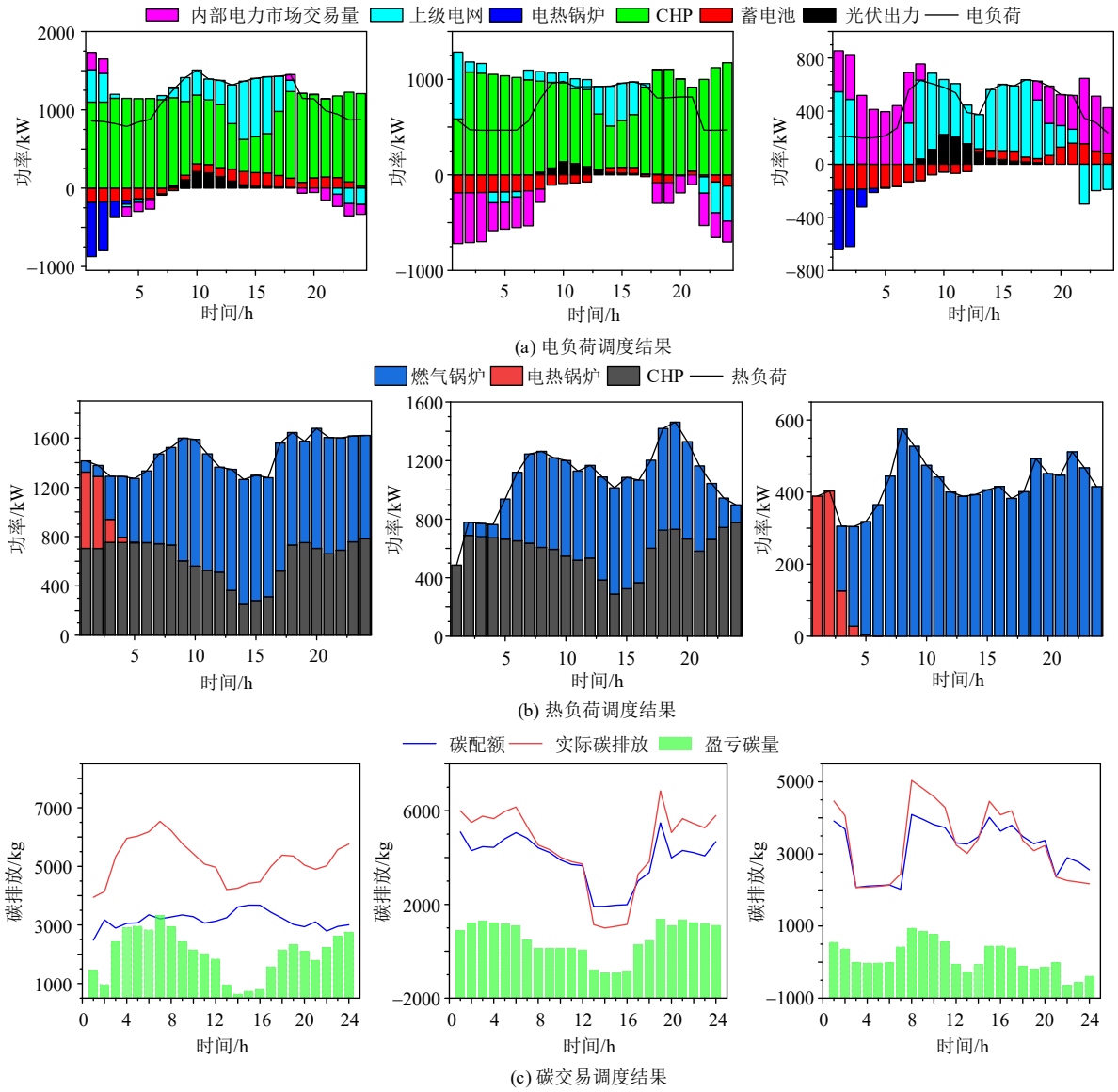

图 4　典型日调度结果

Fig. 4　Typical daily scheduling results

表 2　不同场景下运行结果对比

Table 2　Comparison of operating results in different scenarios

场景	购电成本/元	购气成本/元	碳交易成本/元	总成本/元	碳排放量/t
1	27794.51	3868.17	8478.65	41741.33	321.249
2	26547.45	4497.15	8199.74	40954.34	317.027
3	24876.56	4276.24	7895.07	38997.87	311.761

碳交易成本降低了 2.65%，购电成本降低了 6.29%。这表明在 P2P 能源交易后，各园区间通过直接电能交易，利用中间定价机制有效降低了购电成本，减少了对外部能源的依赖，控制了整体碳排放量。相比场景 1，场景 2 碳交易成本降低 3.29%，购气成本增加 16.26%，而购电成本则降低 4.50%，表明阶梯式碳交易机制提升了 MPIES 降碳积极性。

综上，场景 3 在经济性和环保性方面均表现出显著的优越性。阶梯式碳交易机制和 P2P 能源交易市场的协同作用实现了 MPIES 低碳经济调度目标。

3.4　敏感性分析

在个性化联邦强化架构中，本地 DDPG 模型更新多次之后在聚合中心交换模型参数，为探究嵌入向量的维度和通信间隔对本文所提方法性能影响，分别对其进行不同取值测试。图 5 对比了这两个超参数对典型日平均成本的影响。

结果显示，嵌入向量维度和通信间隔的选择会影响训练结果。随着嵌入向量维度的增加，平均运行成本呈现非线性波动：低维度的嵌入向量不足以捕获并学习到系统的复杂运行特性，过高维度的嵌入向量增加了计算复杂度，导致模型性能下降。因此，选定嵌入向量维度为 6。通信间隔增大时，平均运行成本先下降后上升，过短的通信间隔可能会

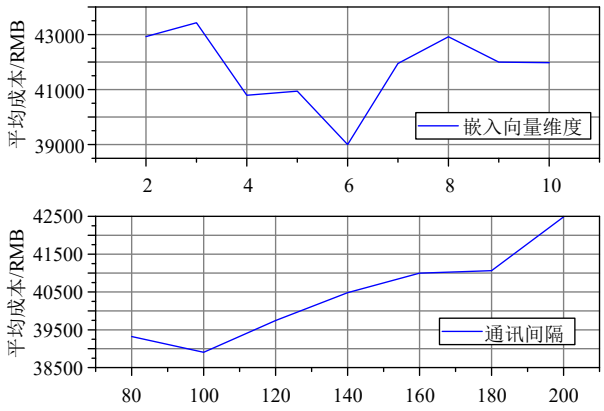

图 5　不同超参数效果对比

Fig. 5　Comparison of different hyperparameter effects

增加通信开销并扰乱局部学习，过长的通信间隔则可能导致模型参数优化不同步，影响全局优化效果。

4　结论

考虑到多主体异构、数据隐私和实时决策等问题，本文提出了一种基于个性化联邦强化学习的 MPIES 低碳经济调度方法。得出了以下结论：

1）受益于阶梯式碳交易市场，测试日调度结果与不考虑碳交易机制和考虑定价碳交易机制相比，碳排放量分别减少了 3.08%和 1.80%。

2）个性化联邦框架下的 MPIES 低碳调度模型通过聚合平均除嵌入向量以外的模型参数实现了协同训练，避免了运行数据的交互，实现了隐私保护。

3）嵌入向量的不断更新使模型准确学习了各园区的运行特性，保证了个性化调度需求，优于传统的联邦强化学习方法。

附录见本刊网络版(http://www.dwjs.com.cn/CN/1000-3673/current.shtml)。

参考文献

[1] MU Chenggang，DING Tao，ZHU Shanying，et al. A decentralized market model for a microgrid with carbon emission rights[J]. IEEE Transactions on Smart Grid，2023，14(2)：1388-1402.

[2] SHI Zhengkun，YANG Yongbiao，XU Qingshan，et al. A low-carbon economic dispatch for integrated energy systems with CCUS considering multi-time-scale allocation of carbon allowance[J]. Applied Energy，2023，351：121841.

[3] WANG Yudong，HU Junjie，LIU Nian. Energy management in integrated energy system using energy-carbon integrated pricing method[J]. IEEE Transactions on Sustainable Energy，2023，14(4)：1992-2005.

[4] LIU Chunling，WANG Chengfu，YAO Wenliang，et al. Distributed low-carbon economic dispatching for multiple park-level integrated energy systems based on improved shapley value[J]. IEEE Transactions on Industry Applications，2024，60(6)：8088-8102.

[5] 骆钊，卢涛，马瑞，等. 可再生能源配额制下多园区综合能源系统优化调度[J]. 电力自动化设备，2021，41(4)：8-14.
LUO Zhao，LU Tao，MA Rui，et al. Optimal scheduling of multi-park integrated energy system under renewable portfolio standard[J]. Electric Power Automation Equipment，2021，41(4)：8-14(in Chinese).

[6] MA Zhoujun，ZHOU Yizhou，ZHENG Yuping，et al. Distributed robust optimal dispatch of regional integrated energy systems based on ADMM algorithm with adaptive step size[J]. Journal of Modern Power Systems and Clean Energy，2024，12(3)：852-862.

[7] YANG Zhao，HU Junjie，AI Xin，et al. Transactive energy supported economic operation for multi-energy complementary microgrids[J]. IEEE Transactions on Smart Grid，2021，12(1)：4-17.

[8] 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(in Chinese).

[9] 杨挺，刘豪，王静，等. 基于深度强化学习的园区综合能源系统低碳经济调度[J]. 电网技术，2024，48(9)：3604-3613.
YANG Ting，LIU Hao，WANG Jing，et al. Deep reinforcement learning-based low-carbon economic dispatch of park integrated energy system[J]. Power System Technology，2024，48(9)：3604-3613(in Chinese).

[10] 高冠中，杨胜春，郭晓蕊，等. 深度强化学习在含分布式柔性资源的电网优化调度中的应用研究综述[J]. 中国电机工程学报，2024，44(16)：6385-6403.
GAO Guanzhong，YANG Shengchun，GUO Xiaorui，et al. A review of research on the application of deep reinforcement learning in optimization dispatch of power grids with distributed flexible resources[J]. Proceedings of the CSEE，2024，44(16)：6385-6403(in Chinese).

[11] XU Xuesong，XU Kai，ZENG Ziyang，et al. Collaborative optimization of multi-energy multi-microgrid system：A hierarchical trust-region multi-agent reinforcement learning approach[J]. Applied Energy，2024，375：123923.

[12] DU Guodong，ZOU Yuan，ZHANG Xudong，et al. Deep reinforcement learning based energy management for a hybrid electric vehicle[J]. Energy，2020，201：117591.

[13] 聂欢欢，张家琦，陈颖，等. 基于双层强化学习方法的多能园区实时经济调度[J]. 电网技术，2021，45(4)：1330-1336.
NIE Huanhuan，ZHANG Jiaqi，CHEN Ying，et al. Real-time economic dispatch of community integrated energy system based on a double-layer reinforcement learning method[J]. Power System Technology，2021，45(4)：1330-1336(in Chinese).

[14] 彭刘阳，孙元章，徐箭，等. 基于深度强化学习的自适应不确定性经济调度[J]. 电力系统自动化，2020，44(9)：33-42.
PENG Liuyang，SUN Yuanzhang，XU Jian，et al. Self-adaptive uncertainty economic dispatch based on deep reinforcement learning [J]. Automation of Electric Power Systems，2020，44(9)：33-42(in Chinese).

[15] ZHOU Hao，ARAL A，BRANDIĆ I，et al. Multiagent Bayesian deep reinforcement learning for microgrid energy management under communication failures[J]. IEEE Internet of Things Journal，2022，9(14)：11685-11698.

[16] 杨照，黄少伟，陈颖. 基于多智能体强化学习的多园区综合能源系统协同优化运行研究[J]. 电工电能新技术，2021，40(8)：1-10.
YANG Zhao，HUANG Shaowei，CHEN Ying. Research on cooperative optimal operation of multi-park integrated energy system based on multi agent reinforcement learning[J]. Advanced Technology of Electrical Engineering and Energy，2021，40(8)：1-10(in Chinese).

[17] 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)：5535-5549．
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)，5535-5549(in Chinese)．

[18] 王金锋，王琪，任正某，等．基于联邦强化学习的电热综合能源系统能量管理策略[J]．上海交通大学学报，2024，58(6)：904-915．
WANG Jinfeng，WANG Qi，REN Zhengmou，et al．Energy management strategy of integrated electricity-heat energy system based on federated reinforcement learning[J]．Journal of Shanghai Jiaotong University，2024，58(6)：904-915(in Chinese)．

[19] WANG Zhenyi，YU Peipei，ZHANG Hongcai．Privacy-preserving regulation capacity evaluation for HVAC systems in heterogeneous buildings based on federated learning and transfer learning[J]．IEEE Transactions on Smart Grid，2023，14(5)：3535-3549．

[20] CHEN Minghao，SUN Yi，XIE Zhiyuan，et al．An efficient and privacy-preserving algorithm for multiple energy hubs scheduling with federated and matching deep reinforcement learning[J]．Energy，2023，284：128641．

[21] DONG Yingchao，WUKEN E，ZHANG Hongli，et al．Bi-level coordinated operation optimization of multi-park integrated energy systems considering categorized demand response and uncertainty：a unified adaptive robust optimization approach[J]．Renewable Energy，2025，241：122331．

[22] QIU Dawei，YE Yujian，PAPADASKALOPOULOS D，et al．Scalable coordinated management of peer-to-peer energy trading：a multi-cluster deep reinforcement learning approach[J]．Applied Energy，2021，292：116940．

[23] QIU Dawei，XUE Juxing，ZHANG Tingqi，et al．Federated reinforcement learning for smart building joint peer-to-peer energy and carbon allowance trading[J]．Applied Energy，2023，333：120526．

[24] LI Fei，WANG Dong，GUO Hengdao，et al．Distributionally Robust Optimization for integrated energy system accounting for refinement utilization of hydrogen and ladder-type carbon trading mechanism[J]．Applied Energy，2024，367：123391．

[25] GAO Jianwei，MENG Qichen，LIU Jiangtao，et al．Thermoelectric optimization of integrated energy system considering wind-photovoltaic uncertainty，two-stage power-to-gas and ladder-type carbon trading[J]．Renewable Energy，2024，221：119806．

[26] LEE S，CHOI D H．Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources[J]．IEEE Transactions on Industrial Informatics，2022，18(1)：488-497．

[27] Ong, S., Clark, N. (2014). Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States. [Data set]. Open Energy Data Initiative (OEDI). National Renewable Energy Laboratory. https://doi.org/10.25984/1788456.

[28] PFENNINGER S，STAFFELL I．Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data[J]．Energy，2016，114：1251-1265．

[29] ZHANG Xiliang，HUANG Xiaodan，ZHANG Da，et al．Research on the pathway and policies for China’s energy and economy transformation toward carbon neutrality[J]．Journal of Management World，2022，38(1)：35-66．



张东升

在线出版日期：2025-03-28。
收稿日期：2024-12-27。

作者简介：
张东升(1999)，男，硕士研究生，研究方向为人工智能在园区综合能源系统中的应用，E-mail：1061312957@qq.com；
黄媛(1974)，女，副教授，硕士生导师，通信作者，研究方向为电力系统的稳定与运行分析、配电网的储能及故障恢复策略研究，E-mail：yuanhuang@scu.edu.cn。

（责任编辑　赵梓含）

附录 A

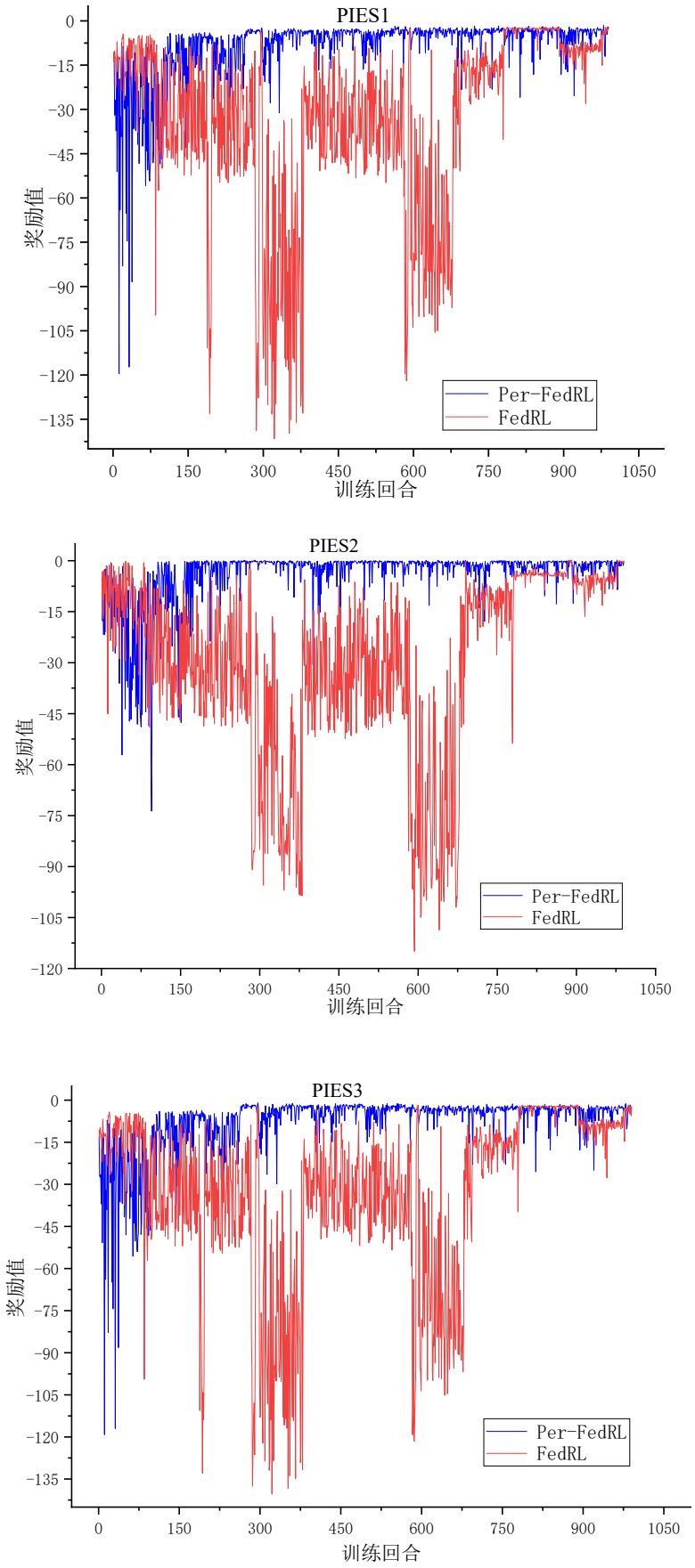

图 A1　不同设备园区奖励函数图

Fig. A1　Reward function diagram of each energy supply subsystem

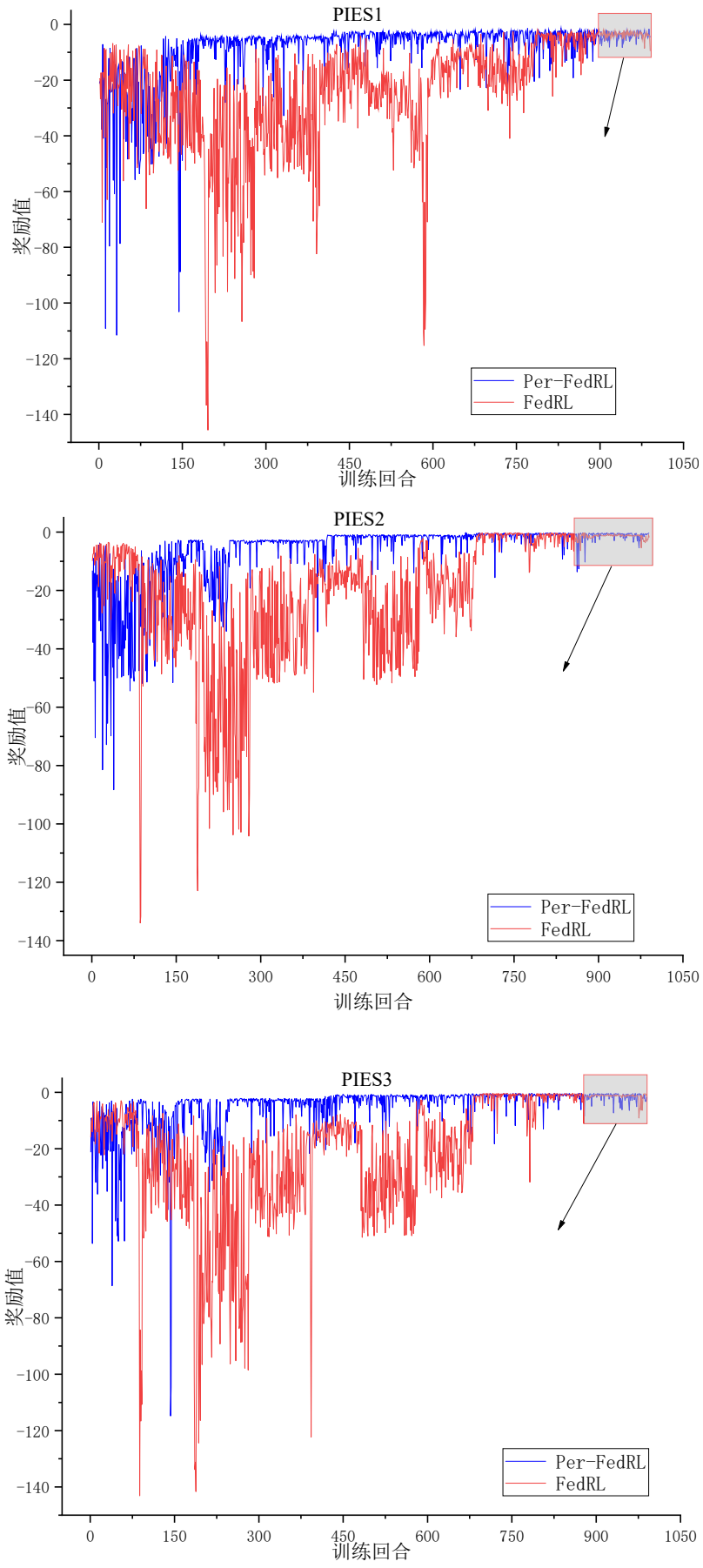

图 A2　相同设备园区奖励函数图

Fig. A2　The same equipment park reward function diagram

表 A1　园区设备类型

Table 1　Types of intelligent building equipment

场景	能源设备类型
园区 1	光伏、CHP 机组、蓄电池、燃气锅炉、电热锅炉
园区 2	光伏、CHP 机组、蓄电池、燃气锅炉
园区 3	光伏、蓄电池、燃气锅炉、电热锅炉

表 A2　园区中设备模型参数

Table 2　Equipment model parameters in intelligent building body

设备类型	最小电/热功率/kW	最大电/热功率/kW	运维单价/(元/kW)
热电联产机组	0	1400	0.053
电热锅炉	0	1200	0.03
燃气锅炉	0	1200	0.03
蓄电池	−200	200	0.04

表 A3　分时电价

Table 3　Time-of-use electricity price

时段	购电电价(元/kW·h)	售电电价(元/kW·h)
峰时段(12:00—16:00、19:00—22:00)	0.92	0.5
平时段(08:00—12:00，16:00—19:00、22:00—24:00)	0.54	0.2
谷时段(00:00—8:00)	0.38	0

表 A4　其他参数

Table 4　Other system parameters

参数	值	参数	值
η_{CHP}	0.32	$B_{soc}^{\min}$	0.2
β	1.25	$B_{soc}^{\max}$	0.8
η_{HE}	0.72	a_1, a_2	36，3
η_{EB}	0.93	b_1, b_2	−0.38，−0.004
η_{EB}	0.82	c_1, c_2	0.0034，0.001
χ_c	0.424	χ_e	0.798
χ_g	0.385		