

# 基于安全强化学习的电力市场发电商低碳报价策略

朱晓晴, 刘智伟<sup>†</sup>, 雷浩, 周长远

(华中科技大学人工智能与自动化学院, 武汉市 430074)

**摘要:** 随着双碳目标的推进, 电力市场发电商在追求经济效益的同时, 面临着日益严格的碳排放配额约束. 针对现有深度强化学习方法难以有效处理硬性物理或环境约束的问题, 提出一种基于拉格朗日松弛的软演员-评论家算法的发电商报价策略优化方法. 首先, 构建考虑直流潮流约束的电力市场出清模型, 并在此基础上建立发电商的约束马尔可夫决策过程模型, 目标是在满足瞬时碳排放配额的前提下最大化长期收益; 其次, 引入原始-对偶更新机制, 通过自适应调整拉格朗日乘子, 将碳排放约束转化为动态惩罚项, 引导智能体在探索过程中自动寻找满足环保合规性的最优报价策略; 最后, 基于 IEEE 3 节点和 IEEE 30 节点系统的仿真结果表明, 所提出方法不仅能够有效逼近纳什均衡, 而且能在严格遵守碳排放配额的同时通过策略性报价实现经济性.

**关键词:** 电力市场; 软演员-评论家算法; 拉格朗日松弛; 碳排放配额; 纳什均衡; 安全强化学习

**中图分类号:** TP273 **文献标志码:** A

**DOI:** 10.13195/j.kzyjc.2025.1180

**引用格式:** 朱晓晴, 刘智伟, 雷浩, 等. 基于安全强化学习的电力市场发电商低碳报价策略 [J]. 控制与决策, xxxx, x(x): xxxx-xxxx.

## Low-carbon bidding strategies for power generators in electricity markets based on safe reinforcement learning

ZHU Xiao-qing, LIU Zhi-wei<sup>†</sup>, LEI Hao, ZHOU Chang-yuan

(School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China)

**Abstract:** With the accelerating implementation of China's dual-carbon targets, power market generators are required to pursue economic benefits while complying with increasingly stringent carbon emission quota constraints. To address the difficulty that existing deep reinforcement learning (DRL) methods face in effectively handling hard physical or environmental constraints, this paper proposes a generator bidding strategy optimization method based on a Lagrangian-relaxed Soft Actor-Critic algorithm (SACLag). First, a market clearing model incorporating DC power flow constraints is constructed, upon which a constrained Markov decision process (CMDP) for each generator is formulated. The objective is to maximize long-term cumulative revenue while satisfying instantaneous carbon emission quota limits. Second, a primal-dual update mechanism is introduced to adaptively adjust the Lagrangian multipliers, transforming the carbon emission constraint into a dynamic penalty term. This mechanism guides the agent to autonomously explore environmentally compliant optimal bidding strategies. Finally, simulation results on the IEEE 3-bus and IEEE 30-bus systems demonstrate that the proposed method not only approximates the Nash equilibrium effectively but also achieves the trade-off between economic returns and environmental compliance by strategic bidding.

**Keywords:** electricity market; soft actor-critic; Lagrangian relaxation; carbon emission allowances; nash equilibrium; safe reinforcement learning

## 0 引言

电力市场是实现电能资源优化配置与系统经济运行的重要机制之一, 其核心目标是通过市场竞争

机制实现电能的合理分配和价格形成<sup>[1]</sup>. 发电公司作为独立的市场主体, 在没有其他竞争对手信息的情况下独立做出战略决策<sup>[2]</sup>. 与此同时, 独立系统运营

收稿日期: 2025-11-14; 录用日期: 2026-02-27.

基金项目: 国家自然科学基金项目 (U24A20268, 62373162, 5003184086).

责任编委: 邢兰涛.

<sup>†</sup>通信作者. E-mail: zwliu@hust.edu.cn.

商作为中立的市场规则执行者,根据发电商报价和负荷需求执行市场出清,确定节点电价及机组出力分配.不同发电商之间通过调整报价竞争市场份额与收益,这一过程本质上构成了一个不完全信息的静态博弈过程<sup>[3]</sup>.

由于发电商的竞标收益依赖于其他竞争对手的策略,而这些策略在不完全信息市场中未知且动态变化,发电商需通过持续的市场交互与收益反馈调整竞标策略,该过程最终收敛至纳什均衡点(Nash Equilibrium Point, NEP)<sup>[4-6]</sup>,即在该状态下任何单个发电商均无法通过单方面改变策略获得更高利润.围绕纳什均衡的求解,现有研究多采用基于博弈论的建模方法,如 Cournot 模型、Bertrand 模型以及供应函数均衡模型等<sup>[7]</sup>,并通过数学规划方法(如对角化算法<sup>[8]</sup>)或启发式算法<sup>[9-10]</sup>求解.然而,这类方法普遍存在两个不足:(1)仅考虑非策略投标,而忽略了动态投标策略修改过程;(2)依赖于所有发电商掌握全局信息(包括其他发电商决策)的强假设,而这些信息在实际电力市场中往往属于不可观测的.因此,获得的纳什均衡必定与真实市场的纳什均衡有很大偏差,难以严格反映不完全信息的电力市场.

在此背景下,深度强化学习(Deep Reinforcement Learning, DRL)为电力市场博弈研究提供了一种新颖的解决方案<sup>[11]</sup>.DRL通过试错探索和与电力市场环境的交互来学习最优交易策略<sup>[12]</sup>.现有研究引入了基于Q学习的强化学习方法来解决第一个限制<sup>[13]</sup>,但隐私保护问题仍然没有得到解决.而隐私保护已通过深度确定性策略梯度方法有效解决<sup>[14-15]</sup>,但由于其确定性策略特性,在多发电商非平稳博弈环境中易出现探索不足和训练不稳定的问题.相比之下,软演员-评论家(Soft Actor-Critic, SAC)算法<sup>[16-17]</sup>通过最大熵强化学习机制在连续动作空间中实现更稳定、鲁棒的策略学习,已被成功应用于电力市场交易机制设计.然而,在全球碳减排日益严格的背景下,发电商的报价决策必须同时满足经济性与碳排放约束,使得发电商的报价决策过程本质上属于带约束的马尔可夫决策过程(Constrained Markov Decision Process, CMDP)<sup>[18-19]</sup>.SAC作为针对标准MDP设计的算法,难以直接处理此类硬性碳排放约束.因此亟需设计安全强化学习方法,在满足碳约束的前提下实现最优竞标策略学习.

综上所述,本文以不完全信息条件下的碳约束电力市场为研究对象,面向发电商竞标决策问题,结合最大熵强化学习的探索优势与拉格朗日松弛的约束处理机制,构建基于Soft Actor-Critic with Lagrangian

的发电商报价模型,使发电商能够在动态博弈环境中,在满足碳排放配额约束的前提下自主学习最优竞标策略.最后,通过仿真算例验证所提出算法在实现安全交易策略的有效性.

## 1 发电商非合作博弈问题

### 1.1 符号定义

问题描述及模型定义中所使用的符号含义如下:索引号.

$g$ : 发电机索引号,  $g \in \{1, 2, \dots, n\}$ ;

$l$ : 负荷索引号,  $l \in \{1, 2, \dots, m\}$ ;

$i$ : 输电网线路节点索引;

集合.

$G$ : 发电机组集合,  $g \in G$ ;

$D$ : 负荷集合,  $g \in D$ ;

变量.

$p_{g,t}^{\text{dg}}$ : 发电商 $g$ 在时刻 $t$ 的实际出力

$p_{l,t}^{\text{load}}$ : 负荷节点 $l$ 的电力需求;

$\lambda_{i,t}$ : 节点电价;

$D_{\text{load}}^{\text{max}}$ : 最大负荷需求;

$\zeta_i$ : 为需求曲线的斜率;

$F$ : 最大线路潮流限制向量;

**PTDF**: 功率传输分布因子矩阵;

$\mathbf{p}_{g,t}^{\text{dg}}$ : 注入功率向量;

$\mathbf{p}_{l,t}^{\text{load}}$ : 负荷向量;

$\partial_{g,t}^{\text{dg}}$ : 边际成本函数的截距;

$\beta_{g,t}$ : 边际成本函数的斜率;

$\zeta_{l,t}$ : 负荷 $l$ 的需求曲线斜率;

$D_l^{\text{max}}$ : 负荷 $l$ 最大需求;

$p_{g,t}^{\text{dg,min}}$ : 发电机组 $g$ 的最小功;

$p_{g,t}^{\text{dg,max}}$ : 发电机组 $g$ 的最大功率;

决策变量.

$\mu_g$ : 发电商 $g$ 的报价截距参数;

### 1.2 问题描述

如图1所示,电力市场的框架由虚拟层和物理层组成.物理层是电力传输的基础,包括输电网,发电机组和智能电表组成.其中,发电机组负责电力生产,并通过公共耦合节点将电力传输至输电网;智能电表则扮演数据采集器的角色,实时测量发电机组的功率信息等,并将其上传给虚拟层的发电商.负荷购买电力.虚拟层则驱动市场的决策与清算过程,主要由发电商、负荷聚合商和系统运营商构成.系统运营商的职责是市场出清和与保障电网运行安全.在决策环节,发电商根据接收到的智能电表功率信息及其他市场状态信息,做出战略决策——即通过选

择一个战略变量来间接影响其供给报价函数的截距, 并将此供给报价函数信息传递给系统运营商。系统运营商作为负责市场出清的主体, 收集汇总所有发电商的供给信息, 并根据这些供给函数信息以及市场负荷需求, 并严格参照输电网的物理约束, 进行市场出清, 最终确定均衡的市场价格和每个发电机组的最终发电量。

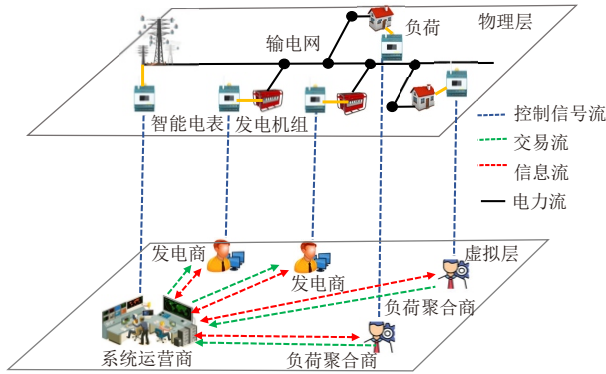


图1 电力市场框架

### 1.3 模型建立

本文考虑以发电商为决策主体的电力批发市场。发电商根据对可再生能源出力的预测结果, 在每个结算时段提交供给报价, 系统运营商在满足电网物理约束的前提下进行市场出清, 得到各发电机组的计划出力和节点电价。由于发电机出力存在预测误差, 实际出力与计划出力之间会产生偏差电量, 需要通过偏差惩罚机制进行结算, 从而使发电商的实际利润具有随机性。为此, 本文将发电商报价问题建模为一个考虑预测误差与偏差惩罚的风险规避型报价优化问题。本文建立的模型是基于如下假设。

**假设 1** 发电商的成本函数假设是输出功率的二次函数。

**假设 2** 发电商的投标策略通过截距参数化来建模, 假设供应函数的斜率与边际成本的斜率相同。

**假设 3** 假设每个负荷的电力消费可以被建模为具有线性需求的曲线。

**假设 4** 市场被设定为不完全信息博弈, 每个发电商仅能观测局部状态, 不能观测竞争对手的成本和当前动作。

**假设 5** 发电商被假设为理性主体。

**假设 6** 假设市场出清过程由直流最优潮流原理驱动, 严格遵守输电网的物理约束。

在上述假设基础上, 相关函数的具体形式如下。

发电机的发电成本 $C(p_g^{\text{dg}})$ 为发电功率的二次函数, 表达式如公式 (1) 所示。

$$C(p_{g,t}^{\text{dg}}) = \partial_g^{\text{dg}} p_{g,t}^{\text{dg}} + \frac{1}{2} \beta_g^{\text{dg}} (p_{g,t}^{\text{gd}})^2. \quad (1)$$

发电机的边际成本的表达式如公式 (2) 所示。

$$\sigma_g^{\text{dg}}(p_{g,t}^{\text{gd}}) = \partial_g^{\text{dg}} + \beta_g^{\text{dg}} p_{g,t}^{\text{gd}}. \quad (2)$$

对于任意的发电商, 它向市场提交的供给曲线是一个关于其计划输出功率 $p_{g,t}^{\text{gd}}$ 的线性函数, 表达式如 (3) 所示。

$$\rho_g^{\text{dg}}(p_{g,t}^{\text{gd}}) = \mu_{g,t}^{\text{dg}} + \beta_g^{\text{dg}} p_{g,t}^{\text{gd}}. \quad (3)$$

式中,  $\mu_{g,t}^{\text{dg}}$  是供给函数的截距, 作为策略变量, 发电商通过调整该变量, 可以直接影响其向市场提交的报价策略, 从而影响市场出清结果及自身的收益。增加 $\mu_{g,t}^{\text{dg}}$  会使供给曲线整体向上平移, 这意味着对于任意给定的发电量 $p_{g,t}^{\text{gd}}$ , 发电商要求更高的价格, 从而增加利润。反之, 降低 $\mu_{g,t}^{\text{dg}}$  会使供给曲线整体向下平移, 这意味着对于任意给定的发电量 $p_{g,t}^{\text{gd}}$ , 发电商会要求更低的价格, 旨在争取更大的市场份额或避免损失。

同时,  $\mu_{g,t}^{\text{dg}}$  需要满足以下约束:

$$0 \leq \mu_{g,t}^{\text{dg}} \leq 2\partial_g^{\text{dg}}. \quad (4)$$

对于负荷 $l$ , 消费者对功率为 $p_{l,t}^{\text{gd}}$ 的负荷 $i$ 愿意支付的价格 $\rho_l^{\text{load}}$ 的表达式如公式 (5) 所示。

$$\rho_l^{\text{load}}(p_{l,t}^{\text{load}}) = \zeta_l (p_{l,t}^{\text{load}} - D_l^{\text{max}}) \quad (5)$$

因此, 基于直流潮流模型建立市场出清模型, 如公式 (6) 所示:

$$\begin{aligned} \max_{p_{g,t}^{\text{dg}}, p_{l,t}^{\text{load}}} \quad & \sum_{l \in L} \int \rho_l^{\text{load}}(p_{l,t}^{\text{load}}) dp_{l,t}^{\text{load}} - \sum_{g \in G} C(p_{g,t}^{\text{dg}}) \\ \text{s.t.} \quad & \sum_{g \in G} p_{g,t}^{\text{dg}} - \sum_{l \in D} p_{l,t}^{\text{load}} = 0 \\ & -\mathbf{F} \leq \mathbf{PTDF}(\mathbf{p}_{g,t}^{\text{dg}} - \mathbf{p}_{l,t}^{\text{load}}) \leq \mathbf{F} \\ & p_g^{\text{dg}, \text{min}} \leq p_{g,t}^{\text{dg}} \leq p_g^{\text{dg}, \text{max}}. \end{aligned} \quad (6)$$

式中, 输电网满足了直流潮流模型、支路功率约束、发电机出力约束, 目标为社会福利最大。

每个发电商都是一个独立的商业实体, 目标是最大化自己的利润。发电商 $g$ 的利润表达式如公式 (7) 所示。

$$C_{g,t}^{\text{gd}} = \lambda_i p_{g,t}^{\text{gd}} - C(p_{g,t}^{\text{gd}}). \quad (7)$$

式中,  $\lambda_i$  是发电机组所接节点 $i$ 的节点电价, 对应于公式 (6) 有功功率平衡约束的拉格朗日乘子。

发电商除了通过调整报价截距 $\mu_{g,t}$  竞争市场份额和出清价格外, 还必须确保所有发电产生的碳排放量总和不超过分配的排放量 $C_{limit}$ 。假设发电商 $g$ 的单位发电碳排放强度为 $\xi_g$  (ton/MWh), 其排放量 $C_{g,t} = \xi_g \cdot (p_{g,t}^{\text{dg}}(\mu_{g,t}))^2 + \chi_g p_{g,t}^{\text{dg}}(\mu_{g,t}) + \delta_g$  取决于市场出清后的中标电量 $p_{g,t}^{\text{dg}}$ 。中标电量 $p_{g,t}^{\text{dg}}$ 是关于报价

策略 $\mu_{g,t}$ 的隐式函数,由系统运营商的市场出清模型决定.分配的碳排放量约束如下:

$$C_{g,t} \leq C_{\text{limit}}. \quad (8)$$

因此,发电商与系统运营商的双层优化框架如图2所示.

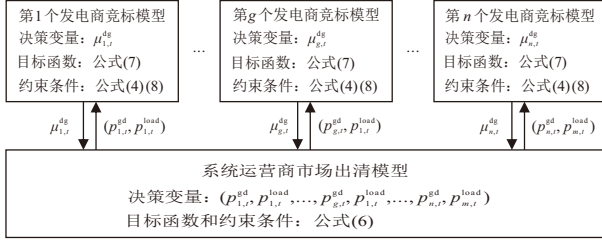


图2 发电商与系统运营商的双层优化框架

#### 1.4 发电商报价问题的马尔科夫决策过程建模

从发电商 $g$ 的视角,其受约束的马尔科夫决策由 $(\mathcal{S}_g, \mathcal{A}_g, \mathcal{O}_g, P_g, R_g, O_g, R_g^c, \gamma)$ 组成,其中:

- $\mathcal{S}_g$ 为真实状态空间,  $s_t \in \mathcal{S}_g$ 描述时段 $t$ 下电力系统和市场的完整信息.

- $\mathcal{A}_g$ 为动作空间,动作 $a_{g,t} = \mu_{g,t} \in \mathcal{A}_g$ .

- $\mathcal{O}_g$ 为观测空间,观测 $o_{g,t} \in \mathcal{O}_g$ 是发电商 $g$ 在时段 $t$ 能够获得的局部市场信息.  $o_{g,t} = (\lambda_{1,t-1}, \lambda_{2,t-1}, \dots, \lambda_{I,t-1}, D_t^{\Sigma})$ ,其中 $\lambda_{i,t-1}$ 为上一时段节点 $i$ 电价,  $D_t^{\Sigma}$ 为当前时段系统总负荷.

- $P_g(s_{t+1} | s_t, a_{g,t})$ 为状态转移概率,刻画在真实状态 $s_t$ 下,发电商 $g$ 采取动作 $a_{g,t}$ 后,在系统运营商出清和负荷不确定性的共同作用下系统转移到 $s_{t+1}$ 的概率.

- $R_{g,t}(s_t, a_{g,t}) = G_{g,t}$ 为单步回报函数.

- $R_{g,t}^c(s_t, a_{g,t}) = [\xi_g \cdot p_{g,t}^{dg}(\mu_{g,t}) - C_{\text{limit}}]^+$ 为碳排放量违背量.

- $\gamma \in [0, 1)$ 为折扣因子,用于平衡当前收益和未来收益.

## 2 Soft Actor-Critic with Lagrangian 算法

### 2.1 算法总体框架

定义 $J_R(\pi)$ 为策略 $\pi$ 下的期望累积折扣利润:

$$J_R(\pi) = \sum_{t=0}^T \mathbb{E}_{(s_t, a_t) \sim \rho_{\pi}} [\gamma^t R_g(s_t, a_{g,t})]. \quad (9)$$

定义 $J_C(\pi)$ 为策略 $\pi$ 下的期望累积碳排放量:

$$J_C(\pi) = \sum_{t=0}^T \mathbb{E}_{(s_t, a_t) \sim \rho_{\pi}} [\gamma^t R_{g,t}^c(s_t, a_t)]. \quad (10)$$

构建拉格朗日目标函数 $\mathcal{L}(\pi, \lambda)$ ,将原约束优化问题转化为如下问题:

$$\begin{aligned} \max_{\pi} \min_{\theta, \lambda} \mathcal{L}(\pi, \lambda, \theta) = & J_R(\pi) + \\ & \theta \sum_{t=0}^T \mathbb{E}[-H - \log(\pi(a_t | s_t))] + \\ & \omega_g (\overline{R}_g^c - R_{g,t}^c). \end{aligned} \quad (11)$$

其中,  $\mathcal{H}(\pi(\cdot | s_t)) = - \sum_a \pi(a | s_t) \ln \pi(a | s_t)$ 为策略在状态 $s_t$ 下的熵,  $\theta$ 为温度参数,用于权衡奖励最大化与策略随机性之间的关系.  $\omega_g$ 是发电机 $g$ 的拉格朗日乘子.  $\overline{R}_g^c$ 是碳排放量违背量上限值.

基于该目标,状态-动作价值函数 $Q^{\pi}(s, a)$ 定义为:

$$\begin{aligned} Q^{\pi}(s, a) = & \mathbb{E}_{\tau \sim \pi} \left[ \sum_{t=0}^T \gamma^t R_g(s_t, a_t, s_{t+1}) + \right. \\ & \left. \alpha \sum_{t=1}^T \gamma^t \mathcal{H}(\pi(\cdot | s_t)) \mid s_0 = s, a_0 = a \right]. \end{aligned} \quad (12)$$

### 2.2 算法设计

为了准确估计拉格朗日函数中的各项,本文采用两组独立的Critic网络:奖励评论家 $Q^{\psi}(s, a)$ 和成本评论家 $Q_C^{\phi}(s, a)$ .  $Q^{\psi}(s, a)$ 估计累积期望利润,  $Q_C^{\phi}(s, a)$ 估计累积期望碳排放.

奖励网络参数 $\psi$ 通过最小化均方差进行更新:

$$L(\psi) = \mathbb{E}_{(s_t, a_t, r_t, s_{t+1}) \sim \mathcal{D}} [(Q^{\psi}(s_t, a_t) - y_t)^2]. \quad (13)$$

其中目标值 $y_t = R_{g,t} + \gamma Q^{\psi}(s_{t+1}, a_{t+1})$ .

同理,成本网络参数 $\phi$ 通过最小化碳排放估计误差进行更新:

$$L(\phi) = \mathbb{E}_{(s_t, a_t, c_t, s_{t+1}) \sim \mathcal{D}} [(Q_C^{\phi}(s_t, a_t) - y_t^c)^2]. \quad (14)$$

其中目标成本值 $y_t^c = R_{g,t}^c + \gamma Q_C^{\phi}(s_{t+1}, a_{t+1})$ .

策略 $\pi$ 通过重参数化技巧表示为 $a_t = \tanh(\mu_{\vartheta}(s_t) + \sigma_{\vartheta}(s_t) \odot \xi)$ ,其中 $\xi \sim \mathcal{N}(0, 1)$ .利用原始-对偶法寻找拉格朗日函数的鞍点 $(\pi^*, \theta^*, \lambda^*)$ .

目标函数重写为:

$$\begin{aligned} L(\vartheta, \theta, \lambda) = & \mathbb{E}[Q^{\psi}(s_t, a_t)] + \\ & \theta(-H - \mathbb{E}[\log \pi(a_t | s_t)]) + \\ & \omega_g (C_{\text{limit}} - \mathbb{E}[Q_C^{\phi}(s_t, a_t)]). \end{aligned} \quad (15)$$

原始变量 $\vartheta_k$ 和对偶变量 $\theta$ 和 $\lambda_k$ 更新为:

$$\vartheta_{k+1} = \vartheta_k + \delta_x \nabla_{\vartheta} L(\vartheta, \theta, \lambda), \quad (16)$$

$$\theta_{k+1} = [\theta_k - \delta_{\theta} \nabla_{\theta} L(\vartheta, \theta, \lambda)]^+, \quad (17)$$

$$\omega_{g,k+1} = [\omega_{g,k} - \delta_{\omega_g} \nabla_{\omega_g} L(\vartheta, \theta, \omega_g)]^+. \quad (18)$$

其中,  $\delta_x$ ,  $\delta_{\theta}$ ,  $\delta_{\omega_g}$ 分别为对应的学习率,  $[\cdot]^+$ 表示向非负实数域的投影.通过该机制,智能体能够在严格遵守增广拉格朗日约束项的前提下,学习到利润最大

化的最优报价策略。

我们使用多个基于 SACLag 的智能体来模拟电力市场, 其中每个智能体独立地学习自己的策略, 并将其他智能体视为环境的一部分。

### 3 实例验证及结果分析

在本节中使用 SACLag 算法对 IEEE 3 节点系统进行仿真. 各节点发电商与负荷参数如表 1 所示. SACLag 的算法参数如下表 1 所示. 所有强化学习算法的代码均在 Python 环境中编写. 所有仿真均在 NVIDIA GeForce RTX 5060 Ti 上运行。

表1 各节点发电商与负荷参数表

节点		发电机组			负荷		
$i$	$g$	$\partial_{g,t}^{dg}$ (\$/MWh)	$\beta_g^m$ (\$/MWh <sup>2</sup> )	$p_g^{\max}$ (MW)	$d$	$f_d$ (\$/MWh <sup>2</sup> )	$D_d^{\max}$ (MW)
1	1	15	0.01	500	1	-0.08	500
2			—		2	-0.06	666.67
3	2	18	0.008	500			—

表2 SACLag 算法主要参数设置

算法	参数	值
SACLag	Actor网络学习率	0.0003
	Critic网络学习率	0.0003
	软更新率	0.003
	折扣因子	0.99

所有发电商的策略参数的收敛结果如图 3 所示. 可以看出, 发电商 1 和发电商 2 都收敛到均衡点. 在第 0-1000 轮, 两个智能体在探索带有随机扰动的行动空间时均表现出高方差. 第 1000-9000 轮呈现出明显的向纳什均衡值收敛的趋势, 随着探索率的降低, 方差逐渐减小. 第 9000-10000 轮, 策略趋于稳定, 探索量极少. 发电商 1 的策略变量在最后 100 轮中, 收敛准确率达到 97.81%, 发电商 2 收敛准确率为 99.83%. SACLag 算法展现出卓越的收敛性能, 成功学习了电力市场中的纳什均衡竞价策略. 并且两个发电商最终的策略变量  $\mu_g^{dg}$  都大于其边际成本截距  $\partial_{g,t}^{dg}$ , 这说明发电商行使了市场力, 进行策略性加

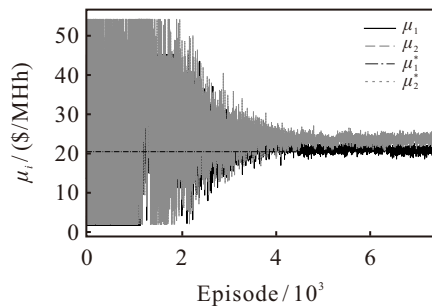


图3 IEEE3 节点的基于 SACLag 的策略变量曲线

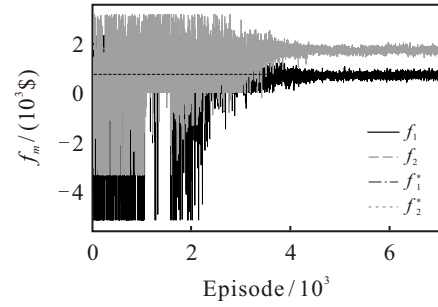


图4 IEEE3 节点的基于 SACLag 的利润曲线

价, 并形成某种程度上的合谋行为。

折扣因子  $\gamma$  反映了发电商对未来收益的重视程度. 当  $\gamma = 0$  时, 表示参与者极度不耐烦, 只关注当前时期的收益, 对未来收益几乎不予考虑; 当  $\gamma$  接近 1 时, 表示参与者十分耐心, 重视长期利益, 愿意为未来收益调整当前行为. 通过比较不同  $\gamma$  值下的结果, 可以定量分析发电商在不同“耐心”水平下可能形成的默契串谋程度及其对市场价格的影响. 理解折扣因子对市场行为的影响, 有助于市场设计者和监管者制定更有效的政策, 以促进竞争或抑制非法串谋. 因此, 设定  $\gamma = 0$ , 其运行结果如图 5 和图 6 所示; 将  $\gamma = 0$  与  $\gamma = 1$  的结果进行对比, 见表 3.

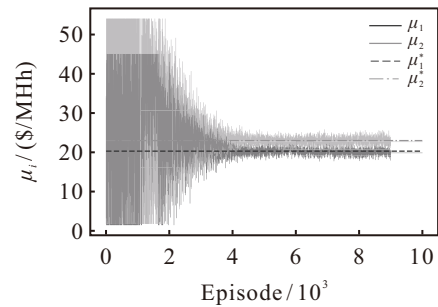


图5  $\gamma = 0$  时 IEEE3 节点的基于 SACLag 的策略变量曲线

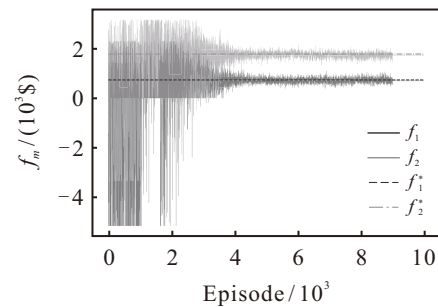


图6  $\gamma = 0$  时 IEEE3 节点的基于 SACLag 的利润曲线.

表3 不同折扣因子下发电商的策略与利润结果

$\gamma$	$\mu_g^{dg}$		$f_g$	
	$g=1$	$g=2$	$g=1$	$g=2$
$\gamma = 0$	21.13	23.09	732.77	1886.98
$\gamma = 1$	21.25	24.12	865.96	1892.51

由表 3 可知, 随着折扣因子由 0 增大到 1, 发电

商策略变量均有所上升,表明其在决策中更加倾向于采取较高出价的行为.同时,两者的利润均呈增长趋势,其中发电商1的收益提升幅度更为显著.这说明,当智能体更加重视长期回报时,模型能够获得

更高的累计收益.总体而言,折扣因子的增加促使各智能体在博弈过程中逐步形成一种兼顾长期效益的协调均衡,体现出一定的默契合谋特征.

图7表示基于 SACLag 的碳排放量的训练过程,发电机组1是煤电,排放率为  $0.8 \text{ t CO}_2/\text{MWh}$ ,发电机组2是汽轮机,排放率为  $0.4 \text{ t CO}_2/\text{MWh}$ .从训练结果来看,系统碳排放满足约束.

为了验证所提出的方法在更复杂的市场环境中的拓展性,我们测试含有6个发电商的IEEE30节点输电网的情况.图8展示了基于 SACLag 算法的IEEE 30节点系统中各发电机组出价策略的收敛过程,图9展示了基于 SACLag 算法的IEEE 30节点

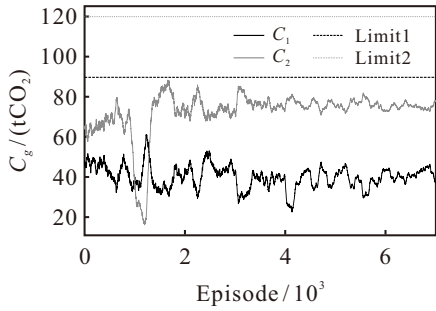


图7 IEEE3节点的基于SACLag的碳排放量

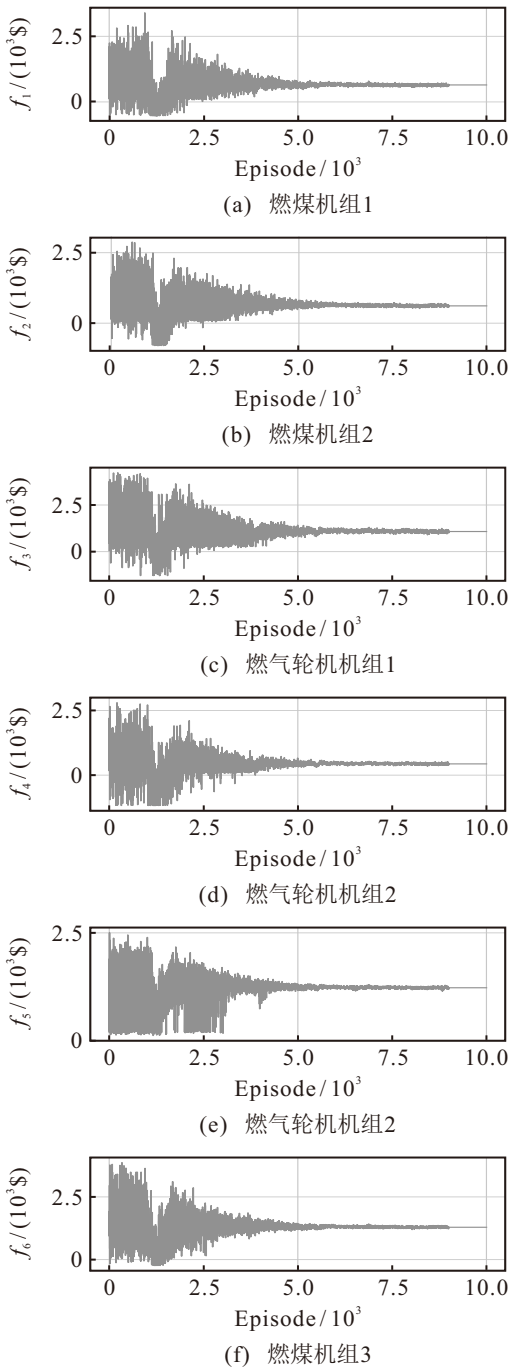


图8 IEEE30节点的基于SACLag的利润曲线

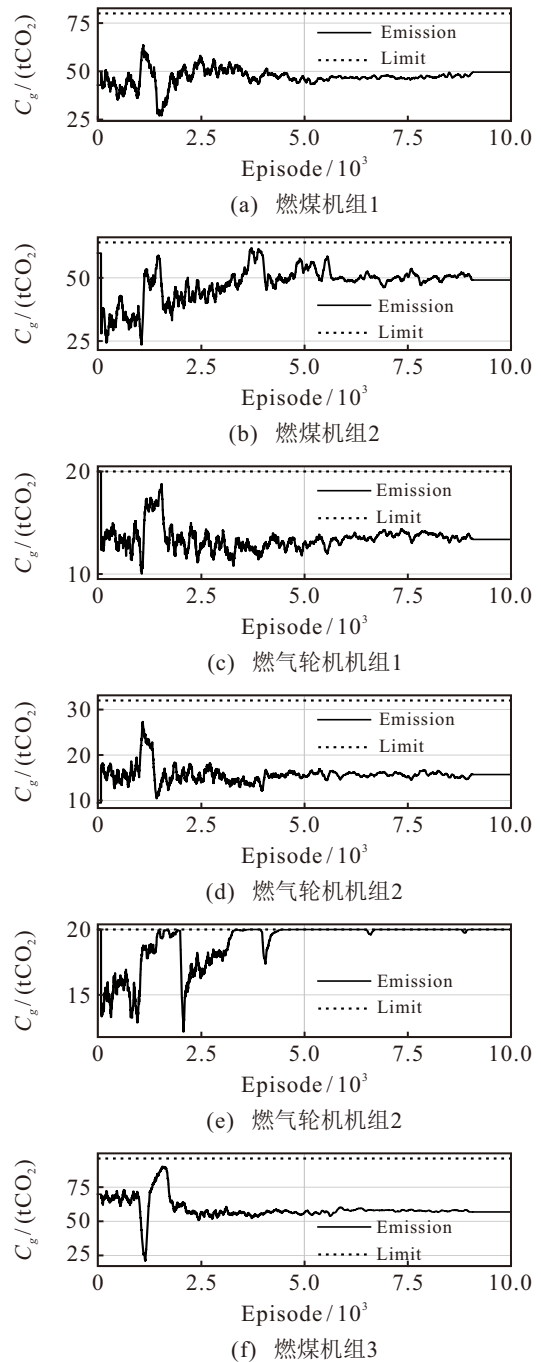


图9 IEEE30节点的基于SACLag的碳排放量

系统中各发电机组的碳排放量. 由图 8 可知, 基于 SAC 算法的 IEEE 30 节点系统中 6 台发电机的利润最终收敛至纳什均衡状态. 由图 9 可知, 所有发电机组的碳排放都满足要求.

## 4 结论

本文面向双碳目标背景下电力市场发电商的低碳竞标问题, 基于安全强化学习方法研究了不完全信息条件下发电商的策略性报价机制. 通过将碳排放配额约束引入强化学习框架, 构建了基于拉格朗日松弛的 Soft Actor-Critic(SACLag) 发电商报价优化模型, 得到以下结论:

1) 所提出的基于 SACLag 的发电商报价模型能够在不完全信息条件下刻画发电商竞标策略的动态演化过程, 揭示了碳排放配额约束对发电商报价行为和市场均衡形成的影响.

2) 通过在 Soft Actor-Critic 框架中引入原始-对偶更新机制, 实现了碳排放约束的有效嵌入, 使强化学习方法能够处理发电商面临的硬性环境约束性.

未来的研究可在以下几个方面进一步拓展: 一是将所提方法推广至多类型发电主体和更大规模市场环境, 研究多智能体交互下的低碳博弈行为; 二是引入碳交易机制和多时间尺度决策过程, 刻画配额跨期调节对竞标策略的影响.

## 参考文献 (References)

- [1] Dai T, Qiao W. Finding equilibria in the pool-based electricity market with strategic wind power producers and network constraints[J]. *IEEE Transactions on Power Systems*, 2017, 32(1): 389-399.
- [2] 刘敏, 王金环. 基于势博弈的智能电网需求侧管理问题[J]. *控制与决策*, 2024, 39(2): 545-550.  
(Liu M, Wang J H. Potential game for demand-side management of smart grids[J]. *Control and Decision*, 2024, 39(2): 545-550.)
- [3] 张启亮, 武建荣. 不完全信息下基于网络演化博弈的微电网能源交易策略[J]. *控制与决策*, DOI: 10.13195/j.kzyjc.2025.0364.  
(Zhang Q L, Wu J R. Microgrid energy trading strategy based on network evolutionary game under incomplete information[J]. *Control and Decision*, DOI: 10.13195/j.kzyjc.2025.0364.)
- [4] 熊炜, 李咸善, 邹宇, 等. 考虑碳排放和综合需求响应的电-气联合运行决策博弈[J]. *控制与决策*, 2023, 38(7): 1979-1987.  
(Xiong W, Li X S, Zou Y, et al. Decision-making game of power-gas joint operation considering carbon emission and integrated demand response[J]. *Control and Decision*, 2023, 38(7): 1979-1987.)
- [5] 王浩丞, 罗贺, 马滢滢, 等. 基于纳什均衡博弈的多无人机对地攻击目标分配方法[J]. *控制与决策*, 2024, 39(4): 1361-1369.  
(Wang H C, Luo H, Ma Y Y, et al. A target assignment method based on Nash equilibrium game for multi UAV ground attack[J]. *Control and Decision*, 2024, 39(4): 1361-1369.)
- [6] 王俐英, 林嘉琳, 宋美琴, 等. 考虑需求响应激励机制的园区综合能源系统博弈优化调度[J]. *控制与决策*, 2023, 38(11): 3192-3200.  
(Wang L Y, Lin J L, Song M Q, et al. Optimal dispatch of park integrated energy system considering demand response incentive mechanism[J]. *Control and Decision*, 2023, 38(11): 3192-3200.)
- [7] Zhang X P. Restructured electric power systems: Analysis of electricity markets with equilibrium models[M]. Hoboken: Wiley, 2010.
- [8] Hobbs B F, Metzler C B, Pang J S. Strategic gaming analysis for electric power systems: An MPEC approach[J]. *IEEE Transactions on Power Systems*, 2000, 15(2): 638-645.
- [9] 邓盛盛, 陈皓勇, 肖东亮, 等. 考虑碳市场交易的寡头电力市场均衡分析[J]. *南方电网技术*, 2024, 18(1): 143-152.  
(Deng S S, Chen H Y, Xiao D L, et al. Equilibrium analysis of oligopoly electricity market considering carbon market trading[J]. *Southern Power System Technology*, 2024, 18(1): 143-152.)
- [10] 刘雨梦, 陈皓勇, 黄龙, 等. 基于多群体协同进化的电力市场均衡模型[J]. *电力系统保护与控制*, 2020, 48(10): 38-45.  
(Liu Y M, Chen H Y, Huang L, et al. Equilibrium model of electricity market based on multi-swarm co-evolution[J]. *Power System Protection and Control*, 2020, 48(10): 38-45.)
- [11] Glover D, Krishnamoorthy G, Ren H D, et al. Deep reinforcement learning for distribution system operations: A tutorial and survey[J]. *Proceedings of the IEEE*, 2025, 113(6): 557-585.
- [12] Qiu D W, Ye Y J, 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.
- [13] 王岩红, 钟颖, 张允华. 基于改进 Q 学习的电动冷藏车多目标跨区域路径优化[J]. *控制与决策*, 2026, 41(3): 741-753.  
(Wang Y H, Zhong Y, Zhang Y H. Multi-objective cross-regional routing optimization of electric refrigerated vehicles based on improved Q-learning[J]. *Control and Decision*, 2026, 41(3): 741-753.)
- [14] 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.
- [15] Liang Y C, Guo C L, Ding Z H, et al. Agent-based modeling in electricity market using deep deterministic policy gradient algorithm[J]. *IEEE Transactions on*

- [Power Systems](#), 2020, 35(6): 4180-4192.
- [16] Zhang Y J, Zhang J, Wu G B, et al. Optimal power dispatch of active distribution network and P2P energy trading based on soft actor-critic algorithm incorporating distributed trading control[J]. [Journal of Modern Power Systems and Clean Energy](#), 2025, 13(2): 540-551.
- [17] Zheng Y Y, Wang H, Wang J L, et al. Optimal scheduling strategy of electricity and thermal energy storage based on soft actor-critic reinforcement learning approach[J]. [Journal of Energy Storage](#), 2024, 92: 112084.
- [18] Shi X Y, Xu Y L, Chen G B, et al. An augmented Lagrangian-based safe reinforcement learning algorithm for carbon-oriented optimal scheduling of EV aggregators[J]. [IEEE Transactions on Smart Grid](#), 2024, 15(1): 795-809.
- [19] Wang W, Yu N P, Gao Y Q, et al. Safe off-policy deep reinforcement learning algorithm for volt-VAR control in power distribution systems[J]. [IEEE Transactions on Smart Grid](#), 2020, 11(4): 3008-3018.

### 作者简介

朱晓晴 (1995-), 女, 博士生, 主要研究方向为基于强化学习的电力市场优化, E-mail: [3021053840@qq.com](mailto:3021053840@qq.com);

刘智伟 (1982-), 男, 教授, 博士, 主要研究方向为协同控制、分布式网络系统优化及其应用, E-mail: [zwliu@hust.edu.cn](mailto:zwliu@hust.edu.cn);

雷浩 (2001-), 男, 硕士研究生, 主要研究方向为控制与优化、深度强化学习在新型电力系统中的应用、边缘计算, E-mail: [haolei@hust.edu.cn](mailto:haolei@hust.edu.cn);

周长远 (2001-), 男, 博士生, 主要研究方向为配电网优化与控制, E-mail: [cyzhou@hust.edu.cn](mailto:cyzhou@hust.edu.cn).