

# 基于两阶段深度强化学习算法的多智能体自由合谋竞价机理研究

刘飞宇<sup>1</sup>, 王吉文<sup>2</sup>, 王正风<sup>2</sup>, 王蓓蓓<sup>1\*</sup>

(1. 东南大学电气工程学院, 江苏省南京市 210096;

2. 国网安徽省电力有限公司, 安徽省合肥市 230022)

## Study on Free Joint Bidding Mechanism in Multi-agent Environment Based on Two-stage Deep Reinforcement Learning Algorithm

LIU Feiyu<sup>1</sup>, WANG Jiwen<sup>2</sup>, WANG Zhengfeng<sup>2</sup>, WANG Beibei<sup>1\*</sup>

(1. School of Electrical Engineering, Southeast University, Nanjing 210096, Jiangsu Province, China;

2. State Grid Anhui Electric Power Co. Ltd, Hefei 230022, Anhui Province, China)

**ABSTRACT:** In the early stages of power market development, the imperfect regulatory mechanism provides opportunities for power generators to secretly communicate and jointly bid. Detecting such collusive behaviors is challenging. In this paper, a bidding model considering the association among generators is established and a novel two-stage deep reinforcement learning algorithm is proposed to tackle the discrete-and-continuous decision problem of choosing collusion partners and bidding coefficients. The joint strategies of generators under different congestion scenarios are analyzed, and the effectiveness of the algorithm is validated in a large-scale case study. The simulation results demonstrate that the proposed method can effectively emulate the free association of generators and identify potential collusion.

**KEY WORDS:** two-stage deep reinforcement learning; free association; multi-agent simulation; joint auction behaviors

**摘要:** 电力市场建设初期, 不完善的监管机制为发电商提供了暗中交流, 联合竞价的机会。然而, 如何找到潜在的发电商合谋组合是相对困难的事情。针对这一问题, 该文建立一种允许发电商自由联合的竞价模型, 并提出全新的两阶段深度强化学习算法, 来求解由离散的合谋对象选择和连续的报价系数确定组合形成的离散、连续动作混合决策问题。在不同阻塞情况下, 对发电商联合策略形成过程进行分析, 并在大算例中验证了算法的有效性。仿真结果表明, 所提出的方

法可以对市场主体的自由联合行为进行有效模拟, 发现潜在的合谋组合。

**关键词:** 两阶段深度强化学习; 自由联合; 多智能体仿真; 合谋竞价

## 0 引言

随着我国电力市场改革的进行, “统一市场、两级运作”的总体框架已基本确定<sup>[1]</sup>, 市场出清结果在系统运行中发挥着越来越重要的作用。然而, 仍不完善市场监管机制使得市场主体有机会私下互通, 策略性地协同竞价影响市场出清结果, 给市场的正常运行带来较大挑战<sup>[2-3]</sup>。在此背景下, 对市场主体策略竞价行为进行深入研究, 探索管控市场力的方法<sup>[4]</sup>就显得愈发重要。

根据行使市场力主体的数量, 市场主体的策略竞价行为可建模为2类: 独立竞价和联合竞价。独立竞价的市场主体被建模为完全理性的个体, 以最大化自身收益为目标优化竞价策略。但实际市场中主体数量众多, 独立竞价行为对市场正常运行的影响有限, 市场主体往往采取联合竞价的方式影响市场出清结果<sup>[5]</sup>。联合竞价的市场主体属于同一个报价联盟, 可相互交流信息并进行协同报价, 以最大化合谋收益作为目标优化联合竞价策略。现有研究重点关注联合竞价建模方式<sup>[6-7]</sup>及最优联合策略<sup>[8]</sup>, 大多基于机组成本信息人为划分合谋组合, 再按照设定的合谋组合研究联合竞价策略<sup>[9]</sup>。这种人为设定合谋组合的方式难以考虑网

**基金项目:** 国网安徽省电力有限公司科技项目: 双碳目标下安徽电网低碳调度及交易关键技术研究(521200220004)。

Project Supported by Science and Technology Project of State Grid Corporation: Research on Key Technologies of Low-Carbon Dispatching and Trading of Anhui Power Grid under the Dual Carbon Goals (521200220004).

络阻塞对市场力的影响，容易忽视潜在的合谋组合，使仿真结果与实际产生偏差。亟需对合谋形成过程进行建模，将联合过程与策略竞价结合，分析联合竞价的机理。

近些年，研究市场主体策略竞价行为的方法主要分为两类：基于博弈论和基于智能算法。基于博弈论的方法侧重于对市场均衡状态进行研究<sup>[9]</sup>，根据使用博弈模型的不同，可细分为主从博弈<sup>[10]</sup>、重复博弈<sup>[11]</sup>、演化博弈<sup>[12-13]</sup>等。虽然该类方法可求解市场主体最优策略，推导市场均衡状态，但为保证理论完备性，在研究中通常需要对市场主体和市场环境进行大量理想化假设，模型中难以考虑交流潮流、网络拓扑等约束条件，这就使得对应研究不可避免与实际场景存在偏差<sup>[14]</sup>。基于智能算法的方法侧重于对市场主体竞价行为进行模拟，分析其竞价策略。根据决策时所需信息量的不同，可细分为完全信息决策算法和非完全信息决策算法。完全信息决策算法主要包括粒子群<sup>[15]</sup>、协同进化<sup>[16]</sup>等，这类算法在求解时需要获取其他主体的报价数据、市场出清模型等信息以进行迭代优化，但这类全局信息在实际中往往难以获取。非完全信息决策算法主要指强化学习等算法，这类算法在训练时仅需机组自身的报价数据及市场出清结果，更加贴近实际情况<sup>[14]</sup>。因此，被广泛用来模拟传统机组<sup>[14,17-24]</sup>、新能源机组<sup>[25]</sup>、负荷<sup>[26-28]</sup>以及储能<sup>[29-31]</sup>等市场主体的竞价行为。表 1 梳理了近 3 年基于强化学习算法研究市场主体竞价策略的文献。从表中可以看出，现有研究中使用强化学习求解的大多是单一决策量的问题。而当问题中包含多个决策量时，主要有以下两种处理方法，其一是简化问题建模<sup>[26]</sup>，直接减少决策动作的数量<sup>[27,31]</sup>，但这需要原问题本身具有可简化的特性；其二是直接输出多维决策动作<sup>[28]</sup>，但如果不同动作间存在耦合关系，会给算法训练带来困难。现有研究中的方法难以求解多决策量问题，更难以求解由离散和连续决策量构成的混合动作空间问题。

针对现有研究的不足，本文提出了一种考虑发电商自由联合行为的多智能体合谋仿真方法。首先对发电商智能体选择合谋对象和策略竞价的行为进行建模；其次，将竞争深度 Q 网络算法 (Dueling deep Q network, Dueling-DQN) 和柔性决策-评价算法 (soft actor critic, SAC) 结合形成两阶

表 1 基于强化学习市场主体竞价行为研究

Table 1 Literature review of market bidding based on reinforcement learning

市场主体	时段	主体数量	决策变量	求解算法
发电商	单时段	多主体	报价系数(离散)	IEWA <sup>[19]</sup>
			报价系数(连续)	WoLF-PHC <sup>[20]</sup> DDPG <sup>[21]</sup> M-CARLA <sup>[23]</sup>
	多时段	单主体	报价系数(离散)	DDQN <sup>[18]</sup>
			报价系数(连续)	DDPG <sup>[14,21]</sup>
	多时段	多主体	报价系数(离散)	DDQN <sup>[18]</sup>
			报价系数(连续)	A3C <sup>[17]</sup> 、 MA-DDPG <sup>[22,24]</sup>
新能源机组	多时段	多主体	报价系数(连续)	Ape-X SAC <sup>[25]</sup>
			报价系数(连续)	M-CARLA <sup>[23]</sup>
售电商 (负荷聚合商)	多时段	单主体	购电量(连续)+ 零售电价(连续)	DDPG <sup>[26]</sup> 、 SAC <sup>[27]</sup>
			DR 上报量(连续)+ 报价系数(连续)	SAC <sup>[28]</sup>
	多时段	单主体	充放功率(连续)	Ape-X SAC <sup>[25]</sup>
充放功率(连续)+ 报价系数(连续)			DDPG <sup>[31]</sup>	
储能 (电动汽车聚合商)	多时段	多主体	充放功率(离散)	WoLF-PHC <sup>[29]</sup> DRMA-DDPG <sup>[30]</sup>

段的求解算法来对模型中的离散、连续混合决策量进行求解。通过这一方法，可以在考虑网络拓扑的情况下，模拟市场主体在长期市场出清过程中可能出现的自由联合竞价行为，进而为电力现货市场中市场力的控制提供一定的参考。

## 1 考虑多智能体自由联合行为的电力现货市场仿真建模

### 1.1 考虑自由联合的发电商竞价模型

在监管不完善的市场中，发电商可以通过私下交流形成合谋组合，通过联合策略竞价获取超额收益。本文在市场主体策略竞价行为建模的基础上，增加了对发电商合谋对象选择过程的建模，与原有竞价模型结合，变为考虑自由联合的发电商策略竞价模型。

#### 1.1.1 合谋对象选择方式

本文引入合谋向量对发电商自发形成合谋组合的过程进行建模，规定有合谋倾向的发电商在提交报价时，额外给出一个合谋向量。需要注意的是，此过程中引入合谋向量并非真实市场流程，仅是作为传达合谋意向的工具。图 1 展示了 4 个

发电商采用合谋向量交流合谋意愿的过程，合谋向量的维度为市场中可以进行合谋的发电商数量，向量中 0-1 元素代表与对应发电商合谋的意愿，若为 1 则表示愿意合谋，约定在合谋向量中仅可选择唯一的发电商进行合谋。同时，为了撮合发电商的合谋意向，基于合谋向量建立一套合谋结果计算机制。该机制可保证合谋意向的交互仅在私下进行，市场运营商无法知晓哪些市场主体进行了合谋，市场主体也仅能在合谋成功时才能知晓合谋对象的信息。

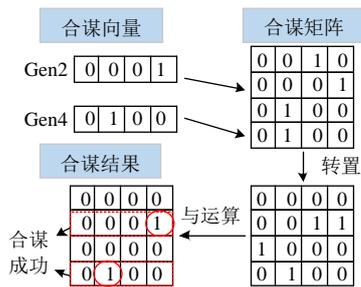


图 1 合谋结果计算方式

Fig. 1 Calculation method of collusion result

1.1.2 竞价方式

发电商成本函数建模为与电量有关的二次函数，其边际成本为一次函数<sup>[18]</sup>。假定发电商向 ISO 提交的报价函数是在其边际成本的基础上乘以一个报价系数，发电商的报价函数为

$$B_i(p_i) = k_i(a_i p_i + b_i), \forall i \in G \quad (1)$$

式中： $B_i(p_i)$ ， $p_i$ ， $k_i$  分别为发电商  $i$  提交的报价曲线、中标电量以及报价系数； $a_i$ ， $b_i$  分别为发电商  $i$  成本的二次项和一次项系数； $G$  为所有发电商的集合。

1.1.3 考虑自由合谋的发电商竞价模型

允许自由联合后，发电商的合谋组合会发生变动。如采用现有的收益分摊方法来计算合谋收益，计算效率较低，且在非完全信息时，发电商无法知晓的全部市场信息难以计算分摊结果。因此，本文以最大化合谋综合效用对合谋行为进行建模<sup>[32]</sup>，设置合谋收益系数来平衡对自身收益与合谋对象收益的重视程度。该合谋效用函数用来对合谋行为进行建模，仅作为智能算法训练的目标函数，不涉及具体的合谋收益分摊结果，能够满足非完全信息下联合竞价的需要。总体竞价模型如式(2)—(3)所示。

$$\max[R_i(k_i, \mathbf{dK}_i) + \varepsilon R_{-i}] \quad (2)$$

$$\text{s.t.} \begin{cases} R_i = \lambda_i p_i - C_i(p_i) \\ k_{i\min} \leq k_i \leq k_{i\max} \\ \min f \end{cases} \quad (3)$$

式中： $R_i$ 为发电商  $i$  自身的收益； $k_i$ 、 $\mathbf{dK}_i$  分别为发电商  $i$  提交的报价系数、合谋向量； $R_{-i}$ 为发电商所选择的合谋对象的收益； $\varepsilon$ 为合谋收益系数，若合谋不成功则为 0，若合谋成功则位于(0,1]区间； $\lambda_i$ 、 $p_i$ 和  $C_i$  分别为发电商  $i$  的节点电价、中标电量和发电成本； $f$ 为市场购电总成本。

1.2 电力现货市场出清模型

本文基于交流潮流模型进行市场出清，以最小化市场购电总成本为目标函数，考虑交流潮流约束，机组出力约束和潮流越限约束，市场出清结果调用 Matpower 进行计算。

2 考虑自由联合的两阶段竞价算法

在对合谋形成过程进行建模后，竞价模型的决策量变为离散的合谋对象选择与连续的报价系数确定，两者结合形成离散-连续混合动作空间。现有的深度强化学习算法只能对离散或连续动作进行决策，难以对混合动作空间进行求解。为了解决这一问题，本文将 Dueling-DQN 和 SAC 算法结合提出了一种两阶段深度强化学习算法框架，构建思路如图 2。将市场中多个发电商替换为基于两阶段深度强化学习算法的竞价智能体，构建多智能体仿真环境。

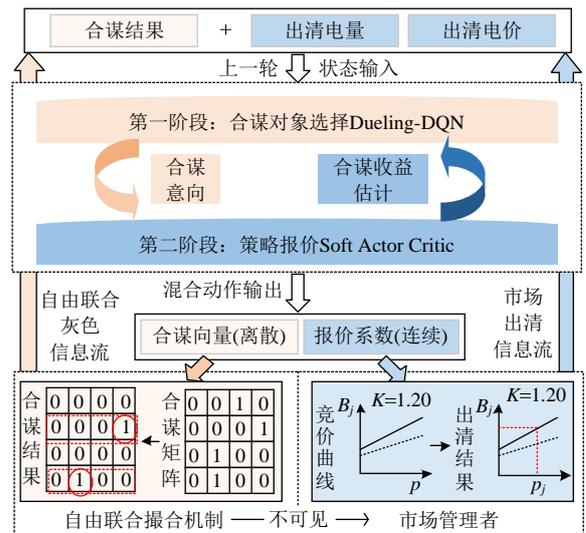


图 2 考虑自由联合的两阶段竞价算法构建思路

Fig. 2 Two-stage bidding algorithm considering free association

2.1 第一阶段：合谋对象选择算法 Dueling-DQN

在第一阶段中，使用 Dueling-DQN 来决定期

望的合谋对象。Dueling-DQN 是 DQN 算法的改进算法，在处理离散动作决策中有良好的效果<sup>[33]</sup>。接下来将介绍算法的具体使用思路。

1) 输入状态。

算法以上一轮市场出清结果和合谋结果作为输入状态  $S_D$ ，具体如式(4)所示：

$$S_D = [\lambda_i, p_i, s_{dk}] \quad (4)$$

式中市场出清结果包括节点电价  $\lambda_i$  和出清电量  $p_i$ ，而合谋结果  $s_{dk}$  基于发电商  $i$  上一轮给出的合谋向量  $dK_i$  计算得到。

2) 输出动作。

算法的输出动作  $a_D$  为合谋向量  $dK_i$ 。算法首先使用网络估计出与不同发电商合谋的期望收益，再基于  $\epsilon$ -greedy 策略选择期望收益最高的作为合谋对象，最后将选择结果转变为合谋向量，即：

$$a_D = dK_i \quad (5)$$

3) 奖励函数。

理论上算法可以将式(2)中的目标函数，即合谋后的综合效用函数  $R$  作为奖励函数，但该函数中包含有第二阶段算法给出的报价系数  $k_i$ ，在计算时需要考虑第二阶段算法训练得到的报价策略。为了促进两阶段算法融合，本文将结合第二阶段算法 Critic 网络的估计结果来计算第一阶段算法的奖励函数，具体将在 3.3 小节中详细介绍。

### 2.2 第二阶段：策略报价算法 Soft Actor Critic

在第二阶段中，使用 SAC 算法来决定具体竞价策略，该算法可以适应连续的状态输入和连续的动作输出<sup>[34]</sup>，且样本利用率高，探索性能好，有效缓解了算法训练过程中过早陷入局部最优的问题<sup>[35]</sup>。算法使用思路如下：

1) 输入状态。

第二阶段算法需要综合考虑上一轮的市场出清结果以及本轮第一阶段算法给出的合谋向量  $dK_i$ ，形成一个综合的输入状态，具体如式(6)所示。

$$S_s = [\lambda_i, p_i, s_{dk}, dK_i] \quad (6)$$

2) 输出动作。

第二阶段算法的输出动作  $a_s$  为发电商的报价系数  $k_i$ ，即：

$$a_s = k_i \quad (7)$$

3) 奖励函数。

第二阶段算法将考虑市场出清收益以及合谋是否成功，以综合效用函数作为算法的奖励函

数，具体如式(8)所示。

$$R = R_i(k_i, dK_i) + \epsilon R_{-i} \quad (8)$$

### 2.3 两阶段深度强化学习算法训练思路

本文将两种不同动作类型的深度强化学习算法结合起来，构建两阶段深度强化学习算法框架，对考虑自由联合的多智能体发电商竞价问题进行求解。首先以 Dueling-DQN 作为第一阶段算法，得到离散动作，即合谋向量；其次，以 SAC 算法作为第二阶段算法，得到连续动作，即报价系数。所提出的两阶段竞价算法网络框架如图 3 所示。

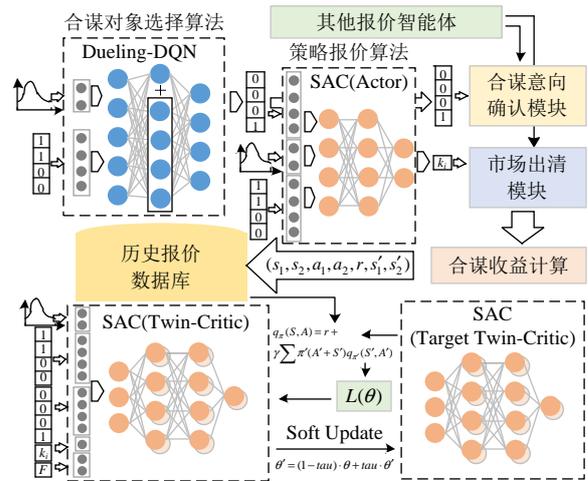


图 3 两阶段深度强化学习算法网络框架

Fig. 3 Framework of two-stage deep reinforcement learning algorithm

为了将两阶段的深度强化学习融合起来，本文对算法的训练思路做出一定改动，主要包括：扩展 SAC 算法的输入以及修改 Dueling-DQN 算法奖励函数的计算方法，接下来将详细说明。

1) SAC 算法 Critic 网络的训练。

原始 SAC 算法中的 Critic 网络用来估计竞价动作的期望收益，其输入为市场状态及报价系数。而在两阶段算法框架中，第一阶段的合谋对象选择会影响最终的收益，这就使得第二阶段算法训练时需要参考第一阶段合谋对象选择的结果。因此，参考多智能体协同训练的方法<sup>[36]</sup>，在算法原有状态动作的基础上，额外增加第一阶段算法给出的合谋向量以及合谋是否成功的信息作为输入，对 Critic 网络的输入状态和动作进行扩展，修改后的训练目标如式(9)—(10)所示。

$$\min[Q'_c(S_D, A, F) - Q_c(S_D, A, F)]^2 \quad (9)$$

$$Q'_c(S_D, A, F) = R(S_D, A) + \gamma \sum \pi(a'_k | S'_D) \cdot [Q_c(S'_{cD}, A', F) - \alpha \log \pi(a'_k | S'_D)] \quad (10)$$

式中： $\mathbf{A}=[a_D, a_S]$ 为完整动作，包括第一阶段的合谋向量和第二阶段的报价系数； $\mathbf{F}$ 为 one hot 形式的向量，是额外补充的市场状态，表示合谋是否成功； $Q_c$ 、 $Q'_c$  分别为 Critic 网络训练前和基于 Bellman 估计的  $Q$  值； $\pi$ 为 Actor 网络现阶段的策略分布， $\pi(a'_k|S')$ 表示状态  $S'$ 下选择动作  $a'_k$  的概率， $S'$ 和  $a'_k$  为下一状态和下一状态 Actor 网络选择的动作； $\gamma$ 为后续收益的折算系数； $\alpha$ 为温度系数，是信息熵与下一状态动作价值的权重，本文中采用梯度下降的方式<sup>[37]</sup>优化 SAC 算法的温度系数。对于目标评价网络，采用软更新的方式进行训练<sup>[38]</sup>。

2) SAC 算法 Actor 网络的训练。

对于 Actor 网络采用与 Critic 网络采用相似的思路，增加第一阶段合谋对象选择的结果，扩展其状态输入。修改后网络的训练目标是最大化动作价值<sup>[37]</sup>，具体如式(11)所示。

$$\max[Q'_c(S, \mathbf{A}, \mathbf{F}) - \alpha \log \pi(a_k | S)], a_k \in \pi \quad (11)$$

式中  $Q'_c$  为 Critic 网络经训练后估计出的动作价值。

3) Dueling-DQN 算法的训练。

对于第一阶段的合谋对象选择，算法网络输出的  $Q$  值代表与不同发电商合谋的期望收益，而合谋收益的多少不仅取决于第一阶段的合谋对象选择，更取决于第二阶段给出的具体竞价动作。在本文中，使用 SAC 算法的 Critic 网络辅助估计第一阶段算法的  $Q$  值。实际测试中发现，训练初期的 Critic 网络估计的结果  $Q'_c$  会和实际有一定偏差，这一偏差会使第一阶段算法训练困难。而如果使用历史数据推算的  $Q'_c$ ，则会存在一定的滞后，同样难以训练。最终，本文在式(13)中设置了一个系数，将两个估计结果结合起来保证第一阶段算法训练的稳定性，修改后的训练目标如式(12)—(13)所示。

$$\min[Q'_D(S, a_D) - Q_D(S, a_D)]^2 \quad (12)$$

$$Q'_D(S, a_D) = [\beta Q'_c(S, \mathbf{A}, \mathbf{F}) + (1 - \beta) \cdot Q'_c(S, \mathbf{A}, \mathbf{F}) + \gamma Q_D(S', a'_D)], a_k \in \pi \quad (13)$$

式中： $\beta$ 为两种估计方式的比值系数； $Q_D$ 、 $Q'_D$  分别为 Dueling-DQN 网络训练前后的  $Q$  值。

算法整体流程如图 4 所示。

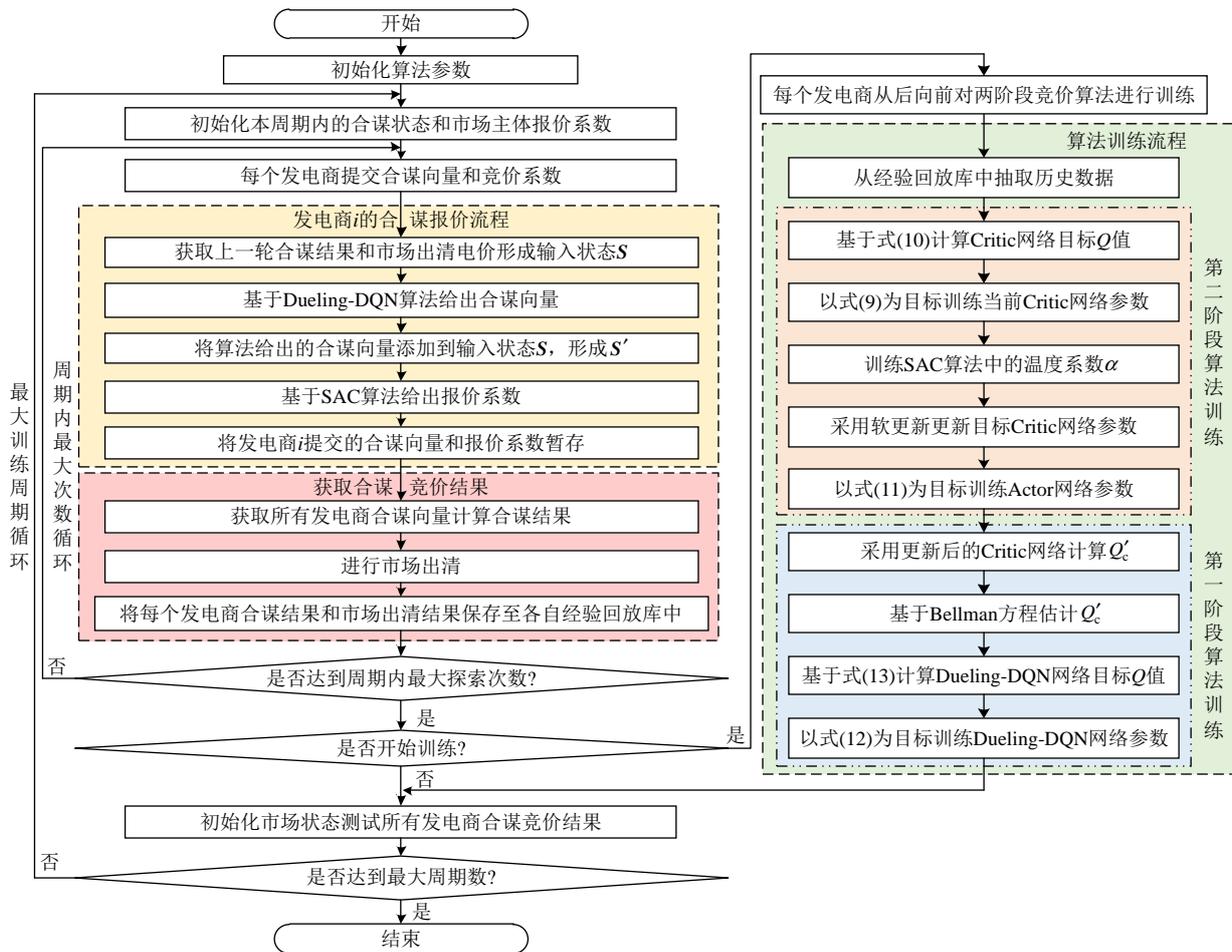


图 4 两阶段竞价算法流程图  
Fig. 4 Flow chart of two-stage bidding algorithm

### 3 算例分析

接下来将在 PJM5 节点使用本文提出的发电商自由联合竞价模型以及两阶段求解算法进行模拟仿真。鉴于 5 节点网络中发电商数量较少，将发电商 1 设置为按照边际成本报价，作为参考基准，选择发电商 2、3、4、5 作为策略报价智能体(记为 gen2、3、4、5)，将式(2)中的合谋收益系数设置为 0.5，对比市场主体独立报价与自由联合后的仿真结果。发电商信息见附表 A1，算法参数的设置见附表 A2。

#### 3.1 不考虑合谋的独立竞价策略仿真

首先对不考虑合谋情况下发电商的独立竞价策略进行研究，以此作为发电商报价系数、出清电价以及出清收益的基准。表 2 给出了无阻塞情况下发电商独立竞价的仿真结果，括号中给出了按照边际成本报价时的收益作为对比。

表 2 不考虑合谋行为时发电商的报价结果

发电商编号	报价系数	节点电价/(元/(MW·h))	收益/元	收益提升率/%
2	1.13	27.27	2 872 (2 259)	27.1
3	1.06	27.74	417 (150)	178.0
4	1.05	27.56	249 (52)	378.8
5	1.12	27.26	1 380 (847)	62.9

对比完全按照边际成本报价，策略报价后 4 个发电商的报价系数都有轻微提升，收益也获得不同程度的提升。进一步来看，虽然 4 个发电商的报价系数均有提高，但最高的报价系数仅为 1.13，总体上仍处在较低水平，且报价行为对节点电价的抬升有限(约 8.3%)。这表明在 5 节点算例中，若不考虑合谋行为，则发电商间竞争较为激烈(赫芬达尔-赫希曼指数(Herfindahl-Hirschman index, HHI)指数为 2 257，最小相对强弱指标(relative strength index, RSI)为 125%)，其独立竞价行为对市场出清结果的影响有限。

#### 3.2 考虑自由联合行为的发电商策略报价仿真

下文将使用提出的两阶段深度强化学习算法，模拟长期竞价中发电商可能出现的自由联合竞价行为，同样将发电商 1 设置为按照边际成本报价，作为市场的基准。

##### 3.2.1 不考虑支路阻塞的情况

表 3 展示了无阻塞时发电商的自由联合竞价结果。在允许合谋后，4 个发电商自发形成了两个合谋组合，发电商 2、5 形成了合谋组合 A，发电商 3、4 形成了合谋组合 B。对比表 1 中独立策略竞价的结

表 3 不考虑阻塞时发电商的合谋竞价结果

Table 3 Collusion simulation results without considering congestion

发电商	报价系数	合谋意向	合谋组合	节点电价/(元/(MW·h))	个体收益/元	综合效用/元
2	1.22	[0,0,0,1]	A	29.46	3 449	4 369
3	1.11	[0,0,1,0]	B	29.48	623	836
4	1.10	[0,1,0,0]	B	29.30	425	737
5	1.21	[1,0,0,0]	A	29.00	1 839	3 564

果，发电商的报价系数、收益进一步提升，节点电价也由约 27.5 元/(MW·h)提升至约 29.2 元/(MW·h)。

下文分析仿真的过程数据，图 5 展示了训练过程中不同发电商对与其他主体合谋期望收益的估计结果，也即第一阶段算法的估计结果，每个周期发电商倾向于选择的合谋对象用红点标出。

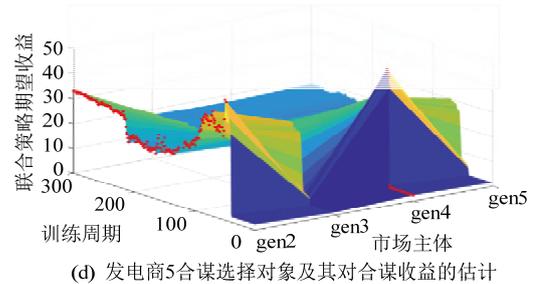
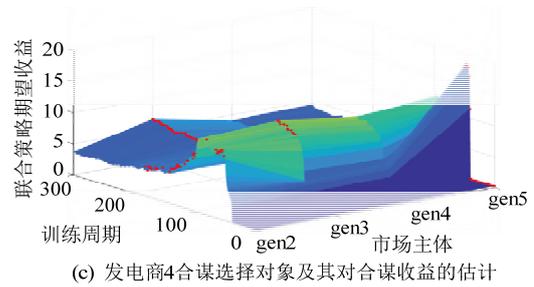
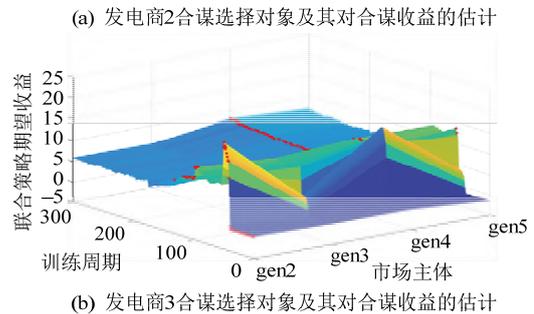
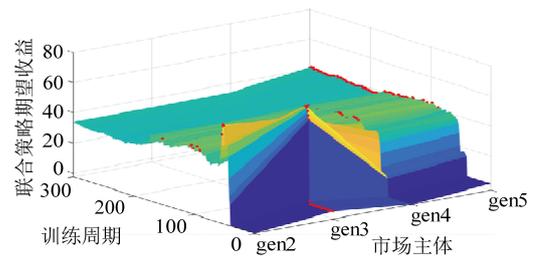


图 5 不考虑阻塞时发电商合谋期望收益的估计结果  
Fig. 5 Expected utilities of collusion between different generators without considering congestion

在前40个周期，由于没有对网络进行训练，与不同主体合谋的期望收益相差很小，基本都在0附近，此时算法给出的合谋选择结果是由网络初始状态决定，即随机生成的。在对算法进行训练后，与不同主体联合竞价的期望收益开始出现不同幅度的增长。从图5(a)、(d)可以看出，发电商2、5在经过80个周期训练后，红点基本都集中在特定的主体上，这表明算法发现与特定市场主体联合竞价的收益明显高于其他主体。在后续训练中，算法会倾向于继续选择该主体，最终形成稳定的合谋组合。而发电商3、4的最佳合谋对象出现过变化，其中发电商4的变化较为明显。在前期，发电商4的最佳合谋对象都是发电商2，但此时发电商2的最佳合谋对象并不是发电商4，这使得合谋无法成功。直到200个周期后，发电商4发现与发电商3进行联合可以提升收益，其最佳合谋对象也因此出现变化。

图6展示了发电商的报价系数在每个周期测试中的变化。前40个周期由于没有进行训练，报价系数同样是固定的。随着训练进行，报价系数逐渐变化。在约200次训练后，报价系数逐渐稳定。最终，发电商2、5的报价系数分别从1.13和1.12提升到1.22和1.21，另外两个发电商的报价系数也有一定程度的提高。

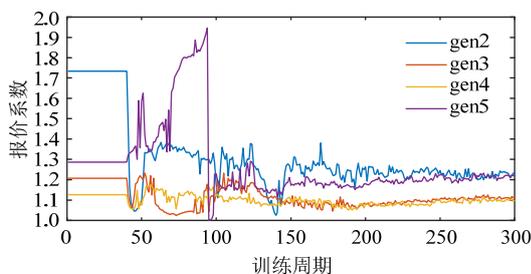


图6 不考虑阻塞时发电商报价系数均值变化

Fig. 6 Variation of bidding coefficient of generators without considering congestion

在上述算例中，有4个发电商可以进行自由联合，如果仅考虑两两联合的情况，共有3种合谋组合划分方式。根据合谋组合中发电商的成本及机组容量，这几种情况可以大致分为强强合谋和强弱合谋两类。为了验证算法的有效性，对所有两两组合情况使用粒子群算法进行仿真<sup>[9]</sup>，表4展示了人为划分组合后的竞价结果。

对比表3和4可以发现，本文所提出的两阶段算法得到的合谋分组结果是表4中的分组三。虽然在具体的竞价策略上有一定从偏差，但在成员收

表4 人为划分组合的合谋竞价结果

Table 4 Simulation results when the collusion combinations are given

合谋划分	发电商	报价系数	节点电价/(元/(MW·h))	个体收益/元	综合效用/元
分组一:	2	1.36	29.40	3511	3899
gen2,3	3	1.12	29.89	775	2531
gen4,5	4	1.15	29.72	402	1373
(强弱联合)	5	1.21	29.39	1942	2143
分组二:	2	1.36	29.37	3501	3769
gen2,4	3	1.16	29.88	633	1595
gen3,5	4	1.11	29.67	535	2286
(强弱联合)	5	1.21	29.36	1924	2241
分组三:	2	1.32	28.77	3320	4128
gen2,5	3	1.10	29.23	662	887
gen3,4	4	1.09	29.05	450	781
(强强联合)	5	1.21	28.76	1618	3278

益及出清电价上，本文所提的考虑自由联合的仿真算法得到的结果与人为划分合谋组合后基于粒子群的仿真结果基本一致，节点电价误差为0.55%，发电商收益的最大误差在4%。

再从收益上来看，本文在对合谋行为建模时约定合谋成功需要双方相互选择，这意味着双方形成合谋组合后的综合效用对比独立报价的提升较大。从图7中可以看出，对于发电商2来说，组合三的综合效用最高，其他3个发电商综合效用最高的分别是组合一、组合二和组合三。而从具体成员上来看，发电商3、4、5均期望与发电商2进行合谋，这意味着发电商2在一定程度上拥有选择权。因此，从最大化综合效用的角度，如果没有人为划分合谋分组，发电商2、5会倾向于组合三的综合分组方式。虽然发电商3、4也想与发电商2进行合谋，但双方不会相互选择，难以形成稳定合谋。

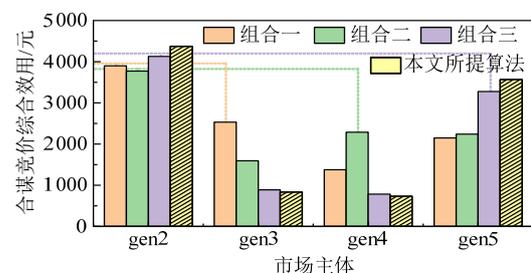


图7 不同合谋组合划分下发电商合谋综合效用对比

Fig. 7 Comprehensive utility of collusion under different collusion combinations

综上所述，在求解结果的准确性上，基于非完全信息的两阶段深度强化学习算法与基于完全

信息的粒子群算法基本一致。同时，本文将合谋形成过程考虑进竞价模型中，解决了难以划分合谋组合的问题，提高了仿真效率。

3.2.2 考虑支路阻塞较重的情况

将支路 1-2 的潮流限制设置为 300MW，对有阻塞情况下发电商的自由联合竞价行为进行仿真，结果如表 5 所示。

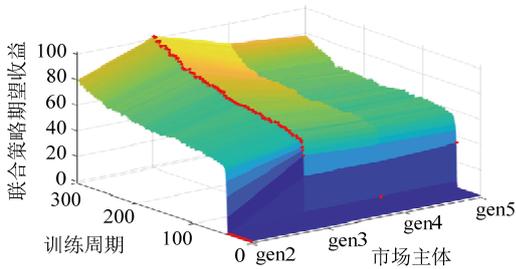
表 5 考虑阻塞时发电商的合谋竞价结果

Table 5 Collusion simulation results when considering congestion

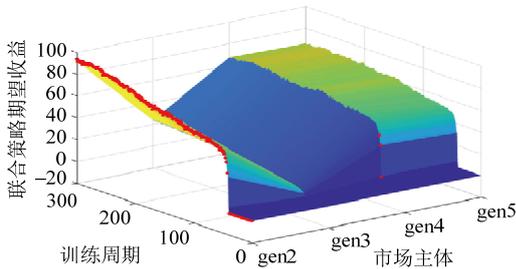
发电机	报价系数	合谋意向	合谋组合	节点电价/(元/(MW·h))	个体收益/元	综合效用/元
2	1.39	[0,1,0,0]	A	39.15	6436	8291
3	1.98	[1,0,0,0]	A	51.98	3710	6928
4	1.28	[0,0,0,1]	B	43.78	5576	5755
5	1.97	[0,0,1,0]	B	39.97	358	3146

从表中可以看出，在网络中出现线路阻塞后，发电商的联合方式相应发生了变化。从合谋组合上来看，发电机 2、3 形成了一个合谋组合，而发电机 4、5 形成了另一个合谋组合。4 个发电机在训练过程中选择的合谋对象及对与其他主体合谋期望收益的估计如图 8 所示。

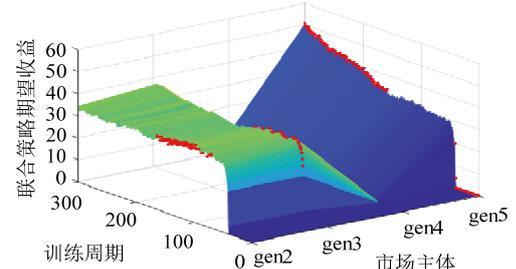
从图 8 可以看出，在网络中出现阻塞后，发电



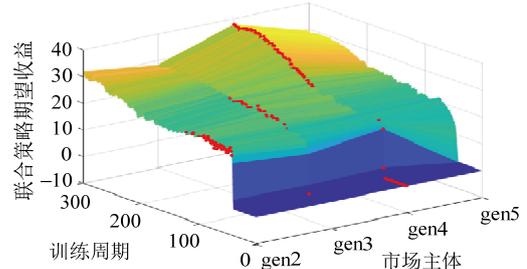
(a) 发电机2合谋选择对象及其对合谋收益的估计



(b) 发电机3合谋选择对象及其对合谋收益的估计



(c) 发电机4合谋选择对象及其对合谋收益的估计



(d) 发电机5合谋选择对象及其对合谋收益的估计

图 8 考虑阻塞时发电商合谋期望收益的估计结果

Fig. 8 Expected utilities of collusion between different generators when considering congestion

商 2、3 形成了稳定的合谋组合，而发电机 4、5 形成合谋组合的过程较为艰难。在前期，发电机 4、5 倾向于选择的合谋对象在发电机 2、3 间波动，这代表双方没有相互选择，无法形成稳定合谋组合。从图 9 也可以看出，在这段周期内，因没有形成稳定合谋组合，发电机 4、5 的报价系数处于 1.5 以下，处于低位且波动明显。而在 180 个训练周期后，发电机 4、5 形成了较为稳定的合谋组合，报价系数也分别降低和升高，且波动减小。这表明本文所提算法中第一阶段合谋向量的变化会促使第二阶段竞价策略随之发生变化，而第一阶段合谋收益的估计也会受具体报价系数的影响，两者相互影响逐渐找到一个合谋策略。

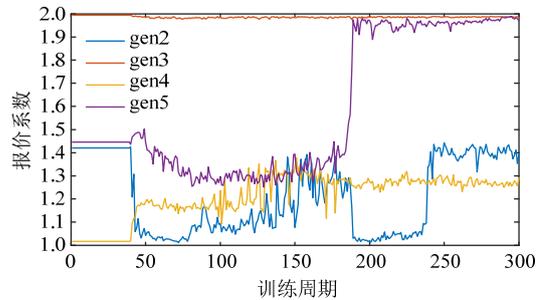


图 9 考虑阻塞情况下发电商报价系数均值变化

Fig. 9 Variation of bidding coefficient of generators when considering congestion

对比没有阻塞和有阻塞的情况，在没有阻塞时，体量较大且发电成本较低的发电机 2、5 优先形成了合谋组合。这两个发电机在独立竞价时就可以获取较高的收益，两者合谋后扩充了整体的市场占比(合谋后 HHI 指数为 4127，最小 RSI 为 95%)，可以通过策略行为更好的行使市场力。这种合谋行为属于现有文献中所描述的强强联合<sup>[9]</sup>，在不考虑网络拓扑影响时，这种因机组成本和容量相近而形成的合谋组合比较容易判断。而在网络中存在阻塞时，发电机 2、3 形成了新的合谋组合。发电机 3 的发电成本较高，在网络中没有阻塞

时, 获取收益较低, 属于实力较弱的发电商。但支路 1-2 的阻塞使发电商的潜在市场力发生了显著改变。不同于因机组容量和成本影响形成的强强联合, 支路阻塞改变了发电商 3 潜在市场力, 进而形成了非传统意义上的合谋组合。通过上述仿真结果可以看出, 虽然没有对阻塞进行特殊建模, 但是依靠无模型深度强化学习算法对环境的学习, 本文所提出算法对支路阻塞也有一定适应能力, 可以发现因阻塞导致的潜在合谋组合。

### 3.2.3 合谋收益系数的影响

本文在式(2)中使用合谋收益系数来反映合谋的紧密程度, 该系数的大小直接影响了发电商竞价模型的目标函数, 进而影响自由联合的竞价行为。接下来将对不同合谋收益系数设置下的自由联合行为进行模拟, 仿真结果见表 A3。

图 10 展示了不同合谋收益系数对于最终节点电价的影响。在没有阻塞的情况下, 合谋收益系数的增加使得市场出清电价逐步增加, 但合谋组合并没有变化。这是因为发电商 2、5 较为强势, 在不同合谋系数下, 两者强强联合均可最大化综合效用。而在线路 1-2 阻塞后, 情况则更加多变。在合谋收益系数大于 0.5 后, 合谋组合发生了变化, 发电商 2 和 4 形成了新的合谋组合。对比原有合谋分组, 节点电价最大值明显增加, 且节点间的价差变大, 这表明网络中阻塞情况加剧, 发电商策略竞价行为对于市场出清结果影响提升。总体来说, 合谋收益系数设置的越大, 合谋行为越严重。同时, 对比合谋收益系数设置为 1.0 (完全合谋竞价)与其他情况的结果可以发现, 合谋收益系数仅影响合谋程度, 并不会从根本上改变市场主体合谋的建模方式, 这与本文对合谋行为建模思路相符。

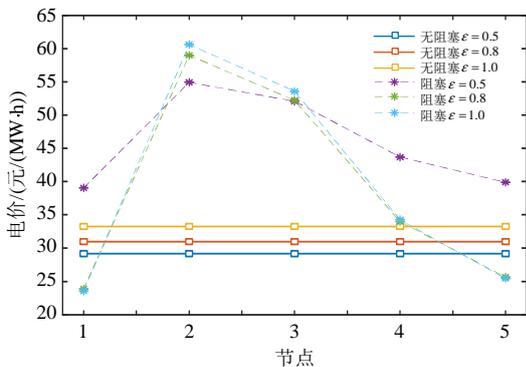


图 10 合谋收益系数对于节点电价的影响

Fig. 10 Influence of collusion coefficients on node price

## 4 进一步的仿真

### 4.1 规模的扩大

为了验证本文所提出算法的适用性, 接下来在 2383 节点系统<sup>[39]</sup>对 10 个规模较大的发电商进行自由联合竞价仿真。仿真周期设置为 600, 为了凸显合谋影响, 合谋收益系数设置为 1.0, 其他参数设置见附表 A2, 仿真结果如表 6 和附表 A4 所示。

表 6 2383 节点自由联合仿真结果

Table 6 Simulation results of the 2383 node network					
发电商	报价系数	合谋组合	节点电价/(元/(MW·h))	个体收益/元	收益提升/元
2	1.23	A	147.1	70725	554
3	1.00	B	147.2	52617	852
4	1.14	—	146.4	18796	1838
7	1.06	—	144.3	73078	-1767
10	1.60	C	147.4	29555	-158
13	1.06	D	141.6	47383	785
15	1.28	C	149.6	50832	2288
32	1.75	A	154.8	50863	10539
33	1.07	B	157.2	8771	7194
37	1.07	D	149.1	89504	3064

总体上来看, 在自由联合后, 10 个发电商中的 8 个两两联合形成了 4 个合谋组合, 仍然有 2 个发电商没有形成合谋组合。在报价行为上, 发电商 15、32 的报价系数对比独立报价有提升, 其中发电商 32 的收益对比独立报价提升了 26.1%。而发电商 3、7、13 的报价系数则出现下降, 其中发电商 7 因合谋不成功, 收益对比独立报价下降了 2.3%。对比独立报价, 合谋后节点电价大多升高, 且不同节点间价差增大, 这表明发电商有效的运用联合行为增强市场力, 进而影响节点电价。在合谋组合上, 如果仅考虑两两联合的情况, 将 10 个发电商划分为 5 个合谋组合, 共有 945 种合谋分组情况。由于网络中存在阻塞, 现有研究方法难以从众多可能的情况中确定合谋分组, 需要对所有可能的情况进行研究, 而这将耗费大量的时间。本文通过对合谋形成过程进行建模, 解决了联合竞价行为研究中合谋组合确定这一难题。从仿真结果来看, 自由联合所形成的合谋组合也并非单纯的强强联合, 如组合 A 和 B 从收益上来看, 均为强带弱, 但两者合谋后的收益增幅却非常明显。这一仿真结果从侧面反映了仅通过机组成本数据难以对因网络拓扑导致的合谋组合进行分析, 而本文所提出的算法则可以发现这些潜在合谋组合, 进一步凸显了本文方法的价值。

## 4.2 合谋的拓展

本文所提方法可以对发电商间的两两合谋行为进行研究，如果想要对多于两个主体的合谋行为进行模拟，可以参考分层聚类的思想<sup>[40]</sup>将已有的合谋组合视为一个主体，再与其他的合谋组合或主体进行自由联合。5.1 节中 10 个发电商形成了 4 个合谋组合，其中组合 A、B 收益提升较高，将其视为稳定的组合。组合 C、D 的收益提升有限，将这两个合谋组合拆分。在此基础上让组合 A、B 与剩下的发电商进行第二次自由联合仿真，结果如附表 A5 所示。

从仿真结果来看，进行第二次自由联合后，原有的合谋组合 A、B 均增加了一个成员，而剩下的 4 个发电商两两形成了合谋组合。从图 11 可以看出，部分发电商的报价系数对比第一次仿真结果进一步提高，节点电价也明显提升。最终组合 A、B、C 的收益在第一次仿真基础上，分别提高了 9860 元和 5309 元，而组合 B、D 的收益则出现了下降。总体来看，当合谋组合成员扩展后，合谋成员策略性报价的倾向更加明显，可以使合谋组合的收益明显提升。

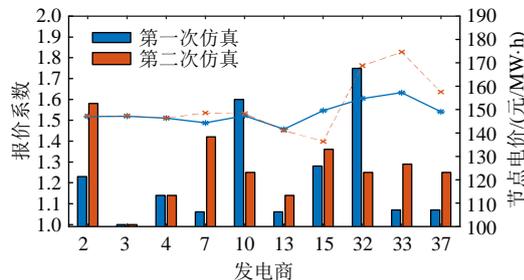


图 11 合谋扩展仿真结果

Fig. 11 Collusion expansion simulation results

## 5 结论与展望

实际的电力市场中网络约束往往较为复杂，在人为划分合谋组合时，难以考虑网络约束对机组潜在市场力的影响，使得划分的合谋组合不准确，不全面。针对这一问题，本文对合谋形成过程进行建模，构建了允许发电商自由联合的竞价模型，并提出了一种全新的基于两阶段深度强化学习算法框架，来对合谋对象和竞价系数选择所形成的离散、连续混合动作空间进行求解。仿真结果表明：

1) 本文所提方法可在考虑网络约束的情况下，模拟发电商在长期竞价过程中可能出现的两两自由联合竞价行为，发现非常规的合谋组合，减少人为划分合谋组合时的困难。

2) 本文的方法在大规模算例中仍然有效，且通过多次仿真，所提出的方法可在一定程度上突破两两合谋的限制，实现对合谋组合的扩展。

3) 考虑到计算速度以及模型复杂程度，本文在对合谋行为建模时采用了综合效用作为目标函数，对特定时段的合谋行为进行仿真分析，但未能对涉及跨时段、不同类型市场主体间收益分摊的合谋行为深入研究

在今后的工作中，将会对现有的合谋模型进行改进，增加合谋收益分摊机制，以分摊收益作为目标函数，对多时段的自由联合竞价行为进行模拟仿真，进一步提高模型的适用性。

## 参考文献

- [1] 国家发展改革委，国家能源局. 关于加快建设全国统一电力市场体系的指导意见[EB/OL]. [2022-01-18]. [http://www.gov.cn/zhengce/zhengceku/2022-01/30/content\\_5671296.htm](http://www.gov.cn/zhengce/zhengceku/2022-01/30/content_5671296.htm). National Development and Reform Commission, National Energy Administration. Guiding opinions on accelerating the construction of a national unified electricity market system[EB/OL]. [2022-01-18]. [http://www.gov.cn/zhengce/zhengceku/2022-01/30/content\\_5671296.htm](http://www.gov.cn/zhengce/zhengceku/2022-01/30/content_5671296.htm)(in Chinese).
- [2] WOLAK F A. Diagnosing the California electricity crisis[J]. The Electricity Journal, 2003, 16(7): 11-37.
- [3] REUTERS. JPMorgan to pay \$410 million to settle power market case[EB/OL]. [2013-06-30]. <https://www.reuters.com/article/us-jpmorgan-ferc-idUSBRE96T0NA20130730>.
- [4] 董礼，王胜华，华回春，等. 中国现货电力市场中发电企业滥用市场力违规识别[J]. 中国电机工程学报, 2021, 41(24): 8397-8407. DONG Li, WANG Shenghua, HUA Huichun, et al. Identification of market power abuse in spot market of Chinese electric market[J]. Proceedings of the CSEE, 2021, 41(24): 8397-8407(in Chinese).
- [5] 孙波，李志恒，谢敬东，等. 现货市场下发电商合谋竞价问题研究——基于云模型构建发电商合谋竞价评判模型的模拟分析[J]. 价格理论与实践, 2020(1): 57-61. SUN Bo, LI Zhiheng, XIE Jingdong, et al. Research on collusion bidding of power generators in spot market——Simulation analysis of constructing a collusion bidding evaluation model of power generators based on cloud model[J]. Price: Theory & Practice, 2020(1): 57-61(in Chinese).
- [6] ACUÑA L G, RÍOS D R, ARBOLEDA C P, et al. Cooperation model in the electricity energy market using bi-level optimization and Shapley value [J]. Operations

- Research Perspectives, 2018, 5: 161-168.
- [7] SUN Guanghui, WANG Xiaowei, YANG Libo, et al. A method for power suppliers' optimal cooperative bidding strategies considering network losses[J]. Global Energy Interconnection, 2020, 3(4): 335-345.
- [8] DOU Chunxia, PAN Xiaodong, ZHANG Zhanqiang, et al. Multi-agent-system-based bi-level bidding strategy of microgrid with game theory in the electricity market[J]. Electric Power Components and Systems, 2019, 47(8): 703-719.
- [9] 高怡静. 市场环境下发电商参与电力交易的报价行为推演研究[D]. 南京: 东南大学, 2020.  
GAO Yijing. Research on the deduction of bidding behavior of generators participating in power transactions under market environment[D]. Nanjing: Southeast University, 2020(in Chinese).
- [10] 吴诚, 高丙团, 汤奕, 等. 基于主从博弈的发电商与大用户双边合同交易模型[J]. 电力系统自动化, 2016, 40(22): 56-62.  
WU Cheng, GAO Bingtuan, TANG Yi, et al. Master-slave game based bilateral contract transaction model for generation companies and large consumers[J]. Automation of Electric Power Systems, 2016, 40(22): 56-62(in Chinese).
- [11] 蒋玮, 吴杰, 冯伟, 等. 日前电力市场不完全信息条件下的电力供需双边博弈模型[J]. 电力系统自动化, 2019, 43(2): 18-24, 75.  
JIANG Wei, WU Jie, FENG Wei, et al. Bilateral game model of power supply and demand sides with incomplete information in day-ahead electricity market [J]. Automation of Electric Power Systems, 2019, 43(2): 18-24, 75(in Chinese).
- [12] 程乐峰, 余涛. 发电市场长期竞价均衡自发形成过程中的一般多策略演化博弈决策行为研究[J]. 中国电机工程学报, 2020, 40(21): 6936-6955.  
CHENG Lefeng, YU Tao. Decision-making behavior investigation for general multi-strategy evolutionary games in the spontaneous formation of long-term bidding equilibria of a power generation market[J]. Proceedings of the CSEE, 2020, 40(21): 6936-6955(in Chinese).
- [13] 杨辉, 莫峻. 发电侧企业群体间竞价行为的随机演化博弈[J]. 电网技术, 2021, 45(9): 3389-3397.  
YANG Hui, MO Jun. Stochastic evolutionary game of bidding behavior for generation side enterprise groups[J]. Power System Technology, 2021, 45(9): 3389-3397(in Chinese).
- [14] YE Yujian, QIU Dawei, SUN Mingyang, et al. Deep reinforcement learning for strategic bidding in electricity markets[J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1343-1355.
- [15] 谢畅, 王蓓蓓, 赵盛楠, 等. 基于双层粒子群算法求解电力市场均衡[J]. 电网技术, 2018, 42(4): 1170-1176.  
XIE Chang, WANG Beibei, ZHAO Shengnan, et al. Equilibrium solution for electricity market based on bi-level particle swarm optimization algorithm[J]. Power System Technology, 2018, 42(4): 1170-1176(in Chinese).
- [16] 石可, 陈皓勇, 李鹏, 等. 基于协同进化的两种电力市场出清机制分析[J]. 电力系统自动化, 2019, 43(9): 68-74.  
SHI Ke, CHEN Haoyong, LI Peng, et al. Analysis on two kinds of electricity market clearance mechanism based on co-evolution[J]. Automation of Electric Power Systems, 2019, 43(9): 68-74(in Chinese).
- [17] 徐尔丰. 基于 A3C 强化学习的电力市场发电商报价策略研究[D]. 北京: 华北电力大学(北京), 2019.  
XU Erfeng. Research on bidding strategy of generators in electricity market based on Asynchronous Advantage Actor-Critic reinforcement learning[D]. Beijing: North China Electric Power University, 2019(in Chinese).
- [18] 高宇, 李昀, 曹蓉蓉, 等. 基于多代理 Double DQN 算法模拟发电侧竞价行为[J]. 电网技术, 2020, 44(11): 4175-4182.  
GAO Yu, LI Yun, CAO Rongrong, et al. Simulation of generators' bidding behavior based on multi-agent Double DQN[J]. Power System Technology, 2020, 44(11): 4175-4182(in Chinese).
- [19] WANG Jidong, WU Jiahui, KONG Xiangyu. Multi-agent simulation for strategic bidding in electricity markets using reinforcement learning[J]. CSEE Journal of Power and Energy Systems, 2023, 9(3): 1051-1065.
- [20] ZHU Ziqing, CHAN K W, BU Siqi, et al. Analysis of evolutionary dynamics for bidding strategy driven by multi-agent reinforcement learning[J]. IEEE Transactions on Power Systems, 2021, 36(6): 5975-5978.
- [21] LIANG Yanchang, GUO Chunlin, DING Zhaohao, 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.
- [22] DU Yan, LI Fangxing, ZANDI H, et al. Approximating Nash equilibrium in Day-ahead electricity market bidding with multi-agent deep reinforcement learning[J]. Journal of Modern Power Systems and Clean Energy, 2021, 9(3): 534-544.
- [23] JIA Qiangang, LI Yiyan, YAN Zheng, et al. A reinforcement-learning-based bidding strategy for power suppliers with limited information[J]. Journal of Modern Power Systems and Clean Energy, 2022, 10(4): 1032-1039.
- [24] 员江洋, 杨明, 刘宁宁, 等. 不完全信息下基于多代理深度确定策略梯度算法的发电商竞价策略[J]. 电网技术, 2022, 46(12): 4832-4842.  
YUN Jiangyang, YANG Ming, LIU Ningning, et al.

- Bidding strategy of generation companies based on multi-agent deep deterministic policy gradient algorithm under incomplete information[J]. *Power System Technology*, 2022, 46(12): 4832-4842(in Chinese).
- [25] WEI Xiangyu, XIANG Yue, LI Junlong, et al. Self-dispatch of wind-storage integrated system: a deep reinforcement learning approach[J]. *IEEE Transactions on Sustainable Energy*, 2022, 13(3): 1861-1864.
- [26] XU Hanchen, SUN Hongbo, NIKOVSKI D, et al. Deep reinforcement learning for joint bidding and pricing of load serving entity[J]. *IEEE Transactions on Smart Grid*, 2019, 10(6): 6366-6375.
- [27] XU Hongsheng, WU Qiuwei, WEN Jinyu, et al. Joint bidding and pricing for electricity retailers based on multi-task deep reinforcement learning[J]. *International Journal of Electrical Power & Energy Systems*, 2022, 138: 107897.
- [28] ZHANG Zhenyuan, CHEN Zihan, LEE W J. Soft actor-critic algorithm featured residential demand response strategic bidding for load aggregators[J]. *IEEE Transactions on Industry Applications*, 2022, 58(4): 4298-4308.
- [29] GAO Xiang, CHAN K W, XIA Shiwei, et al. A multiagent competitive bidding strategy in a pool-based electricity market with price-maker participants of WPPs and EV aggregators[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(11): 7256-7268.
- [30] ZHU Ziqing, CHAN K W, XIA Shiwei, et al. Optimal bi-level bidding and dispatching strategy between active distribution network and virtual alliances using distributed robust multi-agent deep reinforcement learning[J]. *IEEE Transactions on Smart Grid*, 2022, 13(4): 2833-2843.
- [31] TAO Yuechuan, QIU Jing, LAI Shuying. Deep reinforcement learning based bidding strategy for EVAs in local energy market considering information asymmetry [J]. *IEEE Transactions on Industrial Informatics*, 2022, 18(6): 3831-3842.
- [32] 蒋岚翔. 基于互利的发电和售电交易主体竞合均衡分析及优化策略研究[D]. 贵阳: 贵州大学, 2020.  
JIANG Lanxiang. Research on competition and cooperation equilibrium analysis and optimization strategy of power generation and power retailers transactions based on mutual benefit[D]. Guiyang: Guizhou University, 2020(in Chinese).
- [33] WANG Ziyu, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]// *Proceedings of the 33rd International Conference on Machine Learning*. New York: PMLR, 2016: 1995-2003.
- [34] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]// *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018: 1856-1865.
- [35] 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): 819-832.  
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): 819-832(in Chinese).
- [36] LEE D, HE Niao, KAMALARUBAN P, et al. Optimization for reinforcement learning: from a single agent to cooperative agents[J]. *IEEE Signal Processing Magazine*, 2020, 37(3): 123-135.
- [37] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]// *International conference on machine learning*. PMLR, 2018:1861-1870.
- [38] FUJIMOTO S, VAN HOOF H, MEGER D. Addressing function approximation error in actor-critic methods[C]// *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018: 1582-1591.
- [39] UW Electrical Engineering. Power systems test case archive-118 bus power flow test case [EB/OL]. [1993-05-01]. [http://labs.ece.uw.edu/pstca/pf118/pg\\_tca118bus.htm](http://labs.ece.uw.edu/pstca/pf118/pg_tca118bus.htm).
- [40] MURTAGH F, CONTRERAS P. Algorithms for hierarchical clustering: an overview[J]. *WIREs: Data Mining and Knowledge Discovery*, 2012, 2(1): 86-97.

## 附录 A

表 A1 发电商信息

Table A1 Information of generators

发电商 编号	所在 节点	二次项系数/ (元/(MW <sup>2</sup> h))	一次项系数/ (元/(MW·h))	机组容量/ MW
1	1	0.015	16	200
2	1	0.018	15	300
3	3	0.024	23	400
4	4	0.022	24	350
5	5	0.016	20	600

表 A2 两阶段深度强化学习算法参数设置

Table A2 Parameters of two-stage deep reinforcement learning algorithm

参数应用范围	参数名称	参数设置
通用	训练总周期	300
	单个周期训练次数	50
	报价系数上限	2
	初始随机次数	2000
	经验回放池容量	5000
	单次训练样本数	64
	后续奖励折扣	0.25
	软更新率	0.00002

续表

参数应用范围	参数名称	参数设置
Dueling-DQN 算法	隐藏层规模	128*3
	学习率	0.0002
	随机探索概率初值	1.0
	随机探索概率终值	0.05
Soft Actor Critic 算法	actor 隐藏层规模	128*4
	actor 网络学习率	0.0008
	critic 隐藏层规模	512*3
	critic 网络学习率	0.001

表 A3 不同合谋收益系数的发电商合谋竞价结果

Table A3 Collusion simulation results with different collusive profit coefficients

阻塞情况	合谋收益系数	发电商	报价系数	合谋组合	节点电价/(元/(MW·h))	个体收益/元
无阻塞	0.5	2	1.22	A	29.46	3449
		3	1.11	B	29.48	623
		4	1.10	B	29.30	425
		5	1.21	A	29.00	1839
	0.8	2	1.40	A	30.80	3929
		3	1.18	B	31.30	960
		4	1.16	B	31.10	729
		5	1.30	A	30.79	2060
	1.0	2	1.58	A	33.11	4622
		3	1.25	B	33.63	1412
		4	1.25	B	33.43	1006
		5	1.41	A	33.10	2469
支路 1-2 潮流限制 300MW	0.5	2	1.39	A	39.15	6436
		3	1.98	A	51.98	3710
		4	1.28	B	43.78	5576
		5	1.97	B	39.97	358
	0.8	2	1.32	A	23.81	1231
		3	1.95	B	52.32	4327
		4	1.16	A	34.09	1812
		5	1.08	B	25.58	862
	1.0	2	1.26	A	23.54	1372
		3	1.98	B	53.58	4831
		4	1.19	A	34.39	1767
		5	1.09	B	25.42	780

表 A4 2383 节点独立策略竞价结果

Table A4 Simulation results in 2383-node network

发电商编号	报价系数	节点电价/(元/(MW·h))	报价收益/元
2	1.22 (1.00)	146.3 (137.1)	70171 (63481)
3	1.20 (1.00)	146.4 (137.4)	51765 (42017)
4	1.13 (1.00)	146.1 (128.7)	16958 (0)
7	1.91 (1.00)	146.1 (118.1)	74845 (46870)
10	1.14 (1.00)	147.8 (134.2)	29713 (24118)
13	1.72 (1.00)	140.4 (127.6)	46598 (38268)
15	1.18 (1.00)	146.6 (136.3)	48544 (40859)
32	1.00 (1.00)	139.3 (145.3)	40324 (44397)
33	1.00 (1.00)	149.1 (147.6)	1577 (0)
37	1.00 (1.00)	144.1 (142.4)	86440 (85433)

表 A5 2383 节点第二次自由联合仿真结果

Table A5 The second simulation results of the 2383 node network

发电商	报价系数	合谋组合	节点电价/(元/(MW·h))	个体收益/元	收益提升/元
2	1.58	A	147.3	70835	664
3	1.00	B	147.3	52734	969
4	1.14	A	146.4	18784	1826
7	1.42	B	148.6	77395	2550
10	1.25	C	148.3	29885	172
13	1.14	D	141.1	47038	440
15	1.36	D	136.4	40912	-7632
32	1.25	A	168.8	60625	20301
33	1.29	B	174.6	1890	313
37	1.25	C	157.5	94483	8043



刘飞宇

在线出版日期: 2023-06-11。

收稿日期: 2022-10-31。

作者简介:

刘飞宇(1998), 男, 硕士研究生, 主要研究电力市场竞价模拟方向, liufeiyu003@163.com;

\*通信作者: 王蓓蓓(1979), 女, 博士, 副教授, 博士生导师, 主要从事电力市场、需求侧管理方面的研究和教学工作, wangbeibei@seu.edu.cn。

(实习编辑 王正一)

# Study on Free Joint Bidding Mechanism in Multi-agent Environment Based on Two-stage Deep Reinforcement Learning Algorithm

LIU Feiyu<sup>1</sup>, WANG Jiwen<sup>2</sup>, WANG Zhengfeng<sup>2</sup>, WANG Beibei<sup>1</sup>

(1. School of Electrical Engineering, Southeast University;

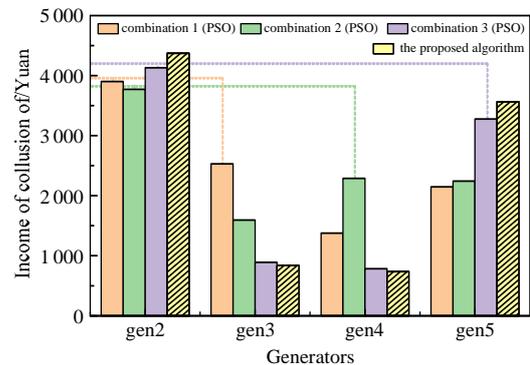
2. State Grid Anhui Electric Power Co., Ltd.)

**KEY WORDS:** two-stage deep reinforcement learning; free association; multi-agent simulation; joint auction behaviors

At the initial stage of power market construction, the imperfect regulation mechanism provides the opportunity of secret communication and joint bidding between generators. In order to find these potential collusions, this paper has proposed a novel collusive bidding model and the corresponding solution algorithm. The proposed method can effectively simulate the free association behaviors of market players and find potential collusion.

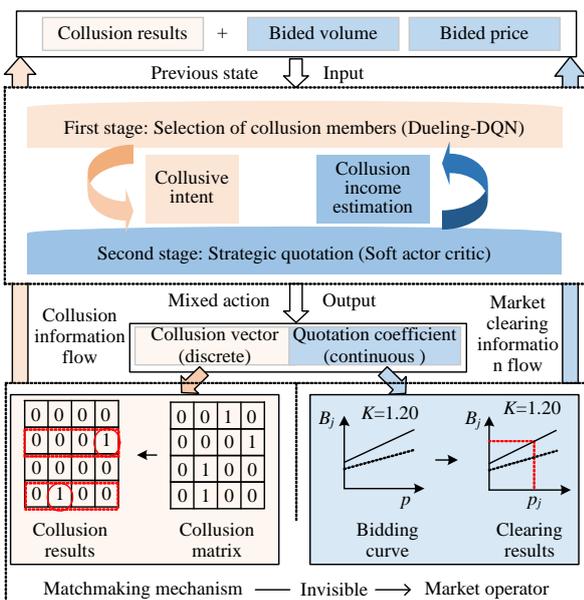
This paper combines the collusion model with the strategic bidding model to model the free association behaviors more accurately. A novel two-stage deep reinforcement learning algorithm is used to solve the discrete-and-continuous decision problem formed by the selections of discrete collusion objects and the determination of continuous bidding coefficient. The algorithm framework is shown in Fig. 1.

To verify the effectiveness of the proposed method, the simulation results based on the two-stage deep reinforcement learning algorithm and particle swarm optimization algorithm in the 5-node network are compared, as shown in Fig. 2. We also study the free association behaviors in the case of line congestion. In order to further verify the scalability of the algorithm, the simulations with larger network scale and larger collusion combinations are studied.



**Fig. 2 Comprehensive utility of collusion under different collusion combinations**

The simulation results show that the method proposed can simulate the free association behaviors of market players in the long-term bidding process. Further simulation results show that the proposed method in this paper is still effective in larger network scale cases. Meanwhile, it can realize the expansion of collusion combinations through multiple simulations. The method proposed in this paper can find unconventional collusion combinations and reduce the difficulty of analyzing potential collusion combinations, which can help market operators to design market power control mechanisms.



**Fig. 1 Two-stage bidding algorithm considering free association**