# A Data-driven Multi-agent Framework for Electricity Market Simulation Using Practical and Heterogeneous Bidding Strategy Spaces

Hanning Mi, Ji Qiao, Ye Li, Zibo Wang, Fujia Han, and Sijie Chen, *Member, IEEE*

***Abstract*—Existing electricity market simulation research assumes that energy providers share theoretical and homogeneous strategy spaces. However, practical bidding strategies differ significantly from those based on theoretical simulation. The factors that determine energy providers' bidding patterns are heterogeneous. Ignorance of practical and heterogeneous bidding strategy spaces distorts real market results. To fill the research gap, this paper proposes a data-driven multi-agent framework for electricity market simulation using practical and heterogeneous bidding strategy spaces. Firstly, a clustering algorithm is proposed for extracting practical and heterogeneous bidding strategy spaces from historical bidding strategies. The bidding strategy spaces are formed by the centers. Then, energy providers are modeled by agents with practical and heterogeneous strategy spaces in a multi-agent framework. Finally, the proximal policy optimization is implemented by agents to model the bidding behaviors in markets. Numerical results based on real data from the Australian Energy Market Operator show that market clearing results under practical and heterogeneous strategy spaces are closer to reality than those under theoretical and homogeneous strategy spaces.**

***Index Terms*—Electricity market, simulation, bidding, strategy space, clustering, data-driven, multi-agent, proximal policy optimization.**

## I. Introduction

COMPLEX electricity market environment [1]-[4] improves the importance of market simulation. Independent system operators (ISOs) can evaluate the rationality of market regulations and market trends. Market participants can investigate the optimal bidding strategies that maximize their profits. Energy providers' strategic bidding process greatly influences the market simulation results. Existing market simulation research can be categorized according to the method to simulate the strategic bidding process. In the first category, game theory is used to model the strategic bidding process. Games can be classified as Cournot games with energy quantity strategies [5], [6], supply function equilibrium games with quadratic supply function strategies [7], and games with price-quantity pair strategies [8]. In the second category, multi-agent frameworks [9]-[12] are used to model the strategic bidding process. Massive market data are used to train agents that represent market participants. The increasing renewable energy penetration has different influences on different market participants' bidding strategies. For example, traditional generators are faced with "missing money" problem caused by low electricity prices [13], while uncertainties of renewable energy constrain their bidding competition [14]. Compared with traditional game theory, the multi-agent framework simulates markets from an individual perspective and provides a method to incorporate heterogeneous market participants. Therefore, this paper focuses on multi-agent frameworks.

Literature has investigated market simulation based on multi-agent frameworks. In [9], a multi-agent framework simulates day-ahead markets where energy providers have incomplete information. Energy providers determine optimal price-quantity pairs with fixed quantities to maximize their profits. In [10], generators in markets are modeled by $Q$-learning approaches, considering their learning and risk aversion. Generators submit one price-quantity pair with the quantity fixed as the maximum capacity. In [11], the Nash equilibrium of markets with multiple strategic energy providers is approximated. Prices of all energy blocks are formatted by multiplying the costs by one strategic coefficient. In [12], an experience-weighted attraction reinforcement learning approach models energy providers with one price-quantity bid. The quantity is fixed and prices are divided into 20 discrete values evenly. These studies simulate electricity markets from a multi-agent perspective.

However, two research gaps remain in existing market simulation research. First, the bidding strategy spaces of energy providers are based on theoretical assumptions. According to bidding behavior analysis based on real bidding data,

Manuscript received: March 13, 2025; revised: June 19, 2025; accepted: August 11, 2025. Date of CrossCheck: August 11, 2025. Date of online publication: September 5, 2025.

This work was supported by the Science and Technology Program of State Grid Corporation of China "Research on Power Grid Situation Cognition and Operational Risk Assessment Technology Based on Probability Statistics and Deep Learning" (No. 5700-202455339A-2-1-ZX).



H. Mi, J. Qiao, Y. Li, Z. Wang, and F. Han are with the China Electric Power Research Institute, Beijing, China (e-mail: mihanning@163.com; qiaoji@epri.sgcc.com.cn; liye@epri.sgcc.com.cn; wangzibo@epri.sgcc.com.cn; hanfujia@epri.sgcc.com.cn).

S. Chen (corresponding author) is with the Key Laboratory of Control of Power Transmission and Conversion, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China, and he is also with the Shanghai Non-carbon Energy Conversion and Utilization Institute, Shanghai Jiao Tong University, Shanghai, China (e-mail: Sijie.chen@sjtu.edu.cn).

DOI: 10.35833/MPCE.2025.000212

practical bidding strategies differ greatly from those based on theoretical assumptions [15]. The difference can be caused by many practical factors such as contract capacity [16] and previous revenues from earlier markets [17]. Second, the bidding strategy spaces of energy providers are homogeneous. The factors that determine energy providers' bidding patterns are heterogeneous, such as the reward functions [18] and risk preference [19]. Ignorance of practical and heterogeneous strategy spaces distorts real markets and undermines the practicality of market simulation results.

The research gaps can be filled by combining data-driven methods and multi-agent frameworks. Many ISOs such as the Australian Energy Market Operator (AEMO) and the Midcontinent Independent System Operator (MISO) publish individual bidding data, which can be used for constructing bidding strategy spaces. The historical bidding strategies can be clustered. The cluster centers can form energy providers' practical and heterogeneous strategy spaces. Energy providers can be modeled by agents with practical and heterogeneous strategy spaces in multi-agent frameworks. These agents can employ reinforcement learning (RL) approaches. Specifically, they interact with environments (electricity markets) with limited information (own market clearing results) [20] to gradually learn the optimal bidding patterns. In this way, markets can be simulated to be closer to reality. To the best of our knowledge, little research simulates electricity market based on practical and heterogeneous bidding strategy spaces from a multi-agent perspective.

The main contributions of this paper are threefold:

1) A data-driven method is proposed for extracting practical and heterogeneous bidding strategy spaces from historical bidding data. Original bidding data published by ISOs are standardized and clustered by a granular computing-based clustering algorithm. Cluster centers form the practical and heterogeneous bidding strategy spaces for energy providers.

2) A multi-agent framework is proposed for electricity market simulation. Energy providers are modeled by agents with practical and heterogeneous bidding strategy spaces. Agents implement proximal policy optimization (PPO) algorithm to participate in markets with limited information.

3) Numerical experiments based on real-world data from the AEMO show that market simulation results under practical and heterogeneous strategy spaces are closer to reality than those under theoretical and homogeneous strategy spaces. The multi-agent PPO (MAPPO) framework outperforms other frameworks in revealing market clearing results.

The remainder of this paper is organized as follows. Section II introduces the practical bidding strategy space extraction based on data-driven methods. Section III details the proposed MAPPO framework. Case studies are shown in Section IV. Section V concludes this paper.

## II. Practical Bidding Strategy Space Extraction Based on Data-driven Methods

This section details the data-driven method for extracting bidding strategy spaces from the historical bidding data published by ISOs. The main idea is to constitute practical and heterogeneous bidding strategy spaces with centers clustered by historical bidding data. First, original bidding data are standardized to be suitable for clustering. Second, a metric that measures the difference between bidding strategies is proposed to be used in the clustering algorithm. Third, a granular computing-based clustering algorithm is designed to cluster bidding strategy centers. Fourth, the method to aggregate small generators into one aggregated generator is introduced. The aggregation decreases the number of agents and the computation burden in the multi-agent framework without losing too much accuracy.

### *A. Bidding Data Standardization*

Bidding data are price-quantity pairs that indicate energy blocks at given prices, constituting step-wise bidding curves. The bidding data in piece-wise form can be converted to price-quantity pairs by the method in our previous work [21]. The original price-quantity pairs are standardized to be suitable for clustering as follows.

#### *1) Price Clipping*

The wide price range of original bidding data significantly degrades the clustering results. The bidding prices of generators range from negative thousands to positive thousands. The number of cluster centers will be huge without any procession of the original bidding prices.

The original bidding prices are clipped to decrease the number of cluster centers. Most of the procured (or not procured) energy blocks are with bidding prices obviously lower (or higher) than the market clearing prices. Whether energy blocks can be procured is determined by whether bidding prices are higher than market clearing prices, rather than the absolute bidding prices. Because the range of market clearing prices is much narrower than that of bidding prices, energy blocks with extremely low or high bidding prices hardly influence market clearing results. Bidding prices can be clipped according to the range of market clearing prices. Let $p^{\min}$ and $p^{\max}$ denote the minimum and maximum bidding prices. Bidding prices lower than $p^{\min}$ are set as $p^{\min}$, and higher than $p^{\max}$ are set as $p^{\max}$.

#### *2) Quantity Padding*

Quantity determines the length of bidding curves. It is challenging to quantify the clustering distance between bidding curves with different lengths. Original bidding quantities of generators vary at different time because capacity withholding is a common strategy used by strategic energy providers.

The original bidding quantities are padded to make the lengths of bidding curves the same. Energy providers may withhold capacity by submitting a quantity smaller than its capacity. Let $q_g^{\max}$ denote the capacity of generator $g$. If the maximum quantity of a submitted bidding curve is less than $q_g^{\max}$, the remaining capacities can be regarded as bidding at the ceiling price. The bidding quantities in bidding curves are padded to $q_g^{\max}$. All bidding curves are padded to the same length, making it possible to quantify the distance between two bidding curves.

#### *3) Price-quantity Pair Normalizing*

Price-quantity pairs are normalized to be suitable for clustering. The distance scale of bidding curves of different generators varies because capacities of generators differ. This makes it hard to select the distance threshold in clustering algorithms. To solve this problem, price-quantity pairs are normalized into the same scale as:

$$\begin{cases} q_{g,t,n} = \dfrac{Q_{g,t,n}}{q_g^{\max}} \\ p_{g,t,n} = \dfrac{P_{g,t,n} - p^{\min}}{p^{\max} - p^{\min}} \end{cases} \quad \forall g \in G, \forall t \in T, \forall n \in N \tag{1}$$

where $Q_{g,t,n}$ and $q_{g,t,n}$ are the bidding quantities of the $n^{\text{th}}$ quantity segment of generator $g$ at time $t$ before and after normalization, respectively; $P_{g,t,n}$ and $p_{g,t,n}$ are the bidding prices of the $n^{\text{th}}$ quantity segment of generator $g$ at time $t$ before and after normalization, respectively; $G$ is the set of generators in the market; $T$ is the set of time intervals in the bidding day; and $N$ is the set of quantity segments in a bidding curve.

## B. Distance Metric

A metric that measures the difference between normalized bidding curves is proposed for clustering. As shown in Fig. 1, normalized price-quantity pairs are plotted as bidding curves (curves 1 and 2). Normalized bidding curves enclose areas ranging from 0 to 1. The sum of areas is used for measuring the distance between bidding curves. The distance metric is calculated as (2).

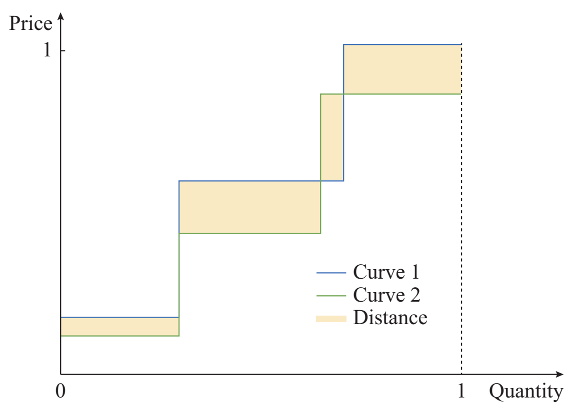


Fig. 1. Illustration of distance metric.

$$d(f_1, f_2, \ldots, f_i) = \int_0^1 (\max(f_1(q), f_2(q), \ldots, f_i(q)) - \min(f_1(q), f_2(q), \ldots, f_i(q)))\mathrm{d}q \tag{2}$$

where $d(f_1, f_2, \ldots, f_i)$ is the distance among bidding curves $f_1, f_2, \ldots, f_i$. The distance metric is small if the bidding curves are similar. Otherwise, the distance metric is large. The enclosed areas are the products of prices and quantities, which can be interpreted as the price difference $dif_p$ at a given quantity difference $dif_q$, which is expressed as:

$$d(f_1, f_2, \ldots, f_i) = dif_p \cdot dif_q \tag{3}$$

According to (3), the price difference of $dif_p$ of the quantity is at most $d(f_1, f_2, \ldots, f_i)/dif_q$. To be more specific, assume that the distance metric is 0.09. For the two most distinct curves among $f_1, f_2, \ldots, f_i$, the price difference of 100% quantities is at most 9%.

## C. Clustering Algorithm

The $K$-means algorithm and its variants are most widely used for clustering electricity market data [15], [22]. However, existing algorithms are not suitable for clustering bidding curves. First, the performance of existing algorithms is sensitive to initialized centers [23]. The bidding patterns of different generators vary greatly. Even for the same generator, bidding patterns at different time can vary greatly. It is hard to pre-estimate the number and position of initial centers to avoid local optima. Second, the lower bound of the representativeness of centers is not constrained. The representativeness of centers is determined by the size of the whole cluster. Existing algorithms allocate each curve to the nearest center [24], without considering how much the size of the cluster is increased during the allocation. The distances between centers and curves in some clusters can be large, resulting in a weak representativeness of centers. The centers with weak representativeness distort the bidding behavior of energy providers.

To overcome the aforementioned challenges, a clustering algorithm based on granular computing is proposed. The basic idea of granular computing is using granular balls to cover the data point space in the clustering process [25]. A granular ball is a data structure that contains many data points. Details about granular computing can be referred to [25], [26]. The proposed clustering algorithm is shown in Algorithm 1. Granular balls with the minimum distance are fused into one granular ball iteratively until the distances between all granular balls are larger than the distance threshold $D$. The curve with the smallest distance to other curves in a granular ball is selected as the center.

**Algorithm 1**: clustering algorithm based on granular computing

**Input**: standardized bidding curves $f_i$ and distance threshold $D$

**Output**: bidding curve granular balls $X$ with centers

Initialize bidding curve granular balls $X_i = \{f_i\}, \forall i$
**Repeat**
  **for all** $X_m$ **do**
    Calculate the distances between $X_m$ and other granular balls $X_n$: $d_{mn} = d(f_i, f_j), f_i \in X_m, f_j \in X_n$
  **end for**
  Find $X_m$ and $X_n$ with the minimum $d_{mn}$
  **if** $d_{mn} < D$ **then**
    Merge $X_m$ and $X_n$ into a new granular ball: $X = X \cup \{X_m \cup X_n\}$
    Remove $X_m$ and $X_n$: $X = X \setminus (X_m \cup X_n)$
  **end if**
**until** $\forall d_{mn} > D$
**for all** $X_m$ **do**
  **for all** $f_i \in X_m$ **do**
    Calculate the sum of distances between $f_i$ and other curves in $X_m$: $d_i = \sum_{f_j \in X_m, j \neq i} d(f_i, f_j)$
  **end for**
  Choose $f_i$ with the minimum $d_i$ as the center of $X_m$
**end for**

The proposed clustering algorithm can overcome the aforementioned challenges. First, no other parameters (such as the number and positions of initial centers) are expected except for $D$. The robustness of the proposed clustering algorithm is strong because there exists no randomness in the initializing process. Second, the lower bound of the representativeness of centers in clusters is regulated by $D$. The fusing process is terminated when the size of the smallest granular ball is larger than $D$ after merging. The distance between the center and other curves in the granular ball is less than $D$.

### *D. Data Aggregation of Small Generators*

Small generators with little market power are aggregated. In most electricity markets, few big generators account for a large share of the generating capacity of the market. These generators are modeled by agents in the multi-agent framework. Compared with these generators, the bidding curves of many other generators with small generating capacity hardly influence market clearing results. It is unnecessary and computationally costly to model all small generators as individual agents. Aggregation of small generators can decrease the number of agents in the multi-agent framework and the computation burden. An aggregated generator is used to represent small generators.

A method of aggregating the bidding data of small generators is proposed. First, the bidding curves of all small generators are constructed as an aggregated bidding curve. This curve can not be used for market clearing because the number of price-quantity pairs is more than the maximum segment regulated by ISOs. Second, the aggregated bidding curve is sampled at a fixed price interval in $[p^{\min}, p^{\max}]$, capturing the corresponding quantities. The number of sampled price-quantity pairs is set as the maximum segment to recover the bidding behavior of small generators with a high resolution. Third, the capacities of all small generators are summed as the capacity of the aggregated generator $q_g^{\max}$, which is used to normalize the sampled bidding curve by (1). So far, the bidding data of small generators are aggregated as the bidding pattern of one agent.

## III. Proposed MAPPO Framework

This section details the proposed MAPPO framework. The centers extracted in Section II constitute practical and heterogeneous bidding strategy spaces of energy providers. Energy providers are modeled by the agents that employ the PPO algorithm to participate in electricity markets with incomplete information.

### *A. Agent Settings*

Generation capacities quantify market power of generators and decide whether they are modeled as agents. Fossil-fueled generators are sorted by their capacities from highest to lowest, and then the top generators are designated as agents once their cumulative capacity accounts for the majority of the market share. The rest generators are regarded as small generators with little market power. These generators are aggregated as one agent by the method in Section II-D. Because the bidding prices of most renewable energy generators are equal to the price floor, their bidding strategy is selling all the predicted energy. All renewable energy is cleared in the proposed MAPPO framework.

The bidding process of agents can be modeled by RL approaches. At time $t$, agent $g$ observes the state $s_{g,t}$ in the environment. The agent takes action $a_{g,t}$ based on the policy learned by the RL approach. The actions of all agents affect the environment. Then, the agent receives the reward $r_{g,t}$ of taking the action, and observes new state $s_{g,t+1}$. The elements in this process are detailed as follows.

1) State. In electricity markets, the market clearing price $Pr_t$ and the cleared energy of an agent are influenced by the bidding strategy of the agent. The agent can refine the strategy based on these observable states. The state can be represented as:

$$s_{g,t}=\left\{Pr_t, \sum_{n\in N} q_{g,t,n}^c\right\} \quad \forall g\in G, \forall t\in T \tag{4}$$

where $q_{g,t,n}^c$ is the cleared quantity of the $n^{\text{th}}$ quantity segment.

2) Action. The actions of agents submit bidding curves to ISOs. The action space of an agent is the bidding strategy space formed by centers clustered by Algorithm 1:

$$a_{g,t}=\{f_i\} \quad \forall g\in G, \forall t\in T \tag{5}$$

3) Environment. The environment determines the state transition and the rewards of agents. After agents submit bidding curves, ISOs implement the economic dispatching (ED) program [27] to clear electricity markets. The results of ED program are the individual cleared energy and $Pr_t$. ED program is the environment and is detailed in Appendix A.

4) Reward. The reward determines the market participation patterns of agents. In the bidding process, the rewards of agents are profits. Agents bid to maximize their profits. The reward function of agent $g$ at time $t$ is designed as:

$$r_{g,t}=\sum_{n\in N} q_{g,t,n}^c(Pr_t-c_{g,n})-Pr_p\cdot(\Delta q_{g,t}^D+\Delta q_{g,t}^U) \tag{6}$$

where $c_{g,n}$ is the cost of the $n^{\text{th}}$ block; $Pr_p$ is the penalty for constraint violation; and $\Delta q_{g,t}^D$ and $\Delta q_{g,t}^U$ are the slack terms for ramp up and ramp down constraints, respectively. If the cleared energy under a strategy violates the ramp rate constraints (A3) and (A4), the dispatching result can not be implemented by the generator. The reward of this strategy is penalized by $Pr_p$. The total energy supply may not satisfy the energy demand, i.e., $\Delta q_t^L>0$, where $\Delta q_t^L$ is the slack term for energy balance constraints. This situation hardly occurs in reality but might occur in the early training process of agents. In this situation, the reward is set as a significant negative value.

### *B. PPO Algorithm*

PPO algorithm [28] is a policy gradient RL approach that explores the action space sufficiently with strong stability by constraining the policy update aptitude between adjacent iterations. This makes it have good performance in terms of learning stability and complexity [29]. PPO algorithm has been widely used in the power system research such as coordination of power distribution networks and microgrids [30], voltage control [31], and charging scheduling [32]. In the proposed MAPPO framework, PPO algorithm is used to train agents to maximize profits. In PPO algorithm, the advantage $A(s_t, a_t)$ denotes how much an action $a_t$ at state $s_t$ is better than other available actions, which is calculated as:

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t) \tag{7}$$

$$Q(s_t, a_t) = \mathbb{E}_{s_{t+1:\infty}, a_{t+1:\infty}} \sum_{m=0}^{\infty} \gamma^m r_{t+m} \tag{8}$$

$$V(s_t) = \mathbb{E}_{s_{t+1:\infty}} \sum_{m=0}^{\infty} \gamma^m r_{t+m} \tag{9}$$

where $Q(s_t, a_t)$ denotes the action value function; $r_{t+m}$ is the reward at time $t+m$; $V(s_t)$ denotes the value function; $\mathbb{E}$ denotes the average of a batch of samples $(S_{t+1:\infty}, a_{t+1:\infty})$ over the collected experience data; and $\gamma$ is the discount factor. A truncated version of generalized advantage estimation (GAE) [33] is used in PPO algorithm to estimate the advantage $A(s_t, a_t)$.

PPO algorithm features an actor-critic structure. The actor is a deep neural network (DNN) denoted by $\pi_\theta(a|s)$, which represents the policy of the agent on choosing action $a$ at state $s$ with parameter $\theta$. At the end of an episode, the agent with parameter $\theta_{old}$ samples a trajectory $\tau = \{s_t, a_t, r_t\}$. The actor employs a clipped surrogate objective function $J(\theta)$ as:

$$J(\theta) = \mathbb{E}_{\tau \sim \theta_{old}}[\min(r(\theta)A_t, clip(r(\theta), 1-\epsilon, 1+\epsilon)A_t)] \tag{10}$$

$$r(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \tag{11}$$

where $r(\theta)$ is the ratio between the updated parameter $\pi_\theta(a_t|s_t)$ and the old parameter $\pi_{\theta_{old}}(a_t|s_t)$; $A_t$ is the abbreviation of $A(s_t, a_t)$; $\epsilon$ is a hyperparameter that constrains the updating aptitude of the actor; and the clip function is shown as:

$$clip(r(\theta), 1-\epsilon, 1+\epsilon) = \begin{cases} 1-\epsilon & r(\theta) \le 1-\epsilon \\ 1+\epsilon & r(\theta) \ge 1+\epsilon \\ r(\theta) & \text{otherwise} \end{cases} \tag{12}$$

According to (12), $J(\theta)$ is limited to prevent the updated policy far from $\theta_{old}$. The gradient ascent method is used to maximize the objective function of the actor [34]. $\theta$ of an actor is updated as:

$$\theta = \theta_{old} + \alpha \nabla_\theta J(\theta) \tag{13}$$

where $\alpha$ is the learning rate; and $\nabla_\theta J(\theta)$ is the gradient of $J(\theta)$ to $\theta$.

The critic is a DNN denoted by $V_\phi(s)$, representing the value of state $s$ with parameter $\phi$. The critic aims to minimize the difference between the estimated value of action and a target value $\hat{R}_t$. The critic employs the loss function $L(\phi)$ based on the mean squared error regression as:

$$L(\phi) = \mathbb{E}_{\tau \sim \theta_{old}} (V_\phi(s_t) - \hat{R}_t)^2 \tag{14}$$

$$\hat{R}_t = r_t + \gamma V_\phi(s_{t+1}) \tag{15}$$

$\phi$ in a critic is updated as:

$$\phi = \phi_{old} + \alpha \nabla_\phi L(\phi) \tag{16}$$

### *C. Market Simulation Framework: MAPPO*

The proposed MAPPO framework is shown in Algorithm 2.

**Algorithm 2**: proposed MAPPO framework

**Input**: historical bidding data, load of target day, and hyperparameters of PPO algorithm

**Output**: market clearing results at time $t$

1. Set market simulation intervals $T$, episodes $E$, update times $K$, and initial cleared energy of generator $g$ $q_g^0$
2. **for** $t = 1{:}T$ **do**
3.   **for** $g \in G$ **do**
4.     Extract bidding strategy spaces of agent $g$ at time $t$ from historical bidding data by algorithm 1
5.     Randomly choose initial action $a_{g,t}^0$
6.     Initialize $\theta$ of actors and $\phi$ of critics
7.   **end for**
8.   Clear markets by the ED
9.   Obtain initial state $s_{g,t}^0, \forall g \in G$
10.   **for** $e = 1{:}E$ **do**
11.     **for** $g \in G$ **do**
12.       Agents act based on policy network of actors
13.     **end for**
14.     Clear markets by ED and update state $s_{g,t}^e, \forall g \in G$
15.     Calculate reward $r_{g,t}^e$ according to (6)
16.     **for** $g \in G$ **do**
17.       Record trajectories $\tau$
18.       **for** $k = 1{:}K$ **do**
19.         Calculate $J(\theta)$ according to (10)
20.         Update $\theta_{g,t}$ according to (13)
21.         Calculate $L(\phi)$ according to (14)
22.         Update $\phi_{g,t}$ according to (16)
23.       **end for**
24.     **end for**
25.   **end for**
26.   Obtain market clearing results with updated $\theta$
27. **end for**

As shown at step 4, the bidding strategy space of an agent is extracted based on the historic bidding data of the generator at this time. Generators can submit different bidding curves at different time. One generator is modeled as different agents at different time. In the markets where the generator can only submit one bidding curve in a day, there is no need to extract different bidding strategy spaces at different time. The hyperparameters and actions of agents are initialized at steps 5 and 6. The market is cleared by the ED program at step 9 after all agents choose the actions, rather than one agent choosing an action. This is because agents make bidding actions simultaneously. Then, agents are trained by PPO algorithm from step 10 to step 25. At each episode, agents choose actions at step 12 and get rewards at step 15. The strategies are improved from step 18 to step 23, as shown in Section III-B. When the bidding process is completed, the market simulation result at time $t$ is obtained at step 26. As shown in Algorithm 2, there is no limit to the number of agents in the proposed MAPPO framework. However, in the markets with a large number of generators, the computational cost of modeling all generators as individual agents is huge. A practical solution is aggregating bidding strategies of generators belonging to the same company as strategies of one agent. In this way, the proposed MAPPO framework models bidding strategies of companies, rather than individual generators. The accuracy of MAPPO inevitably decreases in these markets.

## IV. Case Studies

### A. Case Settings

The market data from the New South Wales in AEMO [35] are used to test the proposed MAPPO framework. The market clearing results from 04:00 on October 7, 2022 to 04:00 on October 8, 2022 are investigated. The time precision is half an hour. The historical bidding data from July 6, 2022 to October 6, 2022 are used to extract the bidding strategy spaces of agents. The distance threshold is set to be $D=0.09$. The first 13 dispatched largest thermal generators and an aggregated generator make up 14 agents. The details about 14 agents can be referred to [36]. Agents 3 and 10 become online as soon as they submit non-zero quantities, and agent 13 keeps online until it submits the zero quantity. This indicates that the three agents actively choose whether to be dispatched. Because the private information to reveal these bidding behaviors is not accessible, the bidding quantities of agents 3, 10, and 13 during their offline time are fixed as 0. $c_{g,n}$ is set to be 29.07 \$/MWh for coal-fueled generators and 174.66 \$/MWh for gas-fueled generators [37]. The maximum and minimum bidding prices are set to be $p^{\max}=300$ \$/MWh and $p^{\min}=50$ \$/MWh, respectively. The hyperparameters of agents and the PPO algorithm are shown in Table I. Real market results are obtained by using real bidding data in the ED model. The ED is programmed by general algebraic modeling system (GAMS) and solved by solver CPLEX. All RL approaches are implemented by Pytorch.

TABLE I
HYPERPARAMETERS OF AGENTS AND PPO ALGORITHM

| Parameter | Value |
|---|---|
| Agent layer number | 3 |
| Activation function | ReLU |
| Hidden layer size | 32 |
| $\epsilon$ | 0.2 |
| $E$ | 2500 |
| $\alpha$ | 0.001 |
| $K$ | 5 |
| $\gamma$ | 0.9 |
| $T$ | 48 |

### B. Practical Bidding Strategy Space Extraction

The every-half-hour bidding strategy spaces of the 14 agents are extracted. The number of clustered centers represents the strategy space size. The strategy space sizes of all agents at different time are shown in Fig. 2. Practical bidding strategy spaces are heterogeneous. Agents 1, 2, and 3 take more bidding strategies. In contrast, agents 10-14 take fewer bidding strategies, indicating that the bidding behavior of these agents is more predictable. The bidding strategy space of one agent varies at different time. It is necessary to extract practical and heterogeneous bidding strategy spaces based on real market data to improve the practicality of the market clearing results.

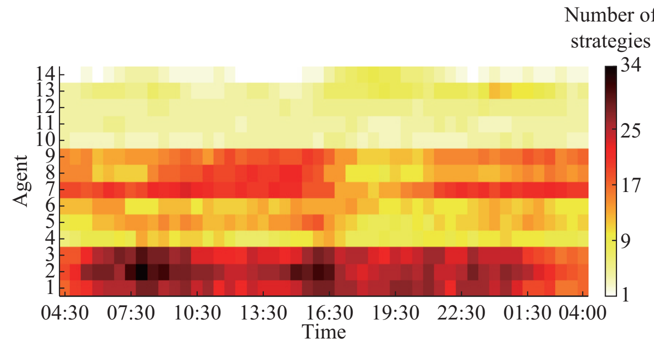


Fig. 2. Strategy space sizes of all agents at different time.

As a variant of the $K$-means algorithm, the $K$-medoids algorithm with the proposed distance metric (denoted as A1) is compared with the proposed clustering algorithm based on granular computing (denoted as A2). The number of clusters in A1 is set as the value in A2 for comparison. The clustering results of agent 10 at 14:30 by different clustering algorithms are shown in Fig. 3.

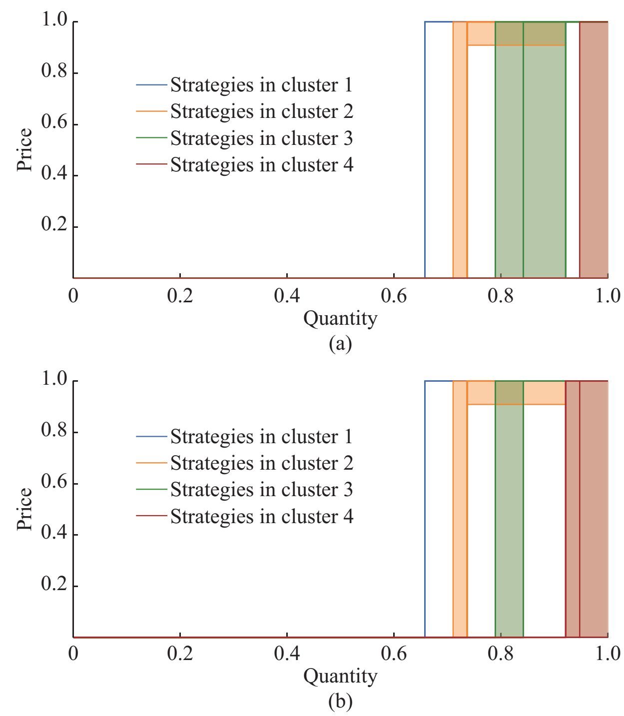


Fig. 3 Clustering results of agent 10 at 14:30 by different clustering algorithms. (a) Results by A1. (b) Results by A2

The areas enclosed by curves in one cluster are shaded. Different from theoretical assumptions, practical bidding strategies only feature 9 shapes. All bidding strategies are clustered into 4 clusters. The average area enclosed by A2 and the variance of areas in A2 is smaller. The performance of A2 is better than that of A1 with suboptimal initialized centers. The average areas enclosed by bidding curves in one cluster for each agent at every time (48 areas per agent) are calculated. The maximums, variances, means, and weighted Calinski-Harabasz and Silhouette coefficients [38] of 48 areas of every agent are shown in Table II, where the best indexes are bold. All maximums in A2 are not larger than $D = 0.09$. The maximums for 12 of 14 agents under A2 are less than those under A1, proving the effectiveness of A2 in limiting the lower bound of the representativeness of centers in clusters. The variances for 12 of 14 agents under A2 are also less than those under A1 because the maximums are restricted in A2. The means of 8 agents are slightly increased under A2. In A2, the cluster with the smallest area continues to embrace the unclassified bidding curves until the maximum area of all clusters reaches $D = 0.09$. The areas of small clusters are increased, and the area differences between different clusters are reduced. The performance of A1 and A2 on weighted Calinski-Harabasz and Silhouette coefficients is similar. This means the intra-cluster compactness and inter-cluster separation of A1 and A2 are similar. To sum up, the proposed clustering algorithm can constrain the lower bound of the representativeness of clusters without sacrificing the intra-cluster compactness and inter-cluster separation, compared with traditional clustering algorithms.

TABLE II
COMPARISON OF ALGORITHMS A1 AND A2

| Agent | Maximum | | Variance | | Mean | | Weighted Calinski-Harabasz coefficient | | Weighted Silhouette coefficient | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A1 | A2 | A1 | A2 | A1 | A2 | A1 | A2 | A1 | A2 |
| 1 | 0.12 | **0.09** | 0.0014 | **0.0006** | **0.045** | 0.050 | 30 | 34 | **0.35** | 0.34 |
| 2 | 0.12 | **0.09** | 0.0015 | **0.0006** | **0.047** | 0.052 | **25** | 24 | **0.33** | **0.33** |
| 3 | 0.12 | **0.09** | 0.0014 | **0.0006** | **0.046** | 0.051 | **23** | **23** | 0.33 | **0.34** |
| 4 | 0.10 | **0.08** | 0.0013 | **0.0006** | **0.044** | 0.048 | 307 | **478** | **0.51** | **0.51** |
| 5 | 0.10 | **0.08** | 0.0012 | **0.0006** | **0.051** | 0.053 | **123** | 77 | 0.40 | **0.42** |
| 6 | 0.10 | **0.08** | 0.0009 | **0.0006** | **0.048** | **0.048** | 96 | **111** | 0.50 | **0.51** |
| 7 | 0.12 | **0.09** | 0.0014 | **0.0006** | **0.048** | 0.049 | 35 | **38** | **0.46** | **0.46** |
| 8 | 0.12 | **0.08** | 0.0015 | **0.0007** | **0.047** | 0.048 | 71 | **94** | 0.46 | **0.47** |
| 9 | 0.11 | **0.09** | 0.0014 | **0.0007** | **0.045** | 0.048 | **55** | 54 | **0.44** | **0.44** |
| 10 | 0.09 | **0.08** | 0.0011 | **0.0007** | 0.058 | **0.053** | **55** | 32 | 0.69 | **0.72** |
| 11 | 0.09 | **0.08** | 0.0014 | **0.0008** | 0.050 | **0.046** | **103** | 47 | **0.73** | 0.71 |
| 12 | **0.06** | **0.06** | **0.0006** | **0.0006** | **0.034** | **0.034** | **360** | 287 | **0.71** | **0.71** |
| 13 | 0.08 | **0.07** | 0.0010 | **0.0008** | **0.036** | **0.036** | 92 | **102** | **0.68** | 0.67 |
| 14 | **0.05** | **0.05** | **0.0004** | **0.0004** | **0.027** | **0.027** | 5160 | **6098** | 0.61 | **0.62** |

A sensitivity analysis of distance threshold $D$ is made. The strategy space distribution under $D = 0.05$, $D = 0.09$, and $D = 0.15$ is tested, respectively. The comparison is shown in Table III. The variance decreases and the number of strategies increases under $D = 0.05$. The representativeness of strategies is improved. Suppose $D$ is an extremely small value, every single strategy is categorized into one cluster, and there is no need to implement the clustering algorithm. However, a large number of strategies increase the computation burden of the proposed MAPPO framework. The variance of strategy increases and the number of strategies decreases under $D = 0.15$, sacrificing the representativeness of strategies. $D = 0.09$ is selected as a compromise between representativeness of clustered centers and computation burden of the proposed MAPPO framework.

TABLE III
SENSITIVITY ANALYSIS OF DISTANCE THRESHOLD $D$

| Agent | $D = 0.05$ | | $D = 0.09$ | | $D = 0.15$ | |
|---|---|---|---|---|---|---|
| | Variance | Number | Variance | Number | Variance | Number |
| 1 | 0.0003 | 38 | 0.0006 | 24 | 0.0011 | 14 |
| 2 | 0.0003 | 43 | 0.0006 | 26 | 0.0010 | 14 |
| 3 | 0.0003 | 38 | 0.0006 | 23 | 0.0010 | 13 |
| 4 | 0.0003 | 15 | 0.0006 | 9 | 0.0012 | 6 |
| 5 | 0.0002 | 21 | 0.0006 | 12 | 0.0012 | 7 |
| 6 | 0.0002 | 18 | 0.0006 | 12 | 0.0016 | 7 |
| 7 | 0.0003 | 29 | 0.0006 | 19 | 0.0011 | 12 |
| 8 | 0.0003 | 24 | 0.0007 | 15 | 0.0013 | 9 |
| 9 | 0.0003 | 24 | 0.0007 | 15 | 0.0012 | 9 |
| 10 | 0.0002 | 6 | 0.0007 | 4 | 0.0008 | 3 |
| 11 | 0.0002 | 6 | 0.0008 | 4 | 0.0015 | 3 |
| 12 | 0.0001 | 7 | 0.0006 | 5 | 0.0012 | 3 |
| 13 | 0.0003 | 8 | 0.0008 | 6 | 0.0019 | 4 |
| 14 | 0.0002 | 4 | 0.0004 | 3 | 0.0008 | 2 |

### C. Strategy Space Comparison

The ability of practical and heterogeneous strategy spaces to reveal market clearing results is tested. If differences between the cleared energy quantity and real market results are minor, the strategy space can better reveal market clearing results. The methods to construct theoretical strategy spaces in [10] and [11] are implemented for comparison.

The strategies in [10] are in the form of one price-quantity pair. In [11], bidding prices are determined by multiplying the cost of each block with one strategic factor. Details can be referred to [36]. The percentage of the absolute difference between cleared energy quantity and real market results relative to real market results is defined as average difference. Average differences under the proposed MAPPO framework by using different strategy spaces are shown in Table IV. The most minor differences in every row are bold. Ten of the 14 agents get the most minor differences with the proposed bidding strategy spaces.

TABLE IV
AVERAGE DIFFERENCES BETWEEN CLEARED ENERGY QUANTITY AND REAL MARKET RESULTS WITH DIFFERENT STRATEGY SPACES UNDER PROPOSED MAPPO FRAMEWORK

| Agent | Average difference (%) | | |
|---|---|---|---|
| | Proposed | Ref. [10] | Ref. [11] |
| 1 | 28.20 | 24.00 | **21.30** |
| 2 | **20.90** | 29.70 | 24.30 |
| 3 | 100.00 | **81.30** | 100.00 |
| 4 | **11.80** | 16.00 | 17.90 |
| 5 | **10.00** | 20.40 | 16.50 |
| 6 | **15.50** | 18.90 | 16.40 |
| 7 | **23.30** | 27.20 | 29.60 |
| 8 | **16.40** | 22.60 | 17.80 |
| 9 | **16.40** | 18.20 | 17.00 |
| 10 | 47.30 | 75.90 | **44.60** |
| 11 | **9.40** | 55.60 | 43.50 |
| 12 | **10.70** | 45.30 | 33.80 |
| 13 | **53.90** | 100.00 | 99.90 |
| 14 | 100.00 | 100.00 | 100.00 |
| All agents | **21.00** | 30.90 | 26.90 |

The average market clearing prices and the average profits of agents with different strategy spaces under the proposed MAPPO framework are shown in Table V.

TABLE V
AVERAGE MARKET CLEARING PRICES AND AVERAGE PROFITS OF AGENTS WITH DIFFERENT STRATEGY SPACES UNDER PROPOSED MAPPO FRAMEWORK

| Strategy space | Average profit ($\$10^6$) | Average market clearing price (\$/MWh) |
|---|---|---|
| Real | 1.11 (+0%) | 154.05 (+0%) |
| Proposed | 1.66 (+49.5%) | 210.12 (+36.4%) |
| Ref. [10] | 0.43 (−61.3%) | 74.74 (−51.5%) |
| Ref. [11] | 0.45 (−59.5%) | 76.24 (−50.5%) |

Note: the value in the parenthesis represents the percentage of the absolute difference between cleared market results and real market results relative to real market results.

The average market clearing prices and the average profits of agents under the proposed bidding strategy spaces are closer to real market results than those under theoretical strategy spaces. The average market clearing prices and the average profits of agents under theoretical strategy spaces are lower than real market results. Under theoretical strategy spaces, the agents bid with all energy capacities homogeneously. The energy supply far exceeds energy demands. The cleared energy quantity of agents can significantly increase by slightly lower bidding prices compared with other homogeneous competitors. The fierce competition forces agents to lower bidding prices, resulting in extremely low average market clearing prices and average profits of agents. The competition in real markets is not as fierce as theoretical assumptions. The proposed bidding strategy spaces outperform other theoretical strategy spaces in revealing market clearing results.

The performance of the proposed MAPPO framework in other scenarios is tested. The ratio of cleared renewable energy and thermal energy on October 9 is 35.6%, which is much higher than 23.8% of the benchmark day. This indicates the renewable energy output is very high. The ratio of load and thermal generator capacity on October 10 is 77.4%, which is much higher than 67.2% of the benchmark day. This indicates the supply and demand are in a state of tension. And the two adjacent dates constitute a multi-period scenario. There are 15 agents bidding in the market in these two days, different from 14 agents on the benchmark day. The results are shown in Table VI, where \ indicate that the agent does not sell energy in the market. The average differences are comparable to those in Table IV. The proposed MAPPO framework can perform well in extreme scenarios.

TABLE VI
PERFORMANCE OF PROPOSED MAPPO FRAMEWORK IN EXTREME SCENARIOS

| Agent | Average difference (%) | |
|---|---|---|
| | October 9 | October 10 |
| 1 | 17.10 | 18.90 |
| 2 | 20.40 | 16.30 |
| 3 | ≥100.00 | 22.30 |
| 4 | 10.70 | 9.70 |
| 5 | 9.30 | 9.00 |
| 6 | 10.50 | 6.60 |
| 7 | 14.50 | 21.20 |
| 8 | 21.90 | 18.10 |
| 9 | 53.20 | 18.50 |
| 10 | 22.50 | 12.70 |
| 11 | 23.70 | 30.70 |
| 12 | ≥100.00 | ≥100.00 |
| 13 | 3.10 | \ |
| 14 | 11.00 | \ |
| 15 | ≥100.00 | ≥100.00 |

### D. Multi-agent Framework Comparison

The following frameworks are implemented on the same bidding strategy spaces to show the merits of the proposed MAPPO framework.

1) Nash equilibrium-based (NE) framework: in game theory, Nash equilibrium is a state where no player can improve the profit by unilaterally changing his action. The strategies at the Nash equilibrium are used.

2) Multi-agent DDQN-based (MADDQN) framework: agents strategically bid with double deep $Q$ network (DDQN). DDQN overcomes the $Q$-value overestimation in

DQN. Compared with PPO algorithm, actions of agents are not clipped in DDQN. Details can be referred to [39].

The cumulative profits of all agents under different frameworks are shown in Fig. 4.

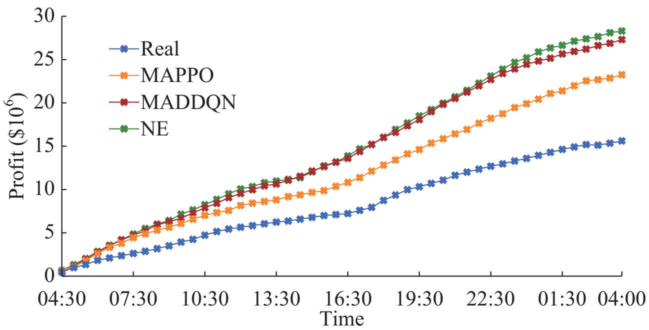


Fig. 4. Cumulative profits of all agents under different frameworks.

Compared with profits of MADDQN and NE frameworks, those under the proposed MAPPO framework are closer to real market profits. The proposed MAPPO framework outperforms other frameworks in depicting real market results. The agents under the NE framework have complete information about market clearing results. This strong assumption results in the highest overall profit. The more realistic incomplete information under the MADDQN and proposed MAPPO frameworks results in lower profits. Policy of agents updating in DDQN is not limited during the training process. The profits under the MADDQN framework are higher than those under the proposed MAPPO framework. Policy of agents updating in PPO algorithm is limited by the clip function during the training process. This is consistent with Fig. 2, which states that agents do not take too many bidding strategies. Most agents tend to take previous strategies to avoid drastic fluctuations in market clearing results. The assumption that agents can change bidding strategies arbitrarily results in deviations from real market results.

The profits of agents under different frameworks are shown in Fig. 5. Again, the average profits of all agents under the proposed MAPPO framework are the closest to real profits. Most agents get the highest profits under the NE framework, except for agents 1, 3, 8, and 13 who earn the highest profits under the MADDQN framework. The incomplete information environment under the MADDQN framework hinders the other ten agents from higher profits. The profits of agents 3, 13, and 14 are obviously less than those of other agents. This is because agent 3 only sells energy at the last 3 times. And the costs of agents 13 and 14 are obviously higher than those of other agents. Agents 4-9 make more profits. By analyzing Fig. 2 and Fig. 5, profitable strategy spaces vary at different time, rather than remaining constant. A too-large or too-small bidding strategy space can not significantly improve profits.

The convergences of the proposed MAPPO and MADDQN frameworks are compared. Market clearing prices during the training process at different time are shown in Fig. 6. The lines represent the mean values of 10 episodes, and the shaded areas represent the standard deviations. Both frameworks can converge after 2500 episodes.

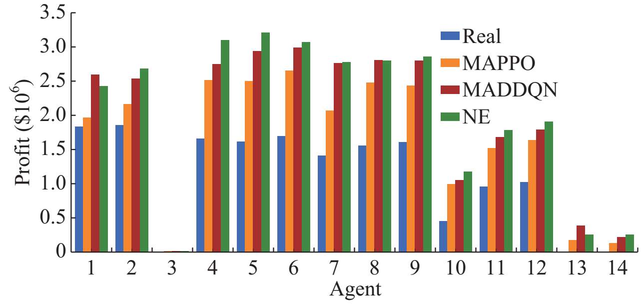


Fig. 5. Profits of agents under different frameworks.

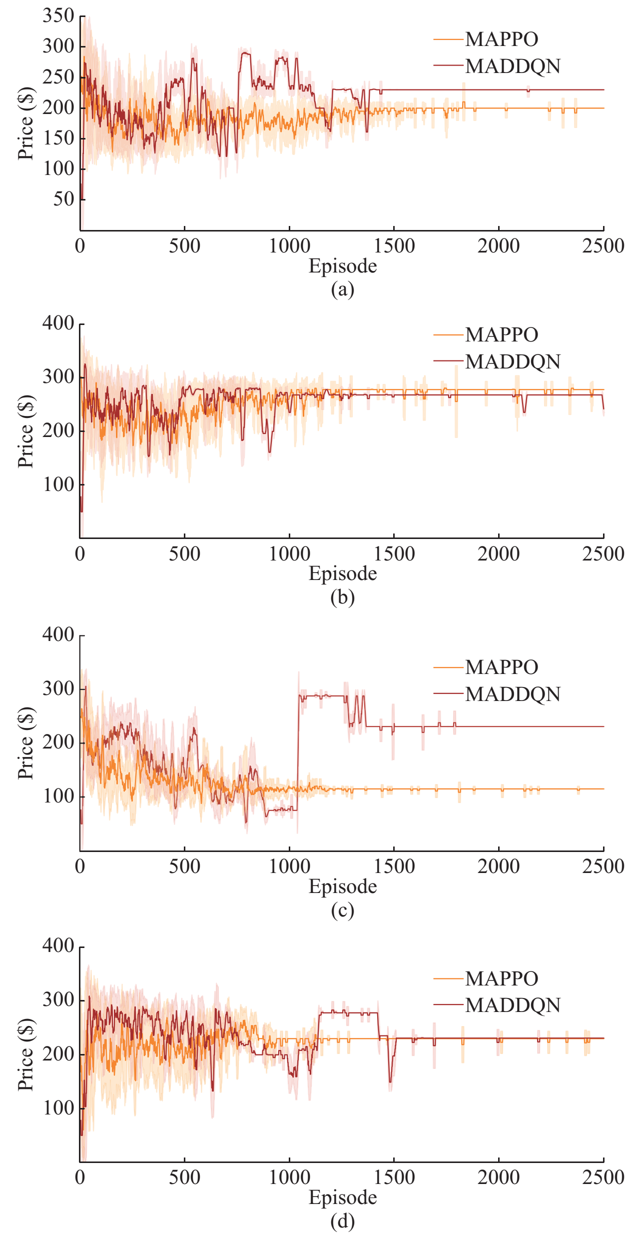


Fig. 6. Market clearing prices during training process at different time under proposed MAPPO and MADDQN frameworks. (a) Prices at 00:00. (b) Prices at 07:30. (c) Prices at 13:00. (d) Prices at 18:30.

The convergence speeds are not distinct. The market clearing prices under the proposed MAPPO framework are less volatile because clip functions are used to avoid violating changes in parameters of agents and stabilize the training process. The convergence of the proposed MAPPO framework is better. The PPO algorithm is more suitable for strategic bidding problems.

### E. Bidding Strategy Analysis

The bidding strategies of agents are investigated. The market clearing prices and loads at different time are shown in Fig. 7.

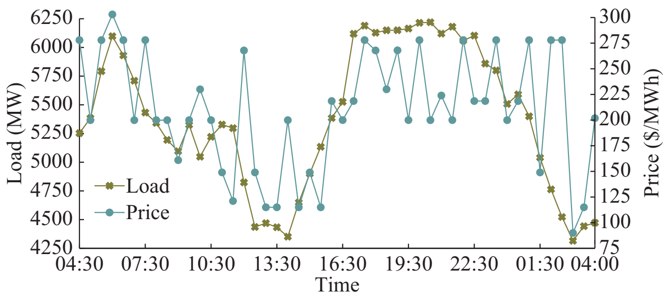


Fig. 7. Market clearing prices and loads at different time.

Strategic bidding makes market clearing prices high during most time periods with high demand, such as 05:00-07:00 and 15:00-21:30. The market clearing prices are low during time periods with low demand such as 02:30-04:00 and 13:00-14:00. The market clearing prices during time periods with high demand are not always high, indicating that the strategy bidding of agents are not always optimal in reality. The bidding strategies of agents at time with high market clearing prices are shown in Fig. 8.

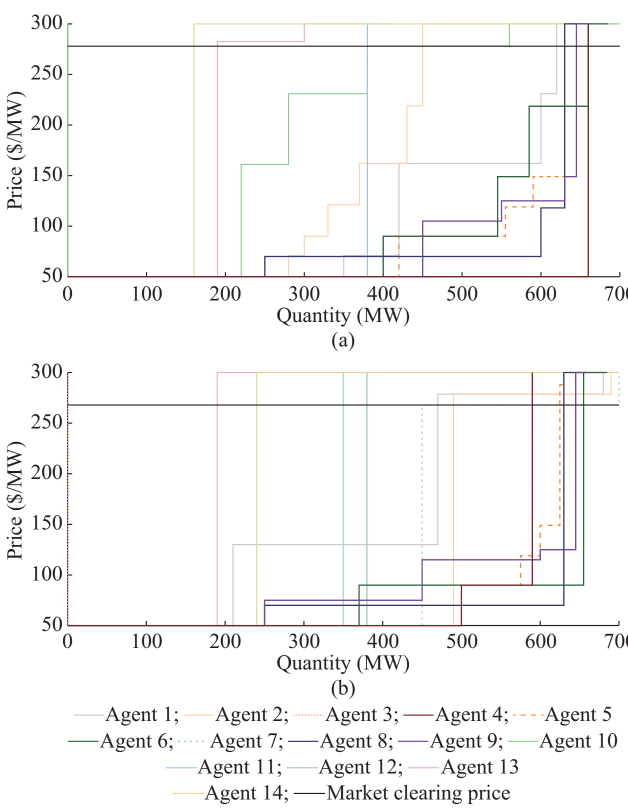


Fig. 8. Bidding strategies of agents at time with high market clearing prices. (a) At 06:30. (b) At 18:00.

Energy quantity with bidding prices below the market clearing price is procured. Agents 4, 5, 6, 8, and 9 make the bidding prices of most energy quantities below the market clearing prices and the bidding prices of a small part of energy quantities high. These strategies can ensure that most energy quantities are cleared, and raise the market clearing prices during time periods with high demand to make excess profits. These five agents are the top five ones with the most cleared energy quantity. As shown in Fig. 5, these five agents make the highest profits among all agents. The marginal agent that determines market clearing prices is agent 7. The procured energy of agent 7 is less than that of the five agents, making its profit lower than that of the five agents. A skillful agent should sell much energy during time periods with high market clearing prices and avoid becoming the marginal agent.

The market clearing prices with pre-aggregated and post-aggregated strategies of small generators are compared to validate the rationality of the data aggregation. Price differences only occur at three time points: 12:00, 16:00, and 17:00, with the corresponding price difference ratios being 13.8%, 26.8%, and 8.5% respectively. As small generators with little market power can hardly determine the market clearing prices, most of them set the energy that they want to be cleared at extremely low prices, and set the energy that they do not want to be cleared at extremely high prices at most time. In other words, most energy quantities concentrate below the lower price bound and above the upper price bound. The sample error caused by data aggregation at these intervals hardly influences the market clearing results. Therefore, the error caused by the data aggregation is decreased.

## V. Conclusion

This paper proposes a data-driven multi-agent framework for electricity market simulation. Numerical results support three findings. First, the representativeness of the strategies clustered by the granular computing-based clustering algorithm is better than other algorithms. The algorithm is more suitable for extracting practical and heterogeneous bidding strategy spaces. Second, market clearing results under the proposed MAPPO framework are closer to reality than those under theoretical and homogeneous strategy spaces. Third, the proposed MAPPO framework outperforms similar frameworks with the same bidding strategy spaces in revealing market clearing results.

The proposed MAPPO framework is not applicable in the markets that provide no individual bidding data. In the markets with few price-makers and huge number of price-takers, enumerating bidding behavior of all agents is costly, and the accuracy of the proposed MAPPO framework decreases.

The proposed MAPPO framework provides a tool for ISOs to analyze the market evolution process. Energy providers can improve their bidding strategies. Future work can focus on more RL approaches with better performances to improve the applicability of this research.

## APPENDIX A

The model of ED program is shown as:

$$\min\left\{\sum_{g\in G}\sum_{n\in N}P_{g,t,n}q^{c}_{g,t,n}+Pr_{p}\cdot\left[\sum_{g\in G}(\Delta q^{D}_{g,t}+\Delta q^{U}_{g,t})+\Delta q^{L}_{t}\right]\right\} \tag{A1}$$

s.t.

$$\sum_{g\in G}\sum_{n\in N}q^{c}_{g,t,n}+\Delta q^{L}_{t}=L_{t}:Pr_{t}\quad \forall t\in T \tag{A2}$$

$$RD_{g}-\Delta q^{D}_{g,t}\leq\sum_{n\in N}q^{c}_{g,t,n}-q^{t-1}_{g}\quad \forall g\in G,\forall t\in T \tag{A3}$$

$$\sum_{n\in N}q^{c}_{g,t,n}-q^{t-1}_{g}\leq RU_{g}+\Delta q^{U}_{g,t}\quad \forall g\in G,\forall t\in T \tag{A4}$$

$$q^{c}_{g,t,n}\leq q_{g,t,n}\quad \forall g\in G,\forall t\in T,\forall n\in N \tag{A5}$$

$$\Delta q^{D}_{g,t},\Delta q^{U}_{g,t},\Delta q^{L}_{t}\geq 0\quad \forall g\in G,\forall t\in T \tag{A6}$$

where $L_t$ is the energy demand; $RD_g$ and $RU_g$ are the ramp up and ramp down rates, respectively; and $q_g^{t-1}$ is the energy output at the last time. The Lagrange multiplier of the system energy balance constraint (A2) is the market clearing price $Pr_t$. The congestion constraint is not included because the detailed topology information is not accessible for generators in most electricity markets.

## REFERENCES


[1] H. Mi, Q. Li, M. Shi *et al*., "A stacking framework for online locational marginal price prediction considering concept drift," *IEEE Transactions on Energy Markets*, *Policy and Regulation*, vol. 2, no. 2, pp. 254-264, Jun. 2024.

[2] H. Mi, S. Chen, J. Ping *et al*., "Traffic equilibrium considering heterogeneity across electric vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 14944-14956, Dec. 2023.

[3] S. Chen, H. Mi, J. Ping *et al*., "A blockchain consensus mechanism that uses proof of solution to optimize energy dispatch and trading," *Nature Energy*, vol. 7, no. 6, pp. 495-502, Jun. 2022.

[4] S. Chen, Z. Shen, L. Zhang *et al*., "A trusted energy trading framework by marrying blockchain and optimization," *Advances in Applied Energy*, vol. 2, p. 100029, May 2021.

[5] M.-C. Hu, S.-Y. Lu, and Y.-H. Chen, "Stochastic-multiobjective market equilibrium analysis of a demand response program in energy market under uncertainty," *Applied Energy*, vol. 182, pp. 500-506, Nov. 2016.

[6] M. Motalleb, P. Siano, and R. Ghorbani, "Networked Stackelberg competition in a demand response market," *Applied Energy*, vol. 239, pp. 680-691, Apr. 2019.

[7] S. Chen, C. Xu, Z. Yan *et al*., "Accommodating strategic players in distributed algorithms for power dispatch problems," *IEEE Transactions on Cybernetics*, vol. 52, no. 11, pp. 12594-12603, Nov. 2022.

[8] E. G. Tsimopoulos and M. C. Georgiadis, "Optimal strategic offerings for a conventional producer in jointly cleared energy and balancing markets under high penetration of wind power production," *Applied Energy*, vol. 244, pp. 16-35, Jun. 2019.

[9] N. Rashedi, M. A. Tajeddini, and H. Kebriaei, "Markov game approach for multi-agent competitive bidding strategies in electricity market," *IET Generation*, *Transmission & Distribution*, vol. 10, no. 15, pp. 3756-3763, Nov. 2016.

[10] D. E. Aliabadi, M. Kaya, and G. Sahin, "Competition, risk and learning in electricity markets: an agent-based simulation study," *Applied Energy*, vol. 195, pp. 1000-1011, Jun. 2017.

[11] Y. Du, F. Li, H. Zandi *et al*., "Approximating Nash equilibrium in day-ahead electricity market bidding with multi-agent deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 3, pp. 534-544, May 2021.

[12] J. Wang, J. Wu, and X. Kong, "Multi-agent simulation for strategic bidding in electricity markets using reinforcement learning," *CSEE Journal of Power and Energy Systems*, vol. 9, pp. 1051-1065, May 2023.

[13] H. Guo, Q. Chen, Q. Xia *et al*., "Modeling strategic behaviors of renewable energy with joint consideration on energy and tradable green certificate markets," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1898-1910, May 2020.

[14] J. Cheng, Z. Yan, X. Xu *et al*., "Bivariate stochastic optimization model for bidding strategies considering competition among renewable power producers," *CSEE Journal of Power and Energy Systems*, vol. 10, pp. 2539-2550, Nov. 2024.

[15] H. Guo, Q. Chen, Y. Gu *et al*., "A data-driven pattern extraction method for analyzing bidding behaviors in power markets," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3509-3521, Jul. 2020.

[16] J. Bustos-Salvagno, "Bidding behavior in the Chilean electricity market," *Energy Economics*, vol. 51, pp. 288-299, Sept. 2015.

[17] Q. Li, Z. Yang, J. Yu *et al*., "Impacts of previous revenues on bidding strategies in electricity market: a quantitative analysis," *Applied Energy*, vol. 345, p. 121304, Sept. 2023.

[18] H. Guo, Q. Chen, Q. Xia *et al*., "Deep inverse reinforcement learning for objective function identification in bidding models," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5684-5696, Nov. 2021.

[19] H. Zhao, J. Zhao, J. Qiu *et al*., "Data-driven risk preference analysis in day-ahead electricity market," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2508-2517, May 2021.

[20] Q. Jia, Y. Li, Z. Yan *et al*., "A reinforcement-learning-based bidding strategy for power suppliers with limited information," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 1032-1039, Jul. 2022.

[21] H. Mi, S. Chen, Q. Li *et al*., "Strategic bidding by predicting locational marginal price with aggregated supply curve," *Energy*, vol. 304, p. 132109, Sept. 2024.

[22] F. Wang, K. Li, L. Zhou *et al*., "Daily pattern prediction based classification modeling approach for day-ahead electricity price forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 105, pp. 529-540, Feb. 2019.

[23] S. Xia, D. Peng, D. Meng *et al*., "Ball *K*-means: fast adaptive clustering with no bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 87-99, Jan. 2022.

[24] D. Cheng, J. Huang, S. Zhang *et al*., "*K*-means clustering with natural density peaks for discovering arbitrary-shaped clusters," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 11077-11090, Aug. 2024.

[25] S. Xia, Y. Liu, X. Ding *et al*., "Granular ball computing classifiers for efficient, scalable and robust learning," *Information Sciences*, vol. 483, pp. 136-152, May 2019.

[26] S. Xia, X. Dai, G. Wang *et al*., "An efficient and adaptive granular-ball generation method in classification problem," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 5319-5331, Apr. 2024.

[27] S. Chen, L. Zhang, Z. Yan *et al*., "A distributed and robust security-constrained economic dispatch algorithm based on blockchain," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 691-700, Jan. 2022.

[28] J. Schulman, F. Wolski, P. Dhariwal *et al*. (2017, Aug.). Proximal policy optimization algorithms. [Online]. Available: https://arxiv. org/pdf/1707.06347

[29] S. Lee, J. Seon, Y. Sun *et al*., "Novel architecture of energy management systems based on deep reinforcement learning in microgrid," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 1646-1658, Mar. 2024.

[30] J. Jin, G. Sun, S. Chen *et al*., "A multi-agent robust deep reinforcement learning approach for coordination of power distribution networks and microgrids with limited information exchange," *Journal of Modern Power Systems and Clean Energy*, doi: 10.35833/MPCE.2025.000035

[31] X. Liu, P. Zhang, H. Xie *et al*., "Graph attention network based deep reinforcement learning for voltage/var control of topologically variable power system," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 1, pp. 215-227, Jan. 2025.

[32] J. Jin, S. Mao, and Y. Xu, "Optimal priority rule-enhanced deep reinforcement learning for charging scheduling in an electric vehicle battery swapping station," *IEEE Transactions on Smart Grid*, vol. 14, no. 6, pp. 4581-4593, Nov. 2023.

[33] C. Guo, X. Wang, Y. Zheng *et al*., "Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning," *Energy*, vol. 238, p. 121873, Jan. 2022.

[34] B. Huang and J. Wang, "Deep-reinforcement-learning-based capacity scheduling for PV-battery storage system," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2272-2283, May 2021.

[35] AEMO. (2025, Aug.). Nemweb data. [Online]. Available: https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/

data-nem/market-data-nemweb
[36] H. Mi, J. Qiao, Y. Li *et al*. (2025, Aug.). Case data. [Online]. Available: https://github.com/miller0331/Case-data-for-MAPPO
[37] AEMO. (2025, Aug.). 2022 forecasting assumptions workbook. [Online]. Available: https://aemo. com. au/consultations/current-and-closed-consultations/2022-consultation-on-forecasting-assumptions-update
[38] Y. Liu, Y. Jiang, T. Hou *et al*., "A new robust fuzzy clustering validity index for imbalanced data sets," *Information Sciences*, vol. 547, pp. 579-591, Feb. 2021.
[39] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double *Q*-learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, pp. 2094-2100, Feb. 2016.

**Hanning Mi** received the B.E., M.E., and Ph.D. degrees in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2017, 2020, and 2024, respectively. He is currently working in China Electric Power Research Institute, Beijing, China. His research interests include artificial intelligence (AI) application in power system analysis and operation.

**Ji Qiao** received the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2018. He is currently working in China Electric Power Research Institute, Beijing, China. His research interests include big data and AI application in power system analysis and operation.

**Ye Li** received the B.S. and M.Eng. degrees from North China Electric Power University, Beijing, China, in 2012 and 2015, respectively. She is currently working in China Electric Power Research Institute, Beijing, China. Her research direction is AI technology for power distribution network.

**Zibo Wang** received the B.S. degree from North China Electric Power University, Baoding, China, in 2014, and the Ph.D. degree from Tianjin University, Tianjin, China, in 2022. He is currently working in China Electric Power Research Institute, Beijing, China. His research interests include peer-to-peer energy market, integrated energy system, and reinforcement learning.

**Fujia Han** received the B.Eng. and M.Eng. degrees in electrical engineering from Nanchang University, Nanchang, China, in 2012 and 2015, respectively. He received the Ph.D. degree from Brunel University, London, U.K., in 2019. He is currently a Research Engineer with China Electric Power Research Institute, Beijing, China. His research interests include big data analytics for smart grid and deep learning application.

**Sijie Chen** received the B.E. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2009 and 2014, respectively. He is a Professor of electrical engineering with Shanghai Jiao Tong University, Shanghai, China. His research interests include electricity market, transactive energy system, blockchain, and machine learning application in power system.