Journal of Thermal Science

https://doi.org/10.1007/s11630-020-1308-0

A Deep Reinforcement Learning Bidding Algorithm on Electricity Market

JIA Shuai[1*], GAN Zhongxue[2*], XI Yugeng[1], LI Dewei[1], XUE Shibei[1], WANG Limin[3]

1. Department of Automation, Key Laboratory of System Control and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China
2. State Key Laboratory of Coal-based Low-carbon Energy, ENN Science and Technology Development Co. Ltd., Langfang 065001, China
3. ENN Energy Power Technology (Shanghai) Co. Ltd., Shanghai 201306, China

Abstract: In this paper, we design a new bidding algorithm by employing a deep reinforcement learning approach. Firms use the proposed algorithm to estimate conjectural variation of the other firms and then employ this variable to generate the optimal bidding strategy so as to pursue maximal profits. With this algorithm, electricity generation firms can improve the accuracy of conjectural variations of competitors by dynamically learning in an electricity market with incomplete information. Electricity market will reach an equilibrium point when electricity firms adopt the proposed bidding algorithm for a repeated game of power trading. The simulation examples illustrate the overall energy efficiency of power network will increase by 9.90% as the market clearing price decreasing when all companies use the algorithm. The simulation examples also show that the power demand elasticity has a positive effect on the convergence of learning process.

Keywords: electricity market, reinforcement learning, energy efficiency, conjectural variation, bidding strategy

1. Introduction

An electricity market refers to competitive market, which enables purchases bidding to buy, sales offering to sell and short-term trades, generally in the form of financial or obligation swaps. Power producers and users trade electricity through negotiation and bidding, and set the price and quantity through market competition using supply and demand principles. In electricity market, each electricity firm will adopt optimal bidding strategy to maximum its profit. Therefore, it is worth studying how power firms should construct bidding strategies.

A class of competitive electricity market models can be divided into price competition models, production competition models and supply function competition models according to different competitive variables [1–4]. Firms use strategic competition in pursuit of maximizing their own profits. It is well known that due to the market mechanism and asymmetric information, firms often ignore the changes of the market price (no response or delay even there is a response), resulting in the inelastic demand of electricity in the electricity market features. Therefore, using the price competition model to analyze the strategic behavior of power generation companies is far from the actual market operation. The competition model based on Supply Function Equilibrium theory (SFE) [5, 6] is more suitable for the operation analysis of inelastic electricity market compared.

Received: Apr 02, 2019

Corresponding author: JIA Shuai;
GAN Zhongxue

E-mail: liu2596615@sjtu.edu.cn
ganzhongxue@fudan.edu.cn

The core issue in SFE model is accurately predicting strategic behavior and interactions of the other firms, but the main difficulty in studying the optimal bidding strategy lies in incomplete information, that is, each firm could not get full information about the bidding strategy and the cost of every electricity firm. Therefore, an optimal bidding strategy for an electricity firm is to estimate the response of each participant to its competitors and market environment changes where the conjectural variation is the main concept [7−9]. Due to the application of the conjectural variation theory [10], market participants can make the best decision for themselves based on their estimates of the response of their competitors to their changes in market strategy, and thus it is suitable for analyzing the power generation firm's bidding strategic behavior with incomplete information. The conjectured supply function (CSF) has been applied to electricity market operations simulation [11, 12]. Assuming the linearity of the supply curve of the firm directly, they applied the CSF near the equilibrium point of the market with the slope or intercept of the linear supply curve fixed and used the firm's output as a competitive strategy causing unsuitable for inelastic electricity demand. Even though rich historical demand and supply data are available, using the data to seek an optimal bidding policy is not an easy task. The major issue is that changes in one-stage bidding game will affect future demand-supply, and it is hard for supervised learning approaches to capture and model these real-time changes [13]. Due to these difficulties, there may be multiple equilibrium solutions in the market using CSF model to predict supply and could not learn accurately enough from historical data.

In recent years, deep reinforcement learning achieves tremendous success in modeling intellectual challenging decision-making problems [14, 15]. In the light of such advances, in this paper, we propose a novel conjectural variation supply function equilibrium model based deep reinforcement learning approach to learn highly efficient bidding strategies so as to seize the maximal profit. There are significant technical challenges when modeling firms' quotation behavior:

(1) Feasibility of problem setting. The Reinforcement Learning framework is reward-driven, meaning that a sequence of actions from the policy is evaluated solely by the reward signal from environment. The definitions of firms, reward and bidding action space are essential. The action space could be prohibitively large since an action needs to decide every firm's situation and next action to make its own strategy. We specifically defined these variables in our framework.

(2) Continuous action spaces. The majority of model-free reinforcement learning algorithms are based on generalized policy iteration: interleaving policy evaluation with policy improvement. In continuous action spaces, greedy policy improvement becomes problematic, requiring a global maximization at every step. Hence, we used the reward value of action to estimate conjectural variation instead of action to solve the problem.

In this paper, we propose an efficient deep reinforcement framework to resolve these challenges and predict the conjectural variation to make quote decision.

This paper is organized as follows. Section 2 provides an overview of electricity market model and participant firm model. Section 3 details the conjectural supply function using conjectural variation approach and proposes a deep reinforcement learning algorithm which can accommodate incomplete information to estimate value of conjectural variation. Section 4 theoretical analyzes the above model and algorithm while Section 5 describes an application of the proposed approach to a standard IEEE 6-generator 30-bus power system and illustrates the feasibility of the algorithm, the impact of the bidding strategy on clearing price and influence on the power network overall energy efficiency. Finally, section 6 provides the conclusions drawn from the study.

2. Model

We focus on real-time trades in electricity market, where bids use supply and demand principles to set the market-clearing price. Electricity market includes market participants and electricity market trading operation. Electricity market trading operation and generation output are executed for a number of intervals in increments of 30 minutes. Market participants are firms that sell electricity in power network, and they make decision on electricity sales and electricity price in single-stage market operation.

2.1 Electricity market operation model

A complete repeated bidding process of electricity market trading is composed of a series of single-stage trading. A single-stage trading consists of bidding stage and clearing stage. Firms make quotation strategy during bidding stage and submit a quote at the end of bidding stage.

In the whole market trading process, firms first predict actions of other firms and make a preliminary strategy. Then they observe the first few stages of market operation data to modify model parameters for more accurate predictions. After that, firms quote price using modified strategy at the end of bidding stage. Market adopts consistent clearing mechanism. When a single-stage game finished, next single-stage game would begin, and firms could quote during a new stage.

Single Stages of Bidding

This trading model can be described in mathematical model based on game theory. A multi-stage trading game with N participants is represented as G=(A,U,S,s), where A=(A_1,…,A_t) is the joint action space; A_t is the one-time quotation action of all the firms at time t, A_t=($a_{t,1}$,$a_{t,2}$,…,$a_{t,n}$); U=(U_1,…,U_t) represents profits of N

participants, U_t=($u_{t,1}$,$u_{t,2}$,…,$u_{t,n}$); S=(S_1,…,S_t) is state space, and S_t is whole game state at time t, S_t=($s_{t,1}$,$s_{t,2}$,…,$s_{t,n}$); $s_{t,i}$ means the state of participant i including its output and quote action at time t. Each participant knows its action, state, and profit, but is unaware of the information of other participants.

2.2 Participant firm model

The electricity firms are the trading market entities that participate in bidding game for selling electricity. Supply function equilibrium (SFE) [16] is used to characterize participant firm model. SFE considers that firms compete through their bidding supply curves and make action A_t. In this model, output by rivals is response to current state S_t at time t. The decision variables for electricity Firm i are parameters v of its bid strategy function $q_i(p, v)$, including the output of other firms $q_{i,j\neq i}$ and clearing price p. This means that Firm i is willing to supply q_i with price p. Then market clearing mechanism determines clearing price p and sets q_i=q_i(p, v).

In the electricity market with N electricity companies competing, the reverse demand curve is

$$p=(e/f)-Q \tag{1}$$

where p is market clearing price; Q is the total power of electricity market; e and f are constant coefficients of reverse demand curve observed by historical data.

Assume that each electricity firm just pursues to maximize its own benefit in time t. For Firm i, the corresponding utility, that is the goal of this electricity firm's goal, is defined as below:

$$U(t)=p(t)q_i(p,t)-Cost_i\left[q_i(p,t)\right] \tag{2}$$

$$D(p,t)=\sum_{i=1}^{N} q_i(p,t) \tag{3}$$

$$Cost_i(q_i)=a_i+b_iq_i+\frac{1}{2}c_iq_i^2 \tag{4}$$

Assuming that supply and demand balance in the end of stage, D(p, t) is load demand at time t, and equals to the total output; a_i, b_i, c_i are coefficients of second power production cost function of Firm i, and q_i is its output.

3. Bidding Strategy Algorithm

The bidding period in electricity market is hour or half-hour due to real-time power balance and periodic load changes, with short cycle of a repeated game under same load demand. Thus, power generation companies need to dynamically evaluate competitors' strategies based on historical information in the market operation in order to accurately predict competitors' strategies to make the best strategies for themselves.

However, the available historical information including total load D(t), market clearing price p(t), and clearing power q_i(t) of each power generation firm at time t, is incomplete. Here we propose a prediction supply function model from supply function model using conjectural variation to describe electricity market trading model. First, a firm predicts output changes of other market participants as market clearing price changes, then calculates its optimal bidding price and output to maximize profit, and updates the predicted result when one bidding period is over. After the multi-stage game, this firm will get the maximal profit after a serious of bidding, and electricity market will reach equilibrium as well.

3.1 Conjectural variation

In this part, we will define how Firm i predicts actions of the other firms $a_{t,j,j\neq i}$ and state $s_{t,j}$. First, we describe the concept of conjectural variation.

(1) Concept of conjectural variation

Conjectural variation is the estimation that market participants' estimates of the changes on its competitors' strategy and expressed as follows:

$$CV_{ij}=\partial q_j/p \tag{5}$$

where p is market price and q_j is product of the j-th participant. Eq. (5) shows that when the i-th market participant's strategy change is ∂p which is quotation price change, the j-th participant's estimating strategy change ∂q_j is the output change. $CV_{ij}=\partial q_j/p$ is the Firm i's conjectural variation of the j-th participant. It is supply conjectural variation function. Thus, the conjectural variation response of Firm i to all its competitors is represented as following formula:

$$CV_i=\frac{\partial q_{\text{others}}}{\partial p}=\sum_{j=1,j\neq i}^{N}\frac{\partial q_j}{p} \tag{6}$$

where $q_{\text{others}}=\sum_{j=1,j\neq i}^{N} q_j(p)$ is the total output of all the competitors.

According to conjectural variation theory, Firm i can predict other firms' action changes with state changes, and then update their own quotation strategy and quote price.

(2) Conjectural equilibrium

Suppose each participant estimates state $\tilde{s}_t[a_i(t)]$, and it will takes the action a_i(t) that maximizes its own profit, where a_i(t) in this game is prediction and quotation; $\tilde{s}_t[a_i(t)]$ is state conjectured after taking action a_i(t), and $\tilde{s}_t[a_i(t)]$ is conjectural state of Firm i, then define conjectural equilibrium as follows.

In a single-stage game G, market reaches conjectural equilibrium if estimated state of any participant i after its action $a_i(t)\subseteq A_t$ satisfies $\tilde{S}_{i,t}[a_i(t)]=S_t[a_1(t),\ldots,a_N(t)]$ when time t is fixed. Estimated state set and action set at this equilibrium point are $[\tilde{s}_1(t),\ldots,\tilde{s}_N(t)]$ and $[a_1(t),\ldots,a_N(t)]$, respectively.

In a repeated game, action taken by participant i at stage t is $a_i(t)$. Denote corresponding estimated state as $\tilde{s}_i(t)$, and the state set of actual participants in time t with joint action is $S_t[a_1(t),\ldots,a_N(t)]$.

When a market trading game reaches conjectural equilibrium, that is, every firm correctly predicts others' actions and states, then it will keep the equilibrium state and go on.

3.2 Deep reinforcement learning algorithm in iterative bidding processes

Assume that the goal of each generation firm is to maximize the profit during the next bidding stage with the same load demand curve, then CV_i updates using the following algorithm.

We use deep reinforcement learning algorithm to initial and update conjectural variation. First, it takes output changes of all the competitors of Firm I according to clearing price changes in stage t as basic conjectural in stage t+1, then updates using CV_i of first few stages with the same load demand curve. We define the future discounted return at time t as Eq. (9), and $a_i(t)$=CV_i(t) which means action $a_i(t)$ at time t is predicting the others output changes with clearing price changes. After estimating, each firm quotes according to its conjectural result, and stage t is over. Reinforcement learning is learning what to do, and how to map situations to actions so as to maximize a numerical reward signal. In fact, the best competitive power firm can do is learning for the aggregate action it faces. Then we demonstrate that a convolutional neural network can learn successful bidding strategies from raw data in complex game. The network is trained with a variant of the Q-learning algorithm, with stochastic gradient descent to update the weights. The optimal action-value function defined as the maximum expected return achievable by following any strategy, after seeing some sequence S and then taking some action a, means one-time predict:

$$Q^*(S,a)=\max_\pi E\left[R_t|S_t=S,a_t=a,\pi\right] \qquad (7)$$

where π is a policy distributions over actions. It obeys identity Bellman equation.

$$Q^*(S,a)=\mathbb{E}_{S'\sim\varepsilon}[r+\gamma max_{a'}Q^*(S',a')\,|\,S,a] \qquad (8)$$

if the optimal value $Q^*(S', a')$ of the sequence S' at the next stage was known for all possible actions values a', then the optimal strategy is to select the action a' maximizing the expected value of r+$\gamma\max_{a'}Q^*(S',a')$. And future discounted return at time t is

$$R_t=U_t+\gamma U_{t+1}+\gamma^2U_{t+2}\ldots+\gamma^{n-t}U_n=U_t+\gamma R_{t+1} \qquad (9)$$

where R_t is the future discounted return at time t; γ is discounted factor, and U_t is profit in each bidding stage.

We refer to a neural network function approximator with weights θ as a Q-network to solve the problem that the state space is too large which is also named curse of dimensionality. We use learning process to train the Q-network, along with extracting features as input and computing the value function as output. A Q-network can be trained by minimizing a sequence of loss functions $L_i(\theta_i)$ that changes at each iteration i, and the error loss is:

$$L_i(\theta_i)=\frac{1}{2}\left[r+\max_{a'}Q(S',a';\theta_{i-1})-Q(S,a;\theta_i)\right]^2 \qquad (10)$$

The parameters from the previous iteration q_{i-1} are held fixed when optimizing the loss function $L_i(\theta_i)$. In Q-learning, we use reward of each iteration and current q value to update Q-table. Then we also use the q value computed as a tag of learning to design loss function, which is the mean square error of approximation value and the real value. Note that the targets depend on the network weights; this is in contrast with the targets used for supervised learning, which are fixed before learning begins. Differentiating the loss function with respect to the weights, we arrive at the following gradient:

$$\nabla_{\theta_i}L_i(\theta_i)=\mathbb{E}_{S'\sim\varepsilon}\left[r+\gamma\max_{a'}Q(S',a';\theta_{i-1})\right.$$
$$\left.-Q(S,a;\theta_i)\nabla_{\theta_i}Q(S,a;\theta_i)\right] \qquad (11)$$

Then the conjectural variation learning algorithm is shown in Algorithm 1:

Algorithm 1 Deep reinforcement learning conjecture

```
Initialize replay memory $D$ to capacity $N$
Initialize action-value function $Q$ with random weights
  for episode = 1, $M$ do
    Initialize sequence $S_1=\{x_1\}$ and preprocessed sequenced $\phi_1=(S_1)$
    for $t=1,T$ do
        With probability $\varepsilon$ select a random action $a_t$ otherwise select $a_t=\max_a Q^*[\phi(S_t), a; \theta]$
        Execute action $a_t$ in emulator and observe reward $r_t$ and image $x_{t+1}$
        Set $S_{t+1}=S_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1}=\phi(S_{t+1})$
        Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in $D$
        Sample random minibatch of transitions $(\phi_t, a_t, r_t, \phi_{t+1})$ from $D$
        $y(t)=\begin{cases} r_t, if\ \phi_{t+1}-\phi_t<\varepsilon \\ r_t+\gamma\max_{a'}Q(\phi_{t+1},a';\theta), else \end{cases}$   $\varepsilon$ approaching 0
        Perform a gradient descent step on $[y_j-Q(\phi_j, a_j; \theta)]^2$ according to Eq. (11)
    end for
  end for
```

Each step of experience is potentially used in many weights updates, which allows for greater data efficiency. And learning directly from consecutive samples is inefficient, due to the strong correlations between the samples; randomizing the samples breaks these correlations and therefore reduces the variance of the updates. By using experience replay, the behavior distribution is averaged over many of its previous states, smoothing out learning and avoiding oscillations or divergence in the parameters. Note that when learning by experience replay, it is necessary to learn off-policy (because our current parameters are different to those used to generate the sample), which motivates the choice of Q-learning.

We now describe the exact architecture used for the conjecture progress. The input to the neural network consists of a (n+1)×3×4 vector produced by ϕ, where n is the number of firms. The first hidden layer convolves some filters with stride 4 with the input vector and applies a rectifier nonlinearity. The second hidden layer convolves double filters with stride 2, again followed by a rectifier nonlinearity. The final hidden layer is fully-connected and consists of rectifier units. The output layer is a fully-connected linear layer with a single output for valid action. We refer to convolutional networks trained with our approach as Deep Q-Networks (DQN).

3.3 Optimal quotation strategy

After calculating conjectural variation CV_i, we need to use it to get the optimal quotation strategy. In each bidding stage, each electricity firm just pursues to maximize its own benefit in stage t. For Firm i, the corresponding optimization problem is shown as below:

$$\max_{q_i} U(t) = p(t) q_i(p,t) - Cost_i\left[q_i(p,t)\right]$$

$$\text{s.t.} \begin{cases} D(p,t) = \sum_{i=1}^{N} q_i(p,t) \\ Cost_i(q_i) = a_i + b_i q_i + \frac{1}{2} c_i q_i^2 \\ q_{i,\min} \le q_i \le q_{i,\max} \end{cases} \tag{12}$$

where $D(p, t)$ is load demand in the stage t, and equals to the total output; a_i, b_i, c_i are coefficients of the i-th firm's second power production cost function, and q_i is its output. In the meantime, the output of Firm i must satisfy power generation constraints [$q_{i,\min}$,$q_{i,\max}$].

For p=(e/f)–D and $D(p,t)$=$q_i(p,t)$+$q_{\text{other}}(p,t)$, Eq. (12) can be rewritten as:

$$\max_{q_i} U(t) = -\left(\frac{1}{2} c_i + 1\right) q_i^2(p,t) + \left(\frac{e}{f} - b_i - q_{\text{others}}(p,t)\right) q_i(p,t) - a_i$$

So, it is a convex optimization problem for q_i. If the solution is out of range, q_i must locate at the boundary.

We can use the first-order optimization condition to solve the optimization problem in supply function equilibrium:

$$\frac{\partial U_i(t)}{\partial p(t)} = 0 \tag{13}$$

Then we can deduce the following formula:

$$q_i(p,t) + \left[p(t) - b - c q_i(p,t)\right] \frac{\partial q_i\left[p(t)\right]}{\partial p(t)} = 0 \tag{14}$$

Suppose supply and demand will eventually balance at the end of stage t, which means $D(p,t) = \sum_{i=1}^{N} q_i(p,t)$, then we can get:

$$q_i(p,t) + \left[p(t) - b - c q_i(p,t)\right] \times \left(\frac{\partial D\left[p(t)\right]}{\partial p(t)} - \frac{\partial \left[\sum_{j=1, j\neq i}^{N} q_j(p,t)\right]}{\partial p(t)} \right) = 0 \tag{15}$$

Let q_{others}=Q–q_i be the output of the other competitors. So, the general conjectural variation of the i-th market participant to the others is:

$$CV_i(p,t) = \frac{\partial q_{\text{others}}(t)}{\partial p(t)} \tag{16}$$

With Eqs. (13–16), we can get the optimal output-price quotation function as:

$$q_i(p,t) = \alpha_i(p,t) + \beta_i(p,t) p(t)$$

$$\begin{cases} \alpha_i(p,t) = \dfrac{-b_i\left[CV_i(p,t) - D_p(p,t)\right]}{1 + c_i\left[CV_i(p,t) - D_p(p,t)\right]} \\ \beta_i(p,t) = \dfrac{CV_i(p,t) - D_p(p,t)}{1 + c_i\left[CV_i(p,t) - D_p(p,t)\right]} \end{cases} \tag{17}$$

Further, it adopts a unified clearing price mechanism, and each firm makes a quotation according to Eq. (17) to pursue the maximum profit with market load demand function D=e–fp, where e and f are fixed coefficients. Then market will reach following equilibrium:

$$\begin{cases} p(t) = \dfrac{e - \sum_{i=1}^{N} \dfrac{-b_i\left[CV_i(t) + f\right]}{1 + c_i\left[CV_i(t) + f\right]}}{\sum_{i=1}^{N} \dfrac{-CV_i(t) + f}{1 + c_i\left[CV_i(t) + f\right]} + f} \\ q_i(t) = \alpha_i(t) + \beta_i(t) p(t) \end{cases} \tag{18}$$

As we can see, Firm i only needs to estimate the total output of all the other competitors and the total response of them to Firm i's output changes. Given the linear power demand curve as D=e–fp which can be got from Eq. (1) with Q=D, the generation Firm i can estimate the conjectural variation response of competitors to the market clearing price based on past information. Then we can get the quotation price and output of the i-th firm.

3.4 Estimation of end of stage

The power Firm i estimates $CV_i(t+1)$ of the other firms according to the conjectural variation CV_i at the beginning of stage $t+1$, historical running data, and information of stage t, including total load, its clearing output $q_i(t)$ and market clearing price $p(t)$. $CV_i(t+1)$ satisfies

$$\frac{\tilde{q}(t+1)-q_{\text{others}}(t)}{\tilde{p}(t+1)-p(t)}=CV_i(t+1) \tag{19}$$

Hence the strategy quotation function power Firm i estimated is:

$$\begin{aligned}\tilde{q}_{\text{others}}(t+1)=\big[&q_{\text{others}}(t)-p(t)CV_i(t+1)\\ &+CV_i(t+1)\tilde{p}(t+1)\big]\end{aligned} \tag{20}$$

Combined Eq. (19) with Eq. (20), the clearing price and stage conjectural variation estimated by Firm i are:

$$\tilde{p}(t+1)=\frac{e-\alpha_i(t+1)-\left[q_{\text{others}}(t)-CV_i(t+1)p(t)\right]}{f+CV_i(t+1)+\beta_i(t+1)} \tag{21}$$

$$\tilde{q}_i(t+1)=\alpha_i(t+1)+\beta_i(t+1)\eta_i(t+1) \tag{22}$$

$$\eta_i(t+1)=\frac{e-\alpha_i(t+1)-\left[q_{\text{others}}(t)-CV_i(t+1)p(t)\right]}{f+CV_i(t+1)+\beta_i(t+1)} \tag{23}$$

And conjectural profit is:

$$\tilde{U}_i(t+1)=\tilde{p}(t+1)\tilde{q}_i(t+1)-Cost_i\left[\tilde{q}_i(t+1)\right] \tag{24}$$

Then the whole dynamic simulation of electricity market under the reinforcement learning method is shown in Algorithm 2.

Algorithm 2 Market operation algorithm

1. Initialize the parameters (e, f) in the demand function (1) obtained by each firm based on the published historical data of the market and initialize the initial $CV_i(t_0)$;
2. Each firm dynamically learns and adjusts the conjectural variation $CV_i(t+1)$ of all competitors to market electricity price changes according to the proposed reinforcement learning Conjecture Algorithm 1;
3. According result of step 2, each firm estimates $CV_i(t+1)$ and cost factor (a_i, b_i, c_i) to determine an optimized supply function;
4. Each firm estimates market equilibrium result using Eqs. (21−22);
5. Clear the market according to the supply function submitted by all firms, and release market clearing price $p(t)$, the total load $Q(t)$, and output of each firm;
6. If the difference between the market clear result and market equilibrium estimated by each firm is less than ε, then the learning process reaches convergence; otherwise, go to step 2.

We also evaluate efficiency of the power trading market by following formula:

$$E_{\text{ptm}}=\frac{p'Q'-pQ}{pQ} \tag{25}$$

where E_{ptm} denotes efficiency; p' and Q' are the clearing price and power generation of the new equilibrium point reached by the proposed algorithm, respectively, while p and Q are the clearing price and power generation in other case.

4. Theoretical Analysis

In this section, we will focus on the concept of conjectural equilibrium proposed in the previous chapter and the bidding strategy with reinforcement learning of the iterative game process, and prove that there is a conjectural equilibrium in above electricity market, and the iterative bidding process will eventually converge.

For a more concise description, the optimization problem can be rewritten as:

$$\begin{aligned}&\max\nolimits_{q_i} U_i\left[x_i(t)\right]\\ &\text{s.t. } \tilde{P}_i(t)\cdot x_i(t)\le \tilde{P}_i(t)\cdot e_i(t)\end{aligned} \tag{26}$$

where $x_i(t)$ denotes the demand of the i-th participant at time t, and $\tilde{P}_i(t)$ represents its conjectured price at time t. For bidding process, participants submit quotation, observe the clearing price, and predict others conjectured prices to adjust their bidding strategies accordingly.

We adopt a model that effect of participant on price is linear:

$$\tilde{P}_i(t)=\alpha_i+\beta_i(t)q_i(t) \tag{27}$$

The parameters are configured according to difference between the actual price and conjectural price at each time in $[t–k, t]$. To simplify the model, we just consider the difference between current and last time, then the coefficients are as below:

$$\alpha_i(t+1)=\alpha_i(t)+\eta_1\left[P_i(t)-\tilde{P}_i(t)\right] \tag{28}$$

$$\beta_i(t+1)=\beta_i(t)+\frac{\eta_2}{q_i(t)}\left[P_i(t)-\tilde{P}_i(t)\right] \tag{29}$$

where η_1 and η_2 are positive constant parameters. After substituting Eqs. (28) and (29) into Eq. (26), we obtain the optimization problem of each participant represented as below:

$$\begin{aligned}&\max\nolimits_{q_i} U_i\left[q_i(t)+e_i\right]\\ &\text{s.t.}\left[\alpha_i+\beta_i q_i(t)\right]\cdot z_i\le 0\end{aligned} \tag{30}$$

Proposition 1. Let q_i and the set of demand are all nonnegative and α_i, β_i are constant at fixed time t, and let U be a continuous function on X. Then the optimization problem of participant has a solution.

Proof.

According to Weierstrass Maximum Theorem [17]: if X is a nonempty compact set in R^m, and $f(x)$ is a continuous function on X, then $f(x)$ has at least one global optimum point in X.

The function U is continuous on X, then it is also continuous on $Q=\{q\,|\,q+e\in X\}$. Let X be $X=Q\cap$ $\{q\,|\,(\alpha+\beta q)\cdot z\le 0\}$. All we need to prove is that X is a nonempty compact set in R^m.

$\sum_i\left(\alpha_i q_i+\beta_i q_i^2\right)\le 0$, which implies

$$\sum_i \beta_i\left(z_i+\frac{\alpha_i}{2\beta_i}\right)^2 \le K \text{, where } K=\sum \beta_i \frac{\alpha_i^2}{4\beta_i^2}.$$

Let $\hat{\beta}=\min(\beta_1,\ldots,\beta_m)$, we can get

$$\sum_i \left(q_i+\frac{\alpha_i}{2\beta_i}\right)^2 \le \frac{K}{\hat{\beta}},$$

$$\left|q_i+\frac{\alpha_i}{2\beta_i}\right| \le \left(\frac{K}{\hat{\beta}}\right)^{\frac{1}{2}},$$

$$|z_i| \le \left(\frac{K}{\hat{\beta}}\right)^{\frac{1}{2}}+\left|\frac{\alpha_i}{2\beta_i}\right|,$$

Hence X is bounded, and obviously closed. Therefore, X is compact.

Applying the Weierstrass theorem, U has at least one global maximum in X.

The set of clearing output calculated by each firm constructs conjectural stage set $\tilde{s}_{i,t+1}=\{\tilde{q}_{1,t+1},\ldots,\tilde{q}_{N,t+1}\}$. If conjectural variation of each power generation on previous stage is consistent with the actual market settlement status, then the market reaches conjectural equilibrium point.

Each generation firm dynamically learns and adjusts the Conjectural variation $CV_i(t+1)$ of all competitors to market electricity price change according to the published reinforcement learning rules based on the past market operation data of the previous stage. If the conjectural stage of Firm i $\tilde{s}_{i,t+1}$ is consistent with actual market clearing stage $s(t+1)=\{q_{1,t+1},\ldots,q_{N,t+1}\}$, then market reaches conjectural equilibrium point.

Suppose the market adopts a unified clearing price settlement mechanism, and each participant quotes price using conjectural variation method to maximize its own profit, and market load demand function is $D=e-fp$. Then market will reach the following equilibrium:

$$\begin{cases} p(t)=\dfrac{e-\sum_{i=1}^{N}\dfrac{-b_i\left[CV_i(t)+f\right]}{1+c_i\left[CV_i(t)+f\right]}}{\sum_{i=1}^{N}\dfrac{-CV_i(t)+f}{1+c_i\left[CV_i(t)+f\right]}+f} \\ q_i(t)=\alpha_i(t)+\beta_i(t)p(t) \end{cases} \tag{31}$$

Remark 1. The strategic supply function for power generation Firm i will directly depend on the power demand curve factor (f), its own cost factor (a_i, b_i, c_i) , and conjectural variation response to market clearing price of all its competitors CV_i.

Remark 2. If conjectural variation response of power Firm i to its competitors is a constant, its optimized supply curve is linear. If the conjectural variation response varies with the market clearing price p, the supply curve is optimized to be non-linear.

5. Application to IEEE 6-generator 30-bus Power System

In this paper, the proposed algorithm was tested on the standard IEEE 6-generator 30-bus power system, and the topological structure is shown in Fig. 1. The system

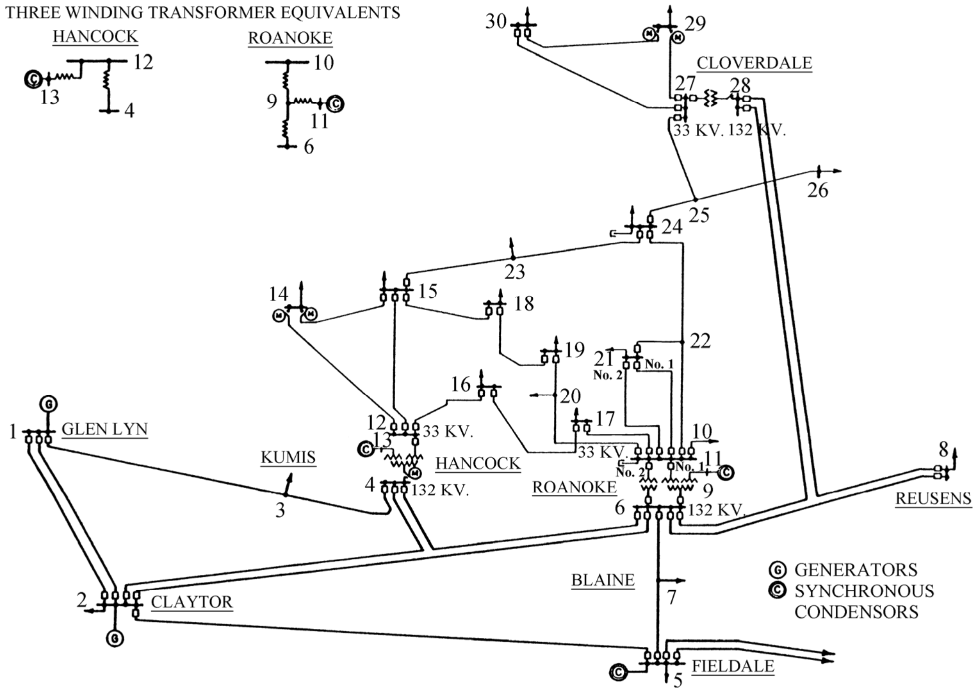

Fig. 1 30-bus power system topological structure

Table 1 Power Generation limits and quadratic cost coefficients for IEEE 30-bus system

Firm	$Pg_{\min}$/MW	$Pg_{\max}$/MW	a/USD·h^{-1}	b/USD·(MW·h)$^{-1}$	c/10^{-4} USD·(MW^2·h)$^{-1}$
1	50	200	0	2.00	200.0
2	20	80	0	1.75	175.0
3	10	30	0	3.00	250.0
4	12	40	0	3.00	250.0
5	15	50	0	1.00	625.0
6	10	35	0	2.25	83.4

parameters are shown in Table 1, and the market reverse demand curve is p=35–0.018D(USD/MWh), where a, b, c are the parameters in Eq. (4). The six generators belong to six firms. We conduct different analysis based on the elasticity and inelasticity of demand, and emphasize the applicability of the algorithm proposed in the real electricity market.

5.1 Case 1: only Firm 1 using reinforcement learning

Fig. 2 (i.e., only Firm 1 dynamically improves its conjectural variation response) shows the profit of Firm 1 in each stage of the process of dynamic learning and improvement of CV.

As can be seen from the figure, Firm 1 constantly improves its understanding of the changes in strategy of all its competitors based on the useful information released by the market, to obtain a more accurate estimation of the competitors' response to market price changes, and thus get the maximal profit. Hence, Firm 1 has the motivation to dynamically learn and adjust the conjectural response.

5.2 Case 2: Different numbers of firms dynamically learn

Case 2 is that all firms continue to improve their conjectural variation response to market clearing price and other competitors' strategy changes.

Fig. 3 gives a comparison of case 1 where only Firm 1 has performed learning iterations and Case 2 where all firms have performed learning iterations. It shows that the market clearing price will reduce when all power generation companies learn. At the same time, compared with Case 1, with number of firms which dynamic learn and adjust conjectural response increasing, that is, when the number changes from 1 to 6, the market clearing price of the market equilibrium has significantly reduced by 7.61%. In Case 2, each firm uses Eq. (18) to estimate clearing price and product of others so as to maximum its profit. The more firms learn, the more accurate estimations of firms have. Hence, the conjectural variation of each firm will be more accurately estimated which will be less than the case that one unique firm learns in the game of power trading market, so it obviously can be seen that the clearing price is also reduced with this procedure. From the perspective of the benefits of the entire network, when the market information is selectively open and electricity firms are encouraged to learn dynamically based on this publicly available information, the equilibrium price of market has been greatly reduced; therefore, the efficiency of the entire energy network is improved by 9.90% according to Eq. (25).

Fig. 4 takes Firm 1 as an example with the dynamic learning process that shows conjectural generation and actual clearing generation at each stage. As can be seen from the figure, the proposed reinforcement-learning algorithm can accurately estimate the clearing price.

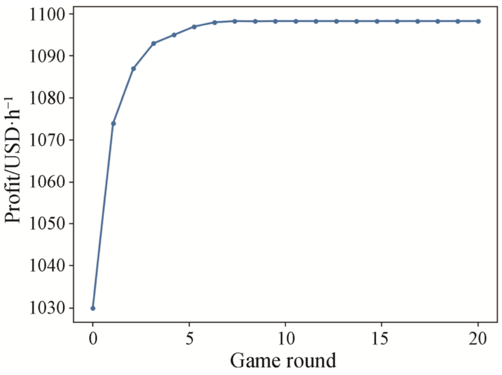

Fig. 2 Profits of Firm 1 provided that other companies hold constant initially conjectured CVs

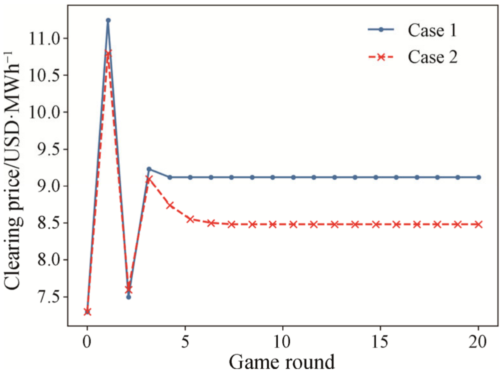

Fig. 3 Market Clearing Prices in Cases 1&2

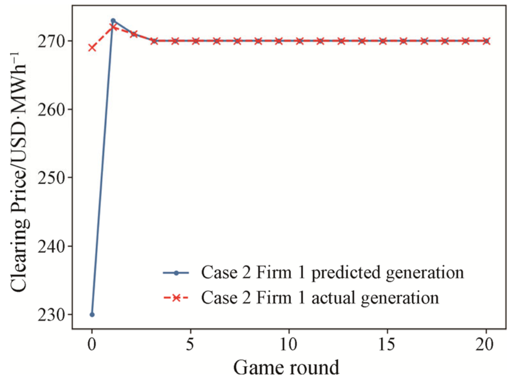

Fig. 4 Conjectural output of company 1 and its actual generation in Cases 2 without electricity demand elasticity

Each firm adopts the reinforcement-learning algorithm, and dynamically estimates the supply function in the first few stages under the same load and adjusts the conjectural response at this stage to determine the optimized supply function. Fig. 5 shows the results of the proposed algorithm comparing with q-learning approach for electricity markets employed by Ashkan Rahimi-Kian [18]. From the results, we can see that the proposed algorithm outperforms the q-learning based supplier-agents algorithm. Moreover, it achieves higher-efficiency than the state-of-the-art techniques.

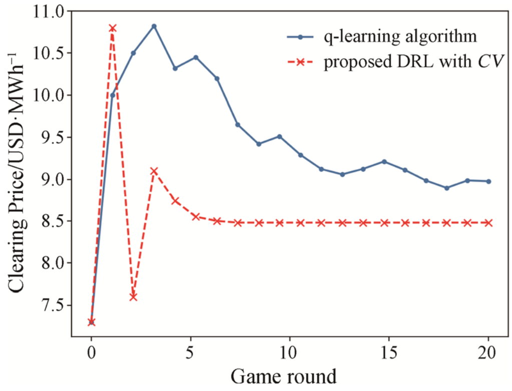

Fig. 5 Results compared with q-learning algorithm

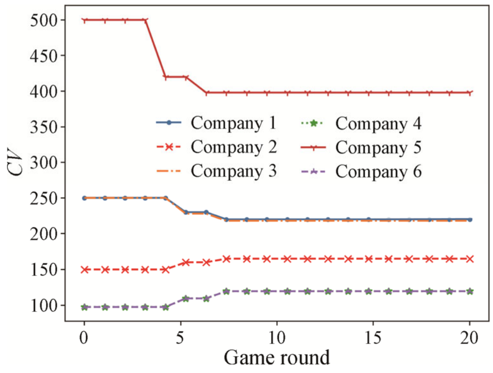

Fig. 6 *CV*s of individual firms using reinforcement learning

When each electricity firm adopts the proposed reinforcement learning algorithm to dynamically learn and adjust the conjectural variation response function, the estimated *CV* value of the past estimate game round plays a certain role in the *CV* prediction of the current stage. Thus, the change in *CV* of firms is relatively gentle and tends to a value close to the initial value, as shown in Fig. 6.

5.3 Case 3: electricity demand is elasticity

As we all know, the electricity demand elasticity influences a lot the operation of the electricity market and the electricity firms in the market. And the lower elasticity, the relatively larger influence. However, the lack of elasticity in electricity demand has always been an important issue in the electricity market, and it is also an important reform link in the domestic electricity market to implement a standardized market mechanism.

It can be seen from the example results in Fig. 7 that when using the dynamic learning mechanism of conjectural equalization function to simulate the operation of the electricity market, the change of power elasticity has an impact on the market equilibrium, but the market clearing price has risen due to the inelastic electricity demand.

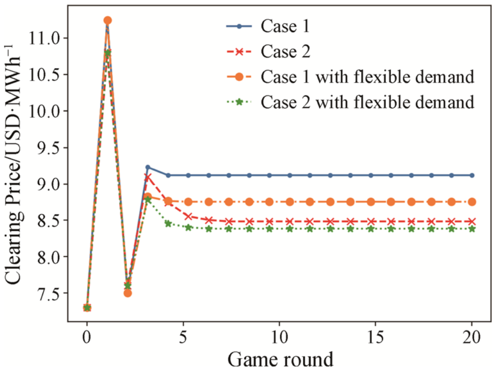

Fig. 7 Market clearing prices in Cases 1 and 2 with or without electricity demand elasticity

Fig. 7 is the market clearing price change chart of Case 1 and Case 2 with or without electricity demand elasticity. It can be seen that when the demand is elastic, the market clearing price is lower than the inelasticity case where decreasing by 5.43% with flexible demand in case 1 and 1.18% in case 2, but the difference is relatively small. Moreover, when the number of companies which dynamically learn and adjust conjectural response increasing, the market clearing price of the market equilibrium has significantly decreased by 7.61% with fixed demand and 3.45% with flexible demand, while total power network power efficiency increases by 9.90% with fixed demand and 2.40% with flexible demand accordingly. Influence of the number of firm on market price is also smooth.

6. Conclusions

This paper mainly proposed a novel deep reinforcement learning approach algorithm based on conjectural variation supply function equilibrium model on repeated game of power trading market. The algorithm uses reinforcement learning to update conjectural variation of a firm to the others to improve the firm's quotation strategy and gain more profit.

The proposed algorithm used by power companies is the estimation of their competitors' bidding strategies. Each electricity firm dynamically adjusts conjectural variation employing the proposed deep reinforcement learning algorithm according to the historical market operation data, so that it can accurately estimate output changes of all competitors with market clearing price change and adjust the supply function to obtain the maximal profit.

Through the system operation simulation and analysis of the results of IEEE standard 6-generators and 30-buses power system, each power firm has the motivation to dynamic learn. The results show that it is much more rewarding to analyze the operation of the electricity market under incomplete information market by using DRL and construct the strategic behavior of a power firm. After dynamic learning and optimizing of the supply function by various power firms, market will eventually reach a conjectural equilibrium state. To a great extent, it also objectively reflects the actual operation situation of power generation firms and forecasts supply curve. With the number of firms, which dynamically learn and modify their conjectural response increasing, the market clearing price of the market equilibrium has significantly decreased, and total power efficiency increases accordingly.

References

[1] Grossman S.J., Nash equilibrium and the industrial organization of markets with large fixed costs. Econometrica: Journal of the Econometric Society, 1981, 49(5): 1149–1172.

[2] Conejo A.J., Nogales F.J., Arroyo J.M., Price-taker bidding strategy under price uncertainty. IEEE Transactions on Power Systems, 2002, 17(4): 1081–1088.

[3] Dai T., Qiao W., Trading wind power in a competitive electricity market using stochastic programing and game theory. IEEE Transactions on Sustainable Energy, 2013, 4(3): 805–815.

[4] Su W., Huang A.Q., A game theoretic framework for a next-generation retail electricity market with high penetration of distributed residential electricity suppliers. Applied Energy, 2014, 119: 341–350.

[5] Klemperer P.D., Meyer M.A., Supply function equilibria in oligopoly under uncertainty. Econometrica: Journal of the Econometric Society, 1989, 57(6): 1243–1277.

[6] Starr R.M., General equilibrium theory: An introduction. Cambridge University Press, 2011.

[7] Díaz C.A., Villar J., Campos F.A., et al. Electricity market equilibrium based on conjectural variations. Electric Power Systems Research, 2010, 80(12): 1572–1579.

[8] Delgadillo A., Reneses J., Conjectural-variation-based equilibrium model of a single-price electricity market with a counter-trading mechanism. IEEE Transactions on Power Systems, 2013, 28(4): 4181–4191.

[9] Lagarto J., Sousa de J., Martins A., et al., Price forecasting in the day-ahead Iberian electricity market using a conjectural variations ARIMA model, 9th International Conference on the European Energy Market, 2012, pp: 1–7.

[10] Figuières C., Theory of conjectural variations. World Scientific, 2004.

[11] Day C.J., Hobbs B.F., Pang J.S., Oligopolistic competition in power networks: a conjectured supply function approach. IEEE Transactions on power systems, 2002, 17(3): 597–607.

[12] García-Alcalde A., Ventosa M., Rivier M., et al., Fitting electricity market models: A conjectural variations approach. 14th Proceedings Systems Computation Conference (PSCC), 2002, 1:403–408.

[13] Rick Chang J.H., Li C.L., Poczos B., et al., One Network to Solve Them All-Solving Linear Inverse Problems Using Deep Projection Models. Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, 2017, pp: 5888–5897.

[14] Silver D., Lever G., Heess N., et al., Deterministic policy gradient algorithms. 2014.

[15] Frosst N., Hinton G., Distilling a neural network into a soft decision tree. arXiv preprint arXiv, 2017, pp: 1711.09784.

[16] Holmberg P., Philpott A.B., On supply-function equilibria in radial transmission networks. European Journal of Operational Research, 2018, 271(3): 985–1000.

[17] Sury B., Weierstrass's theorem—Leaving no 'Stone' unturned. Resonance, 2011, 16(4): 341.

[18] Rahimi-Kian A., Sadeghi B., Thomas R J., Q-learning based supplier-agents for electricity markets. IEEE Power Engineering Society General Meeting, 2005, 1: 420–427.