Paper

A Q-Learning-Based Supplier Bidding Strategy in Electricity Auction Market

Gaofeng Xiong* Student Member
Tomonori Hashiyama** Member
Shigeru Okuma* Member

One of the most important issues for power suppliers in the deregulated electric industry is how to bid into the electricity auction market to satisfy their profit-maximizing goals. Based on the Q-Learning algorithm, this paper presents a novel supplier bidding strategy to maximize supplier's profit in the long run. In this approach, the supplier bidding strategy is viewed as one kind of stochastic optimal control problem and each supplier can learn from experience. A competitive day-ahead electricity auction market with hourly bids is assumed here, where no supplier possesses the market power and all suppliers winning the market are paid based on their own bid prices. The dynamics and the incomplete information of the market are considered. The impact of suppliers' strategic bidding on the market price is analyzed. Agent-based simulations are presented. The simulation results show the feasibility of the proposed bidding strategy.

Keywords: Deregulation, day-ahead electricity auction market, Q-Learning algorithm, supplier bidding strategy

1. Introduction

In the past decade, the electric utility industry in many countries around the world has been undergoing fundamental structural changes to introduce competition and enhance efficiency. The traditional vertically integrated utility is deregulated to open up the system to the market, in response to the pressures of privatization and customer demands. Electricity and services can be sold and purchased as a commodity through different market structures. Under this deregulated and competitive environment, economics and profitability have become the major concern of every market participant, and each of them will act in his/her own self-interest in this new environment.

Among the proposed market structures, the electricity auction market has been widely experienced and implemented in different countries with different protocols. Market participants—electricity suppliers, and distribution companies—are required to submit their sealed bids to the auction market to compete for power energy. All participants winning the auction will be paid based on the rules agreed upon by the participants. Thus, the bidding strategy, which is essential for a successful business in this auction market, is becoming one of the most important issues in deregulated electric industry. Market participants can greatly improve their benefits by strategic bidding.

* Okuma Lab., Dept. of Information Electronics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603
** Hashiyama Lab., Institute of Natural Science, Nagoya City University
Mizuho-cho, Mizuho-ku, Nagoya 467-8501

Developing bidding strategies for competitive suppliers has been studied by many researchers in recent years. Game theory [1] is naturally the first choice to deal with this issue and lots of works have been done using this traditional theory. In Ref. (2), a Nash game approach is used to study the pricing strategy in the deregulated power marketplace, where each participant has incomplete information about others. A method using Cournot non-cooperative game theory to determine the optimal supply quantity for each power producer in an oligopoly electricity market is presented in Ref. (3). The results show that the estimation accuracy of production cost functions of rivals plays an important role in this market. Different electricity market rules and their effects on bidding behaviors in a non-congestion grid are analyzed in Ref. (4). The authors conclude that generators can take advantage of congestion in their strategic bidding behavior.

But game theory is not the only solution to this problem. In fact, due to the complexity, dynamics and uncertainty of the restructured electricity market, evolutionary computation algorithms and reinforcement learning are receiving increasing attention recently and becoming major tools in solving this problem. A genetic algorithm is developed in Ref. (5) to evolve the bidding strategies of participants in a double auction market. Markov Decision Process is used to optimize the bidding decisions to maximize the expected reward over a planning horizon in Ref. (6). The optimal bidding problem is modeled as a stochastic optimization problem in Ref. (7), and, a Monte Carlo approach based method and an optimization based method are developed to solve this problem. An agent-based simulation model of a wholesale electric market is developed in Ref. (13) to provide a source for

strategic insight into the diverse aspects of the emerging electricity marketplace, and Q-Learning algorithm is used to generate the price offers for generation companies in a bilateral contracts market for electricity.

In this paper, the bidding strategy is viewed as one kind of stochastic optimal control problem known as the Markovian Decision Problem (MDP) [16], and Q-Learning algorithm [16]~[18] is used to develop an optimal bidding strategy for suppliers to maximize their long-term profits in a daily repeated electricity auction market. It is assumed that no supplier possesses the market power, which can be used to manipulate the market price to satisfy his/her own interest. Each market participant in this market is assumed to have only information on his/her own cost and the publicly available information of the market, but lack information on other participants. The market participants are also assumed to be so many that it is very difficult for each supplier to estimate other participants' bidding behaviors. But, each participant is designed to have the ability to use the public information of the market and to learn from experience,

Currently, there are two major market pricing rules adopted in the electricity auction markets around the world. One is the uniform pricing rule, the other is the discriminatory pricing rule ("pay-as-bid"). Which one is the better mechanism for electricity auction markets is still an open question. It is also widely believed that, developing bidding strategy under different market designs can provide a deep insight into the complex new electricity markets and identify how rules can be altered to improve the performance of the market. Based on these considerations, in this paper, we use the electricity auction market, which is under discriminatory pricing rule, as the stage on which we develop bidding strategy for electricity suppliers. And simulation results will show that, even under the discriminatory pricing rule, when the power supply is bigger than the power demand, intensive competition among suppliers forces them to bid prices close to their true costs.

This paper is organized as follows: Section 2 describes the model of a day-ahead electricity auction market. Section 3 presents the Q-Learning algorithm and the proposed supplier bidding strategy. Section 4 shows the simulation results, which is based on a multi-agent simulation method. Section 5 gives the conclusion and future work.

2. A Day-ahead Electricity Auction Market

A day-ahead electricity auction market with no demand-side bidding is assumed here. In this day-ahead auction market, all suppliers wishing to sell power tomorrow must submit their bids today to an Independent Contract Administrator (ICA) [15], who will clear the market, determine which supplier should be used to meet the forecasted load, and check if the security and reliability constraints of the power system are satisfied. The relationship of the ICA and suppliers is shown in Fig. 1.

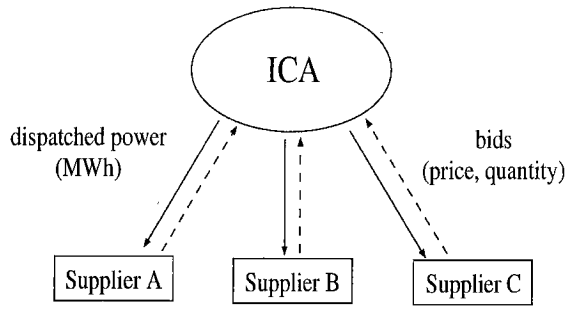


Fig. 1. The relationship of suppliers and the ICA

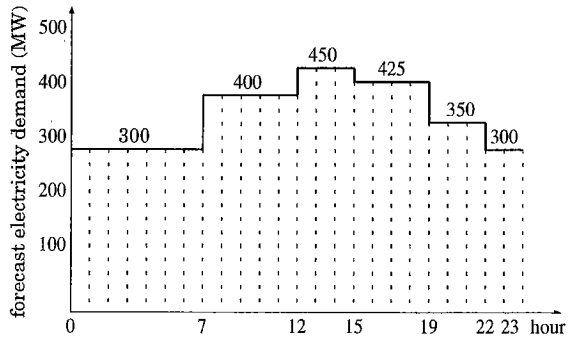


Fig. 2. An example of power load forecasted by the ICA

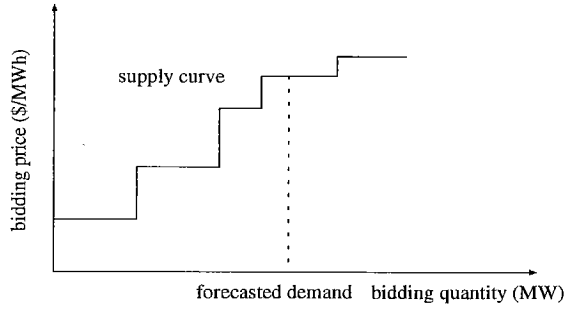


Fig. 3. An example of supply curve at hour h of the next day

Everyday suppliers submit their sealed bids with price ($/MWh) and quantity (MW) at which they are willing to sell during the next day to compete for the power load forecasted by the ICA. An example of forecasted power load by the ICA is shown in Fig. 2. In this paper, hourly bids rule is used, that is, each supplier submits 24 separate hourly bids everyday to compete for power load over the 24-hour of the next day.

The bids from suppliers are ranked by the ICA from the cheapest to the most expensive to construct a supply curve on an hourly basis. Fig. 3 gives an example of the supply curve. The ICA will then select the cheapest supplier until the load of each hour of the next day is met. It should be pointed out that we regulate in this clearing algorithm, when the bidding prices of several suppliers are the same, the supplier with smaller bidding quantity is given the first priority to be accepted to protect the medium-and-small size enterprises.

At the end of every trading day, each supplier is notified of his hourly dispatched power (MWh), which is the quantity called into operation during the next day, and

the hourly market price ($/MWh), which is assumed to be the only publicly available information to each supplier in this paper. The market price at hour h is defined to be the average bidding price $P_{avg}(h)$ of dispatched suppliers at hour h as follows:

$$P_{avg}(h) = \frac{\sum_{i=0}^{i=n-1} Dp(i,h) * P(i,h)}{\sum_{i=0}^{i=n-1} Dp(i,h)} \cdots\cdots\cdots (1)$$

where n denotes the number of the suppliers in the electricity auction market, $P(i,h)$ represents the bidding price ($/MWh) of supplier i at hour h, and $Dp(i,h)$ is the dispatched power (MWh) of supplier i at hour h.

In general, increasing the amount of information available to all bidders could increase the efficiency of the auction market. Therefore, it is better for the Independent Contract Administration (ICA) to publish other information, such as the maximum and minimum bid prices of everyday, as well as the hourly market price. But, the problem is that increasing the amount of publicly available information could at the same time lead to the risk of making the unexpectedly collusive behavior between bidders easier to implement. Therefore, we assume in this paper that the hourly market price is the only publicly available information to all bidders in an attempt to reduce the potentiality of collusive behavior between bidders.

Each supplier winning the market is paid based on a discriminatory pricing rule. Although the discriminatory pricing rule is not so much popular, it is used in UK balancing market [8]. According to this pricing rule we adopt here, winners are paid at their own bidding prices. The reward $\pi(i,h)$ from the bid of each supplier i at hour h is calculated based on the bidding price $P(i,h)$, unit production cost C_i and the dispatched power $Dp(i,h)$:

$$\pi(i,h) = (P(i,h) - C_i) * Dp(i,h) \cdots\cdots\cdots (2)$$

where C_i ($/MWh) is the unit production cost of supplier i's power supply. It should be noted that, in this simplified model, each supplier's production cost is represented as a linear function of his dispatched power, and no startup costs and shut down costs are considered here. In practice, the unit production cost of each supplier's power supply varies with the total output of power supply, and startup costs, shut down costs and ramp rates of generator cannot be ignored.

3. Developing Bidding Strategy Through Q-Learning Algorithm

Q-Learning (QL) algorithm is a reinforcement learning algorithm [16] proposed by Watkins for solving the Markovian Decision Problems with incomplete information. It does not need an explicit model of its environment and can be used on-line to find the optimal strategy through experience obtained from the direct interaction with its environment. These features make it well suitable for dealing with the decision-making problems in the repeated games against unknown opponents,

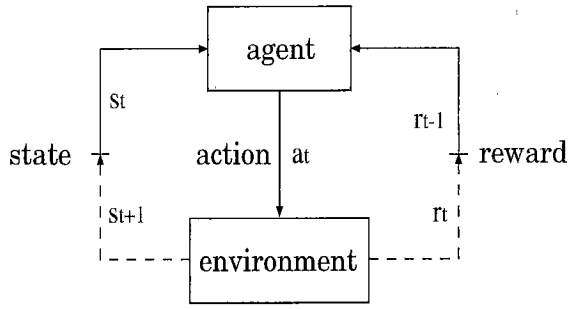


Fig. 4. An illustration of agent's interaction with its environment

such as the bidding strategy in the electricity auction market. Based on the Q-Learning algorithm, this section provides an optimal bidding strategy for suppliers to maximize their profits in the day-ahead electricity auction market.

3.1 Q-Learning Algorithm Assume that a learning agent interacts with its environment at each of a sequence of discrete time steps, $t = 0, 1, 2, \ldots$, as shown in Fig. 4. And let $S = \{s_1, s_2, s_3, \ldots, s_n\}$ be the finite set of possible states of the environment and $A = \{a_1, a_2, a_3, \ldots, a_n\}$ be the finite set of admissible actions the agent can take. At each time step t, the agent senses the current state $s_t = s \in S$ of its environment, and on that basis selects an action $a_t = a \in A$. As a result of its action, the agent receives an immediate reward r_t, and the environment's state changes to the new state $s_{t+1} = s' \in S$ with a transition probability $P_{ss'}(a)$.

The objective of the agent is to find an optimal policy $\pi^*(s) \in A$ for each state s to maximize the total amount of reward it receives over the long run. Q-Learning algorithm provides an efficient on-line approach to determine the optimal policy by estimating the optimal Q-values $Q^*(s,a)$ for pairs of states and admissible actions.

The Bellman optimality equation for $Q^*(s,a)$ is given as follows:

$$Q^*(s,a) = \sum_{s'} P_{ss'}(a)[R^a_{ss'} + \gamma \max_{a'} Q^*(s', a')] \cdots\cdots\cdots (3)$$

where $R^a_{ss'} = r_t$ is the immediate reward from taking action a in the state s and transitioning from state s to s', and γ ($0 \leq \gamma \leq 1$) is a scaling factor used to discount the future rewards. If γ is small, it means that the expected future rewards count for less.

Any policy selecting actions that are greedy with respect to the optimal Q-values is an optimal policy [17]. Thus, the optimal policy is

$$\pi^*(s) = arg \max_a (Q^*(s,a)) \cdots\cdots\cdots (4)$$

Without knowing the $P_{ss'}(a)$, the Q-Learning algorithm can find the $Q^*(s,a)$ in a recursive manner by using the available information s_t, a_t, s_{t+1} and r_t. The update rule for Q-Learning is

$$Q_{t+1}(s,a) = \begin{cases} Q_t(s,a) + \alpha \Delta Q_t(s,a) \\ \qquad \text{if } s = s_t \text{ and } a = a_t \\ Q_t(s,a) \quad \text{otherwise} \end{cases} \quad \cdots\cdots\cdots\cdots\cdots (5)$$

where α is the learning rate and

$$\Delta Q_t(s,a) = \{r_t + \gamma \max_{a'}[Q_t(s_{t+1}, a')]\} - Q_t(s,a) \quad \cdots\cdots\cdots\cdots\cdots (6)$$

The learning rate α $(0 < \alpha \leq 1)$ reflects the degree to which estimated Q-values are updated by new data. High values imply more rapid updates, with a risk of instability [9].

If the Q-value for each admissible state-action pair (s,a) is visited infinitely often, and the learning rate α decreases over the time step t in a suitable way, then as $t \to \infty$, $Q_t(s,a)$ converges with probability one to $Q^*(s,a)$ for all admissible pairs (s,a).

3.2 QL-based Supplier Bidding Strategy In the repeated day-ahead electricity auction market, each supplier will attempt to maximize his/her profit in a long run and to reduce risks. The need to maximize profit and manage risks at the same time is becoming a dominant industry problem [11]. Based on the Q-Learning algorithm, a bidding strategy for suppliers is developed to balance the tradeoff between the expected profit and risks. As we assumed in the above section that the production cost of each supplier at each hour is a linear function of his dispatched power, so each supplier will bid his maximum generation capacity (MW) as his bidding quantity at each hour of every trading day to attempt to maximize his profit in this auction market, with an expectation that his generator will run at the maximum capacity all day. Therefore, the bidding strategy results in an hourly bidding price decision-making problem.

As described earlier, it is assumed that each supplier has information only on his/her own cost and the public information of hourly market price, but lacks of information on the rivals. Thus, the bidding process is a stochastic process. During this stochastic bidding process, each supplier will attempt to meet his/her objectives of:

- increasing his/her profit from day to day,
- satisfying the target utilization rate on his/her generator everyday,

as described in Ref. (10). The target utilization rate is defined as the ratio between the expected dispatched power (MWh) and the maximum power output (MWh) of generator everyday.

To apply the Q-Learning algorithm in the bidding strategy for suppliers to achieve their objectives, it is necessary to define the states, actions, and rewards first.

(1) *States* The state of environment is represented by the market price, and has 20 different levels which is equally distributed between 0 $/MWh and the market ceiling price that is specified to 20 $/MWh here. As shown in Fig. 5, if the market price is within the interval of (19,20] $/MWh, then the environment's state is in state 19. It should be noted that the state of environment should include other market information such as predict power load of the next trading day, but these are not taken into account here for simplicity.

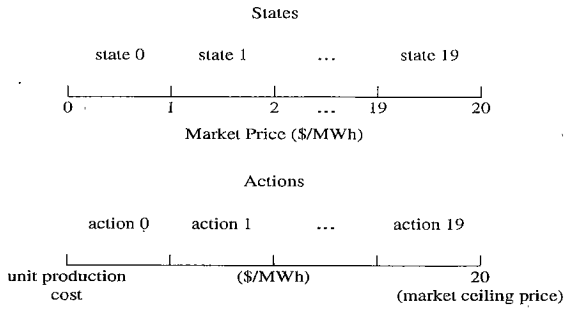


Fig. 5. The definition of environment's states and agent's actions

(2) *Actions* Each rational supplier will generate bidding price between his/her unit production cost and the market ceiling price. Therefore, it is assumed here that each supplier's admissible actions are represented by 20 intervals which are equally distributed between his/her unit production cost and the market ceiling price, as shown in Fig. 5. Applying an action a is to randomly generate a bidding price in the ath interval.

(3) *Rewards* Taking into account the requirement of utilization rate on supplier's generator everyday, it is assumed here that it is required a constant utilization rate on generator for each hour of everyday, for simplicity. The reward of supplier i from his bids at hour h under action a and state s, considering the constant utilization rate for that hour, is calculated based on the following formula, which can be viewed as a penalty function:

$$r_{i,h}(s,a) = \pi(i,h) * \left(\frac{utl_a(h)}{utl_t}\right)^n \quad \cdots\cdots\cdots\cdots (7)$$

where utl_t is the target utilization rate, $utl_a(h)$ is the actual utilization rate at hour h, and n $(n = 0, 1, 2, \ldots)$ is a constant which shows how strictly a supplier tries to satisfy the requirement of utilization rate. If n is large, it implies that the supplier is strict in satisfying the requirement of utilization rate, and can be thought to be of risk averse type. If $n = 0$, there is no penalty effect on the reward, and the supplier can be viewed to be opportunistic.

3.3 Algorithm Implementation As described earlier, startup costs, shut down costs and ramp rates of generators are not considered. It implies that whether the operation status of generator is on or off at any hour, it does not affect the operation status of generator at the next hour. Therefore, the whole day profit-maximizing problem can be decomposed into an hourly profit-maximizing problem. Based on this consideration, in this paper, Q-values for state-action pairs at each hour of a supplier are stored in a lookup table. An example of the Q-values for state-action pairs is given in Table 1, where the Q-values in bold style are the maximum values under each state and the actions associated with them are the optimal actions a supplier would take most likely.

The steps of suppliers' learning and bidding are given as follows:

Table 1. An example of the Q-values of state - action pairs

	...	action 13	action 14	...	action 16	...
...						
state 15		**493**	379		153	
state 16		**485**	257		290	
state 17		479	**529**		298	
state 18		501	552		**602**	
...						

(1) *Step 1: State identification* At the beginning of the current trading day, each supplier uses the publicly available 24 separate hourly market prices on the previous trading day as the 24 hourly states of the current trading day.

(2) *Step 2: Action selection* After having obtained the 24 hourly states, each supplier inquires his/her Q-value lookup tables to select the optimal action with maximum Q-value in each state and generate the bidding price at each hour according to the definition of an action.

To balance the exploration and exploitation of suppliers' learning from the dynamic electricity auction market, ϵ-greedy method [16] is introduced to the QL-based supplier bidding strategy. That is, during the action selection process, the supplier selects most of the time an action a with maximum $Q(s,a)$ in the state s; but, with a small probability ϵ, he also randomly selects an action a from all the admissible actions in the state s, independently of the Q-values $Q(s,a)$, to explore the new optimal bidding strategy in the dynamically competitive market.

(3) *Step 3: Q-value update* At the end of the current trading day, after being notified of the dispatched power and the market price at each hour, each supplier calculates the rewards according to Eq. (7), and updates the Q-values of each hour based on the available rewards and next states which are the hourly market prices of current trading day, according to Eqs. (5), (6).

4. Simulation Results

An agent-based simulation method is developed here to test the bidding strategy proposed in the above section. The application of agent-based simulation method to deal with issues in the deregulated electricity industry is a newly promising research area [12] [13]. In this paper, it is assumed that there are 10 adaptive agents, each of them representing a supplier who participates in the day-ahead electricity auction market and is able to explore and exploit the optimal bidding strategy to meet his/her profit-maximizing goal in the competitive environment. Table 2 gives the maximum capacities, unit production costs and strategic parameters of these agents. As can be seen from the capacity levels of all agents, no supplier possesses the market power since no agent has the dominant market share.

According to the definition of each strategic parameter, these 10 agents can be divided into two categories. Agents with parameters: $\alpha = 0.7$, $\gamma = 0.1$, $\epsilon = 0.1$, $n = 2$ can be viewed to be risk averse; others can be thought to be of opportunistic type. Comparing the agents of opportunistic type, these risk averse agents can adjust their optimal policy quickly as the environment changes, count the expected future rewards for less when making their decisions (selecting the optimal actions), maintain enough ability to explore new optimal bidding strategy in dynamic environment and are less greedy.

Table 2. The maximum capacities (MC), unit production costs (UPC) and strategic parameters of 10 agents

agent	MC (MW)	UPC ($/MWh)	α	γ	ϵ	n	utl_t
0	45	8.0	0.7	0.1	0.1	2	0.90
1	45	8.0	0.7	0.1	0.1	2	0.90
2	45	8.0	0.1	0.5	0.01	1	0.80
3	45	8.0	0.1	0.5	0.01	1	0.80
4	50	10.0	0.7	0.1	0.1	2	0.90
5	50	10.0	0.1	0.5	0.01	1	0.80
6	50	10.0	0.1	0.5	0.01	1	0.80
7	30	12.0	0.7	0.1	0.1	2	0.80
8	30	12.0	0.1	0.5	0.01	1	0.70
9	30	12.0	0.1	0.5	0.01	1	0.70

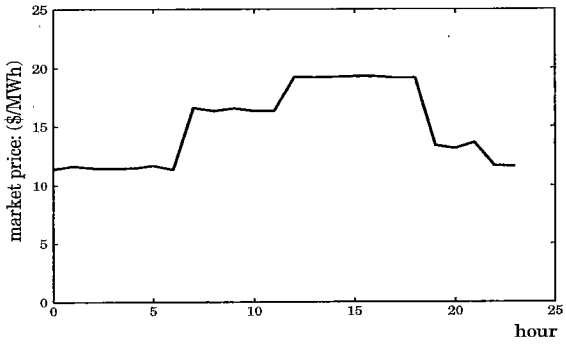


Fig. 6. The average hourly market price of every trading day

To obtain the initial Q-values of each agent at each hour, the simulation process is designed to run first for 10,000 trading days with the forecasted power demand by the ICA shown in Fig. 2. During this initial process, the learning rate α is designed to be state-action dependent varying with time, as used in Ref. (14). That is, the learning rate $\alpha_{d,h}(s,a)$ of each agent at hour h on the trading day d is inversely proportional to the visited number $\beta_{d,h}(s,a)$ of state-action pair (s,a) up to the present trading day, as follows:

$$\alpha_{d,h}(s,a) = \frac{1}{\beta_{d,h}(s,a)} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (8)$$

After this learning process, the learning rate α is set back to the value shown in Table 2, and the learned initial Q-values are used to develop optimal bidding strategy proposed in the Section 3.

Firstly, the impacts of the proposed bidding strategy on the market price are investigated. With the learned Q-values, the simulation process are carried out for another 1,000 trading days under the same power demand condition as shown in Fig. 2.

Fig. 6 shows the average hourly market price during

Table 3. The average daily rewards and actual utilization rate of the 10 agents

agent	0	1	2	3	4
rewards ($)	7316	7333	6957	6952	6009
actual utilization rate	0.96	0.97	0.93	0.93	0.95

agent	5	6	7	8	9
rewards ($)	5702	5744	2456	2338	2322
actual utilization rate	0.87	0.87	0.71	0.61	0.60

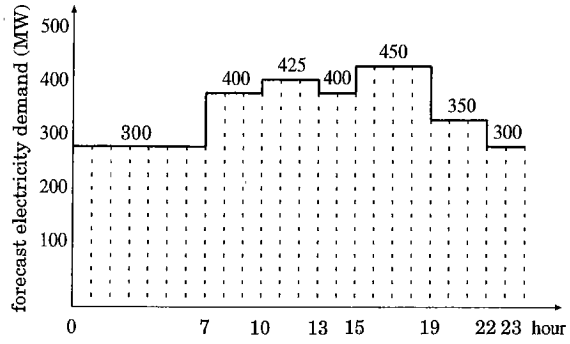


Fig. 7. The power load forecasted by the ICA after 1,000 trading days

the 1,000 trading days. As shown in this figure, the intense competition among agents leads to the lowest market prices during the off-peak load periods such as at hour 0, 1 and so on, where the electricity supply is much bigger than the power demand. However, during the peak load period from hour 12 through 18, at which there are shortages of the power supply, the market prices are very close to the market ceiling price due to the agents' learning to exercise market power. These facts show that the proposed QL-based bidding strategy is successful in generating optimal bidding prices at different hours for agents in the day-ahead electricity auction market.

Table 3 gives the average daily rewards and actual utilization rate of these 10 agents during the 1,000 trading days. As shown in this table, the risk averse agents can get more rewards from their strategic bidding and have higher actual utilization rate on their generators than those of opportunistic type at the same generation capacity level.

Secondly, with the learned Q-values, the proposed bidding strategy is tested for 2,000 trading days with a different power load case, where the power load forecasted by the ICA during the first 1,000 trading days are the same as shown in Fig. 2, but those in the following 1,000 trading days are changed as shown in the Fig. 7. This change of the power load pattern can be thought to be due to the change of seasons.

The average hourly market prices of the final 200 trading days of the first and second 1,000 trading days are shown in Fig. 8. The changes of market price occur only during the periods when the power loads are changed. This fact suggests to some extent that the interactions of these adaptive agents with QL-based bidding strategy lead to a stable market equilibrium over a long term in a stationary power load case, and also shows the proposed approach is capable of dealing with the bidding

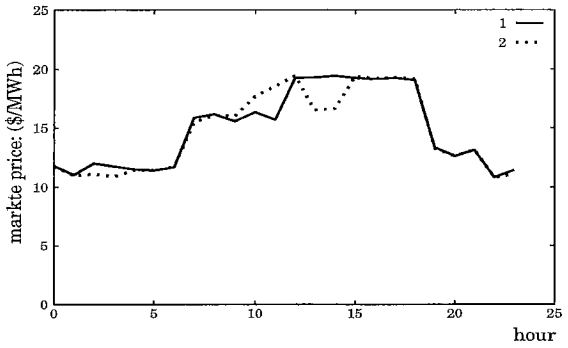


Fig. 8. The average hourly market price of the final 200 trading days 1) during the first 1,000 trading days, and 2) during the second 1,000 trading days

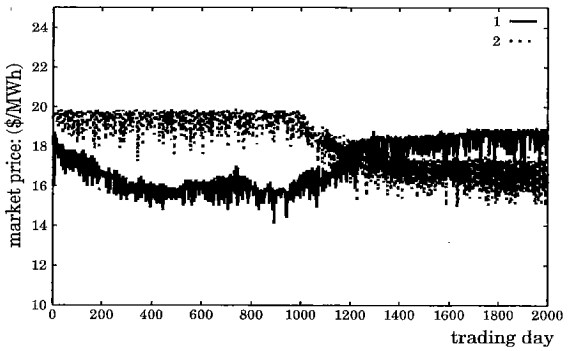


Fig. 9. The market price 1) at hour 11, and 2) at hour 14 of everyday during the 2,000 trading days

price decision-making problem in the dynamic auction market.

How the agents learn can be seen in Fig. 9, which displays the market prices at hour 11 and hour 14 during the 2,000 trading days. Power loads are changed from 400 MW to 425 MW at hour 11, and from 450 MW to 400 MW at hour 14, respectively, on the trading day 1,000. As can be seen from this figure, when the power load changes, the market price changes accordingly as adaptive agents interact with each other and learn from experience to develop their optimal bidding prices. It implies that agents have sensed the change of power demand and adjusted their bidding prices accordingly. It should be noted that, it takes a few days for the market prices to reach new dynamic equilibria at these hours when the power demands are changed. This can be explained that Q-Learning algorithm has a slow convergence. To reduce the influence of this drawback of Q-Learning algorithm to some extent, it is believed to be an effective way for agents to predict the hourly market price and power demand and make their bidding decisions accordingly.

5. Conclusion and Future Work

Based on the Q-Learning algorithm, an optimal bidding strategy was proposed in this paper to provide suppliers an optimal approach to maximize their profits in the long run from the day-ahead electricity auction market. Each supplier with the proposed bidding strategy can learn from experience and make full use of the

public information of the market. A penalty function was introduced to the calculation of the reward from supplier's bids. Supplier's business type—risk averse or opportunistic—were considered, the impacts of corresponding strategy on the rewards and utilization rate on generator were investigated. Also the impacts of the proposed bidding strategy on the market price were analyzed. Simulation results have shown the feasibility of this QL-based supplier bidding strategy.

In this paper, we developed a novel supplier bidding strategy in a day-ahead electricity auction market where discriminatory pricing rule is used. Extending our methodology to study markets where the uniform pricing rule is adopted will be our future work.

(Manuscript received April, 19, 2002,
revised September, 27, 2002)

References

(1) D. Fudenberg and J. Tirole: Game Theory, The MIT Press, Cambridge, Massachusetts (1991)

(2) R.W. Ferrero, J.F. Rivera, and S.M. Shahidehpour: "Application of games with incomplete information for pricing electricity in deregulated power pools," *IEEE Trans. Power Syst.*, Vol.13, No.1, pp.184–189 (1998-2)

(3) F.S. Wen, and A.K. David: "Oligopoly Electricity Market Production under Incomplete Information," *IEEE Power Engineer Rev.*, pp.58–61 (2001-4)

(4) K. Seeley, J. Lawarree, and C.C. Liu: "Analysis of Electricity Market Rules and Their Effects on Strategic Behavior in a Noncongestive Grid," *IEEE Trans. Power Syst*, Vol.15, No.1, pp.157–162 (2000-2)

(5) C.W. Richter, Jr and G.B. Sheble: "Genetic Algorithm Evolution of Utility Bidding Strategies for the Competitive Marketplace," *IEEE Trans. Power Syst.*, Vol.13, No.1, pp.256–261 (1998-2)

(6) H.L. Song, C.C. Liu, J. Lawarree, and R.W. Dahlgren: "Optimal Electricity Supply Bidding by Markov Decision Process," *IEEE Trans. Power Syst.*, Vol.15, No.2, pp.618–624 (2000-5)

(7) F.S. Wen and A.K. David: "Optimal Bidding Strategies and Modeling of Imperfect Information Among Competitive Generators," *IEEE Trans. Power Syst.*, Vol.16, No.1, pp.15–21 (2001-2)

(8) D.W. Bunn and F.S. Oliveria: "Agent-Based Simulation—An Application to the New Electricity Trading Arrangements of England and Wales," *IEEE Trans. Evolutionary Computation*, Vol.5, No.5 (2001-10)

(9) SEPIA's web site, http://www.htc.honeywell.com/projects/sepia

(10) J. Bower and D. Bunn: "Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the England and Wales electricity market," *J. Economic Dynamics & Control*, Vol.25, pp.561–592 (2001-3)

(11) M. Ilic and P. Skantze: "Electric Power Systems Operation by Decision and Control," *IEEE Control Systems Magazine*, Vol.20, No.4, pp.25–39 (2000-8)

(12) C.C. Liu, J.H. Jung, G.T. Heydt, V. Vittal, and A.G. Phadke: "The Strategic Power Infrastructure Defense (SPID) System," *IEEE Control Systems Magazine*, Vol.20, No.4, pp.40–52 (2000-8)

(13) S.A. Harp, A. Brignone, B.F. Wollenberg, and T. Samad: "SEPIA: A Simulator for Electric Power Industry Agents," *IEEE Control Systems Magazine*, Vol.20, No.4, pp.53–69 (2000-8)

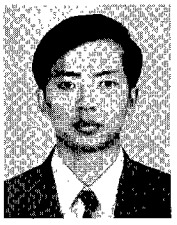
(14) J. Nie and S. Haykin: "A Dynamic Channel Assignment Policy Through Q-Learning," *IEEE Trans. Neural Networks*, Vol.10, No.6, pp.1443–1455 (1999-11)

(15) A.R. Bergen and V. Vittal: Power Systems Analysis, 2nd ed., Prentice-Hall (2000)

(16) R.S. Sutton and A.G. Barto: Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA (1998)

(17) A.G. Barto, S.J. Bradtke, and S.P. Singh: "Learning to act using real-time dynamic programming," *Artificial Intelligence*, Vol.72, pp.81–138 (1995)

(18) C.J.C.H. Watkins: Learning from Delayed Rewards, PhD thesis, Cambridge University, Cambridge, England (1989)



Gaofeng Xiong (Student Member) received his B.S., M.S. degrees in Electrical Engineering from Hunan University, P.R.China, in 1992 and 1995, respectively. He was a faculty member at College of Mechanical and Automotive Engineering, Hunan University, from 1995 to 1999. He is currently a Ph.D. student at Electrical Engineering, Nagoya University, Japan. His research interests include power system economics, evolutionary computation, and reinforcement learning. He is a student member of the IEE of Japan, and the IEEE.

Tomonori Hashiyama (Member) received his Ph.D. degree in Electrical Engineering from Nagoya University, Japan, in 1996. He is currently an associate professor at the Institute of Natural Sciences, Nagoya City University. His research interest include evolutionary computation, optimization, and intelligent system applications. He is a member of the IEE of Japan, and the IEEE.

Shigeru Okuma (Member) received his M.E. degree in systems engineering from Case Western Reserve University, OH, U.S.A, and his Ph.D. degree in Electrical Engineering from Nagoya University, Japan, in 1974 and 1978, respectively. Since 1990, he has been a Professor of Electrical Engineering at Nagoya University. His research interests are in the areas of power electronics, robotics, and evolutionary computation. He is a member of the IEE of Japan, and the IEEE.