Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2014) 22: 34 – 42
© TÜBİTAK
doi:10.3906/elk-1205-12

A learning method to evaluate a generation company's bidding strategy in the electricity market

Shengjie YANG, Jiangang YAO*
College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, P.R. China

Received: 07.05.2012 • **Accepted:** 14.09.2012 • **Published Online:** 20.12.2013 • **Printed:** 20.01.2014

Abstract: In the electricity market, generation companies (GenCos) are usually faced with the problem of choosing a better bidding strategy. They often have to evaluate each possible strategy according to its potential reward. In a competitive market environment, the electricity price is stochastic and volatile, and the GenCo's mixed strategies also make the problem more complicated. In this paper, we model the market price with a Markov regime-switching model and propose the temporal difference learning method in the Markov decision process to approximate the expected reward over an infinite horizon. The simulations based on this method have achieved the evaluation of 2 mixed strategies. The results show the difference of the expected rewards between the strategies, which could be important evidence for choosing a better strategy.

Key words: Bidding strategy, Markov decision process, temporal difference learning, Markov regime-switching

1. Introduction

Since the 1990s, liberalization has been the main trend of the world's electric power industry. It is generally accepted that a competitive market can bring new technologies and stimulate investments. The participants in the market have to deal with more uncertainties and suffer more risks. They deploy increasingly complex bidding strategies in the trading.

In recent years, a few researchers have been devoted to this area. They studied the models and the related methods of optimal supply bidding in the electricity market. Kian et al. [1] used dynamic game theory to study optimal bidding strategies in dynamic oligopolistic electricity markets. Galiana et al. [2] compared the performances of mixed strategies in pool and bilateral trading. Song et al. [3] modeled the bidding in the spot market as a standard Markov decision process (MDP). Bagnall [4] developed a simulated model of the UK market with learning classifier systems. Guerci et al. [5] modeled the generation companies (GenCos) as learning agents in an artificial power exchange, and the Marimon and McGrattan learning and Q learning methods were compared by simulation. Hence, the optimal strategies could be acquired in several ways. An efficient method is needed to compare them before making the final choice. In the stochastic environment, the parameters are not directly observable, and the action taken in the current state may also have an effect on the future reward. Thus, it is not easy to evaluate a bidding strategy, especially if it is a mixed one.

The reward of a bidding strategy is associated with many factors, including the market rules, nonconvex cost curve, opponents' strategies, and network constraints. Researchers could achieve the strategy evaluation

[*] Correspondence: wwmmyang@gmail.com

based on the modeling of the physical structure [6]. However, the problem becomes complicated when the analysis is applied to a large system in a practical case. In this paper, we consider the clearing price as the aggregation of these factors. Based on the Markov regime-switching (MRS) model of price forecasting, we propose a learning method in the MDP framework to evaluate the bidding strategy. Taking the regimes as the states of the market, temporal difference (TD) learning could achieve the approximations of the value functions in the stochastic environment, which has no requirement about the explicit expression of the states' internal transition mechanism.

The rest of this paper is organized as follows. In Section 2, we introduce the problem of strategic bidding in the framework of the MDP. In Section 3, TD learning is proposed to approximate the value functions of the mixed strategy. In Section 4, the simulation is achieved based on the evaluation method. Section 5 is devoted to the conclusions.

2. MDP framework

A standard MDP is usually defined as a tuple $(X,\ A,\ Pr,\ R,\ \beta)$. X is the finite state set of the environment. A is the finite action set of the decision maker. Pr is the transition matrix and $Pr(x,\ a,\ y)$ represents the probability of the transition from state x to y when the decision maker takes action a. $R$is the reward received from the environment after each transition. $\beta$is the discount factor.

It is usually considered to model the GenCo's repeated bidding as a MDP. The GenCo submits the strategic bid to the market, receives the reward, observes the market state, and then prepares for the next bidding. There are some cases about whether the GenCo's actions would have an effect on the market, which is true in most of the oligopolistic markets. However, in the electricity market, as the number of participants and the capacity of units both grow large enough, the GenCo, as an individual participant, is more likely to be a price taker and could change little of the clearing price in a large-scale market environment. Therefore, in this paper, we do not consider the effect of action on the transition. The transition matrix is redefined as:

$$Pr(x,a,y) = Pr(x,y). \tag{1}$$

2.1. Market state

In the electricity market, after the period of submitting bids from GenCos to the market, the independent system operator (ISO) will clear the market and return the reward to the GenCos according to the clearing price and their bidding curves. Actually, the clearing price would determine the reward, and the bidding strategy is mainly based on the forecasting of the clearing price. However, in a competitive market, the price is usually highly stochastic and volatile, and it is affected by many factors, such as demand, network constraints, or the submitted bids. For several years, researchers have developed plenty of methods to forecast price. The state-of-the-art models include the autoregressive integrated moving average [7,8], transfer function [9], support vector machine [10], and neural networks [11,12].

The seasonal fluctuations in electricity prices have been acknowledged for years, which is especially true in the spot market. Researchers have studied the spot prices in markets like EEX, OMEL, PJM, and Nord Pool [13,14]. It is evident that structural changes in the market environment usually cause seasonal fluctuations of the price. These changes mainly include temperature, economic situation, weekly holiday or working, quantity of rainfall, and periodic maintenance of large units. They usually affect the trend of price more predictably. However, even if the environment remains stable for a short-term period, the price could still be highly volatile and abrupt, which may be caused by unpredictable factors such as market competition, failure of units, or transmission lines. Based on that, a price model of the MRS is widely examined by researchers. Vucetic et al. [15] characterized the price time series by a number of regimes according to the price-load relationship. Alizadeh et al. [16] employed a MRS approach for determining the time-varying minimum variance hedge ratio in energy futures markets. Mount et al. [17] proposed a 2-regime model to represent the volatile behavior of electricity prices. Three-regime models of the electricity price were proposed in [18,19]. Higgs et al. [20] also showed that the regime-switching model outperforms the basic stochastic and mean-reverting models.

Here we consider the regimes in the price MRS model as the market state. According to the traditional MRS model, the spot price $Pric_t$ at stage t could be decomposed into 2 independent parts, as follows, a stable and predictable component μ_t and a stochastic component η_t:

$$Pric_t = \mu_t + \eta_t, \tag{2}$$

where μ_t is mainly determined by long-term structural changes throughout weeks or seasons, and η_t, as a stochastic variable, could be modeled by some kinds of probability distribution, such as normal distribution, lognormal distribution [18], or beta distribution [21].

2.2. Bidding action and reward

At stage t, the GenCo chooses an action from its action space to take part in the bidding. A valid supply bid usually contains 2 elements, the price curve and capacity caps. The GenCo's true marginal cost is formulated as:

$$MC(P) = mP + n, \tag{3}$$

where m and n are the nonnegative coefficients. The GenCo's reported marginal cost is formulated similarly as:

$$MC_t(P) = m_tP + n_t, \tag{4}$$

where m_t and n_t are the coefficients as the reported values of m and n. The bidding action could be summarized as $a_t = (m_t,\ n_t,\ C_{\max,t},\ C_{\min,t})$. Each action has several choices about its elements. As illustrated in Figure 1, we usually assume that the elements of a valid bid in the electricity market satisfy the following conditions.

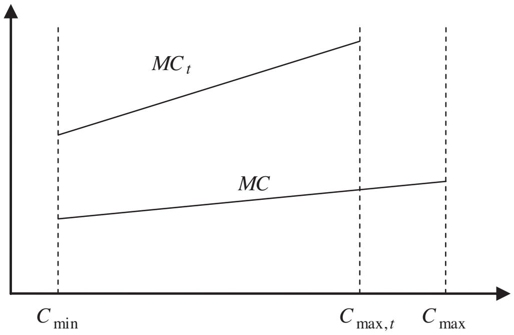

Figure 1. Illustration of the GenCo's bidding.

1) The reported maximum capacity $C_{\max,t}$ is less than the true maximum capacity $C_{\max}$, and the reported minimum capacity $C_{\min,t}$ could be equal to the true minimum capacity $C_{\min}$, like in Eq. (5). Next, a_t is simplified to $(m_t,\ n_t,\ C_{\max,t})$.

$$C_{\min} = C_{\min,t} \leq P \leq C_{\max,t} \leq C_{\max} \tag{5}$$

2) The reported cost MC_t is equal to or greater than the true marginal cost MC.

3) The reported cost MC_t should be in nondecreasing order, which means:

$$m_t = \partial MC_t / \partial P \geq 0. \tag{6}$$

After the period of accepting bids, the ISO will clear the market to meet the demand-supply balance and then publish the clearing price and the corresponding clearing capacity. Here we model the market clearing by sampling the price from the stochastic model of Eq. (2). Thus, the reward R_t is:

$$R_t = \int_0^{P_t} \left[Pric_t(P) - MC(P) \right] \mathrm{d}P, \tag{7}$$

where P_t is the clearing capacity.

2.3. Mixed strategy

We have defined the regime in the price model of the MRS as the market state. The action has also been defined as the reported cost curve and the minimum and maximum capacities. Each time, the GenCo chooses an action according to the current market state and its policy. The deterministic policy represents the deterministic choice of action a in state x_i and is defined as:

$$a = \pi(x_i). \tag{8}$$

More generally, the stochastic policy maps the state set to the distributions over the action set:

$$\pi(x_i) = (\sigma_{i,1}, \sigma_{i,2}, \cdots, \sigma_{i,m}), \tag{9}$$

where $\sigma_{i,j}$ is the probability of taking action a_j in state x_i. In every round of decision making, the action is sampled from the distribution like:

$$a \sim \pi(x_i). \tag{10}$$

The stochastic policies in all of the states form a mixed strategy, which could be described as:

$$\pi(x_1, \cdots x_n) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nm} \end{bmatrix}. \tag{11}$$

3. TD learning

Reinforcement learning is a series of machine learning methods that are mainly based on the theory of the Bellman equation. Different from supervised and unsupervised learning, the learning agent in reinforcement learning would not be told the right or wrong answers, but it would constantly evaluate the performance in the environment. As the agent keeps trying, it will become more experienced at dealing with the invisible internal mechanism and will give more practical evaluations. It is an efficient tool for an agent to learn and control in stochastic conditions. The popular reinforcement learning methods include Q learning and TD learning.

The state representation was described in Section 2, but in reality the transition probabilities and the distribution characteristics of the price are unknown. Thus, traditional methods like dynamic programming are not feasible, and reinforcement learning is adopted to deal with the uncertainty. First, we define a value function to evaluate the strategy according to its performance. Underlying the strategy π, the value function V_π represents the highest possible expected reward in the process, which starts from the state x_0.

$$V_\pi(x) = E[\sum_{t=0}^{\infty} \gamma^t R_{t+1} \,|x = x_0\,] \tag{12}$$

From a forward view, the approximation from n steps is:

$$r_t^{(n)} = R_{t+1} + \gamma R_{t+2} + \ldots + \gamma^{n-1} R_{t+n} + \gamma^n V(s_{t+n}). \tag{13}$$

To provide a unified form of various look-ahead methods, Sutton and Barto [22] introduced a bootstrapping learning method called TD(λ), $0 \leq \lambda \leq 1$. It learns by reducing the discrepancies between the approximations made by the agent at different times. The weighted average of n-step returns, called λ-return, is defined as:

$$r_t^\lambda = (1-\lambda) \sum_{n=1}^{\infty} \lambda^{n-1} r_t^{(n)}. \tag{14}$$

When $\lambda = 0$, the return can be rewritten as Eq. (15), which is known as TD(0). When $\lambda = 1$, there is no approximation about the reward. Next, TD(λ) turns to the every-visit Monte Carlo method and no value will be given until the process is completed.

$$r_t^\lambda = r_t^1 = R_{t+1} + \gamma V(s_{t+1}) \tag{15}$$

To achieve the multistep learning, eligibility traces works as the temporary record of the occurrence of an event, such as the visiting of a state or the taking of an action. The trace is updated at each step for every state. It marks the memory parameters associated with the event as eligible for the current learning changes. The eligibility trace is incremented by 1 for the current state and decayed by $\gamma\lambda$, as shown in Figure 2.

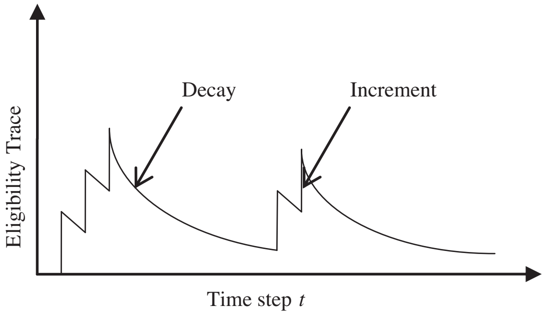

Figure 2. Illustration of the accumulating eligibility trace.

Above all, the pseudocode of the TD(λ) method is represented as follows:

```
1. Initialize $s_0$, $z_0$, $N$, $V$
2. Loop1 for $N$ times
3. Take action $a_t$ according to its strategy
```

```
4. Receive the reward $R_{t+1}$, and observe next state $x_{t+1}$
5. $\delta_{t+1} = R_{t+1} + \gamma V_t(x_{t+1}) - V_t(x_t)$    (16)
6. Loop2 for all states
    7. $z_{t+1}(x) = \begin{cases} 1 + \gamma\lambda z_t(x), & if\ x = x_t \\ \gamma\lambda z_t(x), & if\ x \neq x_t \end{cases}$    (17)
    8. $V_{t+1}(x) = V_t(x) + \alpha\delta_{t+1} z_{t+1}(x)$    (18)
9. End loop2
10. End loop1.
```

4. Simulation

In this section, we construct a repeated bidding market environment with a pay-as-bid auction, and we test the proposed method to evaluate strategies in the electricity market. Without loss of generality, we could assume that there are 3 states, S_1, S_2, and S_3, in the market, which correspond to prices at low, medium, and high levels. The prices in different states are also stochastic, maybe following normal or logarithmic normal distribution, as indicated by some researchers. Here we assume that the lower and upper caps of the clearing price are US\$28/MWh and \$48/MWh. The price of each state follows a normal distribution N (μ, σ^2), where μ is the seasonal component, and $\eta \sim$ N (0, σ^2) is the stochastic component. The values of μ, σ^2, and transition probability P_r are defined in Eqs. (19) and (20). However, these values, as the parameters of the market, are invisible to the GenCo.

$$(\mu, \sigma^2) = \begin{cases} (32, 4), & \text{if} \quad s = S_1 \\ (36, 4), & \text{if} \quad s = S_2 \\ (40, 4), & \text{if} \quad s = S_3 \end{cases} \tag{19}$$

$$P_r = \left\{ \begin{array}{ccc} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.3 & 0.1 \end{array} \right\} \tag{20}$$

As defined in Eqs. (3)–(6), the actual parameters of the GenCo's production are listed in Table 1. We assume there are 3 bidding actions available for the GenCo, Act_1, Act_2, and Act_3. The reported parameters of the actions are listed in Table 2. The parameters of TD learning are listed in Table 3.

Table 1. Actual parameters of the GenCo's production.

m (\$/(MW^{2}h))	n (\$/(MWh))	$C_{\min}$ (MW)	$C_{\max}$ (MW)
0.08	20	100	300

Table 2. Reported parameters of the GenCo's actions.

Action name	m_t (\$/(MW^{2}h))	n_t (\$/(MWh))	$C_{\max,t}$ (MW)
Act_1	0.060	24.80	240
Act_2	0.072	26.45	270
Act_3	0.068	32.40	300

Table 3. Parameters of TD learning.

α	γ	λ	N
0.1	0.9	0.9	2000

In the next step, we evaluate 2 mixed strategies of the GenCo' supply bidding, $Stra_1$ and $Stra_2$. They are defined as:

$$Stra_1 = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}, \tag{21}$$

$$Stra_2 = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.2 & 0.6 \\ 0.3 & 0.5 & 0.2 \end{bmatrix}, \tag{22}$$

where the value in the ith row and the jth column is the probability of taking the jth action in the ith state.

The simulations of the 2 strategies are achieved in the environment of MATLAB R2011b, which costs 0.1601 s and 0.1703 s, respectively. Though the approximated values do not converge to exact points because the rewards and state transitions are both stochastic, they remain on a relatively stable level after 2000 iterations. The results are shown in Figures 3 and 4.

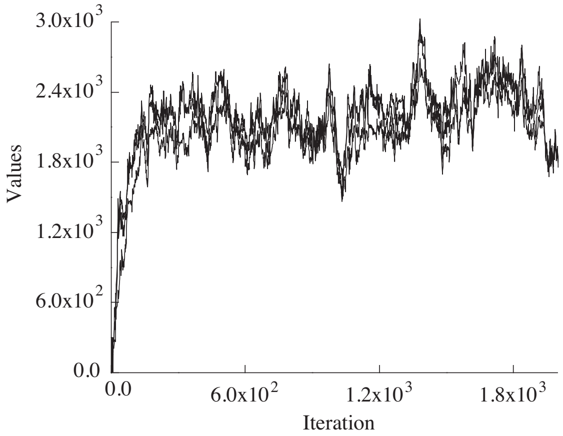

Figure 3. Approximated values of all of the states with $Stra_1$.

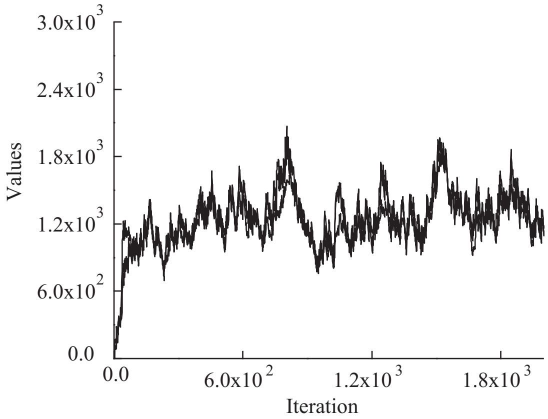

Figure 4. Approximated values of all of the states with $Stra_2$.

In these figures, the approximated values of all of the states are plotted together as the series of curves. There are messages given intuitively here. First, with the same strategy, the values of different states make similar changes in the repeated bidding process, which means that the initial state has little effect on the expected reward over an infinite horizon. Second, the values in Figure 3 are obviously higher than the ones in Figure 4. In other words, the long-term reward with $Stra_1$ is larger than that with $Stra_2$. The simulations are repeated several times. Almost all of them show similar results, so we could generally consider that the mixed strategy $Stra_1$ is better than $Stra_2$.

The simulation gives a simple example to evaluate the bidding strategies. However, it has exhibited the obvious advantage of evaluating the strategy through a price model. The real electricity market may include hundreds of generators and tens of transmission lines, which bring a large number of physical constraints to be satisfied. The interactions between the participants are also stochastic and dynamic. When the analysis takes all of them into account, the computation would be of high dimension and high complexity. As confirmed by the simulation, the proposed method provides a general way to alleviate us from computation pressure. The strategies are evaluated through learning from the price model. By comparing the expected reward acquired in the learning process, we could make the decision of a better strategy.

5. Conclusions

In this paper, we propose the TD learning method to evaluate the mixed strategies in the MDP framework. It achieves the evaluation by approximating the expected reward over the infinite horizon. The basic ideas are based on the price model of the MRS and TD learning, which are widely accepted by the related researchers. The combination of them as an evaluation method in the electricity market has not yet been considered by others. This method integrates plenty of the stochastic factors existing in the realistic occasion, and it also avoids being troubled by the complicated nonlinear model of the electricity grid. Overall, it is attractive to explore the electricity market from a general view of economics, while electricity is more and more revealing of the common features of other commodities in the liberalized market.

References

[1] A.R. Kian, J.B. Cruz, “Bidding strategies in dynamic electricity markets”, Decision Support Systems, Vol. 40, pp. 543–551, 2005.

[2] F.D. Galiana, I. Kockar, P.C. Franco, “Combined pool/bilateral dispatch—Part I: performance of trading strategies”, IEEE Transactions on Power Systems, Vol. 17, pp. 92–99, 2002.

[3] H. Song, C. Liu, J. Lawarrée, R.W. Dahlgren, “Optimal electricity supply bidding by Markov decision process”, IEEE Transactions on Power Systems, Vol. 15, pp. 618–624, 2000.

[4] A. Bagnall, G. Smith, “A multi-agent model of the UK market in electricity generation”, IEEE Transactions on Evolutionary Computation, Vol. 9, pp. 522–536, 2005.

[5] E. Guerci, S. Ivaldi, S. Cincotti, “Learning agents in an artificial power exchange: tacit collusion, market power and efficiency of two double-auction mechanisms”, Computational Economics, Vol. 32, pp. 73–98, 2008.

[6] D. Huang, X. Han, “Study on generation companies’ bidding strategy based on hybrid intelligent method”, Ninth International Conference on Hybrid Intelligent Systems. Vol. 3, pp. 409–412, 2009.

[7] J. Contreras, R. Espínola, F.J. Nogales, A.J. Conejo, “ARIMA models to predict next-day electricity prices”, IEEE Transactions on Power Systems, Vol. 18, pp. 1014–1020, 2003.

[8] A.J. Conejo, M.A. Plazas, R. Espinola, A.B. Molina, “Day-ahead electricity price forecasting using the wavelet transform and ARIMA models”, IEEE Transactions on Power Systems, Vol. 20, pp. 1035–1042, 2005.

[9] F.J. Nogales, J. Contreras, A.J. Conejo, R. Espínola, “Forecasting next-day electricity prices by time series models”, IEEE Transactions on Power Systems, Vol. 17, pp. 342–348, 2002.

[10] D.C. Sansom, T. Downs, T.K. Saha, “Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants”, Journal of Electrical and Electronics Engineering, Vol. 22, pp. 227–234, 2003.

[11] B.R. Szkuta, L.A. Sanavria, T.S. Dillon, “Electricity price short-term forecasting using artificial neural networks”, IEEE Transactions on Power Systems, Vol. 14, pp. 851–857, 1999.

[12] C.P. Rodriguez, G.J. Anders, “Energy price forecasting in the Ontario competitive power system market”, IEEE Transactions on Power Systems, Vol. 19, pp. 366–374, 2004.

[13] R. Weron, “Heavy-tails and regime-switching in electricity prices”, Mathematical Methods of Operations Research, Vol. 69, pp. 457–473, 2008.

[14] I. Simonsen, “Volatility of power markets”, Physica A: Statistical Mechanics and its Applications, Vol. 355, pp. 10–20, 2005.

[15] S. Vucetic, K. Tomsovic, Z. Obradovic, “Discovering price-load relationships in California’s electricity market”, IEEE Transactions on Power Systems, Vol. 16, pp. 280–286, 2001.

[16] A. Alizadeh, N. Nomikos, P. Pouliasis, “A Markov regime switching approach for hedging energy commodities”, Journal of Banking & Finance, Vol. 32, pp. 1970–1983, 2008.

[17] T. Mount, Y. Ning, X. Cai, “Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters”, Energy Economics, Vol. 28, pp. 62–80, 2006.

[18] J. Janczura, R. Weron, “An empirical comparison of alternate regime-switching models for electricity spot prices”, Energy Economics, Vol. 32, pp. 1059–1073, 2010.

[19] R. Huisman, R. Mahieu, “Regime jumps in electricity prices”, Energy Economics, Vol. 25, pp. 425–434, 2003.

[20] H. Higgs, A. Worthington, “Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: the Australian wholesale spot electricity market”, Energy Economics, Vol. 30, pp. 3172–3185, 2008.

[21] R. Becker, S. Hurn, V. Pavlov, “Modelling spikes in electricity prices”, Economic Record, Vol. 83, pp. 371–382, 2007.

[22] R. Sutton, A. Barto, Reinforcement Learning: An Introduction, Cambridge, MIT Press, 1998.