Energy Trading Game for Microgrids Using Reinforcement Learning

Xingyu Xiao, Canhuang Dai, Yanda Li, Changhua Zhou, and Liang Xiao *

Department of Communication Engineering, Xiamen Univ., 361005 China,
Email:lxiao@xmu.edu.cn

Abstract. Due to the intermittent production of renewable energy and the time-varying power demand, microgrids (MGs) can exchange energy with each other to enhance their operational performance and reduce their dependence on power plants. In this paper, we investigate the energy trading game in smart grids, in which each MG chooses its energy trading strategy with its connected MGs and power plants according to the energy generation model, the current battery level, the energy demand, and the energy trading history. The Nash equilibria of this game are provided, revealing the conditions under which the MGs can satisfy their energy demands by using local renewable energy generations. In a dynamic version of the game, a Q-learning based strategy is proposed for an MG to obtain the optimal energy trading strategy with other MGs and the energy plants without being aware of the future energy consumption model and the renewable generation of other MGs in the trading market. We apply the estimated renewable energy generation model of the MG and design a hotbooting technique to exploit the energy trading experiences in similar scenarios to initialize the quality values in the learning process to accelerate the convergence speed. The proposed hotbooting Q-learning based energy trading scheme significantly reduces the total energy that the MGs in the smart grid purchase from the power plant and improves the utility of the MG.

Key words: Energy trading, game theory, reinforcement learning, smart grids.

1 Introduction

As important entities in smart grids, microgrids (MGs) are small-scale power supply networks that consist of renewable energy generators, such as wind turbines and solar panels, local electrical consumers and energy storage devices [1]. Each MG is aware of the local energy supply and the demand profiles of other MGs and the nearby power plant such as the energy selling prices using wireless networks [2]. Therefore, microgrids with extra energy can sell energy to other

* This research was supported in part by National Natural Science Foundation of China under Grant 61671396 and the CCF-Venustech Hongyan Research Initiative (2016-010).

microgrids with insufficient energy to reduce their dependence on the energy generated by the power plants with fossil fuel and save the long-distant energy transmission loss.

Game theory is an important tool to study the energy trading in smart grids [3–8]. For example, the energy demand of consumers and the response of utility companies are formulated as a Stackelberg game in [4], yielding a reserve power management scheme to decide energy trading price. The energy trading of a power facility controller to buy energy from the power plant and multiple residential users was studied in [6], which yields a charging-discharging strategy to minimize the total energy purchase cost. The energy exchange game for MGs formulated in [7] analyzes the subjectivity decision of end-users in the energy exchange with prospect theory. The energy exchange game developed in [8] addresses energy cheating with the indirect reciprocity principle.

However, to our best knowledge, the game theoretical study on energy trading among multiple MGs with heterogeneous and autonomous operators and renewable energy supply are still open issues. In this paper, we formulate the energy exchange interactions among interconnected MGs and the power plant as an energy trading game, in which each MG chooses the amount of energy to sell to or purchase from the connected MGs and the power plants in the smart grid based on its battery level, the energy generation model and the trading history. The MGs negotiate with each other on the amount of trading energy according to the time-varying renewable energy generation and power demand of the MGs. The energy generation model such as [13] is incorporated in the energy trading game to estimate the renewable energy generation. The Nash equilibrium (NE) of this game is derived, disclosing the conditions that the MGs are motivated to provide their extra renewable energy to other MGs and purchase less energy from the power plants.

Reinforcement learning techniques, such as Q-learning can be used by smart grids to manage the energy storage and generation. For example, a temporal difference-learning based storage control scheme proposed in [9] for the residential users can minimize the electric bill without knowing the power conversion efficiencies of the DC/AC converters. The Q-learning algorithm based heterogeneous storage control system with multiple battery types proposed in [10] improves the system efficiency. In a two-layer Markov model based on reinforcement learning investigated in [11], generators choose whether to participate in the next days generation process in the power grid to improve both the day-ahead and real-time reliability. However, these works focus on the energy storage and generation rather than the energy trading among the MGs.

In this paper, a Q-learning based energy trading strategy is proposed for the MG to derive the optimal policy via trial-and-errors without being aware of the energy demand model and the storage level of other MGs in the dynamic game. To accelerate the learning speed, we exploit the renewable energy generation model in the learning process and design a hotbooting technique that applies the trading experiences in similar smart grid scenarios to initialize the quality values of the Q-learning algorithm at the beginning of the game. Simulation

results show that the hotbooting Q-learning based energy trading scheme further promotes the energy trading among the connected MGs in a smart grid, reduces the reliance on the energy from the power plants, and significantly improves the utility of the MGs.

The rest of this paper is organized as follows: The energy trading game is formulated in Section 2, and the NE of the game is provided in Section 3. A hotbooting Q-learning based energy trading strategy is proposed for the dynamic game in Section 4. Simulation results are provided in Section 5, and conclusions are drawn in Section 6.

2 Energy Trading Game

We consider an energy trading game consisting of N MGs that are connected with each other and a power plant in the main grid via a substation. Each MG is equipped with renewable power generators, active loads, electricity storage devices, and the power transmission lines connecting with other MGs and the power plant. A microgrid has energy supply from other microgirds, the power plant, and local renewable energy generators based on wind, photovoltaic, biomass, and tidal energy.

The renewable energy generation such as wind power is local-independent, intermittent and time-varying. The amount of the energy generated by renewable power generators in MG i at time k denoted by $g_i^{(k)}$ can be estimated via the power generation history and the modeling method such as [13], yielding an estimated amount of the generated power denoted by $\hat{g}_i^{(k)}$. For simplicity, the estimation error regarding $g_i^{(k)}$ is assumed to follow a uniform distribution, given by

$$g_i^{(k)} - \hat{g_i}^{(k)} \sim G \cdot \mathrm{U}(-1, 1), \tag{1}$$

where G is the maximum estimation error.

In a smart grid, the energy trading interaction among the MGs can be formulated as an energy trading game that consists of N players. The amount of energy that MG i intends to sell to (or buy from) MG j before the bargaining is denoted by $x_{ij}^{(k)}$, which is chosen by MG i based on the observed state of the smart grid, such as its battery level, the energy trading prices, and its current energy production, and the energy demand. The trading strategy of MG i at time k is denoted by $\boldsymbol{x}_i^{(k)} = [x_{ij}^{(k)}]_{1 \le j \le N} \in \boldsymbol{X}$, where $\boldsymbol{X}$ is the feasible action set of the MGs and $x_{ii}^{(k)}$ is the amount of energy that MG i intends to trade with the power plant. If $x_{ij}^{(k)} > 0$, MG i intends to sell its extra energy to MG j or the power plant. If $x_{ij}^{(k)} < 0$, MG i aims to buy energy.

Note that sometimes two MGs intend to sell energy to each other at the same time, i.e., $x_{ij}^{(k)} x_{ji}^{(k)} > 0$. This problem has to be addressed with the energy trading bargaining. The resulting actual trading strategy of MG i at time k is

denoted by $\boldsymbol{y}_i^{(k)} = [y_{ij}^{(k)}]_{1\leq j\leq N}$, where $y_{ii}^{(k)}$ and $y_{ij}^{(k)}$ denote the amounts of the energy sold if positive by MG i to the power plant and MG j, respectively, or the amount of the energy purchased from them if negative, with $|y_{ij}^{(k)}| \leq C$, in which C is the maximum amount of energy exchange between two MGs. The time index k is omitted, if no confusion incurs. Therefore, the actual amount of trading energy between MG i and MG j after the bargaining is based on their intention trading interactions and given by

$$y_{ij} = \begin{cases} -\min(-x_{ij}, x_{ji}), & \text{if } x_{ij} < 0,\ x_{ji} > 0 \\ \min(x_{ij}, -x_{ji}), & \text{if } x_{ij} > 0,\ x_{ji} < 0 \\ 0, & \text{o.w.} \end{cases} \tag{2}$$

In this way, we can ensure that $y_{ij} + y_{ji} = 0$, $\forall\, i \neq j$. The amount of the energy that MG i trades with the energy plant is given by

$$y_{ii} = \sum_{1\leq i\neq j\leq N} x_{ij} - \sum_{1\leq i\neq j\leq N} y_{ij}. \tag{3}$$

Energy storage devices, such as batteries, can charge energy if the load in the MG is low and discharge if the load is high. The battery level of MG i, denoted by $b_i^{(k)}$, cannot exceed the storage capacity denoted by B, with $0 < b_i^{(k)} \leq B$. The estimated amount of the local energy demand is denoted by $d_i^{(k)}$, with $0 \leq d_i^{(k)} \leq D_i$, where D_i represents the maximum amount of local energy required by MG i. The battery level of MG i depends on the amount of trading energy, the local energy generation, and the energy demand at that time. For the smart grid with N MGs, we have

$$b_i^{(k)} = b_i^{(k-1)} + g_i^{(k)} - d_i^{(k)} + \sum_{j=1}^{N} y_{ij}^{(k)}. \tag{4}$$

The energy gain of MG i, denoted by $G_i(b)$, is defined as the benefit that MG i obtains from the battery level b, which is nondecreasing with b with $G(0) = 0$. As the logarithmic function is widely used in economics for modeling the preference ordering of users and for decision making [4], we assume that

$$G_i(b) = \beta_i \ln(1 + b), \tag{5}$$

where the positive coefficient β_i represents the ability that MG i satisfies the energy demand of the users.

To encourage the energy exchange among MGs, the local market provides a lower selling price for the trade between MGs denoted by $\rho^{-(k)}$ and a higher buying price denoted by $\rho^{+(k)}$, compared with the prices offered by the power plant which are denoted by $\rho_p^{-(k)}$ and $\rho_p^{+(k)}$, respectively, i.e., $\rho^{-(k)} > \rho_p^{-(k)}$ and $\rho^{+(k)} < \rho_p^{+(k)}$.

The utility of MG i at time k, denoted by $u_i^{(k)}$, depends on the energy gain and the trading profit, given by

$$u_i^{(k)}(\boldsymbol{y}) = \beta \ln\left(1 + b_i^{(k-1)} + g_i^{(k)} - d_i^{(k)} + \sum_{j=1}^{N} y_j\right) - \sum_{j\neq i}^{N} y_j \left(\mathrm{I}(y_j \leq 0)\rho^{-(k)} + \mathrm{I}(y_j > 0)\rho^{+(k)}\right) - y_i\left(\mathrm{I}(y_i \leq 0)\rho_p^{-(k)} + \mathrm{I}(y_i > 0)\rho_p^{+(k)}\right), \tag{6}$$

where $\mathrm{I}(\cdot)$ be an indicator function that equals 1 if the argument is true and 0 otherwise.

3 NE of the Energy Trading Game

We first consider the NE of the energy trading game with $N = 2$ MGs, which is denoted by $\boldsymbol{x_i^*} = [x_{ij}^*]_{1\leq j\leq 2}$. Each MG chooses its energy trading strategy at the NE state to maximize its own utility, if the other MG applies the NE strategy. By definition, we have

$$u_1(\boldsymbol{x_1^*}, \boldsymbol{x_2^*}) \geq u_1(\boldsymbol{x_1}, \boldsymbol{x_2^*}), \forall \boldsymbol{x_1} \in X \tag{7}$$

$$u_2(\boldsymbol{x_1^*}, \boldsymbol{x_2}) \leq u_2(\boldsymbol{x_1^*}, \boldsymbol{x_2^*}), \forall \boldsymbol{x_2} \in X. \tag{8}$$

Theorem 1. *The energy trading game with $N = 2$ microgrids and a power plant has an NE ($\boldsymbol{x_1^*}$, $\boldsymbol{x_2^*}$) given by*

$$\boldsymbol{x_1^*} = \left[0, \frac{\beta}{\rho} - 1 - b_1^{(k-1)} - g_1^{(k)} + d_1^{(k)}\right] \tag{9}$$

$$\boldsymbol{x_2^*} = \left[\frac{\beta}{\rho - 1} - 1 - b_2^{(k-1)} - g_2^{(k)} + d_2^{(k)}, 0\right], \tag{10}$$

if

$$\begin{cases} \rho^- = \rho^+ = \rho_p^+ - 1 = \rho_p^- + 1 = \rho & \text{(11a)} \\ 0 < \dfrac{\beta}{\rho} - 1 - b_1^{(k-1)} - g_1^{(k)} + d_1^{(k)} & \\ \qquad < -\dfrac{\beta}{\rho - 1} + 1 + b_2^{(k-1)} + g_2^{(k)} - d_2^{(k)} & \text{(11b)} \\ |x_{12}| \leq |x_{21}| & \text{(11c)} \\ x_{12} > 0,\ x_{21} < 0. & \text{(11d)} \end{cases}$$

Proof. If (11) holds, by (2) and (3), we have $x_{11} = x_{22} = 0$ and $y_{12} = \min(x_{12}, -x_{21}) = x_{12}$, and thus (6) can be simplified into

$$u_1(\boldsymbol{x_1}, \boldsymbol{x_2^*}) = \beta \ln\left(1 + b_1^{(k-1)} + g_1^{(k)} - d_1^{(k)} + x_{12}\right) - x_{12}\rho, \tag{12}$$

$$u_2(\boldsymbol{x_1^*}, \boldsymbol{x_2}) = \beta \ln\left(1 + b_2^{(k-1)} + g_2^{(k)} - d_2^{(k)} + x_{21}\right) - x_{21}(\rho - 1) + x_{12}^*. \tag{13}$$

Thus, we have

$$\frac{du_1(\boldsymbol{x_1}, \boldsymbol{x_2^*})}{dx_{12}} = \frac{\beta}{1 + b_1^{(k-1)} + g_1^{(k)} - d_1^{(k)} + x_{12}} - \rho, \tag{14}$$

and

$$\frac{d^2u_1(\boldsymbol{x_1}, \boldsymbol{x_2^*})}{dx_{12}^2} = -\frac{\beta}{\left(1 + b_1^{(k-1)} + g_1^{(k)} - d_1^{(k)} + x_{12}\right)^2} < 0, \tag{15}$$

indicating that $u_1(\boldsymbol{x_1}, \boldsymbol{x_2^*})$ is convex in terms of $\boldsymbol{x_1}$. Thus the solution of $du_1(\boldsymbol{x_1}, \boldsymbol{x_2^*})/dx_{12} = 0$ is given by (10). Thus $u_1(\boldsymbol{x_1}, \boldsymbol{x_2^*})$ is maximized by $\boldsymbol{x_1^*}$ in (9), indicating that (7) holds. Similarly, we can prove that (8) holds.

Corollary 1. *At the NE of the energy trading game with $N = 2$ MGs if (11) hold, MG 1 buys y_{12}^* amount of energy from MG 2, and the latter sells $-y_{22}^*$ energy to the power plant, with*

$$y_{12}^* = \frac{\beta}{\rho} - 1 - b_1^{(k-1)} - g_1^{(k)} + d_1^{(k)} \tag{16}$$

$$-y_{22}^* = \beta\frac{2\rho - 1}{\rho(\rho - 1)} + 2 + \sum_{i=1}^{N}\left(b_i^{(k-1)} + g_i^{(k)} - d_i^{(k)}\right), \tag{17}$$

and the utility of MG 1 and that of MG 2 are given respectively by

$$u_1 = \beta\left(\ln\frac{\beta}{\rho} - 1\right) + \rho\left(1 + b_1^{(k-1)} + g_1^{(k)} - d_1^{(k)}\right) \tag{18}$$

$$u_2 = \beta\left(\ln\frac{1}{\rho - 1} - 1 + \frac{1}{\rho}\right) + \rho\left(1 + b_2^{(k-1)} + g_2^{(k)} - d_2^{(k)}\right) - 2 - \sum_{i=1}^{2}\left(b_2^{(k-1)} + g_2^{(k)} - d_2^{(k)}\right). \tag{19}$$

4 Energy Trading based on Hotbooting Q-learning

The repeated interactions among N MGs in a smart grid can be formulated as a dynamic energy trading game. The amounts of the energy that MG i trades with the power plant and other MGs impact on its future battery level and the future trading decisions of other MGs as shown in (2) and (4). Thus the next state observed by the MG depends on the current energy trading decision, indicating a Markov decision process. Therefore, an MG can use Q-learning to derive the optimal trading strategy without knowing other MGs' battery levels and energy demand models in the dynamic game. More specifically, the amount of the energy that MG i intends to sell or purchase in the smart grid at time k, i.e. $\boldsymbol{x_i^{(k)}}$,

Algorithm 1 Hotbooting process for MG i.

Initialize α, γ, $Q_i^*(\boldsymbol{s}_i, \boldsymbol{x}_i)$=0, and $V_i^*(\boldsymbol{s_i})$=0, $\forall \boldsymbol{s}_i, \boldsymbol{x}_i$
Set $b_i^{(0)} = 0$
For $t = 1, 2, \cdots, I$
  Emulate a similar energy trading scenario for N MGs
  For $k = 1, 2, \cdots, K$
    Observe $\hat{g}_i^{(k)}$ and $d_i^{(k)}$
    Obtain state $\boldsymbol{s}_i^{(k)} = \left[d_i^{(k)}, \hat{g}_i^{(k)}, b_i^{(k-1)}\right]$
    Choose $\boldsymbol{x}_i^{(k)} \in \boldsymbol{X}$ via Eq. (22)
    For $j = 1, 2, \cdots, N$
      If $j \neq i$
        Negotiate with MG j to obtain $y_{ij}^{(k)}$ via (2)
        Sell or purchase $|y_{ij}^{(k)}|$ amount of the energy to or from MG j
      Else
        Calculate $y_{ii}^{(k)}$ via (3)
        Sell or purchase $|y_{ii}^{(k)}|$ amount of the energy to or from the power plant
      End if
    End for
    Obtain $u_i^{(k)}$
    Observe $b_i^{(k)}$
    Calculate $Q_i^*\left(\boldsymbol{s}_i^{(k)}, \boldsymbol{x}_i^{(k)}\right)$ via (20)
    Calculate $V_i^*\left(\boldsymbol{s}_i^{(k)}\right)$ via (21)
  End for
End for

is chosen based on its quality function or Q-function denoted by $Q_i(\cdot)$, which describes the expected discounted long-term reward for each state-action pair. The state observed by MG i at time slot k, denoted by $\boldsymbol{s}_i^{(k)}$, consists of the current local energy demand, the estimated amount of the renewable energy generated at time k and the previous battery level of the MG, i.e., $\boldsymbol{s}_i^{(k)} = \left[d_i^{(k)}, \hat{g}_i^{(k)}, b_i^{(k-1)}\right]$.

The value function $V_i(\boldsymbol{s})$ is the maximal Q function over the feasible actions at state $\boldsymbol{s}$. The Q function and the value function of MG i are updated, respectively, by the following:

$$Q_i\left(\boldsymbol{s}_i^{(k)}, \boldsymbol{x}_i^{(k)}\right) \leftarrow (1-\alpha) Q_i\left(\boldsymbol{s}_i^{(k)}, \boldsymbol{x}_i^{(k)}\right) + \alpha\left(u_i^{(k)} + \gamma V_i\left(\boldsymbol{s}_i^{(k+1)}\right)\right) \tag{20}$$

$$V_i\left(\boldsymbol{s}_i^{(k)}\right) = \max_{\boldsymbol{x} \in \boldsymbol{X}} Q_i\left(\boldsymbol{s}_i^{(k)}, \boldsymbol{x}\right), \tag{21}$$

where $\alpha \in (0, 1]$ is the learning rate representing the weight of current experience in the learning process, and the discount factor $\gamma \in [0, 1]$ indicates the uncertainty of the microgrid regarding the future utility.

Algorithm 2 Hotbooting Q-learning based energy trading of MG i.

Initialize α, γ, Q_i=Q_i^*, and V_i=V_i^*
Set $b_i^{(0)} = 0$
For $k = 1, 2, \cdots$
  Estimate $\hat{g}_i^{(k)}$ and $d_i^{(k)}$
  Obtain state $\boldsymbol{s}_i^{(k)} = \left[d_i^{(k)}, \hat{g}_i^{(k)}, b_i^{(k-1)}\right]$
  Select the trading strategy $\boldsymbol{x}_i^{(k)}$ via Eq. (22)
  For $k = 1, 2, \cdots, K$
    If $j \neq i$
      Negotiate with MG j to obtain $y_{ij}^{(k)}$ via (2)
      Sell or purchase $|y_{ij}^{(k)}|$ amount of the energy to or from MG j
    Else
      Calculate $y_{ii}^{(k)}$ via (3)
      Sell or purchase $|y_{ii}^{(k)}|$ amount of the energy to or from the power plant
    End if
  End for
  Obtain $u_i^{(k)}$
  Observe $b_i^{(k)}$
  Update $Q_i\left(\boldsymbol{s}_i^{(k)}, \boldsymbol{x}_i^{(k)}\right)$ via Eq. (20)
  Update $V_i\left(\boldsymbol{s}_i^{(k)}\right)$ via Eq. (21)
End for

The standard Q-learning algorithm initializes the Q-function with an all-zero matrix, which is usually not the optimal value and thus degrades the learning performance at the beginning. Therefore, we design a hotbooting technique to initialize the Q-value based on the training data obtained from the large-scale experiments performed in similar smart grid scenarios. This saves the random explorations at the beginning of the game and thus accelerates the convergence rate. More specifically, we perform I similar energy trading experiments before the start of the game, as shown in Algorithm 1.

To balance the exploitation and exploration in the learning process, an ϵ-greedy policy is applied to choose the amount of the energy to trade with other MGs and the energy plant, i.e., $\boldsymbol{x}_i^{(k)}$ is given by

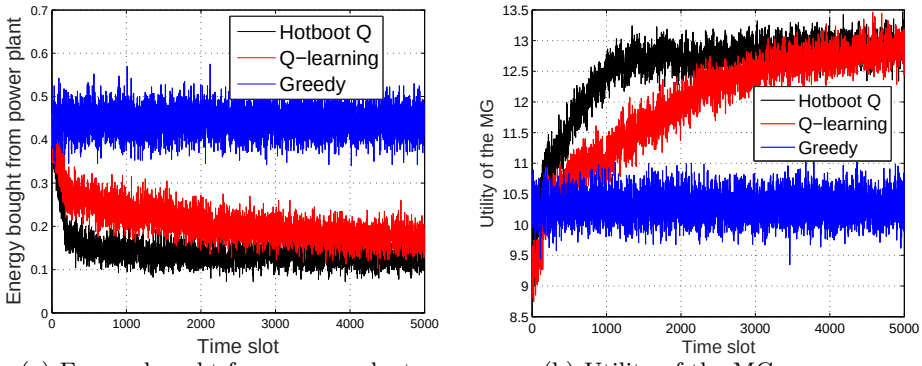
$$\Pr(\boldsymbol{x}_i^{(k)} = \boldsymbol{\Theta}) = \begin{cases} 1-\epsilon, & \boldsymbol{\Theta} = \arg\max\limits_{\boldsymbol{x}\in\boldsymbol{X}} Q_i\left(\boldsymbol{s}_i^{(k)}, \boldsymbol{x}\right) \\ \frac{\epsilon}{|\boldsymbol{X}|}, & \text{o.w.} \end{cases} \tag{22}$$

MG i chooses $\boldsymbol{x}_i^{(k)}$ according to ϵ-greedy strategy and negotiates with other MGs to determine the actual amounts of the energy in the trading $\boldsymbol{y}_i^k$ according to (2). As shown in Algorithm 2, the MG observes the reward and the next state.

According to the resulting utility $u_i^{(k)}$, the MG updates its Q function via (20) and (21).

5 Simulation Results

Simulations have been performed to evaluate the performance of the hotbooting Q-learning based energy trading strategy in the dynamic game with $N = 2$ MGs. In the simulation, if not specified otherwise, the energy storage capacity of each MG is $B = 4$, and the energy gain is $\beta = 8$. The local energy demands, the energy trading prices, and the renewable energy generation models of each MG in the simulations are retrieved from the energy data of microgrids in Hong kong in [13]. As benchmarks, we consider the Q-learning based trading scheme and the greedy scheme, in which each MG chooses the amount of selling/buying energy according to its current battery level to maximize its estimated immediate utility.



(a) Energy bought from power plant (b) Utility of the MG

Fig. 1. Performance of the energy trading strategies in the dynamic game with $N = 2$, $B = 4$ and $\beta = 8$.

As shown in Fig. 1, the proposed Q-learning based energy trading strategy outperforms the greedy strategy with less energy bought from the power plant and a higher utility of the MG. For example, the Q-learning based strategy decreases the average amount of the energy purchased from the power plant by 47.7% and increases the utility of the MG by 11.6% compared with the greedy strategy at the 1500-th time slot in the game. The performance of the Q-learning based strategy is further improved with the hotbooting technique that exploits similar energy trading experiences to accelerate the learning speed. As shown in Fig. 1, the hotbooting Q-learning based energy trading strategy decreases the amount of the energy purchased from the power plant by 33.7% and increases

the utility of the MG by 9.5% compared with the Q-learning based strategy at the 1500-th time slot.

6 Conclusion

In this paper, we have formulated an MG energy trading game for smart grids and derived the NE of the game, disclosing the conditions under which the MGs in a smart grid trade with each other and reduce the dependence on the power plant. A Q-learning based energy trading strategy has been proposed for each MG to choose the amounts of the energy to trade with other MGs and the power plant in the dynamic game with time-varying renewable energy generations and power demands. The learning speed is further improved by the hotbooting Q-learning technique. Simulation results show that the proposed hotbooting Q-learning based energy trading technique improves the utility of MG and reduces the amount of the energy purchased from the power plant, compared with the benchmark strategy.

References

1. Amin, S.M., Wollenberg, B.F.: Toward a Smart Grid: Power Delivery for the 21st Century. IEEE Trans. Smart Grid, vol. 3, no. 5, pp. 34–41 (2005)
2. Farhangi, H.: The Path of the Smart Grid. IEEE Power and Energy Magazine, vol 8, no. 1, pp. 18–28 (2010)
3. Baeyens, E., Bitar, E., Khargonekar, P.P., Poolla, K.: Wind Energy Aggregation: A Coalitional Game Approach. Decision and Control and European Control Conference, pp. 3000-3007 (2011)
4. Maharjan, S., Zhu, Q., Zhang, Y., Gjessing, S., Basar, T.: Dependable Demand Response Management in the Smart Grid: A Stackelberg Game Approach. IEEE Trans. Smart Grid, vol. 4, no. 1, pp. 120–132 (2013)
5. Wang, Y., Saad, W., Han, Z., Poor, H.V., Basar, T.: A Game-theoretic Approach to Energy Trading in the Smart Grid. IEEE Trans. Smart Grid, vol. 5, no. 3, pp. 1439–1450 (2014)
6. Tushar, W., Chai, B., Yuen, C., et al.: Three-party Energy Management with Distributed Energy Resources in Smart Grid. IEEE Trans. Industrial Electronics, vol. 62, no. 4, pp. 2487–2498 (2015)
7. Xiao, L., Mandayam, N.B., Poor, H.V.: Prospect Theoretic Analysis of Energy Exchange Among Microgrids. IEEE Trans. Smart Grid, vol. 6, no. 1, pp. 63–72 (2015)
8. Xiao, L., Chen, Y., Liu, K.R.: Anti-cheating Prosumer Energy Exchange based on Indirect Reciprocity. In: IEEE Int'l Conf. Commun., pp. 599C604. Sydney (2014)
9. Guan, C., Wang, Y., Lin, X., Nazarian, S., Pedram, M.: Reinforcement Learning-based Control of Residential Energy Storage Systems for Electric Bill Minimization. In: IEEE Consumer Commun. and Netw. Conf., pp. 637–642. Las Vegas (2015)
10. Qiu, X., Nguyen, T.A., Crow, M.L.: Heterogeneous energy storage optimization for microgrids. IEEE Trans. Smart Grid, vol. 7, no. 4, pp. 1453–1461 (2016)
11. Dalal, G., Gilboa, E., Mannor, S.: Hierarchical Decision Making in Electricity Grid Management. In: Int'l Conf. Machine Learning, pp. 2197–2206. New York (2016)

12. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement Learning: A Survey. J. Artificial Intelligence Research, vol. 4, pp. 237–285 (1966)
13. Wang, H., Huang, J.: Incentivizing Energy Trading for Interconnected Microgrids. IEEE Trans. Smart Grid (2016)