Global Energy Interconnection
Development and Cooperation Organization

ELSEVIER
Production And Hosting By Elsevier On Behalf Of KeAi
KeAi

Volume 3 Number 3 June 2020 (227-236)
DOI: 10.14171/j.2096-5117.gei.2020.03.004

Global Energy Interconnection

Contents lists available at ScienceDirect
https: //www.sciencedirect.com/journal/global-energy-interconnection

Full-length article

Nash-Q learning-based collaborative dispatch strategy for interconnected power systems

Ran Li, Yi Han, Tao Ma, Huilan Liu
State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Lianchi District, Baoding 071003, P.R. China

Scan for more details

Abstract: The large-scale utilization and sharing of renewable energy in interconnected systems is crucial for realizing "instrumented, interconnected, and intelligent" power grids. The traditional optimal dispatch method can not coordinate the economic benefits of all the stakeholders from multiple regions of the transmission network, comprehensively. Hence, this study proposes a large-scale wind-power coordinated consumption strategy based on the Nash-Q method and establishes an economic dispatch model for interconnected systems considering the uncertainty of wind power, with optimal wind-power consumption as the objective for redistributing the shared benefits between regions. Initially, based on the equivalent cost of the interests of stakeholders from different regions, the state decision models are respectively constructed, and the noncooperative game Nash equilibrium model is established. The Q-learning algorithm is then introduced for high-dimension decision variables in the game model, and the dispatch solution methods for interconnected systems are presented, integrating the noncooperative game Nash equilibrium and Q-learning algorithm. Finally, the proposed method is verified through the modified IEEE 39-bus interconnection system, and it is established that this method achieves reasonable distribution of interests between regions and promotes large-scale consumption of wind power.

Keywords: Interconnected region, Noncooperative game, Q-learning, Wind power accommodation.

1 Introduction

Wind energy has become one of the most promising new energy sources in the context of global energy interconnection. In China, as the distance between the wind power resources and load center is considerable, large-scale wind-power grid integration leads to shortage of backup resources in the regional power grid, rendering complete consumption of the wind power difficult on the spot. Besides, the fluctuation and anti-peak characteristics lead to insufficient system peak regulation capacity, resulting in serious wind-power curtailment [1]. Therefore, the maximization of wind-power consumption during grid dispatch has become a research hotspot [2–4].

With the development and expansion of the power system, "instrumentation, interconnection, and intelligence" is the development direction of the power grid, and each isolated power system will inevitably become interconnected through the tie-line [5]. Furthermore, the

Received: 9 January 2020/ Accepted: 3 April 2020/ Published: 25 June 2020

✉ Huilan Liu
liuhuilan111@163.com

Yi Han
m18636256181@163.com

Ran Li
liranlelele@163.com

Tao Ma
13066050371@163.com

concept of cross-regional allocation of new energy between nations has been proposed [6–8]. With the traditional centralized algorithm, it is difficult to effectively solve the problems of hierarchical and partitioned dispatch in an interconnected power grid. Hence, robust optimization [9] has been adopted to describe wind-power output, and a decentralized and coordinated scheduling model based on the objective cascade analysis method has been proposed for the decentralized and coordinated dispatch of an interconnected power system. Moreover, the synchronous alternating direction multiplier method has been applied to generate the optimization model for each region, and the large-scale consumption of wind-power dispatch involving energy storage has been considered [10]. A cut-plane consistency algorithm has been proposed to decentralize the economic dispatch models in interconnected areas [11]; however, timely deletion of the nonfunctional cut-planes during the iteration process is necessary. To speed-up the convergence speed, the Ward equivalence method has been applied for decomposing interconnected systems, and the objective function has been processed using the modified generalized Benders decomposition method [12]. In addition, a distributed algorithm has been proposed based on Newton's method to solve the problem of multiregional economic scheduling [13], which can reduce the iteration time and effectively increase the calculation efficiency.

The aforementioned studies on interconnected system scheduling mainly focus on problems such as large-scale complex constraints and the difficulty in combining wind-power interconnected power grids with security constraints; hence, they mainly adopt mathematical methods for objective function optimization. In most available studies, in case of multiregional participation in dispatch, the regions are regarded as economic interest communities with the same objective function in pursuit of overall maximum interest, unreasonably distributing the benefits and even sacrificing the interests of certain regions. The coordinated dispatch of interconnected systems is a process in which two regions pursue their own interests to maximize not only the economic interests but also the overall social and economic interests. Therefore, game theory [14–16] can be used to analyze the respective interests of the interconnected regions and the Nash equilibrium that the two may achieve. Additionally, the aforementioned studies optimize the power output of the unit through mathematical models based on the load and wind-power forecast data at each moment, in isolation, ignoring the fact that power system dispatch is a continuous and repeated process. For each process, the uncertainty prediction and wind-power dispatch can be based on experience. Therefore, reinforcement learning [17–19] provides a new solution for using such experience to optimize the power output of the unit more efficiently, maximize wind-power consumption, and complete the game between regions based on system experience and conditions. With online learning ability, the wind-power consumption capacity of an integrated wind storage system has been strengthened [20]. For coordination and dispatch in an integrated energy microgrid, a dispatch model has been constructed based on the multiagent game and Q-learning algorithm to maximize the operating incomes and reduce the costs of all the parties [21].

This study proposes a large-scale wind-power coordinated consumption strategy to realize the optimal operation of interconnected systems. The main contributions of this study are summarized as follows:

• A coordinated economic dispatch model based on the noncooperative game Nash equilibrium, considering the interconnected wind-power uncertainty, is established. The interactions between the regions and tie-line for various operational cases are presented in detail.

• The Q-learning algorithm is used to optimize the state of two regions, separately, to determine the Nash equilibrium of the largest fleet of wind power, most economic unit combination, and tie-line power output strategies.

• By analyzing the modified IEEE 39-bus interconnection system, the Nash-Q learning method proposed in this study is compared with the traditional dispatch method, establishing that the proposed method achieves reasonable distribution of interests among regions and promotes large-scale consumption of wind power.

The reminder of this paper is organized as follows. Sections 2 and 3 introduce the detailed dispatch model and Nash-Q learning method, respectively. The performed case study and discussions are presented in Section 4. Section 5 highlights the conclusions of this study.

2 Coordinated dispatch model based on noncooperative game Nash equilibrium

2.1 Two-region economic optimization decision model

In interconnected system dispatch, for a certain area, the start-up/shut-down of conventional thermal power units can be arranged based on load data, wind-power forecast, and tie-line schedule to calculate the equivalent economic cost of each region, and that of the entire system, subsequently. In the dispatch process of an interconnected system, one region does not necessarily cause another region to suffer the same economic loss, while optimizing its equivalent

economic cost. The two interconnected parties may have strategies to achieve a common interest balance, i.e., to further emphasize the maximization of their own benefits from the perspective of cooperation; hence, the coordination of the economic interests of the two regions constitutes a noncooperative game Nash equilibrium model. This study focuses on improving and emphasizing the autonomy of the region itself, and the collaboration and cooperation between regions. The start-up/shut-down strategies are separately formulated in the two regions to realize coordination and cooperation between them through the adjustment of the tie-line power. The game factor is defined as the equivalent cost of each region, assuming that region I is a high-generation area for wind power, and region J is a wind-power consumption area; the following equations result:

$$f_I = \sum_{t=1}^{T}\left\{\sum_{n=1}^{N_G^I}[u_{n,t}^I F_n^I(P_{n,t}^{G,I}) + u_{n,t}^I(1-u_{n,t-1}^I)S_{n,t}^I] + \sum_{j=1}^{N_W} C_W \Delta P_{j,t}^W\right\} - \gamma\sum_{t=1}^{T} P_t^l \tag{1}$$

$$f_J = \sum_{t=1}^{T}\left\{\sum_{n=1}^{N_G^I}[u_{n,t}^J F_n^J(P_{n,t}^{G,J}) + u_{n,t}^J(1-u_{n,t-1}^J)S_{n,t}^J]\right\} + \gamma\sum_{t=1}^{T} P_t^l \tag{2}$$

$$S_{n,t} = \begin{cases} S_{n,hot} & T_{n,\min} \leqslant T_{n,t,off} \leqslant T_{n,\min} + T_{n,cold} \\ S_{n,cold} & T_{n,t,off} > T_{n,\min} + T_{n,cold} \end{cases} \tag{3}$$

$$F_n\left(P_{n,t}^G\right) = a_n P_{n,t}^{G\,2} + b_n P_{n,t}^G + c_n \tag{4}$$

$$\Delta P_{j,t}^W = P_{j,t}^W - P_{j,t}^{WS} \tag{5}$$

where f_I and f_J are the equivalent cost functions of the region I and region J game players, respectively; T is the number of dispatch periods; N_G^I and N_G^J are the total number of thermal power units in each region, respectively; N_W is the number of wind farms in region I; $u_{n,t}^I$ and $u_{n,t}^J$ are the start-up states of thermal power unit n in each period, which range from 0–1; F_n^I and F_n^J are the coal consumption function of each region, respectively; $P_{n,t}^{G,I}$ and $P_{n,t}^{G,J}$ are the output of thermal power unit n in each period; $S_{n,t}^I$ and $S_{n,t}^J$ are the starting costs of thermal power unit n in each period of each region; C_W is the penalty cost coefficient of abandoned wind; $\Delta P_{j,t}^W$ is the abandoned wind power of wind turbine j in each period; γ is the tie-line price; P_t^l is the power on the tie-line; $S_{n,hot}$ and $S_{n,clod}$ are the hot and cold start-up cost, respectively; $T_{n,t,off}$ is the continuous downtime of each period of unit n; $T_{n,min}$ is the minimum stop time; $T_{n,cold}$ is the cold start time; a_n, b_n, and c_n are the operating cost coefficients of unit n; $P_{j,t}^W$ is the predicted wind power of wind turbine j in each period; and $P_{j,t}^S$ is the wind power of wind turbine j paralleling in the grid in each period.

2.2 Constraints

The unit output constraints are as follows:

$$u_{n,t} P_n^{G\min} \leqslant P_{n,t}^G \leqslant u_{n,t} P_n^{G\max} \tag{6}$$

where $P_n^{G\min}$ and $P_n^{G\max}$ are the minimum and maximum technical outputs of unit n, respectively.

The unit ramp rate constraints are as follows:

$$P_{n,t-1}^G - R_{D,n} \leqslant P_{n,t}^G \leqslant P_{n,t-1}^G + R_{U,n} \tag{7}$$

where $R_{D,n}$ and $R_{U,n}$ are the up and down climbing rates of unit n, respectively.

The minimum start-off time constraints are as follows:

$$\begin{cases} \left(T_{n,t-1,off} - T_{n,off}\right)\left(U_{n,t-1} - U_{n,t}\right) \leqslant 0 \\ \left(T_{n,t-1,on} - T_{n,on}\right)\left(U_{n,t-1} - U_{n,t}\right) \leqslant 0 \end{cases} \tag{8}$$

where $T_{n,on}$ and $T_{n,off}$ are the minimum on and off time of unit n, respectively.

The constructed wind-power output uncertainty set is expressed as follows [22]:

$$\begin{cases} P_{j,t}^W = P_{j,t}^e + \left(\tau_{j,t}^+ - \tau_{j,t}^-\right) P_{j,t}^h \\ P_{j,t}^e = 0.5\left(P_{j,t}^{W\max} + P_{j,t}^{W\min}\right) \\ P_{j,t}^h = 0.5\left(P_{j,t}^{W\max} - P_{j,t}^{W\min}\right) \\ \tau_{j,t}^+ + \tau_{j,t}^- \leqslant 1 \\ \sum_{j=1}^{N_W}\left(\tau_{j,t}^+ + \tau_{j,t}^-\right) \leqslant \Gamma_S \\ \sum_{t=1}^{T}\left(\tau_{j,t}^+ + \tau_{j,t}^-\right) \leqslant \Gamma_T \end{cases} \tag{9}$$

where $P_{j,t}^{W\max}$ and $P_{j,t}^{W\min}$ are the maximum and minimum predicted outputs, respectively, of wind farm j in each period; $\tau_{j,t}^-$ and $\tau_{j,t}^+$ range from 0 to 1; Γ_T is the maximum number of times that the specific wind farm j is forced to reach the predicted boundary value within the dispatch period T ; and Γ_S is the maximum number of times that all the wind farms reach the predicted boundary value in each dispatch period t.

The power balance constraints are as follows:

$$\sum_{n=1}^{N_G} P_{n,t}^G + \sum_{j=1}^{N_W} P_{j,t}^W - \chi_t p_t^l = D_t \tag{10}$$

where χ_t is the power flow direction of the interconnected regions. If p_t^l is the same as the specified positive direction, $\chi_t = 1$; else, $\chi_t = -1$; and D_t is the predicted value of the load in the respective area.

The load demand constraints are as follows:

$$\sum_{n=1}^{N_G} P_{n,t}^G + \sum_{j=1}^{N_W} P_{n,t}^W - \chi_t p_t^l = (1+r) D_t \tag{11}$$

where r is the load reserve ratio of the respective area.

The transmission capacity constraints are as follows:

$$\begin{cases} p_t^l = 0 & P_{j,t}^W \leqslant \eta_t P_j^{\max} \\ p_t^l \neq 0 & P_{j,t}^W > \eta_t P_j^{\max} \\ 0 \leqslant p_t^l \leqslant p_l^{\max} \end{cases} \tag{12}$$

where $P_j^{\max}$ is the maximum wind-power prediction value in a dispatch period; η_t is the proportional control coefficient in the adjustment period, $0 \leqslant \eta_t \leqslant 1$; and $P_l^{\max}$ is the maximum limit of the transmission power on the tie-line.

The aforementioned interconnected system economic dispatch model, considering wind-power uncertainty, is a multivariate mixed-integer nonlinear optimization problem; due to computational limitations, it is difficult to determine the optimal solution for this model. In this study, the method presented in [23] is used to linearize the coal consumption and start-up cost function, as well as the nonlinear constraint variables in the constraint function; this problem is then transformed into a mixed-integer linear one.

2.3 Interarea Nash equilibrium

Regions I and J constitute a noncooperative game-based model in the process of pursuing their own best interests. The Nash equilibrium points achieved by each region independently are the optimal strategies of their own objective functions [24–25]. The following can be expressed as

$$G^* = g\left(I_g, A_g^*, f_g^*\right) \tag{13}$$

$$A_g^* = \left\{A_{I,t}^*, A_{J,t}^*, P_t^{l*}\right\} \tag{14}$$

$$\left(A_{I,t}^*, P_t^{l*}\right) = \arg\min f_I\left(A_{I,t}, A_{J,t}, P_t^l\right) \tag{15}$$

$$A_{J,t}^* = \arg\min f_J\left(A_{I,t}, A_{J,t}, P_t^l\right) \tag{16}$$

where G is the equilibrium point of the game; g is the game function; I_g is the individual involved in the game; f_g^* is the Nash equilibrium winning function of the players participating in the game, namely, the equivalent cost; A_g^* is the Nash equilibrium action strategy set of the game player; $A_{I,t}^*$ is the Nash equilibrium strategy for the start-up output of the units in region I; $A_{J,t}^*$ is the Nash equilibrium strategy for the start-up output of the units in region J; and P_t^{l*} is the Nash equilibrium strategy for the tie-line power values. Equations (15) and (16) represent a region's own optimal strategy, when the other selects the optimal strategy.

The process of solving the Nash equilibrium problem is described as follows:

Step 1: Input raw data, including the load prediction data, wind-power prediction data, and the various data and parameters required by objective functions f_I and f_J.

Step 2: The initial value of the Nash equilibrium solution $\left(A_{I,t}^0, A_{J,t}^0, P_t^{l0}\right)$ is given.

Step 3: Iterative search: The k-th round optimization result is solved according to its own equivalent cost objective function and the optimization results of the previous round, i.e.,

$$A_{I,t,k} = \arg\min f_I\left(A_{I,t,k-1}, A_{J,t,k-1}, p_{t,k-1}^l\right) .$$

$$p_{t,k}^l = \arg\min f_J\left(A_{I,t,k}, A_{J,t,k-1}, p_{t,k-1}^l\right) .$$

$$A_{J,t,k} = \arg\min f_J\left(A_{I,t,k}, A_{J,t,k-1}, p_{t,k}^l\right) .$$

Step 4: Determine whether the Nash equilibrium solution has been found. If the k-th round optimization result is consistent with the (k-1)-th round, i.e.,

$$\left(A_{I,t}^*, A_{J,t}^*, p_t^{l*}\right) = \left(A_{I,t,k}, A_{J,t,k}, p_{t,k}^l\right) = \left(A_{I,t,k-1}, A_{J,t,k-1}, p_{t,k-1}^l\right) .$$

If the equation is satisfied, the Nash equilibrium solution has been found, i.e., the Nash equilibrium equivalent cost f_I^* and f_J^* of each region, the tie-line power value P_t^{l*} at each moment, and the start-up output strategies $A_{I,t}^*$ and $A_{J,t}^*$ can be determined.

3 Interconnected system economic dispatch based on the Nash-Q learning method

3.1 Basic principle of the Q-learning algorithm

Reinforcement learning is a process of repeatedly learning and repetitively interacting with the environment to strengthen certain decisions. In this process, as the agent acts on the environment through action A, the environment changes and the reward value R is generated. The agent receives the reward value R, and selects the next action to be performed according to the enhanced signal and the current state S, with the goal of finding the optimal strategy to accomplish the target task. A typical reinforcement learning model is depicted in Fig. 1.

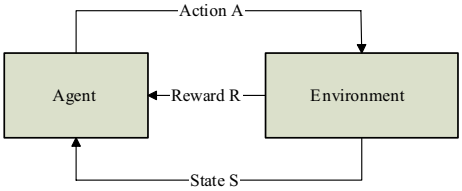


Fig. 1 Reinforcement learning model

Q-learning is a type of reinforcement learning, commonly based on the Markov decision process [26–28]. In this study, the Q-learning algorithm used in the Nash game results in more efficient commitment and dispatch decisions. It sets the previous empirical Q-value as the initial value of the subsequent iterative calculation, improving the convergence efficiency of the algorithm. The value function and iterative process of the Q-learning algorithm are expressed as follows:

$$Q(s,a)=R(s,s',a)+\beta\sum_{s'\in S}P(s'\mid s,a)\max_{a\in A_g}Q(s',a) \quad (17)$$

$$Q^{k+1}(s_k,a_k)=Q^k(s_k,a_k)+\alpha[R(s_k,s_{k+1},a_k)+\beta\max_{a'\in A_g}Q^k(s_{k+1},a')-Q^k(s_k,a_k)] \quad (18)$$

where s and $s'$are the current and next state, respectively; S is the state space set; β is the discount factor; $R(s,s',a)$ is the reward value obtained by performing action a from state s to state s' ; $P(s'\mid s,a)$ is the transition probability to state s' after performing action a in state s; $Q(s,a)$ is the Q-value of performing action a in state s; α is the learning factor; and $Q^k(s_k,a_k)$ is the k-th iteration value of the optimal value function Q^*.

3.2 Principle of the Nash-Q learning algorithm

In a noncooperative game, due to the constant change of state, the action strategy of the game players changes accordingly. At time t, player i iteratively learns to update the value of Q_t^i and equalizes the Q_t^j value of player j, in addition; a Nash game equilibrium $(Q_t^1(s),...,Q_t^n(s))$ is then formed in state s. By defining the noncooperative game Nash equilibrium solution as $(\pi^1(s),...,\pi^n(s))$, the updated iterative equation of the Nash-Q learning algorithm is provided as follows [29]:

$$Q_{k+1}^i(s,a^1,\ldots,a^n)=(1-\partial_k)Q_k^i(s,a^1,\ldots,a^n)+\partial_k\left[R_k^i(s,a^1,\ldots,a^n)+\beta NashQ_k^i(s')\right] \quad (19)$$

where $NashQ_k^i(s')=\pi^1(s'),...,\pi^n(s')Q_k^i(s')$ indicates the payoff of player i in state s' for the selected equilibrium; Q_{k+1}^i is the (k+1)-th iteration value in the Q-value function for player i; and R_k^i is the k-th iteration value in the immediate reward function for player i.

3.3 Selection of the state space and action strategy set

In general, there are two methods for implementing the Q-function: the neural network method and the lookup table [26]. In this study, the latter is used to implement Q-learning, where the utility or worth of each corresponding action-state is solely expressed by aquantified value (Q-value), which is determined by an action-value function. Therefore, it is necessary to first determine S and A.

In the objective function model, the state variable includes the load prediction value S_{Load} and the wind-power prediction value S_W in each period; the action variable set A includes the start-up outputs a_I and a_J of the thermal power units in regions I and J, respectively, and the tie-line power value a_l. Before generating the Q-value table, we first discretize the continuous state variable and action variable to form a (state, action) pair function. The state variables can be discretized by the following expression [30]:

$$\Delta P_i=\frac{P_i^{\max}-P_i^{\min}}{N_i} \quad (20)$$

where ΔP_i is the interval length of the i-th variable; $P_i^{\max}$ and $P_i^{\min}$ are the maximum and minimum values of each variable, respectively; and N_i is the interval number of the i-th variable. All the state variables can be discretized into interval forms using (20), and the state $S_k=\{S_{Load},S_W\}$ to which each region belongs can be uniquely determined. The action strategy variable needs to be discretized into a fixed value form, after which a set of action strategies $a_k=\{a_I, a_J, a_l\}$ can be uniquely determined according to the unit state and interval of the tie-line power value.

3.4 Coordinated dispatch based on Nash-Q learning

After state space S and action strategy A are determined, prelearning and online learning can be performed. Before reaching the optimal Q-value, prelearning accumulation of experience is necessary to generate a Q-table that approximates the optimal solution; online learning can then be performed to obtain the best action strategy [31].

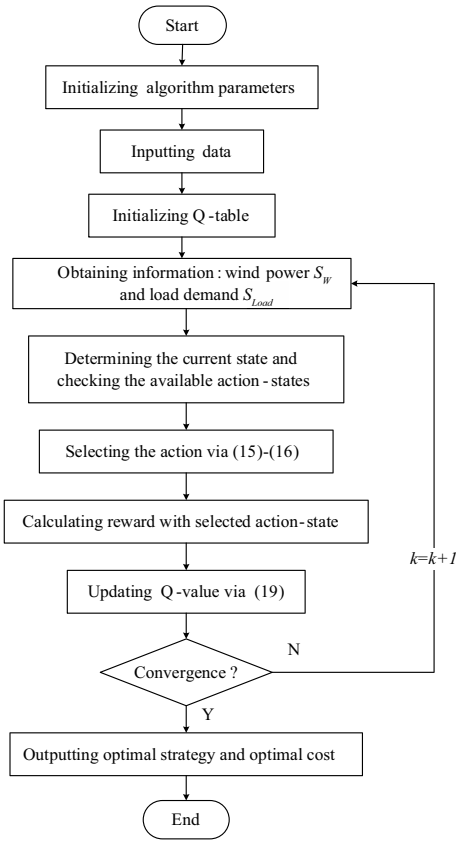


Fig. 2 Nash-Q learning flow chart

4 Analysis of examples

4.1 Description of the examples

In this study, the two-region economic dispatch model is solved through Yalmip programming in MATLAB; the Gurobi solver is applied, in addition. The modified IEEE 39-bus two-region system is used to verify the model depicted in Fig. A1. This interconnected system contains four wind turbines and 20 thermal power units. All the four wind turbines are located in region *I*. Fig. 3 displays the wind-power prediction value, and its upper and lower bounds. The parameters of the thermal power units and load demand data in each region are available in [32], and the upper limit of the tie-line power transmission is 500 MW. The penalty cost of abandoned wind is 100 $/MW, the electricity price of the tie-line γ=20 $/MW, Γ_S=4, and Γ_T=12.

With respect to parameter setting for the Q-learning algorithm, it is supposed that the learning factor α =0.01 and the discount factor β =0.8. For state space division, the load power is divided into 16 discrete spaces at intervals of 50 MW, and the wind-power output is divided into six discrete spaces at intervals of 50 MW. Therefore, corresponding to a 24-h dispatch period, regions *I* and *J* correspond to 2304 and 384 states, respectively. For action space division, the operation of the units is divided into two fixed states, start or stop, and the tie-line power value is divided into {0, 100, 200, 300, 400, 500} six fixed values. Therefore, regions *I* and *J* contain 6144 actions. Utilizing the annual historical load and wind-power data, the Nash-Q learning model is prelearned, and a Q-table approximating the optimal solution is established, which gives Q-learning a higher decision-making ability. The range of the regional cost is $ 5.47–6.23 million and $5.51–6.35 million, respectively, in regions *I* and *J* during the prelearning stage.

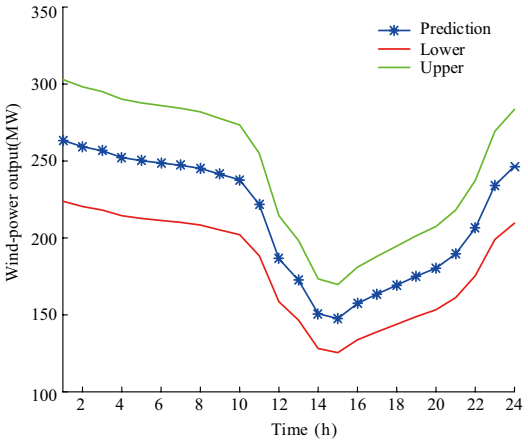


Fig. 3 Prediction value, and upper and lower bounds of the wind-power output

4.2 Impact of tie-line dispatch on the system based on Nash-Q learning

To better compare the analysis results, we designed the following three cases:

Case 1: The tie-line power $P_t^l = 0$ at any time, and the interconnected system is divided into two isolated systems in which economic dispatch is implemented, respectively.

Case 2: η_t =0, which allows the tie-line power to be adjusted at any time. The traditional method is used to solve the interconnected dispatch model.

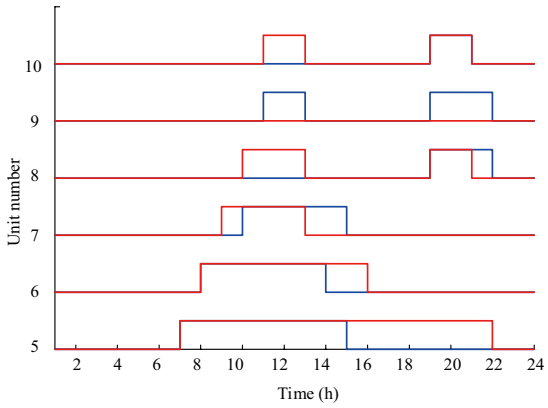
Case 3: η_t =0, and the Nash-Q learning method proposed in this study is applied to solve the interconnected dispatch model in which Q-learning has been prelearned.

The total cost of the three cases is shown in Table 1.

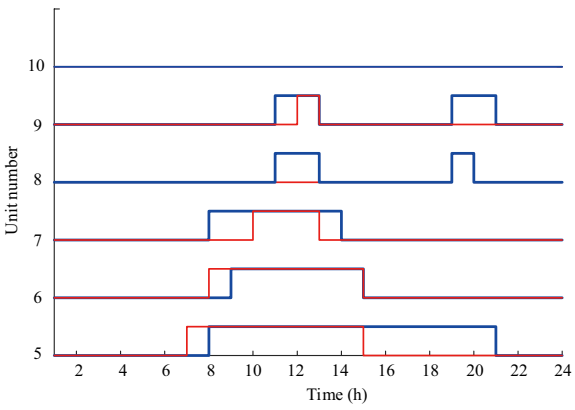
Table 1 Cost of each region for Cases 1–3

Case	f_I/$	f_J/$	Total costs/$
Case 1	562337	585293	1147630
Case 2	551394	571358	1122752
Case 3	555391	569163	1124554

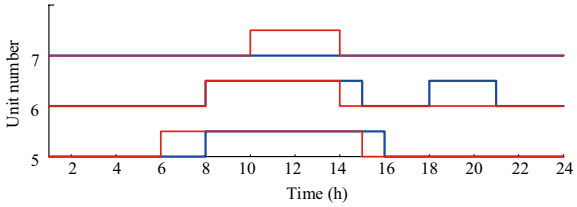
1) Analysis of Cases 3 and 1



(a) Start-up status of the units in Case 1



(b) Start-up status of the units in Case 3



(c) Start-up status of the units in Case 2

region I region J

Fig. 4 Unit combinations in different cases

In Cases 3 and 1, units 1–4 in both regions are always in the start-up group at any time, and the states of units 5–10 are different, as shown in Figs. 4(a) and 4(b). The total system dispatch cost in Case 3 is 1124554 $, which is 23076$ lesser than that of Case 1. It can also be observed that there is a decrease in the dispatch costs of both regions in Case 3. Comparing Figs. 4(a) and 4(b), the number of start-ups and running time in Case 3 are reduced in the two regions, indicating that the interconnected system can not only reduce the operating cost but also promote the consumption of wind power.

2) Analysis of Cases 3 and 2

In Cases3 and 2, the start-up/shut-down of the units in the two regions differ for units 5–7, as depicted in Figs. 4(b) and 4(c). Therefore, on comparing the start-up/shut-down status and the equivalent cost of each region between Cases 2 and 3, it can be observed that the number of start-ups in region *I* and the run-time of the units increase, and the number of start-ups in region *J* and the run-time of the units reduce in Case 3; correspondingly, the equivalent cost of region *J* decreases, whereas that of the region *I* increases. This demonstrates that Nash-Q learning can redistribute the economic benefits of the two regions again through iterative solutions to enable more wind-power consumption.

The transmission power of the tie-line in Cases 3 and 2 and the wind power dispatch are displayed in Fig. 5, where the total system dispatching cost in Case 2 is 1122752 $, and the total cost in Case 3 is 0.16% more than that of Case 2. From Fig. 5, it can be calculated that the wind power consumption in Case 3 increases by 1.76% compared to Case 2. Moreover, with respect to Case 2, there is a decrease in the tie-line power value in the peak load periods of 10–13h and 20–21 h in Case 3; the tie-line power value reduces in the two low-load periods of 1–5h and 23–24 h. In both cases, the wind power can be completely consumed in the 6–22h period. In the two load periods of 1–5h and 23–24 h, there is wind-power curtailment in both cases; however, the wind-power consumption capacity in Case 3 is higher. This demonstrates that Nash-Q learning can alleviate the peak pressure in the peak load period by coordinating the interests between the two regions, ensuring the benefit of the wind-power receiving end region *J* in the low-load period, and promoting the consumption of wind power, but at the expense of the overall economics of the system.

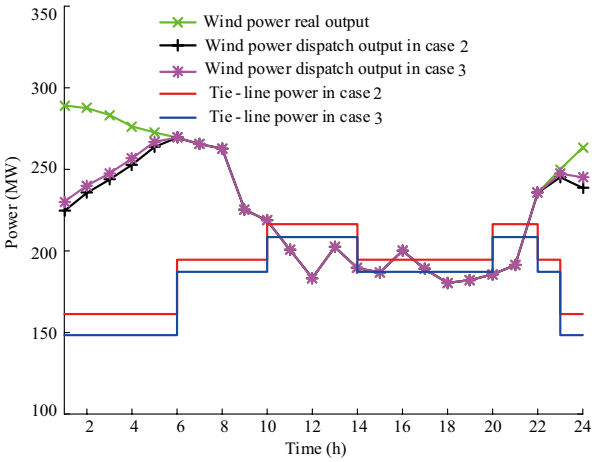


Fig. 5 Comparison of the tie-line power and wind-power dispatch output between Cases 2 and 3

4.3 Influence of γ and η_t on the economics of interconnected systems

This study analyzes the influence of the control coefficient η_t on the wind-power adjustment period and the tie-line electricity price γ on the economics. For η_t =0 and γ ranging from 0–25 $/MW, the equivalent and total cost of each region are listed in Table 2. It can be seen that as the tie-line electricity price γ increases, the equivalent cost of region *J* gradually increases compared to region *I*; correspondingly, there is a deterioration in the overall economy and wind-power consumption. For γ =20 $/MW and η_t ranging from 0–1, the changes in the total cost are presented in Table 3. It can be seen that with the increase in η_t, the time-period in which the wind power can be shared between the regions decreases, deteriorating the wind-power consumption and the overall economy.

Table 2 Influence of γ on the economic dispatch of interconnected systems

γ	f_I/$	f_J/$	Total costs/$
0	557844	565654	1123498
5	556789	567139	1123928
10	555713	568555	1124268
15	555539	568762	1124301
20	555391	569163	1124554
25	552083	573735	1125818

Table 3 Influence of η_t on the economic dispatch of interconnected systems

η_t	0	0.3	0.6	0.9	1
Total costs/$	1124554	1125280	1126583	1127824	1128030

4.4 Algorithm performance comparison

The discrete particle swarm optimization (DPSO) algorithm is used to solve the optimal dispatch strategy by optimizing the unit output to obtain the equivalent cost of each region. For the overall game process, the Nash equilibrium point is solved using the iterative search method. The solution procedures are different from those of the Q-learning method. The same time section is used for 10 repeated calculations in both algorithms. Four indicators are used for comparison, including the mean value of the objective function, variance (D), standard deviation (SD), and relative standard deviation (RSD) of each algorithm, and the results are shown in Table 4. The stability of the solution method proposed in this study is slightly better than that of the DPSO, and the objective function value based on particle swarm optimization has a relatively uniform distribution, increasing the uncertainty of the result.

Table 4 Optimization results of the two algorithms

Algorithm	Mean value of objective function /$	D	SD	RSD
DPSO	1138281	0.2964	0.5444	0.0510
Q-learning	1124393	0.0102	0.1427	0.0135

The iterative convergence process is illustrated in Fig.6. The DPSO algorithm converges to the optimal value with 1,200 iterations, whereas the Q-learning algorithm after prelearning can converge to the optimal cost value in approximately 600 iterations. Besides, the initial total cost of Q-learning is approximately 30% less than the DPSO, and it is also more economical after convergence. This shows that after prelearning, the Q-learning algorithm has reasonable ability to make an optimal decision based on the learned and accumulated experience; hence, its initial solution is close to the optimal value. Compared to the DPSO heuristic algorithm, the Q-learning algorithm has a more efficient solution speed and optimal decision solution for multivariate high-dimensional complex scheduling optimization problems.

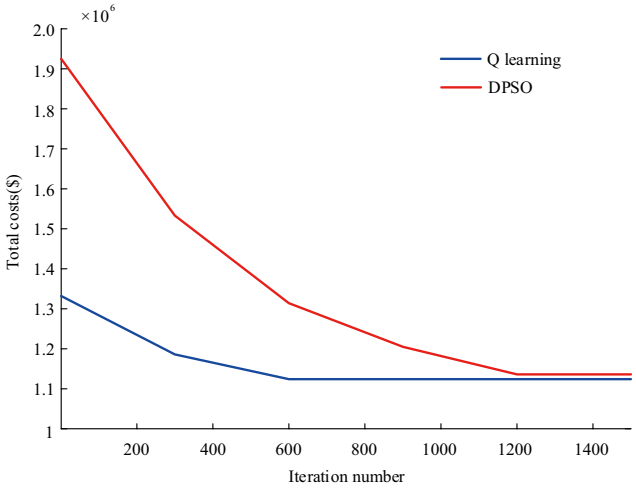


Fig. 6 Comparison of the iterative convergence between the Q-learning and DPSO algorithms

5 Conclusions

In this study, a coordinated economic dispatch model, considering the interconnected wind-power uncertainty, was presented, which integrates and exploits the synergy between game theory and reinforcement learning algorithms. The rationality and feasibility of this model in dispatch decision-making was analyzed in detail and discussed. After verification through a calculation example, the following conclusions were drawn:

1) The economic dispatch of interconnected systems based on the Nash-Q learning algorithm can not only effectively deal with the uncertainty of the wind-power output and improve wind-power consumption but also redistribute the shared benefits between regions.

2) The power line parameter γ of the tie-line has significant effect on the dispatch cost of the interconnected system. The larger the value of γ, the higher is the cost of purchasing wind power from the sending end, and the lower is the wind power consumption. The larger the value of η_t, the lesser is the wind power shared between the interconnected systems in each period, and the overall economics worsen.

3) The pre-learning Q-learning algorithm has better convergence and computational efficiency than the intelligent algorithm.

Further research can focus on improving the generalization of the proposed method and online transfer learning can be used for application to various scenarios.

Acknowledgements

This work is supported by the Fundamental Research

Funds For the Central Universities (No. 2017MS093).

References

[1] Shu Y, Zhang Z, Guo J, Zhang Z (2017) Study on key factors and solution of renewable energy accommodation. Proceedings of the CSEE 37(01):1-9

[2] Cui X, Zou C, Wang H, Zhou B (2018) Source and load coordinative optimal dispatching of combined heat and power system considering wind power accommodation. Electric Power Automation Equipment 38(07):74-81

[3] Zhang Q, Wang X, Yang T, Ren J, Zhang X (2017) A robust dispatch method for power grid with wind farms.Power Syst Technol 41(5):1451-1459

[4] Liu W, Wen J, Xie C, Wang W, Liu F (2014) Multi-objective fuzzy optimal dispatch based on source-load interaction for power system with wind farm. Electric Power Automation Equipment 34(10):56-63

[5] Fu Z, Li X, Yuan Y (2019) Research on technologies of ubiquitous power internet of things. Electric Power Construction 40(05):1-12

[6] Han J, Yi G, Xu P, Li J (2019) Study of future power interconnection scheme in ASEAN. Global Energy Interconnection 2(6):548-558

[7] Kåberger T (2018) Progress of renewable electricity replacingfossil fuels. Global Energy Interconnection 1(1):48-52

[8] Voropai N, Podkovalnikov S, Chudinova L, Letova K (2019) Development of electric power cooperation in Northeast Asia. Global Energy Interconnection 2(1):1-6

[9] Ren J, Xu Y (2018) Decentralized coordinated scheduling model of interconnected power systems considering wind power uncertainty.Automation of Electric Power Systems 42(16): 41-47+160+201-205

[10] Ren J, Xu Y, Dong S (2018) A decentralized scheduling model with energy storage participation for interconnected power system with high wind power penetration. Power Syst Technol 42(4):1079-1086

[11] Zhao W, Liu M, Zhu J, Li L (2016) Fully decentralised multi-area dynamic economic dispatch for large-scale power systems via cutting plane consensus. IET GENERTRANSMDIS 10(10): 2486-2495

[12] Li Z, Wu W, Zhang B, Wang B (2015) Decentralized multi-area dynamic economic dispatch using modified generalized benders decomposition. IEEE T Power Syst 31(1):526-538

[13] Lyu K, Tang H, Wang K, Tang B, Wu H (2019) Coordinated dispatching of source-grid-load for inter-regional power grid considering uncertainties of both source and load sides. Automation of Electric Power Systems 43(22):38-45

[14] Moris P (1994) Introduction to game theory. Springer-Verlag, New York

[15] Richard S, Andrew G(1998) Reinforcement Learning: An Introduction. MIT press, Cambridge

[16] Lu Q, Chen L, Mei S (2014) Typical applications and prospects of game theory in power system. Proceedings of the CSEE 34(29): 5009-5017

[17] Cheng L, Yu T, Zhang X, Yin L (2019) Machine learning for energy and electric power systems: state of the art and prospects. Automation of Electric Power Systems 43(1):15-43

[18] Yamada K,Takano S (2013) A Reinforcement learning approach using reliability for multi-agent systems. The Society of Instrument and Control Engineers 49(1):39-47

[19] Huang Q, Uchibe E, Doya K (2016) Emergence of communication among reinforcement learning agents under coordination environment. 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), Cergy-Pontoise,19-22 Sept 2016

[20] Liu G, Han X, Wang S, Yang M, Wang M (2016) Optimal decision-making in the cooperation of wind power and energy storage based on reinforcement learning algorithm. Power Syst Technol 40(9):2729-2736

[21] Liu H, Li J, Ge S, Zhang P, Chen X (2019) Coordinated scheduling of grid-connected integrated energy microgrid based on multi-agent game and reinforcement learning. Automation of Electric Power Systems 43(01): 40-50

[22] Wei W, Liu F, Mei S (2013) Robust and economical scheduling methodology for power systems:part two application examples. Automation of Electric Power Systems 37(18):60-67

[23] Carrion M, Arroyo JM (2006) A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. IEEE T Power Syst 21(3): 1371-1378

[24] Wang Y, Wang X, Shao C, Gong N (2020) Distributed energy trading for an integrated energy system and electric vehicle charging stations: A Nash bargaining game approach. Renew Energy 155:513-530

[25] Chuang A, Wu F, Varaiya P (2001) A game-theoretic model for generation expansion planning: problem formulation and numerical comparisons. IEEE T Power Syst 16(4):885-891

[26] Wu X, Tang Z, Xu Q, Zhou Y (2020) Q-learning algorithm based method for enhancing resiliency of integrated energy system. Electric Power Automation Equipment 40(04):146-152

[27] Amato C, Chowdhary G, Geramifard A, Ure N, Kochenderfer M (2013) Decentralized control of partially observable Markov decision processes. 52nd IEEE Conference on Decision and Control, Florence, Italy, 10-13 Dec2013

[28] Araabi B, Mastoureshgh S, Ahmadabadi M (2007) A study on expertise of agents and its effects on cooperative Q-learning. IEEE T Syst ManCy B 37:398-409

[29] Hu J, Michael P W (2003) Nash Q-learning for General-Sum Stochastic Game. J Mach Learn Res 4: 1039-1069

[30] Wu X, Tang Z,Xu Q,Zhou Y (2020) Q-learning algorithm based method for enhancing resiliency of integrated energy system. Electric power automation equipment40(04):146-152

[31] Liu J, Ke Z, Zhou W (2020) Energy dispatch strategy and control optimization of microgrid based on reinforcement learning. Journal of Beijing University of Posts and Telecommunications 43(01): 28-34

[32] Ongsakul W, Petcharaks N (2004) Unit commitment by enhanced adaptive Lagrangian relaxation. IEEE T Power Syst 19(1):620-628

Biographies

Ran Li received her bachelor degree in Electronic Engineering from North China Electric Power University, Baoding, China, in 1986, and master and Ph.D. degrees in Electronic Engineering in 1990 and 2009 respectively from the North China Electric Power University, Baoding, China. She is now a Professor with the School of North China Electric Power University, Baoding, China. Her main research interests focus on the power system analysis and dispatch of the new energy in the power system.

Yi Han received B.S. degree in Electronic Engineering, from North China Electric Power University, Baoding, China, in 2018 and is now working toward a master degree at North China Electric Power University, Baoding, China. Her current research interest includes the power system analysis and dispatch of the new energy in the power system.

Tao Ma received B.S. degree in Electronic Engineering, from Shandong Agricultural University, Taian, China, in 2017 and is now working toward a master degree at North China Electric Power University, Baoding, China. His current research interest includes the power system analysis and dispatch of the new energy in the power system.

Huilan Liu received master degree in Power System and Automation from North China Electric Power University in 2014. From 2014 to 2016, she was assistant engineer with Department of Electric Engineering, North China Electric Power University Baoding, where currently she is engineer. Her research interest is equipment life prediction and distributed energy storage technology.

(Editor Dawei Wang)

Appendix A

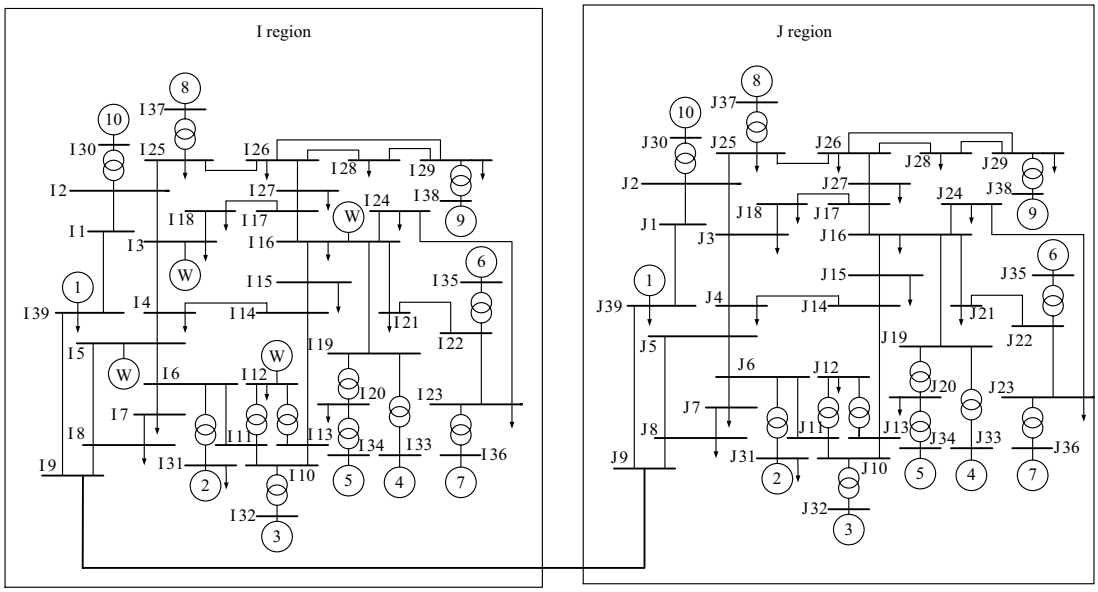


Fig. A1 Interconnection structure of the modified IEEE-39 node system