Research on hierarchical control and optimisation learning method of multi-energy microgrid considering multi-agent game

Hong Liu¹, Jifeng Li²,²,², Shaoyun Ge¹

¹Key Laboratory of Smart Grid, Tianjin University, Ministry of Education, Nankai District, Tianjin 300072, People’s Republic of China
²E-mail: lijifeng2014@163.com

Abstract: Due to the depletion of traditional fossil energy, to improve energy efficiency and build a cost-effective integrated energy system has become an inevitable choice. Aimed at the problem that the traditional centralised scheduling method is difficult to reflect the multi-dimensional interests of different agents in the multi-energy microgrid system, and the application of artificial intelligence technology in integrated energy scheduling still needs further exploration, this manuscript proposed a hierarchical control optimisation learning method with consideration of multi-agent game. Firstly, the multi-energy microgrid was taken as the research object, the microgrid system architecture was analysed, and the multi-agent partition in the system was pursued based on different economic interests. Secondly, for the technical aspects involved in the integrated energy regulation and management, the management layers of the multi-energy microgrid were divided, and the functions of different management layers were analysed. Based on this, the regulation functions were realised by considering the Nash Q-learning and the artificial intelligence method of Petri-net. Finally, the learning and decision-making ability of the method through practical cases were analysed, and the effectiveness and applicability of the proposed method were explained. This study explores the application of artificial intelligence technology in energy Internet energy management.

1 Introduction

Energy is the foundation of human survival and development and the source of social and economic development. Improving energy efficiency and ensuring effective supply of energy has become an inevitable choice to solve the contradiction between social development and environmental protection [1]. In recent years, with the continuous deepening of the construction of energy Internet, the existing modes of separate planning and independent operation of the traditional energy supply systems have been broken, and the coordinated supply of electricity/gas/heat (cold) and multi-energy has been realised. A integrated energy supply system with the goal of reliable, cost-effectiveness, clean and environmental protection will surely be the future development trend [2]. However, due to the inclusion of multiple energy supply carrier units, the composition of the integrated energy system will be more complicated, not only in physical equipment, but also in the transmission of data information. The massive multi-energy information poses new challenges to the rational decision making and management of system operations. Therefore, combined with the different management organisations in the system, on the basis of obeying the overall construction operation objectives, the hierarchical analysis of the integrated energy system can make the function of the control system more clear and the information processing efficiency higher.

In the process of building an integrated energy system, how to improve energy efficiency and comprehensively coordinate the use of multiple energy sources are key issues to be considered and an important basis for improving system efficiency. At present, the energy management and regulation control of the integrated energy system usually adopts the centralised control method [3, 4]. By mathematically modelling the energy efficiency and economy of the equipment in the system, the global optimisation target and the optimised iterative algorithm are used for scheduling. However, when facing multiple distributed energy sources, large amounts of control data, and flexible control methods in the integrated energy system, centralised management is difficult to achieve flexible and efficient deployment. At the same time, integrated energy systems usually contain multiple energy supply service providers with their own interests which makes it difficult to make concessions to the global optimisation goals of the overall system. Therefore, the feasibility of the above method remains to be studied. Based on the above considerations, many studies have introduced multi-agent [5] and game theory [6, 7] into energy management technology by using multiple agents autonomy, interactively and distributed computing features to reflect coordinated control between multiple stakeholders; use game theory to analyse the interests of different agents and the possible equilibrium of interests. However, at present, the division of the agent or game subject in the integrated energy system is usually based on the “source”/”load” as the boundary, and the division method is slightly rough. Therefore, the in-depth multi-agent interest game relationship still needs to be explored.

At the same time, considering the complexity of the integrated energy system, the traditional scheduling process is susceptible to personnel experience, although many of the scheduling tasks in the system have fixed processes, such as monitoring and alarming, which can be accomplished using traditional procedures and algorithms. However, there are still some jobs that require human experience and abstract thinking; at the same time, there are many control variables that need to be considered in the process of integrated energy regulation and management, also there are problems with peak and valley difference rates for different energy loads. The traditional numerical iterative solution method may be limited by the dimensional disaster problem to a certain extent, and the future regulatory requirements are shifted to the more refined real-time rolling optimisation trend. Therefore, how to use artificial intelligence (AI) technology to develop more intelligent control methods is an important technical problem that needs to be solved urgently [8]. Reinforced learning techniques in the field of AI do not require rigorous and accurate environmental models and system identification. For the stochastic optimal control problem, the closed-loop strategy is constructed by the interaction with the real environment or a set of trajectories obtained in the simulation. As long as sufficient pre-offline training is performed, the problem of slow analysis time caused by traditional recursive analysis methods can be solved. At present, the application of AI, especially learning algorithms, in energy systems mainly focuses on fault monitoring and diagnosis [9, 10], load/power generation prediction [11, 12]. However, in these studies, AI technology is only applied
to some part of the solution to the problem, the application is in the initial stage, and further research and application in planning and operation are still to be studied. Li et al. [13] takes the optimisation of power system operating cost as the goal, comprehensively considers the power balance, equipment constraint, electricity price trend and other information, and proposes a dynamic hierarchical reinforcement learning method based on the rate of change of key state variables, which solves the control problem of power system. Li and Jayaweera [14] proposed an implicit mode Markov decision process, based on the preliminary prediction results, it is effectively combined with the demand response strategy to provide guidance for users’ real-time energy consumption decision making. In [15], the multi-agent microgrid system is used as the research object. Through the deep learning method, the continuous state space of the equipment unit is effectively regulated, thereby improving the economics and reliability of the microgrid system operation. Kazmi et al. [16] proposed a constant temperature load control method based on multi-agent learning collaboration. By increasing the interactive application of knowledge and perception between agents, the learning efficiency of control strategy and the energy utilisation efficiency are improved.

The above studies have laid a certain foundation for the research in this manuscript. However, the current research has the following problems: firstly, the hierarchical analysis of the energy regulation of integrated energy systems and the application of AI to various management links remain to be explored. Secondly, at present, AI technology, especially reinforcement learning methods, are mostly applied in the field of single energy systems represented by electric energy, and it has been difficult to meet the multi-energy coupling cooperative scheduling requirements in the context of energy Internet. Thirdly, the more effective intelligent division and interest game relationship analysis in the multi-energy microgrid still needs to be further explored.

Therefore, in view of the above problems, this paper takes the multi-energy microgrid as the research object, and proposes a hierarchical control optimisation learning method considering multi-agent game. The main contributions are as follows:

(i) Introducing the idea of multi-agent and pursuing multi-agent partitioning and benefit function analysis based on the different interests existing in the multi-energy microgrid;
(ii) A hierarchical control method for multi-energy microgrid is proposed, so as to solve complex energy management problems clearly. In the control method, this paper adopts the Nash equilibrium Q-learning and Petri-net AI method to realise the regulation function, and the feasibility of the AI method in energy management scheduling decision alongside with the advantages of traditional heuristic algorithms are verified by practical cases.

The rest of this paper is organised as follows. Section 2 introduces the multi-energy microgrid architecture and multi-agent partitioning method; Section 3 introduces the functional architecture of the hierarchical control of multi-energy microgrid; Section 4 proposes the hierarchical control strategy based on AI methods; In Section 5, the effectiveness and practicability of the proposed method are verified by specific cases, and the algorithm is compared; Section 6 summarises this paper.

2 Architecture and multi-agent partitioning of multi-energy microgrid

As the key node of energy Internet, the multi-energy microgrid has received extensive attention due to its flexible operation. This paper selects the multi-energy microgrid as the research object and gives full play to its flexible and intelligent characteristics. The multi-energy microgrid is an autonomous energy system that consists of an energy management system, distributed energy resources, energy storage devices, energy conversion devices and diverse energy loads and in terms of structure it can be divided into energy input, conversion, storage and output. This paper combines the typical industrial park and energy bus concept [17] to build a multi-energy microgrid which consists of combined cold heat and power system (CCHP), gas heat pump (GHP), distributed photovoltaic (PV), central air-conditioning (CAC), electricity storage (ES), heat storage (HS) and other equipment, as well as a variety of energy sources such as electricity/gas/cold/heat. The structure of the multi-energy microgrid system is shown in Fig. 1.

On this basis, this paper takes the different interests pursued in the microgrid system as the reference and combines the actual situation to divide the multi-energy microgrid into the following agents: renewable energy provider (REP) is responsible for the specific control of PV and ES equipment in the multi-energy microgrid. Microgrid energy provider (MEP) is responsible for the control of CCHP system, GHP, CAC and HS equipment in the microgrid system. The electric vehicle owner (EVO) is responsible for the charge and discharge control of the EVs in the microgrid system. The terminal energy user (TEU) is responsible for transmitting the demand information of the electric/heat/cold load of the microgrid system and performing load reduction during the necessary time. Introducing the multi-agent concept into the multi-energy microgrid scheduling can on one hand better conform to the actual operation of the microgrid system and establish the basis for the subsequent game analysis; also on the other hand, to a certain extent, it can avoid the state or action space dimension disaster that may be brought about by the subsequent reinforcement learning method. Meanwhile, in the process of studying energy management issues, this paper divides the multi-energy microgrid into decision layer, operation layer and equipment layer. The functional objectives and control methods of different management layers will be described in detail later.

3 Function of hierarchical regulation and control in multi-energy microgrid

Based on the multi-energy microgrid framework introduced in Section 1, the hierarchical control and optimisation learning method of multi-energy microgrid considering multi-agent game in this paper mainly includes three functional management layers: decision layer, operation layer and equipment layer. The hierarchical control function of multi-agent is shown in Fig. 2. This section focuses on the functions of each layer.
3.1 Analysis on function of decision layer of multi-agent game optimisation

The decision layer of multi-energy microgrid is the key part to realise its own optimal control and intelligent coordination control with utility energy network. Considering that the actual multi-energy microgrid is composed of multi-agents, the stakeholders will consider their own economy more in the actual scheduling process, so the traditional centralised optimisation method is difficult to be implemented. Therefore, based on the agent partitioning method proposed in Section 1, this paper comprehensively considers REP, MEP and EVO, which are three agents with controllable resources, and obtains the Nash equilibrium point by analysing the game relationship and interest pursuit between the three agents. This is used as the basis for the decision layer to formulate the energy management and scheduling strategy of microgrid system. Among them, MEP agent is responsible for the supply of electric/heat/cold load. Apart from being part of the electric load supply in the microgrid system, the REP agent adjusts the strategy of charging and discharging for the ES equipment according to the price formulated by the utility power grid. At the same time, with reasonable price, REP can also choose to sell electricity directly to the utility power grid; EVO can choose whether or not to discharge the utility power grid according to the price set by the utility power grid.

MEP agent plays a relatively leading role in the microgrid system. By optimising the output of controllable resources during each period of time, the operation cost and the power purchasing cost from the utility power grid can be reduced; at the same time, when the power supply is insufficient or the electricity price is suitable, MEP is responsible for purchasing electricity from the utility power grid. The controlling goal of MEP is to minimise the total operating cost, the specific formula is

$$I_{MEP} = \min \sum_{t=1}^{N_T} \left[ (p_D(t)P_{ED}(t) + C_{CCHP}(t) + C_{GHP}(t)) + C_{OM}^{REP} + C_{OM}^{ST} \right]$$

(1)

where $N_T$ is the total number of scheduling periods; $p_D(t)$ is the power purchase cost of the microgrid system to the utility power grid during time period $t$; $P_{ED}(t)$ is the purchase power from the microgrid to the utility power grid during time period $t$; $C_{CCHP}(t)$ and $C_{GHP}(t)$ are the fuel cost of CCHP system and GHP during time period $t$; $C_{OM}^{REP}$ is the start and stop cost of controllable unit in MEP agent during time period $t$; $C_{OM}^{ST}$ is the operation and maintenance cost of MEP agent during time period $t$.

$$C_{OM}^{REP}(t) = \Delta t \sum_{\alpha = 1}^{N_{REP}} K_m P_m(t)$$

(2)

where $N_{REP}$ is the total number of devices in MEP agent; $K_m$ is the operation and maintenance cost coefficient of equipment $m$; $P_m(t)$ is the output of equipment $m$ during time period $t$. In addition, for the formula of fuel cost calculation for the operation of CCHP system and GHP, one can refer to [18].

Based on the different electricity prices in different periods, the REP agent reasonably arranges the feed-in power supply of the PV-ES system by adjusting the state of charging/discharging of ES device. The net economic benefit is the income of electricity sales deducting the operation and maintenance cost of all kinds of equipment. It can be expressed as

$$I_{REO} = \max \sum_{t=1}^{N_T} \left[ (p_D(t)P_{ED}(t) - C_{OM}^{REP}(t) - C_{OM}^{ST}(t)) \right]$$

(3)

where $p_D(t)$ is the feed-in price set by the utility power grid department during time period $t$; $P_{ED}(t)$ is the feed-in power of REP to the utility power grid during time period $t$; $C_{OM}^{REP}(t)$ is the operation and maintenance cost of REP agent during time period $t$; $C_{OM}^{ST}(t)$ is the depletion cost of ES device during time period $t$. For the calculation, one can refer to [18].

EVO agent, which is mainly aimed to profit, decides whether to sell power to the utility power grid according to the electricity price. For the EVO price sensitivity response model, the greater the difference is between the EV feed-in price and the EV scheduling cost, the higher the probability there is for EVO to respond to V2G scheduling. Using consumer psychology [19], the user price sensitivity response model of EV is established. The response characteristics of EVO can be fitted as follows:

$$\alpha = 1.431\Delta p - 6.962\Delta p^2 + 11.12\Delta p^3$$

$$- 6.096\Delta p^3 + 1.004\Delta p$$

(4)

where $\alpha$ is the probability of response scheduling of EVO, and $\Delta p$ is the difference between the grid price for EV and the normal electricity price. Based on the uncertainty of the responding probability, the EVO benefit function is

$$I_{EVO} = \max \sum_{t=1}^{N_T} \left[ (p_D(t)P_{EV}(t) \cdot (r - \alpha(\Delta p(t)))) \right]$$

(5)

where $P_{EV}(t)$ is the discharge power of EVO to the utility power grid during time period $t$, $r$ is a random number between 0 and 1, if $r$ is greater than $\alpha(\Delta p(t))$, the EVO agent choose to participate in the V2G scheduling; $\lceil \cdot \rceil$ is the round up function.

In the energy management of the decision layer, MEP and REP agents mainly undertake the supply tasks of different energy loads in the microgrid system, as well as achieving the optimisation of the balance between economic cost/benefit of the agent itself. On the other hand, REP and EVO agents have the ability to exchange power with the utility power grid in order to obtain benefits, but while protecting their own interests, they need to consider the power capacity constraints of grid-connected tie line transmission with the utility power grid. In the process of pursuing the best interests, each agent constitutes the static cooperative game with complete information. The cooperation agreement is that each agent should meet the normal and stable operation of the equipment and microgrid system. In this process, by formulating its own operation strategy, they optimise their own balance of costs and benefits. Therefore, from the above game relationship, it can be seen that the game problem at the decision layer is not a traditional unified optimisation problem, but a number of optimisation problems for each participant (or alliance) to optimise its own objectives independently. The Nash equilibrium point achieved is the optimisation result.

In order to solve the game problem mentioned above, this paper will use Q-learning algorithm combined with evolutionary game theory to solve the Nash equilibrium point of multi-agent game as the scheduling plan of decision layer. It can be expressed as

$$G = \left\{ N_g, A_g, u_g \right\}$$

(6)

where $N_g$ is the number of agents participating in the game, which is set as 3 in this paper; $A_g$ is the set of agents behaviour participating in the game; $u_g$ is the utility set of agents participating in the game, in this paper, the economic interests of each agent are represented by $I$.

3.2 Analysis on the function of operation layer for system operation constraint verification

Although the decision layer regulation can optimise the economic benefits of multi-agents in the microgrid system under the game environment, however, most of the considerations are the security and operation constraints of the equipment, which may ignore the protection of the overall operation security of the microgrid system. Therefore, on the basis of the cooperating with the distributed optimisation of the decision layer, the operation layer is
responsible for controlling the running state of the microgrid system in order to improve the cooperative game protocol of the decision layer. This paper focuses on verifying the running state of the microgrid system from three aspects: the first is the problem of power balance in the microgrid system. If it is not satisfied, it will give feedback to the decision layer in time, help the decision layer to adjust and accumulate learning experience. Different types of energy power balance formulas are shown in the following equations:

\[ P_{\text{E,i}}(t) + P_{\text{CHP}}(t) + P_{\text{PV}}(t) + P_{\text{L,i}}(t) = P_{\text{Load}}(t) \]  
\[ \eta_{\text{CHP}} P_{\text{CHP}}(t) + \eta_{\text{PV}} P_{\text{PV}}(t) + P_{\text{L,i}}(t) = P_{\text{Load}}(t) \]  
\[ P_{\text{CHP}}(t) + \eta_{\text{CHP}} P_{\text{CHP}}(t) + P_{\text{L,i}}(t) = P_{\text{Load}}(t) \]

where \( P_{\text{E,i}}(t) \) is the output of different units during time period \( t \); \( \eta_{\text{PV}} \) is the energy production/conversion efficiency of different units; \( P_{\text{Load}}(t) \) is the demand for different energy loads during time period \( t \). Energy storage equipment, \( \eta_{\text{PV}} \) indicates the charging and discharging status of the equipment, which is distinguished by positive and negative signs in the process of simulation.

The second is the energy flow constraint in the microgrid system, which can reasonably monitor the energy flow, evaluate the global operation state of the microgrid system and avoid the influence of multi-agent distributed optimisation on the global running state. At the same time, it also provides guidance for the learning experience of improving the operation economy and stability of microgrid system. Guided by the method in [1], this paper combines the operation signal commands transmitted from the decision layer to the equipment layer, the mixed energy flow of the whole microgrid system is calculated to realise the running state monitoring and energy supply of the microgrid system. The mathematical model for the calculation of mixed energy flow in the multi-energy microgrid is shown as

\[
\begin{align*}
    f_{\text{PS}}(x_e, x_e, x_e, x_{MG}) &= 0 \\
    f_{\text{HS}}(x_e, x_e, x_e, x_{MG}) &= 0 \\
    f_{\text{CS}}(x_e, x_e, x_e, x_{MG}) &= 0 \\
    f_{\text{MG}}(x_e, x_e, x_e, x_{MG}) &= 0
\end{align*}
\]

where \( f_{\text{PS}}, f_{\text{HS}}, f_{\text{CS}} \) and \( f_{\text{MG}} \) represent the state equations of energy flow in electric, heat, cold and global microgrid system, while \( x_e, x_e \) and \( x_e \) represent the relevant variables affecting energy flow in electric, heat, cold subsystems. \( x_{MG} \) is the energy use and distribution coefficient of microgrid system, which is obtained through the decision layer. Based on the decoupling solution method, the energy flow of the microgrid system is calculated by issuing the instruction and distribution information to the operation layer and the equipment layer, and whether the voltage amplitude, flow and other factors meet the operation constraints is checked. If in the situation of exceeding the limit or dissatisfaction, the adjustment instruction will be issued to the equipment layer in time, and the information will be fed back to the decision layer, so as to effectively avoid the situation that the multi-agent game at the decision layer may not satisfy the overall operation limit. In view of the adjustment problems involved in the constraint verification of the operation layer, the Petri-net method will be used to explain the regulation and control instruction logic of different agents in different situations.

The third is the transmission power constraint limit of grid-connected tie line between multi-energy microgrid and utility power grid, which has been explained in Section 3.1.

3.3 Analysis on the function of optimisation of multi-agent game on equipment layer

The equipment layer of the multi-agent hierarchical control system constructed in this paper includes all kinds of energy production/conversion devices, all kinds of loads, grid switches, control units and so on. The main functions are as follows:

(i) Equipment operation control. The device control unit receives instructions issued by the decision layer or operation layer of the microgrid system to set the operation state of the production/storage/conversion devices.

(ii) All kinds of load control. The control unit in the microgrid system will control the load according to the demand of the user, the importance of the load, the grade of energy, the constraint layer of energy flow and so on. If necessary, the load is switched according to the instructions of the operation layer.

(iii) Information transmission and interaction. The equipment layer is responsible for providing the running status of the components and loads in the system to the operation layer of the microgrid system, as well as receiving the instructions sent to the control unit by the decision layer and the operation layer.

4 Hierarchical regulation and control strategy of multi-energy microgrid

As the equipment layer is only responsible for the transmission of information and the actual control of the unit equipment in the microgrid system, therefore, this section mainly introduces the control strategy of the upper decision layer and the middle operation layer. The multi-agent hierarchical control diagram of the multi-energy microgrid is shown in Fig. 3.

4.1 Regulation and control strategy on decision layer based on Q-learning method

As mentioned in Section 3.1, the main task of the decision layer is to take the Nash equilibrium point as the optimisation result to manage the energy in the microgrid system, and to formulate the preliminary operation strategy on the premise of considering the multi-agent optimisation game factors. The multi-agent partition and interest pursuit function in the microgrid system have been introduced in the previous section. In this section, the energy scheduling problem of the multi-energy microgrid decision layer is modelled as a Markov decision process, and the evolutionary game theory is combined with the Q-learning algorithm in order to achieve the energy management of decision layer and the formulation of multi-agent operation strategy.

The Q-learning algorithm [20, 21] is one of the commonly used reinforcement learning algorithms, and it is also an online learning and dynamic optimisation technique based on value function iteration. The principle is to use the previous empirical \( \hat{Q} \) value table as the initial value of the subsequent iteration calculation, to shorten the convergence time of the algorithm. The value function and iterative process of the Q-learning algorithm are expressed as

\[
Q(s, a) = R(s, s', a) + \gamma \sum_{a' \in A} P(s'|s, a) \max_{a \in A} Q(s', a)
\]

\[
Q^{(k+1)}(s, a) = Q^k(s, a) + \alpha \left[ R(s, s', a) + \gamma \max_{a' \in A} Q^k(s', a') - Q^k(s, a) \right]
\]

where \( s \) and \( s' \) are the current state and the state of the next period; \( R(s, s', a) \) is the immediate reward function value obtained after the state \( s \) is transferred to the state \( s' \) by the action \( a \); \( \gamma \) is the discount factor; \( P(s'|s, a) \) is the probability that state \( s \) shifts to state \( s' \) after control action \( a \); \( Q^k \) is the \( k \)th iteration value of optimal value function \( Q^* \); \( \alpha \) is the learning factor, which represents the degree of trust in the updated part; \( Q(s, a) \) is the Q value when performing action \( a \) in state \( s \).

This paper further introduces evolutionary game theory into the Q-learning method [21]. The constant change of the state will make the strategy of the game agent update constantly in the evolutionary game. Since the historical combination of actions will...
be considered in each step of the state in the evolutionary game, the transfer of state for each game agent is also the transfer of its chosen action.

The game environment between multiple agents can be established based on (6): the participants are three agents that contain controllable resources in the microgrid system; the strategy set is the operation decision making of each agent on the energy supply or energy storage devices; the utility functions correspond to Formulas (1), (3) and (5). In the Nash Q-learning method, for the game participation agent $N_g$, if its action $a_i^g$ is the best reflection dynamic when the combination of action strategies of other agents are given, that is, all the agent participants can meet the best operating cost/benefit economy of $a_i^g \in A_g$ for each game, namely

$$u_i^g(a_i^g, a_{i-1}^g) \geq u_i^g(a_i^g, a_{i-1}^g)$$

(13)

It is said that Q-learning reaches the Nash equilibrium point, where $(a_i^g, a_{i-1}^g)$ is the Nash equilibrium solution, $a_{i-1}^g$ is the action strategy of other agents except agent $i$.

Under the environment of considering game theory, the expected total reward and updated value of the game participation agent $N_g$ can be expressed by the $Q$ value function as

$$Q_{i+1}^{N_g} = (s, a^1, \ldots, a^N) = R_{i+1}^{N_g} + \gamma \sum_{s'} P(s' | s, a^1, \ldots, a^N) Q_{i+1}^{N_g}(s', a', \ldots, a^N)$$

(14)

$$Q_{i+1}^{N_g} = (s, a^1, \ldots, a^N) = (1 - \alpha_i)Q_{i+1}^{N_g}(s', a', \ldots, a^N) + \alpha_i \left[ R_{i+1}^{N_g} + \gamma \sigma(s') \ldots \sigma(s') Q_{i+1}^{N_g}(s') \right]$$

(15)

where $Q_{i+1}^{N_g}$ is the $i$th iteration value of the $Q$ value function $Q_{i+1}^{N_g}$ of the agent $i$; $R_{i+1}^{N_g}$ is the $i$th iteration value of the immediate reward function $R_{i+1}^{N_g}$ of the agent $i$; $N_g$ is the number of agents participating in the game; $[a^1(s'), \ldots, a^N(s')]$ is a mixed strategy equilibrium solution of Nash equilibrium.

The control action needs to be selected according to the current state in the iterative process. Boltzmann probability distribution method [22] is used in this paper to describe the transition probability of the state in the evolutionary game. The Boltzmann probability distribution method selects the action by probability, and the probability of selecting the action $a_i$ in the state $s$ is

$$p(a_i) = \frac{e^{Q(s, a_i) / \lambda}}{\sum_{a \in A} e^{Q(s, a) / \lambda}}$$

(16)

where $\lambda$ is the exponential function of the iteration period-$\lambda$ in evolutionary game. When $\lambda$ increases, the agent's decision randomness also increases; and when $\lambda$ decreases, the decision randomness also decreases. It can be seen that the Boltzmann probability distribution method combined with the Q-learning algorithm has the ability of adaptive learning.

For the problem to be studied in this paper, the decision layer control process based on the Q-learning method is:

Step (1): Initialise the $Q$ value table. The initial value of each element $(s, a)$ in the $Q$ value table of the offline pre-learning stage is taken as 0. In the online learning stage, the initial value is transformed into the feasible $Q$ value table reserved by pre-learning.

Step (2): Discretise the continuous state and the action variable to form a (state, action) pair value function, generate a sample by Markov simulation, combine the multi-agent interest Nash equilibrium target of the decision layer, and select the current running state and action strategy $a' = \arg \max_{a' \in A} \{ N_g \cdot A_g \cdot a_g \}$.

In the selection of the state space, the actual value of the PV output/load demand of the microgrid system in each period, the output of CCHP system, GHP, CAC, the charge and discharge power of ES/HS equipment, the charge and discharge power of EVs are used as state input. The above variables are continuous variables. In order to cooperate with the Q-learning method, the
above variables are discretised into interval forms, and the lengths of the intervals can be expressed as

$$\Delta P_m = \frac{P_{m, \text{max}} - P_{m, \text{min}}}{M_m}$$  \hspace{1cm} (17)

The output of the \( m \)-type equipment is divided into \( M_m \) intervals according to the output characteristics, and combine it with the algorithm for subsequent learning, and the output constraint condition of the equipment unit is also guaranteed. The state space of the microgrid system during the time period is the combination of the PV power output, load demand and the state of the energy production/storage equipment. The only state can be determined according to the output of the unit in the microgrid system and the demand of the load \( S_k = (k, SPV, S_{\text{Load}}, SCCHP, S_{\text{GHP}}, S_{\text{CAC}}, S_{\text{ES}}, S_{\text{HS}}, S_{\text{EV}}) \), where \( S_{\text{Load}} \) is a combination of three terminal load states power/heat/cold.

Similarly, in the selection of the action space, the action strategy in this paper includes whether the equipment unit of CCHP system, GHP, CAC is involved in operation and the charge/discharge behaviour of the ES and HS equipment. The only action strategy \( a_k = (k, a_{\text{CCHP}}, a_{\text{GHP}}, a_{\text{CAC}}, a_{\text{ES}}, a_{\text{HS}}, a_{\text{EV}}) \) can be determined according to the output of the unit in the micro-grid system and the operation of the energy storage during the time period.

The possible \((\text{state, action})\) combination instruction is sent to the operation layer for verification, and the \((\text{state, action})\) combination that does not satisfy the constraint will be removed, and the \((\text{state, action})\) space that satisfies the operation layer constraint condition and reaches the Nash equilibrium through the decision layer is the \((\text{state, action})\) space of microgrid system during the time period. The \( Q \) value of different agents can be calculated after the state space \( S_k \) of the iteration \( k \) and the action strategy \( a_k \) are determined.

**Step (3):** The calculation of the immediate reward value of each agent corresponds to the Formulas (1)–(5); at the same time, the future state \( S' \) is predicted.

**Step (4):** After the future state \( S' \) is obtained, the \( Q \) value table will be updated according to the iterative formula that considers the Nash Q-learning method, and set \( S = S' \). At the same time, note that at different periods of iteration, the energy stored by the ES/HS devices need to be calculated in combination with the corresponding dynamic model [18] and the \((\text{state, action})\) pairs.

**Step (5):** Determine whether the learning process converges, whether the \( Q \)-learning reaches the Nash equilibrium and whether the \( Q \) value of each agent reaches convergence; or the given number of learning steps or time limits have been reached. If it does not converge, then \( k = k + 1 \) and return to Step (2).

The control flow of the algorithm is shown in Fig. 4.

4.2 Control strategy of operation layer based on Petri-net

The microgrid system needs to cater to the price factor of utility energy network and the internal demand of different types of energy load, and also to deal with the impact of environmental factors on the output of renewable energy. Therefore, the effective monitoring and timely adjustment of the operation layer act as a connection link to the safe and stable operation of the whole microgrid system. Considering that the Q-learning method is more to adjust the state and strategy of each agent, the security problem of the overall operation of the system may be ignored. Therefore, this section will introduce the coordinated operation strategy between multiple agents and control instruction logic of the operation layer and equipment layer based on the control diagram in Fig. 3. The Petri-net model is used in this section to describe the multi-agent and the control instruction logic of the operation layer and equipment layer. The Petri-net model gives full play in a variety of expressions such as strict logic statements and graphics, which can perform good analysis and description of the advantages and characteristics about the coupling behaviour of multi-agent system [23]. Fig. 5 shows the schematic diagram of the control strategy. The specific state control/transition logic for multi-agents is shown in Tables 1–4.

In the Petri-net model, the operation switching process of the agent in the microgrid system depends not only on its own event-driven, but also on the interactive behaviour between other agents. For example, an agent that includes an energy storage device switches its charge and discharge state through interaction with other agents; as for PV, it is necessary to obtain support from other agents to meet load requirements through interaction due to the volatility of its output power. When an agent needs to change its operation state through interactive behaviour, it can send requests to other agents. If it gets a reply from other agents, it can achieve the switching of the operation state.

5 Case analysis and comparison

5.1 Case overview

This study took typical industrial park in China as an example case, and the physical structure and equipment constitution of the industrial park is illustrated in Fig. 1. In the terms of energy supply and demand, the cooling season continues from April to October, demonstrating a relatively huge demand on cold load. Heat load mainly consists of demands for drying, fresh air, hot water for daily life and so on existing in the product process, which is relatively seasonal although do not continue for a certain season. Electricity demand lasts for the whole year, and the characteristics curve of the demands on electric, heat and cold load in the microgrid system for the whole year is illustrated in Fig. 6, and the PV output for the whole year is illustrated in Fig. 7. The equipment operation parameters are illustrated in Table 5 [24], and the start-up/shut-down cost for the controllable equipment is 81.94 [24]. The microgrid system contains 100 EVs of the model Nissan leaf [25]. TOU power price in this paper are divided: 10:00–15:00 and 18:00–21:00 as peak period; 07:00–10:00, 15:00–18:00 and 21:00–23:00 as flat period; 00:00–07:00 and 23:00–24:00 as valley period; which are illustrated in Table 6, and the gas price is 2.28 ¥/m³.

In terms of Q-learning algorithm parameters, the learning factor \( \alpha \) is set as 0.01, the discount factor \( \gamma \) is set as 0.8 and the dispatch interval is set as \( \Delta t = 15 \) min. In the terms of state space division: respectively divide the outputs of PV, CCHP, GHP, CAC, ES device and HS device into 6/5/5/8/6/4 discrete spaces, and respectively divide the demands on the electricity, heat and cold into 5/3/4 discrete spaces; divide the output of the EVs with constant charging and discharging power into 1 discrete space; and
In the terms of the action spaces: the PV output is controlled by the external environment and hard to be manually altered, so it possesses only 1 state; the controllable unit possesses 2 states which are on and off; the energy-storage device and the EVs possesses 3 states which are charging, idle and discharging. Each agent establishes its own state/action space according to the included equipment.

### 5.2 Existence proof of Nash equilibrium and analysis of pre-learning

By the continuous iterative search of the pre-learning and the multi-agent game utility functions, the $Q$ value $(Q_{MEP}, Q_{REP}, Q_{EVO})$ of different agents during the energy management scheduling periods will converge to the Nash equilibrium point $(Q_{MEP}^*, Q_{REP}^*, Q_{EVO}^*)$. For the lemma and proof process involved in the Nash equilibrium convergence in the specific Q-learning algorithm, one can refer to [21].

Combined with the case overview in Section 5.1, the offline learning and simulations were performed based on historical data that are shown in Figs. 6 and 7 more than 4000 times. The total cost/benefit changes of different multi-agents are shown in Fig. 8.

The analysis on Fig. 8 leads to the analysis that during the pre-learning phase, the cost/benefit of all different agents are in high values, main reasons of which are: at the initial learning stage, MEP agent will try to reduce the operation cost by making the output as little as possible; on the contrary, REP and EVO agents will try to make more output for discharging to the utility power grid as much as possible for increasing their revenue, but seldom consider their necessary contributions to the energy balance and peak load shifting of the microgrid. However, the related operation layer constraint demands cannot be met, the action strategy explorations of different agents are not comprehensive, the relatively dominant MEP of agents have to sacrifice their own economic interests to ensure the operation layer constraint demands be met, which will bring the a relatively higher cost in the early phase. In the early period of the pre-learning, the values in the state/action $Q$ value table constructed by different agents possesses a relatively big difference from the final equilibrium values, so the optimal action strategy ought to be learnt in the constant exploration with the support of the operation layer

---

**Table 1** REE agent state control/transition logic

| State/transfer | State/transfer description |
|---------------|---------------------------|
| S1            | receive control instructions of the operation layer, |
| S2            | since the PV output is affected by the weather, it is difficult to perform artificial control, so only the ES reception adjustment command is considered here |
| S3            | ES power adjustment command |
| S4            | ES charge/discharge state adjustment command and send ES state adjustment request to other agents |
| S5            | ES operation state is adjusted as needed or to cooperate with other agents |
| S6            | ES stops adjustment and feeds back to the operation layer and decision layer to perform readjustment |
| tr1           | operation adjustments are not needed by the REP agent |
| tr2           | operation adjustments are needed by the REP agent |
| tr3           | ES power adjustment |
| tr4           | ES charge/discharge state adjustment |
| tr5           | adjustment instruction satisfies the ES$_{SOC}$ state constraint |
| tr6           | adjustment instruction does not satisfy the ES$_{SOC}$ state constraint |
| tr7           | other agents respond actively |
| tr8           | other agents do not respond |
### Table 2  MEP agent state control/transition logic

| State/transfer | State/transfer description |
|----------------|---------------------------|
| S1             | receive control instructions of the operation layer |
| S2             | since only the CCHP system in the MEP agent will affect the electric energy supply, only the CCHP reception adjustment command is considered here |
| S3             | since CCHP and GHP in the MEP agent will affect the heat energy, CCHP and GHP are considered to receive coordinated adjustment command |
| S4             | since CCHP and CAC in the MEP agent will affect the cold energy, CCHP and CAC are considered to receive the coordinated adjustment command |
| S5             | the CCHP system operation state is adjusted as needed or to cooperate with other agents |
| S6             | the CCHP system stops the adjustment and feeds back to the operation layer and decision layer to perform readjustment |
| S7             | the CCHP and GHP operation state is adjusted as needed or to cooperate with other agents |
| S8             | the CCHP and GHP operation state is adjusted as needed or to cooperate with other agents |
| S9             | the CCHP system stops the adjustment and feeds back to the operation layer and decision layer to perform readjustment |
| S10            | the CCHP and GHP operation state is adjusted as needed or to cooperate with other agents |
| S11            | the CCHP and CAC operation state is adjusted as needed or to cooperate with other agents |
| S12            | the CCHP system stops the adjustment and feeds back to the operation layer and decision layer to perform readjustment |
| S13            | the CCHP and CAC system operation state is adjusted as needed or to cooperate with other agents |
| S14            | the CCHP and CAC system stops the adjustment and feeds back to the operation layer and decision layer to perform readjustment |
| tr1            | operation adjustments are not needed by the MEP agent |
| tr2            | operation adjustments of electric energy are needed by the MEP agent |
| tr3            | operation adjustments of heat energy are needed by the MEP agent |
| tr4            | operation adjustments of cold energy are needed by the MEP agent |
| tr5            | the adjustment instruction satisfies the CCHP state constraint |
| tr6            | the adjustment instruction does not satisfy the CCHP state constraint |
| tr7            | other agents do not respond |
| tr8            | other agents respond actively |
| tr9            | the adjustment instruction satisfies the CCHP and GHP state constraint |
| tr10           | the adjustment instruction does not satisfy the CCHP and GHP state constraint |
| tr11           | other agents do not respond |
| tr12           | other agents respond actively |
| tr13           | the adjustment instruction satisfies the CCHP and CAC state constraint |
| tr14           | the adjustment instruction does not satisfy the CCHP and CAC state constraint |
| tr15           | other agents do not respond |
| tr16           | other agents respond actively |

### Table 3  EVO agent state control/transition logic

| State/transfer | State/transfer description |
|----------------|---------------------------|
| S1             | receive control instructions of the operation layer |
| S2             | EVs charge and discharge behaviour adjustment instruction |
| S3             | the EVO agent sends instructions to other agents, and requests for coordinated adjustment |
| S4             | on-demand adjustment of charge and discharge behaviour of EVs |
| S5             | EVs stop the adjustment and feedback to the operation layer and decision layer to perform readjustment |
| tr1            | operation adjustments are not needed by the EVO agent |
| tr2            | operation adjustments are needed by the EVO agent |
| tr3            | EVO response adjustment command |
| tr4            | EVO does not respond to adjustment command |
| tr5            | other agents respond actively |
| tr6            | other agents do not respond |

### Table 4  TEU agent state control/transition logic

| State/transfer | State/transfer description |
|----------------|---------------------------|
| S1             | receive control instructions of the operation layer |
| S2             | electric/heat/ cold load energy-consumption adjustment command |
| S3             | the TEU agent sends instructions to other agents, and requests for coordinated adjustment |
| S4             | on-demand adjustment of consumption behaviour of electric/ heat/ cold load |
| S5             | electric/heat/cold load stop the adjustment and feedback to the operation layer and decision layer to perform readjustment |
| tr1            | operation adjustments are not needed by the TEU agent |
| tr2            | operation adjustments are needed by the TEU agent |
| tr3            | TEU responds to adjustment commands for load translation or reduction |
| tr4            | TEU does not respond to the adjustment commands |
| tr5            | other agents respond actively |
| tr6            | other agents do not respond |
calibration. With the continuous deepening of learning, all agent revenues will approximate to optimal solution of Nash equilibrium and decision-making capability for Q-learning will constantly improve.

After 2000 times of learning on historical data and the analysis on the multi-agent economic game, the cost/benefit of all agents have gradually grown into be stable, which indicates that at present, Q-learning algorithm has passed the exploration trial and error, accumulated certain experience and obtained the capability to make relatively proper energy management policies.

5.3 Simulation result analysis of online energy management

To further verify the online decision-making capability for the optimisation learning method of hierarchical control on the multi-energy microgrid, based on the pre-learning process, this study further verified the online learning decision capability of the method. Specifically, combined with the scheduling sequence shown in Fig. 4 and the algorithm characteristics of Q-learning, the method proposed in this paper can also be applied to inter-day scheduling.

Considering the low-cold load demand at the selected period, this section mainly analyses the power/heat scheduling management, and the prediction curves of electric/heat load demands are illustrated in Fig. 9.

The power/heat energy management balance and dynamic output changes of the energy supply equipment and the energy storage equipment obtained by the method of this study are shown in Figs. 10 and 11. Where, $P_{\text{load},e1}$ curve is the initial demand curve of electric load; $P_{\text{load},e2}$ curve is the power supply curve considering the ES, EVs charging and CAC power consumption demands.

The result shows a vigorous PV output at the period of 11:00–13:00, which is also the peak period of the electricity price. Therefore, the REP agent would deliver the unused PV output in the load supply to the utility grid to gain a benefit. The electricity price valley happens at 0:00–7:00 and 23:00–24:00. Now, the price is relatively low. Thus, the REP agent would deliver a signal for the ES charging and the purchase electricity from the utility grid. The power load is high and the heat load is low during 18:00–21:00. Accordingly, the CCHP system would deliver an extra heat for HS after satisfying the heat load demand. This can relieve the power supply pressure caused by the insufficient PV output at night.

Considering the fluctuating output of the PV and based on the energy management ability of the method introduced here from the perspective of load supply and demand balance, the output distribution of the REP agent with the PV and ES is further analysed, which is shown in Fig. 12. During the vigorous output period of PV, the REP agent realises the supply of a partial electric load. At the peak period of the electricity price, it sells electricity to the utility power grid to obtain the income. Besides, the ES equipment charges the electricity during the valley price period or the vigorous output period of the PV, and it discharges the electricity at the necessary time.

In summary, by the result analysis on the online learning scheduling in the typical day, the method proposed in this paper can ensure the supply–demand balance and the safe operation of the system, and possesses decision-making ability for rational energy management.

5.4 Comparison of algorithm performance

Chaos particle swarm optimisation algorithm [26] is selected to achieve the economic optimisation by optimising the equipment output and compared with Q-learning method introduced in this paper. Based on the online learning environment in Section 5.3 and taking the daily operation cost of MEP agent as an example, the

---

**Table 5** Operation parameters for each type of unit

| Devices type | Rated capacity/power | $P_{\text{min}}$, kW | $P_{\text{max}}$, kW | Conversion efficiency | $K_m$, ¥/kW |
|--------------|----------------------|----------------------|----------------------|----------------------|-------------|
| CCHP         | 2000 kW              | 20                   | 2000                 | 1                    | 1.547       |
| PV           | 2600 kW              | 0                    | 2600                 | —                    | 0.02        |
| GHP          | 1000 kW              | 10                   | 1000                 | 0.8                  | 0.40        |
| CAC          | 2000 kW              | 20                   | 2000                 | 0.88                 | 0.30        |
| ES           | 3000 kW/h            | —                    | 2400                 | 0.9                  | 0.03        |
| HS           | 1000 kW/h            | —                    | 800                  | 0.9                  | 0.03        |
| EV           | 30 kW/h              | —                    | 66                   | 0.9                  | —           |
| tie lines    | —                    | —                    | 800                  | —                    | —           |

**Table 6** Electricity price

| Period            | Purchase, ¥/(kW h) | Sell, ¥/(kW h) |
|-------------------|---------------------|----------------|
| peak period       | 0.83                | 0.65           |
| flat period       | 0.49                | 0.38           |
| valley period     | 0.17                | 0.13           |
iterative operation processes of different algorithms are illustrated in Fig. 13.

According to Fig. 13, for the complex optimisation problems of higher dimension solution space, heuristic algorithm requires a long search process and 1600 times of iteration to reach convergence range takes 7893.8 s for online optimisation, which is hard to fulfil the timeliness demand of energy management, in addition, heuristic algorithm is easy to fall into local optimum. However, Q-learning method after the pre-learning process can directly locate the search space of states and actions to be near the final solution by learning from experience, and deeply explore to acquire the final Nash equilibrium solution, which can converge by about 800 times of iteration taking 5537.6 s online, possesses a relative advantages in algorithm efficiency and the optimal result.

6 Conclusions

In the context of frequent energy interaction and diversified means of information communication, multi-energy microgrids including energy production/transmission/conversion and utilisation are significant components of the energy Internet. How to improve the conventional form of energy utilisation and achieve effective management and coordinated scheduling of multiple energy sources is the top priority of future researches. In this thesis, the multi-energy microgrid was taken as the research object, and a hierarchical control optimisation learning method considering multi-agent game was proposed. The main conclusions of this manuscript are as follows.

Based on the different interests existing in the microgrid system, the multi-agent partitioning and game relationship analysis were carried out, and the transition trend of behaviour strategy of different agents in the coordinated scheduling process of multi-energy microgrid was analysed, from only considering their own economic optimisation to considering global operation stability by practical cases.

The multi-agent economic decision-making model for different interests and the game decision-making model considering multi-agent interest equilibrium were constructed, respectively. Through the analysis of the actual day-ahead scheduling cases, each agent can realise the optimisation of the self-benefit by reasonably arranging the working strategy of the controllable resources under the premise of ensuring the overall operation safety of the system.
AI method was applied that considered Nash Q-learning and Petri-net to realise the coordinated scheduling of multi-energy microgrid, and introduced the AI method into the control link of multi-energy microgrid. By the practical cases, the reinforcement learning method after pre-learning was superior to the traditional heuristic algorithm in the aspects of convergence speed and calculation time and so on.

In the future research, adding the TEU agent to the game model will be further considered, and the impact of the terminal load demand response and how will it affect the agent's economy will be considered; at the same time, considering the transaction between microgrid system and the utility power grid, the benefits brought by the load supply in the microgrid system will be refined in the future, and the regulation strategy will be improved; in addition, the economic benefits under the multi-agent game situation will be compared with the economic benefits of the traditional centralised whole-social one to provide guidance for the future operation mode of the microgrid system.

7 Acknowledgment

This work was supported by the Natural Science Foundation of China (grant no. 51777133).

8 References

[1] Xu, X., Jin, X., Jia, H., et al.: ‘Hierarchical management for integrated community energy systems’, Appl. Energy, 2015, 160, pp. 231–243
[2] Wu, J., Yan, J., Jia, H., et al.: ‘Integrated energy systems’, Appl. Energy, 2016, 167, pp. 155–157
[3] Davatgaran, V., Saniei, M., Mortazavi, S.S.: ‘Optimal bidding strategy for an energy hub in energy market’, Energy, 2018, 148, pp. 482–493
[4] Hossain, M.A., Pota, H.R., Squartini, S., et al.: ‘Energy scheduling of community microgrid with battery cost using particle swarm optimization’, Appl. Energy, 2019, 254, p. 113723
[5] Keshta, H.E., Ali, A.A., Saied, E.M., et al.: ‘Real-time operation of multi-micro-grids using a multi-agent system’, Energy, 2019, 174, pp. 576–590
[6] Fernandez, E., Hossain, M.I., Nizami, M.S.H.: ‘Game-theoretic approach to demand-side energy management for a smart neighbourhood in Sydney incorporating renewable resources’, Appl. Energy, 2018, 232, pp. 245–257
[7] Yue, C., Wei, W., Feng, L., et al.: ‘Analyzing and validating the economic efficiency of managing a cluster of energy hubs in multi-carrier energy systems’, Appl. Energy, 2018, 230, pp. 403–416
[8] Mehmood, M.U., Chan, D., Zeeshan., et al.: ‘A review of the applications of artificial intelligence and big data to buildings for energy-efficiency and a comfortable indoor living environment’, Energy Build., 2019, 202, p. 109383
[9] Liu, R., Yang, B., Zio, E., et al.: ‘Artificial intelligence for fault diagnosis of rotating machinery: a review’, Mech. Syst. Signal Process., 2018, 108, pp. 33–47
[10] Zhao, Y., Li, T., Zhang, X., et al.: ‘Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future’, Renew. Sust. Energy Rev., 2019, 109, pp. 85–101
[11] Afraasiabi, M., Mohammadi, M., Rastegar, M., et al.: ‘Multi-agent microgrid energy management based on deep learning forecaster’, Energy, 2019, 186, p. 115873
[12] Wei, N., Li, C., Peng, X., et al.: ‘Conventional models and artificial intelligence-based models for energy consumption forecasting: a review’, J. Pet. Sci. Eng., 2019, 181, p. 106187
[13] Li, F., Wu, M., He, Y., et al.: ‘Optimal control in microgrid using multi-agent reinforcement learning’, ISA Trans., 2012, 51, (6), pp. 743–751
[14] Li, D., Jayaweeru, S.K.: ‘Machine-learning aided optimal customer decisions for an interactive smart grid’, IEEE Syst. J., 2015, 9, (4), pp. 1529–1540
[15] Kofinas, P., Dounis, A.I., Voros, G.A.: ‘Fuzzy Q-learning for multi-agent decentralized energy management in microgrids’, Appl. Energy, 2018, 219, pp. 53–67
[16] Kazimi, H., Syukri, J., Balint, A., et al.: ‘Multi-agent reinforcement learning for modeling and control of thermostatically controlled loads’, Appl. Energy, 2019, 238, pp. 1022–1035
[17] Chengshahan, W., Chaoxian, L., Peng, L., et al.: ‘Modeling and optimal operation of community integrated energy systems: a case study from China’, Appl. Energy, 2018, 230, pp. 1242–1254
[18] Ge, S., Li, J., Liu, H., et al.: ‘Research on operation-planning double-layer optimization design method for multi-energy microgrid considering reliability’, Appl. Sci., 2018, 8, (11), p. 2062: 1–21
[19] Wu, Q., Wu, J., Wang, L., et al.: ‘Determination and analysis of TOU (time-of-use) power price based on DSM (demand side management) and MCP (marketing clearing price)’, the 6th Int. Conf. on Advances in Power System Control, Operation and Management, APSCOM2003, Hong Kong, November 2003
[20] Rocchetta, R., Bellani, L., Compare, M., et al.: ‘A reinforcement learning framework for optimal operation and maintenance of power grids’, Appl. Energy, 2019, 241, pp. 291–301
[21] Hu, J.L., Wellman, M.P.: ‘Nash Q-learning for general-sum stochastic games’, J. Mach. Learn. Res., 2004, 4, (6), pp. 1039–1069
[22] Tovar, J., Olmo, F.J., Aladoss-Arboledas, L.: ‘One-minute global irradiance probability density distributions conditioned to the optical air mass’, Sol. Energy, 1998, 62, pp. 397–417
[23] Zhong, C., He, W., Li, Z., et al.: ‘Deadlock analysis and control using Petri net decomposition techniques’, Inf. Sci., 2019, 482, pp. 440–456
[24] Tsikalakis, A.G., Hatziargyriou, N.D.: ‘Centralized control for optimizing microgrid operation’, IEEE Trans. Energy Convers., 2008, 23, (1), pp. 241–248
[25] Specifications of 2016 Nissan LEAF[OL]. Available at http://www.nissanusa.com/electric-cars/leaf/versions-specs/version.0.html
[26] Tang, X., Ling, Z., Cai, J., et al.: ‘Multi-fault classification based on support vector machine trained by chaos particle swarm optimization’, Knowl.–Based Syst., 2010, 23, pp. 486–490

489

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)