Model-Free Energy Optimization for Energy Internet

Energy Internet and We-Energy pp 299-325 | Cite as

- Qiuye Sun (1) Email author (sunqiuye@mail.neu.edu.cn)

1. College of Information Science and Engineering, Northeastern University, , Shenyang, China

Chapter
First Online: 13 July 2018

Part of the Renewable Energy Sources & Energy Storage book series (RESES)

Abstract

With the high penetration of distributed energy, the scale of current energy network becomes larger. At the same time, it also has the problems of complex computation and slow convergence. In order to realize the rational planning and utilize various energy resources and improve the reliability and economy of the overall system, the problems of system stability, economic operation and power flow calculation must be considered comprehensively. As a kind of machine learning, reinforcement learning has strong intelligence and rapidity, which can realize the optimal control of the system. Aiming at the energy management model of regional energy Internet, this chapter studies how to transform energy management into Q learning model, and uses Q learning algorithm to verify the validity of the model. In the meantime, for the optimization scheduling problem of large-scale system, this chapter expands the optimal power flow model of energy internet into the optimal operation structure composed of multiple We-Energies based on the previous one, and uses the distributed reinforcement learning algorithm to optimize the large-scale energy internet scheduling which considers the average consistent information search to achieve the optimization process for cooperating and communicating multiple We-Energies.

10.1 Introduction

Optimal energy flow (OEF) is widely used to realize the interconnected multicarrier system (IMS) of economic and security operation to reduce the network loss [1]. The IMS is conventionally modeled as a multitude of subsystems in which the operation of energy subsystems was scheduled and optimized individually [2]. "We-Energy" (WE), as a novel energy interacting area for energy internet, exchanges energy (electricity, district heat and natural gas) with others by the advanced communication, electronic conversion and automatic control technology [3]. In order to enhance the performance of environment, economic and security, many papers have a comprehensive research on it. In [4, 5, 6, 7, 8, 9], the optimal power flow of electricity and natural gas combined system is discussed. Paper [10] discussed the scheduling model of multiple energy system which is based on distributed CHP device. In this model, the city network of electricity, heating and natural gas is coupled. Energy production and consumption matching problem is summarized at CHP system level in [11]. The optimal operation of electricity and heating combined system is studied in [12], which considered the heating network constrains and proved that CHP system can promote to dispose the wind power.

In this chapter, two issues concerning optimal performance of energy internet are considered. The first research is conducted on energy management of We-Energy. The combined of multiple types of energy is one of the specific characteristics of Energy Internet. The Energy Internet can be assumed as a cluster of distributed energy resources and loads, which contains various types of energy resources such as electricity, gas, heat and so on [13]. The use of different kind of energy brings great benefit to Energy Internet, which allows multiple end users to make options according to their own power demands, hence increasing the flexibility of the power system and weakening the impact of traditional energy supplier. However, using distributed generations indiscriminately may also impose undesirable effects on power system. Therefore, issues on optimal energy management come into play. A lot of researches concerning control and operation of power system have been done in recent years. Several common optimization objectives including lower cost of carbon and minimum operating cost have been discussed in [14]. In [15], authors proposed a smart energy management system in order to minimize the operating cost of the micro-grid. Only electricity is discussed during optimization process while other types of energy resources are not considered in the paper. The authors in [16] proposed a micro-grid scenario consists of combined heat and power generation, as well as power and thermal energy storage devices. And an online algorithm has been put forward to optimize the cost of whole system.

However, the optimized economic dispatch does not always satisfy the demands when taking pollutants emission into account. So, multi-objective energy management has drawn attention from researchers so as to realize optimization both economically and environmentally. The authors in [17] proposed an intelligent energy management system (IEMS) for a CHP-based micro-grid, and minimized the operation cost and the net emission simultaneously. An efficient modified bacterial foraging optimization algorithm was used to find the optimal set points of the system. Reference [18] proposed a Stackelberg game-based optimization model, and a differential evolution-based heuristic algorithm was designed to reach the Stackelberg equilibrium. But in the previous studies, there is a lack of consideration of specific characteristics of Energy Internet, such as

openness, sharing and peer-to-peer integration.

Another research is conducted on optimal power flow. Paper [19] proposed the generalized heuristic algorithm to study the optimal power flow of multiple energy system. While with the increasing utilization of co-generation plants such as electricity, natural gas and local district heating systems that make a strong coupling in IMS [20], the structure of network becomes more complex. There is a challenge to find the optimal strategy in such a way for this class of complex nonlinear multi objective optimization problems. The traditional optimal power flow algorithm such as linear programming [21], interior-point method [22], is unable to obtain the global optimal solution with these problems and conventional artificial intelligence such as algorithms genetic algorithm, particle swarm optimization have the disadvantages of slow computation speed for the large-scale network. The domestic and foreign scholars made in-depth study in this question for more efficient algorithm. In addition to improve the basic heuristic algorithm for OEF [23], distributed algorithm has become a research focus [24]. Meanwhile, some hybrid algorithm has been discussed. Paper [25] presents the reinforcement learning combined with simulated annealing (SA) algorithm to solve the optimal reactive power dispatch.

Recently, reinforcement learning algorithm (RL) as a kind of machine learning algorithms attracts people's attention. Some learning strategies on the basis of RL to solve deterministic optimal control problems in continuous state spaces can be found in some studies such as [26, 27, 28, 29, 30, 31, 32, 33, 34, 35]. The distributed reinforcement learning (DRL) which is a new branch of reinforcement learning algorithm has been developed rapidly in various areas including distributed control, robotic teams, collaborative decision support systems, and economics [36]. DRL is defined to be composed of multiple agents, the whole system will achieve the learning goals through each agent executing part of reinforcement mission independently. All performance of DRL exhibits its advantage on the col problems, and the features aim at strategic decision make DRL widely used [37]. Each Pareto optimal solution is also a Nash equilibrium for a fully cooperative game [38], which means that if one agent is provided with greatest possible reward in a combined action, the reward received by the other agents must also be maximized.

In this chapter, the energy management of We-Energy is discussed. An Energy Internet model consisting of combined heat and power unit (CHP), photovoltaic unit, heating only unit and storage device is constructed. To construct an environmental-friendly and low-operating cost energy consumption structure, a multi-objective optimization model is proposed. Furthermore, in order to satisfy the power and heat demands of the We-Energy simultaneously as well as realize minimum operating cost and pollutant emission, an intelligent energy management system (IEMS) is presented. In particular, reinforcement learning method has been implemented to formulate the optimal operating strategy. Eligibility trace theory is also been introduced to accelerate the computational process.

What's more, the optimal energy flow in interconnected multicarrier systems where

electric, heat and natural gas systems are coordinated. We propose a double-deck optimal model to improve the performance of security and environment for IMS. The proposed formulation for large-scale systems can be solved by machine learning algorithms which could find the optimization strategy intelligently. We present a hybrid reinforcement learning algorithm (HRL) for distributed multicarrier energy network, which computes a global optimal policy in cooperative subsystems on the basis of the implementation of independent optimization for subsystems. A policy is defined as a set of actions deriving from the reward function connecting the environment.

10.2 Reinforcement Learning Applied to Energy Management

10.2.1 Reinforcement Learning on Markov Decision Processes

As a main class of machine learning methods, reinforcement learning (RL) is an effective means for making sequential decision under uncertainties. In a reinforcement learning system, a reinforcement learning agent aims to find an optimal action policy by trial-and-error interaction with its uncertain environment. At each time step, the learning agent perceives the state of the environment but it is not provided with explicit information of the action to take. The agent autonomously selects a random action with certain probability and the current state of the environment therewith transits into its successive state. After that, the learning agent can receive a reward signal that evaluates the effect of this action.

A Markov decision process can be characterized as the formulation of a sequential decision-making problem. Therefore, reinforcement learning can be described by a finite *Markov* decision process.

A finite Markov decision process can be characterized as a 4-tuple $\{S, A, R, P\}$, where S is the state space of a finite set of states, A is the action space of a finite set of actions, R is the reward function, P is the matrix of state transition probability. $r\left(s, a, s'\right)$ and $p\left(s, a, s'\right)$ represents the instant reward and probability of the state transition from state s to state s' taking action a. An action policy of the MDP can be defined as $\pi : S \rightarrow A$.

The objective function of reinforcement learning is to receive the largest discounted reward. Therefore the state value function can be defined as

$$V^{\pi}(s) = \sum_{t=0}^{\infty} \gamma^{t} r\left(s_t, a_t\right) | s_0 = s, a_t = \pi\left(s_t\right), \tag{10.1}$$

where $\gamma \in (0, 1)$ is the discount factor, s_0 is the initial environment state. According to the equation above, the optimal action policy can be described as

$$\pi^* = \arg\max_{\pi} V^{\pi}(s),$$
(10.2)

The action-state value function can be defined as
$$Q^{\pi}(s_t, a_t) = r(s_t, a_t) + \gamma V^{\pi}(s_{t+1}),$$
(10.3)

According to the equation above, the optimal action policy can be described as
$$\pi^* = \arg\max_{a \in A} Q^{\pi}(s, a),$$
(10.4)

10.2.2 The Q Learning Applied to Energy Management

10.2.2.1 Modeling of Energy Management and Multi-objective Optimization

In the Energy Internet, different structures of energy prosumers bring about various energy demands. Consequently, as a novel energy interaction area of Energy Internet, We-Energy no longer follows the track of traditional energy network where different types of energy are supplied independently. WE is capable to transform various types of energy such as electricity, district heat and natural gas into desired energy and exchange with others using advanced communication, electronic conversion and automatic control technology. That is to say, WE is no longer a passive consumer, but also a potential energy supplier during the energy interaction process.

As shown in Fig. 10.1, WEs are connected to information bus and energy bus simultaneously which can realize bi-directional interaction of information and bi-directional transmission of energy. Information of each WE can be submitted to the information bus and useful data of other WEs located in the Energy Internet can be abstracted from information bus as well, which achieves bi-directional information interaction. WE is able to supply excess energy to others through energy bus and can get compensatory energy when needed. Various types of energy can be transmitted through energy bus. In this section, we take electricity and heat into account simultaneously.

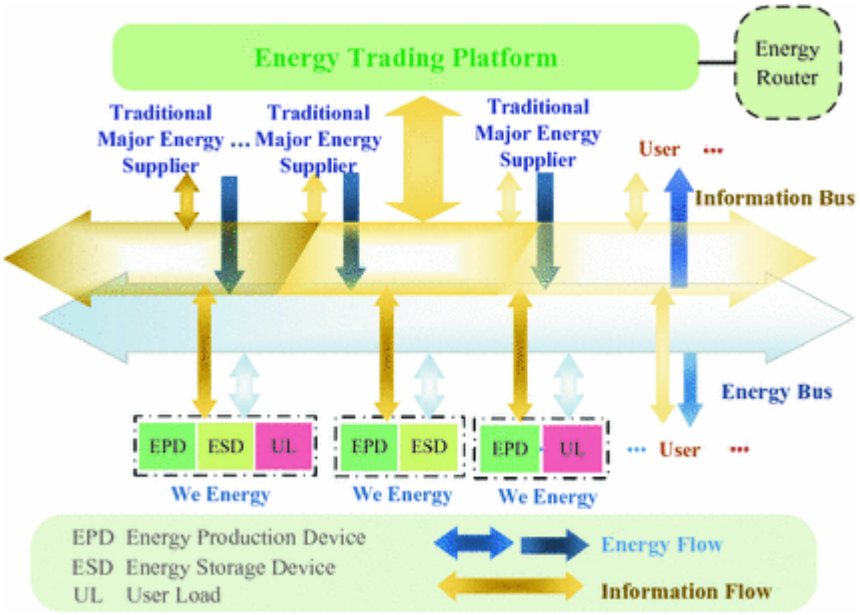

Fig. 10.1

Operation mode of We-Energy in the energy internet

Basically, a WE is comprised of three units including energy production unit, energy storage unit and user load. Therefore, WE can be classified into seven types in which traditional pattern of energy network is embraced as well. WE only consists of user load and energy supplier can be regarded as traditional energy users and traditional energy supplier respectively. Energy storage unit plays an important role in energy management of We-Energy and can improve its performance and economic efficiency. In this section, we adopt the WE structure consists of three units at the same time.

In this research, we consider a WE composed of both electrical and thermal producers, users and storage devices.

The goal of IEMS is to satisfy electrical and thermal load demands considering both economic and environmental criteria. We-Energy participates in the open market, buying and selling power to the Energy Internet via energy bus and information bus.

1. A.

Objective function

In expectation of a more sustainable and environmental friendly dispatch of We-Energy, we consider a WE holds more renewable energy resources and less fossil fuels. Thus, the objective function should be formulated as

$$f = \min(f_1, f_2),$$

(10.5)

where f_1 is the operating cost function of energy dispatch, f_2 is the pollutant emission function of energy dispatch.

1. B.

Operating cost

As We-Energy is a comprehensive combination of energy producers in the Energy Internet, multi types of primary energy resources including fossil fuels, natural gas and renewable energy are used to satisfy the demand of users. The optimization of economic benefit is to minimize the operating cost of energy suppliers. The operating cost of We-Energy includes the cost of photovoltaic, gas-fired CHP and heat only unit. The objective function is introduced as

$$f_1 = \min \sum_{t=1}^{T} \left(\sum_{i=1}^{N_{CHP}} F_{CHPi,t} + \sum_{i=1}^{N_{PV}} F_{PVi,t} + \sum_{i=1}^{N_{Heat}} F_{Heati,t} + F_{Grid,t} \times bid_t \right),$$

(10.6)

where N_{CHP}, N_{pv}, N_{Heat} is the total number of natural gas-fired CHP, photovoltaic and heat only unit respectively. $F_{PV_{i,t}}$, is the cost of the ith photovoltaic and the cost is related to the of active power. $F_{CHP_{i,t}}$,, $F_{Heati,t}$ is the cost of the ith CHP and heating only unit at time step t respectively. $F_{Grid,t}$ is the active power abstracted from the Energy Internet, while bid_t is the corresponding electrical price.

The nonlinear cost function of a CHP unit can be explained as below

$$F_{CHPi,t} = a_i \times P_{CHPi,t}^2 + b_i \times F_{CHPi,t} + c_i + d_i \times H_{CHPi,t}^2 + e_i \times H_{CHPi,t} + f_i \times H_{CHPi,t} \times p_{CHPi,t},$$

(10.7)

where a_i,b_i, c_i, d_i, e_i, f_i are the generation parameters of the ith natural gas-fired CHP, $F_{CHP_{i,t}}$ is the active power of the ith natural gas-fired CHP at time t.

The cost function of a heat only unit is expressed as a quadratic function

$$F_{Heatj,t} = a_j \times H_{Heatj,t}^2 + b_j \times H_{Heatj,t} + c_j,$$

(10.8)

where a_j, b_j, c_j are the generation parameters of the jth heating only unit, $H_{Heatj,t}$ is the active power of the jth natural gas-fired CHP at time step t.

1. C.

Pollutant emission

Due to the aggravation of environmental pollution and energy shortage, there is a rising trend of reducing consumption of coal, natural gas and other traditional fossil fuels. In order to minimize the pollutant emission, the use of clean renewable energy should be

maximized, and the objective function is expressed as

$$f_2 = \min \sum_{t=1}^{T} \left[\left(\sum_{i=1}^{N_{CHP}} \alpha_i H_{CHPi,t}^2 + \beta_i H_{CHPi,t} + \gamma_i \right) + \left(\sum_{j=1}^{N_{Heat}} \alpha_j H_{Heatj,t}^2 + \beta_j H_{Heatj,t} + \gamma_j \right) + E_{Grid,t} \right],$$

(10.9)

where $\alpha_i, \beta_i, \gamma_i$ are the emission parameters of the ith gas-fired CHP, and $\alpha_j, \beta_j, \gamma_j$ are the emission parameters of the jth heating only unit. As can be seen in (10.9), the emission function of CHP unit and heat only unit is quadratic function.

In this section, the emission during the production of electricity is also taken into consideration. The emission mainly caused by burning fossil fuels in thermal power plants, considering coverage fraction of these plants, E_{Grid} can be defined as follows

$$E_{Grid,t} = \alpha P_{Grid,t}^2 + \beta F_{Grid,t} + \gamma',$$

(10.10)

where α, β, γ' are emission parameters of thermal power plants.

1. D.

Constraints

There are several constraints should be taken into consideration. Power balance between electrical demand and electrical supply, and balance between thermal demand and thermal supply are expressed as follows

$$\sum_{i=1}^{N_{CHP}} F_{CHPi,t} + \sum_{i=1}^{N_{PV}} F_{PVi,t} + F_{Grid,t} + F_{ch,t} - F_{dis,t} = F_{load,t},$$

(10.11)

$$\sum_{i=1}^{N_{CHP}} H_{CHPi,t} + \sum_{i=1}^{N_{Heat}} H_{Heati,t} = H_{load,t},$$

(10.12)

where $F_{ch,t}, F_{dis,t}$ are the charging and discharging rate of the storage unit at time step t respectively. $F_{load,t}, H_{load,t}$ are the electrical load demand and heat demand respectively.

Equations (10.13) and (10.14) can be expressed as follows as well

$$\sum_{i=1}^{N_{CHP}} F_{gasi,t} \times \eta_e^C + \sum_{i=1}^{N_{PV}} F_{PVi,t} + F_{Grid,t} + F_{ch,t} - F_{dis,t} = F_{load,t},$$

(10.13)

$$\sum_{i=1}^{N_{CHP}} F_{gasi,t} \times \eta_h^C + \sum_{i=1}^{N_{Heat}} H_{Heati,t} = H_{load,t},$$

(10.14)

where $F_{gasi,t}$ is the gas input of the natural gas-fired CHP at time step t, η_e^C, η_h^C are the output ratio of electric power and heat respectively.

In addition, the output power of all the units should satisfy its upper and lower bound, which can be expressed as

$$P_{PVi,t} \leq P_{PV}^{\max},$$
(10.15)

$$P_{ch,t} \leq P_{ch}^{\max},$$
(10.16)

$$P_{dis,t} \leq P_{dis}^{\max},$$
(10.17)

$$P_{CHP}^{\min} \leq P_{CHP,t} \leq P_{CHP}^{\max},$$
(10.18)

$$H_{CHP}^{\min} \leq H_{CHP,t} \leq H_{CHP}^{\max},$$
(10.19)

$$H_{Heat,t} \leq H_{Heat}^{\max},$$
(10.20)

$$E_{stor} \leq E_{stor}^{\max},$$
(10.21)

where $P_{PV}^{\max}$, $P_{ch}^{\max}$, $P_{dis}^{\max}$, $P_{CHP}^{\max}$, $H_{CHP}^{\max}$, $H_{Heat}^{\max}$, $E_{stor}^{\max}$ are the upper limit of each device respectively. $P_{CHP}^{\min}$, $H_{CHP}^{\min}$ are the lower bound of output electrical power and heat of CHP respectively.

10.2.2.2 Reinforcement Learning Method

1. A.

   Q-learning with eligibility trace

The aim of Q-learning is to learn the value of each action taken from the action space at each state, which is defined to be the predicted total discounted reward received by the agent over the future as a result of taking that action from the action space. The one-step Q-learning is defined as follows [11]:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \eta[r_{t+1} + \gamma \max_{a \in A} Q_{t+1}(s_{t+1}, a) - Q_t(s_t, a_t)],$$
(10.22)

At each time step t, the action value $Q_t(s_t, a_t)$ is recorded. After selecting a subsequent state s_{t+1}, an immediate reward r_{t+1} is obtained and $\max_{a \in A} Q_{t+1}(s_{t+1}, a)$ is picked out by searching a lookup table that stores the action values for each state. The parameter $\gamma \in [0, 1)$ is the discount factor. If the discount factor is small, the agent tends to care more about the immediate reward rather than the rewards received in the future. Thus, in order to make the agent more "farseeing", a large discount factor is chosen in this section. The parameter $\eta \in [0, 1)$ is the learning rate which determines how far the agent

is adjusted towards the estimated value. A large factor allows the agent to learn faster, and vice versa. In this section, a large learning rate is selected to shorten the learning process.

In order to accelerate the training process, the eligibility trace theory is introduced to the Q-learning algorithm. The updating rule of the eligibility trace is expressed as follows:

$$e_{t+1}(s,a) = \begin{cases} \gamma\lambda e_t(s,a) + 1 & s = s_t, a = a^* \\ 0 & s = s_t, a \neq a^* \\ \gamma\lambda e_t(s,a) & s \neq s_t \end{cases},$$

(10.23)

where $e_t(s, a)$ is the eligibility trace of the state-action pair at time step t, λ is the trace decay parameter, a^* is the optimal action at time step t. A larger decay parameter makes the algorithm converge faster, so a large λ is adopted in this section.

Therefore, in consideration of eligibility, the updating rule of Q-learning can be rewritten as

$$Q_{t+1}(s_t,a_t) = Q_t(s_t,a_t) + \eta[r_{t+1} + \gamma \max_{a\in A} Q_{t+1}(s_{t+1},a) - Q_t(s_t,a_t)]e_t(s,a).$$

(10.24)

1. B.

State space and action space

The state space of the intelligent energy management system is as follows:

$$S = \{s \mid bid_t\}.$$

(10.25)

The action space of the intelligent energy management system can be described as

$$A = \left\{a \mid F_{gas,t}\right\}.$$

(10.26)

1. C.

Reward function

After taking an action, the IEMS receives an immediate reward to evaluate the selected action. Since our goal is to optimize the operation of We-Energy overall rather than merely optimize a single objective, the reward function should take the aforementioned

two objectives into consideration simultaneously. In addition, constrains should be satisfied as well. Therefore, the reward function is defined as

$$r = -\left((1-\omega)\, f_1 + \omega f_2 + KN\right),$$

(10.27)

where parameter $\omega \in (0, 1)$ is weight, f_1 is the value of the operating cost function, f_2 is the value of the pollutant emission function, K is a positive number, N is the number of the inequality constraints that not be satisfied.

1. D.

   Action selection policy

The action selection policy allows the agent to select an action a_i at state s with a probability of $p(s, a_i)$ according to the action values. The policy can be described as

$$p(s, a_i) = \frac{e^{Q(s,a_i)/\tau}}{\sum_{a_i} e^{Q(s,a_i)/\tau}},$$

(10.28)

where τ is a parameter called temperature, which determines the randomness of the exploration. If a lower temperature is selected, the agent tends to select the action with higher action value, while a higher temperature makes the agent act more randomly.

10.2.3 Simulation and Results

In order to verify the effectiveness of the proposed energy management strategy for the We-Energy, the following simulation model is built as shown in Fig. 10.2. We consider a We-Energy model consists of a combined heat and power unit, a photovoltaic unit, a heating only unit and a storage unit.

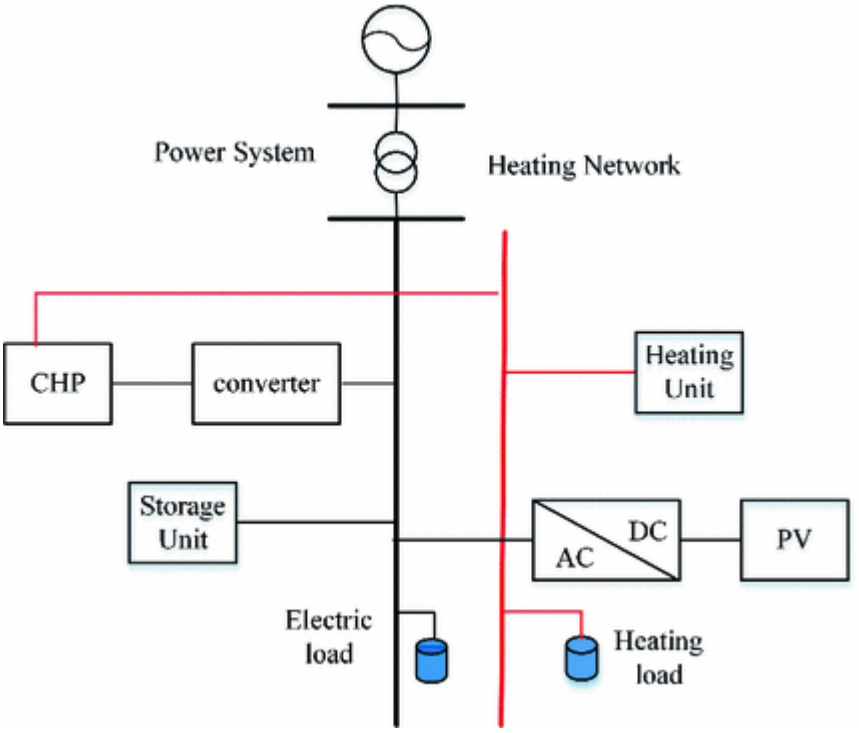

Fig. 10.2

Simulation model of We-Energy

Table 10.1 explains the limits of output power of each device in the proposed system. Cost coefficients and emission coefficients are presented in Tables 10.2 and 10.3.

Table 10.1

Installed device

ID	Type	Min power (kW)	Max power (kW)
1	PV	0	25
2	CHP (electrical)	0	60
3	CHP (heat)	0	80
4	Heat	0	50

Table 10.2

Cost fuction coeffecients

Devices	Cost parameters					
	a_i	b_i	c_i	d_i	e_i	f_i
CHP	0.0065	1.21	2	0.003	4	0.61
Heat	0.038	2.011	65	–	–	–

Table 10.3

Emission function coeffecients

Devices	Emission parameters		
	α_i	β_i	γ_i
CHP	0.08	−2	11
Heat	0.7	−2	5
Power plants	0.46	−1.3	3.27

Power demand and heat demand are shown in Figs. 10.3 and 10.4 respectively. The output power of photovoltaic is shown in Fig. 10.5. The day ahead market price of electricity is proposed in Fig. 10.6. The operating cost of photovoltaic is shown in Fig. 10.7.

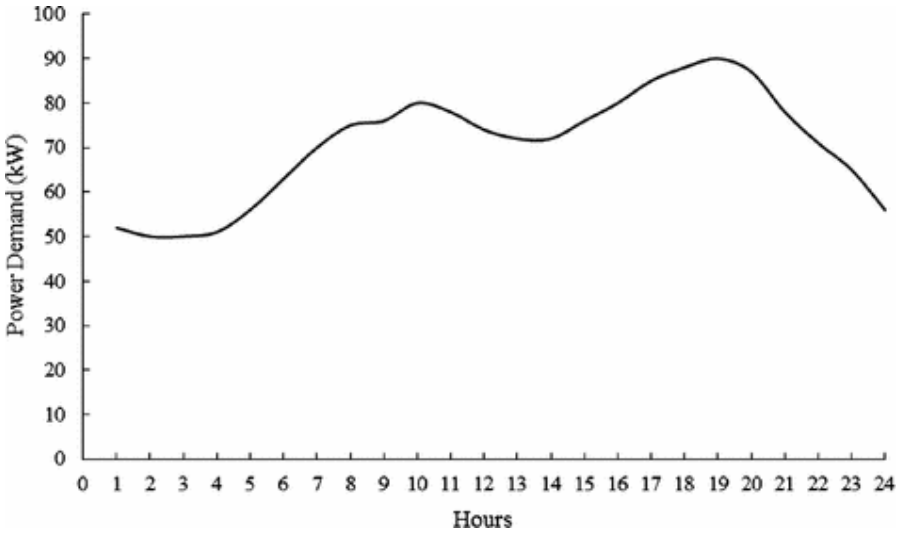

Fig. 10.3

Electrical power demands of a day

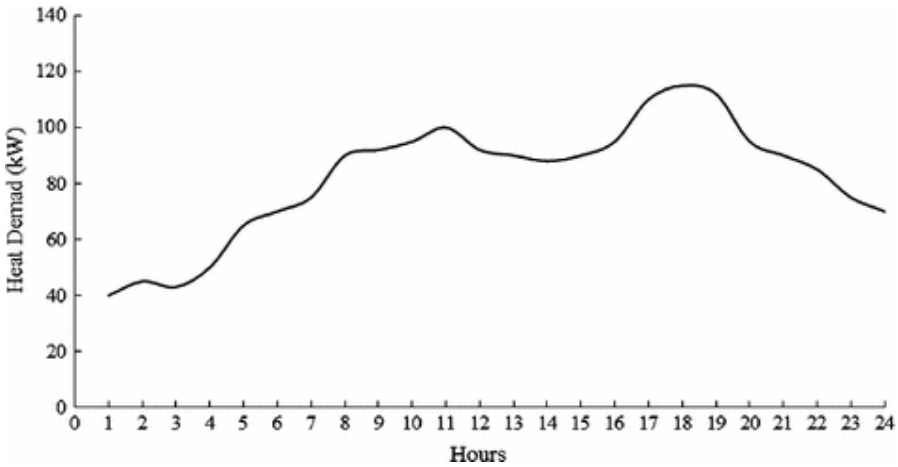

Fig. 10.4

Heat demands of a day

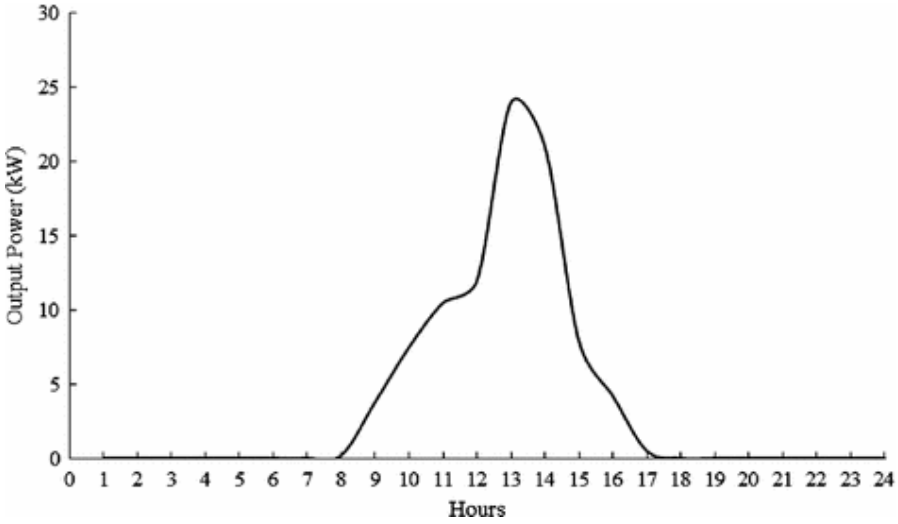

Fig. 10.5

Output power of a photovoltaic

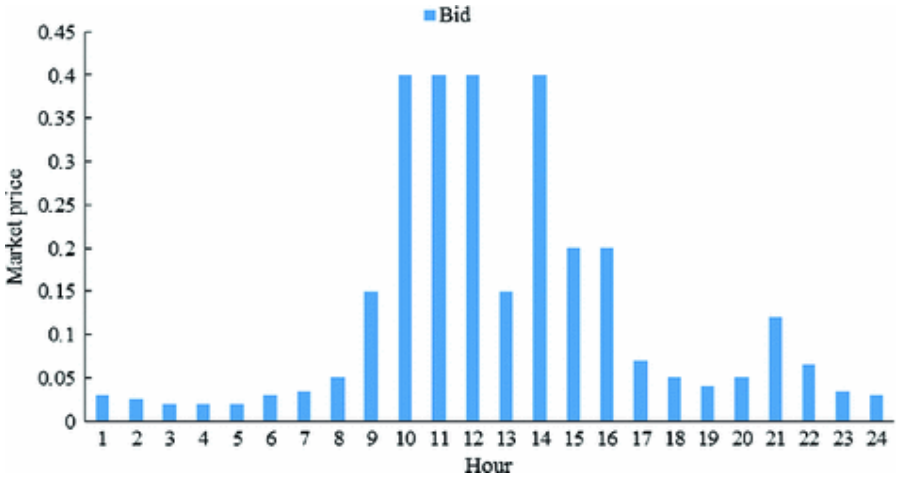

Fig. 10.6

Day ahead market price

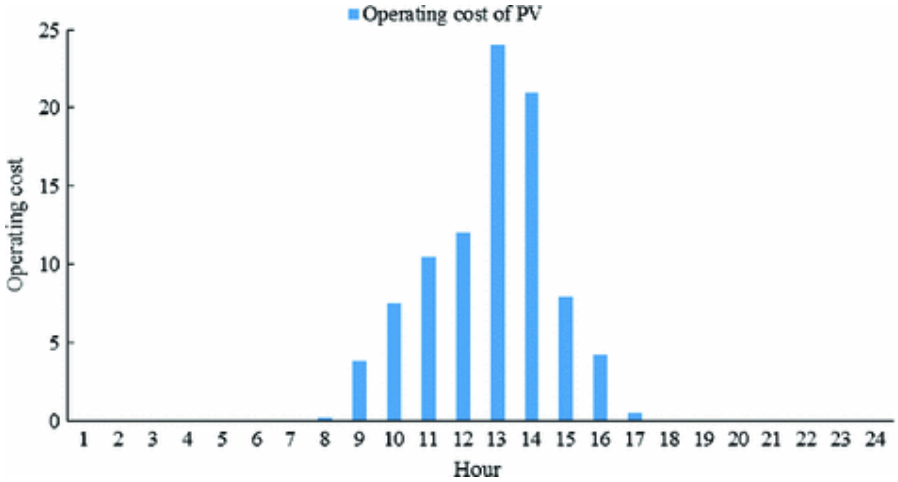

Fig. 10.7

Operating cost of photovoltaic

In this section, we consider CHP produces electrical power and heat at a fixed ratio of 1:1.3. Therefore, electrical power and heat generation scheduling for proposed WE is shown in Figs. 10.8 and 10.9.

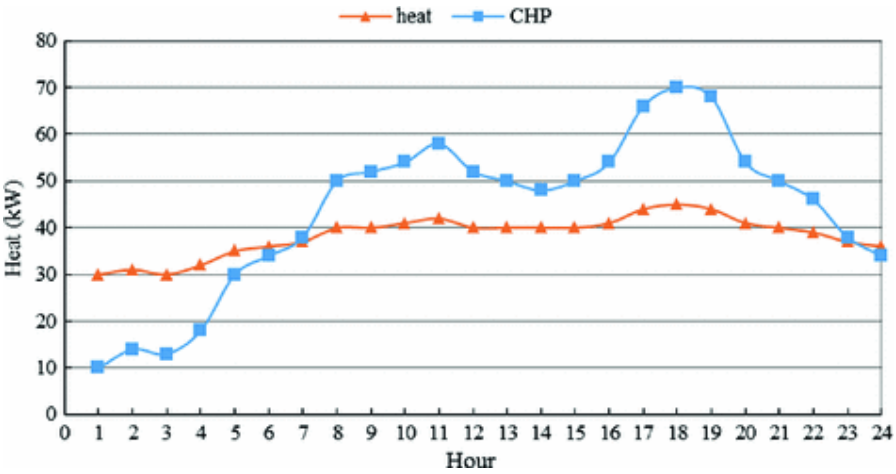

Fig. 10.8

Heat scheduling of We-Energy

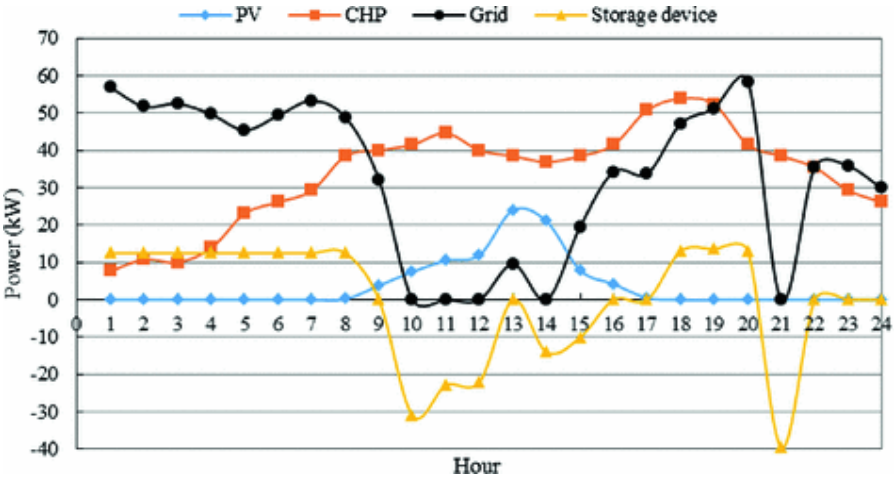

Fig. 10.9

Power scheduling of We-Energy

As can be seen in Figs. 10.8 and 10.9, reinforcement learning algorithm is incorporated into the energy management system and realizes optimal power scheduling. Considering the benefits of renewable sources on reducing pollutant emission, power generated by photovoltaic is consumed at first. In the peak hours, when the market prices are high (9:00–21:00), storage device transfers power to load while less power is obtained from Energy Internet. When the market price is low, power is bought from Energy Internet to fully charge the storage device. Therefore, optimization is realized both environmentally and economically.

10.3 MFEO Energy Internet Scheme with Distributed Reinforcement Learning

10.3.1 Distributed Reinforcement Learning Algorithm

As a branch of reinforcement learning, distributed reinforcement learning is defined as a global structure composed of a plurality of agents. Each agent independently performs some or all of the reinforcement learning tasks while allowing the entire system to achieve the set learning goals. In the current research field, the theory and method of agent research is an important method to solve large-scale and complex information interconnection system, while multi-agents with different structures, distributed performance, dynamic characteristics and large-scale autonomy multi-agent systems can not only achieve individual intelligent optimization, but also make the whole system more responsive, intelligent and social.

At present, the intelligence and fast speed of the system operation are higher and higher, and the information technology and distributed processing power have been developed rapidly. The current study of distributed reinforcement learning model is divided into four types considering the feature of the established system structural differences, including the central reinforcement learning (RLC), independent reinforcement learning (RLI), group reinforcement learning (RLG) and social reinforcement learning (RLS). A multi-agent network based on distributed reinforcement learning is established, and the information of rewards and punishments received by adjacent agent can be obtained by only a small amount of transmission information between each agent. At the same time, in the learning process, according to the overall performance of the system, the application of iterative methods to influence the characteristics of the non-adjacent agent has reached the global optimization objective of the system.

The main feature of the central reinforcement learning model is the learning objective set by the collaborative feature of multi-agent. Based on the classical reinforcement learning algorithm, the optimal interconnection strategy is obtained. In essence, the central reinforcement learning is to make the system distributed optimization problem as a learning objective and to study centralized learning.

The feature of independent reinforcement learning model is that each agent is an independent study module in the whole system, in the process of reinforcement learning, each agent takes the action strategy with the best reward value according to the interaction feedback between itself and the environment. Each agent as an independent agent has no contact with other agents and the agents receive signals from the system's distribution mechanism. This kind of learning mechanism to retain the agent independence of the individual, the optimal target for the whole system is not easy to balance, but also to change the agent number given high degree of freedom, number of convergence problem in the process of reinforcement learning agent. It is suitable for complex system with more agents.

The characteristic of group reinforcement Learning is to set the state of all agents as a state set and define all the actions as interconnected actions, the Q-table of each agent's

reinforcement learning is formed by the correspondence between the interconnected state, interconnected actions and the Q value. Therefore, group reinforcement learning is essentially a group behavior in the learning process. Each agent needs to consider the status and actions of other agents which leads to the state set and action set of the whole system are relatively large and the learning speed is slow. Group reinforcement learning is more suitable for the case of a smaller number of agents. The most essential difference of the mathematical model between the group reinforcement learning and the independent reinforcement learning lies in the definition of multi-agent state sets.

The social reinforcement learning model is an extension of independent reinforcement learning, which combines the independent reinforcement learning model with the social structural system and economic model. The combination of social attributes and economic structure enhances the intelligence of multi-agents to conduct more compatible collaborative work and competition at a time to promote the optimization of the system.

10.3.2 Related Issues of Model-Free Energy Optimization

With the high proportion of new energy terminals such as distributed power, electric vehicles and distributed energy storage components which have diversified features including energy production, storage and consumption in power grids, energy internet is characterized by complex nonlinearity, strong uncertainty and strong coupling. As a result, there may be situations where a complete and accurate model of energy internet could not be established in solving practical problems. The reinforcement learning method is based on the autonomy-based evaluation of the learning process considering the action that feedback from the environment without accurate model. This method of replacing rapid mathematical feedback with fast autonomic feedback has a good application effect for solving complex, uncertain and unstructured environmental system problems. Therefore, this chapter adopts the reinforcement learning to realize the model-free optimization of energy internet.

A collaborative optimization control structure composed of multiple We-Energies is established. The optimal scheduling of energy internet is realized through the distributed parallel optimal calculation of multiple We-Energies.

According to the basic description of distributed reinforcement learning model and the establishment of the structure of the collaborative optimization, the analysis to solve the problem of optimal operation for energy internets is discussed as follows.

1. (1)

   For a large-scale energy internet with distributed power, the architecture of the energy internet is more complex, and the energy subject with the form of We-Energy will increase. On this basis, the classical centralized optimization method for solving the optimal operation is complicated to deal with the larger system.

2. (2)

   In solving complex target function, the traditional planning method, such as Newton method is unable to solve the limitations which are short of flexibility for dealing with such network characteristics and robustness is weak, and it cannot achieve ideal results.

To sum up, the method of distributed parallel coordination optimization is applied to solve the problem of large-scale hybrid optimal operation. This chapter presents a new approach combining with distributed reinforcement learning algorithm to solve hybrid optimal operation problem. According to the Q-learning algorithm which is used to solve the energy management problem has certain feasibility, therefore, the integration of reinforcement learning with the energy Internet structure based on multiple We-Energies interconnection constitutes a distributed reinforcement learning optimization model, but the following key issues need to be considered in model establishment:

1. (1)

   For each We-Energy form the independent optimization individual, we make the energy action space not exceed 50,000 according to the actual situation to avoid the influence of the large action space for the learning process.

2. (2)

   In the process of reinforcement learning, the objective function of each We-Energy and the interconnection action of We-Energy must be considered at the same time. The actions taken from the We-Energy will affect adjacent We-Energy, thus affecting the overall performance.

3. (3)

   The objective function is calculated objectively with the method of parallel computation and serial computation. In the actual energy internet model, the objective function considers the energy loss of each We-Energy while considering the energy loss of the whole energy internet.

10.3.3 Distributed Reinforcement Learning Model for Hybrid Optimal Energy Flow

10.3.3.1 Mathematical Model

The energy internet is a large-scale energy coupling network with electricity, gas and heat. Compared with the power systems, optimal power flow problem of the IMS system becomes more complicated caused by the complex structure. In order to improve efficiency of the method to solve optimization planning problems in energy system, a double-deck multicarrier energy network model is proposed in this chapter.

With interconnections of the IMS system, it is well known that the large-scale systems can be changed as the presence of different political parties. Each party which is defined as “We-Energy” plays the double role of producers and consumers (prosumers) in energy internet. The main body of WE will be the individual, company or community that consists of energy production or storage devices such as distributed generation, energy storage, CCHP and so on. WEs coordinate with each other to guarantee multi energy to reliable transport. In addition, each WE is connected to be considered as the nodes of interconnected multicarrier systems.

Several kinds of WE are presented in this model which consists of energy production device, energy storage device, and user load. Some of them will be connected to district heating plant and gas source. Meanwhile some WE will contain distributed generation such as wind power plant electricity, gas and heat using the coupling way to transmit, while energy hub is defined as an energy carrier to coupling link electricity/gas/heat. Multiple energy inputs will be transformed to other forms of energy as the output of the system. The model of energy hub can be described as

$$\underbrace{\begin{bmatrix} L_e \\ L_g \\ L_h \end{bmatrix}}_{L} = \underbrace{\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} P_e \\ P_g \\ P_h \end{bmatrix}}_{P}, \tag{10.29}$$

where matrix L stands for electricity, gas and heating output, matrix P are input of electricity, gas and heating from the corresponding grid respectively. Matrix C is the conversion coupling matrix which is the mapping from energy input to the energy output. Then, using energy hub to analyze multi energy condition will implement the collaborative energy flow optimization calculation in IMS.

In order to highlight the importance of the environmental economic and secure benefits of IMS, a double nested optimization model is constructed for both minimizing the WE consumption and optimizing the voltage stability of IMS.

The first layer is considered to minimize the energy cost of system operating for each WE at different energy types so as to make the best use of renewable energy. The total cost of operation is sum of the multi-fuel type consumed by the IMS multiplied by the fuel cost. The objective of the total cost study is as follows:

$$F_1 = \min\left(\sum_{t=1}^{T}\sum_{i=1}^{N^G}(a_i^G(P_i^G)^2 + b_i^G P_i^G + c_i^G) + \sum_{t=1}^{T}\sum_{i=1}^{N^G} Q_i^G + \sum_{t=1}^{T}\sum_{i=1}^{N^F} Q_i^F\right),$$

(10.30)

T stands for the number of scheduling time. a_i^G, b_i^G and c_i^G are the coefficients of CHP generators, P_i^G is the active power of generator, N_i^G is the number of CHP generator sets. Also N_i^H is the total number of coal-fired boiler sets. Q_i^G is the heat output from CHP, while Q_i^F is the heat output from traditional coal-fired boiler.

The aim of second layer is to maintain voltage stability of the global system based on voltage constrains. It can be presented as follows:

$$F_2 = \min\frac{1}{n}\sum_{j=1}^{n}|V_j - V^*|,$$

(10.31)

where V_j means the voltage of node j. V^* is the node voltage rating and n is the number of node. Through considering the voltage deviation of each WE, the security of IMS will be improved.

With double-deck optimization structure, these objectives are to be met with such thoroughness and confidence as to be embedded into planning or operation problems in multicarrier energy systems.

In order to achieve this objective, an OEF model must meet the following requirements with the consideration of electrical network, heating network and natural gas network operation constraints. The corresponding constraints are presented in the following.

In electrical network, the active power balance equation and reactive power balance equation connected to the ith bus can be calculated as follows:

$$P_{E,i}^G = P_{E,i}^D + \sum_{j=1}^{N} V_i V_j(G_{E,ij}\cos\theta_{ij} + B_{E,ij}\sin\theta_{ij}),$$

(10.32)

$$Q_{E,i}^G = Q_{E,i}^D + \sum_{j=1}^{N} V_i V_j(G_{E,ij}\sin\theta_{ij} - B_{E,ij}\cos\theta_{ij}),$$

(10.33)

It can be seen that power balance condition is met with that generator power injection is equal to load demand plus losses in IMS.

In heating network, the energy balance in IMS is expressed by

$$P_{H,i}^{S} = P_{H,i}^{D} + \sum_{j=1}^{N_H} P_{H,ij}^{L},$$

(10.34)

It should be noted that the equation satisfies the equality of heating power producers with the load demand plus heat energy users losses.

While in natural gas network, the gas power balance is expressed by

$$P_{G,i}^{S} = P_{G,i}^{D} + \sum_{j=1}^{N_G} P_{G,ij}^{comp} + \sum_{j=1}^{N_G} P_{G,ij}^{L},$$

(10.35)

In this equation, gas power injection is commensurate with the load demand plus the compressor power and natural gas power.

Meanwhile, it also should be subjected to some inequality constraints of the whole network.

10.3.3.2 Proposed RL Method

Considering the energy network structure as well as the optimal energy flow model, the hybrid reinforcement learning algorithm was applied to solve the problem innovatively.

According to the distributed model of IMS, each WE coordinates and interacts with each other to solve the complex problems. Unfortunately, the centralized approach does not fit well with this narrowly defined double-deck model. They tend to enhance the calculation difficulty and require consideration of multiple aspects. Hybrid Reinforcement Learning (HRL) is an effective way to improve the learning efficiency and solve the problem of “dimension disaster”. For the characteristics of the model, the first layer can use the distributed RL for each WE while centralized RL is put into use for the second layer.

1. A.

   Implementation of DRL for IMS

DRL is a method which expands the single-agent RL. In DRL, each agent can obtain rewards from adjacent agent with a little information. The global system use iteration to influence non-adjacent agent so as to optimize the performance of the whole system based on reinforcement learning.

Combining the IMS and the distributed reinforcement learning, the implementation of

optimal energy flow is the generalization of the Markov decision process

Definition 1

A OEF of IMS is a tuple $(S, A_1, \ldots, A_n, P, R_1, \ldots, R_n)$ where n is the number of WEs, S is the discrete set of environment states, $A_i (i = 1, \ldots, n)$ are the sets of actions available to the WEs, yielding the joint action set $A = A_1 + \cdots + A_n$ that every WEs parallel compute for reinforcement learning which is different from the centralized algorithm. $P : S \times A \times S \to [0, 1]$ means the state transition probability. $R_i (i = 1, \ldots, n)$ are the direct reward functions of each WE.

In the OEF algorithm, through taking into consideration the objects and constrains, control variables in IMS are output power of each WE by adjusting the energy storage equipment and so on. The vector of state variables can be defined as follows:
$u = [\theta, |V|, \dot{m}, v_g, T_{s,load}, T_{r,load}]^T$,
(10.36)
where π is the vector of pressure. θ, $|V|$ are vectors of unknown angles and magnitudes of voltage. $\dot{m}$ stands for the vector of pipeline mass flows and T^S, T^r are vectors of supply and return temperatures.

For DRL, in order to achieve the goal of OEF, rewards in reinforcement learning should be combined with the objective function and constraint conditions.

The local immediate reward value R of each WE need satisfy the constraint conditions to ensure the validity of the calculation results for each subsystem. Each WE will obtain the optimal strategy by maximizing reward function values.

Definition 2

The local reward for WE is defined as

$$R_{i,0}^K = \begin{cases} 0, & if\ constraints\ are\ violated \\ \frac{1}{F_2^K(X)}, & otherwise \end{cases},$$

(10.37)

Every WE will check control variables through connected transmission lines to see whether they meet the corresponding boundary conditions. If all of constraints are satisfied, the local reward signal will be set to the negative objective function. Otherwise, it will be zero. The local rewards are applied to each WE to guide action strategy.

The aim of OEF is to seek a best strategy from the action space, so that the global reward is presented as an average value of summation of local rewards from each WE.

$$R^K = \frac{1}{n}\sum_{i=1}^{n} R_{i,0}^K$$

(10.38)

The structure of DRL is shown below as Fig. 10.10. Through the multi-energy flow calculation for IMS, the running status of each WE will be acquired. Afterwards, the local reward of WE will be obtained from the information interaction with environment according to Definition 2. Then, the global reward will be updated with the local reward if all the information is available in information fusion unit. The Q-learning unit will operate based on RL iteration rule to find the optimal strategy. Meanwhile, combined with the prior knowledge for initial action set, the learning state and learning efficiency could be improved.

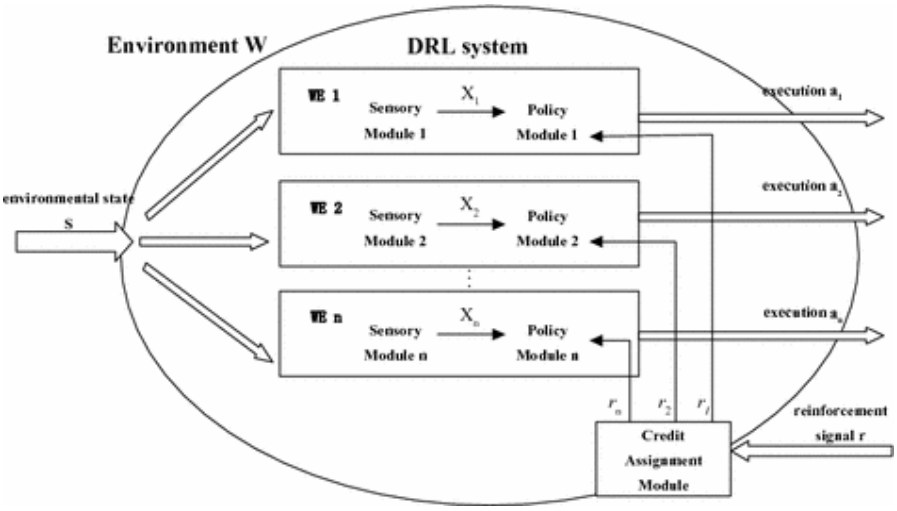

Fig. 10.10

Distributed reinforcement learning optimal energy flow structure

1. B.

   Implementation of RL for each WE

In order to minimum energy cost of each WE, RL algorithm is utilized to search for the optimal strategy which considers the operating state of internal equipment from WE.

RL is a method for single-agent which can be achieved by Markov decision process (MDP) modeling. A four-tuple (S, A, P, r) is defined to express the approach, where S is the limited environmental state space, A stands for the limited action sets. $P(s, a, s')$ stands for the probability that state S transfers to the state s' under control action a and $r(s, a, s')$ is the immediate reinforcement signal given by environment when the state s transfers to the state s' after action a. In an optimal energy flow algorithm of IMS, Q-learning is used to evaluate state of system after an action without an environmental

model.

The function Q refers to the optimal reward discount value of WE with action a at state s, denoted by

$$Q(s,a) = r(s,a,s') + \gamma \max_{a \in A} Q(s',a), \tag{10.39}$$

where s, s' are the current state and the state of the next moment. γ is the discount factor which the value of it often is $0 < \gamma < 1$.

Definition 3

Given the RL iteration rule for WE, the Q-learning operation is defined as

$$Q_0(s,a) = 0 \quad for \text{ all } s \in S, u \in U, \tag{10.40}$$

$$Q_t(s,a) = \begin{cases} Q_t(s,a) & if\ s \neq s_t\ or\ a \neq a_t \\ (1-\alpha_t)Q_{t-1}(s,a) + \alpha_t[r_t(s,a) + \gamma \max Q_{t-1}(s',a)] & if\ s = s_t\ and\ a = a_t \end{cases}, \tag{10.41}$$

where α_t is the learning factor with $0 < \alpha_t < 1$,α_t indicates the proportion of update part in Q value. If its corresponding state s or the action a_t has no samples in state space, the value of this pair will not update. Otherwise, the update rule will be used to approach its value.

Definition 4

The reward for WE is defined as

$$r_{i,0}^k = \begin{cases} 0, & if\ constraints\ are\ violated \\ \frac{1}{F_1^k(X)}, & otherwise \end{cases}, \tag{10.42}$$

The action sets are made up of equipment actions in WE and the reward is designed as the reciprocal of reward function. It can be seen that the less energy consumption, the more rewarding under the condition of satisfying the constraints. Greedy strategy is adopted for always choosing the highest Q-value movement in the current state.

In the second layer OEF algorithm, each WE uses two types of equipment to realize operation control strategy, conventional energy components and the energy coupling unit. The first one is power source and the second one is conversion units of each We-Energy such as electrical transformers, power electronic devices, gas compressors, heat exchangers or boilers and others.

RL method only needs to respond to the control effect of assessment information on the

basis of the above equipment adjustment. Using RL for each WE to implement energy optimization of energy cost respectively has a higher robustness.

10.3.4 Simulation and Results

In this section, we apply the proposed HRL algorithm to the OEF problem for IMS with the structure of nine interconnection WE which is shown in Fig. 10.11. All the parameters are expressed in per-unit value. In power system, the apparent power per-unit value is 100 MVA, voltage basic value is defined in 1 kV and the scope of bus voltage is [0.9, 1.1]. Power basic value in natural gas network is 100 MW and the pressure is set for 10 bar. Thermal power is 100 MW and the temperature basic value is 100 °C, while time delay of heating pipe network is set for 1 h. Table 10.4 shows the efficiencies of devices for each WE (Fig. 10.12).

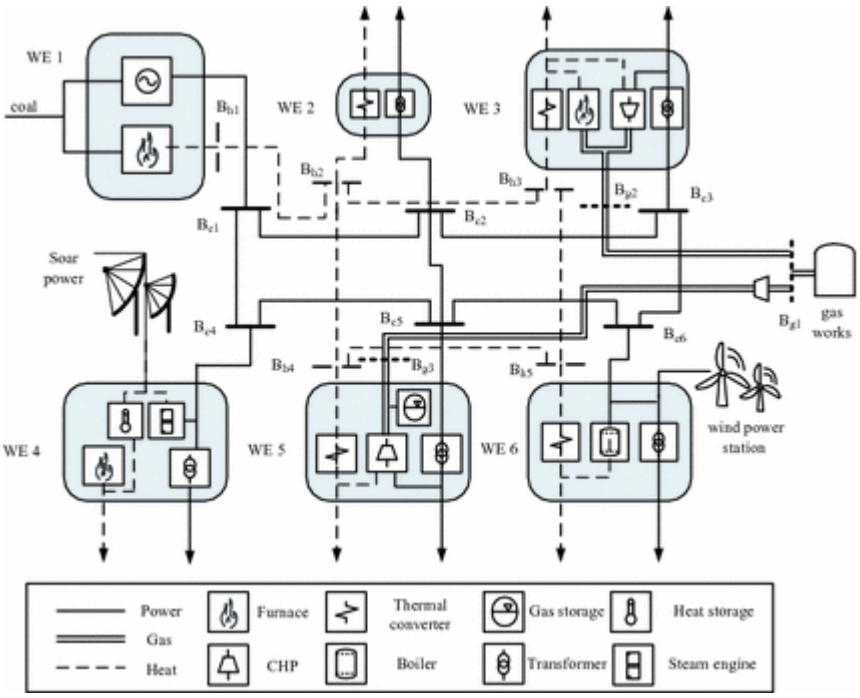

Fig. 10.11

Energy internet simulation diagram

Table 10.4

Test network conversion efficiency of energy equipment

Devices	Efficiency	Capacity
The transformer in WE 1	$\eta_{ge} = 0.3$	2000
The furnace in WE 1	$\eta_{gh} = 0.85$	700
The furnace in WE 3	$\eta_{gh} = 0.7$	700
The CHP in WE 3,5	$\eta_{ge} = 0.35, \eta_{gh} = 0.45$	700
The furnace in WE 4	$\eta_{gh} = 0.9$	300
The heat storage in WE 4	$\eta_h^{ch} = 0.9, \eta_h^{dis} = 0.9$	700
The boiler in WE 4	$\eta_{gh} = 0.9$	150
The gas storage in WE 4	$\eta_g^{ch} = 0.9, \eta_g^{dis} = 0.9$	300

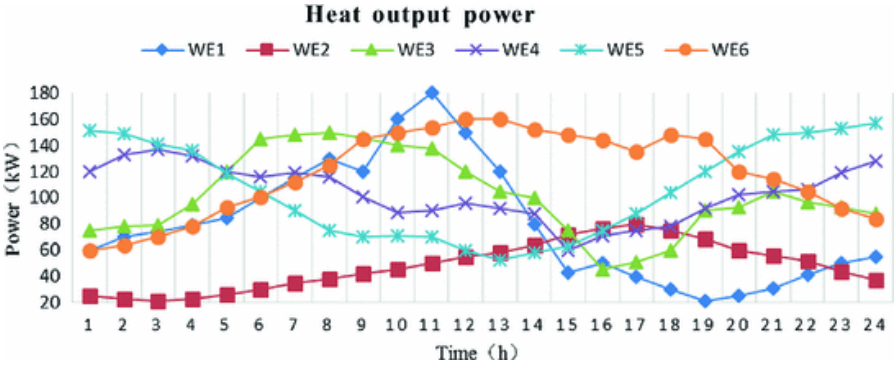

Fig. 10.12

The heat output power of We-Energies

In layer 1, the curve of the heat output power given by WEs is shown in Fig. 10.14. In order to minimum voltage deviation in IMS, we should adjust the output of each WE under the system constrains. According to maximum limits of each WE, the power output values of each device is divided into 10 grades on the basis of their maximum limits with 20% fluctuating value.

The action set size of DRL in IMS is $10 \times 6 = 60$ which is smaller than the size of centralized RL in 10^6. Figure 10.13 shows the optimal power flow result of the WE in the first layer. Figure 10.14 shows the learning step of layer 1 for IMS. The reward is the

reciprocal of the average voltage deviation. The learning factor α of iteration rule is set for 0.85 and the discount factor γ is 0.2. It can be seen in the chapter that the reward converges to the reward of 25 after 1680 steps. Meanwhile, if we modify the learning factor α to be 0.6, the results are presented in Fig. 10.15. Compared with Fig. 10.14, the learning steps rise to 1960 steps. It shows that learning factor affects the convergence speed that the higher the learning factor, the better the convergence speed.

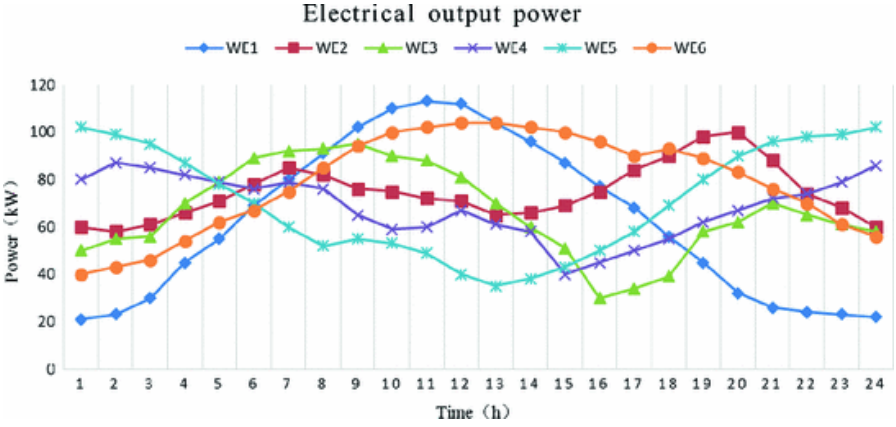

Fig. 10.13

The electrical output power of We-Energies

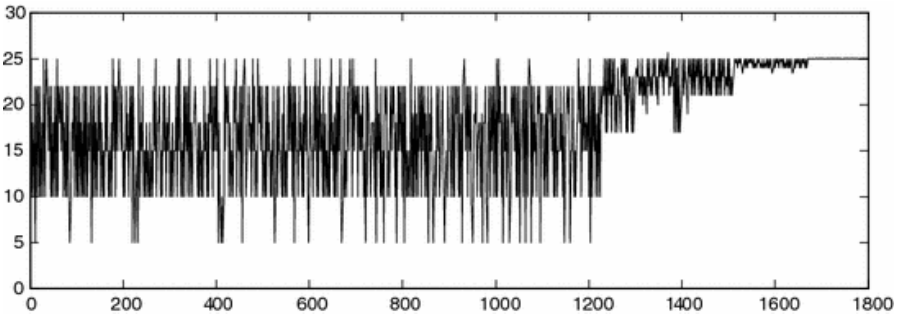

Fig. 10.14

Distributed RL process at $\alpha = 0.85, \gamma = 0.2$

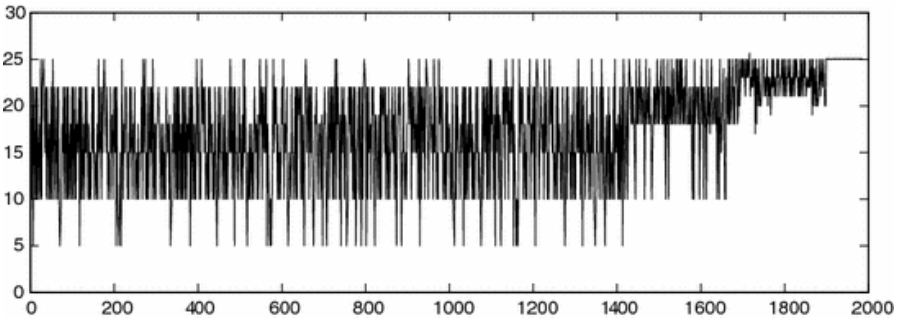

Fig. 10.15

Distributed RL process at $\alpha = 0.6, \gamma = 0.2$

Then, compared with voltage deviation before optimization, the indicator of DRL shows that voltage deviation has reduced from 0.045 to 0.038.

In layer 2, According to the devices in each WE, all kinds of control equipment are considered as control variable including generator, boil, CHP, electric boiler, thermal storage devices and gas storage device. In order to set each WE action space with unified standard, the action variables will be discretized in this chapter.

The output of generator is divided into 5 grades according to its limits. The power input values of boil will be in 10 grades based on its capacity. The same as boil, the power input of CHP is divided into 5 grades. The power input of electric boil is in 10 grades on average based on their maximum limits. The power input values of thermal storage devices are in 3 grades and the power input values of gas storage device is in 5 grades

As we can see that there are 5 control variables in WE1. With the grades defined above, the number of actions for WE1 can be calculated as $5 \times 10 \times 5 \times 10 \times 3 = 7500$. WE2 can be calculated as $3 \times 10 = 30$. Other WE follows a similar pattern to WE1.

Through the RL for each WE, the optimal energy cost has reduced 21%-26% from the objective function (Fig. 10.16).

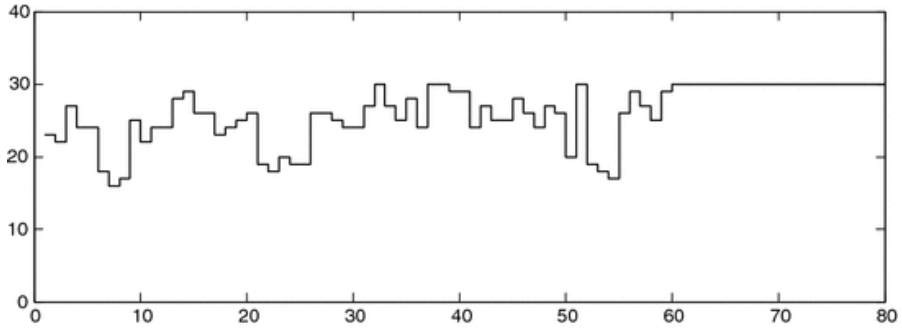

Fig. 10.16

DRL optimization process

Based on the system simulation result for Modified 6-bus, hybrid reinforcement learning is verified to be satisfied with the optimal energy flow in interconnected multicarrier systems.

10.4 Conclusion

This chapter proposes a model free optimal model and presents the RL algorithm that can drive these agents to parallel learn behaviors. The model free optimal model not only reduces the loss of the system but also improves the safety and reliability, especially considering the accuracy of information. In addition, using the RL algorithm for this optimal energy model has some advantages. The method of utilizing each WE to undertake the task independently to reach the coordinated system is suitable for hierarchical control mode of energy internet. It has the ability to solve the Inadaptability

of multi-objective function and real-time performance of large-scale network computing.

References

1. A. Shabanpour-Haghighi, A.R. Seifi, Energy flow optimization in multicarrier systems. IEEE Trans. Ind. Inf. **11**(5), 1067–1077 (2015)
CrossRef (https://doi.org/10.1109/TII.2015.2462316)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Energy%20flow%20optimization%20in%20multicarrier%20systems&author=A.%20Shabanpour-Haghighi&author=AR.%20Seifi&journal=IEEE%20Trans.%20Ind.%20Inf.&volume=11&issue=5&pages=1067-1077&publication_year=2015)

2. J. Hu, Q. Sun, T. Fei, A game-theoretic pricing model for energy internet in day-ahead trading market considering distributed generations uncertainty. in *IEEE Symposium Series on Computational Intelligence* (2016)
Google Scholar (https://scholar.google.com/scholar?q=J.%20Hu%2C%20Q.%20Sun%2C%20T.%20Fei%2C%20A%20game-theoretic%20pricing%20model%20for%20energy%20internet%20in%20day-ahead%20trading%20market%20considering%20distributed%20generations%20uncertainty.%20in%20IEEE%20Symposium%20Series%20on%20Computational%20Intelligence%20%282016%29)

3. R.Z. R´ıos-Mercado, C. Borraz-S´anchez, Optimization problems in natural gas transportation systems: a state-of-the-art review. Appl. Energy **147**, 536–555 (2015)
Google Scholar (https://scholar.google.com/scholar?q=R.Z.%20R%C2%B4%C4%B1os-Mercado%2C%20C.%20Borraz-S%C2%B4anchez%2C%20Optimization%20problems%20in%20natural%20gas%20transportation%20systems%3A%20a%20state-of-the-art%20review.%20Appl.%20Energy%20147%2C%20536%E2%80%93555%20%282015%29)

4. C.M. Correa-Posada.; P. Sanchez-Martin, Integrated power and natural gas model for energy adequacy in short-term operation. IEEE Trans. Power Syst., **30(**6), 3347–3355 (2015)
CrossRef (https://doi.org/10.1109/TPWRS.2014.2372013)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Integrated%20Power%20and%20Natural%20Gas%20Model%20for%20Energy%20Adequacy%20in%20Short-Term%20Operation&author=Carlos%20M..%20Correa-Posada&author=Pedro.%20Sanchez-Martin&journal=IEEE%20Transactions%20on%20Power%20Systems&volume=30&issue=6&pages=3347-3355&publication_year=2015)

5. A. Alabdulwahab, A. Abusorrah, X. Zhang, X, Coordination of interdependent natural gas and electricity infrastructures for firming the variability of wind energy in stochastic day-ahead scheduling. IEEE Trans. Sustain. Energy **6**(2), 606–615 (2015)
CrossRef (https://doi.org/10.1109/TSTE.2015.2399855)

Google Scholar (http://scholar.google.com/scholar_lookup?title=X%2C%20Coordination%20of%20interdependent%20natural%20gas%20and%20electricity%20infrastructures%20for%20firming%20the%20variability%20of%20wind%20energy%20in%20stochastic%20day-ahead%20scheduling&author=A.%20Alabdulwahab&author=A.%20Abusorrah&author=X.%20Zhang&journal=IEEE%20Trans.%20Sustain.%20Energy&volume=6&issue=2&pages=606-615&publication_year=2015)

6. M. Chaudry, N. Jenkins, M. Qadrdan et al., Combined gas and electricity network expansion planning. Appl. Energy **113**(6), 1171–1187 (2014)
CrossRef (https://doi.org/10.1016/j.apenergy.2013.08.071)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Combined%20gas%20and%20electricity%20network%20expansion%20planning&author=M.%20Chaudry&author=N.%20Jenkins&author=M.%20Qadrdan&journal=Appl.%20Energy&volume=113&issue=6&pages=1171-1187&publication_year=2014)

7. X. Xu, H. Jia, H.D. Chiang et al., Dynamic modeling and interaction of hybrid natural gas and electricity supply system in microgrid. IEEE Trans. Power Syst. **30**(3), 1212–1221 (2015)
CrossRef (https://doi.org/10.1109/TPWRS.2014.2343021)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Dynamic%20modeling%20and%20interaction%20of%20hybrid%20natural%20gas%20and%20electricity%20supply%20system%20in%20microgrid&author=X.%20Xu&author=H.%20Jia&author=HD.%20Chiang&journal=IEEE%20Trans.%20Power%20Syst.&volume=30&issue=3&pages=1212-1221&publication_year=2015)

8. X. Zhang, M. Shahidehpour, A. Alabdulwahab et al., Optimal expansion planning of energy hub with multiple energy infrastructures. IEEE Trans. Smart Grid **6**(5), 2302–2311 (2015)
CrossRef (https://doi.org/10.1109/TSG.2015.2390640)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Optimal%20expansion%20planning%20of%20energy%20hub%20with%20multiple%20energy%20infrastructures&author=X.%20Zhang&author=M.%20Shahidehpour&author=A.%20Alabdulwahab&journal=IEEE%20Trans.%20Smart%20Grid&volume=6&issue=5&pages=2302-2311&publication_year=2015)

9. C. Liu, M. Shahidehpour, J. Wang, Coordinated scheduling of electricity and natural gas infrastructures with a transient model for natural gas flow. *Chaos* **21(**21) (2011)
CrossRef (https://doi.org/10.1063/1.3600761)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Coordinated%20scheduling%20of%20electricity%20and%20natural%20gas%20infrastructures%20with%20a%20transient%20model%20for%20natural%20gas%20flow&author=Cong.%20Liu&author=Mohammad.%20Shahidehpour&author=Jianhui.%20Wang&journal=Chaos%3A%20An%20Interdisciplinary%20Journal%20of%20Nonlinear%20Science&volume=21&issue=2&pages=025102&publication_year=2011)

10. X. Zhang, G.G. Karady, S.T. Ariaratnam, Optimal allocation of CHP-based distributed generation on urban energy distribution networks. IEEE Trans. Sustain. Energy **5**(5), 246–253 (2014)
CrossRef (https://doi.org/10.1109/TSTE.2013.2278693)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Optimal%20allocation%20of%20CHP-based%20distributed%20generation%20on%20urban%20energy%20distribution%20networks&author=X.%20Zhang&author=GG.%20Karady&author=ST.%20Ariaratnam&journal=IEEE%20Trans.%20Sustain.%20Energy&volume=5&issue=5&pages=246-253&publication_year=2014)

11. X. Fei, M. Yong, Combined electricity-heat operation system containing large capacity thermal energy storage. Proc. CSEE **34**(29), 5063–5072 (2014)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Combined%20electricity-heat%20operation%20system%20containing%20large%20capacity%20thermal%20energy%20storage&author=X.%20Fei&author=M.%20Yong&journal=Proc.%20CSEE&volume=34&issue=29&pages=5063-5072&publication_year=2014)

12. G. Zepeng, K. Chongqing, Operation optimization of integrated power and heat energy systems and the benefit on wind power accommodation considering heating network constraints. Proc. CSEE **35**(14), 3596–3604 (2015)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Operation%20optimization%20of%20integrated%20power%20and%20heat%20energy%20systems%20and%20the%20benefit%20on%20wind%20power%20accommodation%20considering%20heating%20network%20constraints&author=G.%20Zepeng&author=K.%20Chongqing&journal=Proc.%20CSEE&volume=35&issue=14&pages=3596-3604&publication_year=2015)

13. Z. Pan, H. Sun, Q. Guo, Energy Internet oriented security analysis method for multi-energy flow. Power Syst. Technol. **40**, 1627–1634 (2016)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Energy%20Internet%20oriented%20security%20analysis%20method%20for%20multi-energy%20flow&author=Z.%20Pan&author=H.%20Sun&author=Q.%20Guo&journal=Power%20Syst.%20Technol.&volume=40&pages=1627-1634&publication_year=2016)

14. Q. Huang, M.L. Crow, G.T. Heydt, J.P. Zheng, S.J. Dale, The future renewable electric energy delivery and management (FREEDM) system: the energy internet. Proc. IEEE **99**(1), 133–148 (2011)
CrossRef (https://doi.org/10.1109/JPROC.2010.2081330)
Google Scholar (http://scholar.google.com/scholar_lookup?title=The%20future%20renewable%20electric%20energy%20delivery%20and%20management%20%28FREEDM%29%20system%3A%20the%20energy%20internet&author=Q.%20Huang&author=ML.%20Crow&author=GT.%20Heydt&author=JP.%20Zheng&author=SJ.%20Dale&journal=Proc.%20IEEE&volume=99&issue=1&pages=133-148&publication_year=2011)

15. C. Chen, S. Duan, T. Cai, B. Liu, G. Hu, Smart energy management system for optimal microgrid economic operation. IET Renew. Power Gener. **5(**3), 258–267, (2011)
CrossRef (https://doi.org/10.1049/iet-rpg.2010.0052)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Smart%20energy%20management%20system%20for%20optimal%20microgrid%20economic%20operation&author=C..%20Chen&author=S..%20Duan&author=T..%20Cai&author=B..%20Liu&author=G..%20Hu&journal=IET%20Renewable%20Power%20Generation&volume=5&issue=3&pages=258&publication_year=2011)

16. X. Ou, Y. Shen, Z. Zeng, G. Zhang, L. Wang, Cost minimization online energy management for microgrids with power and thermal storages. in *24th International Conference on Computer Communication and Networks (ICCCN), Las Vegas, NV* (2015), pp. 1–6
Google Scholar (https://scholar.google.com/scholar?q=X.%20Ou%2C%20Y.%20Shen%2C%20Z.%20Zeng%2C%20G.%20Zhang%2C%20L.%20Wang%2C%20Cost%20minimization%20online%20energy%20management%20for%20microgrids%20with%20power%20and%20thermal%20storages.%20in%2024th%20International%20Conference%20on%20Computer%20Communication%20and%20Networks%20%28ICCCN%29%2C%20Las%20Vegas%2C%20NV%20%282015%29%2C%20pp.%201%E2%80%936)

17. M. Motevasel, A.R. Seifi, T. Niknam, Multi-objective energy management of CHP (combined heat and power)-based micro-grid. Energy **51**, 123–136 (2013)
CrossRef (https://doi.org/10.1016/j.energy.2012.11.035)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Multi-objective%20energy%20management%20of%20CHP%20%28combined%20heat%20and%20power%29-based%20micro-grid&author=M.%20Motevasel&author=AR.%20Seifi&author=T.%20Niknam&journal=Energy&volume=51&pages=123-136&publication_year=2013)

18. L. Ma, N. Liu, J. Zhang, W. Tushar, C. Yuen, Energy management for joint operation of CHP and PV prosumers inside a grid-connected microgrid: a game theoretic approach. IEEE Trans. Ind. Inf. **12**(5), 1930–1942 (2016)
CrossRef (https://doi.org/10.1109/TII.2016.2578184)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Energy%20management%20for%20joint%20operation%20of%20CHP%20and%20PV%20prosumers%20inside%20a%20grid-connected%20microgrid%3A%20a%20game%20theoretic%20approach&author=L.%20Ma&author=N.%20Liu&author=J.%20Zhang&author=W.%20Tushar&author=C.%20Yuen&journal=IEEE%20Trans.%20Ind.%20Inf.&volume=12&issue=5&pages=1930-1942&publication_year=2016)

19. A. Shabanpour-Haghighi, A. R. Seifi, Energy flow optimization in multicarrier systems. *IEEE Trans. Ind. Inf.* **11(**5) (2015)
CrossRef (https://doi.org/10.1109/TII.2015.2462316)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Energy%20Flow%20Optimization%20in%20Multicarrier

%20Systems&author=Amin.%20Shabanpour-Haghighi&author=Ali%20Reza.%20Seifi&journal=IEEE%20Transactions%20on%20Industrial%20Informatics&volume=11&issue=5&pages=1067-1077&publication_year=2015)

20. T. Krause, G. Andersson, K. Frohlich, A. Vaccaro, Multiple-energy carriers:modeling of production, delivery, and consumption. *Proc IEEE,* **99,** 15–27 (2011)
CrossRef (https://doi.org/10.1109/JPROC.2010.2083610)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Multiple-Energy%20Carriers%3A%20Modeling%20of%20Production%2C%20Delivery%2C%20and%20Consumption&author=T.%20Krause&author=G%C3%B6ran.%20Andersson&author=K.%20Fr%C3%B6hlich&author=A.%20Vaccaro&journal=Proceedings%20of%20the%20IEEE&volume=99&issue=1&pages=15-27&publication_year=2011)

21. B. Stott, J.L. Marinho, OAlsac, Review of linear programming applied to power system rescheduling. in *Power Industry Computer Applications Conference* (1979), pp. 142–154
Google Scholar (https://scholar.google.com/scholar?q=B.%20Stott%2C%20J.L.%20Marinho%2C%20OAlsac%2C%20Review%20of%20linear%20programming%20applied%20to%20power%20system%20rescheduling.%20in%20Power%20Industry%20Computer%20Applications%20Conference%20%281979%29%2C%20pp.%20142%E2%80%93154)

22. K.A. Clements, P.W. Davis., K.D. Frey, An interior point algorithm for weighted least absolute value power system state Eestimation (1991)
Google Scholar (https://scholar.google.com/scholar?q=K.A.%20Clements%2C%20P.W.%20Davis.%2C%20K.D.%20Frey%2C%20An%20interior%20point%20algorithm%20for%20weighted%20least%20absolute%20value%20power%20system%20state%20Eestimation%20%281991%29)

23. R.Z. Ríos-Mercado, C. Borraz-Sánchez, Optimization problems in natural gas transportation systems: a state-of-the-art review. Appl. Energy **147**, 536–555 (2015)
CrossRef (https://doi.org/10.1016/j.apenergy.2015.03.017)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Optimization%20problems%20in%20natural%20gas%20transportation%20systems%3A%20a%20state-of-the-art%20review&author=RZ.%20R%C3%ADos-Mercado&author=C.%20Borraz-S%C3%A1nchez&journal=Appl.%20Energy&volume=147&pages=536-555&publication_year=2015)

24. Q. Sun, R. Han, H. Zhang et al., A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet. IEEE Trans. Smart Grid **6**(6), 3006–3019 (2015)
CrossRef (https://doi.org/10.1109/TSG.2015.2412779)
Google Scholar (http://scholar.google.com/scholar_lookup?title=A%20multiagent-based%20consensus%20algorithm%20for%20distributed%20coordinated%20control%20of%20distributed%20generators%20in%20the%20energy%20internet&author=Q.%20Sun&author=R.%20Han&author=H.%20Zhang&

journal=IEEE%20Trans.%20Smart%20Grid&volume=6&issue=6&pages=3006-3019&publication_year=2015)

25. Y. Xu, W. Zhang, W. Liu et al., Multiagent-based reinforcement learning for optimal reactive power dispatch. IEEE Trans. Syst. Man Cybern. Part C **42**(6), 1742–1751 (2012)
CrossRef (https://doi.org/10.1109/TSMCC.2012.2218596)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Multiagent-based%20reinforcement%20learning%20for%20optimal%20reactive%20power%20dispatch&author=Y.%20Xu&author=W.%20Zhang&author=W.%20Liu&journal=IEEE%20Trans.%20Syst.%20Man%20Cybern.%20Part%20C&volume=42&issue=6&pages=1742-1751&publication_year=2012)

26. K. Doya, Reinforcement learning in continuous time and space. Neural Comput. **12**(1), 219–245 (2000)
CrossRef (https://doi.org/10.1162/089976600300015961)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Reinforcement%20learning%20in%20continuous%20time%20and%20space&author=K.%20Doya&journal=Neural%20Comput.&volume=12&issue=1&pages=219-245&publication_year=2000)

27. Z. Chen, S. Jagannathan, Generalized Jamilton–Jacobi–Bellman formulation -based neural network control of affine nonlinear discrete-time systems. *IEEE Trans. Neural Netw.* **19(**1), 90–106 (2008)
CrossRef (https://doi.org/10.1109/TNN.2007.900227)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Generalized%20Hamilton%E2%80%93Jacobi%E2%80%93Bellman%20Formulation%20-Based%20Neural%20Network%20Control%20of%20Affine%20Nonlinear%20Discrete-Time%20Systems&author=.%20Zheng%20Chen&author=S..%20Jagannathan&journal=IEEE%20Transactions%20on%20Neural%20Networks&volume=19&issue=1&pages=90-106&publication_year=2008)

28. T. Dierks, B. TThumati, S. Jagannathan, Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence. Neural Netw. **22**(5–6), 851–860 (2009)
CrossRef (https://doi.org/10.1016/j.neunet.2009.06.014)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Optimal%20control%20of%20unknown%20affine%20nonlinear%20discrete-time%20systems%20using%20offline-trained%20neural%20networks%20with%20proof%20of%20convergence&author=T.%20Dierks&author=B.%20TThumati&author=S.%20Jagannathan&journal=Neural%20Netw.&volume=22&issue=5%E2%80%936&pages=851-860&publication_year=2009)

29. D. Liu, Q. Wei, Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. IEEE Trans. Neural Netw. Learn. Syst. **25(**3), 621–634 (2014)
Google Scholar (https://scholar.google.com/scholar?q=D.%20Liu%2C%20Q.%20Wei

%2C%20Policy%20iteration%20adaptive%20dynamic%20programming%20algorithm%20for%20discrete-time%20nonlinear%20systems.%20IEEE%20Trans.%20Neural%20Netw.%20Learn.%20Syst.%2025%283%29%2C%20621%E2%80%93634%20%282014%29)

30. H. Modares, F.L. Lewis, Naghibi-Sistani M B. Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems. Automatica **50**(1), 193–202 (2014) MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3157741) CrossRef (https://doi.org/10.1016/j.automatica.2013.09.043) Google Scholar (http://scholar.google.com/scholar_lookup?title=Naghibi-Sistani%20M%20B.%20Integral%20reinforcement%20learning%20and%20experience%20replay%20for%20adaptive%20optimal%20control%20of%20partially-unknown%20constrained-input%20continuous-time%20systems&author=H.%20Modares&author=FL.%20Lewis&journal=Automatica&volume=50&issue=1&pages=193-202&publication_year=2014)

31. H. Zhang, L. Cui, X. Zhang et al., Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. IEEE Trans. Neural Netw. **22**(12), 2226–2236 (2011) CrossRef (https://doi.org/10.1109/TNN.2011.2168538) Google Scholar (http://scholar.google.com/scholar_lookup?title=Data-driven%20robust%20approximate%20optimal%20tracking%20control%20for%20unknown%20general%20nonlinear%20systems%20using%20adaptive%20dynamic%20programming%20method&author=H.%20Zhang&author=L.%20Cui&author=X.%20Zhang&journal=IEEE%20Trans.%20Neural%20Netw.&volume=22&issue=12&pages=2226-2236&publication_year=2011)

32. R. Kamalapurkar, H. Dinh, S. Bhasin et al., Approximate optimal trajectory tracking for continuous-time nonlinear systems. Automatica **51**, 40–48 (2015) MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3284751) CrossRef (https://doi.org/10.1016/j.automatica.2014.10.103) Google Scholar (http://scholar.google.com/scholar_lookup?title=Approximate%20optimal%20trajectory%20tracking%20for%20continuous-time%20nonlinear%20systems&author=R.%20Kamalapurkar&author=H.%20Dinh&author=S.%20Bhasin&journal=Automatica&volume=51&pages=40-48&publication_year=2015)

33. B. Kiumarsi, F.L. Lewis, H. Modares et al., Reinforcement [formula omitted]-learning for optimal tracking control of linear discrete-time systems with unknown dynamics. Automatica **50**(14), 1167–1175 (2014) MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3191387) CrossRef (https://doi.org/10.1016/j.automatica.2014.02.015) Google Scholar (http://scholar.google.com/scholar_lookup?title=Reinforcement%20%5Bformula%20omitted%5D-learning%20for%20optimal%20tracking%20control%20of%20linear%20discrete-time%20systems%20with%20unknown%20dynamics&author=B.%20Kiumarsi&author=FL.%20Lewis&author=H.%20Modares&journal=Automatica&volume=50&issue=14&pages=1167-1175&publication_year=2014)

34. H. Zhang, Q. Wei, Y. Luo, A novel infinite-time optimal tracking control scheme

for a class of discrete-time nonlinear systems via the Greedy HDP iteration algorithm. IEEE Trans. Syst. Man Cybern. Part B Cybern. **38**(4), 937–942 (2008)
CrossRef (https://doi.org/10.1109/TSMCB.2008.920269)
Google Scholar (http://scholar.google.com/scholar_lookup?title=A%20novel%20infinite-time%20optimal%20tracking%20control%20scheme%20for%20a%20class%20of%20discrete-time%20nonlinear%20systems%20via%20the%20Greedy%20HDP%20iteration%20algorithm&author=H.%20Zhang&author=Q.%20Wei&author=Y.%20Luo&journal=IEEE%20Trans.%20Syst.%20Man%20Cybern.%20Part%20B%20Cybern.&volume=38&issue=4&pages=937-942&publication_year=2008)

35. R. Kamalapurkar, L. Andrews, P. Walters et al., Model-based reinforcement learning for infinite-horizon approximate optimal tracking. in *Decision and Control*. (IEEE, 2015), pp 5083–5088
Google Scholar (https://scholar.google.com/scholar?q=R.%20Kamalapurkar%2C%20L.%20Andrews%2C%20P.%20Walters%20et%20al.%2C%20Model-based%20reinforcement%20learning%20for%20infinite-horizon%20approximate%20optimal%20tracking.%20in%20Decision%20and%20Control.%20%28IEEE%2C%202015%29%2C%20pp%205083%E2%80%935088)

36. C. Yu, M. Zhang, F. Ren et al., Emotional multiagent reinforcement learning in spatial social dilemmas. IEEE Trans. Neural Netw. Learn. Syst. **26**(12), 3083–3096 (2015)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3453259)
CrossRef (https://doi.org/10.1109/TNNLS.2015.2403394)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Emotional%20multiagent%20reinforcement%20learning%20in%20spatial%20social%20dilemmas&author=C.%20Yu&author=M.%20Zhang&author=F.%20Ren&journal=IEEE%20Trans.%20Neural%20Netw.%20Learn.%20Syst.&volume=26&issue=12&pages=3083-3096&publication_year=2015)

37. P. Plamondon, B. Chaib-Draa, A.R. Benaskeur, A Q-decomposition and bounded RTDP approach to resource allocation. in *Autonomous Agents & Multiagent Systems/Agent Theories, Architectures, and Languages* (2007), pp 1–8
Google Scholar (https://scholar.google.com/scholar?q=P.%20Plamondon%2C%20B.%20Chaib-Draa%2C%20A.R.%20Benaskeur%2C%20A%20Q-decomposition%20and%20bounded%20RTDP%20approach%20to%20resource%20allocation.%20in%20Autonomous%20Agents%20%26%20Multiagent%20Systems%2FAgent%20Theories%2C%20Architectures%2C%20and%20Languages%20%282007%29%2C%20pp%201%E2%80%938)

38. L. Matignon, G.J. Laurent, N.L. Fort-Piat, Coordination of independent learners in cooperative Markov games. Piat (2009)
Google Scholar (https://scholar.google.com/scholar?q=L.%20Matignon%2C%20G.J.%20Laurent%2C%20N.L.%20Fort-Piat%2C%20Coordination%20of%20independent%20learners%20in%20cooperative%20Markov%20games.%20Piat%20%282009%29)

Copyright information

About this chapter

Cite this chapter as:

Sun Q. (2019) Model-Free Energy Optimization for Energy Internet. In: Energy Internet and We-Energy. Renewable Energy Sources & Energy Storage. Springer, Singapore

- First Online 13 July 2018
- DOI https://doi.org/10.1007/978-981-13-0523-8_10
- Publisher Name Springer, Singapore
- Print ISBN 978-981-13-0522-1
- Online ISBN 978-981-13-0523-8
- eBook Packages Energy

Personalised recommendations