Available online at www.sciencedirect.com

ScienceDirect

Energy Reports 8 (2022) 1122–1129

www.elsevier.com/locate/egyr

2021 The 2nd International Conference on Power Engineering (ICPE 2021), December 09–11, 2021, Guangxi, China

Dynamic economic dispatch of power system based on DDPG algorithm

Zhicheng Liu[a], Yipeng Liu[b], Hao Xu[a], Siyang Liao[b,*], Kefan Zhu[a], Xinxiong Jiang[b]

[a] *Central China Electric Power Dispatching and Control Sub-center of State Grid, Wuhan 430000, China*
[b] *School of Electrical Engineering and Automation, Wuhan University, Wuhan 430000, China*

Received 26 January 2022; accepted 20 February 2022
Available online 14 March 2022

Abstract

With the rapid developments of renewable energy sources, the uncertainty of the power supply will bring new challenges to scheduling problems, and conventional scheduling strategies may lose their effectiveness. The current general scheduling strategies need to model the uncertainty of the environment, while it is difficult to achieve a high degree of accuracy in the power system with high penetration rate of new energy, which will directly affect the scheduling result. In response to this problem, this paper studies the economic dispatch of power systems based on deep deterministic policy gradient (DDPG),which avoids the uncertainty modeling of the environment in principle. Combined with the basic economic dispatch model, this paper has defined a learning mode of the algorithm and built an algorithm framework of economic dispatch of power system based on DDPG. The results of the experiment show that proposed algorithm is highly adaptable to random fluctuation of renewable energy.

Peer-review under responsibility of the scientific committee of the 2021 The 2nd International Conference on Power Engineering, ICPE, 2021.

Keywords: DDPG; Renewable energy; Artificial intelligence; Power system dispatch

1. Introduction

China has put forward the dual carbon target of carbon neutrality and carbon peak. With the realization of the dual carbon target, there will be a high percentage of renewable energy and power electronics and a trend of large-scale interconnection of new energy sources such as photovoltaics and wind power [1]. Compared with the deterministic traditional thermal power generation which obeys the power generation strategy arrangement, photovoltaic, wind power and other renewable energy power sources exhibit randomness, volatility, and uncertainty. To deal with the uncertainty problem brought by the power supply side, the traditional deterministic scheduling strategy is viable with reserving enough conventional generator sets when new energy accounts for a small proportion. However,

* Corresponding author.
E-mail address: liaosiyang@whu.edu.cn (S. Liao).

https://doi.org/10.1016/j.egyr.2022.02.231

Peer-review under responsibility of the scientific committee of the 2021 The 2nd International Conference on Power Engineering, ICPE, 2021.

according to China's "3060" carbon peak target, the percentage of non-fossil energy consumption needs to exceed 30% by 2030, when the traditional deterministic scheduling strategy will lose its original effectiveness. At that time, economic cost defects and security loopholes of conventional strategy will be highlighted. Therefore, how to complete the economic dispatch of the power system will be an important issue in an environment with uncertainties.

The first problem that mainstream researches have encountered is modeling uncertainty on the power supply side. Based on the uncertainty theory, Ref. [2] uses interval analysis for wind power and photovoltaic forecasts, combined with normal distribution, and proposes a normal uncertainty distribution to describe the output forecast errors of wind power and photovoltaics, but the wind power error forecast does not obey the normal distribution [3]. Ref. [4] believes that the probabilistic characterization of wind turbine output originates from the wind speed that obeys the Weibull distribution, but the conversion process from the prediction error of wind speed to the wind turbine output prediction error will magnify the error. At present, there are still few related researches on the prediction error of photovoltaic power output. Ref. [5], based on Copula theory, estimates the conditional prediction error distribution of photovoltaic power generation output, but the accuracy is still limited by the model.

Considering the uncertainty of wind and solar output, to solve economic dispatch problems, the first idea is using a deterministic scheduling model. The deterministic scheduling model formulates a power generation plan based on the determined wind and solar output forecast curves and reserves enough reserve capacity to deal with the uncertainty. Reserve capacity can be provided by conventional generator sets, demand response [6], electric vehicles [7], and energy storage [8]. Another way of thinking is to use the uncertainty scheduling model, which considers the power characteristics of the intermittent power supply by modeling the probability of uncertainty. Ref. [9] and Ref. [10] apply scene methods to realize the environmental differentiation of power output, and complete the description of the random output of the power supply through the feature identification of the scenes. Ref. [11], which takes the rolling correction of wind power and photovoltaic forecast errors as the premise, restricts the reserve capacity of conventional units and evaluates its confidence level. The core point of using the uncertainty scheduling model is to model the uncertainty, while the modeling under the background of complex randomness is very difficult and a high accuracy is always hard to reach.

Different from the methods mentioned above, this paper adopts a deep reinforcement learning method to avoid modeling environmental uncertainty, which is an environment interactive learning method with a clear goal orientation. Based on the Q-learning algorithm, whose action space and state space are discrete, Ref. [12] and Ref. [13] respectively proposed a microgrid scheduling method that considers integrated energy and a method of using energy storage capacity to compensate for wind power forecast errors. Ref. [14] uses deep convolutional neural network (CNN) to approximate its state space based on a Deep Q network algorithm, but its actions are still all discrete. Compared with discrete action space, the continuity of action space of DDPG will achieve economic dispatch with better effects and higher accuracy, which is used in this paper. But the research of applying DDPG to economic dispatch of power system is rather little. Ref. [15] realizes optimal dispatch of integrated electricity-gas with soft actor-critic deep reinforcement learning, with the result being satisfactory.

Above all, this paper uses a DDPG algorithm, which is able to avoid modeling uncertainty of environment and achieve precise actions, to realize economic dispatching of power system with renewables.

2. Basic model of economic dispatch

2.1. Objective function

The economic dispatch of the power system should fully consider the time sequence correlation between the various dispatch periods. On the condition of ensuring safety and stability, this paper will adjust the output and reserve capacity of the generator set to reduce the phenomenon of wind abandonment, abandonment of light, and load shedding. The objective function should be set to minimize the total cost of scheduling:

$$\min E\{\sum_{t\in T} F_{G,t} + F_{R,t} + F_{S,t}\}, \tag{1}$$

where $F_{G,t}$ is the power generation cost of the conventional generator sets, $F_{R,t}$ is the reserve cost, and $F_{S,t}$ is the random cost.

(1) Power generation cost of conventional generator sets. The power generation of conventional generator sets is the sum of generating costs of each generator set during the dispatch period:

$$F_{G,t} = \sum_{i=1}^{N} a_i + b_i P_{G,i,t} + c_i {P_{G,i,t}}^2, \tag{2}$$

where a_i, b_i, c_i are the cost coefficients corresponding to the ith unit, and N conventional units are considered in the model.

(2) Reserve cost. The reserve cost is the cost incurred by the system's pre-spinning reserve:

$$F_{R,t} = \sum_{i=1}^{N} (k_U U_{i,t} + k_D D_{i,t}), \tag{3}$$

$U_{i,t}$ and $D_{i,t}$ are the upward reserve capacity and the downward reserve capacity of the ith generator, respectively. k_U and k_D are the corresponding cost coefficients.

(3) Random cost. The random cost is caused by the forecast error of photovoltaic and wind power output. The power difference caused by the forecast error can be made up by the spinning reserve, but when the reserve margin is exceeded, it will lead to load shedding or abandonment of wind and solar power. Then random cost can be expressed as:

$$F_{S,t} = (c_U P_{U,t} + c_D P_{D,t}) + \left(c_L P_{L,t} + c_A P_{A,t}\right), \tag{4}$$

where $P_{U,t}$, $P_{D,t}$, $P_{L,t}$, $P_{A,t}$ are respectively the upward reserve power, the downward reserve power, the total power of the removed load, and the total power of wind and light curtailment. c_U, c_D, c_L, c_A are the corresponding cost coefficients.

It is worth noting that in this paper, all renewable energy outputs and loads in the environment is equivalent to random variables $P^{\Gamma}_{PV,t}$, $P^{\Gamma}_{WT,t}$ and $P^{\Gamma}_{L,t}$, which correspond to photovoltaic, wind power, and load power, respectively. We do not care about the distribution of these random variables, but realize the agent's self-adaptation to the random variables by continuous iterative learning.

2.2. Restrictions

The constraints in this article are taken from the conventional economic dispatch model. In a scheduling period **t**, the following constraints below should be satisfied.

(1) Power constraint:

$$P_{PV,t} + P_{WT,t} + \sum_{i=1}^{N} P_{G,i,t} = P_{L,t}, \tag{5}$$

where $P_{PV,t}$, $P_{WT,t}$, $P_{G,i,t}$, $P_{L,t}$ are the outputs of photovoltaic, wind power, the ith conventional generator set and load in the model, respectively.

(2) Output constraints:

$$P_{G,i,\min} \le P_{G,i,t} \le P_{G,i,\max}, \tag{6}$$

where $P_{G,i,\min}$, $P_{G,i,\min}$, $P_{G,i,t}$ are the lower limit, the upper limit, and the output value of the generator unit, respectively.

(3) Climbing constraints:

$$\begin{cases} P_{G,i,t} - P_{G,i,t-1} \le s_{U,i} \\ P_{G,i,i-1} - P_{G,i,t-1} \le s_{D,i} \end{cases}, \tag{7}$$

where $s_{U,i}$ and $s_{D,i}$ respectively represent the upward and downward ramp rate limits of the corresponding unit.

(4) Reserve constraints:

$$\begin{cases} 0 \le U_{i,t} \le P_{G,i,\max} - P_{G,i,t} \\ 0 \le D_{i,t} \le P_{G,i,t} - P_{G,i,\min} \end{cases}, \tag{8}$$

which means the reserve capacity of a conventional generator set is constrained by its current output value and output limit.

3. Dynamic economic dispatch method based on DDPG

3.1. Initial definition of algorithm

Within the framework of deep reinforcement learning, the interaction between the agent and the environment can be described by Markov Decision Process (MDP). MDP contains 5 elements, namely state, action, reward, state transition probability and discount factor, forming a five-tuple (S, A, R, P, γ). In order to make the algorithm applicable in economic dispatch, it is necessary to define elements below with reference to the economic dispatch model.

(1) Define the state space

The state space will determine the environment perception content of the agent. For conventional generator sets, the status should include the output value, upward reserve capacity, and downward reserve capacity of all generator sets, which can be expressed as:

$$S_G = \left(P_{G,1}, P_{G,2} \cdots P_{G,N}, U_1, U_2 \cdots U_3, D_1, D_2 \cdots D_3\right). \tag{9}$$

For random photovoltaic output, wind power output, and load, not only the real-time value can be considered, but also the first derivative and second derivative can be introduced in consideration of the time series correlation:

$$S_M = \left(P_{PV}, P_{WT}, P_L, P_{PV,d1}, P_{WT,d1}, P_{L,d1}, P_{PV,d2}, P_{WT,d2}, P_{L,d2}\right). \tag{10}$$

Then this state space with ($3N$+9) dimensions can finally be expressed as:

$$S = \{S_G, S_M\}. \tag{11}$$

(2) Define the action space

The decision-making actions of the agent are the output of the algorithm which will change the state of environment with probability P. The controllable actions considered in this paper include output value, upward reserve, and downward reserve of all conventional generator sets, with a dimension of $3N$. The action space is represented as:

$$A = \left\{A_{P,1}, A_{P,2} \cdots, A_{P,N}, A_{U,1}, A_{U,2} \cdots, A_{U,3}, A_{D,1}, A_{D,2} \cdots, A_{D,3}\right\}, \tag{12}$$

where $A_{P,i}$, $A_{U,i}$, $A_{D,i}$ respectively represent the decision-making action values of the ith unit output, upward standby, and downward standby.

(3) Define the reward

According to the economic dispatch objective function, the agent will obtain rewards by changing the state of the environment during the interaction process with the environment. The instant reward is defined as:

$$R = -0.05(F_G + F_R + F_S). \tag{13}$$

(4) Set the discount factor γ

The larger the discount factor, the more sensitive the algorithm is to future decision-making effects. Setting a larger discount factor can enhance the agent's foreseeability in the scheduling process. This paper sets γ to 0.85.

3.2. Training method

The DDPG algorithm is in "Actor-Critic" mode. After sensing the state of the environment, the Actor will output actions according to the current policy. The Actor includes an online policy network and a target policy network. Critic, which includes an online Q network and a target Q network, will use value Q to evaluate the correctness of decision-making actions. DDPG completes training by updating the parameters of the neural networks, and the update methods of online networks and target networks are gradient updating strategy and soft updating strategy respectively.

(1) Update online Q network:

$$\theta^Q \leftarrow \theta^Q + \alpha_Q \delta \cdot \nabla_{\theta^Q} Q\left(s, a \left|\theta^Q\right.\right), \tag{14}$$

where θ^Q is the network parameter of the update online Q network, α_Q is corresponding learning rate, and δ is temporal difference error.

(2) Update online policy network:

$$\theta^{\mu} \leftarrow \theta^{\mu} + \alpha_{\mu} \cdot \nabla_{a} Q\left(s, a \left|\theta^{Q}\right.\right)\Big|_{a=\mu(s|\theta^{\mu})} \cdot \nabla_{\mu} Q\left(s \left|\theta^{\mu}\right.\right), \tag{15}$$

where θ^{μ} is the network parameter of the online policy network, and α_{μ} is learning rate.

(3) Update target network:

$$\begin{cases} \theta^{Q} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'} \\ \theta^{\mu} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'} \end{cases}, \tag{16}$$

where τ determines the network update speed in soft update mode, which is generally far less than 1. The soft update mode causes the target network to be updated very slowly, but it also makes learning more stable and easier to converge.

3.3. Model framework

Online policy network observes state j, and then takes actions according to the policy. After that, the environment will return to the next state $(j+1)$ and the rewards obtained from this action. The target policy network also selects the virtual optimal action according to the state $(j+1)$ and feeds it to the target Q network to calculate the target Q value. Enter the target Q value, which is y_j, then the online Q network will calculate the temporal difference error and complete the gradient update of its network parameter θ^{Q}. Under the supervision of the online Q network, the online policy network will also complete the update of the policy, that is, the gradient update of its parameter θ^{μ}. Finally, the soft update of the target Q network and target policy network is completed according to θ^{Q} and θ^{μ} respectively.

It is worth noting that the framework uses an empirical playback mechanism to break the temporal correlation of training samples (see Fig. 1).

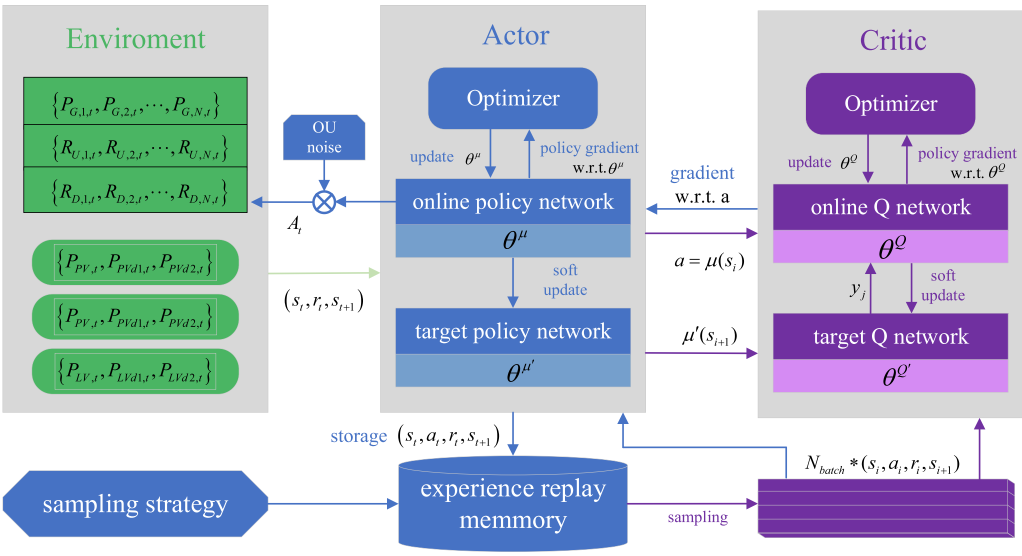

Fig. 1. Model framework.

4. Case studies

This paper is based on the IEEE 30-node system to verify the effectiveness and economy of the above method in dynamic economic dispatch. The new energy and load data refer to the data of North Hebei Power Grid, whose length is 357 days and interval is 15 min. As we take 100 MW as the power reference value, the total load in the system is 3.7 p.u., the rated output of photovoltaic and wind power is 0.418 p.u. and 0.521 p.u., the number of conventional generator sets is 6, and the relevant parameters of $G_1 \sim G_6$ [16] (see Table 1).

According to the above settings, the dimension of the state space of the DDPG algorithm of this example is 27, and the dimension of the action space is 18.

Table 1. Conventional generator parameters.

Unit	$P_{G,i,\min}$/p.u.	$P_{G,i,\max}$/p.u.	a_i/(\$/h)	b_i/(\$/(MW h))	c_i/(\$/((MW)^2 h))	D_i(p.u./h)	U_i(p.u./h)
G_1	0.45	2.00	786.80	38.5397	0.1524	0.5	0.5
G_2	0.30	1.00	945.70	46.2678	0.1058	0.3	0.3
G_3	0.15	0.60	1050.10	40.1591	0.0280	0.15	0.2
G_4	0.15	0.80	1244.00	38.3055	0.0354	0.2	0.15
G_5	0.20	0.40	1658.60	36.3278	0.0211	0.15	0.15
G_6	0.10	0.4	1356.70	38.2704	0.0179	0.15	0.15

4.1. Training process

In this case, we define the length of a dispatch period t is 24 h. When we are training, one day's data will be randomly selected as training data in each dispatch period, which means a training cycle will complete 96 iterations, corresponding to 96 sets of empirical data. And the experience data will be stored in the experience replay memory finally. Each group of experience includes the corresponding state, action, reward, and the next state after the action. Then the update mechanism will randomly extract the experience from it for the parameter update of the neural network. When the number of training scheduling periods reaches 2×10^6, the algorithm basically completes convergence, and the rewards obtained by the agent in each scheduling period are basically close to the optimal value.

4.2. Dynamic scheduling results

After the training, this case will randomly select one day's data in the non-training database as test data. Under the framework of the algorithm in this paper, the trained agent seems to complete the perception of the current environment through test data and make the most valuable actions within its cognitive range. One is to adjust the output of generators, and the other is to adjust the upward and downward reserve capacity to avoid severe penalties caused by abandoning wind, abandoning light, and load shedding.

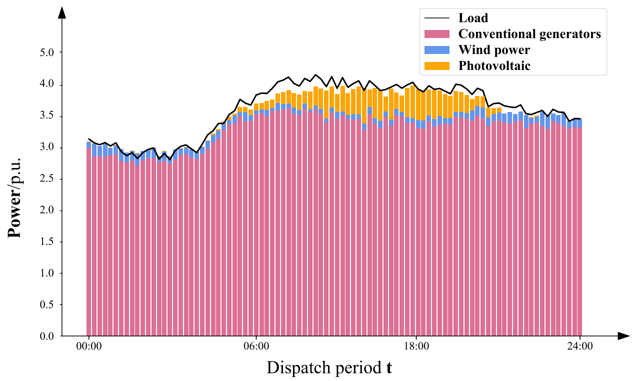

Fig. 2. Daily dispatch results.

Fig. 2 shows the results of dynamic economic dispatch on one day. It was concluded that the dispatch output of conventional generator sets can effectively adapt to the power fluctuations of wind power and photovoltaics and automatically track the trend of load changes.

Fig. 3 shows the adjustment of the reserve capacity by the agent. Only at four points A, B, C, D, the wind and solar abandonment phenomenon caused by the sharp decrease of the power shortage occurred.

4.3. Comparative analysis

Model predictive control (MPL) is commonly used optimization method in deterministic dispatching, which does not describe the uncertainty of environment. In this method, the current predictive value is used to roll out

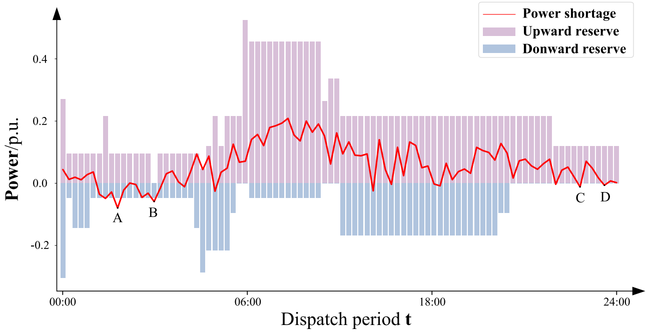

Fig. 3. Power balance result.

the optimal control strategy based on a predictive model. However, MPL relies so much on the predicted value that the unavoidable error of renewables forecast make it impossible to achieve the optimal effect at the source. Q-learning, the same as DDPG, can not only adapt to uncertainty of environment, but also avoid probabilistic modeling of environment. But its discretized action space makes it far inferior to DDPG in action selectivity. Below is a dispatching cost comparison of the three different methods (see Table 2).

Table 2. Dispatching cost comparison.

Method	Action cost/$	Uncertain cost/$	Total cost/$
DDPG	625213	31687	656900
Q-learning	656139	42042	698181
Model prediction	662872	43929	706801

As a new generation of artificial intelligence algorithm, DDPG has great advantages in the accuracy of actions and the adaptability of the environment, which is also reflected in economic analysis. Compared with Q-learning and the model prediction method mentioned above, its action costs and uncertain costs are less. Action cost refers to the cost incurred by the agent adjusting outputs and reserve capacities of generators during the dispatch process, which is determined by the role of the agent for DDPG. The uncertain cost is caused by uncertainty of power system, including the startup cost of up and down standby, and the cost of abandoning wind, abandoning light, and load shedding.

5. Concluding remarks

Aiming at the system with high penetration rate of renewable energy, this paper proposes an adaptive uncertain dynamic economic dispatch method based on DDPG. On the basis of the economic dispatch model, this paper has completed the initial definition of the DDPG algorithm, and built the overall framework of the algorithm, finally developed the agent's ability to complete the dynamic economic dispatch of power system.

The algorithm adapts to random fluctuations of photovoltaic, wind power, and load after training, and can accurately control generator output and reserve capacity. These experimental results show that DDPG has better forecasting and control capabilities, which can be directly verified by its lowest dispatch cost.

The DDPG algorithm used in this article, which avoids the probabilistic modeling of complex environments in principle, implements the efficient update of the strategy in the "Actor-Critic" mode and introduces neural networks to ensure accurate actions by avoiding the discretization of the state space and the action space, are also worth considering in other optimization problems of the power system.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was funded by the CENTRALCHINA BRANCH OF STATE GRID CORPERATION OF CHINA, grant number SGHZ0000DKJS2100256.

References

[1] Erdiwansyah, Mahidin, Husin H, et al. A critical review of the integration of renewable energy sources with various technologies. Protection and Control of Modern Power Systems 2021;6:37–54. http://dx.doi.org/10.1186/s41601-021-00181-3.

[2] Cao Dawei. Research on multi time scale energy optimal dispatch of microgrid considering uncertainty. 2018, p. 1–56, North China Electric Power University, 2018.

[3] Bouffard F, Galiana FD. Stochastic security for operations planning with significant wind power generation. IEEE Trans Power Syst 2008;23(2):301–6, 2008.

[4] Hetzer J, Yu DC, Bhattarai K. An economic dispatch model incorporating wind power. IEEE Trans Energy Convers 2008;23(2):603–11, 2008.

[5] Zhao Weijia, Zhang Ning, Kang Chongqin, et al. A method of probabilistic distribution estimation of conditional forecast error for photovoltaic power generation. Automat Electr Power Syst 2015;39(16):08–15, 2015.

[6] Zhao C, Zang J, Watson JP, et al. Multi-stage robust unit commitment considering wind and demand response uncertainties. IEEE Trans Power Syst 2013;28(3):2708–17, 2013.

[7] Zhao J, Wan C, Xu Z, et al. Spinning reserve requirement optimization considering integration of plug-in electric vehicles. IEEE Trans Smart Grid 2017;2017(8):2009–21, 2017.

[8] Kiran BDH, Kumari MS. Demand response and pumped hydro storage scheduling for balancing wind power uncertainties: A probabilistic unit commitment approach. Int J Electr Power Energy Syst 2016;2016(81):114–22, 2016.

[9] Gao Hongjun, Liu Junyong, Wei Zhenbo, et al. A security-constrainted dispatching model for wind generation units based on extreme scenario set optimization. Power Syst Technol 2013;37(06):1590–5, 2013.

[10] Fu Yimu, Liu Mingbo. Scenario decomposition method for multi-objective stochastic dynamic economical dispatch problem. Automat Electr Power Syst 2014;38(9):34–40, 2014.

[11] Li Zhiwei. Research on stochastic optimization scheduling of multi-energy power system with high proportion of wind power and photovoltaic power generation. 2019, p. 1–130, North China Electric Power University, 2019.

[12] Akin T, Borhan M, Gianfranco C, et al. A short-term spatio-temporal approach for photovoltaic power forecasting. In: Power systems computation conference. 2016, p. 1–07, 2016.

[13] Li Yanqing, Du Yingying. Short-term photovoltaic power forecasting based on double-dimensional sequential imputation framework and improved kohonen clustering. Electr Power Automat Equip 2019;39(1):60–5, 2019.

[14] Zhang Zidong, Qiu Caiming, Zhang Dongxia, et al. A coordinated control method for hybrid energy storage system in microgrid based on deep reinforcement learning. Power Syst Technol 2019;43(6):1914–21, 2019.

[15] Qao Ji, Wang Xinying, Zhang Qing, et al. Optimal dispatch of integrated electricity-gas system with soft actor-critic deep reinforcement learning. Proc CSEE 2021;41(3):819–32, 2021.

[16] Sun Yuanzhang, Wu Jun, Li Guojie, et al. Dynamic economic dispatch considering wind power penetration based on wind speed forecasting and stochastic programming. Proc CSEE 2009;29(4):41–7, 2009.