Many-Objective Distribution Network Reconfiguration Via Deep Reinforcement Learning Assisted Optimization Algorithm

Yuanzheng Li, *Member, IEEE*, Guokai Hao, Yun Liu, *Senior Member, IEEE*, Yaowen Yu, Zhixian Ni, *Graduate Student Member, IEEE*, and Yong Zhao

***Abstract*—With the increasing penetration of renewable energy (RE), the operations of distribution network are threatened and some issues may appear, i.e., large voltage deviation, deterioration of statistic voltage stability, high power loss, etc. In turn, RE accommodation would be significantly impacted. Therefore, we propose a many-objective distribution network reconfiguration (MDNR) model, with the consideration of RE curtailment, voltage deviation, power loss, statistic voltage stability, and generation cost. This aims to assess the trade-off among these objectives for better operations of distribution networks. As the proposed model is a non-convex, non-linear, many-objective optimization problem, it is difficult to be solved. We further propose a deep reinforcement learning (DRL) assisted multi-objective bacterial foraging optimization (DRL-MBFO) algorithm. This algorithm combines the advantages of DRL and MBFO, and is targeted to find the Pareto front of proposed MDNR model with better searching efficiency. Finally, we conduct case study on the modified IEEE 33-bus, 69-bus, and 118-bus power distribution systems, and results verify the effectiveness of the MDNR model and outperformance of the proposed DRL-MBFO.**

***Index Terms*—Distribution network reconfiguration, renewable energy, many-objective optimization, deep reinforcement learning.**

NOMENCLATURE

Abbreviations

RE	Renewable energy
DN	Distribution network
DNR	Distribution network reconfiguration
MDNR	Many-objective distribution network reconfiguration
RL	Reinforcement learning
DRL	Deep reinforcement learning
DRL-MBFO	Deep reinforcement learning assisted multi-objective bacterial foraging optimization algorithm
NSGA-II	Nondominated sorting genetic algorithm II
MOPSO	Multi-objective particle swarm optimization algorithm
MBFO	Multi-objective bacterial foraging optimization algorithm
DQN	Deep-Q-Net
PV	Photovoltaic
Cur	Curtailment of PV power
VD	Voltage deviation
PL	Power loss
L	L-index
GC	Generation cost
HV	Hypervolume
SP	Spacing index
MD	Mean distance
NPS	Number of Pareto solutions
SD	Standard deviation
PROMETHEE	Preference ranking organization method for enrichment of evaluations

Distribution Network Constants

N_B	Number of distribution network buses
N_S	Number of sampling scenarios of PV power
N_K	Number of branches
N_G	Number of generators
β	Integrated PV power
λ	The accommodation of PV power
R_k	Resistance of branch k
a_g, b_g, c_g	Cost coefficients of generator g
d_g, e_g	Sinusoidal term coefficients of generator g
$P_g^{\min}$	Minimum power output of generator g
G_{ij}, B_{ij}	Conductance and susceptance between buses i and j
C_g	The curtailed PV power in each step
V_{ref}	Reference value of voltage
σ	Standard deviation of PV power forecast error
K_c	Illumination when the PV conversion efficiency reaches the maximum value

Manuscript received February 8, 2021; revised May 26, 2021; accepted August 3, 2021. Date of publication August 27, 2021; date of current version May 24, 2022. This work was supported by the National Natural Science Foundation of China under Grant 62073148 and Tencent Rhinoceros Foundation of China under Grant CCF-Tencent RAGR20210102. Yuanzheng Liu's work was supported by the National Natural Science Foundation of China under Grant 51807120. Paper no. TPWRD-00235-2021. *(Corresponding author: Yun Liu.)*

Yuanzheng Li, Guokai Hao, Yaowen Yu, and Yong Zhao are with the School of Artificial Intelligence and Automation, and Ministry of Education Key Laboratory of Image Processing and Intelligence Control, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: yuanzheng_li@hust.edu.cn; h17803899119@163.com; yaowen_yu@hust.edu.cn; zhiwei98530@hust.edu.cn).

Yun Liu is with the School of Electric Power, South China University of Technology, Guangzhou 510640, China (e-mail: liuyun19881026@gmail.com).

Zhixian Ni is with the China-EU Institute for Clean and Renewable Energy, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: m201971297@hust.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRD.2021.3107534.

Digital Object Identifier 10.1109/TPWRD.2021.3107534

I_s	Stand illumination of PV
P_r	Rated active power of PV

Distribution Network Variables

x	The vector of DN state variables
u	The vector of DN decision variables
S	Set of sampling scenarios of PV power
P_i	Injected active power at bus i
Q_i	Injected reactive power at bus i
ξ_s	The s-th sample of PV power scenarios
$P_{\xi_s,i}$	Integrated active power at bus i with respect to the sample ξ_s
V_i, V_j	Voltages at buses i and j
$\theta_{i,j}$	The difference of voltage angle between buses i and j
$V_{i,min}, V_{i,max}$	The minimal and maximum voltage magnitudes at bus i
S_k	Transmission capacity of branch k
$S_{k,max}$	Maximum transmission capacity of branch k
P_g	Active power of generator g
Q_g	Reactive power of generator g
$P_{g,min}, Q_{g,min}$	Minimal active and reactive power of generator g
$P_{g,max}, Q_{g,max}$	Maximum active and reactive power of generator g
ζ	Topology of DN
Γ	Set of topologies that satisfy the radial structure
N_C	The total number of gradual curtailments
p_s	Probability occurrence for scenario s
C_s	PV power curtailment under scenario s
U_k	Bus voltage at the end of branch k
V_g	Voltage at generator g
Y_{LL}	Admittance matrix among load buses
Y_{LG}	Admittance matrix between the load bus and generator
I_t	Forecast illumination of PV
$P_{pre}(I_t)$	PV power output when the forecast illumination is I_t
P_{pre}	Forecast PV power
P_{err}	PV power forecast error
P_t	The actual PV power

DRL-MBFO Parameters

s, s'	Current state and the next state of agent
a, a'	Current action and the next action of agent
r	Reward
θ, θ^-	Parameters of Q_eval_net and Q_target_net
$Q(s, a, ; \theta)$	Q-value when state, action and parameters of networks are s, a, and θ
$L(\theta)$	Loss function
α	Learning rate
γ	Discount factor
$\mathbf{D}$	Replay memory
$\mathbf{N}$	The capacity of $\mathbf{D}$
$f_{t,i}$	The fitness of state $s'_{t,i}$
$\boldsymbol{\mathcal{A}}$	Non-dominated archive
$\mathcal{A}_j$	The j-th Pareto solution in $\boldsymbol{\mathcal{A}}$
X_{best}	Set of Pareto solutions
B	Solution set from $\mathcal{A}$, but currently dominated by X_{best}
N_D, N_{nonD}	The parts of dominated and non-dominated bacteria

I. Introduction

In RECENT years, energy and environmental crisis are major obstacles to the sustainable development of our society. To lessen these crisis, renewable energy (RE) has been widely utilized throughout the world [1]. For instance, the overall installed capacity of RE generation have reached 2537 GW by 2019 in our world, and the capacity of wind and solar power are as high as 623 GW and 586 GW, respectively [2]. Therefore, the power system with highly penetrated RE is gradually formed.

However, the high penetration of RE would lead to significant challenges to operations of power systems, such as the distribution network (DN) [3], [4]. That is, uncertain RE power output impacts the distribution and direction of DN power flow [5], which may cause the increase of power loss [6], voltage deviations [7], etc.

In order to deal with these challenges, one of the strategies is the curtailment of RE. For instance, [8] proposes a method of RE curtailment, in order to solve the problem of over-voltage prevention. In addition, [9] studies the influence of RE curtailments on the security and reliability of the power system, based on different scenarios of RE curtailment. Furthermore, [10] proposes an efficient distributed control algorithm, and adopts RE curtailment to maintain the bus voltage within an acceptable range, considering that the varying solar power would lead to voltage violations. However, it should be noted that the curtailment severely restricts RE accommodation, which is discouraged as more RE is expected to be utilized.

As one of the most crucial tools, the distribution network reconfiguration (DNR) has been widely used to improve DN operations, such as minimizing the power loss [11], the voltage deviation [12], etc. For instance, [11] proposes a DNR model for annual configuration scheduling, which determines the reconfiguration period to achieve the minimum power loss. Furthermore, an adaptive fuzzy-based parallel genetic algorithm is used in [12] to solve the DNR problem. It aims to minimize the power loss, the voltage deviation and the number of switches. Besides, [13] aims to optimize the power loss and load balancing, simultaneously, and the optimal configurations on different test networks are obtained. With the integration of RE, some scholars also study the DNR while considering uncertainties of RE power output. For example, [14] proposes a DNR framework to minimize the active power loss, and RE uncertainties are taken into account. Similarly, [15] adopts the DNR to reduce the operation cost while satisfying the acceptable level of risk brought by the highly penetrated RE. In addition, [16] proposes a stochastic DNR to handle the RE uncertainties and minimizes the power loss of the DN.

The above study verifies the good performance of DNR in improving DN operations. However, with the increasing integration of RE, how to well accommodate more of them and supporting the operational performance of DN is worth investigating. The main reason lies in that RE is expected to be utilized as much

as possible according to the requirements of various countries (some even decreed laws). However, existing approaches have not investigated this issue, to the best of authors' knowledge. Thus, an issue is naturally inspired, is it possible to enhance the accommodation of RE via DNR while supporting the operation of the DN? In other words, the voltage deviation, network power loss, statistic voltage stability, generation cost, etc., should not be significantly affected by the RE accommodation. It means the above indexes that manifest operations of DN, together with the amount of RE accommodation should be simultaneously optimized to achieve a balance. The issue refers to the many-objective optimization problem. Therefore, different from existing approaches, we propose a many-objective distribution network reconfiguration (MDNR) model to investigate the issue. It aims to obtain the trade-off relationship among the accommodation and the performance of DN operations. This would help decision making to promote RE accommodation while supporting DN operations.

It is worth noting that, unlike traditional multiple-objective optimization problems, the many-objective one contains no less than 4 objectives. Accordingly, as the number of objectives increases, the number of Pareto solutions[1] also increases substantially, which makes traditional multi-objective optimization algorithms lose the selection driver [17]. Specifically, it is known that the population of traditional algorithms would evolve to better searching directions, and update themselves for obtaining better objective values. However, the number of Pareto/non-dominated solutions may rise significantly when the objectives increase. It may lead to the situation that there exist no better searching directions for these non-dominated solutions to move. In addition, a large number of non-dominated solutions also make dominated ones not easy to update themselves with clear directions. The above issues are referred to the loss of the selection driver, and would weaken the performance of optimization algorithms. In other words, when confronted with many objectives, the searching efficiency of conventional algorithms, such as non-dominated sorting genetic algorithm II (NSGA-II) [18], multi-objective particle swarm optimization (MOPSO) [19], and the multi-objective bacterial foraging optimization algorithm (MBFO) [20], would be reduced. Note that MBFO is proposed in recent years, and it can obtain better non-dominating solutions with faster convergence speed, compared with traditional algorithms, such as NSGA-II, etc. [20].

Therefore, to deal with the problem of losing selection driver, we propose a deep reinforcement learning (DRL) assisted optimization algorithm. First, on the basis of the MBFO algorithm, we adopt the Deep-Q-Net (DQN) based DRL to help their updating. Note that DQN is a popular form of DRL. The reason why we use DQN is that it can output specific actions to help MBFO members move toward better searching directions, according to their states. That is, it takes advantage of reinforcement learning, and could help lessen the loss of the selection driver with the guide of better directions.

However, it should be mentioned that the traditional DQN outputs all the Q-values once, which makes the action space large. It would cause the inefficiency of DQN. To overcome this shortcoming, we further propose an improved DQN scheme. In detail, we set a number of DQNs to cooperatively deal with the action space, and the number is dependent on quantities of decision variables in our proposed MDNR model. Then, the action space is decomposed into several subspaces, and each subspace is processed by a DQN, respectively. Consequently, in this paper, we propose a deep reinforcement learning assisted optimization algorithm based on the improved DQN scheme, in order to well solve our proposed MDNR model.

To this end, at first, this paper aims to establish a MDNR model, in which five objectives are investigated: 1) RE curtailments, 2) voltage deviation, 3) power loss, 4) statistic voltage stability, and 5) generation cost. Then, we further propose a deep reinforcement learning (DRL) assisted MBFO (DRL-MBFO) algorithm, in order to efficiently address the MDNR model. Main contributions of this paper are shown as follows.

1) Unlike existing DNR approaches that have mainly been used to improve DN operational performance, such as reducing the power loss, the voltage deviation, etc., our work aims to investigate the trade-off relationship among the RE accommodation and DN operations. This would help operators to make better decision for enhancing RE accommodation while supporting the DN secure and economic operations, under the background that more RE is integrated into power systems and encouraged to be utilized significantly. Therefore, we propose a MDNR model to solve this issue, and many objectives are simultaneously considered for obtaining the Pareto solutions. In this way, the relationship among RE curtailments, voltage deviation, power loss, statistic voltage stability, and generation cost can be investigated for an attempt to balance the RE accommodation and DN operations.

2) Compared with the existing approaches, another contribution of our work is proposing a DRL assisted optimization algorithm. It is targeted to solve our proposed MDNR model, as traditional optimization algorithms would lose selection driver and weaken the searching performance when the number of optimized objectives increases. Therefore, we propose a DRL-MBFO algorithm based on an improved DQN scheme. Specifically, we propose this improved scheme to overcome the problem that traditional DQN makes the action space large. Then, we use the improved DQN to help MBFO algorithm to evolve with better directions, and enhance its selection driver.

Finally, three systems including the IEEE 33-bus, IEEE 69-bus and IEEE 118-bus DNs are tested to verify the effectiveness of our proposed MDNR model, as well as the efficiency of the DRL-MBFO algorithm based on the improved DQN scheme. The reminder of this paper is organized as follows. Section II presents the MDNR model while considering uncertain RE integrated. Section III shows details of the DRL-MBFO. Then, numerical simulations are conducted in Section IV. In the end, conclusions are drawn in Section V.

II. Many-Objective Distribution Network Reconfiguration Model

A. Problem Formulations

For the MDNR model, objective functions are minimized while satisfying equality and inequality constraints. Without loss

[1]Pareto solutions are obtained to reflect the trade-off among multiple objectives. None of them can bring the optimal solution for these objectives but show a trade-off. Pareto solutions are also called as non-dominated solutions.

of generality, we take photovoltaic (PV) power as an example to study RE curtailments. In this paper, objective functions are shown as follows: 1) PV power curtailment, 2) voltage deviation, 3) power loss, 4) statistic voltage stability, and 5) generation cost. The optimization of such objectives aims to enhance the PV accommodation and better support the DN operations. On the other hand, corresponding constraints include power flow equations, power outputs of generators, branch capacities, node voltages, and the network topology. The decision variables are states of tie switches. Therefore, the MDNR model is formulated as follows.

$$\min_{u} \ F(x, u, S) \tag{1}$$

$$P_i + P_{\xi_s,i} = V_i \sum\nolimits_{j=1}^{N_B} V_j (G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij}) \tag{2}$$

$$Q_i = V_i \sum\nolimits_{j=1}^{N_B} V_j (G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij}) \tag{3}$$

$$V_{i,min} \le V_i \le V_{i,max} \tag{4}$$

$$S_k \le S_{k,max} \tag{5}$$

$$P_{g,min} \le P_g \le P_{g,max} \tag{6}$$

$$Q_{g,min} \le Q_g \le Q_{g,max} \tag{7}$$

$$\zeta \in \Gamma \tag{8}$$

where (1) is the vector of objective functions, x represents the vector of state variables in the DN, including P_i, Q_i, V_i, $\theta_{i,j}$, P_g and Q_g, etc. u stands for the vector of decision variables. Each dimension of u indicates the serial number of the open switch regarding the corresponding loop in the DN, which is determined via our used loop encoding method. Details of this method can be referred to the Appendix A of this paper. S stands for the set of sampling scenarios of PV power, and it can be obtained by the two-point estimation method [21]. This method is widely used and very efficient [21]–[24]. Therefore, we adopt it in our work, and obtain the PV power scenarios and their occurrence probabilities. Then, the obtained scenarios represent PV power uncertainties, which could be described by the scenarios ξ_s and the corresponding probability p_s. $s = 1, 2, \ldots, N_S$, and N_S is the number of sampling scenarios.

In addition, (2)–(3) stand for power flow equations, which are non-convex. P_i and Q_i are denoted as the injected active and reactive power at the i-th bus ($i = 1, 2, \ldots N_B$), and N_B is the number of DN buses. $P_{\xi_s,i}$ represents the integrated active PV power at bus i with respect to sample ξ_s. Note that the PV system is usually designed to operate at a unity power factor, thus it provides only active power [25]–[28], thus the reactive power is not considered in this paper. V_i and V_j are voltages at the i-th and the j-th bus. G_{ij} and B_{ij} stand for the conductance and susceptance, and θ_{ij} denotes the voltage angle difference between buses i and j. Inequality constraints (4)–(7) stand for the operational limits of the DN. $V_{i,min}$ and $V_{i,max}$ are the minimal and maximum voltage magnitudes at bus i. S_k denotes the transmission capacity of branch k ($k = 1, 2, \ldots, N_K$), N_K is the number of the branches. $S_{k,max}$ is the corresponding rated value. P_g and Q_g stand for active and reactive power outputs of the generator g ($g = 1, 2, \ldots, N_G$), N_G denotes the total number of generators. Besides, $P_{g,min}$, $P_{g,max}$, $Q_{g,min}$ and $Q_{g,max}$ are their corresponding lower and upper bounds, respectively. Finally, (8) is the topology constraint, i.e., there is only one possible path between each bus and the generator [29]. ζ represents the topology of DN after reconfiguration, and Γ is the set of topologies that satisfy the radial structures.

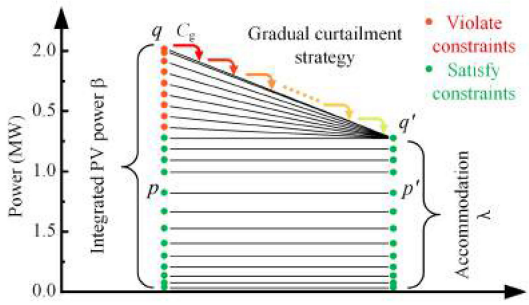


Fig. 1. Gradual curtailment strategy.

B. Objectives

1) PV Power Curtailment: As mentioned earlier, PV power is expected to be accommodated as much as possible. However, high PV power penetration would bring a great threat to the DN operations, which may directly cause some constraints (2)–(7) violated. In order to assure the PV accommodation while satisfying the constraints, this paper adopts a gradual curtailment strategy, and the detailed process is shown in Fig. 1.

In this figure, the points in the left vertical line represent the integrated PV power β, and the ones in the right vertical line stand for its accommodation λ. When the integrated PV power β would not violate operational constraints (2)–(7), the corresponding accommodation is equal to the integrated power, for instance, regarding the points of p and p'. However, once the integrated PV power cause some constraints violated, the curtailment strategy has to be activated for ensuring the normal operation of DN. For example, in terms of the point q, the β will be gradually decreased with several curtailments until the constraints (2)–(7) are satisfied. In this way, the integrated PV power q would not equal to power accommodation q', and the difference between q and q' is the PV power curtailment. The detailed steps of gradual curtailment strategy is shown in following procedures.

Step 1: Initialize the counter of several curtailments $N_C = 0$, and set the curtail step size of PV power C_g.

Step 2: Calculate power flow equations (2)-(3), if the power flow is convergent, we can obtain the power system state variable x, then check constraints (4)–(7). If these constraints are simultaneously satisfied, go to ***Step 4***.

Step 3: If the power flow is not convergent or (4)–(7) is violated, then the C_g power will be curtailed. Meanwhile, update $N_C = N_C + 1$ and calculate $P'_{\xi_s,i} = P_{\xi_s,i} - C_g$, then replace $P_{\xi_s,i}$ with $P'_{\xi_s,i}$ in Eq. (2). Afterwards, return to ***Step 2***.

Step 4: Calculate the total curtailments regarding scenario based on the counter N_c, i.e., $C_s = N_C \cdot C_g$.

Finally, taking the stochastic PV power into account, the expected curtailment can be calculated as follows.

$$\min_{u} \mathrm{Cur}(x, u, S) = \sum_{s=1}^{N_S} p_s \cdot C_s \tag{9}$$

where p_s represents the probability occurrence for scenario ξ_s, and C_s is the PV power curtailment under this scenario.

2) Voltage Deviation: Voltage deviation (VD) is an important index in the power quality. However, the highly integrated RE may worsen the voltage deviation in the DN [30]. It would will not only threaten secure operations, but also directly affects the service life of electric power equipment. Therefore, we minimize the VD as follows.

$$\min_{u} \ \mathrm{VD}(x,u,S) = \sum_{s=1}^{N_S} p_s \cdot \frac{1}{N_B} \sum_{i=1}^{N_B} |V_i - V_{ref}| \tag{10}$$

where V_{ref} is the reference value (usually set as 1.0). In (10), we first calculate the averaged VD for all N_B buses in the DN, and obtain the expected VD under the PV scenarios using their probability p_s.

3) Power Loss: Power loss (PL) is one of the important indexes to measure the operational economics of DNs. The RE with high penetration would affect the distribution and direction of power flow, which may cause PL increase significantly. It is calculated as follows.

$$\min_{u} \ \mathrm{PL}(x,u,S) = \sum_{s=1}^{N_S} p_s \cdot \sum_{k=1}^{N_K} R_k \frac{P_k^2 + Q_k^2}{U_k^2} \tag{11}$$

where R_k is the resistance of branch k, and U_k stands for the bus voltage at the end of this branch. In (11), we first compute the sum of PL for each branch, and obtain its expectation on the basis of PV power scenarios.

4) Statistic Voltage Stability: The statistic voltage stability can be manifested by L-index [31]. The larger L-index, the worse statistic voltage stability. It can reflect the distance between the load bus and the voltage collapse point, and each load bus is with respect to a L-index. Therefore, we aim to minimize the maximum L-index to support the statistic voltage stability of DN.

$$\min_{u} \ \mathrm{L}(x,u,S) = \max_{i \in N_B} \sum_{s=1}^{N_S} p_s \cdot \left| 1 - \frac{\sum_{g=1}^{N_G} F_{i,g} V_g}{V_i} \right| \tag{12}$$

where V_g and $F_{i,g}$ are the voltage at generator g and the element of matrix F, respectively. F is presented as follows.

$$F = -Y_{LL}^{-1} Y_{LG} \tag{13}$$

where Y_{LL} represents the admittance matrix among load buses, while Y_{LG} stands for the admittance matrix between the load bus and generator.

5) Generation Cost: In addition, reducing generation cost is also significant for the economic benefits in the DN. The generation cost (GC) is calculated as follows.

$$\min_{u} \ \mathrm{GC}(x,u,S) = \sum_{s=1}^{N_S} p_s \cdot \sum_{g=1}^{N_G} \left\{ c_g P_g^2 + b_g P_g + a_g + \left| d_g \sin(e_g(P_g^{\min} - P_g)) \right| \right\} \tag{14}$$

where a_g, b_g, and c_g are cost coefficients, d_g and e_g are the sinusoidal term coefficients that manifest the valve point effect [32]. In addition, $P_{G_g}^{\min}$ represents the minimum power output of generator g.

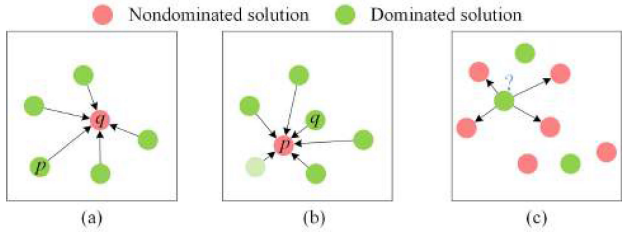


Fig. 2. The loss of selection driver.

III. Deep Reinforcement Learning Assisted Multi-Objective Bacterial Foraging Optimization Algorithm

As discussed in the previous section, the proposed model is a non-convex many-objective optimization problem. It can be converted to a single-objective optimization problem via the weighted sum method [33]. However, it is unsuitable for solving the proposed MDNR model in this paper, as this method could not well solve the non-convex optimization problem [33].

Therefore, multi-objective evolutionary algorithms provide an alternative approach to deal with this problem, such as MOPSO, NSGA-II, MBFO, etc. However, these algorithms have encountered great difficulties in solving many-objective optimization problems, although they have shown excellent performance in the problems with two or three objectives. The primary reason is that some candidate solutions become non-dominated with the increasing number of objectives. It makes traditional algorithms lose the selection driver, which could be shown in Fig. 2. In this figure, we take the solutions p and q as examples. Initially, p is dominated by q, due to the selection driver, p would move toward q, as shown in Fig. 2(a). After some iterations in the optimization process, p might become a better nondominated solution compared with q. Then, it makes q and other dominated solutions move towards p, which is presented in Fig. 2(b). Via above operations, the population would evolve to a better searching direction. However, as the number of objectives increases, the number of nondominated solutions also rises substantially, as shown in Fig. 2(c). Since there may exist no better directions for nondominated solutions to move, which would greatly reduce the selection driver of the population. In addition, a large number of non-dominated solutions also make the dominated solutions not easy to update themselves with clear directions. This is referred to the loss of the selection driver, and would weaken the performance of optimization algorithms [17].

In order to solve this problem, we introduce the DRL technique to assist optimization algorithms find suitable searching directions. Specifically, DRL could conduct the off-line learning on the basis of training data. It would help provide references for actions of candidate solutions, in order to obtain better optimization performance. In this paper, we select the DRL to assist MBFO, as MBFO is a recently proposed optimization algorithm and performs better than NSGA-II and MOPSO. It is briefly presented in the following subsection.

A. The Algorithm of MBFO

MBFO is an emerging optimization algorithm, which simulates the foraging behavior of bacteria. This algorithm solves

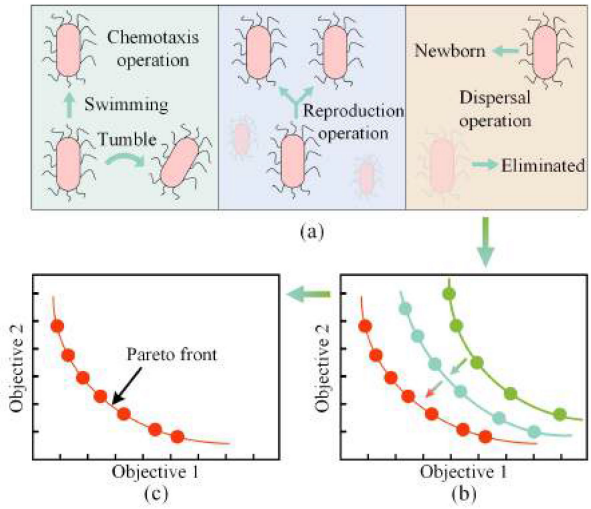


Fig. 3. Operations of MBFO.

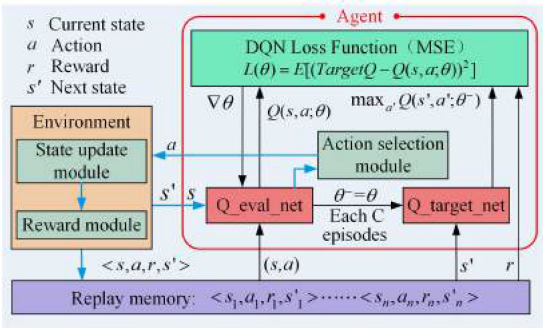


Fig. 4. The procedure of DQN.

optimization problems via the following operations, i.e., chemotaxis, reproduction and elimination, as shown in Fig. 3(a). Therein, the chemotaxis operation has two basic movements, i.e., the swimming and the tumbling. Bacteria tumble randomly to change the direction of their movement, while moving forward via swimming. The reproduction operation is to simulate the evolutionary process of bacteria, and maintains the good individuals. Afterwards, some bacteria may mutate and cause the elimination, and new individuals will be randomly generated in the feasible region. Via above operations, the bacteria will evolve with iterations as shown in Fig. 3(b), and the Pareto front can be obtained after several iterations via the nondominated sorting, as presented in Fig. 3(c).

B. Deep Reinforcement Learning

DRL is an artificial intelligence algorithm that combines the deep learning and reinforcement learning (RL). DQN [34] is one of the pioneering work of DRL, on the basis of deep neural networks and RL. Therein, RL is used to achieve the technique of autonomous learning stimulated by the goal of agents, and the deep neural network is adopted to address the problem of agent perception. The procedure of DQN is shown in Fig. 4.

There are three primary elements in this figure, i.e., the agent, environment and replay memory. The agent first obtains the state s from the environment. Then, it processes the state via the neural network Q_eval_net and the action selection module, and outputs action a to the environment. After that, the environment updates the state of agent to the next state s' through the state update module, and calculates the reward r by the reward module. Finally, we store $<s, a, r, s'>$ in the replay memory, and output the next state s' to the agent again.

For details of the agent, there exist two neural networks, i.e., Q_eval_net and Q_target_net. They are with the same structure but different parameters, in which θ and θ^- represent the weight parameters, respectively. Note that these two networks are adopted to predict Q-values of all actions with the states s and s', respectively. The agent will randomly extract a fixed number of samples from the replay memory to train Q_eval_net. When the training sample is $<s, a, r, s'>$, we first input s to Q_eval_net to obtain the Q-value of all actions, and a is used to determine the Q-value $Q(s, a; \theta)$. Specifically, it represents the Q-value output by Q_eval_net, when the state and action are s and a, respectively. Meanwhile, we also input s' to Q_target_net to obtain the Q-value of all actions when the state is s'. Then, the maximum Q-value $\max_{a'} Q(s,' a'; \theta^-)$ will combine with r to form $TargetQ$, as shown in (15). Therein, $\max_{a'} Q(s,' a'; \theta^-)$ stands for the maximum of all Q-values which are output by Q_target_net, when the state is s'. a' represents the action corresponding to $\max_{a'} Q(s,' a'; \theta^-)$. After that, we could calculate the loss function according to (16). On this basis, the weight parameter θ is updated depending on the loss function $L(\theta)$ at each iteration, which is shown in (17).

$$TargetQ = r + \gamma \max_{a'} Q(s,' a'; \theta^-) \tag{15}$$

$$L(\theta) = E[(TargetQ - Q(s, a; \theta))^2] \tag{16}$$

$$\theta_{i+1} = \theta_i + \alpha \cdot \nabla_\theta L(\theta_i) \tag{17}$$

where γ and α stand for the discount factor and learning rate, which would have a great influence on the training process. In addition, i is the current number of iterations.

Note that the learning rate α controls the learning speed. In other words, it determines the changing extents of weights of the neural network, concerning the loss function. When the learning rate is small, the weights of the neural network change slowly, which may result in a slow training process. On the contrary, a large learning rate value may lead to the weights diverging away from the optimal values [35]. In this paper, the DQN starts training from a relatively high learning rate, and we gradually decrease it during the training, which could make the training process faster. In addition, the discount factor γ is used to measure the present value of the future reward. The purpose of DQN is to determine the optimal policy, therefore, it is necessary to consider the current and future rewards. Besides, the value of current reward is usually more important than the future one. Therefore, we need to multiply the future reward by the discount factor [36], [37].

In the training process, after a constant iterations C, θ is assigned to θ^- for the update of Q_target_net. Repeat the above iterations until the termination condition is satisfied, the trained DQN can output the Q-value of all actions according to the current state s. Also, an attractive reward which is more likely to be obtained when the Q-value of action is large.

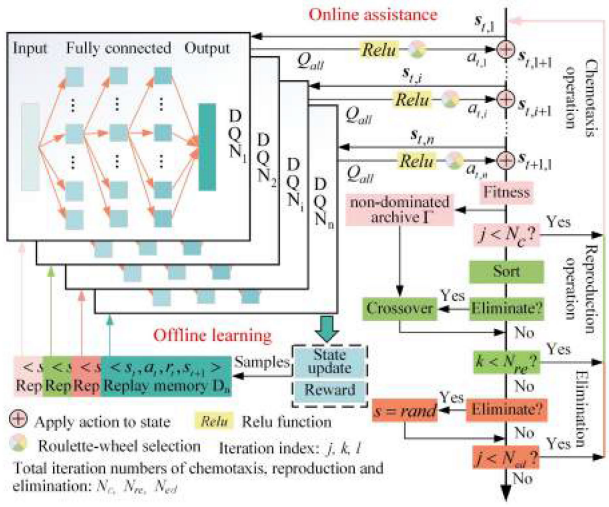


Fig. 5. Multi-objective bacterial foraging optimization algorithm based on DQN.

C. Deep Reinforcement Learning Assisted MBFO Algorithm

MBFO can effectively solve various multi-objective optimization problems. However, it is not effective for dealing with many objectives as it would lose the selection driver. In this paper, we propose a novel optimization algorithm, i.e., DRL-MBFO, in which the DRL and MBFO are well combined to enhance the selection ability and find suitable optimization directions. Specifically, the candidates of traditional MBFO are not easy to update themselves with clear directions when confronted with many objectives. Therefore, we adopt the DQN based DRL to help these candidates, as the DQN can output specific actions for MBFO. However, the traditional DQN outputs all the Q-values once, which makes the action space large. To solve this problem, we further propose an improved deep-Q-learning algorithm. In detail, we set a number of DQNs to cooperatively deal with the action space, and the number is dependent on quantities of decision variables in our proposed MDNR model. Then, the action space is decomposed into several subspaces, and each subspace is processed by a DQN, respectively.

The details are presented in Fig. 5, which comprises two main parts, i.e., the offline learning and online assistance. First, offline leaning is used for training DQNs via the improved deep-Q-learning algorithm.

Specifically, we input the randomly generated state (decision variable) to the 1-st DQN (DQN_1), and the DQN_1 would output the Q-values for all actions. Then, an action is selected to update the state and calculate rewards. In this way, the 1-st sample is generated including the previous state, updated state, reward and action, which is saved into the 1-st replay memory. Afterwards, the updated state is input into the 2-nd DQN (DQN_2), and we could obtain the 2-nd sample. Afterwards, the n-th sample is generated and the n-th replay memory is used to store it. Note that the state in the n-th sample is then input to the DQN_1 again, and we repeat the procedure until all DQNs are well trained.

Second, the trained DQNs are used to assist update candidate solutions of MBFO, this process is referred to online assistance. In this process, we first input states to DQN_1, and DQN_1 would output the Q-values of all actions. After that, an action is selected to update the first dimension of state. Then, we input the state into the next DQN, i.e., DQN_2. Repeat above process until all dimension of states are updated, and the chemotaxis operation is completed. This process is related to the online assistance, which helps the chemotaxis operation of MBFO. In this way, the chemotaxis operation with the online assistance is combined with the reproduction and elimination operations, in order to enhance the performance of MBFO. The details of offline learning and online assistance are presented as follows.

1) Offline Learning: In this part, we present the improved deep-Q-learning algorithm, which is adopted for training DQNs in the offline leaning. The procedure of the algorithm is shown in **Algorithm 1**. First, we initialize $\mathbf{D}, \mathbf{N}, \boldsymbol{\theta}, \boldsymbol{\theta}^{-}$, where $\mathbf{D} = [\mathrm{D}_1, \mathrm{D}_2, \ldots, \mathrm{D}_n]^{\mathrm{T}}$ denotes the replay memory, $\mathbf{N}=[\mathrm{N}_1, \mathrm{N}_2, \cdots, \mathrm{N}_n]^{\mathrm{T}}$ represents the capacity of $\mathbf{D}$. $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]^{\mathrm{T}}$ and $\boldsymbol{\theta^{-}} = [\theta_1^{-}, \theta_2^{-}, \ldots, \theta_n^{-}]^{\mathrm{T}}$ are the weight parameters of Q_eval_net and Q_target_net, respectively. n is the number of dimensions of $\boldsymbol{u}$. Accordingly, the number of DQN in Q_eval_net is also n, i.e., Q_eval_net=$[\mathrm{DQN}_1, \mathrm{DQN}_2, \cdots, \mathrm{DQN}_n]^{\mathrm{T}}$, and Q_target_net=$[\mathrm{DQN}_1^{-}, \mathrm{DQN}_2^{-}, \cdots, \mathrm{DQN}_n^{-}]^{\mathrm{T}}$.

In **Algorithm 1**, there exist three loops. The outermost loop is the first one, and its iteration starts from 1 to M. The second loop stems from 1 to T, and the innermost one begins from 1 to n, which is the third loop. In addition, m, t and i are the current number of iterations regarding these three loops, respectively. When the first loop starts, the state s_0 is randomly initialized, it should be noted that the state also stands for the vector of decision variables u. Therefore, each dimension of the states indicates the serial number of the open switch in the corresponding loop. The details could be referred to the Appendix A of this paper. When the third loop starts, we input $s_{t,i} = [s_{t,i}^1, \ldots, s_{t,i}^i, \ldots, s_{t,i}^n]$ to Q_eval_net, and it then outputs Q-values of all actions regarding state $s_{t,i}$. Afterwards, the ϵ-greedy strategy chooses an action $a_{t,i}$, and calculate $s'_{t,i}$ according to following formula.

$$\mathbf{s}'_{t,i} = [s_{t,i}^1, \ldots, s_{t,i}^i + a_{t,i}, \ldots, s_{t,i}^n] \tag{18}$$

where $s'_{t,i}$ stands for the next state of agent when the iterations of the second and the third loop are t and i. $a_{t,i}$ represents the corresponding action, which also indicates that the change in the i-th dimension variable $s_{t,i}^i$ of $s_{t,i}$. For example, when $s_{2,3} = [s_{2,3}^1, s_{2,3}^2, s_{2,3}^3, s_{2,3}^4, s_{2,3}^5] = [1, 2, 3, 4, 5]$ and $a_{2,3} = 1$, according to (18), the next state $s'_{2,3} = [1, 2, 3+1, 4, 5] = [1, 2, 4, 4, 5]$.

Next, we estimate the fitness (objective value) $f_{t,i}$ of $s'_{t,i}$, and calculate the reward $r_{t,i}$ as follows.

$$r_{t,i} = \begin{cases} -1, & \text{if } \exists\ j, \mathcal{A}_j \prec f_{t,i} \\ 1, & \text{otherwise} \end{cases} \tag{19}$$

where $\mathcal{A}_j$ denotes a Pareto solution in $\mathcal{A}$, and $\mathcal{A}$ represents the non-dominated archive. $\mathcal{A}_j \prec f_{t,i}$ means that $f_{t,i}$ is dominated by $\mathcal{A}_j$.

If $r_{t,i} = 1$, we store $f_{t,i}$ in $\boldsymbol{\mathcal{A}}$. Afterwards, the tuple $<\mathbf{s}_{t,i}, a_{t,i}, r_{t,i}, \mathbf{s}'_{t,i}>$ is stored in D_i, and $\mathbf{s}_{t,i+1} = \mathbf{s}'_{t,i}$ is set. Afterwards, the next step is to train the Q_eval_net. In the training, we first randomly extract a fixed number of samples from D_i to train DQN_1, and update θ_i by the gradient descent. After each C

Algorithm 1: Improved Deep-Q-leaning Algorithm.

Initialize replay memory $\mathbf{D}$ with capacity $\mathbf{N}$
Initialize Q_eval_net with random weights $\boldsymbol{\theta}$
Initialize Q_target_net with random weights $\boldsymbol{\theta}^{-}$
for m=1 to M **do**
  Random initialize state s_0
  for t=1 to T **do**
    for i=1 to n **do**
      Input $s_{t,i}$ to DQN_i to obtain the Q-values of all actions
      Choose action $a_{t,i}$ according to the ε-greedy strategy
      Update the state to $\mathbf{s}'_{t,i}$ via $\mathbf{s}_{t,i}$ and $a_{t,i}$
      Calculate the fitness value $f_{t,i}$ of $\mathbf{s}'_{t,i}$
      Set $r_{t,i} \leftarrow \begin{cases} -1, & \text{if } \exists\ j, \mathcal{A}_j \prec f_{t,i} \\ 1, & \text{otherwise} \end{cases}$
      if $r_{t,i} = 1$ **then**
        Store $f_{t,i}$ in $\boldsymbol{\mathcal{A}}$
      end if
      Store transition $<\mathbf{s}_{t,i}, a_{t,i}, r_{t,i}, \mathbf{s}'_{t,i}>$ in D_i
      Set $\mathbf{s}_{t,i+1} \leftarrow \mathbf{s}'_{t,i}$
      Sample random minibatch of transitions $<\mathbf{s}_{k,i}, a_{k,i}, r_{k,i}, \mathbf{s}'_{k,i}>$ from D_i
      Set $y_{k,i} \leftarrow r_{k,i} + \gamma max_{a'_{k,i}} Q^{-}(\mathbf{s}'_{k,i}, a_{k,i}{}'; \theta_i^{-})$
      Perform a gradient descent step on $(y_{k,i} - Q(\mathbf{s}_{k,i}, a_{k,i}; \theta_i))^2$ with respect to the network parameters θ_i
      Reset $\theta_i{}^{-} \leftarrow \theta_i$ for each C steps
    end for
    Set $\mathbf{s}_{t+1,1} \leftarrow \mathbf{s}'_{t,n}$
  end for
end for

steps, we then copy θ_i to $\theta_i{}^{-}$. Later, when the innermost loop completes, $\mathbf{s}_{t+1,1} = \mathbf{s}'_{t,n}$ is set for the next cycle. Finally, we repeat the above loop until the termination condition is satisfied.

2) Online Assistance: In the online assistance, bacteria are regarded as agents. In addition, the meanings of state *s* and action a are the same as those of the offline learning. Moreover, in the chemotaxis operation, the bacteria input their states $s_{t,i}$ into the DQN_1, regarding the first dimension of s. Then, DQN_1 will output the Q-value of all actions Q_{all} in this dimension. Then, we choose the action that will be executed through the roulette-wheel selection, in order to prevent the state of bacteria from getting worse, the $Relu$ function is used to filter the Q-value less than 0. Therefore, only actions with a Q-value greater than 0 will be executed to the state $s_{t,i}$, and update the state to $\mathbf{s}_{t,i+1}$, as shown in (18). Repeat the above process until the state of all dimensions has been updated. Finally, we calculate the fitness f of the state $\mathbf{s}_{t+1,1}$. Then, the dominance of bacteria and all the Pareto solutions are preserved in a non-dominated archive Γ as shown in (20), and Γ will be updated according to (21).

$$\Gamma = \Gamma \cup X_{best} \tag{20}$$

$$\Gamma = \Gamma - B \tag{21}$$

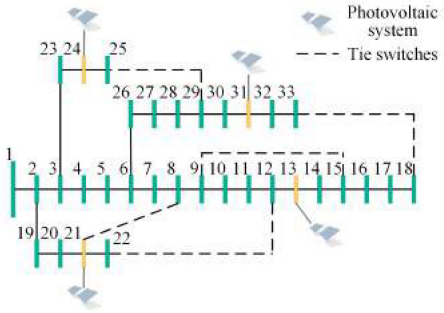


Fig. 6. The modified IEEE 33-bus distribution system.

where X_{best} stands for the set of Pareto solutions, and B represents the solution set from Γ, but currently dominated by X_{best}.

After conducting the chemotaxis operation, we calculate the crowding distance [19] of each bacterium. First, according to the dominance, we divide bacteria into two parts, i.e., N_D and N_{nonD}. Therein, N_D and N_{nonD} represent the parts of dominated and non-dominated bacteria. Then, we sort these two parts of bacteria via the crowding distance, respectively. The smaller the crowding distance, the higher the ranking. After that, we put N_D after N_{nonD}, and eliminate half of the bacteria. The remaining bacteria will crossover to generate new bacteria to keep the population size [20]. The above procedure is the reproduction operation of DRL-MBFO. Finally, when the bacterium satisfies the probability of elimination, it will be eliminated, and a new individual will be randomly generated in the feasible region. Repeat the above process until the termination conditions are satisfied.

In summary, although MBFO would lose the selection driver for solving our proposed many-objective DNR problem, bacteria will still move in a better direction due to the assistance of DQN. This makes DRL-MBFO more effective than other optimization algorithms, which will be further verified in Section IV (Simulation Studies) of this paper.

IV. Simulation Studies

A. Simulation Settings

In order to verify the effectiveness of proposed model of MDNR and the algorithm of DRL-MBFO, we conduct simulation studies on a modified IEEE 33-bus distribution system [38]. The topology of this system is shown in Fig. 6. Therein, the dotted lines represent the initial tie switches (opened). The slack generator is located at bus 1, and 4 PVs at the 13-th, 21-th, 24-th, and 31-th buses, respectively. Meanwhile, the corresponding rated active powers of each PV are assumed to be 0.375MW, 0.525MW, 0.525MW, and 1.05MW, respectively.

In addition, the forecast illuminations are set as 736 W/m^2, 888 W/m^2, 744 W/m^2 and 960 W/m^2 for the 4 PVs, respectively. Furthermore, we set forecast errors of illumination as 8% of the corresponding forecasting values. The details of PV illuminations and output power could be referred to Appendix B. Based on these settings, the integrated forecast PV power is approximately 2.14 MW, taking up to 57.63% of the total system electricity demand. We run our simulations on an Intel(R)

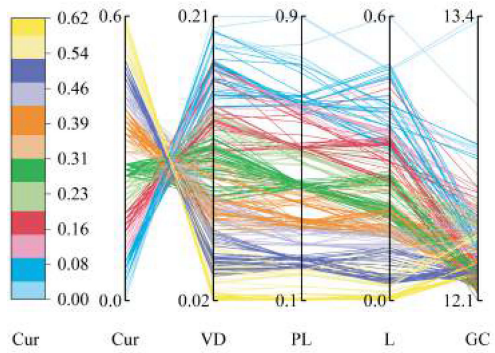


Fig. 7. Parallel coordinates plot of five objective values.

TABLE I
OBJECTIVE VALUES REGARDING ALTERNATIVES SOLUTIONS

Solution	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
Cur (MW)	0.0	0.0144	0.1647	0.4858	0.6184	0.2718	0.6052	0.6052
VD	0.2121	0.2015	0.1400	0.0529	0.0233	0.0962	0.0237	0.0260
PL (MW)	0.8803	0.7686	0.5174	0.2029	0.0852	0.4198	0.0904	0.0883
L	0.5227	0.4411	0.2994	0.0996	0.0520	0.3016	0.0530	0.0486
GC ($/h)	13.3402	12.7455	12.1293	12.1702	12.2603	12.1873	12.2118	12.1993

Core(TM) i7-9750H CPU @2.60 GHz, and all code is written in Matlab.

B. Effectiveness of MDNR Model

In this part, we mainly validate the effectiveness of proposed MDNR model. First, we use the algorithm of DRL-MBFO to solve this model, and a series of Pareto solutions can be obtained. However, as the number of objectives in our proposed MDNR is 5, it is difficult to manifest the trade-off of these objectives in the traditional cartesian coordinate. In this way, we present the Pareto front through the parallel coordinates plot [39], as shown in Fig. 7. This kind of plot is a visualization tool for presenting each objective on a separate axis, and crossing lines are formed to describe the trade-off relationship among different Pareto solutions. We can easily see that all the lines are crossed, which means that the trade-off between two arbitrary solutions. It demonstrates there exists a conflicting relationship among the 5 objectives of Cur, VD, PL, L, and GC. For instance, the high curtailment of PV power is usually accompanied by a low VD, PL, and L. On the contrary, when the amount of curtailment of PV power decreases, it would lead to the rise in some objectives, such as VD and L, which threaten secure operations of DN. Therefore, we should comprehensively consider the many objectives, rather than a single or few ones.

Furthermore, it is worth mentioning that we should choose a final solution from the obtained Pareto solutions. In this paper, Preference Ranking Organization Method for Enrichment of Evaluations (PROMETHEE) [40] is adopted to select the final solution. However, the relative weights of Cur, VD, PL, L, and GC need to be set, and the values of such weights represent the corresponding importances of objectives. That is, the greater the relative weight, the more importance of the objective. In order to qualitatively investigate the trade-off for these 5 objectives, the relative weights are set as different combinations, and the obtained results are shown in Table I. Wherein, $S_1, S_2, \ldots$, and S_8 represent different alternative solutions, their relative weights are set to be $[w_1, w_2, w_3, w_4, w_5]$ =[0.996, 0.001, 0.001, 0.001, 0.001], [0.75, 0.0625, 0.0625, 0.0625, 0.0625], [0.5, 0.125, 0.125, 0.125, 0.125], [0.25, 0.1875, 0.1875, 0.1875, 0.1875], [0.001, 0.25, 0.25, 0.25, 0.25], [0.499, 0.499, 0.001, 0.001, 0.001], [0.25, 0.25, 0.167, 0.167, 0.167], and [0.001, 0.001, 0.333, 0.333, 0.333], respectively.

TABLE II
DECISION VARIABLES OF EACH SOLUTION

Solution	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_0
Decision variable (Open switches)	28-29	24-25	6-26	3-23	28-29	23-24	28-29	25-29	3-23
	30-31	30-31	30-31	9-15	17-18	27-28	17-18	18-33	16-17
	14-15	14-15	13-14	14-15	14-15	13-14	9-15	9-15	13-14
	2-3	2-3	2-3	8-21	7-8	8-21	8-21	6-7	8-21
	8-21	8-21	11-12	21-22	9-10	11-12	10-11	10-11	9-10

[1]The topology for each solution is shown in Appendix C of this paper.

From Table I, it is easy to observe from S_1 to S_5 that the objective value of Cur is increasing with the decrease of its relative weight. Specifically, for S_1, the relative weight of Cur is 0.999, which is close to 1. Therefore, we almost merely consider minimizing the value of Cur, thus obtaining its minimum as 0 MW. However, the values of other 4 objectives (i.e., VD, PL, L, and GC) reach 0.2121, 0.8803 MW, 0.5227, and 13.3402 $/h, respectively. They are higher than those of $S_2 \sim S_5$. It shows S_1 performances worse on these 4 objectives, although it has the best value of Cur. Regarding S_3, the relative weight of Cur is decreased to 0.5, which leads to the issue that weights for other 4 objectives are increased. Therefore, their objective values for S_3 are better than those of S_1 and S_2.

On the contrary, the weight of Cur regarding S_5 is only 0.001, it means this objective is seldomly considered and the value is as high as 0.6184 MW, In this case, although S_5 could obtain the best values of VD, PL, and L, the large amount of PV power curtailments is discouraged for sustainable development of RE. From the above analysis, we can conclude that only one objective is not suitable to be considered in the DN operations with highly penetrated RE.

To conduct further comparisons, for $S_6 \sim S_8$, we reduce the relative weights of Cur and VD from 0.499 to 0.001. Specifically, the values of Cur and VD are 0.2718 MW and 0.0962 for S_6, when corresponding weights are larger. Meanwhile, PL and L are 0.4198 MW and 0.3016. On the contrary, for S_8, Cur becomes worse as 0.6052 MW, and VD is slightly better as 0.0260. However, on the hand, values of PL and L are significantly improved, i.e., 0.0883 MW and 0.0486, compared with those of S_6. In other words, if we only consider few objectives in DNR, i.e., Cur and VD, others (PL and L) would not be in desirable operations. Therefore, it is also inappropriate to consider only few objectives in our study.

Note that relative weights of these objectives can determine the final DNR scheme from the obtained Pareto solutions. However, such weights are determined by the attitude of DN operator. For simple consideration, we can set the weights as [0.2, 0.2, 0.2, 0.2, 0.2], and the solution S_0 is obtained with the corresponding objective values [0.4311, 0.0948, 0.2725, 0.1735, 12.1718]. Then, S_0 could be an alternative of the final DNR scheme. In addition, decision variables corresponding to $S_0 \sim S_8$ are also shown in Table II.

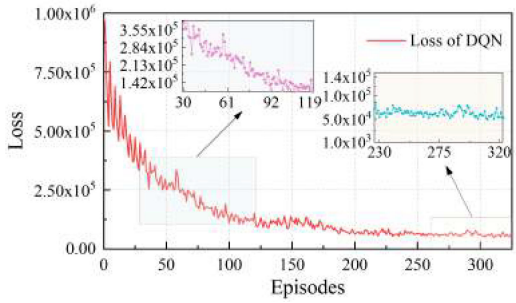


Fig. 8. The value of the loss function of DQN.

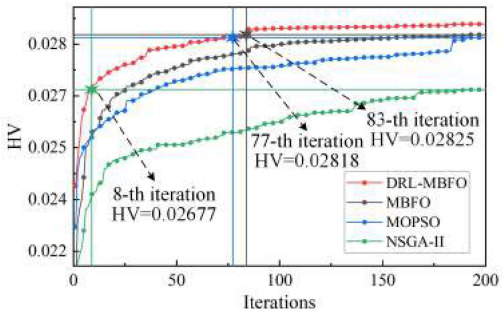


Fig. 9. Evolution of HV for the five algorithms with increasing iterations.

TABLE III
THE COMPARISON OF HV, SP, MD, AND NPS METRICS

Algorithm	HV	SP	MD	NPS
DRL-MBFO	**0.02854**	**0.01358**	**1.3457**	**369**
MBFO	0.02825	0.02033	1.3331	327
MOPSO	0.02818	0.01643	1.3370	334
NSGA-II	0.02677	0.10787	1.3012	165

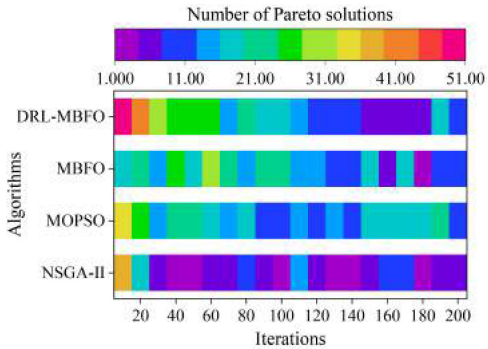


Fig. 10. Distribution of Pareto solutions.

C. Outperformance of the Proposed DRL-MBFO

In this subsection, the effectiveness of proposed DRL-MBFO is verified. First, we present the convergent curve of loss values of DQN, as shown in Fig. 8. It is seen that with the process of training, the loss value has dropped significantly and eventually tends to converge. This demonstrates that the trained DQN can well output the Q-value of each action according to the state of bacteria in DRL-MBFO. Then, the bacteria will search toward more rewarding directions. Therefore, it shows that the strategy which uses DQN to assist MBFO is feasible.

In order to further test the outperformance of DRL-MBFO, we use traditional MBFO, MOPSO, and NSGA-II for comparisons, via adopting 100 independent runs under the same circumstances. For comparing performances of these algorithms more clearly, the hypervolume (HV), the spacing (SP) index, the mean distance (MD), and the number of Pareto solutions (NPS) are used to conduct metric comparisons [41], [42]. Therein, HV is an index to assess both diversity and convergence of the obtained Pareto front, which is the volume between the front and a selected reference point in the objective space. In addition, SP is used to evaluate the extent of even distribution of obtained Pareto front. Furthermore, MD represents the mean Euclidian distance between Pareto solutions and the reference point, and NPS is used to assess the searching efficiency of finding more Pareto solutions.

First, results of the best HVs with respect to these four optimization algorithms are adopted to make comparisons, as shown in Table III. It is seen that the proposed DRL-MBFO can obtain the largest HV value, i.e., 0.02854. This means it obtains the best Pareto front with convergence and diversity than those of other three traditional algorithms. Also, DRL-MBFO outperforms regarding other indexes, i.e., SP, MD and NPS, and the corresponding values are 0.01358, 1.3457 and 369, respectively. This means our proposed algorithm has a better searching performance.

In order to further compare convergent speeds of the four algorithms, Fig. 9 shows their HV values in each iteration. It is easily observed that DRL-MBFO obtains a higher HV value in the whole stage of iterations. Moreover, compared with MBFO, DRL-MBFO only needs 83 iterations to achieve a larger value, i.e., 0.02825. Analogously, compared with MOPSO and NSGA-II, the number of iterations for DRL-MBFO to reach HV values of 0.02818 and 0.02677 are merely 77 and 8, respectively. Therefore, these results indicate the outstanding convergent speed of our proposed optimization algorithm.

The reason why the proposed DRL-MBFO obtains a better HV value within a few iterations is that the DQN will help bacteria select the best searching action based on evaluations of Q-values, via (18). In order to further show the efficiency of DRL-MBFO, we counter the number of obtained Pareto solutions for each 10 iterations, regarding these four algorithms. The results are shown in Fig. 10. It is easily seen that the number of Pareto solutions obtained by DRL-MBFO are concentrated in the first 30 iterations. Especially, 93 solutions are obtained in the first 20 iterations, which are 2 times of MOPSO and NSGA-II, and 3 times of MBFO, approximately. Therefore, the DRL-MBFO is significantly better than other algorithms in terms of searching efficiency.

For further validating the stability of DRL-MBFO, we save all the HV values of each algorithm in the 100 independent runs, and their distributions are shown in Fig. 11. The result demonstrates that DRL-MBFO has the largest mean value 2.831×10^{-2} and the smallest standard deviation (SD) 7.320×10^{-5}. This indicates that DRL-MBFO not only has the best convergence, but also excellent performance on the stability. Meanwhile, NSGA-II has the smallest mean 2.590×10^{-2} and the largest SD 3.944×10^{-4}, and the mean and SD regarding MBFO and MOPSO are between those of the DRL-MBFO and NSGA-II.

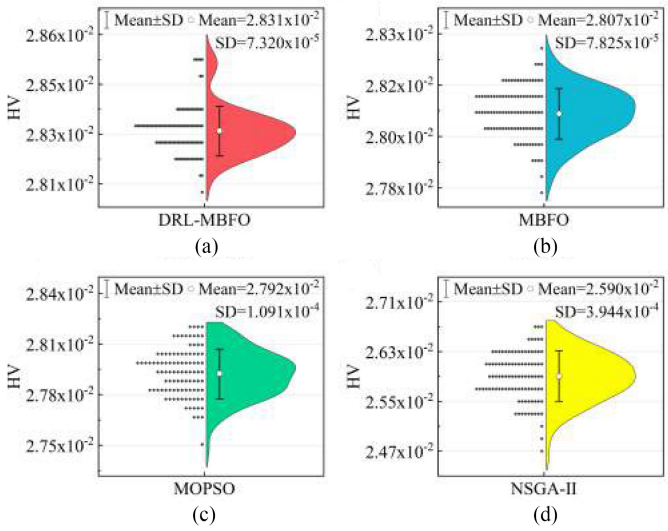


Fig. 11. The distribution of HV value.

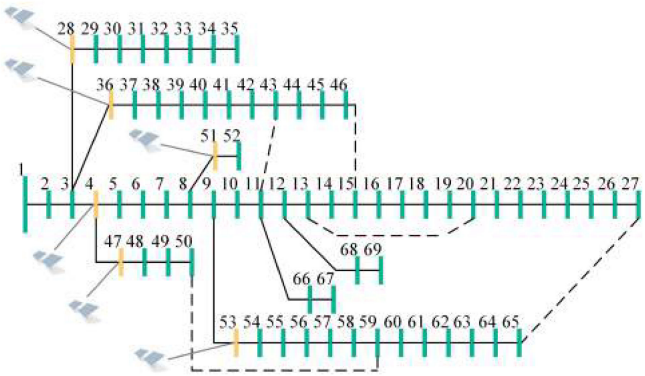


Fig. 12. The modified IEEE 69-bus system.

This verifies that DRL-MBFO is superior to MBFO, MOPSO and NSGA-II in the convergence and stability for solving our proposed MDNR model. Besides, the Kruskal-Wallis test [43] is used to discuss whether the HV values of four algorithms come from the same distribution. The obtained p-value is 4.32×10^{-75}, therefore, there exist significant differences among the results of DRL-MBFO and other three algorithms.

Moreover, similar to other algorithms, the proposed DRL-MBFO also has a certain randomness. Thus the obtained results are slightly different regarding each run. Since the HV indicator is claimed to be a popular measure [44] and it has various advantages [45] compared with other indicators, we choose the result with the largest HV indicator among dependent runs as the final one.

D. Handling With Large Systems

In order to test the scalable of MDRN and DRL-MBFO, the similar numerical simulations are conducted on the modified IEEE 69-bus system and IEEE 118-bus system [46], [47], and their topologies are shown in Fig. 12 and Fig. 13, respectively. For the IEEE 69-bus system, we set that 6 PVs are installed on the buses 4, 28, 36, 47, 51 and 53, respectively. Corresponding

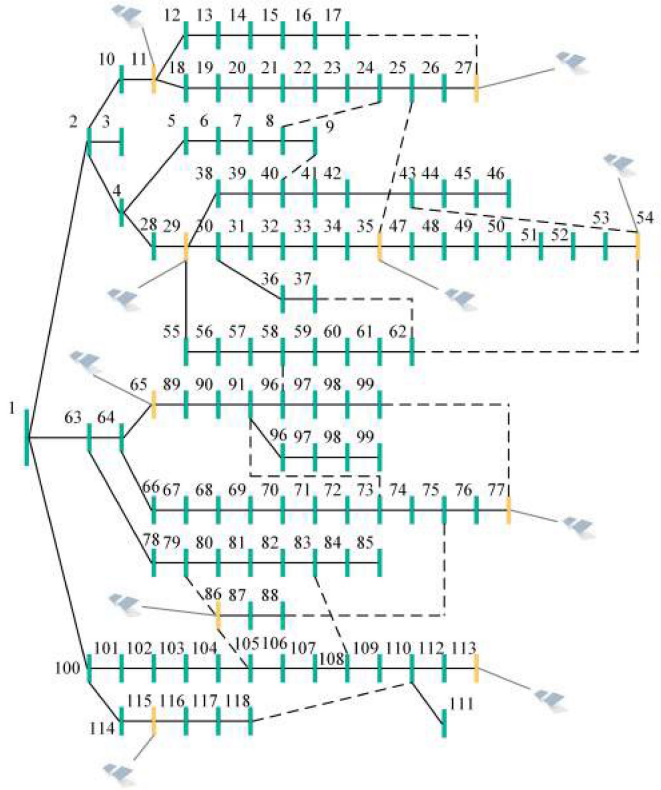


Fig. 13. The modified IEEE 118-bus system.

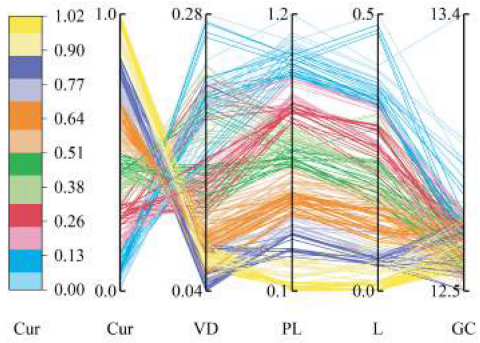


Fig. 14. Parallel coordinates plot of 5 objective values for IEEE 69-bus system.

forecast illuminations are set to be 810 W/m^2, 450 W/m^2, 530 W/m^2, 750 W/m^2, 760 W/m^2, 820 W/m^2, and the rated active power of each PV are assumed to be 1.05 MW, 0.45 MW, 0.45 MW, 0.6 MW, 0.45 MW, 0.6 MW, respectively. For the IEEE 118-bus system, the 10 PVs are set on the 11-th, 27-th, 29-th, 35-th, 54-th, 65-th, 77-th, 86-th, 113-th and 115-th buses. Meanwhile, the forecast illuminations are set to be 810 W/m^2, 450 W/m^2, 530 W/m^2, 750 W/m^2, 760 W/m^2, 820 W/m^2, 560 W/m^2, 780 W/m^2, 430 W/m^2, 660 W/m^2, and the rated active power of all PVs is assumed to be 3.0 MW.

First, in order to validate the scalability of our proposed MDNR model, we also obtain the Pareto solutions and Pareto fronts based on the modified IEEE 69-bus system and IEEE 118-bus system. In addition, we present these Pareto fronts via the parallel coordinates plot, as respectively shown in Fig. 14

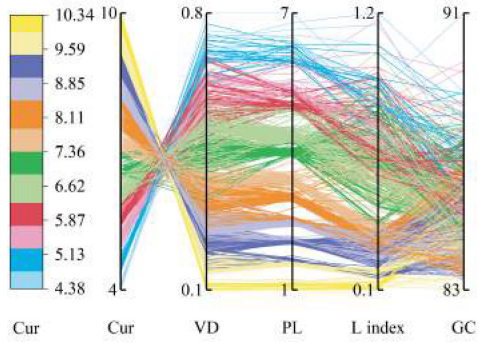


Fig. 15. Parallel coordinates plot of 5 objective values for IEEE 118-bus system.

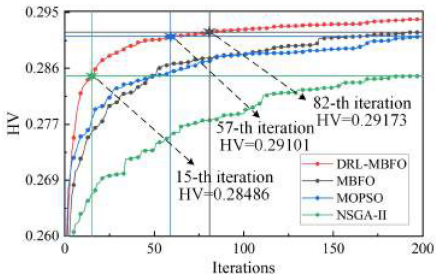


Fig. 16. Evolution of HV for the four algorithms with increasing iterations for IEEE 69-bus system.

and Fig. 15. For the IEEE 69-bus system, we can easily see that all the lines are crossed, which means that the trade-off between two arbitrary solutions. Meanwhile, most of the lines are crossed between coordinates Cur and VD, this indicates there exists a conflict relationship between objective Cur and VD. Similarly, there also exists a conflict relationship between objectives VD and PL. In contrast, the lines between coordinates PL and L index are rarely crossed, which shows that there is almost no conflict between these two objectives. The results of the IEEE 118-bus system are similar to those of the IEEE 69-bus system and IEEE 33-bus system, the number of crossed lines between other objectives is less than that of Cur and VD. Therefore, the conflict relationship mainly exists for objectives Cur and VD, and there are only slight conflicts among other objectives.

Based on the above results, the trade-off relationships among different Pareto solutions are reflected for the IEEE 33-bus, 69-bus and 118-bus distribution power networks. Overall, the high curtailment of PV power is usually accompanied by a low VD, PL, and L. Meanwhile, when the amount of curtailment of PV power decreases, it would lead to the rise in some objectives, such as VD and L, which threaten secure operations of DN. It shows that we should comprehensively consider the many objectives rather than a single or few ones. Therefore, based on the three test systems, our proposed reconfiguration model has a good scalability.

In addition, to verify the scalability of our proposed DRL-MBFO, we further compare the quality of Pareto solutions and convergent speed of the DRL-MBFO with the other algorithms. Fig. 16 and Fig. 17 show their corresponding HV values of each iteration for the IEEE 69-bus and IEEE 118-bus systems, respectively. Moreover, for the IEEE 69-bus system, DRL-MBFO obtains a higher HV value at about the 82nd iteration compared

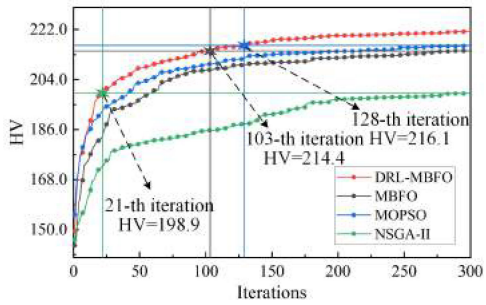


Fig. 17. Evolution of HV for the four algorithms with increasing iterations for IEEE 118-bus system.

with MBFO, i.e., 0.29173. Analogously, compared with MOSPO and NSGA-II, the number of iterations for DRL-MBFO to reach HV values of 0.29101 and 0.28486 are merely 57 and 15, respectively. This shows that the convergent speed of the proposed DRL-MBFO is faster than other algorithms. At the same time, the HV value of DRL-MBFO reaches 0.29372 when the number of iterations achieves 200, which shows that the obtained Pareto solution set by DRL-MBFO has a better quality.

In the IEEE 118-bus system, when the number of iterations researches 300, the HV values of DRL-MBFO, MBFO, MOPSO, NSGA-II are 221.3, 214.4, 216.1 and 198.9, respectively. This indicates the Pareto solutions obtained by DRL-MBFO has the best performance. Besides, compared with MBFO, MOPSO and NSGA-II, the number of iterations for DRL-MBFO to reach their best HV are merely 103, 128 and 21, respectively.

Therefore, the above results together with that of IEEE 33-bus system show that the proposed DRL-MBFO not only obtains the best Pareto solutions, but also has a faster convergence speed. This indicates that the proposed DRL-MBFO is also effective for large systems. To be concluded, the above results have verified that our proposed approach has a good scalability.

In summary, based on above analysis, the searching performance and convergent speed of DRL-MBFO have the notable advantage over MBFO, MOPSO, and NSGA-II in solving the proposed MDNR model. Also, we have verified that the necessity of considering many objectives in the distribution network reconfiguration considering the environment of highly penetrated renewable energy.

V. Conclusion

In this paper, we have proposed a many-objective distribution network reconfiguration (MDNR) model for obtaining the relationship among photovoltaic power curtailment, voltage deviation, power loss, statistic voltage stability, and generation cost. As traditional optimization algorithms perform poorly in dealing with the MDNR model, a deep reinforcement learning assisted multi-objective bacterial foraging optimization algorithm (DRL-MBFO) is further proposed.

Case studies are conducted on the modified IEEE 33-bus, 69-bus, and 118-bus power distribution systems. By comparing the obtained Pareto solutions, the trade-off among different objectives is verified, which indicates that it is unsuitable to consider only few objectives in the DN operations with highly penetrated RE. Therefore, the effectiveness of the proposed MDNR model is verified. Then, the comparisons among DRL-MBFO and other

three conventional optimization algorithms are conducted. It is also verified that the proposed DRL-MBFO have notable advantage in searching performance and convergent speed.

However, there are still some limitations of the DRL-MBFO. For instance, the DRL-MBFO consists of various parameters, such as the learning rate, discount factor, the number of iterations of DRL, and the number of layers of neural network, etc. Therefore, it is not easy to well train this algorithm, and various attempts and trails are needed. Furthermore, in the DRL-MBFO, the non-dominated sorting is used to obtain Pareto solutions. However, this kind of sorting is not efficient, as many objectives are studied in our proposed MDRN model. Therefore, using more efficient sorting methods would reduce the computational complexity of DRL-MBFO, which is also our future research direction.

APPENDIX

A. The Loop Encoding Method and Radiality Constraint

1) The Loop Encoding Method: The efficient encoding of the distribution network topology could significantly reduce the computational burden of the optimization procedure. In this paper, the loop encoding method is used, as it is more efficient than traditional 0/1 binary encoding [48]–[50]. In this method, each branch in the DN is equivalent to a switch. As the DN operates radially, it should be considered that the number of closed switches is one less than the number of buses [48]. In other words, the number of open switches should be equal to the number of loops in the DN [48]. Meanwhile, the radial structure also indicates that there exist no loops in the DN, thus each loop at least has one open switch [50]. For above reasons, all the possible structures could be formed through the way that we can respectively open one switch in each loop. It means that each dimension of the decision variables indicates the serial number of the open switch in the corresponding loop [49]. Compared with the traditional 0/1 binary encoding method, our used loop encoding could effectively reduce the dimension of decision variables [51].

2) Radiality Constraint: In order to ensure the radiality constraint, we should distinguish whether the DN is connected and looped. First, according to the graph theory [52], if A' represents the adjacency matrix of the structure of DN, the structure is connected if and only if:

$$\min\left(\sum_{i=1}^{N_B-1}\left(A'\right)^i + E(N_B)\right) \neq 0 \qquad \text{(A-1)}$$

where $E(N_B)$ is N_B-dimension unit matrix, and $\min(\cdot)$ means the minimum number of matrix elements.

Afterwards, we further test whether the DN is loop, if not, then it is radial. However, in our work, the loop encoding method is used. Due to the number of closed switches is one less than the number of buses, therefore, if the structure of DN is connected, it has no loops [52]. Therefore, in this paper, we just need to distinguish the DN is connected or not, and the DN is radial if it is connected.

Specifically, we take the IEEE 33-bus system as a simple example, there are 5 loops as shown in Fig. 18, the dimension

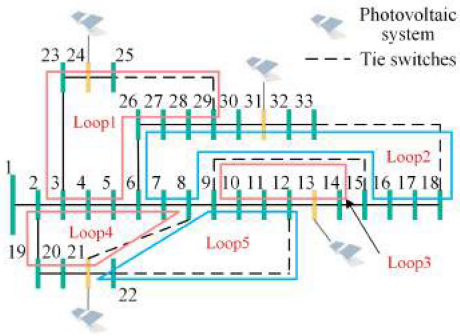


Fig. 18. The loops in IEEE 33-bus system.

TABLE IV
DIMENSION OF DECISION VARIABLES AND CORRESPONDING SWITCHES

Dimension of independent variables	Corresponding switches
1	3-4-5-6-26-27-28-29-25-24-23-3
2	6-7-8-9-15-16-17-18-33-32-31-30-29-28-27-26-6
3	9-10-1-12-13-14-15-9
4	2-3-4-5-6-7-8-21-20-19-2
5	8-9-10-11-12-22-21-8

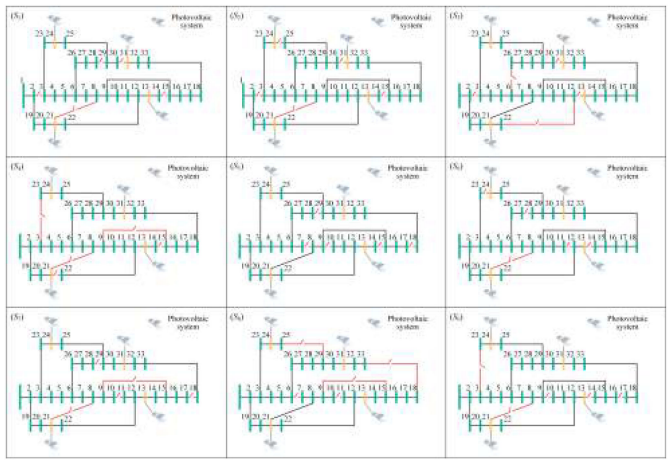


Fig. 19. The topology regarding each solution.

of decision variables and corresponding switches are listed in Table IV. Moreover, when decision the decision variables $u = [1, 1, 1, 1, 1]$, it indicates the corresponding switches 3-4, 6-7, 9-10, 2-3, 8-9 are opened. On the contrary, it means that the corresponding switches 4-5, 7-8, 10-1, 3-4, 9-10 are opened when $u = [2, 2, 2, 2, 2]$.

B. PV Power

The following model is presented in this paper to calculate the PV power, referred to [53].

$$P_{pre}(I_t) = \begin{cases} \dfrac{P_r}{I_s K_c} I_t^2 & 0 \le I_t \le K_c \\ \dfrac{P_r}{I_s} I_t & K_c \le I_t \le I_s \\ P_r & I_t \ge I_s \end{cases} \qquad \text{(A-2)}$$

where $P_{pre}(I_t)$ stands for the output power when the forecast illumination is I_t. K_c is the illumination when the PV conversion efficiency reaches the maximum, and I_s represents the standard illumination of the PV system. They are set as 300 W/m^2 and 1000 W/m^2, respectively. In addition, P_r is the rated active power of the PV system. Then, the actual PV power P_t is formulated as follows.

$$P_t = P_{pre} + P_{err} \tag{A-3}$$

$$P_{err} \sim N(0, \sigma^2) \tag{A-4}$$

where P_{pre} and P_{err} stand for the forecast power and the corresponding forecast error, respectively. Meanwhile, P_{err} is usually assumed to obey the Normal distribution [54], [55], and the expectation and the standard deviation are 0 and σ, respectively.

C. The Topology Regarding Each Solution

The topologies of the obtained solutions $S_0 \sim S_8$ are shown in Fig. 19.

REFERENCES

[1] X. Wu, X. Wang, and C. Qu, "A hierarchical framework for generation scheduling of microgrids," *IEEE Trans. Power Del.*, vol. 29, no. 6, pp. 2448–2457, Dec. 2014.

[2] A. Whiteman, S. Rueda, D. Akande, N. Elhassan, G. Escamilla, and I. Arkhipova, "Renewable capacity statistics 2020," *Int. Renewable Energy Agency*, 2019.

[3] Z. Li, S. Jazebi, and F. de Leon, "Determination of the optimal switching frequency for distribution system reconfiguration," *IEEE Trans. Power Del.*, vol. 32, no. 4, pp. 2060–2069, Aug. 2017.

[4] M. Zeraati, M. E. H. Golshan, and J. M. Guerrero, "Distributed control of battery energy storage systems for voltage regulation in distribution networks with high PV penetration," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3582–3593, Jul. 2018.

[5] Y. Wu, C. Lee, L. Liu, and S. Tsai, "Study of reconfiguration for the distribution system with distributed generators," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1678–1685, Jul. 2010.

[6] A. M. Tahboub, V. R. Pandi, and H. H. Zeineldin, "Distribution system reconfiguration for annual energy loss reduction considering variable distributed generation profiles," *IEEE Trans. Power Del.*, vol. 30, no. 4, pp. 1677–1685, Aug. 2015.

[7] Y. Xu, Z. Y. Dong, R. Zhang, and D. J. Hill, "Multi-timescale coordinated voltage/var control of high renewable-penetrated distribution systems," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4398–4408, Nov. 2017.

[8] S. Chalise, H. R. Atia, B. Poudel, and R. Tonkoski, "Impact of active power curtailment of wind turbines connected to residential feeders for overvoltage prevention," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 471–479, Apr. 2016.

[9] S. Martn-Martnez, E. Gmez-Lazaro, A. Molina-Garcia, and A. Honrubia-Escribano, "Impact of wind power curtailments on the spanish power system operation," in *Proc. IEEE PES Gen. Meeting Conf. Expo.*, 2014, pp. 1–5.

[10] J. Li, Z. Xu, J. Zhao, and C. Zhang, "Distributed online voltage control in active distribution networks considering PV curtailment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5519–5530, Oct. 2019.

[11] Y. Takenobu, N. Yasuda, S. Kawano, S. Minato, and Y. Hayashi, "Evaluation of annual energy loss reduction based on reconfiguration scheduling," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1986–1996, May 2018.

[12] A. Asrari, S. Lotfifard, and M. Ansari, "Reconfiguration of smart distribution systems with time varying loads using parallel computing," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2713–2723, Nov. 2016.

[13] Q. Peng, Y. Tang, and S. H. Low, "Feeder reconfiguration in distribution networks based on convex relaxation of OPF," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1793–1804, Jul. 2015.

[14] H. Wu, P. Dong, and M. Liu, "Distribution network reconfiguration for loss reduction and voltage stability with random fuzzy uncertainties of renewable energy generation and load," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5655–5666, Sep. 2020.

[15] M. R. Dorostkar-Ghamsari, M. Fotuhi-Firuzabad, M. Lehtonen, A. Safdarian, and A. S. Hoshyarzade, "Stochastic operation framework for distribution networks hosting high wind penetrations," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 344–354, Jan. 2019.

[16] J. Zhan, W. Liu, C. Y. Chung, and J. Yang, "Switch opening and exchange method for stochastic distribution network reconfiguration," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 2995–3007, Jul. 2020.

[17] J. Cheng, G. G. Yen, and G. Zhang, "A many-objective evolutionary algorithm with enhanced mating and environmental selections," *IEEE Trans. Evol. Comput.*, vol. 19, no. 4, pp. 592–605, Aug. 2015.

[18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[19] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256–279, Jun. 2004.

[20] B. Niu, H. Wang, J. Wang, and L. Tan, "Multi-objective bacterial foraging optimization," *Neurocomputing*, vol. 116, pp. 336–345, 2013.

[21] M. Alhazmi, P. Dehghanian, S. Wang, and B. Shinde, "Power grid optimal topology control considering correlations of system uncertainties," *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 5594–5604, Nov./Dec. 2019.

[22] A. Kavousi-Fard and T. Niknam, "Optimal distribution feeder reconfiguration for reliability improvement considering uncertainty," *IEEE Trans. Power Del.*, vol. 29, no. 3, pp. 1344–1353, Jun. 2014.

[23] M. Rastegar, M. Fotuhi-Firuzabad, H. Zareipour, and M. Moeini-Aghtaieh, "A probabilistic energy management scheme for renewable-based residential energy hubs," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2217–2227, Sep. 2017.

[24] Y. Huang, L. Wang, W. Guo, Q. Kang, and Q. Wu, "Chance constrained optimization in a home energy management system," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 252–260, Jan. 2018.

[25] A. Spring, G. Wirth, G. Becker, R. Pardatscher, and R. Witzmann, "Grid influences from reactive power flow of photovoltaic inverters with a power factor specification of one," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1222–1229, May 2016.

[26] Y. Liu, J. Bebic, B. Kroposki, J. de Bedout, and W. Ren, "Distribution system voltage performance analysis for high-penetration PV," in *Proc. IEEE Energy 2030 Conf.*, 2008, pp. 1–8.

[27] J. Park, W. Liang, J. Choi, A. A. El-Keib, M. Shahidehpour, and R. Billinton, "A probabilistic reliability evaluation of a power system including solar/photovoltaic cell generator," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2009, pp. 1–6.

[28] F. Ding and K. A. Loparo, "Hierarchical decentralized network reconfiguration for smart distribution systems-part II: Applications to test systems," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 744–752, 2015.

[29] S. Civanlar, J. Grainger, H. Yin, and S. Lee, "Distribution feeder reconfiguration for loss reduction," *IEEE Trans. Power Del.*, vol. 3, no. 3, pp. 1217–1223, 1988.

[30] Y. Guo, Q. Wu, H. Gao, S. Huang, B. Zhou, and C. Li, "Double-time-scale coordinated voltage control in active distribution networks based on MPC," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 294–303, Jan. 2020.

[31] P. Kessel and H. Glavitsch, "Estimating the voltage stability of a power system," *IEEE Trans. Power Del.*, vol. 1, no. 3, pp. 346–354, Jul. 1986.

[32] Y. Li, M. Li, and Q. Wu, "Energy saving dispatch with complex constraints: Prohibited zones, valve point effect and carbon tax," *Int. J. Elect. Power Energy Syst.*, vol. 63, pp. 657–666, 2014.

[33] X. Su, M. A. S. Masoum, and P. J. Wolfs, "PSO and improved BSFS based sequential comprehensive placement and real-time multi-objective control of delta-connected switched capacitors in unbalanced radial MV distribution networks," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 612–622, Jan. 2016.

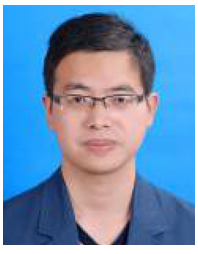
[34] Y. Li, "Deep reinforcement learning: An overview," 2017, *arXiv:1701.07274*.

[35] Q. Zhang and C. Shu, "Performance investigation of learning rate decay in LMS-based equalization," *IEEE Photon. Technol. Lett.*, vol. 33, no. 2, pp. 109–112, Jan. 2021.

[36] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[37] V.-H. Bui, A. Hussain, and H.-M. Kim, "Double deep Q-learning-based distributed operation of battery energy storage system considering uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 457–469, Jan. 2020.

[38] Z. Tian, W. Wu, and B. Zhang, "A mixed integer quadratic programming model for topology identification in distribution network," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 823–824, Jan. 2016.
[39] H. Julian and W. Daniel, "State of the art of parallel coordinates," in *STAR Proc. Eurographics*, 2013, pp. 95–116.
[40] J.-P. Brans and Y. De Smet, "PROMETHEE methods," *Multiple Criteria Decis. Anal.: State Art Surv.*, vol. 78, pp. 187–219, 2016.
[41] K. Shang and H. Ishibuchi, "A new hypervolume-based evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 24, no. 5, pp. 839–852, Oct. 2020.
[42] M. de Athayde Costa e Silva, C. E. Klein, V. C. Mariani, and L. dos Santos Coelho, "Multiobjective scatter search approach with new combination scheme applied to solve environmental/economic dispatch problem," *Energy*, vol. 53, pp. 14–21, 2013.
[43] D. K. de Vries and Y. Chandon, "On the false-positive rate of statistical equipment comparisons based on the Kruskal-Wallis h statistic," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 3, pp. 286–292, Aug. 2007.
[44] K. Bringmann and T. Friedrich, "Approximation quality of the hypervolume indicator," *Artif. Intell.*, vol. 195, pp. 265–290, 2013.
[45] K. Shang, H. Ishibuchi, L. He, and L. M. Pang, "A survey on the hypervolume indicator in evolutionary multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 25, no. 1, pp. 1–20, Feb. 2021.
[46] M. Baran and F. Wu, "Optimal capacitor placement on radial distribution systems," *IEEE Trans. Power Del.*, vol. 4, no. 1, pp. 725–734, Jan. 1989.
[47] D. Zhang, Z. Fu, and L. Zhang, "An improved TS algorithm for loss-minimum reconfiguration in large-scale distribution systems," *Electric Power Syst. Res.*, vol. 77, no. 5/6, pp. 685–694, 2007.
[48] A. Asrari, T. Wu, and S. Lotfifard, "The impacts of distributed energy sources on distribution network reconfiguration," *IEEE Trans. Energy Convers.*, vol. 31, no. 2, pp. 606–613, Jun. 2016.
[49] A. Asrari, S. Lotfifard, and M. S. Payam, "Pareto dominance-based multi-objective optimization method for distribution network reconfiguration," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1401–1410, May 2016.
[50] S. Tan, J.-X. Xu, and S. K. Panda, "Optimization of distribution network incorporating distributed generators: An integrated approach," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2421–2432, Aug. 2013.
[51] A. Mendes, N. Boland, P. Guiney, and C. Riveros, "Switch and tap-changer reconfiguration of distribution networks using evolutionary algorithms," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 85–92, Feb. 2013.
[52] J. Zhu *et al.*, "An improved PSO algorithm based on statistics for distribution network reconfiguration to increase the penetration of distributed generations," in *Proc. IET Int. Conf. Resilience Transmiss. Distrib. Netw.*, 2015, pp. 1–6.
[53] C. O. Incio and C. L. T. Borges, "Stochastic model for generation of high-resolution irradiance data and estimation of power output of photovoltaic plants," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 952–960, Apr. 2018.
[54] Z. Ziadi *et al.*, "Optimal voltage control using inverters interfaced with PV systems considering forecast error in a distribution system," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 682–690, Apr. 2014.
[55] L. Zhang, W. Tang, J. Liang, P. Cong, and Y. Cai, "Coordinated day-ahead reactive power dispatch in distribution network based on real power forecast errors," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2472–2480, May 2016.



Yuanzheng Li (Member, IEEE) received the M.S. degree and Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, and South China University of Technology (SCUT), Guangzhou, China, in 2011 and 2015, respectively. He is currently an Associate Professor with HUST. He has authored or coauthored several peer-reviewed papers in international journals. His current research interests include deep learning, reinforcement learning, hydrogen energy, smart grid, optimal power system/microgrid scheduling and decision making, stochastic optimization considering large-scale integration of renewable energy into the power system and multi-objective optimization.

Guokai Hao received the B.S. degree in mechanical engineering from the Zhengzhou University, Zhengzhou, China. He is currently working toward the M.S. degree with Artificial Intelligence and Automation, the Huazhong University of Science and Technology (HUST), Wuhan, China. His research interests include power system optimization, grid integration of renewable energy, deep reinforcement learning, and graph neural network.

Yun Liu (Senior Member, IEEE) received the B.Eng. (First Class Hons.) and Ph.D. degrees from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2011 and 2016, respectively. His work experience includes the University of Central Florida, Nanyang Technological University, and Shenzhen University. He is currently an Associate Professor with the School of Electrical Power Engineering, South China University of Technology, Guangzhou, China. His research interests include power system stability analysis, microgrid, integrated energy system, and distributed control/optimization.

Yaowen Yu received the B.S. degree in automation from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2011, and the Ph.D. degree in electrical engineering from the University of Connecticut, Storrs, CT, USA, in 2016. He is currently an Associate Professor with HUST. His research interests include power system optimization, grid integration of renewable energy, and economics of electricity markets.

Zhixian Ni (Graduate Student Member, IEEE) received the B.S. degree in automation from the Wuhan University of Technology, China. He is currently working toward the M.S. degree with China-EU Institute for Clean and Renewable Energy, Huazhong University of Science and Technology, Wuhan, China. His current research interests include planning, optimal scheduling of large-scale renewable energy integrated power system, demand response and artificial intelligence and its application in the smart grid.

Yong Zhao received the Ph.D. degree in automation and system engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1996. He is currently a Professor with the Huazhong University of Science and Technology. His current research interests include system engineering and artificial intelligence and its application in the smart grid. He was the Deputy Dean of Research Institute of Future Power Grid supported by State Grid Corporation of China and Huazhong University of Science and Technology.