

Ensemble learning for optimal active power control of distributed energy resources and thermostatically controlled loads in an islanded microgrid

Xiaoshun Zhang^{a,b}, Dezhi Wang^a, Tao Yu^{a,*}, Zhao Xu^b

^a College of Electric Power, South China University of Technology, 510640 Guangzhou, China

^b Department of Electrical Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong

Abstract-- To achieve an effective coordination between the secondary control and the tertiary control of load frequency control (LFC), a new optimal active power control (OAPC) is constructed for real-timely changing the operating points of distributed energy resources (DERs) and thermostatically controlled loads (TCLs) in an islanded microgrid. A large number of TCLs are integrated as a load aggregator (LA) for participating the secondary control of LFC, which can enhance the dynamic response performance due to their much faster response speeds compared with that of distributed generators. Since OAPC is a nonsmooth nonlinear optimization with a quite short implementation period, a novel model-free ensemble learning (EL) is proposed to rapidly obtain a high-quality optimal solution for it. EL based OAPC is composed of multiple sub-optimizers and a learning concentrator, where each sub-optimizer is responsible for providing the exploitation and exploration samples to the learning concentrator, while the reinforcement learning based concentrator is mainly used for knowledge learning and knowledge transfer. Case studies are thoroughly carried out to verify the performance of EL based OAPC in an islanded microgrid with 12 DERs and 900 TCLs.

Keywords –Ensemble learning, Optimal active power control, Load frequency control, thermostatically controlled loads, Islanded microgrid

* Corresponding author. Tel.: +86 13002088518.
E-mail address: taoyu1@scut.edu.cn (Tao Yu).

Nomenclature			
Variables		Parameters	
T^{in}	indoor temperature	P_{AC}^{r}	rated power of the air conditioner
T_{R}	refrigerator temperature	$P_{\text{EWH}}^{\text{r}}$	rated power of the electric water heater
T_{HW}	hot water temperature	ΔP_m^{min}	minimum reserve capability of the m th generator
s_{AC}	switch state of air conditioner	ΔP_m^{max}	maximum reserve capability of the m th generator
s_{R}	switch state of refrigerator	$\Delta P_{\text{LA}}^{\text{min}}$	minimum reserve capability of LA
s_{EWH}	switch state of electric water heater	$\Delta P_{\text{LA}}^{\text{max}}$	maximum reserve capability of LA
Δf	frequency deviation	ΔP_m^{rate}	maximum ramp rate of the m th generator
ΔP_{Σ}	total generation command	C_{LA}	regulation cost coefficient of LA
ΔP_{M}	power mismatch	a_m, b_m, c_m	fuel cost coefficients of the m th generator
ΔP_m	generation command of the m th generator of secondary control	α	knowledge learning factor
P_m^0	scheduled operating point of the m th generator of tertiary control	γ	discount factor
ΔP_m^1	real-time corrective power of the m th generator of primary control	ε	exploitation rate
ΔP_{LA}	generation command of LA of secondary control	p_{m}	positive multiplier
f_m^{c}	fuel cost of the m th generator	E	number of learning agents
Q	Q-value matrix (knowledge)	M	penalty factor
ΔQ	knowledge increment	N	number of sub-optimizers
R	reward function	L	length of binary bit string
a_{rand}	a random action	k_{max}	maximal iteration number
Q^{n0}	initial knowledge of the new task	Abbreviations	
Q^{h*}	optimal knowledge of the h th source task	OAPC	optimal active power control
r_h	similarity between the h th source task and the new task	LFC	load frequency control
SA_p^k	state-action pairs set of the best individual	DERs	distributed energy resources
F	fitness function	TCLs	thermostatically controlled loads
Indices		LA	load aggregator
m	index of generator	WT	wind turbine
i	index of controllable variable	PV	photovoltaic
j	index of binary bit	TOU	time of use
k	index of iteration	AC	air conditioner
e	index of individual	Re	refrigerator
h	index of source task	WH	water heater
o	index of current best sub-optimizer	EL	ensemble learning
p	index of sub-optimizer	GA	genetic algorithm
		PSO	particle swarm optimization
		PROP	proportional
		GSO	group search optimizer
		IPM	interior point method

1. Introduction

Microgrid has become a popular way to effectively integrate various small-scale distributed energy resources (DERs) and associated loads [1]. When the microgrid is operated in the islanded mode, load frequency control (LFC) will become one of the most crucial operation tasks for the microgrid without a strong support from the main grid [2]. In general, LFC for an islanded microgrid is a hierarchical control with three levels [3], i.e., primary control (droop control), secondary control, and tertiary control.

The primary control commonly maintains the system frequency stability via the active power-frequency droop control, which is achieved by a local control manner [4]. In order to improve the droop control performance, various modified droop

control techniques have been investigated, including e.g., adaptive and robust droop controls [5],[6]. However, the primary control will inevitably lead to the frequency deviation. Hence, the secondary control is used to restore the frequency to the nominal value with a centralized controller [7], which usually employs a proportional-integral (PI) controller with fixed participation factors to calculate the supplementary active power set points of all the reserve resources [8]. Finally, the tertiary control aims to achieve an optimal operation according to the forecast data of load demand and renewables generation [9], and the implementation period commonly ranges from 15 min to several hours [10].

In the past decades, though extensive investigations have been carried out for these three controls of LFC in an islanded microgrid, most of these studies [7]-[15] did not address two important issues, as follows:

- *Online optimization of participation factors*: for the secondary control, the output of the centralized controller (total generation command) is simply distributed among all the reserve resources with the fixed participation factors, which is usually optimized by an offline manner. Hence, the fixed participation factors easily lead to an unsatisfied dynamic response performance of secondary control when the system operation state change greatly (e.g., outage of a unit).
- *Coordination between secondary control and tertiary control*: the secondary control will change the units' operating points due to the unexpected power disturbance, but it often neglects the objective function of tertiary control, which may lead to a low economic operation for an islanded microgrid as the implementation period of the tertiary control is much longer than that of the secondary control. Although the total generation cost of tertiary control was incorporated in real-time LFC in [15], it did not consider the improvement of dynamic response performance for the secondary control, and various constraints (e.g., GRC) of different units for participating the secondary control.

To handle these issues, this paper proposes a novel optimal active power control (OAPC) of participation factors optimization of secondary control via an effective coordination with primary and tertiary controls. By considering the dynamic response performance of secondary control [16], OAPC is essentially a nonsmooth nonlinear optimization with a min-max objective function. Consequently, the traditional gradient-based algorithms are difficult to find a high-quality optimal solution for OAPC with a given initial solution [17] as they are highly dependent on its accurate mathematical model. Compared with that, the model-free heuristic optimization algorithms are more flexible and more efficient for global optimization of OAPC, but often consume more computation time and can hardly meet the requirement of online implementation (1 to 16 s). In order to overcome these problems, this paper proposes a novel ensemble learning (EL) for rapidly searching a high-quality optimal solution of OAPC with the following features

- EL is composed of multiple sub-optimizers and a learning concentrator, in which various sub-optimizers can effectively enhance the exploration ability with different optimization mechanisms, while the learning concentrator can efficiently implement a deep exploitation by utilizing the current searching results from the sub-optimizers. Hence, the high-quality of

the obtained optimum can be guaranteed.

- The learning concentrator is not only able to learn the knowledge via the self-learning and the guidance by the sub-optimizers, but also can achieve the knowledge transfer from the source tasks to a new task. Therefore, the computation time of EL can be dramatically reduced, which is adequate to satisfy the online optimization of OAPC.

The remaining of this paper is organized as follows: Section 2 presents the mathematical model of OAPC. Section 3 provides the basic principle of EL. The design of EL for OAPC is developed in Section 4, while simulation results and discussions are given in Section 5. Finally, Section 6 concludes the paper.

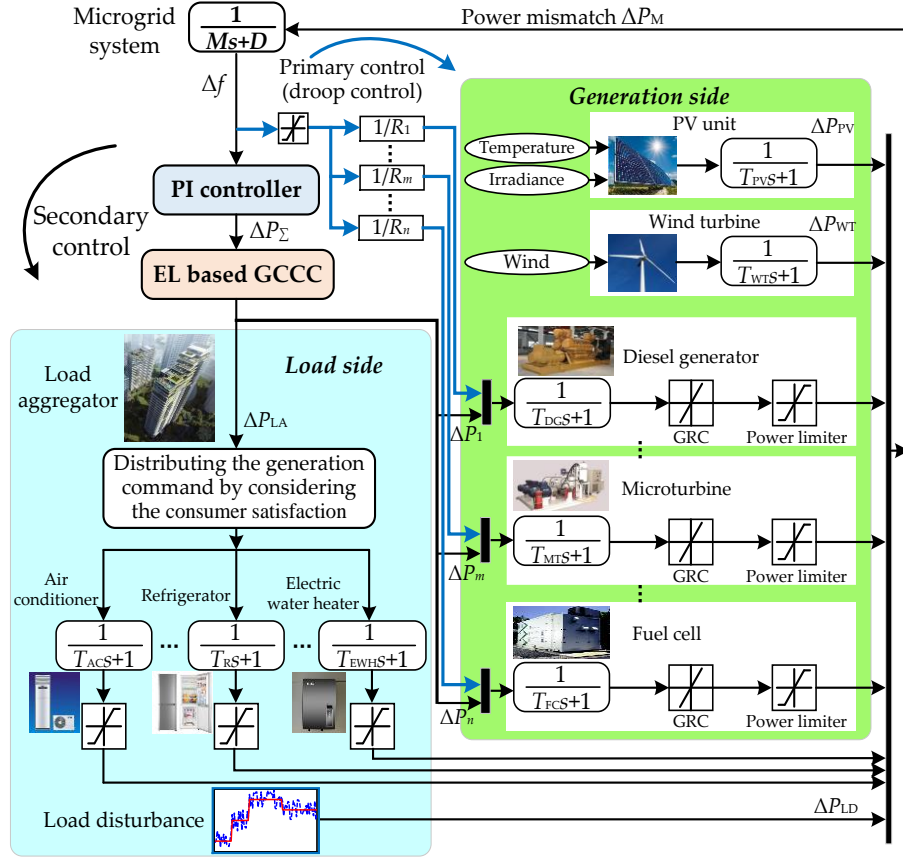


Fig. 1. The LFC framework in an islanded microgrid.

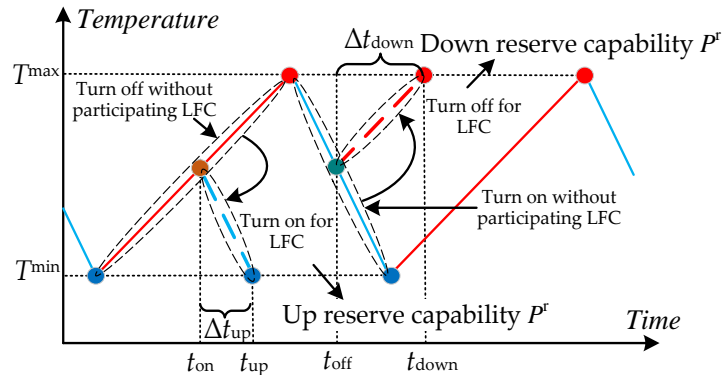


Fig. 2. The reserve capability of air conditioner.

2. Optimal active power control of LFC

2.1. LFC framework in an islanded microgrid

In an islanded microgrid, the power mismatch ΔP_M may results from the power output fluctuations of wind turbine (WT) and photovoltaic (PV) unit with the changeable weather, load disturbance, and operation faults. As the frequency deviates from its nominal value, then the frequency regulation generators or loads will change their operating points according to the distributed primary control and the centralized secondary control [2],[3] based on the initial operating points of tertiary control, as given in Fig. 1. Note that the PI controller is used for tracking the power mismatch according to the dynamic frequency deviation Δf , where ΔP_Σ is the controller output, i.e., the total generation command. Then EL based OAPC will optimally distribute ΔP_Σ among the controllable DERs and the load aggregator (LA), thus the frequency can be quickly restored with a low operation cost.

2.2. Reserve capability of LA

In this study, a large number of household consumers are aggregated as a LA for secondary control of LFC [18], thus the optimization difficulty of OAPC can be significantly reduced with much fewer controllable variables. Particularly, only three types of common thermostatically controlled loads (TCLs) including air conditioner, refrigerator, and electric water heater, are employed for LFC as they can excellently act as energy storage [19]. When they are disconnected, the temperature will slightly change in a few minutes, thus the consumer comfort can be guaranteed at a high level.

To evaluate the reserve capability of LA, it needs to acquire the real-time operating states of all the TCLs [10]. For satisfying the consumer comfort, each TCL should keep the temperature within an ideal range via the constant spaced switches. For example, the air conditioner needs to maintain the indoor temperature within the range from T^{\min} to T^{\max} , which will be automatically turned on if the indoor temperature T^{in} exceeds T^{\max} , and turned off if the indoor temperature is lower than T^{\min} , as shown in Fig. 2. When the thermostatically controllable load is off, it can provide the down reserve capability for LFC via turning on. Similarly, it can provide the up reserve capability from the ON state to the OFF state.

Note that the maximal participation time of each TCL for LFC is determined by its current operating state and temperature variation feature, as

a) Air conditioner: The indoor temperature mainly depends on its previous value, outdoor temperature, rated temperature, and the house heat resistance, as [20]

$$T^{\text{in}}(t_c + \Delta t) = T^{\text{in}}(t_c) e^{-\Delta t / (RC_{\text{air}})} + [T^{\text{out}}(t_c) - R \cdot P_{\text{AC}}^r \cdot s_{\text{AC}}(t_c)] \cdot [e^{-\Delta t / (RC_{\text{air}})} - 1] \quad (1)$$

where $T^{\text{out}}(t_c)$ is the outdoor temperature at time t_c ; Δt is the time length of ON or OFF state; R is the heat resistance of the house; C_{air} is the specific heat capacity of air; P_{AC}^r is the rated power of the air conditioner; and s_{AC} is the switch state of air conditioner, which is 1 at ON state and 0 at OFF state.

b) Refrigerator: The refrigerator temperature is mainly determined by its cooling effect of the ON state, and warming effect of the OFF state, as [21]

$$T_{\text{R}}(t_c + \Delta t) = T_{\text{R}}(t_c) + \Delta t [\gamma_{\text{R}} - \alpha_{\text{R}} \cdot s_{\text{R}}(t_c)] \quad (2)$$

where $T_{\text{R}}(t_c)$ is the refrigerator temperature at time t_c ; γ_{R} is the warming effect of the OFF state; α_{R} is the cooling effect of the ON state; and s_{R} is the switch state of refrigerator.

c) Electric water heater: According to the law of conservation of energy, the hot water temperature can be calculated as [22]

$$T_{\text{HW}}(t_c + \Delta t) = \frac{T_{\text{HW}}(t_c)(V_{\text{tank}} - f_{\text{r}}(t_c) \cdot \Delta t)}{V_{\text{tank}}} + \frac{T_{\text{IW}}(t_c) \cdot f_{\text{r}}(t_c) \cdot \Delta t}{V_{\text{tank}}} + C_1 \left[C_2 \cdot P_{\text{EWH}}^r \cdot s_{\text{EWH}}(t_c) \cdot \eta_{\text{EWH}} - \frac{A_{\text{tank}} \cdot (T_{\text{HW}}(t_c) - T^{\text{in}}(t_c))}{R_{\text{tank}}} \right] \frac{\Delta t}{V_{\text{tank}}} \quad (3)$$

where $T_{\text{HW}}(t_c)$ is the hot water temperature at time t_c ; $T_{\text{IW}}(t_c)$ is the inlet water temperature at time t_c ; V_{tank} is the volume of the tank; $f_{\text{r}}(t_c)$ is the hot water flow rate at time t_c ; P_{EWH}^r is the rated power of the electric water heater; and s_{EWH} is the switch state of electric water heater; η_{EWH} is the efficiency factor; A_{tank} is the surface area of the tank; and R_{tank} is the tank heat resistance.

It can be found from (1)-(3) that the maximal participation time Δt_{up} or Δt_{down} of refrigerator and electric water heater can be directly solved due to the simple linear equations, while that of air conditioner requires a iterative method (e.g., Newton method) to approximate the real solution as it is essentially a transcendental equation.

Assume the unexpected power disturbance occurs at time t_c , then the TCL can participate secondary control of LFC if its maximal participation time is longer than the minimal participation time requirement Δt_{fc} , as

$$t_{\text{m}} = \begin{cases} \Delta t_{\text{up}} - \Delta t_{\text{fc}} \geq 0, & \text{if } \Delta P_{\Sigma} < 0 \\ \Delta t_{\text{down}} - \Delta t_{\text{fc}} \geq 0, & \text{if } \Delta P_{\Sigma} > 0 \end{cases} \quad (4)$$

where t_{m} is the time margin of participation secondary control

Hence, the up and down reserve capabilities of LA are equal to the total rated power of all the TCLs with positive t_{m} under the current ΔP_{Σ} , respectively. For the sake of consumer comfort, the generation command ΔP_{LA} of LA is distributed to all the TCLs according to the descending order of t_{m} , i.e., the one with a larger positive t_{m} will be dispatched in priority until the power balance constraint can be satisfied.

It is worth noting that each TCL could participate in secondary control LFC only if it is fitted with control device, temperature sensor, user interface, and communication device [10]. Particularly, the control device is not only used for switching the load to response the control command from the LA, but also for electric measurements. Besides, the communication device

can achieve the interaction between the TCL and the LA.

2.3 Mathematical model of OPAC

In this study, the proposed OAPC mainly focuses on the dynamic response performance and the operation cost. The first one can be improved by minimizing the maximum of all the regulation generators' ramp time f_1 , while the second one f_2 consists of the fuel cost of the generation side and the regulation cost of LA. In order to reduce the solving difficulty and shorten the computation time, these two objectives are transformed into a single overall objective f by the multiplication method, thus OAPC can be constructed with the power balance constraints, power output limits, and generation rate constraints (GRC), as

$$\text{Minimize } f(x) = f_1(x) \cdot f_2(x) \quad (5)$$

$$\text{with } \begin{cases} f_1(x) = \max_{m=1,2,\dots,n} (|\Delta P_m + \Delta P_m^1| / \Delta P_m^{\text{rate}}) \\ f_2(x) = \sum_{m=1}^n f_m^c (\Delta P_m + \Delta P_m^1) + C_{\text{LA}} \Delta P_{\text{LA}} \\ f_m^c (\Delta P_m) = a_m (P_m^0 + \Delta P_m + \Delta P_m^1)^2 + b_m (P_m^0 + \Delta P_m + \Delta P_m^1) + c_m \end{cases} \quad (6)$$

$$\text{subject to } \begin{cases} \Delta P_{\Sigma} = \sum_{m=1}^n \Delta P_m + \Delta P_{\text{LA}} \\ \Delta P_{\Sigma} \cdot \Delta P_{\text{LA}} \geq 0 \\ \Delta P_{\text{LA}}^{\min} \leq \Delta P_{\text{LA}} \leq \Delta P_{\text{LA}}^{\max} \\ \Delta P_{\Sigma} \cdot \Delta P_m \geq 0, m = 1, 2, \dots, n \\ |P_m(t) - P_m(t+1)| \leq \Delta P_m^{\text{rate}}, m = 1, 2, \dots, n \\ P_m = P_m^0 + \Delta P_m + \Delta P_m^1 \\ \Delta P_m^{\min} \leq \Delta P_m + \Delta P_m^1 \leq \Delta P_m^{\max}, m = 1, 2, \dots, n \end{cases} \quad (7)$$

where ΔP_m is the generation command of the m th generator of secondary control; P_m^0 is the scheduled operating point of the m th generator of tertiary control; ΔP_m^1 is the real-time corrective power of the m th generator of primary control; P_m is the real-time operating point command of the m th generator; ΔP_{LA} is the generation command of LA of secondary control; ΔP_m^{rate} is the maximum ramp rate of the m th generator; f_m^c is the fuel cost of the m th generator; C_{LA} is the regulation cost coefficient of LA; a_m , b_m , and c_m are the fuel cost coefficients of the m th generator; t is the time period index; ΔP_m^{\min} and ΔP_m^{\max} are the minimum and maximum reserve capability of the m th generator, respectively; $\Delta P_{\text{LA}}^{\min}$ and $\Delta P_{\text{LA}}^{\max}$ are the minimum and maximum reserve capability of LA, respectively; and n is the number of generators which participate LFC.

3. Ensemble learning

3.1. Optimization framework

As shown in Fig. 3, EL consists of multiple sub-optimizers and a learning concentrator. For each new optimization task, each

sub-optimizer will search a potential higher quality optimum based on its own searching mechanism, while the learning concentrator will approximate the optimal knowledge of the current task via a knowledge transfer from the previous optimal knowledge of the source tasks. Then the knowledge of the learning concentrator can be updated by continuous interaction with the environment, which includes three steps, i.e., 1) implementing different actions (solutions) to the environment based on exploitation and exploration, and the guidance from multiple sub-optimizers; 2) getting the feedback reward and state from the environment; and 3) updating the knowledge with the reinforcement learning.

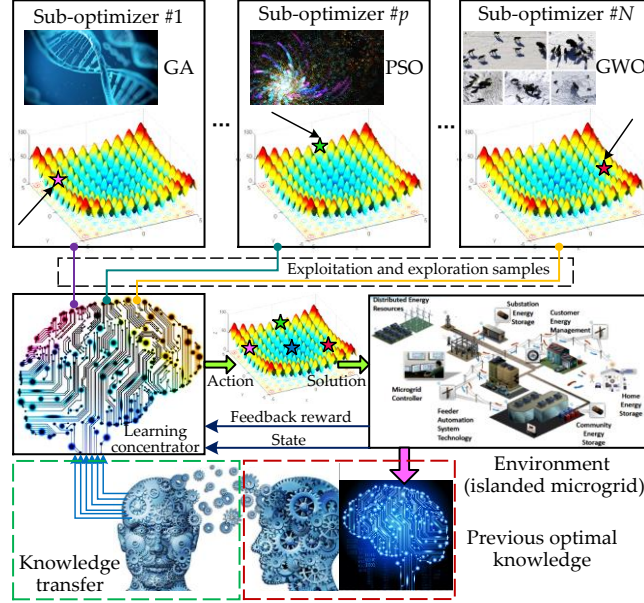


Fig. 3. Optimization framework of EL.

3.2. Learning concentrator

3.2.1. Knowledge learning

In order to construct the knowledge for each continuous controllable variable, a binary Q-learning [23] with associative memory is adopted for storing and learning the knowledge. Hence, the knowledge for each continuous controllable variable can be represented as a binary bit string, where each bit (0 or 1) has a knowledge value. Like other reinforcement learning, the binary Q-learning can update the knowledge according to the feedback reward after implementing an action to the environment at the current state, as follows:

$$\begin{cases} Q_{il}^{k+1}(s_{il}^{ke}, a_{il}^{ke}) = Q_{il}^k(s_{il}^{ke}, a_{il}^{ke}) + \alpha \Delta Q_{il}^k \\ \Delta Q_{il}^k = R_{il}^e(s_{il}^{ke}, s_{il}^{k+1,e}, a_{il}^{ke}) + \gamma \max_{a_{il} \in A_{bb}} Q_{il}^k(s_{il}^{k+1,e}, a_{il}) - Q_{il}^k(s_{il}^{ke}, a_{il}^{ke}) \end{cases} \quad (8)$$

where α is the knowledge learning factor; γ is the discount factor; the subscript i and l represent the i th controllable variable and the j th binary bit, respectively; superscripts k and e denote the k th iteration and the e th individual, respectively, with $e=1,2, \dots, E$; E is the number of learning agents; Q_{il} is the Q-value matrix (knowledge) of the l th binary bit for the i th

controllable variable; ΔQ is the knowledge increment; (s, a) means the state-action pair; $R(s^k, s^{k+1}, a^k)$ is the reward function of a transition from state s^k to s^{k+1} used under a selected action a^k ; a_{il} is any alternative actions (0 or 1); and A_{bb} is the action space of each binary bit.

In general, a learning agent prefers choosing a binary bit with a large knowledge value as it can get a larger feedback reward with a higher probability, which also easily leads to a low quality optimum. To properly balance the exploitation and exploration, the ε -Greedy rule [24] is used for determining the actions, as follows:

$$a_{il}^{ke} = \begin{cases} \arg \max_{a_{il} \in A_{bb}} Q_{il}^k(s_{il}^{ke}, a_{il}), & \text{if } q_0 \leq \varepsilon \\ a_{\text{rand}}, & \text{otherwise} \end{cases} \quad (9)$$

where q_0 is a random number with a probability uniformly distributed in the range $[0, 1]$; ε is the exploitation rate; and a_{rand} denotes a random action (exploration).

3.2.2. Knowledge transfer

For most optimization algorithms, one of their common drawbacks is incapable of rapidly handing a new task by utilizing the knowledge of the source tasks (previous tasks), thus it often takes a long computation time. To fill up this gap, the knowledge transfer from the source tasks to a new task is introduced in EL, which can be described as follows [25]:

$$Q_{il}^{n0} = \sum_{h=1}^H r_h Q_{il}^{h*}, \quad i = 1, 2, \dots, n+1; l = 1, 2, \dots, L \quad (10)$$

where Q_{il}^{n0} is the initial knowledge of the new task; Q_{il}^{h*} is the optimal knowledge of the h th source task; L is the length of the binary bit string; and r_h denotes the similarity between the h th source task and the new task, which will be a larger number if the corresponding source task is more similar with the new task, i.e., the new task will exploit more knowledge from the optimal knowledge of the h th source task, with $0 \leq r_h \leq 1$ and $\sum_{h=1}^H r_h = 1$.

3.3. Interaction between sub-optimizers and learning concentrator

EL is an ensemble system with various optimization algorithms in nature, which can reach a most satisfactory solution via a comprehensive evaluation of multiple decision strategies [26], as illustrated in Fig. 3. And this is also the greatest advantage of EL. Generally speaking, a more diversity of sub-optimizers is beneficial to the quality of the final optimal solution, thus various types of optimization algorithms are encouraged to be employed as the sub-optimizers, e.g., genetic algorithm (GA) [27], PSO [28], and grey wolf optimizer (GWO) [29]. To achieve an effective comprehensive evaluation, the interaction between sub-optimizers and learning concentrator can be designed as follows:

$$\begin{cases} \mathbf{Q}_{il}^{k+1}(s_{il}^{kp}, a_{il}^{kp}) = \mathbf{Q}_{il}^k(s_{il}^{kp}, a_{il}^{kp}) + \alpha \Delta \mathbf{Q}_{il}^k \\ \Delta \mathbf{Q}_{il}^k = R_{il}^p(s_{il}^{kp}, s_{il}^{k+1,p}, a_{il}^{kp}) + \gamma \max_{a_{il} \in A_{il}} \mathbf{Q}_{il}^k(s_{il}^{k+1,p}, a_{il}) - \mathbf{Q}_{il}^k(s_{il}^{kp}, a_{il}^{kp}) \end{cases} \quad (11)$$

$$\mathbf{SA}_p^k = \left\{ (s_{il}^{kp}, a_{il}^{kp}) \mid i=1,2,\dots,n+1; l=1,2,\dots,L \right\} \quad (12)$$

$$o = \arg \min_{p=1,2,\dots,N} \left[F(\mathbf{SA}_p^k) \right] \quad (13)$$

$$\mathbf{SA}_{lc}^k = \begin{cases} \mathbf{SA}_o^k, & \text{if } F(\mathbf{SA}_o^k) < F(\mathbf{SA}_{lc}^k) \\ \mathbf{SA}_{lc}^k, & \text{otherwise} \end{cases} \quad (14)$$

where \mathbf{SA}_p^k is the state-action pairs set of the best individual obtained by the p th sub-optimizer, $p=1,2,\dots,N$; N is the number of sub-optimizers; \mathbf{SA}_{lc}^k is the state-action pairs set of the best individual obtained by the learning concentrator; F is the fitness function; and o is the index of the current best sub-optimizer with the smallest fitness function.

4. Design of EL for OAPC

4.1. Feedback reward

The feedback reward needs to be designed by fully integrated the mathematical model of OAPC from (5)-(7), while a feasible solution with a smaller f will obtain a higher feedback reward. To accelerate the knowledge learning of EL, the cooperative mechanism of ant colony [30] is introduced for the calculation of feedback reward, as

$$R_{il}^e(s_{il}^{ke}, s_{il}^{k+1,e}, a_{il}^{ke}) = \begin{cases} \frac{P_m}{F(\mathbf{SA}_{lc}^k)}, & \text{if } (s_{il}^{ke}, a_{il}^{ke}) \in \mathbf{SA}_{lc}^k \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

$$F = \begin{cases} f + \sum_{u=1}^{N_c} M(Z_u - Z_u^{\lim})^2, & \text{if violated} \\ f, & \text{otherwise} \end{cases} \quad (16)$$

where p_m is a positive multiplicator; M is the penalty factor to guarantee a feasible solution; Z_u denotes the u th constraint in (7); and Z_u^{\lim} is the constraint limit of Z_u .

4.2 Knowledge transfer between different tasks

Based on the mathematical model of OAPC (5)-(7), the distinctions between different optimization tasks mainly includes the total generation command ΔP_Σ and the reserve capabilities of all units. Hence, the knowledge transfer of EL should be designed with these two elements.

Firstly, the deviation of total generation command can be regarded as the similarity between the sources tasks and a new task, which can be calculated as [16]

$$r_h = \begin{cases} \frac{\Delta P_{\Sigma}^b - \Delta P_{\Sigma}^{\text{nt}}}{\Delta P_{\Sigma}^b - \Delta P_{\Sigma}^{b-1}}, & \text{if } h = b-1 \\ \frac{\Delta P_{\Sigma}^{\text{nt}} - \Delta P_{\Sigma}^{b-1}}{\Delta P_{\Sigma}^b - \Delta P_{\Sigma}^{b-1}}, & \text{if } h = b \end{cases} \quad (18)$$

where $\Delta P_{\Sigma}^{\text{nt}}$ is the total generation command of a new task, with $\Delta P_{\Sigma}^b \leq \Delta P_{\Sigma}^{\text{nt}} < \Delta P_{\Sigma}^{b+1}$; ΔP_{Σ}^b and ΔP_{Σ}^{b+1} are the total generation commands of the most two similar source tasks, respectively.

Secondly, a continual pre-learning is carried out to update the optimal knowledge of the dynamic source tasks with changing reserve capabilities. In this paper, the implementation period of pre-learning is set to be 15 min as the reserve capability of each resource is basically unchanged during this period.

4.3 Design of sub-optimizers

For the sake of an efficient interaction between sub- optimizers and learning concentrator, all the sub-optimizers need to be compatible with the binary Q-learning. Therefore, ten various binary heuristic algorithms are employed as the sub-optimizers of EL, including GA [27], six various binary PSO with different transfer functions [31], binary bat algorithm, binary dragonfly algorithm, and binary GWO.

4.4 Overall execution procedure

In summary, the overall execution procedure of EL for OAPC is provided in Table 1, where k_{\max} is the maximal iteration number. In order to shorten the computation time, all the sub-optimizers and the learning concentrator are executed in parallel based on multi-core CPU.

Table 1

Overall Execution Procedure of EL for OAPC

EL inputs: ΔP_{Σ} and operating parameters of each reserve resource;
EL outputs: optimal knowledge matrices and optimal generation commands;
Initialize the learning parameters;
If the current task is a source task
Initialize the knowledge without knowledge transfer;
Else
Initialize the knowledge with knowledge transfer using (10) and (18);
End-if
Repeat
1) Select an action for each binary bit at the current state by (9);
2) Searching a potential high-quality optimum with each sub-optimizer;
3) Calculate the fitness function of each solution using (5)-(7) and (16);
4) Calculate the feedback reward of each state-action pair by (15);
5) Implement the interaction between the sub-optimizers and the learning concentrator using (11)-(14);
6) Update the knowledge of each controllable variable using (8);
7) Let $k = k + 1$; If $k > k_{\max}$, then iteration terminates, otherwise return to step 1).
End

5. Case Studies

The performance of EL for OAPC is thoroughly evaluated on an islanded microgrid, which is compared with that of five algorithms, including proportional (PROP) method [32], GA [27], PSO [28], group search optimizer (GSO) [33], and interior point method (IPM), where PROP is an engineering method with fixed participation factors; the population size and the maximal iteration number of three heuristic algorithms are set to be 150 and 150, respectively. Through the trial-and-error, the main parameters of EL are given in Table 2. Moreover, all the simulations are undertaken in Matlab R2016a by a small server with Intel(R) Xeon (R) E5-2670 v3 CPU at 2.3 GHz with 64 GB of RAM.

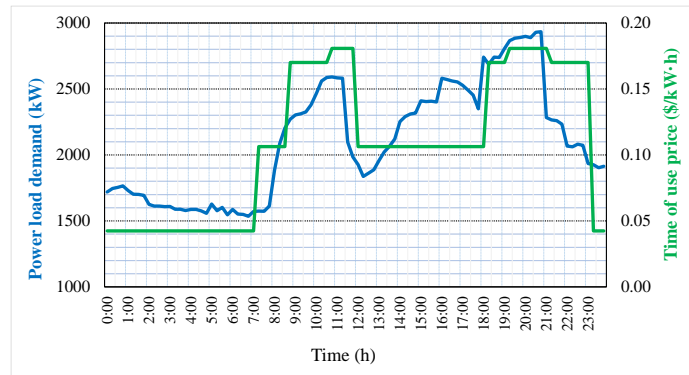


Fig. 4. Power load demand curve and time of use price at the current day.

Table 2

The Parameters used in EL for OAPC

Parameter	Range	Pre-learning	Online optimization
α	$0 < \alpha < 1$	0.1	0.8
γ	$0 < \gamma < 1$	0.1	0.1
ε	$0 < \varepsilon < 1$	0.8	0.9
p_m	$p_m \geq 0$	1	1
E	$E \geq 1$	50	10
M	$\chi > 0$	10^8	10^8
N	$N > 0$	10	10
L	$L \geq 1$	16	16
k_{\max}	$k_{\max} \geq 2$	150	30

Table 3

Main parameters of various DERs.

Type	No.	ΔP_m^{rate} (kW/s)		Fuel cost coefficient		
		Up	Down	a_m	b_m	c_m
Diesel generator (DG)	#1	1	1	0.0004	0.2348	10.9952
	#2	1	1	0.0004	0.2348	10.9952
	#3	1	1	0.0004	0.2348	10.9952
	#4	1	1	0.0004	0.2348	10.9952
Microturbine (MT)	#1	1.8	2.4	0.0002	0.1164	5.2164
	#2	1.8	2.4	0.0002	0.1164	5.2164
	#3	1.2	1.6	0.0002	0.1088	5.2164
	#4	1.2	1.6	0.0002	0.1088	5.2164
	#5	1.8	2.4	0.0002	0.1164	5.2164
	#6	1.8	2.4	0.0002	0.1164	5.2164
Fuel cell (FC)	#1	6	6	0.0003	0.1189	3.5442
	#2	6	6	0.0003	0.1189	3.5442

Table 4

Main Parameters of TCLs.

Type	Number	Rated power (kW)	Ideal temperature range (°C)	
			T^{\min}	T^{\max}
Air conditioner (AC)	300	2.5	24	26
Refrigerator (Re)	300	0.5	2	8
Electric water heater (WH)	300	1.5	40	50

5.1 System Model

The islanded microgrid consists of 12 DERs and 300 household consumers for secondary control of LFC, where the main parameters of all the regulation units and TCLs are given in Table 3 and Table 4, respectively. Note that the reserve capabilities of all the regulation units can be calculated according to their maximum power outputs and the current operating points, while the number of TCLs can be determined by their time margins of participation secondary control of LFC in (4). Moreover, the power load demand curve and time of use (TOU) price are given in Fig. 4, where the regulation cost coefficient C_{LA} is set to be 30% of TOU price for the electricity compensation fee of LA; each household consumer has an air conditioner (AC), a refrigerator (Re) and an electric water heater (WH). In the following case studies, the implementation period of the secondary control is set to be 4 s.

5.2 Pre-learning for offline optimization of source tasks

EL requires a pre-learning to obtain the optimal knowledge matrices of all source tasks, which will then be transferred to the initial knowledge matrices of a new task for online optimization of OAPC. For the testing system, the deviation of total generation command ΔP_{Σ} is divided into 12 intervals, i.e., $\{[-300, -250), [-250, -200), \dots, [250, 300]\}$ kW, with one endpoint for each source task. As illustrated in Fig. 5, EL can converge to the optimal culture matrices and the optimal solution of the source task ($\Delta P_{\Sigma}=200$ kW) within 25 iterations, where ΔQ represents the matrix 2-norms of knowledge matrix difference. It also can be found that each sub-optimizer can search an optimal solution with a low fitness function for the current source task, which can guarantee a high-quality optimal solution of learning concentrator due to diversity of sub-optimizers. Following the same calculation manner, the optimal knowledge matrices of other source tasks can be obtained by EL.

5.3 Online optimization of new tasks

5.3.1 Study of knowledge transfer

Fig. 6 shows the convergence of different algorithms for the new task ($\Delta P_{\Sigma}=225$ kW) in online optimization. According to (10) and (18), the initial knowledge matrices of this new task can be generated via a linear weighted sum of two most similar sources tasks ($\Delta P_{\Sigma}=200$ kW and $\Delta P_{\Sigma}=250$ kW). It can be clearly observed that the proposed EL can rapidly approximate a

higher quality optimal solution compared with other algorithms, which is also adequate to satisfy the online optimization of OAPC as the execution time (about 1.5 s) is less than the implementation period of secondary control (4 s). This fully proves that the knowledge transfer can dramatically accelerate the convergence rate of EL.

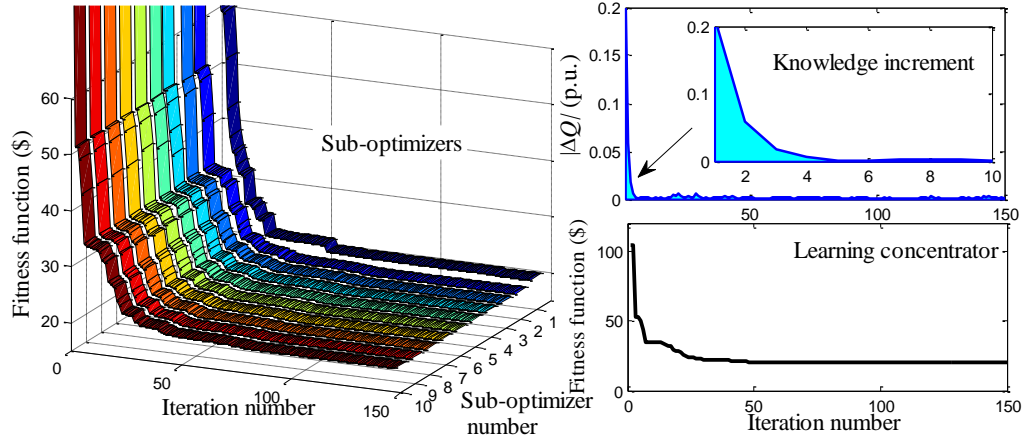


Fig. 5. Convergence of EL for the source task ($\Delta P_{\Sigma}=200$ kW) in pre-learning.

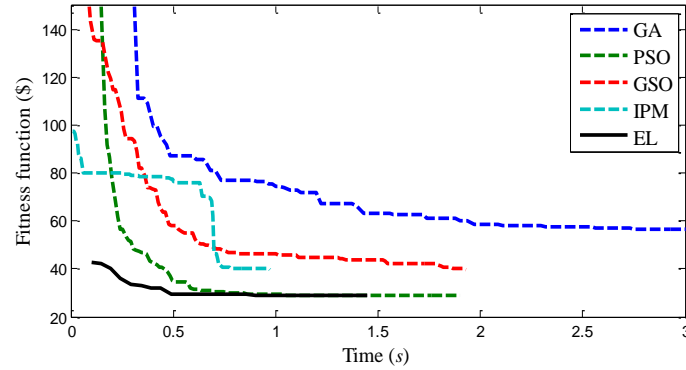


Fig. 6. Convergence of different algorithms for the new task ($\Delta P_{\Sigma}=225$ kW) in online optimization.

5.3.2 Study on a step power disturbance

In order to further test the performance of EL, it is put into the closed-loop secondary control for online optimization of different real-time new tasks in an islanded microgrid. Fig. 7 provides the online optimization results of different algorithms for a step power disturbance ($\Delta P_M=225$ kW) from 12:00 p.m. to 12:10 p.m. It can be seen from Fig. 7(a) that all the total active power output deviations can match well with the power disturbance, thus the frequency deviation can be rapidly recovered to zero (See Fig. 7(b)). Among them the frequency deviation obtained by EL is the smallest, while the fluctuation of the total active power output deviation obtained by IPM is the largest as its optimal solution is mainly determined by the initial solution. As shown in Fig. 7(c), the total operation cost of EL is the lowest, which also confirms that EL can converge to the high-quality optimums for different new tasks. Besides, the total operation cost obtained PROP is the highest as it is lack of dynamic optimization of OAPC and adopt the fixed participation factors for different new tasks. Furthermore, it can be observed from Fig. 7(d) that all of LA, FC #1, and FC #2 bear more power disturbance than that of other units due to their faster ramp rates and

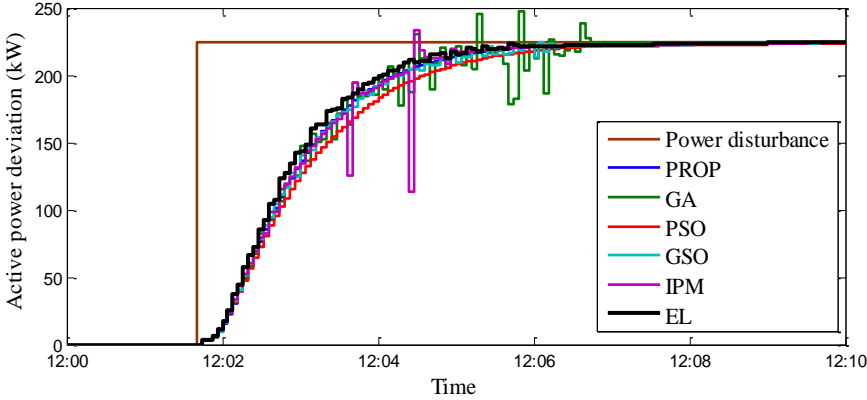
288 relatively low operation costs.

289 As the step power disturbance is positive, thus the controllable TCLs should be turned off from the ON state for providing
290 the up reserve capability. It can be found from Fig. 8 that three types of TCLs can adequately satisfy the consumers comfort
291 because all of the indoor temperature, refrigerator temperature, and hot water temperature can be controlled in the ideal range
292 after participating secondary control of LFC. This validates the effectiveness of the dispatch strategy for TCLs by considering
293 the time margin in (4).

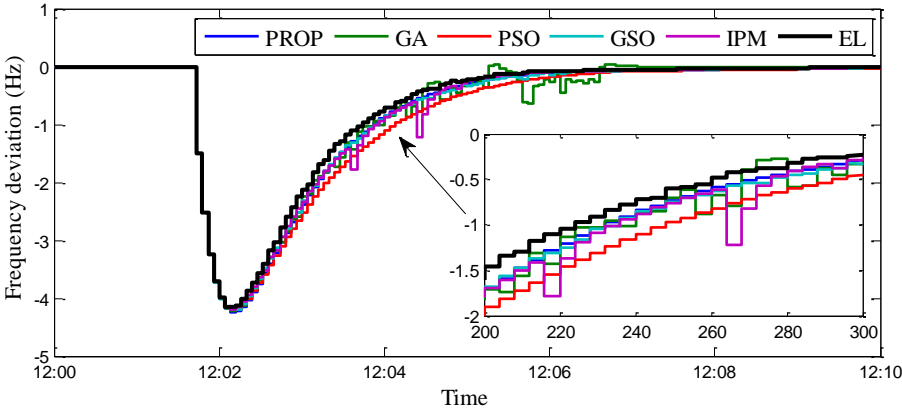
294
295

296
297

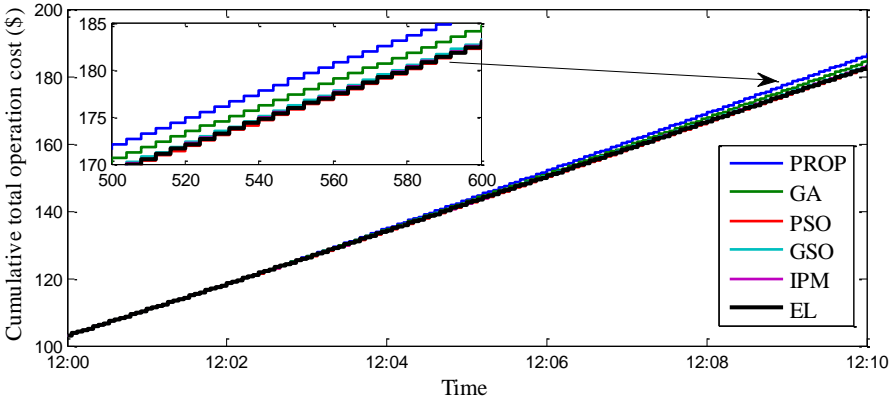
298
299



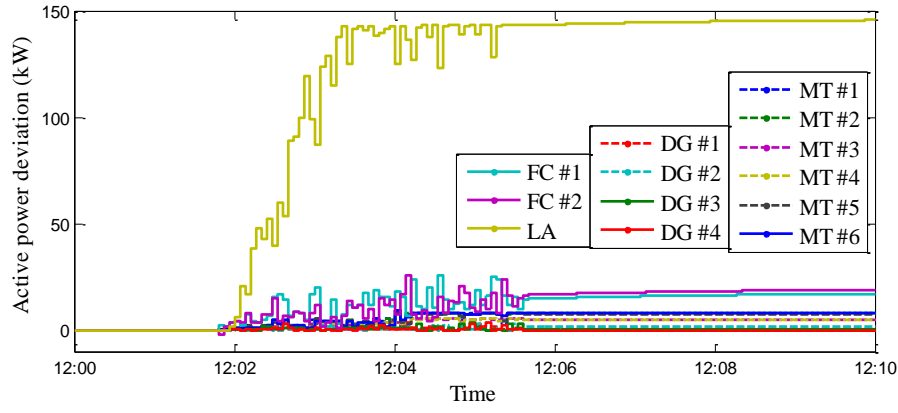
(a) Deviation of total active power output



(b) Frequency deviation

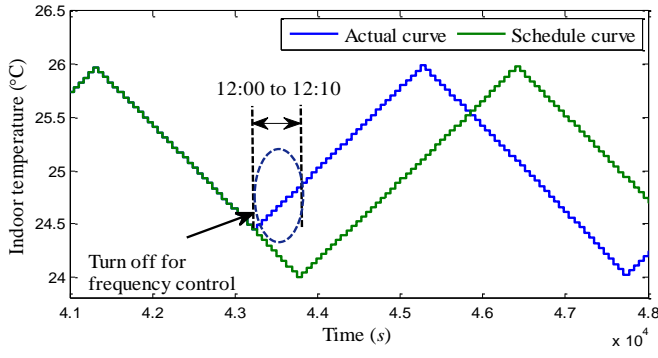


(c) Cumulative total operation cost

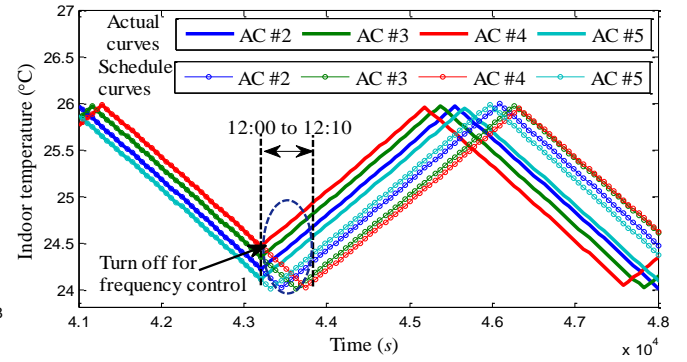


(d) Active power output deviation of different units obtained by EL

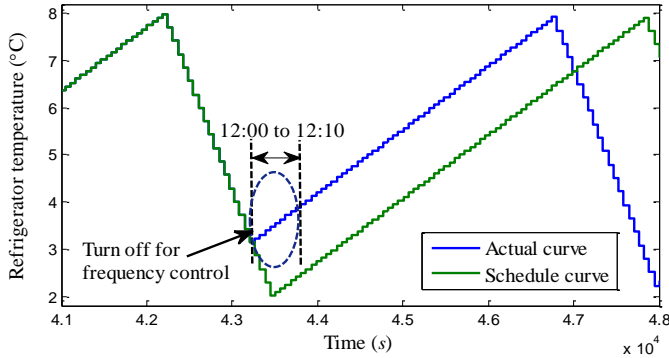
Fig. 7. Online optimization of different algorithms for a step power disturbance ($\Delta P_M=225$ kW) from 12:00 p.m. to 12:10 p.m.



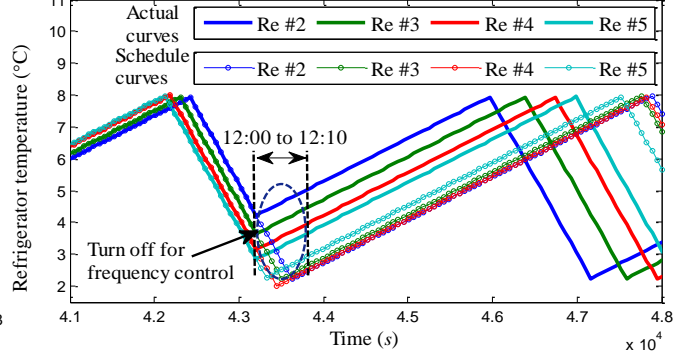
(a) Indoor temperature of AC #1



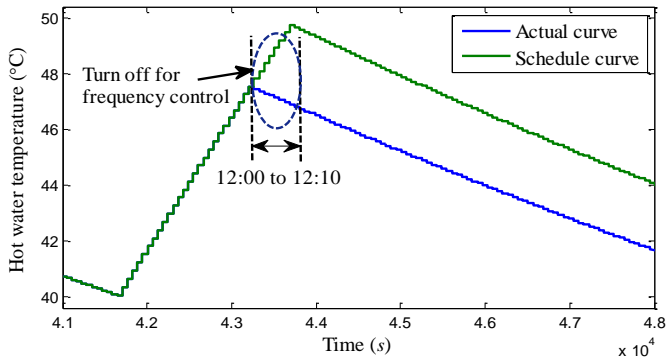
(b) Indoor temperatures of AC #2 to AC #5



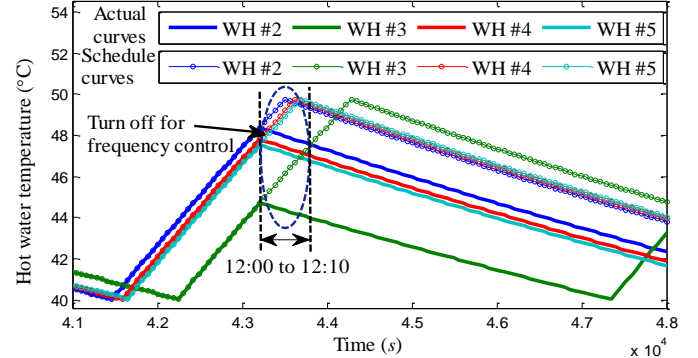
(c) Refrigerator temperature of Re #1



(d) Refrigerator temperatures of Re #2 to Re #5



(e) Hot water temperature of WH #1



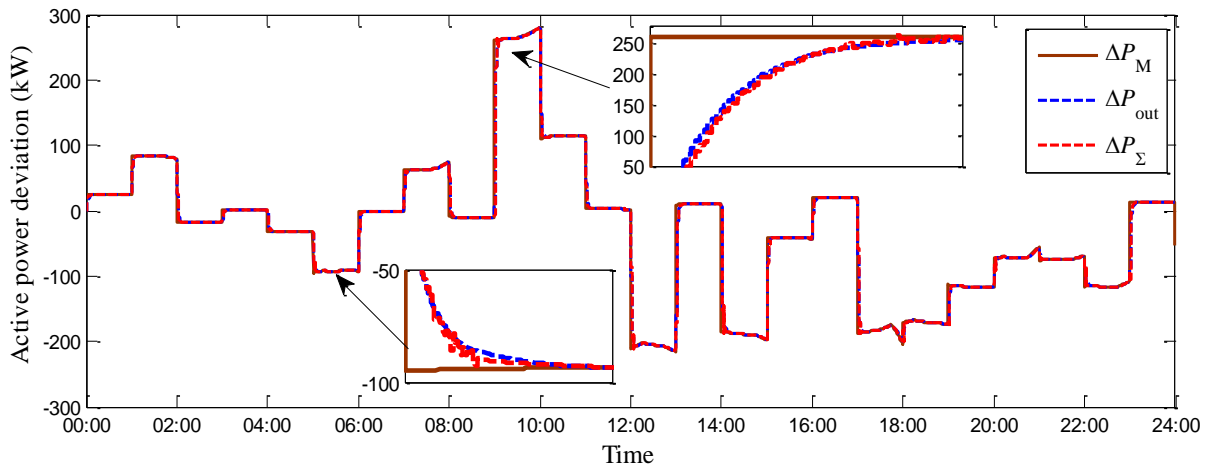
(f) Hot water temperatures of WH #2 to WH #5

Fig. 8. Temperature variation curves of some TCLs in online optimization of EL for a step power disturbance ($\Delta P_M=225$ kW) from 12:00 p.m. to 12:10 p.m.

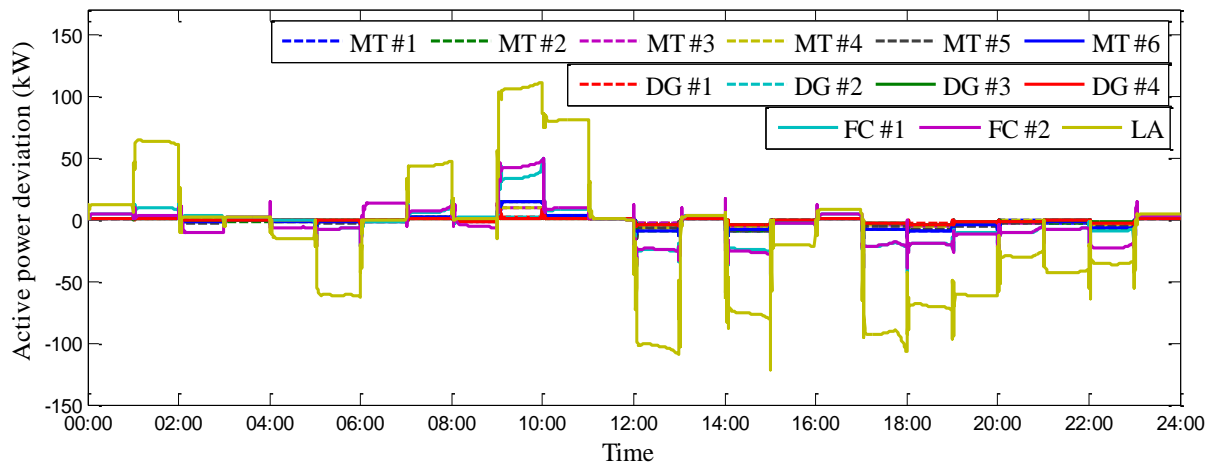
5.3.3 Study on a stochastic power disturbance

For thoroughly evaluating the performance of EL, a stochastic power disturbance is added in the testing islanded microgrid over a 24-hour period, where the online optimization results obtained by EL are given in Fig. 9(a) to (d). It can be obviously observed that the total active power output deviation obtained by EL can still match well with the stochastic power disturbance, which also leads to a smaller dynamic frequency deviation and a higher cumulative total operation cost. Moreover, LA is assigned to the largest generation commands, which indicates that LA can significantly improve the dynamic response performance and reduce the total operation cost due to its fast response speed and low regulation cost coefficient.

Compared with other algorithms, EL obtains the smallest frequency deviation and a low cumulative total operation cost over a day, as illustrated in Fig. 10. This also demonstrates that EL can greatly guarantee the optimum quality via an efficient search combination between various sub-optimizers and a learning contractor. Note that the cumulative total operation cost obtained by PROP is the lowest as the units with larger reserve capabilities just have the lower cost coefficients.



(a) Deviation of total active power output



(b) Active power output deviation of different units

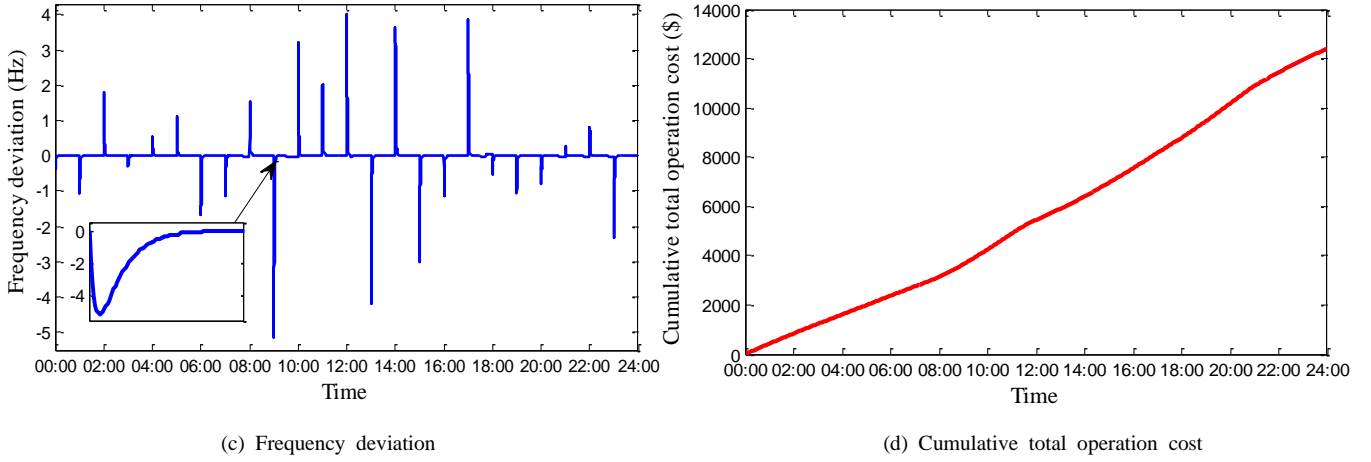


Fig. 9. Online optimization of EL for a stochastic power disturbance over 24-hour period.

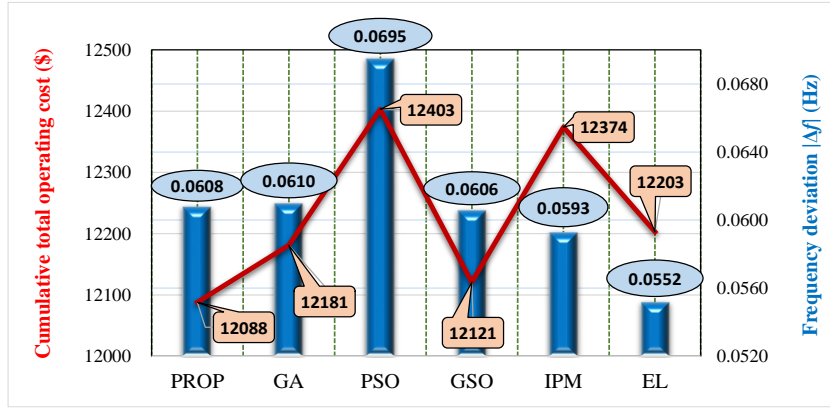


Fig. 10. Comparative results of different algorithms for a stochastic power disturbance over 24-hour period.

6. Conclusion

In this paper, a novel EL is presented for OAPC of DERs and TCLs in an islanded microgrid, which has the main contributions as

- 1) A mathematical model of OAPC is firstly presented for an effective coordination between secondary control and tertiary control of DERs and TCLs in an islanded microgrid, thus a high dynamic response performance and a low operation cost can be achieved.
- 2) Large number of TCLs with ultra-fast ramp rates can be effectively utilized for secondary control of LFC by aggregating them as a LA, while the consumer comfort can be completely satisfied via a real-time evaluation of their maximal participation time.
- 3) Through an efficient integration between various sub-optimizers and a learning concentrator, a high-quality optimum obtained by EL can be guaranteed with a wide exploration and a deep exploitation.
- 4) The optimal knowledge matrices of the source tasks can be reused for different real-time new tasks via knowledge transfer, such that EL is adequate to satisfy the online optimization of OAPC with a quite short implementation period.

Acknowledgment

This work was jointly supported by National Natural Science Foundation of China (51477055, 51777078), and Hong Kong RGC Theme based Research Scheme Grants No.T23-407/13N and T23-701/14N.

References

- [1] Hossain E, Kabalci E, Bayindir R, Perez Ronald. Microgrid testbeds around the world: State of art. *Energy Convers Manage* 2014;86:132-53.
- [2] Koohi-Kamali S, Rahim NA. Coordinated control of smart microgrid during and after islanding operation to prevent under frequency load shedding using energy storage system. *Energy Convers Manage* 2016;127:623-46.
- [3] Rezaei Navid, Kalantar M. Smart microgrid hierarchical frequency control ancillary service provision based on virtual inertia concept: An integrated demand response and droop controlled distributed generation framework. *Energy Convers Manage* 2015;92:287-301.
- [4] Rezaei Navid, Kalantar M. Economic-environmental hierarchical frequency management of a droop-controlled islanded microgrid. *Energy Convers Manage* 2014;88:498-515.
- [5] Mohamed YAI, El-Saadany EF. Adaptive decentralized droop controller to preserve power sharing stability of paralleled inverters in distributed generation microgrids. *IEEE Trans Power Electron* 2008;23(6):2806-16.
- [6] Zhong QC. Robust droop controller for accurate proportional load sharing among inverters operated in parallel. *IEEE Trans Ind Electron* 2013;60(4):1281-90.
- [7] Bevrani H, Habibi F, P Babahajyani, Watanabe M, Mitani Y. Intelligent frequency control in an AC microgrid: Online PSO-based fuzzy tuning approach. *IEEE Trans Smart Grid* 2012;3(4):1935-44.
- [8] Liu S, Wang X, Liu PX. Impact of communication delays on secondary frequency control in an islanded microgrid. *IEEE Trans Ind Electron* 2015;62(4):2021-31.
- [9] J. Guerrero, J. C. Vasquez, and L. G. D. Vicuna. Hierarchical control of droop-controlled AC and DC microgrids-A general approach toward standardization. *IEEE Trans. Ind. Elect.* 2011; 58(1):158-66.
- [10] Lakshmanan V, Marinelli M, Hu J, Binder HW. Provision of secondary frequency control via demand response activation on thermostatically controlled loads: Solutions and experiences from Denmark. *Appl Energy* 2016;173:470-80.
- [11] Tang X, X Hu, Li N, Deng W, Zhang G. A novel frequency and voltage control method for islanded microgrid based on multienergy storages. *IEEE Trans Smart Grid* 2016;7(1):410-19.
- [12] Pahasa J, Ngamroo I. Coordinated control of wind turbine blade pitch angle and PHEVs using MPCs for load frequency control of microgrid. *IEEE Syst J* 2016;10(1):97-105.
- [13] Sekha PC, Mishra S. Storage free smart energy management for frequency control in a diesel-PV-fuel-cell based hybrid AC microgrid. *IEEE Trans Neural Netw Learn Syst* 2016;27(8):1657-71.
- [14] Kim YS, Kim ES, Moon SI. Distributed generation control method for active power sharing and self-frequency recovery in an islanded microgrid. *IEEE Trans Power Syst* 2017;32(1):544-51.

- [15] Liu Y, Z Qu, Xin H, Gan D. Distributed real-time optimal power flow control in smart grid. *IEEE Trans Power Syst* 2017;32(5):3403-14.
- [16] Zhang XS, Li Q, Yu T, Yang B. Consensus transfer Q-learning for decentralized generation command dispatch based on virtual generation tribe. *IEEE Trans Smart Grid* 2016; DOI: 10.1109/TSG. 2016.2607801.
- [17] Qiu X, Xu JX, Xu Y, Tan KC. A new differential evolution algorithm for minimax optimization in robust design. *IEEE Trans Cybern* 2017; DOI: 10.1109/TCYB.2017.2692963.
- [18] Hu J, Cao J, Guerrero JM, Yong T, Yu J. Improving frequency stability based on distributed control of multiple load aggregators. *IEEE Trans Smart Grid* 2017;8(4):1553-67.
- [19] Xu Z, Østergaard J, Tøgeby M. Demand as frequency controlled reserve. *IEEE Trans Power Syst* 2011;26(3):1062-71.
- [20] Wang J, Li Y, Zhou Y. Interval number optimization for household load scheduling with uncertainty. *Energy Build* 2016; 130:613-24.
- [21] Chehreghani Bozchalui M. Optimal operation of energy hubs in the context of smart grids. Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Waterloo, Waterloo, ON, Canada, 2011.
- [22] Shao S, Pipattanasomporn M, Rahman S. Development of physical-based demand response-enabled residential load models. *IEEE Trans Power Syst* 2013;28(2):607-14.
- [23] Zhang X, Bao T, Yu T, Yang B, Han C. Deep transfer Q-learning with virtual leader-follower for supply-demand Stackelberg game of smart grid. *Energy* 2017;133:348-65.
- [24] Bianchi RAC, Celiberto LA, Santos PE, Matsuura JP, Lopez de Mantaras R. Transferring knowledge as heuristics in reinforcement learning: A case-based approach. *Artif Intell* 2015;226:102-21.
- [25] Pan J, Wang X, Cheng Y, Cao G. Multi-source transfer ELM-based Q learning. *Neurocomputing* 2014;137:57-64.
- [26] Dietterich TG. Ensemble learning. in *The Handbook of Brain Theory and Neural Networks*, 2nd ed., M. A. Arbib, Ed. Cambridge, MA: MIT Press, 2002.
- [27] Mohamed FA, Koivo HN. Online management genetic algorithm of microgrid for residential application. *Energy Convers Manage* 2012;64:562-8.
- [28] Zou D, Li S, Li Z, Kong X. A new global particle swarm optimization for the economic emission dispatch with or without transmission losses. *Energy Convers Manage* 2017;139:45-70.
- [29] Yang B, Zhang X, Yu T, Shu H, Fang Z. Grouped grey wolf optimizer for maximum power point tracking of doubly-fed induction generator based wind turbine. *Energy Convers Manage* 2017;133:427-43.
- [30] Zhang X, Yu T, Yang B, Cheng L. Accelerating bio-inspired optimizer with transfer reinforcement learning for reactive power optimization. *Know Based Syst* 2017;116:26-38.
- [31] Mirjalili S, Lewis A. S-shaped versus V-shaped transfer functions for binary particle swarm optimization. *Swarm Evol Comput* 2013;9:1-14.
- [32] Yu X, Zhou Q. Practical implementation of the SCADA + AGC/EDC system of the Hunan power pool in the central China power network. *IEEE Trans Energy Conver* 1994;9(2):250-55.
- [33] Hagh MT, Teimourzadeh S, Alipour M, Aliasghary P. Improved group search optimization method for solving CHPED in large scale power systems. *Energy Convers Manage* 2014;80:446-56.