IET Generation, Transmission & Distribution

Research Article

IET Journals
The Institution of Engineering and Technology

Artificial emotional reinforcement learning for automatic generation control of large-scale interconnected power grids

ISSN 1751-8687
Received on 1st November 2016
Revised 24th February 2017
Accepted on 17th March 2017
E-First on 12th June 2017
doi: 10.1049/iet-gtd.2016.1734
www.ietdl.org

Linfei Yin[1], Tao Yu[1] ✉, Lv Zhou[2], Linni Huang[1], Xiaoshun Zhang[1], Baomin Zheng[1]

[1]*School of Electric Power, South China University of Technology, Guangzhou Guangdong, People's Republic of China*
[2]*Department of Electrical and Computer Engineering, University of Auckland, Auckland 1010, New Zealand*
✉ *E-mail: taoyu1@scut.edu.cn*

Abstract: This study proposes an artificial emotional reinforcement learning (ERL) controller for automatic generation control (AGC) of large-scale interconnected power grids. In the scheme of ERL, the agent consists of two parts, a mechanical logical part and a humanistic emotional part, which essentially develop the control strategies of the agent. These two parts in proposed controller are introduced by reinforcement learning (RL) and artificial emotion (AE), respectively. The ERL controller can generate different control strategies depending on the operating scenarios, by highly integrating AE functions, which are quadratic function, exponential function, and linear function, respectively, with the elements of RL, such as action, learning rate, and reward function. The effectiveness of ERL controller with nine control strategies has been demonstrated considering AGC on a two-area load frequency control power system and China Southern Power Grid (CSG) power system. Results of simulation show the superior performance of ERL over that of proportional–integral control and four RL techniques.

1 Introduction

Frequency and power control has been recognised as one of the most important issues in large-scale interconnected power grids. Automatic generation control (AGC) is conventionally employed to maintain the nominal system frequency and the scheduled tie-line power interchanges by regulating the total generation output of AGC units. Therefore, the implementation of control algorithm for obtaining the total generation output can critically determine the overall control performance of AGC.

Considerable research has been dedicated to addressing the control methods of AGC in the past decades. In [1–3], proportional–integral (PI) controller has been widely implemented into AGC due to its simple structure and practicality. To obtain the optimal gains of PI controller, various meta-heuristic optimisation techniques have been successfully investigated for AGC of interconnected power grids, such as genetic algorithm [4, 5], particle swarm optimisation [6], imperialist competitive algorithm [7], firefly algorithm [8], bacterial foraging optimisation algorithm [9], and disrupted oppositional-based gravitational search algorithm [10]. However, these PI controllers with fixed optimal gains could not always achieve a satisfactory performance of AGC with the constantly changing operation scenarios in power system with increasing penetration of renewable energy.

To handle this issue, various intelligent controllers that can excellently perform when the system is operating over the non-linear operating range [11], including fuzzy logic controller [12], artificial neural network controller [13], and reinforcement learning (RL) controllers [14–18], have been developed to replace the classic controllers. Among these approaches, RL techniques, such as Q(λ) learning [14], R(λ) learning [15], state-action-reward-state-action (SARSA) [16], correlated-Q(λ) learning [17], and win or learn fast policy hill-climbing [18], have been demonstrated that they could effectively enhance the robustness and dynamic control performance of AGC under various complex operation scenarios due to their strong online-learning abilities. Nevertheless, these RL techniques always inevitably face three challenges in the application of AGC: (i) low control precision or long computation time resulted from the discrete actions; (ii) low learning efficiency with a fixed reward function and a fixed learning rate; (iii) low intelligence of RL agent due to its only one single mechanical logical part for learning. Hence, this paper aims to explore a novel RL technique to address these three problems.

A promising approach is to introduce emotional learning to RL, which could deal with continuous variables and generate the control policies by not only a logical but also a humanistic intelligent part. In recent years, emotional learning-based intelligent (ELI) controller has been presented to handle the above issue, which showed the superior control performance in various real-world applications, including non-linear control of an interconnected power system [19], doubly-fed induction generator [20], power-flow control [21], sensorless speed control of switched reluctance motor [22], dynamic voltage regulator [23], interline power-flow controller [24], asymmetrical six-phase induction motor [25], real-time position control of a servo-hydraulic rotary actuator [26], unmanned ground vehicle navigation [27] and so on. Moreover, emotional learning was developed with combining Q-learning for traffic flow forecasting of multi-agent systems [28, 29].

In this study, a novel artificial emotional RL (ERL), which combines RL with emotional learning, is proposed for AGC of large-scale interconnected power grids. The main contributions can be summarised as follows:

- ERL can generate the continuous control strategy for AGC through a humanistic emotional part from the artificial emotion (AE), thus 'dimension disaster' of traditional RL resulted from the discrete actions can be solved, while a more precise control strategy (i.e. output of AGC controller) can be acquired with a short computation time.
- A dynamic reward function and a dynamic learning rate can make the agent update the $\boldsymbol{Q}$-value matrix more efficiently, which can result in a faster convergence and a high-quality control strategy, thus it can achieve a superior AGC performance while satisfying the requirement of AGC period.
- Three AE functions including quadratic function, exponential function, and linear function are developed to be highly integrated with the action, learning rate, and reward function, thus ERL can generate various control strategies with a more intelligent RL agent, and the AGC controller will be more adaptable to a wide range of operation scenarios of large-scale interconnected power grids.

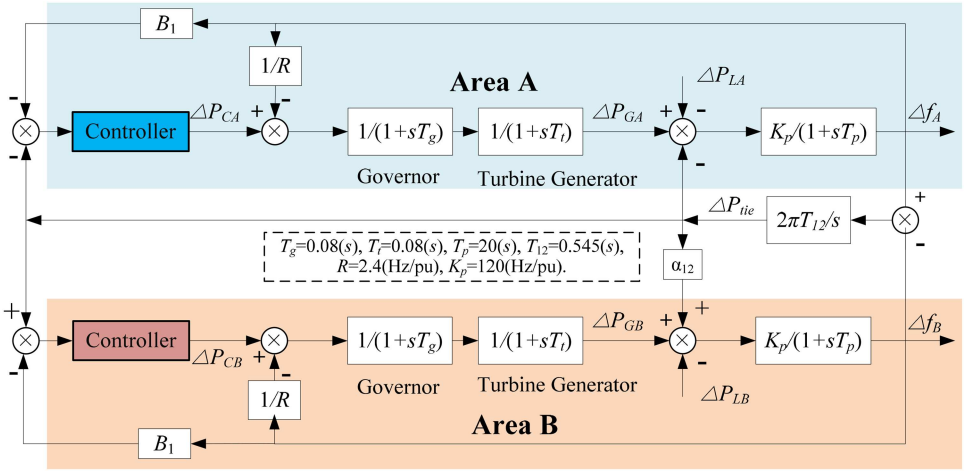

Fig. 1 *Two-area power system LFC model*

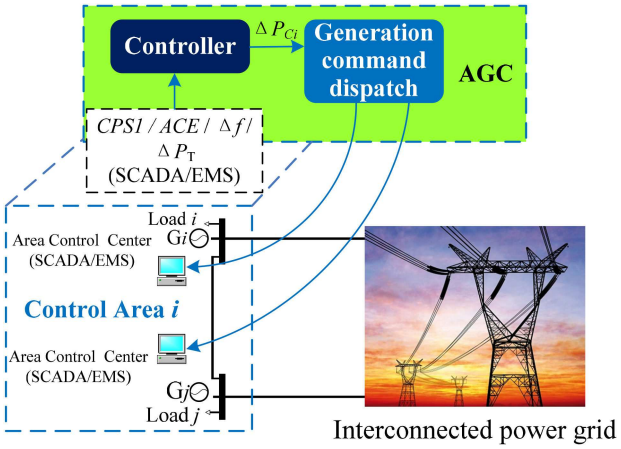

Fig. 2 *Control framework of AGC*

The paper is organised as follows. The control model of AGC is given in Section 2. Section 3 describes basic principle of ERL and the detailed design procedures for AGC. Simulation results of the proposed controller in a two-area load frequency control (LFC) power system and China Southern Power Grid (CSG) power system are presented in Section 4. Finally, Section 5 briefly concludes this paper.

2 Control model of AGC

2.1 Control framework

As shown in Fig. 1, a basic two-area power system with lossless tie line [30] is adopted to investigate the process of AGC, where each area consists of generator, governor, non-reheat steam turbine, and load. A control area with AGC aims to achieve three major objectives: (i) maintain the system frequency at its nominal value; (ii) maintain the tie-line power interchanges at its scheduled value; and (iii) maintain each unit's generation at its economic value. Here, the concept of area control error (ACE) is introduced as the adjustment generation power for promoting both the frequency deviation and the scheduled tie-line power deviation to be zero. Therefore, AGC can calculate power mismatch between the generation side and load side with various real-time ACE through a specific controller.

AGC mainly contains two procedures. One is to calculate the total generation command ΔP_{Ci} so as to balance the power disturbance by a controller; the other is to generate command dispatch that distributes the total generation command among the AGC units, as shown in Fig. 2. Note that, this paper focuses on the design procedure of controller in which the inputs of controller includes control performance standard (CPS), ACE, frequency deviation Δf, scheduled tie-line power deviation ΔP_{T} and so on.

2.2 Control objective

In each control area, the controller of AGC aims to minimise the absolute value of ACE, which is calculated as follows:

$$\mathrm{ACE} = \Delta P_t - 10B\Delta f \tag{1}$$

where B means frequency response coefficient of each area, which is generally measured in MW/0.1 Hz, thus the right-most part in (1) needs to be multiplied by 10 since the frequency deviation Δf is often measured in Hz.

On the other hand, indices of CPS including CPS1 and CPS2, which are established by North American Electric Reliability Council (NERC) [14], need to be considered in the design of AGC controller as they can enhance the frequency supporting efforts from control areas and relax the pressure of regulating ACE to zero [31, 32], whereas it can provide opportunity to realise significant savings in fuel cost, and unit wear and tear through allowing less unit manoeuvring. In general, CPS1 evaluates the impact of ACE deviation on the frequency of system, whereas CPS2 is used to restrict the ACE magnitude. They can be calculated as follows:

- CPS1 requires the ACE and frequency deviation in ith control area to satisfy the following constraints during an assessment period:

$$\frac{\sum(E_{\mathrm{AVE-min}} \cdot \Delta F_{\mathrm{AVE-min}})}{(-10B_i) \cdot n} \le \varepsilon_1^2 \tag{2}$$

where $E_{\mathrm{AVE\text{-}min}}$ is the clock-1-min average of ACE, and $\Delta F_{\mathrm{AVE\text{-}min}}$ is the clock-1-min average of frequency deviation; ε_1 is the targeted frequency bound for CPS1; n is the number of minutes in the assessment period; and B_i represents the frequency bias of the ith control area, expressed in MW/0.1 Hz.

To calculate the CPS1 compliance, a 1-min average compliance factor (CF1) can be identified and calculated to quantitatively represent a control area's contribution to the reliability of the ith interconnected power grid, as

$$\mathrm{CF1} = \frac{E_{\mathrm{AVE-min}} \cdot \Delta F_{\mathrm{AVE-min}}}{(-10B_i) \cdot \varepsilon_1^2} \tag{3}$$

Then, the values of 1-min CPS1 or 10-min CPS1 metrics can be obtained as below:

$$\mathrm{CPS1} = \left(2 - \mathrm{AVG}_{\mathrm{period}}\{\mathrm{CF1}\}\right) \times 100\% \tag{4}$$

- CPS2 is used to limit ACE magnitude, which requires the 10-min averages of ACE in a control area to be less than a threshold (L_{10}) as given in the following:

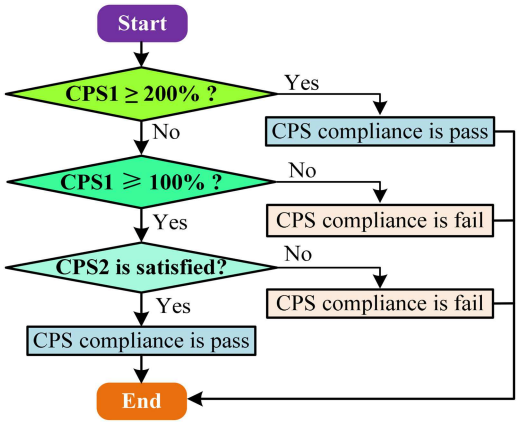

Fig. 3 *Flowchart of determining CPS compliance*

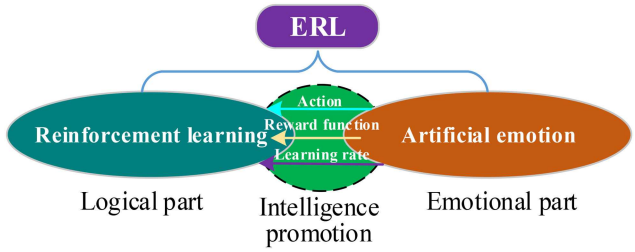

Fig. 4 *Framework of ERL*

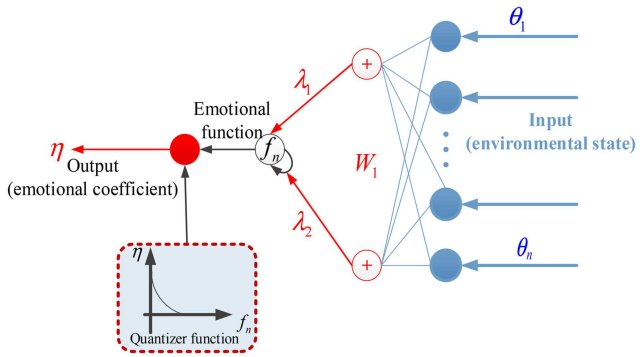

Fig. 5 *AE quantiser*

$$|E_{\text{AVE}-10-\text{min}}| \le L_{10} \tag{5}$$

$$L_{10} = 1.65 \cdot \varepsilon_{10} \cdot \sqrt{(-10B_i) \cdot (-10B_s)} \tag{6}$$

where $E_{\text{AVE-10-min}}$ is the clock-10-min average of ACE; B_s is the summation of the frequency bias settings of all the control areas considered in the interconnection; and ε_{10} is the targeted frequency bound for CPS2.

- CPS compliance: In China Power Industry, CPS1 and CPS2 compliances are assessed in association with NERC's rules on the basis of daily, monthly, and yearly data. In addition, CPS compliance is formulated based on the CPS1/CPS2 metrics to evaluate the overall AGC performance under CPS standards. According to the grid code of CSG, the logical flowchart of determining CPS compliance over a specific assessment period (typically 10 min) is illustrated in Fig. 3.

  i. If CPS1 ≥ 200%, then there is no need to consider CPS2, and CPS compliance rating passes.
  ii. If 100% ≤ CPS1 < 200% and CPS2 compliance is satisfied, then the CPS compliance rating passes.
  iii. If 100% ≤ CPS1 < 200% and CPS2 standard is in violation, then the CPS compliance rating fails.
  iv. If CPS1 ≤ 100%, then the CPS compliance rating fails.

Therefore, CPS compliance percentage on a daily, monthly or yearly basis can be calculated as follows:

$$\text{CPS}(\%) = \left[1 - \frac{\text{violation periods}}{\text{total periods} - \text{unavailable periods}}\right] \times 100\% \tag{7}$$

3 ERL and application design for AGC

3.1 Basic principle of ERL

ERL is composed of a mechanical logical part and a humanistic emotional part, which can be achieved by RL and AE, respectively, as illustrated in Fig. 4. Moreover, the action, reward function and learning rate of RL can be equipped with AE, thereby the agent of ERL will be more intelligent than that of RL.

3.1.1 Reinforcement learning: As one of the most popular RL, Q-learning is a model free algorithm which provides interactions between the agent and the environment, and it mainly consists of five elements: Q-value matrix $\boldsymbol{Q}$, probability distribution matrix $\boldsymbol{P}$, reward function R, action set $\boldsymbol{A}$, and state set $\boldsymbol{S}$. Generally, a RL agent implements an action a at the current state s in an environment based on the Q-value matrix $\boldsymbol{Q}$ and probability distribution matrix $\boldsymbol{P}$, and these two matrices can be subsequently updated with the reward function, as follows:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(R(s,s',a) + \gamma \max_{a \in A} Q(s',a) - Q(s,a)) \tag{8}$$

$$P(s,a) \leftarrow \begin{cases} P(s,a) - \beta(1 - P(s,a)) & \text{if } a' = a \\ P(s,a)(1-\beta) & \text{if } a' \neq a \end{cases} \tag{9}$$

where α is the learning rate; γ is the discount factor; β is the search factor; s, s' denote the current state and the next state, respectively; $R(s,s',a)$ is the reward function obtained from the current state s to the next state s' with a selected action a.

3.1.2 Artificial emotion: AE quantiser: Different from the agent of RL techniques, the agent of ERL can create a sense of emotion according to different states obtained from the environment. To construct this process, a quantiser is presented to transform the state $\boldsymbol{\theta}$ (input) to the emotional coefficient η (output), as shown in Fig. 5. The mathematical model of the quantiser is described as below:

$$f_n = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \theta_i \omega_i \tag{10}$$

$$\eta \leftarrow \begin{cases} k_\eta & \text{if } 1/f_n \ge \eta_{\max} \\ k_\eta / f_n & \text{if } 1/f_n < \eta_{\max} \end{cases} \tag{11}$$

where θ_i represents the input information; ω_i denotes the emotional weight; λ_i is the converted value of the ith input information; f_n is the emotion energy; and k_η is the upper limit of emotional coefficient.

Intelligence promotion for RL: In order to enhance the intelligence of RL, the emotion coefficient η has to be converted to the actual effect on RL. Hence, three converting functions, quadratic converting function, exponential converting function, and linear converting function, are introduced to realising this process, which can be described as follows:

$$C_{\mathrm{f}}(\eta) = \begin{cases} k_a\eta^2 + k_b\eta + k_c & \text{for quadatic converting function} \\ \mathrm{e}^{\eta} + k_d & \text{for exponential converting function} \\ k_e\eta + k_f & \text{for linear converting function} \end{cases} \quad (12)$$

where k_a, k_b, k_c, k_d, k_e, and k_f are the converting coefficients.

Based on the converting function (12), the action, reward function, and learning rate of RL can be modified according to different emotional coefficients, which can be updated as follows:

$$a_\eta \leftarrow aC_f(\eta) \quad (13)$$

$$R_\eta \leftarrow RC_f(\eta) \quad (14)$$

$$\alpha_\eta \leftarrow \alpha C_f(\eta) \quad (15)$$

where a_η, R_η, and α_η are the action, reward function, and learning rate modified with AE, respectively.

Table 1 Advantage comparison of nine control strategies of ERL

Type of converting function	Control strategies	Modified part of RL	Feature	Advantage
quadratic	ERL-I	action *a*	continuous action	more precise control strategy
	ERL-II	learning rate *α*	dynamic learning rate	faster convergence
	ERL-III	reward function *R*	dynamic reward function	higher learning efficiency
exponential	ERL-IV	action *a*	continuous action	more precise control strategy
	ERL-V	learning rate *α*	dynamic learning rate	faster convergence
	ERL-VI	reward function *R*	dynamic reward function	higher learning efficiency
linear	ERL-VII	action *a*	continuous action	more precise control strategy
	ERL-VIII	learning rate *α*	dynamic learning rate	faster convergence
	ERL-IX	reward function *R*	dynamic reward function	higher learning efficiency

3.1.3 Discussion of nine control strategies: Note that the performance of RL is mainly depended on the design of action, learning rate, and reward function [14–18], so that ERL employs three converting function (12) to modify these three parts, respectively, as illustrated in Table 1. Hence, ERL contains nine control strategies (i.e. ERL-I, ERL-II, …, and ERL-IX) for realising a better control performance, which will be fully compared through extensive simulations on a specific control problem. In theory, the control strategies with modifying action, i.e. ERL-I, ERL-IV, and ERL-VII can achieve a more precise control strategy as they can generate a continuous action by (13) between its upper and lower bounds. Besides, ERL-II, ERL-V, and ERL-VIII with a dynamic learning rate can accelerate the convergence, thus the proposed algorithm can rapidly calculate an optimal control action for AGC. Lastly, ERL-III, ERL-VI, and ERL-IX with a dynamic reward function can result in a higher learning efficiency, i.e. the Q-value matrix can be updated more efficiently in (8).

3.2 Control design for AGC

3.2.1 Design of state and action set: As illustrated in Fig. 1, the controller of AGC needs to collect the frequency deviation Δf and scheduled tie-line power deviation ΔP_{T} from the interconnected power grid, and then calculate the indices of ACE, CPS1, and CPS2 by (3)–(7). During the design process, the interconnected power grid is regarded as the environment of ERL, and the state set is designed based on the range of ACE, as given in Table 2. The action set of ERL is determined by the controller output of AGC, and is carefully designed by trial-and-error method. In this paper, the action set is selected based on different states through various simulations, which is also provided in Table 2.

3.2.2 Design of reward function: The design of reward function can significantly influence the performance of ERL. Since the controller of AGC aims to regulate ACE to zero, it is designed in (8) in this paper

$$R(k) = \begin{cases} 10 & \text{if } \left|\mathrm{ACE}^k\right| \le 1 \\ \mathrm{ACE}^k - \mathrm{ACE}^{k-1} & \text{if } \mathrm{ACE}^k > 1 \\ \mathrm{ACE}^{k-1} - \mathrm{ACE}^k & \text{if } \mathrm{ACE}^k < -1 \end{cases} \quad (16)$$

where ACE^k denotes the value of ACE at kth iteration.

3.2.3 Execution procedure: In summary, the execution procedure of ERL-based controller for AGC is given in Fig. 6.

4 Numerical simulation

Simulations of ERL for AGC are carried out on two-area LFC power system and the CSG power system in the Simulink with the MATLAB software platform compared with PI control, Q-learning

Table 2 State set of ERL for AGC

State No.	Range of ACE, MW	Action set, MW	State No.	Range of ACE, MW	Action set, MW
$s = 1$	(−∞, −50]	{50, 100, 150, 200, 250, 300, 350, 400, 450, 500}	$s = 8$	(1, 10]	{10, 7, 4, 1, −2, −5, −8, −11, −14, −17, −20}
$s = 2$	(−50, −40]	{30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60}	$s = 9$	(10, 20]	{0, −3, −6, −9, −12, −15, −18, −21, −24, −27, −30}
$s = 3$	(−40, −30]	{20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50}	$s = 10$	(20, 30]	{−10, −13, −16, −19, −22, −25, −28, −31, −34, −37, −40}
$s = 4$	(−30, −20]	{10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40}	$s = 11$	(30, 40]	{−20, −23, −26, −29, −32, −35, −38, −41, −44, −47, −50}
$s = 5$	(−20, −10]	{0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30}	$s = 12$	(40, 50]	{−30, −33, −36, −39, −42, −45, −48, −51, −54, −57, −60}
$s = 6$	(−10, −1]	{−10, −7, −4, −1, 2, 5, 8, 11, 14, 17, 20}	$s = 13$	(50, ∞)	{−50, −100, −150, −200, −250, −300, −350, −400, −450, −500}
$s = 7$	(−1, 1]	{−5, −4, −3, −2, −1, 0, 1, 2, 3, 4, 5}			

[30], R(λ) [15], SARSA [16], and SARSA(λ) [33], where the simulation time of AGC is set to be 4 s.

4.1 Case study on two-area LFC power system

In order to obtain the optimal parameters of ERL, a large number of simulations are carried out to analyse the control performance, where the indices $|\Delta f|$ and |ACE| are used for evaluation. As illustrated in Fig. 7, different parameters combinations of ω_1 and k_η, as well as k_a and k_b, can significantly determine the indices $|\Delta f|$ and |ACE| as it generates different AE quantisers for the emotional part of ERL. Hence, the parameters combination with the minimal |ACE| can be chosen for AGC. Based on the simulation results from trial-and-error with different parameters combinations, all the parameters of ERL are given in Table 3.

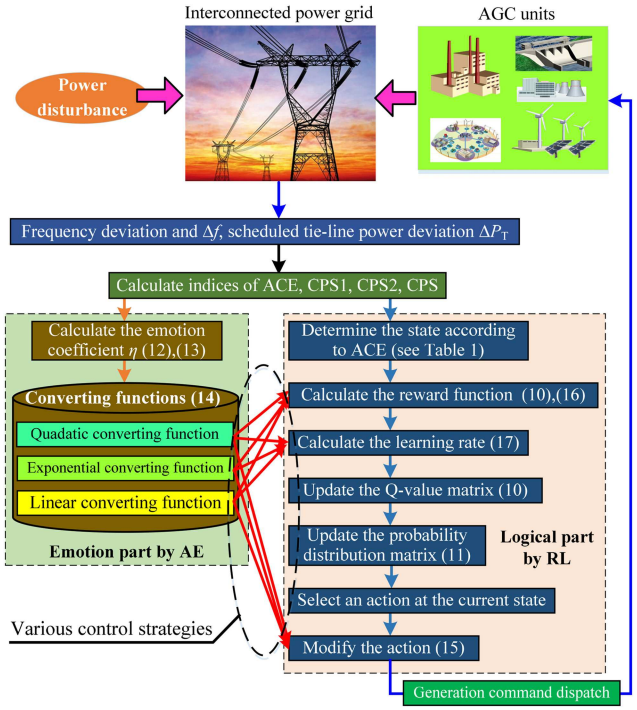

Fig. 6 *Execution procedure of ERL for controller of AGC*

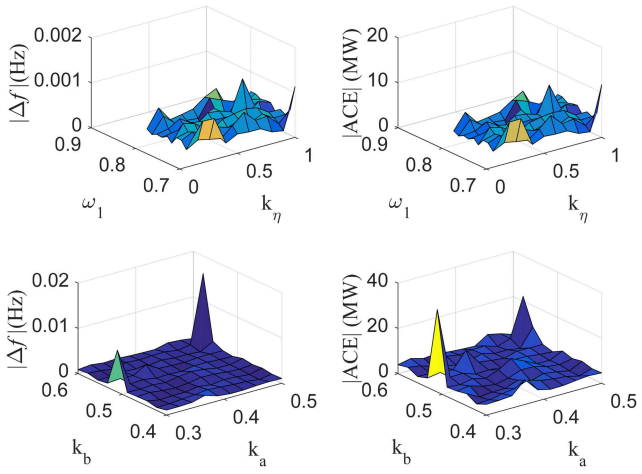

Fig. 7 *Simulation results obtained by ERL with different parameters*

The AGC of two-area LFC power system [14–18] modelled in Simulink is described in Fig. 1, where each control technique will be implemented in the controller for performance comparisons. A typical sinusoidal load disturbance and a stochastic load disturbance are applied in area A, respectively, as shown in Fig. 8. Furthermore, the period and the amplitude of the first load disturbance are 1200 s and 1000 MW, respectively, while the amplitude of the second load disturbance is no >000 MW.

By imposing the first load disturbance, the control performance of PI control, four RL techniques, and a group of ERL with nine strategies is shown in Fig. 9. It can be seen that all the control algorithms can adequately balance the power disturbance (see Fig. 9*a*), and the frequency deviation Δf and ACE (see Figs. 9*b* and *c*) can be restricted in a small range so as to improve the index of CPS1 (see Fig. 9*d*). Moreover, the conventional PI control can achieve a moderate performance with the proper fixed gains. Compared with that, Q-learning, R(λ), SARSA, and SARSA(λ) can continuously improve the AGC performance due to their online learning ability. By contrast, ERL with different strategies can rapidly obtain a control strategy with a higher CPS1, a smaller $|\Delta f|$ and |ACE| through accelerating the learning efficiency and enhance the control precision with an intelligent emotional part.

The statistic results with two different load disturbances on various AGC performance metrics are listed in Table 4, where all the indices are the average in the execution time. It is observed in Table 4 that the control performance of ERL-I is the best among that of other algorithms, which verifies that the quadratic converting function of AE can successfully improve the action of RL, which results from that the continuous action with emotion coefficient can closely approximate the optimal strategy of AGC controller. The classical RL techniques, including *Q*-learning, R(λ), SARSA, and SARSA(λ), can achieve a superior AGC performance (i.e. the higher CPS1, the smaller |ACE| and $|\Delta f|$) compared with that of PI control, which demonstrates the online learning ability of RL. By imposing the stochastic load disturbance, the AGC performance of ERL-II, ERL-V, and ERL-VIII degrades dramatically as the dynamic learning rate will lead to a suboptimal control strategy for a rapidly changing load disturbance when pursuing a faster convergence. However, other six ERL algorithms can also achieve a satisfactory AGC performance with a stochastic load disturbance, which indicates that ERL with a continuous action and a dynamic reward function can generate more optimal control strategies against to PI and RL with a single logical part, thus it can be competently implemented for AGC to balance different load disturbances.

4.2 Case study on CSG power system

The second simulation is based on CSG power system, which consists of four provincial control areas interconnected by parallel HVDC-HVAC (high-voltage direct current and high-voltage alternating current) transmission systems [34], i.e. Yunnan Power Grid, Guizhou Power Grid, Guangxi Power Grid, and Guangdong Power Grid. The interconnected framework of CSG is given in Fig. 10. As the most developed province in South China, Guangdong Power Grid is selected as AGC implementation with different algorithms. In this case study, three sinusoidal load disturbances (amplitude: 1000 MW; period: 20 min; noises: 0, 10,

Table 3 Parameters in ERL for AGC

Parameter	Range	Value	Parameter	Range	Value
α	$0 < \alpha < 1$	0.1	k_b	$0.1 < k_b < 1$	0.4
γ	$0 < \gamma < 1$	0.3	k_c	$0 < k_c < 0.2$	0.01
β	$0 < \beta < 1$	0.5	k_d	$0 < k_d < 0.2$	0.01
ω_i ($i = 1, 2\ldots$)	$0 < \omega_i < 1$	[0.6, 0.4]	k_e	$0 < k_e < 0.2$	0.01
k_η	$0.1 < k_\eta < 2$	1	k_f	$0 < k_f < 1$	0.4
k_a	$0.1 < k_a < 1$	0.5			

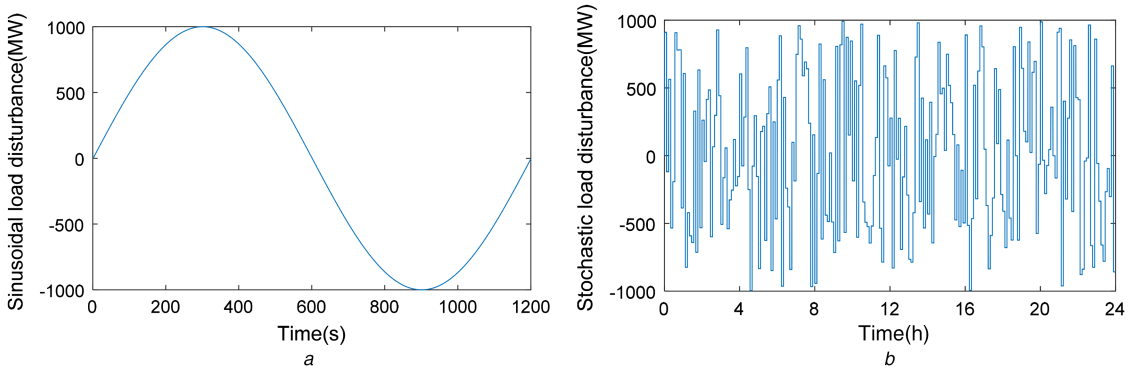

Fig. 8 *Two applied load disturbances in area A*
(a) Sinusoidal load disturbance, ***(b)*** Stochastic load disturbance

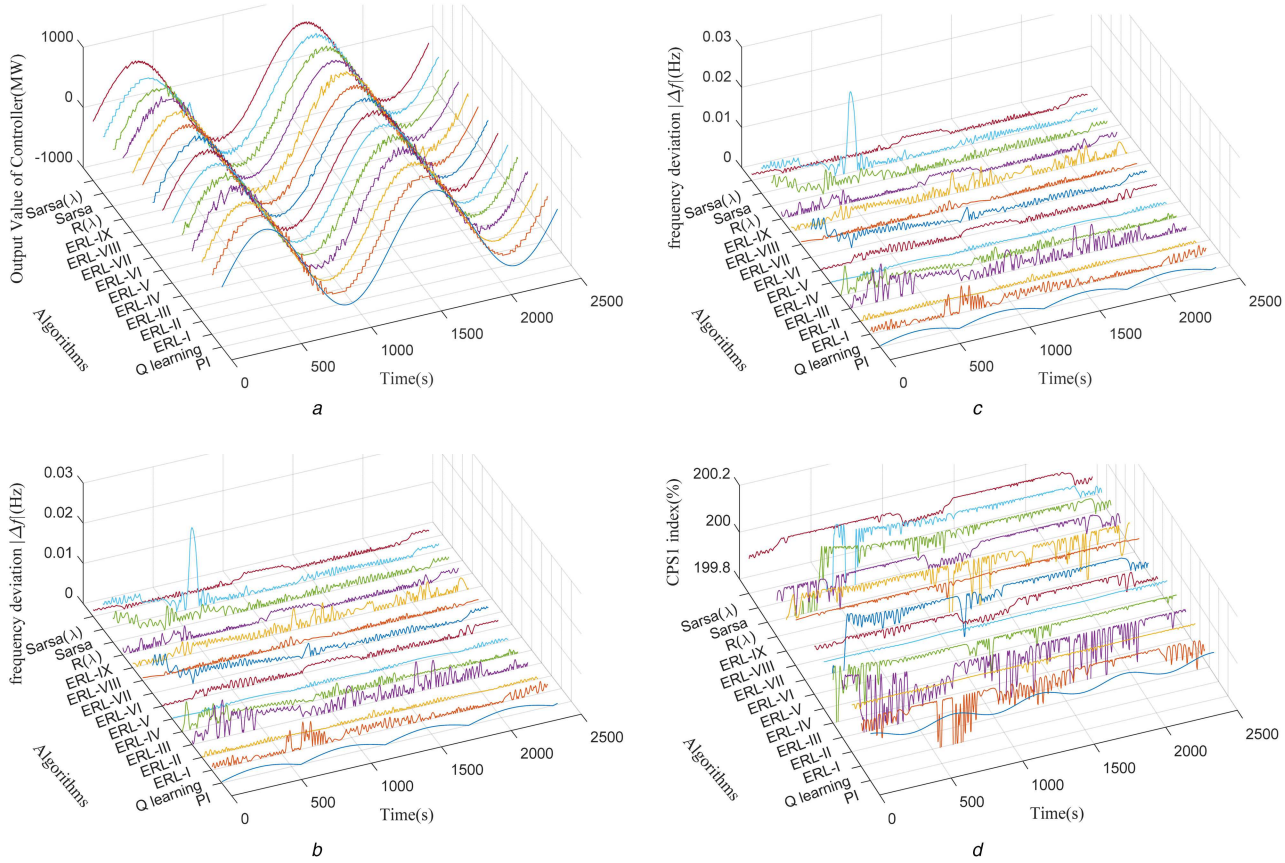

Fig. 9 *Control performance of two-area LFC power system with a sinusoidal load disturbance obtained by different algorithms*
(a) Output of controller, ***(b)*** 10-minute average of Δ*f*, ***(c)*** 10-minute average of ACE, ***(d)*** 10-minute average of CPS1

20% finite bandwidth white noise, respectively) and a stochastic load disturbance (amplitude: ≤1000 MW), are applied in Guangdong Power Grid.

Table 5 shows the statistic results of different algorithms in the model with different load disturbances. Similarly, ERL-I has the most superior control performance with highest CPS1, smallest |Δ*f*| and |ACE| among all the algorithms, which again confirms the merits of the combination between quadratic converting function of AE and continuous action of RL. Specifically, the differences of control performance among ERL and other algorithms are provided as follows:

- *Sinusoidal load disturbance with 0% white noise*: the index of |ACE| of ERL-I is 66.209 and 71.087% less than that of PI and that of R(λ), respectively, and index of |Δ*f*| of ERL-I is 37.567 and 62.686% less than that of PI and that of *Q*-learning, respectively, while all the ERL algorithms have smaller |ACE| than that of PI and that of four RL techniques, where six ERL algorithms have the smaller |Δ*f*| compared with other control methods.
- *Sinusoidal load disturbance with 10% white noise*: the index of |ACE| of ERL-I is 68.915 and 58.463% less than that of PI and that of Q-learning, respectively, and index of |Δ*f*| of ERL-I is 36.686 and 61.246% less than that of PI and that of SARSA, respectively, while all the ERL algorithms have smaller |ACE| than that of PI and that of four RL techniques, where five ERL algorithms have the smaller |Δ*f*| compared with other control methods.
- *Sinusoidal load disturbance with 20% white noise*: the index of |ACE| of ERL-I is 52.825 and 34.409% less than that of PI and that of SARSA, respectively, and index of |Δ*f*| of ERL-I is 57.568 and 33.246% less than that of PI and that of Q-learning, respectively, while all the ERL algorithms have smaller |ACE| than that of PI and that of four RL techniques, where seven ERL algorithms have the smaller |Δ*f*| compared with other control methods.
- *Stochastic load disturbance*: the index of |ACE| of ERL-I is 59.660 and 52.400% less than that of PI and that of Q-learning, respectively, and index of |Δ*f*| of ERL-I is 26.994 and 65.071% less than that of PI and that of SARSA(λ), respectively, while

Table 4 Statistic results of two-area LFC power system obtained by different algorithms

Load disturbance	Algorithms	CPS1, %	CPS2, %	|ACE|, MW	$\|\Delta f\|$, Hz	CPS, %
sinusoidal	PI	199.98	100.00	4.9113	0.001153	100.00
	Q-learning	199.98	100.00	5.0516	0.000848	100.00
	R(λ)	199.99	100.00	4.2089	0.000786	100.00
	SARSA	199.99	100.00	4.1067	0.000797	100.00
	SARSA(λ)	199.98	100.00	4.0047	0.000787	100.00
	ERL-I	**200.00**	**100.00**	**1.4139**	**0.000333**	**100.00**
	ERL-II	199.98	100.00	3.8440	0.000862	100.00
	ERL-III	199.99	100.00	2.6166	0.000581	100.00
	ERL-IV	200.00	100.00	2.0396	0.000460	100.00
	ERL-V	199.98	100.00	4.2360	0.000810	100.00
	ERL-VI	199.99	100.00	3.3012	0.000771	100.00
	ERL-VII	199.99	100.00	2.3176	0.000488	100.00
	ERL-VIII	199.97	100.00	4.9871	0.001123	100.00
	ERL-IX	199.98	100.00	3.7281	0.000859	100.00
stochastic	PI	199.98	100.00	5.3396	0.001018	100.00
	Q-learning	199.99	100.00	3.2729	0.000769	100.00
	R(λ)	199.99	100.00	2.5444	0.000603	100.00
	SARSA	199.99	100.00	1.7387	0.000736	100.00
	SARSA(λ)	200.00	100.00	2.1701	0.000568	100.00
	ERL-I	**200.00**	**100.00**	**0.6489**	**0.000182**	**100.00**
	ERL-II	200.00	100.00	18.1769	0.000542	100.00
	ERL-III	200.00	100.00	2.3581	0.000247	100.00
	ERL-IV	200.00	100.00	2.6016	0.000446	100.00
	ERL-V	200.01	100.00	14.2661	0.000469	100.00
	ERL-VI	200.00	100.00	2.0789	0.000504	100.00
	ERL-VII	200.00	100.00	0.8825	0.000274	100.00
	ERL-VIII	199.99	100.00	26.9057	0.000437	100.00
	ERL-IX	200.00	100.00	1.7235	0.000502	100.00

Bold indicates best values corresponding to that indices

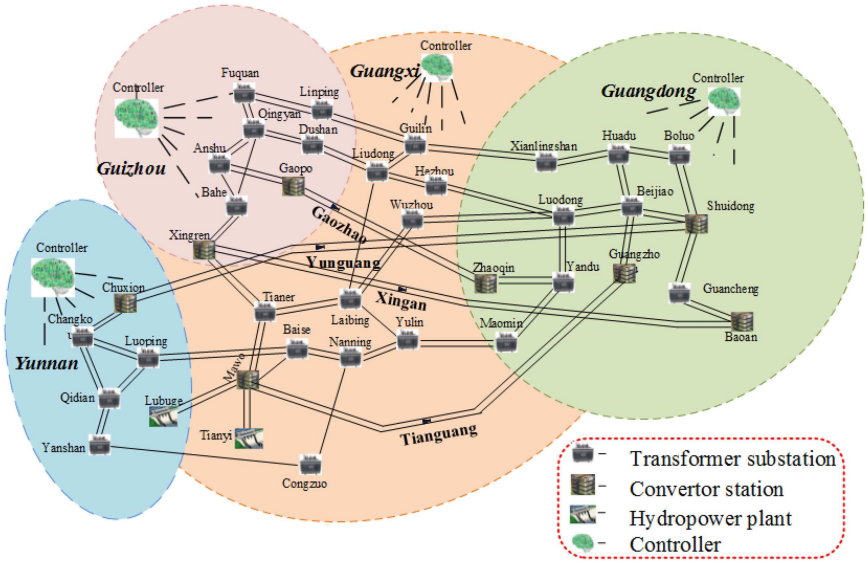

Fig. 10 *Interconnected framework of CSG power system*

six ERL algorithms have smaller |ACE| than that of PI and that of four RL techniques, where only two ERL algorithms have the smaller $|\Delta f|$ compared with other control methods.

In summary, most ERL algorithms with different nine control strategies can outperform PI and four conventional RL techniques in most cases, the performance of that dramatically degrade for the rapidly changing stochastic load disturbance. Both Tables 4 and 5 demonstrate that ERL-I with quadratic converting function of AE to improve the action of RL can approximate the optimal control strategy for controller of AGC, thus index of CPS1 can be improved and indices of $|\Delta f|$ and |ACE| can be minimised.

5 Conclusion

In this paper, a novel ERL with nine control strategies has been proposed for designing controllers of AGC implemented in both two-area LFC power system and CSG interconnected power system. The designed controllers exploit more intelligent feature of AE and integrate it into controller based on RL. The main contributions can be concluded as follows:

i. The intelligence of RL can be significantly improved by AE, thus the corresponding mechanical logical part and humanistic

emotional part can make the agent of ERL approximate a more optimal control strategy for AGC.

ii. ERL can generate different control strategies by combining different converting functions with action, reward function, and learning rate of RL, in which the continuous action can

Table 5 Statistic results of Guangdong power grid obtained by different algorithms

Load disturbance	Algorithms	CPS1, %	CPS2, %	|ACE|, MW	$\|\Delta f\|$, Hz	CPS, %
sinusoidal with 0% white noise	PI	191.41	81.05	200.7485	0.027612	81.05
	Q learning	177.79	94.03	154.0899	0.046200	94.03
	R(λ)	201.99	87.21	234.6179	0.028967	94.46
	SARSA	178.46	95.53	151.8444	0.045562	95.53
	SARSA(λ)	187.62	92.78	149.3353	0.033593	94.20
	ERL-I	196.97	**100.00**	**67.8340**	**0.017239**	**100.00**
	ERL-II	185.32	95.83	128.2360	0.039482	95.83
	ERL-III	194.36	98.57	92.2408	0.020530	98.57
	ERL-IV	195.29	**100.00**	71.4858	0.020814	**100.00**
	ERL-V	193.82	91.69	139.0966	0.022589	94.38
	ERL-VI	202.35	95.44	100.2425	0.020234	**100.00**
	ERL-VII	192.51	98.29	86.9814	0.028083	98.15
	ERL-VIII	189.49	92.94	140.0630	0.041667	94.38
	ERL-IX	202.39	97.18	110.1723	0.022810	99.07
sinusoidal with 10% white noise	PI	191.06	77.13	205.7547	0.028267	77.13
	Q learning	177.89	92.39	153.9791	0.046042	92.39
	R(λ)	178.30	93.39	152.5177	0.045756	93.39
	SARSA	177.92	93.70	153.6907	0.046181	93.70
	SARSA(λ)	178.05	93.58	153.677	0.045981	93.58
	ERL-I	195.24	**100.00**	**63.9583**	**0.017897**	**100.00**
	ERL-II	185.32	92.87	129.6828	0.038568	93.13
	ERL-III	193.78	**100.00**	96.4025	0.021744	**100.00**
	ERL-IV	196.13	**100.00**	70.3140	0.018779	**100.00**
	ERL-V	197.29	94.38	134.7554	0.030325	96.71
	ERL-VI	197.13	93.24	111.1909	0.018022	97.50
	ERL-VII	189.21	97.43	99.3898	0.028886	97.43
	ERL-VIII	214.04	92.76	142.7518	0.029943	97.50
	ERL-IX	194.55	95.95	118.9428	0.023104	98.08
sinusoidal with 20% white noise	PI	190.53	71.44	214.2669	0.029252	72.62
	Q learning	178.11	92.20	153.5688	0.046020	92.20
	R(λ)	178.55	92.59	152.3895	0.045426	92.59
	SARSA	178.05	92.84	154.1073	0.045980	92.84
	SARSA(λ)	178.75	92.99	151.2604	0.045114	92.99
	ERL-I	194.59	**100.00**	**101.0811**	**0.019527**	**100.00**
	ERL-II	183.66	91.83	134.6681	0.040472	92.13
	ERL-III	191.98	**100.00**	120.4560	0.024654	**100.00**
	ERL-IV	193.74	**100.00**	109.5012	0.022502	**100.00**
	ERL-V	200.55	94.70	139.2029	0.036099	98.10
	ERL-VI	196.24	96.95	122.3921	0.029136	**100.00**
	ERL-VII	194.74	**100.00**	113.3126	0.020369	**100.00**
	ERL-VIII	198.13	93.01	144.6103	0.028843	96.71
	ERL-IX	193.44	89.31	127.7830	0.023331	91.95
stochastic	PI	182.72	89.53	203.6770	0.030329	89.53
	Q learning	186.97	86.03	172.6104	0.030654	86.03
	R(λ)	178.49	92.77	147.7151	0.044895	92.77
	SARSA	181.73	91.52	126.4819	0.038019	91.52
	SARSA(λ)	185.55	90.02	156.6333	0.063391	93.77
	ERL-I	193.51	**100.00**	**82.1634**	**0.022142**	**100.00**
	ERL-II	190.24	97.76	146.0929	0.041530	100.00
	ERL-III	211.07	93.77	79.6555	0.026736	100.00
	ERL-IV	183.71	93.52	134.7329	0.039417	93.52
	ERL-V	188.35	100.00	111.1423	0.031929	100.00
	ERL-VI	226.40	87.03	161.2770	0.045175	100.00
	ERL-VII	188.85	100.00	103.4488	0.033030	100.00
	ERL-VIII	209.65	93.27	104.9748	0.036190	100.00
	ERL-IX	192.16	96.51	111.5574	0.049279	100.00

Bold indicates best values corresponding to that indices

tackle the 'curse of dimensionality' in traditional RL. Furthermore, a dynamic learning rate and a dynamic reward function can achieve a faster convergence.

iii. Compared with PI and four RL techniques, ERL can achieve more superior control performance with a higher CPS1, smaller $|\Delta f|$ and |ACE|. Therefore, it is very convincing to implement ERL technique in designing controllers for AGC in large-scale interconnected power grids.

6 Acknowledgments

This work was supported by National Key Basic Research Program of China (973 Program: 2013CB228205), National Natural Science Foundation of China (51477055).

7 References

[1] Khuntia, S.R., Panda, S.: 'Simulation study for automatic generation control of a multi-area power system by ANFIS approach', *Appl. Soft Comput.*, 2012, **12**, (1), pp. 333–341

[2] Mohanty, B., Panda, S., Hota, P.K.: 'Controller parameters tuning of differential evolution algorithm and its application to load frequency control of multi-source power system', *Int. J. Electr. Power Energy Syst.*, 2014, **54**, (1), pp. 77–85

[3] Parmar, K.P.S., Majhi, S., Kothari, D.P.: 'Load frequency control of a realistic power system with multi-source power generation', *Int. J. Electr. Power Energy Syst.*, 2012, **42**, (1), pp. 426–433

[4] Demiron, A., Zeynelgil, H.L.: 'GA application to optimization of AGC in three-area power system after deregulation', *Int. J. Electr. Power Energy Syst.*, 2007, **29**, (3), pp. 230–240

[5] Hota, P.K., Mohanty, B.: 'Automatic generation control of multi-source power generation under deregulated environment', *Int. J. Electr. Power Energy Syst.*, 2016, **75**, pp. 205–214

[6] Bhatt, P., Roy, R., Ghoshal, S.P.: 'Optimized multi area AGC simulation in restructured power systems', *Int. J. Electr. Power Energy Syst.*, 2010, **32**, (4), pp. 311–322

[7] Parmar, K.P.S., Majhi, S., Kothari, D.P.: 'LFC of an interconnected power system with multi-source power generation in deregulated power environment', *Int. J. Electr. Power Energy Syst.*, 2014, **57**, (5), pp. 277–286

[8] Shekhar, G.T.C., Sahu, R.K., Baliarsingh, A.K., *et al.*: 'Load frequency control of power system under deregulated environment using optimal firefly algorithm', *Int. J. Electr. Power Energy Syst.*, 2016, **74**, pp. 195–211

[9] Debbarma, S., Saikia, L.C.: 'Bacterial foraging based FOPID controller in AGC of an interconnected two-area reheat thermal system under deregulated nvironment'. Int. Conf. on Advances in Engineering, Science and Management, Nagapattinam, Tamil Nadu, India, March 2012, pp. 303–308

[10] Dahiya, P., Sharma, V., Naresh, R.: 'Automatic generation control using disrupted oppositional based gravitational search algorithm optimised sliding mode controller under deregulated environment', *IET Gener. Transm. Distrib.*, 2016, **10**, (16), pp. 3995–4005

[11] Ibraheem Kumar, P., Kothari, D.P.: 'Recent philosophies of automatic generation control strategies in power systems', *IEEE Trans. Power Syst.*, 2005, **20**, (1), pp. 346–357

[12] Wu, S., Meng, J.E., Gao, Y.: 'A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks', *IEEE Trans. Fuzzy Syst.*, 2001, **9**, (4), pp. 578–594

[13] Zeynelgil, H.L., Demiroren, A., Sengor, N.S.: 'The application of ANN technique to automatic generation control for multi-area power system', *Int. J. Electr. Power Energy Syst.*, 2002, **24**, (5), pp. 345–354

[14] Yu, T., Zhou, B., Chan, K.W., *et al.*: 'Stochastic optimal relaxed automatic generation control in non-Markov environment based on multi-step Q(λ) learning', *IEEE Trans. Power Syst.*, 2011, **26**, (3), pp. 1272–1282

[15] Yu, T., Zhou, B., Chan, K.W., *et al.*: 'R(λ) imitation learning for automatic generation control of interconnected power grids', *Automatica*, 2012, **48**, (9), pp. 2130–2136

[16] Yu, T., Zhang, S.P.: 'Optimal CPS control for interconnected power systems based on SARSA on-policy learning algorithm', *Power Syst. Protection Control*, 2013, **41**, (1), pp. 211–216 (in Chinese)

[17] Yu, T., Wang, H.Z., Zhou, B., *et al.*: 'Multi-agent correlated equilibrium Q(λ) learning for coordinated smart generation control of interconnected power grids', *IEEE Trans. Power Syst.*, 2015, **30**, (4), pp. 1669–1679

[18] Xi, L., Yu, T., Yang, B., *et al.*: 'A novel multi-agent decentralized win or learn fast policy hill-climbing with eligibility trace algorithm for smart generation control of interconnected complex power grids', *Energy Convers. Manage.*, 2015, **103**, pp. 82–93

[19] Farhangi, R., Boroushaki, M., Hosseini, S.H.: 'Load–frequency control of interconnected power system using emotional learning-based intelligent controller', *Int. Journal of Electrical Power & Energy Systems*, 2012, **36**, (1), pp. 76–83

[20] Valikhani, M., Sourkounis, C.: 'A brain emotional learning-based intelligent controller (BELBIC) for DFIG system'. Int. Symp. on Power Electronics, Electrical Drives, Automation and Motion, Ischia, 18–20 June 2014, pp. 713–718

[21] Hosseinzadeh Soreshjani, M., Arab Markadeh, G., Daryabeigi, E., *et al.*: 'Application of brain emotional learning-based intelligent controller to power flow control with thyristor-controlled series capacitance', *IET Gener. Transm. Distrib.*, 2015, **9**, (14), pp. 1964–1976

[22] Dehkordi, B.M., Parsapoor, A., Moallem, M., *et al.*: 'Sensorless speed control of switched reluctance motor using brain emotional learning based intelligent controller', *Energy Convers. Manag.*, 2011, **52**, (1), pp. 85–96

[23] Khooban, M., Javidan, R.: 'A novel control strategy for DVR: optimal bi-objective structure emotional learning', *Int. J. Electr. Power Energy Syst.*, 2016, **83**, pp. 259–269

[24] Jafari, E., Marjanian, A., Solaymani, S., *et al.*: 'Designing an emotional intelligent controller for IPFC to improve the transient stability based on energy function', *J. Electr. Eng. Technol.*, 2013, **8**, (3), pp. 478–489

[25] Daryabeigi, E., Abjadi, N.R., Markadeh, G.R.A.: 'Automatic speed control of an asymmetrical six-phase induction motor using emotional controller (BELBIC)', *J. Intell. Fuzzy Syst.*, 2014, **26**, (4), pp. 1879–1892

[26] Sadeghieh, A., Sazgar, H., Goodarzi, K., *et al.*: 'Identification and real-time position control of a servo-hydraulic rotary actuator by means of a neurobiologically motivated algorithm', *ISA Trans.*, 2012, **51**, (1), pp. 208–219

[27] Vargas-Clara, A., Redkar, S.: 'Unmanned ground vehicle navigation using brain emotional learning based intelligent controller (BELBIC)', *Smart Sci.*, 2015, **3**, (1), pp. 10–15

[28] Abdi, J., Moshiri, B., Abdulhai, B.: 'Emotional temporal difference Q-learning signals in multi-agent system cooperation: real case studies', *IET Intell. Transp. Syst.*, 2013, **7**, (3), pp. 315–326

[29] Abdi, J., Moshiri, B., Abdulhai, B., *et al.*: 'Short-term traffic flow forecasting: parametric and nonparametric approaches via emotional temporal difference learning', *Neural Comput. Appl.*, 2013, **23**, (1), pp. 141–159

[30] Zhou, B., Chan, K.W., Yu, T.: 'Q-learning approach for hierarchical AGC scheme of interconnected power grids', *Energy Proc.*, 2011, **12**, pp. 43–52

[31] Yao, M., Shoults, R.R., Kelm, R.: 'AGC logic based on NERC's new control performance standard and disturbance control standard', *IEEE Trans. Power Syst.*, 2000, **15**, (4), pp. 1455–1456

[32] Jaleeli, N., VanSlyck, L.S.: 'NERC's new control performance standards', *IEEE Trans. Power Syst.*, 1999, **14**, (3), pp. 1092–1099

[33] Yu, T., Zhang, S.P.: 'Automatic control of electricity generation based on 5-component update learning algorithm SARSA (λ)', *Control Theory Appl.*, 2013, **30**, (10), pp. 1246–1251 (in Chinese)

[34] Deforest, N., Mendes, G., Stadler, M., *et al.*: 'Optimal deployment of thermal energy storage under diverse economic and climate conditions', *Appl. Energy*, 2014, **119**, (119), pp. 488–496