Scientific Research

Q-learning Based Dynamic Optimal Relax Automatic Generation Control

Tao Yu, Ye Yuan, Haihua Liang

College of Electric Power, South China of Technology, Guangzhou 510640, Guangdong, China

Email: taoyu1@scut.edu.cn, handsomeyuanye@163.com, scutlhh@qq.com

Abstract: Relax control is to relieve the control of the plants while CPS compliances are ensured. This paper describes an application of the Q-learning algorithm in Automatic Generation Control (AGC) to achieve relax control. As the Q-learning algorithm always pursuits the maximum reward in long term, the number of pulse reversals, the value of CPS, and the change of the power outputs are introduced as the control variables in the reward function of the Q-learning controller. To get the maximum long-term reward, Q-learning controller will try to reduce the number of pulse reversals, to ensure the CPS compliances, and to decrease the change of the power outputs. When the coefficients of the control variables are suitable, the CPS compliances are ensured, the number of pulse reversals are reduced , and the power outputs are kept changing smoothly. Cases show that the proposed controllers can obviously enhance the performance of relax control of AGC systems while the CPS compliances are ensured.

Keywords: Reinforcement learning, Q-learning algorithm, Control performance standard, Relax Automatic Generation Control

I. Introduction

AGC is an indispensable mean of technology, which can maintain power system frequency and tie-line power flow in requested range. In order to evaluate the control performance of AGC, the North American Electric Reliability Council (NERC) released control performance standards (CPS) for AGC in 1997. After this, how to design the AGC system's optimum control policy under the CPS standard becomes a brand-new research subject.

To ensure CPS compliances, the general idea is to improve the AGC control strategy under the old CPC standard [1]. The NARI Group in China has done a series of practical work [2-3]. And the CPS1 and CPS2 controller that based the PI structure are designed. Fuzzy control is also used to research the CPS control strategy [4]. "Wedge-Shaped" control law and model predictive control (MPC) method are combined in paper [5].

A new CPS based AGC concept called Relax Control methodology is proposed by the authors in [6], in which both the relax and tighten control directions for AGC system is study by using a simple supervised learning method. Q-learning based dynamic optimal CPS control is introduced in paper [7]. Q-learning method is one of the most important methods in reinforcement learning based on the Markov Decision Process (MDP) theory which further enhances the on-line learning and dynamic optimized ability of the control strategy. The paper [7] compared Q-learning controller with PI controller, and the simulation results show that Q-learning controller can obviously enhance the robustness and adaptability of AGC systems while the CPS compliances are ensured. In this paper, the relax control methodology is further studied by using Q-learning method. And the number of pulse reversals is taken as a control variable of the reward function in Q-learning. By adjusting the coefficient of the number of pulse reversals and the value of λ/μ in the reward function, the effect of relax AGC can be achieved.

The paper is organized as follows: Section II formulates a brief introduction of NERC's control performance standards; Section III describes the Q-learning approach and its relax controller design; Section IV discusses the simulation of the two-area power system model; Section V concludes this paper.

II.Q-learning Method for Relax Control

The relaxed control performance is reflected in the amount of control pulses for total power regulation command from the dispatching terminal. The number of pulses is defined as the average number of pulses that are sent to each regulating unit per hour. Similarly, the number of pulse reversals is defined as the average number of direction changes in pulses sent to each regulating unit per hour.

In order to achieve relax AGC, Q-learning controller has to control two variables. One is the number of pulse reversals, the other is the CPS1. The two variables are expected to have a good long-term performance. The Q-learning controller's goal is to maximize the long-term return. So the on-line optimization Q-learning approach

This work is jointly supported by the authors gratefully acknowledge the support of National Natural Science Foundation of China (50807016), Guangdong Natural Science Funds Project (9151064101000049), the Fundamental Research Funds for the Central Universities, The Hong Kong Polytechnic University (Project G-U494).

is very applicable for the relax AGC.

The relax AGC structure is illustrated in Fig.1. It shows input signals of the Q-learning controller. The ACE/ΔF/CPS Real-time Monitor Values Database is responsible for real-time supervision and data acquisition for interconnected power grids, such as the instantaneous and average value of ACE, ΔF, CPS1. And the Long-termed Historical Database records and deposits the statistical CPS compliance data.

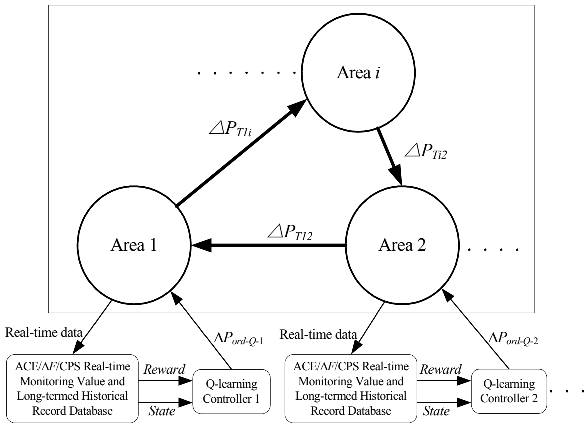


Figure 1. Q-learning based relax control structure.

To design a Q-learning controller, the state-action pairs information should be quantized and the immediate reward function $R(s,s',a)$ should be defined properly for the learning system. It is suitable to choose the CPS1 and the ACE as state information to constitute the state space S (the state space of our Markov chain). Then the state space S should be discretized. The first state variable CPS1 can be discretized as the following 23 levels, $(-\infty, 0)$, $[0, 100)$, $[100, 105)$, $[105, 110)$, $[110, 115)$ ··· $[185, 190)$, $[190, 195)$, $[195, 200)$, $[200, +\infty)$. The second state variable ACE is to distinguish the cause for the change of CPS1 value. It can be discretized as negative and positive. Then, state space S can be discretized as 46 states. The action space A is discrete power regulation introductions of the AGC. How to quantify the action space A depends on the capacity and the type of the generator.

Then the reward function $R_i(s,s',a)$ of control area i can be given as blow:

$$\begin{cases} R_i(k)=\ \sigma_i, \quad \sigma_i\ge 0, & CPS1_i(k)\ge 200 \\ R_i(k)=-\{\lambda_{1i}[ACE_i(k)-ACE^*_i]^2 & \\ \quad +\mu_{1i}[a_{ord-i}(k)-a^*_{ord-i}]^2+v_{1i}N_{per-cycle}\}, & CPS1_i(k)\in[100,200) \\ R_i(k)=-\{\lambda_{2i}[CPS1_i(k)-CPS1^*_i]^2 & \\ \quad +\mu_{2i}[a_{ord-i}(k)-a^*_{ord-i}]^2+v_{2i}N_{per-cycle}\}, & CPS1_i(k)<100 \end{cases} \quad (1)$$

where σ_i can be an arbitrary positive number, here we choose 1000 in the case study. $CPS1_i(k)$ and $ACE_i(k)$ represent the instantaneous value of $CPS1$ and ACE at the kth learning time respectively, CPS_i^* and ACE_i^* express the set target value for the controlled variable CPS1 and ACE respectively. As for the $CPS1_i^*$, our experience shows that a value of 200 works well if higher CPS1 compliance is required, and we can also select the daily or monthly mean of CPS1 to carry out the relaxed control in area i. The value of the ACE^*_i, in our application, is chosen the threshold value of dead-zone in order to improve the CPS2 compliance, reduce inadvertent exchange electric quantity and prevent ACE from crossing zero frequently. $a_{ord-i}(k)$ is the pointer of the selected action from the action sets A at the kth learning step, not the real power control action value, a^*_{ord-i} is the pointer of which the control action is equal to 0, supplementation with the quadratic term of action variation is to avoid mechanical wear-and-tear of AGC generators and a series of economic cost resulting from large fluctuation in the power control signal. λ_{1i}, λ_{2i} and μ_{1i}, μ_{2i} are the optimum weight factors for reward function $R_i(s, s', a)$ in control area i, which are equivalent to the parameters of matrix Q and R in linear quadratic regulator (LQR) algorithm. $N_{per-cylce}$ is the number of pulse reversals in a cycle. And v is the weight of the number of pulse reversals in the reward function.

The relax control can be achieved in the following two ways.

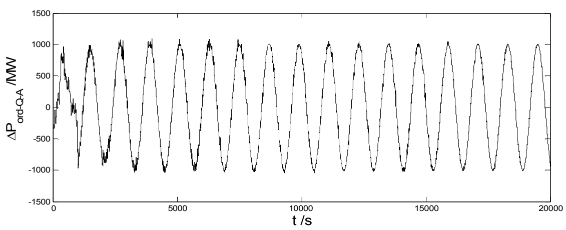
The first way is to change the weight of v related to the number of pulse reversals. If v is given a big weigh, the value of the $R_i(k)$ will become small. In order to get a large value of $R_i(k)$, the Q-learning controller will try to reduce the $N_{per-cylce}$. Then, the number of pulse reversals will be reduced and the relax control can be achieved. However, if the value of v is too big, the number of pulse reversals will be too small. Then we may not get qualified value of CPS1. So it is important to choose a suitable value of v.

The other way is to change the value of λ/μ. The power control outputs will slow down along with the decrease of weight ratio λ/μ, so that the regulating pressure of AGC plants will also be release to achieve the relax control based on NERC's CPS, conversely, Q-learning controller tends toward the tightened control. The proposed controller is an intelligent controller capable of online self-learning and dynamic optimization, power dispatchers can modify the weight ratio λ/μ online to implement relaxed or tightened control for AGC system

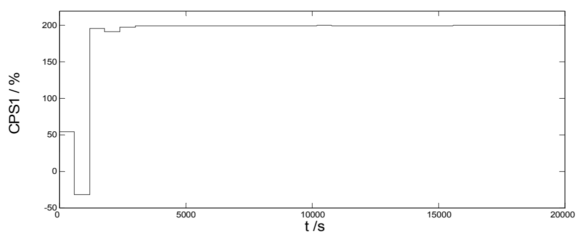
III. Performance Results

The performance of the proposed Q-learning method has been assessed through simulation studies. The test system used for studying the performance of the algorithm is the generally Two-area Power System Load-frequency Control (LFC) Model. The parameters of this system are taken from [8]. AGC decision cycle time is 4s and the three-dimensional variables ($CPS1$, ACE, $N_{per\text{-}cylce}$) are state input signals to Q-learning controller. As the AGC decision cycle time is 4s, the number of pulse reversals is calculated every cycle. So the value of $N_{per\text{-}cylce}$ is an integer between zero and one. The output control vector is discretized in eleven values equal to A = { -300, -200, -100, -50, -20, -5, 0, 5, 20, 50, 100, 200, 300 } MW.

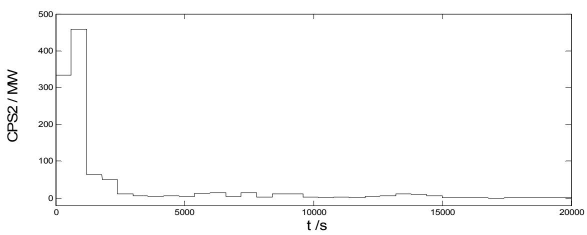
Then, the performance of the controller can be tested on the model. As the process of Q-learning can be divided into two processes, the pre-learning process and the normal regulation process, the controller can be tested in the two processes respectively. In the pre-learning process, sine wave is used to simulate the load disturbance. The amplitude of the sine wave is 1000MW and the cycle of the sine wave is 1200s. Fig.2 shows the response of the system in the pre-learning process of one area.



(a) The outputs of the Q-learning controller



(b) The value of CPS1 for every 10 minutes.



(c)The value of ACE for every 10 minutes

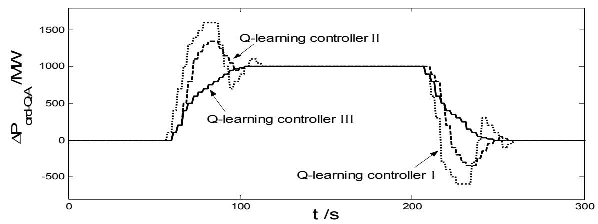
Figure 2. Response of the system during the pre-learning process

From Fig.2, it can be seen that after a period of exploring, the outputs of the Q-learning can trace the fluctuation of the load disturbance. And the CPS1 and CPS2 can be kept constants. Then, the pre-learning process ends.

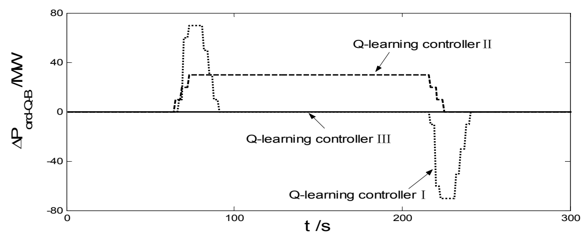
In the normal regulation process, the following three typical Q-learning controllers with different parameters are adopted to test the performance of relax control.

- Q-learning controller I: $\lambda 1$=1, $\lambda 2$=50, $\mu 1$=$\mu 2$=1, $v1$=$v2$=0
- Q-learning controller II: $\lambda 1$=1, $\lambda 2$=50, $\mu 1$=$\mu 2$=20, $v1$=$v2$=0
- Q-learning controller III: $\lambda 1$=1, $\lambda 2$=50, $\mu 1$=$\mu 2$=20, $v1$=$v2$=60

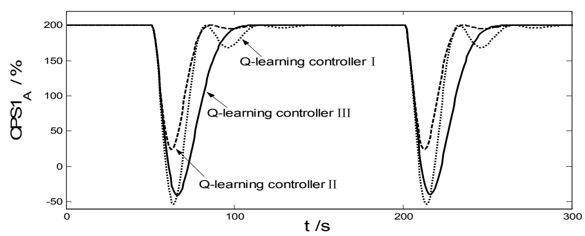
The square-wave load is used to simulate the load disturbance and the response of the system is showed in the figure blow:



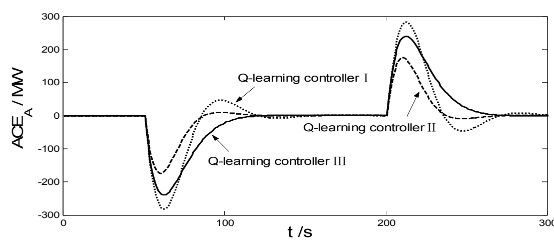
(a) The outputs of the different Q-learning controllers of A area.



(b) The outputs of the different Q-learning controllers of B area.



(c)The value of CPS1of A area at different values of ν and λ/μ



(d) The value of ACE of the A area at different values of ν and λ/μ

Figure 3. Response of the system during normal regulation process.

978-1-935068-17-4 © 2010 SciRes.

From the figure above, it can be seen that different parameters in Eq. (1) influence the performance of relax control. When the weight of the v is big, to get more reward, the Q-learning controller will try to decrease the number of pulse reversals and the performance of the relax control is enhanced. It can be seen from comparing the outputs of Q-learning controller II and III Similarly, to get more reward, the power control outputs will slow down along with the decrease of weight ratio λ/μ. It can be also seen by comparing the outputs of Q-learning controller I and II. However, Fig. 3(c) and (d) show that the value of the CPS1 and ACE is not so good when the performance of the relax control is enhanced.

In the normal regulation process, the the number of pulse and pulse reversals is also calculated as they reflect the performance of the relax control. When they are small, the generators don't regulate frequently, and the performance of the relax control is obvious.

The white noise load disturbance is adopted. The amplitude of the load is less than 1500MW, and the sampling time of the load is 15 minutes. Parameter disturbance is white noise with limited bandwidth of ten percents. In area A, the number of pulse and pulse reversals during 24 hours are calculated at different values of v and λ/μ to show the performance of relax control. They are showed in the following Tab. I and II.

From the tables, it can be seen that the number of pulse and pulse reversals becomes smaller when v becomes bigger and λ/μ becomes smaller. However, the CPS compliances are decreasing at the same time. So it is important to get banlance between the performance of relax control and the CPS compliance.

Table I. The results of the simulation at different value of μ when λ1=1 λ2=50, v1=v2=0, μ1=μ2

Index	μ=0	μ=10	μ=20	μ=30	μ=40	μ=50
CPS1 (%)	175.3	176.7	172.6	169.5	167.8	162.9
CPS2 (%)	96.7	96.1	95.5	94.7	94.1	93.4
CPS (%)	95.1	95.8	94.3	93.9	93.3	91.5
Pulse No.	335	320	287	267	263	273
Pulse Rev.No.	95.9	85.4	67.9	60.2	56.5	51.1

Table II. The results of the simulation at different value of v when λ1=1 λ2=50, μ1=μ2=30, v1=v2

Index	v=10	v=20	v=30	v=40	v=50	v=60
CPS1 (%)	168.	163.	162.	153.	146.	137.
CPS2 (%)	5	2	993.	7	8	5
CPS (%)	94.6	94.0	2	91.6	89.4	85.7
Pulse No.	93.7	93.3	92.5	90.1	87.1	84.2
Pulse	260	255	251	267	280	302
Rev.No.	58.5	56.2	54.7	60.2	42.5	38.9

IV. Conclusion

Relax AGC is significant for conserve energy in grid dispatch on NERC's new CPS. It can reduce the loss and the regulated pressure of generators. In this paper, we use Q-learning controller to implement the relax control and cases show that the proposed controllers can obviously enhance the performance of relax control while the CPS compliances are ensured. The conclusions are showed as blow:

Firstly, by adjusting the parameters v and λ/μ in the reward function of the Q-learning controller, the relax AGC can be achieved. The number of pulse reversals can be reduced and the pressure of the generators can be relieved.

Secondly, though the generators do not regulate frequently as the relax control, the CPS compliances can be ensured. So the Q-learning controller based relax control can be applied in practice.

References

[1] N. Jaleeli, L. S. VanSlyck, "NERC's new control performance standards," IEEE Trans. on Power System, vol. 14, pp. 1091-1099, Aug. 1999.

[2] Gao Zonghe, Teng Xianliang, Tu Liqun, Hierarchical AGC mode and CPS control strategy for interconnected power systems [J]. *Automation of Electric Power Systems*, vol. 28, pp. 78-81, 2004.

[3] Gao Zonghe, Teng Xianliang, Zhang Xiaobai. "CPS control strategy for interconnected power systems." *Automation of Electric Power Systems*, vol. 29, pp. 40-44, 2005.

[4] A.Feliachi, D.Rerkpreedapong, "NERC compliant load frequency control design using fuzzy rules," Electric Power Systems Research, vol. 73, pp. 101–106, 2005

[5] N. Atic, A.Felachi, D.Rerkpreedapong, "CPS1 and CPS2 compliant wedge-shaped model predictive load frequency control," *In the Proceedings of IEEE PES General Meeting* , pp. 855 – 860 Jun. 2004.

[6] Yu Tao, Chen Liang, Cai Guanglin, "CPS Statistic Information Self-learning Methodology Based Adaptive," *Automatic Generation Control. Proceedings of CSEE*, vol. 28, pp. 45-49, 2008.

[7] T. YU, B. Zhou, K. W. Chan, "Q-learning Based Dynamic Optimal CPS Control for Interconnected Power Grids," *Proceedings of CSEE*, vol. 29, pp. 13-19, Jul. 2009.

[8] G.K. Chana, S. Asgarpoor, "Optimum Maintenance Policy with Markov Decision Processes," *Electric Power System Research*, vol. 76, pp. 452–456, 2006..

[9] C. J. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.* , vol. 8, pp. 279-292, Aug. 1992.

[10] Ray G, Prasad A N, Prasad G D, "A new approach to the design of robust load frequency controller for large scale power systems, " *Electric Power System Research*, vol. 51, pp. 13-22, Jul. 1999.

[11] H.Song, C.Liu, J. Lawarrée, W. Dahlgren, "Optimal electricity supply bidding by Markov Decision Process," *IEEE Trans. On Power System*, vol. 15, pp. 618-624, 2000.

[12] N.Jalceli, L.S.VanSlyck, "Control Performance Standards and Procedures for Interconnected Operation," *Technical Report EPRI TR-107813*, April, 1997.