Application Study of Reinforcement Learning Control for Building HVAC System

Sung-Hwan Cho†

Department of Mechanical & Automotive Engineering, JeonJu University, JeonJu 560-759, Korea

Key words: HVAC, PI control, Reinforce learning control, TRNSYS program

ABSTRACT: Recently, a technology based on the proportional integral (PI) control have grown rapidly owing to the needs for the robust capacity of the controllers from industrial building sectors. However, PI controller generally requires tuning of gains for optimal control when the outside weather condition changes. The present study presents the possibility of reinforcement learning (RL) control algorithm with PI controller adapted in the HVAC system. The optimal design criteria of RL controller was proposed in the environment chamber experiment and a theoretical analysis was also conducted using TRNSYS program.

Nomenclature

a : behavior
E : probability
K : control gain
Q : target value
R : reinforce value
s : status
T : temperature [℃]

Greek symbols

γ : damped coefficient

Subscripts

i : integrator controller
p : proportion controller, policy
t : time

1. Introduction

Most of industrial heating, ventilating, and air-conditioning (HVAC) systems and refrigeration systems have a proportional integral (PI) controller. When operational conditions changes, such systems do not maintain their optimal control well, and subsequently the inappropriate control increases the consumption of energy. Therefore, it is necessary to determine the dynamic characteristics of a plant using appropriate tuning. This tuning process, however, takes a lot of time and cost, and is less applicable to a system having strong non-linear properties and long delay time. Since the control performance may change after the tuning, moreover, maintaining the optimal control requires re-tuning. As a solution to these problems, self tuning control algorithms, e.g. neural network control and reinforcement learning (RL) control, have been used. The neural network control has a slow learning rate in a complicated neural network, and does not work in neural saturation, and requires collecting a lot of data in off-line mode.

Watkins et al.[4] developed Q-learning as an optimal learning method. Anderson et al.[5,6] applied RL to a heating coil simulation to compare it with neural network and PI controls in terms of the theories and experimental settings

† Corresponding author
Tel.: +82-63-220-2663; fax: +82-63-220-2663
E-mail address: shcho@jj.ac.kr

of control variables and the abilities of learning. Barto et al.[7] studied the learning and execution of a program having real-time dynamic characteristics. Sutton[8] used a temporal difference model as an evaluation function learning method of RL.

This study developed a model allowing the control of learning in on-line mode and the self-tuning to improve the control performance of HVAC systems by using an RL control algorithm which reinforces output control signals of a PI controller. In the experiments of this study, the model was applied to a HVAC system in an actual construction. As a theoretical study, a method of designing optimal reinforcement control, which is one of the most important factors in a RL controller, was proposed by creating control algorithms (PI, RL) as a module of the TRNSYS program and using dynamic simulations under various different conditions.

2. Control algorithm

2.1 Basic concepts

RL (Reinforce Learning) is an approach to optimal learning from interactions with the environment given by reinforcement signals even without a lot of knowledge of the environment. The Q-learning developed by Watkins et al.[4] is the most popular RL technique, and an algorithm to find an optimal action considering the attenuation in the reinforcement for future actions. This algorithm defines Q-values of state-action pairs to determine the optimal action at each state.

RL has two components of learning: agent and environment. The agent takes an action suitable for the state given in the environment. The environment sends to the agent reinforcement signals indicating whether the action taken will result in a suitable change of the state and action. The agent learns from the repetition of these processes.

2.2 Reinforce learning control algorithm

The basic process of Q-learning can be described as follows:

(1) The status, 's' and the target value, $Q(s, a)$ for action, 'a', is initialized to default, usually zero.

(2) The current state 's' is recognized.

(3) An action is chosen in accordance with the state-action rules.

(4) The action 'a' is performed at the given state, and the followed environment is given as s_t.

(5) The action applied as a result of a reinforcement and the compensation at the state are defined as $R(s_t, a_t)$.

(6) The target value, $Q(s_t, a_t)$ is a value function at the given state, s_t, and action, a_t, and is described as Eq. (1),

$$Q_\pi(s_t, a_t) = E_g\left\{\sum_{k=0}^{T}\gamma^k R(s_{t+k}, a_{t+k})\right\} \tag{1}$$

where γ is a damped coefficient, ranged from 0 to 1 and k is iteration number at the state s_t.

(7) Adding the sum of immediate reinforcement and future reinforcement to Eq. (1) give

$$\begin{aligned} Q_\pi(s_t, a_t) &= E_g\left\{R(s_t, a_t)\sum_{k=1}^{T}\gamma^k R(s_{t+k}, a_{t+k})\right\} \\ &= E_g\left\{R(s_t, a_t) + \sum_{k=0}^{T-1}\gamma^k R(s_{t+k+1}, a_{t+k+1})\right\} \end{aligned} \tag{2}$$

(8) The policy evaluation in a dynamic program can be obtained by an iterative calculation until the value function converges into a wanted sum. This iterative policy evaluation is presented as the current value of the value function as in Eq. (3),

$$\begin{aligned} \Delta Q_\pi(s_t, a_t) = E_g\{R(s_t, a_t) + \gamma Q_\pi(s_{t+1}, a_{t+1})\} \\ - Q_\pi(s_t, a_t) \end{aligned} \tag{3}$$

where an expected value is determined for a possible next state. The expected value requires a single model in state transition probability, and if there is not a model, Monte Carlo learning is to be used.

(9) A value iteration program, which is a combination of policy evaluation and policy improvement, is used to improve the action-choice policy and achieve optimal control

$$\Delta Q_\pi(s_t, a_t) = a_t\{R(s_t, a_t) + \gamma Q_\pi(s_{t+1}, a_{t+1})\} - Q_\pi(s_t, a_t) \quad (4)$$

where learning rate, a_t, ranges $0 \le a_t \le 1$.

(10) Assuming that the total reinforcement value is to be minimized, the Monte Carlo learning expresses the value iteration for the target value as Eq. (5),

$$\Delta Q_\pi(s_t, a_t) = a_t\left\{R(s_t, a_t) + \gamma \min_{a' \in A} Q_\pi(s_{t+1}, a_{t+1})\right\} - Q_\pi(s_t, a_t) \quad (5)$$

2.3 Structure of RL controller

Watkins et al.[4] proved that the optimal sum of reinforcement could be obtained by minimizing the target value, Q, at the state when choosing an action, 'a'. Figure 1 shows a schematic of a RL controller combined with a PI controller. TD error is temporal difference between target value and action value. Q algorithm selects action considering the value of TD error.

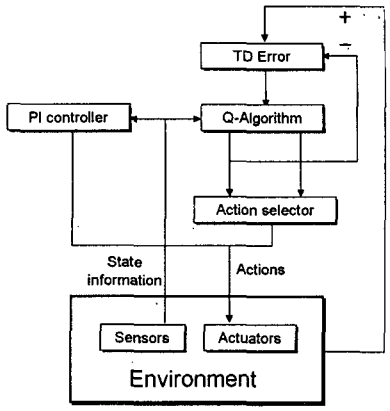


Fig. 1 Structure of RL controller combined with PI controller.

3. Experiments

In order to determine the applicability of RL control technology to actual HVAC systems, a RL controller and a PI controller were tested at an environment chamber in Korea Institute of Energy Research under winter conditions. The RL controller (K_p, K_i) was arranged to reinforce the control gains based on the PI controller.

3.1 Application to environmental chamber

The environment chamber is a test facility used to create artificial atmospheres for the analysis of the performance of HVAC systems and the thermal characteristics of a building, e.g. interior environment, energy consumption, and system capacity. The environment chamber contains a test house which is constructed to

Table 1 Operation range of experimental system

	Operation range
Indoor condition	24℃ (75.2°F)
Outdoor condition	−5~10℃
Supply fan	Max.: 1,000 CMH (0.278 m^3/kg), Min.: 200 CMH (0.055 m^3/kg)
Return fan	Max.: 900 CMH (0.250 m^3/kg), Min.: 200 CMH (0.055 m^3/kg)
Cooling coil	Capacity: 13,608 kcal/h, Condenser: 9,072 kcal/h and 4,536 kcal/h, Inlet cooling water temp.: 7℃, Outlet cooling water temp.: 13℃
Supply set pressure	45 mmAq (448 Pa)

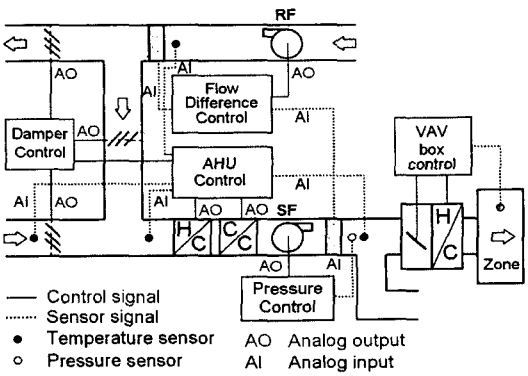


Fig. 2 Schematic diagram of AHU.

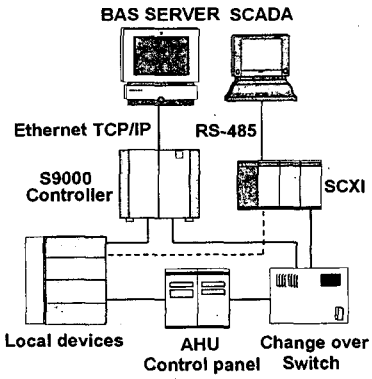


Fig. 3 System set-up for experiment.

monitor the heating and air-conditioning loads heat transfer in structures, and HVAC control. Table 1 describes the operational conditions for the instruments inside the test house.

Figure 2 illustrates the non-heated-floor room in the test house and the variable-air-volume HVAC system installed in the environment chamber. The HAVC system maintained the air supply temperature constant with the change in interior loads, and controlled the room temperature by changing the airflow for each room and area.

3.2 System set-up

The operational control of the HVAC system was composed of the supervisory control by the main computer and the local loop control. Figure 3 shows the operational control system used to automatically control the HVAC system for the test house. This study used both the existing operational control system (BAS) and the operational control system (SCADA) for the RL algorithm testing. For the existing operational control system, an Ethernet TCP/IP-based data interface was used for the supervisory control and the local loop control to monitor and control the data in real-time. As it was impractical to make a test for a variety of actual control algorithms, this study developed an operational control program using the PI control and RL control algorithms suitable for this test to compare their control performances.

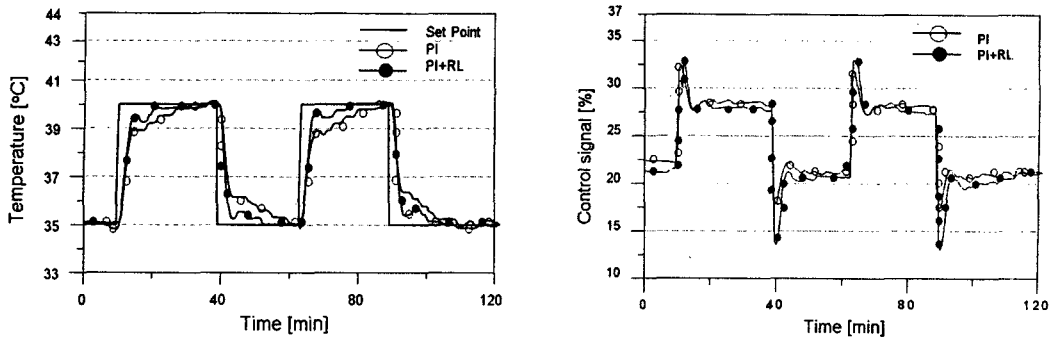
4. Results and discussions

4.1 Evaluation of RL controller

To assess the RL controller, the external and interior temperatures under winter conditions were maintained to 5℃ and 24℃, respectively, by controlling the heater coils using the PI controller and the RL controller. Figure 4(a) shows the reactions of the PI controller and the RL controller to the temperature change from 35℃ to 40℃ by stages. When the air supply temperature was changed as shown in Fig. 4(b), the RL controller worked better than the PI controller: for example, the errors in normal service decreased and the responses were faster. It was also found that with the change of operational settings, the RL controller reinforced the output control signals more than the PI controller.

4.2 Optimization of design parameters

One of the important things in the design of a RL controller is to determine the settings to reinforce the output when the controller has large input and output errors. This study examined the reactions to various reinforcement



(a) Supply air temperature (b) Control signal

Fig. 4 Control performance of PI controller and RL controller combined with PI controller.

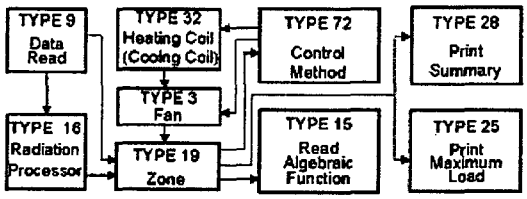


Fig. 5 Flow chart of TRNSYS program.

signals under winter conditions to develop the criteria for designing a RL controller. For the examination, this study programmed the method of analyzing the PI controller and the RL controller using the fin efficiency and effectiveness-NTU and the method of controlling the temperatures in a radiant heating room. These methods were used as modules for the TRNSYS program. Figure 5 shows a schematic of the TRNSYS program.

4.3 Application to TRNSYS

The air is heated by the heating coils (Type 32) and then supplied to the interior (Type 19). The air coming into the interior provides information for the control mechanism (Type 72), and subsequently the PI controller and RL con-

Table 2 TRNSYS input data

System	Parameter	Data
Heating coil (Cooling coil)	Number of row deep / Number of parallel cooling circuits (-)	6/24
	Coil face area (m^2)	4.17
	Inside tube diameter (m)	0.02
	Inlet air dry bulb temp / Inlet air wet bulb temp / Inlet water temp (℃)	15/15/10
	Mass flow rate of air / mass flow rate of water (kg/hr)	500/100
Fan	Max. flow rate / Inlet mass flow rate (kg/hr)	2500/100
	Fluid specific heat rate (kJ/kg℃)	1.012
	Max. power consumption (kJ/hr)	3,500
	Fraction of pump power converted to fluid thermal energy, $0<f_{par}<1$	0.9
	Inlet fluid temp. (℃)	15
	Control function	
Duct	Pipe inside diameter / length (m)	0.35/3.45
	Overall loss coefficient based on inside pipe surface area (kJ/hr m^2℃)	2
	Fluid density (kg/m^3) / specific heat rate (kJ/kg℃)	1.2/1.012
	Initial fluid temp. / Outlet temp. (℃)	15/25
	Mass flow rate (kg/hr)	100

troller send control signals to the heating or air-conditioning coils and the fan to control the air supply suitably for each environment. The information obtained from Type 19 area is delivered to the output processing section (Type 28) to produce the control data for the environment. Table 2 lists the variables and inputs used in the TRNSYS program. The inputs were decided based on the actual settings of each installation.

4.4 Variation on gain value of PI controller

For a theoretical review, the external air temperatures in summer and winter were set to 35 ℃ and 5℃, respectively, and the supply flow was 600 kg/hr. Figures 6 and 7 show the interior temperatures of 3 cases, K_p =1.5, K_i =4.0, K_p =0.1, K_i =0.5 and K_p =6.0, K_i =2.5. Although the interior temperature did not pursue well the settings when the gain value of the PI controller was K_p =0.1, K_i =0.5, but the case of K_p =6.0, K_i =2.5 were pursued well the settings at the both cases of heating and air-conditioning.

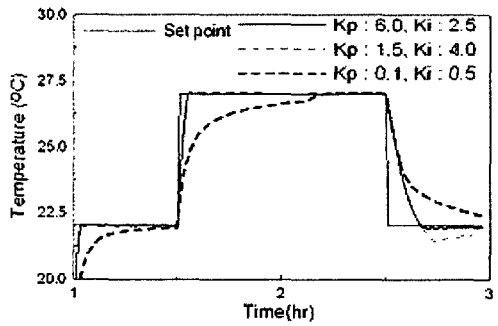


Fig. 6 Response of PI controller in heating mode.

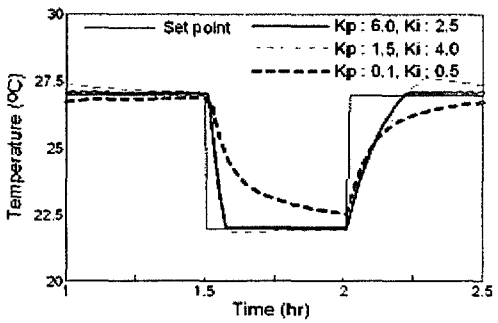
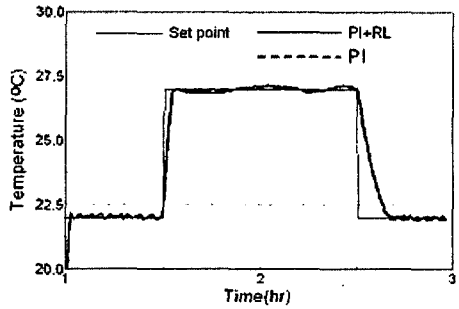


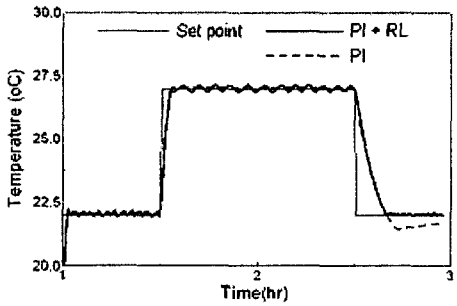
Fig. 7 Response of PI controller in cooling mode.

4.5 Optimization on reinforce signal

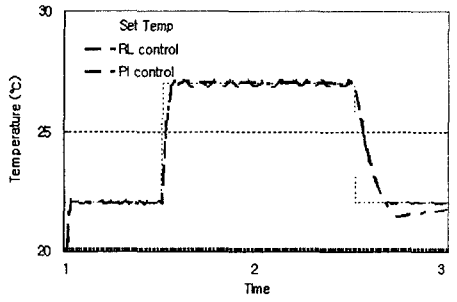
A RL controller combined with a PI con-



(a) K_p =6.0, K_i =2.5



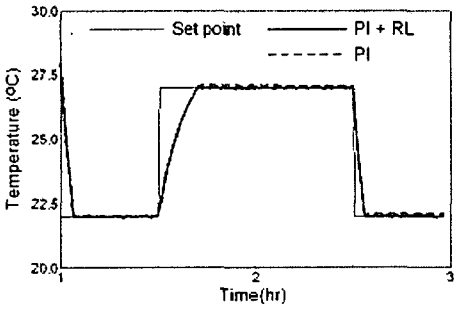
(b) K_p =1.5, K_i =4.0



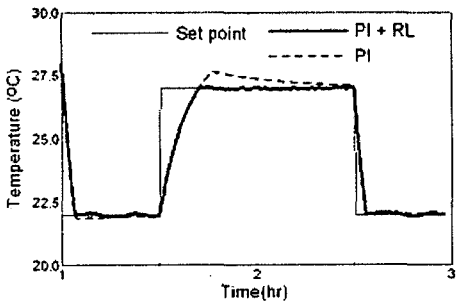
(b) K_p =0.1, K_i =0.5

Fig. 8 Response of PI, RL controller under winter condition.

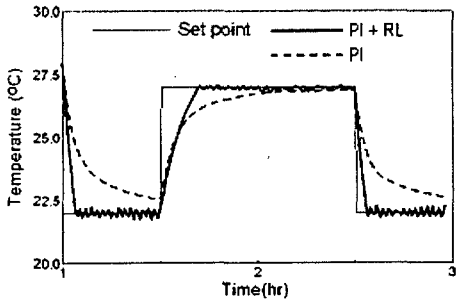
troller can reinforce control signals to respond properly to external conditions and to reach the settings even when control gains (K_p, K_i) are not well controlled or external conditions are rapidly changed. In this study, the gain values (K_p, K_i) reviewed from the PI controller were changed to compare the performances of the PI controller and the RL controller at the winter and summer conditions, i.e. external temperatures of 5 and 35℃, respectively, settings under the winter conditions to 22℃ and 27℃. The reinforcement value for the RL controller was set from −0.5 to +0.5. When the gain value was K_p=6.0, K_i=2.5, the gain was so close to an optimum that the RL controller was not effective. When the gain value reduced to K_p=0.1, K_i=0.5, the PI gain was not suitable for the system. However, the interior temperature took a longer time to reach the set temperature. This result shows that using a RL controller makes it easier to reach a set temperature even when the gain value of a PI controller is not appropriate.



(a) K_p=6.0, K_i=2.5



(b) K_p=1.5, K_i=4.0



(c) K_p=0.1, K_i=0.5

Fig. 9 Response of PI, RL controller under summer condition.

Figure 9 shows the control of air supply with the change of PI gain values when the interior temperature settings were changed to 22℃ and 27℃ under the winter conditions, i.e. the exterior temperature of 35℃. Although the control of gain values K_p, K_i was not appropriate, the RL controller used together with the PI controller increased the ability to pursue the set temperatures. In other words, the RL control facilitated the adaptation to external environments in both winter and summer conditions.

Figures 10 and 11 shows the results of changing the reinforcement signals of the RL controller from 0.05 to 0.90 under the summer and winder conditions, with the gain value of (0.1, 0.5), which was not appropriate. As shown in Fig. 10, the ability to attain the set temper-

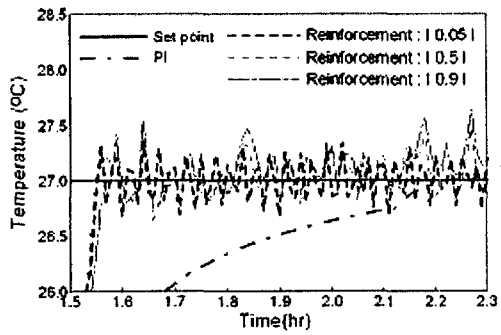


Fig. 10 Response of PI, RL controller under winter condition (K_p=0.1, K_i=0.5).

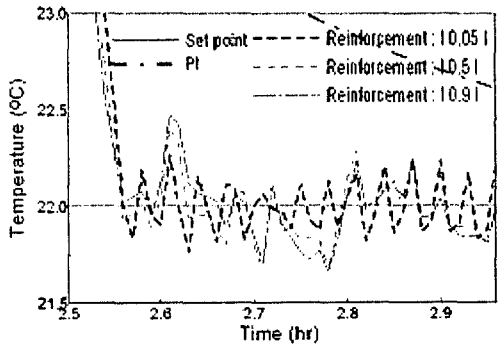


Fig. 11 Response of PI, RL controller under summer condition (K_p =0.1, K_i =0.5).

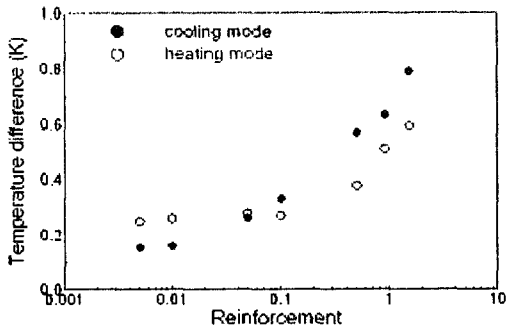


Fig. 12 Temperature difference ($T_{max} - T_{min}$) of indoor temperature with respect to reinforcement variation using PI, RL controller under summer and winter conditions (K_p =0.1, K_i =0.5).

ature increased under the winter conditions when the reinforcement signal was 0.05.

Figure 11 indicates that the reinforcement signal of 0.05 also enhanced the ability to reach the set temperature under the summer conditions. Figure 12 depicts the differences of the maximum and minimum interior temperatures when the reinforcement signals changed from 0.005 to 1.5 under the air-conditioning and heating conditions. In both modes, the differences were almost constant at the reinforcement signal of 0.05. This result indicates for easier control of air-conditioning and heating, it is better to lower the reinforcement signal ratio when designing a RL controller.

5. Conclusions

As one of methods to improve the performance of control in building HVAC system, reinforce learning control algorithm with self-synchronous capacity of PI controller was applied to this study. From the design of RL controller combined with PI controller, experimental and theoretical investigations are performed. From the study, following descriptions are concluded.

(1) Applying a RL control algorithm to a PI controller enhances the reaction rate and the quality of control in both air-conditioning and heating.

(2) Under winter and summer conditions, the reinforcement signal ratio influences the ability to attain a set temperature: when the ratio is lower the ability is better.

(3) Therefore, it is necessary to lower the reinforcement signal ratio when designing a RL controller used together with a PI controller for both air-conditioning and heating systems.

References

1. Ministry of Commerce, Industry and Energy, 2003, Total energy consumption report, pp. 1-80.
2. Virk, G. S. and Loveday, D. L., 1992, A comparison of predictive, PID, and on/off techniques for energy management and control, Proceedings of ASHRAE, pp. 3-10.
3. Hang, C. C. and Åström, K. J. and Ho, W. K., 1991, Refinements of the Ziegler-Nichols tuning formula, IEEE Proceedings Part D-Control Theory Application., Vol. 138, No. 2, pp. 111-118.
4. Watkins, C. and Dayan, P., 1992, Technical note: Q-learning, Machine Learning, Vol. 8, pp. 279-292.
5. Anderson, C. W., Hittle, D. C., Katz, A. D. and Kretchmar, R. M., 1997, Synthesis of reinforcement learning, neural networks, and

PI control applied to a simulated heating coil, Artificial Intelligence in Engineering, Vol. 11, No. 4, pp. 421-429.
6. Anderson, C. W., 1993, Q-learning with hidden-unit restarting, Advances in Neural Information Processing Systems, Vol. 5, Hanson, S. J., Cowan, J. D. and Giles, C. L., eds., Morgan Kaufmann Publishers, San Mateo, CA, pp. 81-88.
7. Barto, A. G., Bradtke, S. J. and Singh, S. P., 1995, Learning to act using real-time dynamic programming, Artificial Intelligence, Special Volume: Computational Research on Interaction and Agency, Vol. 72, No. 1, pp. 81-138.
8. Sutton, R. S, 1988, Learning to predict by the method of temporal difference, Machine Learning, Vol. 9, pp. 9-44.
9. Sutton, R. S. and Barto, A. G., 1998, Reinforcement Learning: an Introduction, Cambridge, MA, MIT Press, pp. 51-85.
10. So, J. H., Cho, S. H., Song, M. H. and Park, M. S., 2001, Experimental study on control performance of reinforcement learning method, Proceedings of the SAREK, pp. 697-701.