강화학습을 이용한 AHU 제어 성능 개선에 관한 연구

정회원 유 승 선*, 김 문 성**, 소 정 훈***, 곽 훈 성****

A Study on Improvement of AHU Control Performance using Reinforcement Learning

Seung-sun Yoo*, Moon-seong Kim**, Jung-hoon So***, Hoon-sung Kwak**** *Regular Members*

요 약

신경회로망을 이용한 대부분의 제어 응용은 인공적인 지각 기능에 의한 자동 조절로 대표된다. 그것은 종종 생물계의 방식과 같은 경험에 의한 지능 제어이고 직관과 패턴인식 제어이다. 현재 많은 자동화된 건물들에서 건물 공조에 대한 제어로서 구조가 단순하고 매우 견실한 특성을 지닌 PI(Proportional Integral) 제어에 의한 공조를 수행하고 있지만 좋은 성능을 유지하기 위해서는 적절한 동조 및 재동조가 필요하다.

본 논문에서는 위의 문제점들을 해결하고 제어기의 제어 성능을 향상시킬 수 있는 방법으로, 신경회로망 학습방법(지도/비지도/강화학습) 중의 하나인 강화학습법을 이용한 강화학습 제어기를 제안하였으며, 이를 환경실험실 내의 실제 운전 중인 공조시스템에서 실험하여 그 타당성을 입증하였다.

ABSTRACT

Most common applications using neural networks for control problems are the automatic controls using the artificial perceptual funtion. These control mechanisms are similar to those of the intelligent and pattern recognition control of an adaptive method frequently performed by the animate nature. Many automated buildings are using HVAC(Heating Ventilating and Air Conditioning) by PI that has simple and solid characteristics. However, to keep up good performance, proper tuning and re-tuning are necessary.

In this paper, as the one of method to solve the above problems and improve control performance of controller, using reinforcement learning method for the one of neural network learning method(supervised/unsupervised/reinforcement learning), reinforcement learning controller is proposed and the validity will be evaluated under the real operating condition of AHU(Air Handling Unit) in the environment chamber.

Ⅰ. 서 론

현대적 제어 이론이 급속하게 발전되고 있지만, 공기 조화 및 냉동기기 등의 산업용 제어기의 대부분이 PID(Proportional Integral Derivative) 형태의 제어기를 사용하고 있다. PID 제어기는 구조의 단순성에도 불구하고 목표치 추종, 외란 및 프로세스 변수에 대한 안정성 및 강인성 등의 우수한 성능을 가지고 있어 산업용 프로세스 제어에 가장 많이 사용되고 있다. 또한 여러 동조 기법을 사용하여 플랜트(plant)의 동특성(dynamics)을 실험적으로 추출하여 최적의 제어를 위한 변수값을 찾아 제어기를 설계할 수 있다[1,2,3].

그러나 프로세스 모델의 불확실성이 존재하거나 운전 환경이 변화하는 경우에는 제어 성능을 미리 예측할 수 없다. 따라서 양호한 제어 성능을 유지하기 위해서는 적절한 동조를 통해서 플랜트의 동특

* (주)비멕(yss2590@hanmir.com),
** 대원과학대학 컴퓨터정보통신과(kms@daewon.ac.kr),
*** 한국에너지기술연구소,
**** 전북대학교 컴퓨터공학과
논문번호 : 010355-1123, 접수일자 : 2001년 11월 23일

성을 추출할 필요가 있지만, 동조 과정은 많은 시간과 경비가 소모될 뿐만 아니라 강한 비선형이나 큰 지연시간을 갖는 시스템에서는 매우 어렵고, 동조 후에도 제어 성능이 감소될 수 있으므로 최적의 제어 성능을 유지하기 위해서는 재동조 과정이 필요하다[1,3].

특히 전형적인 비선형 구조로 운전할 수 밖에 없는 빌딩 자동 제어에 있어서는 수시로 변화하는 제어 환경에 대한 정확한 예측과 자동 동조의 기능이 필수적인 요소라 할 수 있다. 이러한 최적의 제어 성능을 얻을 수 있는 PID 제어기의 파라메터를 결정하는 방법으로 1942년 Ziegler-Nichols[4]에 의해 제안된 Ziegler-Nichols 동조법 이후 Astrom과 Hagglund의 릴레이 실험에 의한 동조법 등 PID 제어기의 동조법에 대한 많은 연구가 이루어졌으며, 현재 Ziegler-Nichols 동조법을 개선한 몇몇 방법들이 사용되고 있다[5,6,7,8,9].

따라서 본 연구에서는 위의 이러한 문제점들을 해결하고 제어기의 제어 성능을 개선할 수 있는 방법의 하나로 확률적 동적 계획법을 기반으로 하는 자유 모델(free model) 강화 학습(Reinforcement Learning)법으로 개발된 Q-학습(Q-Learning)방법을 이용하여 최적 건물 공조 제어기를 설계하여, 인공적으로 외부(outdoor temperature)의 환경을 자유로이 조절할 수 있는 인공 기후 실험동 내부의 건물 공조 시스템에 적용하여 제어기의 정확성과 향후 적용가능성을 검토하였다.

Ⅱ. 제어 이론

1. 비례 제어동작

비례 제어기에서는 제어기의 출력 $m(t)$ 와 오차신호 $e(t)$ 의 관계가

$$m(t) = K_p e(t) \tag{1}$$

이며, 전달함수로 나타내면 다음과 같다.

$$\frac{M(s)}{E(s)} = K_p \tag{2}$$

여기서 K_p 는 비례감도(proportional sensitivity) 혹은 게인(gain)이라 한다. 이러한 비례제어는 오차는 줄여주지만 게인의 증가로 인해 시스템이 불안정하게 될 수도 있다.

2. 적분 제어동작

적분 제어기에서는 제어기의 출력 $m(t)$ 의 값의 변화율이 오차신호 $e(t)$ 에 비례한다. 즉,

$$\frac{dm(t)}{dt} = K_i e(t) \tag{3}$$

또는

$$m(t) = K_i \int e(t)dt \tag{4}$$

여기서 K_i 는 적분(integral) 상수이다. 적분 제어기의 전달함수는 다음과 같다.

$$\frac{M(s)}{E(s)} = \frac{K_i}{s} \tag{5}$$

적분 제어는 정상상태 오차를 줄여주며 파라미터 변동에 대한 강인성을 갖게 하지만 안정성을 감소시킨다.

3. 미분 제어동작

미분 제어기에서는 제어기 출력의 크기 $m(t)$ 가 오차신호 $e(t)$ 의 변화율에 비례한다. 즉,

$$m(t) = K_d \frac{de(t)}{dt} \tag{6}$$

여기서 K_d 는 미분(derivative) 상수이다. 미분 제어기의 전달함수는 다음과 같다.

$$\frac{M(s)}{E(s)} = K_d s \tag{7}$$

미분동작은 일반적으로 감쇠를 증가시켜 안정성을 증가시키며 오버슈트를 감소시킨다. 하지만 미분동작은 고주파 잡음을 증폭시키는 효과도 있다. 또한, 미분제어동작은 단독으로는 사용하지 않는데 그 이유는 이 제어동작이 과도기간 중에만 영향을 미치기 때문이다.

4. 비례-적분-미분 제어동작

비례 제어동작, 미분 제어동작, 적분 제어동작을 조합한 PID 제어기의 방정식과 전달함수는 각각 식(8), 식(9)와 같다.

$$m(t) = K_p e(t) + K_d \frac{de(t)}{dt} + K_i \int e(t)dt \tag{8}$$

$$\frac{M(s)}{E(s)} = K_p + K_d s + K_i \frac{1}{s} \quad (9)$$

비례 제어동작, 미분 제어동작, 적분 제어동작을 조합한 PID 제어동작은 세 가지 기본 제어동작의 유리한 점만을 가지게 할 수 있다.

5. 강화학습(Reinforcement Learning)

강화학습은 크게 두 개의 요소로 구성되어 있다. 하나는 행위자(agent)이고 또 다른 하나는 환경 즉, 외계(environment)이다. 행위자는 외계에서 주어진 상태에 따라 적절한 행위(action)를 외계에 행하며, 외계는 주어진 행위에 따라 변화된 상태와 행위가 적절했는가에 대한 판단을 강화신호로서 행위자에게 보내는데, 이와 같은 과정을 반복하며 학습이 이루어지게 된다. 여기서 중요한 것이 외계의 상태와 행위자의 행위 간의 사상이다. 행위자는 평가 함수(value function)와 정책 함수(policy function)를 가지고 있다. 행위자는 평가함수를 통하여 상태에 따른 정책함수를 평가하여 정책을 변경하고 또한 외계에서 주어진 강화신호에 따라 상태의 평가함수를 적절하게 다시 변경한다.

이러한 강화학습(RL) 방법 중 Sutton의 Temporal Difference(TD)에 의한 actor-critic 구조와 Watkins의 Q-learning 등이 있다. 이 방법들은 현재에 즉각적인 보상(강화신호)이 없는 경우 강화신호를 예측하여 학습한다. 그렇기 때문에 강화학습은 사람이 학습에 관여하지 않아도 제어기의 행위에 대한 환경으로부터의 평가신호를 이용하여 올바른 제어 신호를 제어기가 스스로 배우게 되고 환경이 변하더라도 사람처럼 제어기가 스스로 적응하기 때문에 다양한 모델과 환경이 존재하는 경우에 적합하다.

본 연구에서 사용한 강화학습 방법인 Q-학습 방법은 확률적 종적 계획법에 기반을 둔 자유 모델(model free) 강화 학습법의 하나로서 개발되었다. 이 방법은 마코비언 환경(markovian environment) 하에서 학습능력을 가진 시스템이 최적으로 행동할 수 있도록 해준다. 여기서 $T(s,a,s')$를 현재 상태 s에서 행위 a를 수행할 때 현재상태 s와 행위 a 사이의 관계로부터 현재상태 s가 다음 상태 s'로 변하게 될 상태전이확률이라 하고, 이 때 각 상태와 행위에 대한 보답(reward)을 $r(s,a)$라 한다. 일반적인 강화학습은 시간축에 걸쳐 얻어지는 보답의 합을 극대화하는 행위들의 정책(policy)을 찾는 것으로 정의된다. 이러한 f는 s로부터 a로의 단순한 대입을 의미한다. 여기서 보답들의 합을 다음과 같이 정의한다.

$$\sum_{n=0}^{\infty} \gamma^n r_{p+n} \quad (10)$$

상태를 감지하고 행동하는 1 반복 공정(cyclic process)을 1 스탭(step)으로 정의하고 시작상태로부터 출발하여 목표상태로 도달할 때까지의 과정을 1 반복(iteration)이라 정의한다면, r_p는 상태 s로부터 출발하여 행동정책 f를 따라 진행할 경우 어떤 스텝 p에서 받을 보답이라고 정의된다. γ는 시간 축에 따라 감소하는 감쇠상수이며, 미래의 보답이 행동책략에 얼마만큼의 영향을 끼칠 것인가를 정하기 위해 사용된다. 대개는 1 이하의 값으로 정의된다. 이 때 상태전이확률들과 각 상태전이에 대한 보답분포(reward distribution)를 미리 알 수 있다면 잘 알려진 Dynamic Programming에 의해 최적의 행동정책을 구할 수 있을 것이다. 그러나 이러한 정보를 미리 알 수 없으므로 Wakins는 통계적으로 적절한 행위를 배울 수 있는 Q-학습을 개발하게되었다.

$Q(s,a)$는 어떤 상태 s에서 행동 a를 취하고 이후에 최적의 행동정책 f를 따르기 위한 응답값 혹은 행위값이라 하고 다음과 같이 정의한다.

$$Q(s,a) = r(s,a) + \gamma \sum_{s' \in s} T(s,a,s') \max Q(s',a') \quad (11)$$

Watkins는 초기에 T와 r에 대해 아는 바가 없으므로 최적의 Q값으로 점증적으로 접근해가기 위해 온라인으로 Q값을 산정함으로써 Q값을 구하고자 하였다. 이러한 Q값의 갱신은 다음과 같다.

$$Q(s,a) = aQ(s,a) + (1-a)(r(s,a) r\max Q(s',a')) \quad (12)$$

여기서 보답 r은 상태 s에서 행위 a를 수행하는 것에 대한 실질적인 보상값이고 s'는 다음상태를, a는 0과 1사이의 값으로 학습속도를 각각 나타낸다. 여기서 Q값이 수렴함에 따라 최종적인 수렴상태에서 우변의 현재상태의 Q값과 좌변의 현재상태 Q값이 같으므로 위의 Q값 산정 방법은 다음과 같이 풀어 쓸 수 있다.

$$Q(s,a) = r(s,a), \max Q(s',a') \quad (13)$$

위 식은 Q-학습이 현재상태에서 다음상태로 전이

될 확률이 1이라는 가정 하에서 이루어진 방법임을 의미한다. 즉, 동일한 상태에서 특정 행위에 의해 나타날 수 있는 상태는 오직 하나라는 것을 의미하며, 현재 상태에서 어떤 행위에 의해 다음 상태로 전이될 확률이 1인 환경에서 전체 행위값의 합을 최대화시키고자 하는 응용분야에 적용될 수 있다는 것을 의미한다.

이러한 Q-학습을 이용하기 위해 이산된 환경의 상태 집합을 S, 행동 집합을 A라고 놓았을 때 기본적인 Q-학습 알고리즘은 다음과 같다.

① 모든 환경 s와 행위 a에 대하여 $Q(s,a)$를 임의의 값(일반적으로 0)으로 초기화한다.
② 현재의 환경 s를 인식한다.
③ 환경-행위 규칙에 따라 행위 a를 선택한다.
④ 주어진 환경에서 행위 a를 수행하고, 다음 환경을 s', 즉각적인 보상을 r로 놓는다.
⑤ s, a, s' 그리고 r로부터 환경-행위 규칙을 갱신한다.

$$\triangle Q_\pi(s_t, a_t) = a_t[R(s_t, a_t) + \gamma \max_{a' \in A} Q_\pi(s_{t+1}, a') - Q_\pi(s_t, a_t)]$$

a_t : 학습률, γ : 0과 1 사이의 감쇠 계수

⑥ ②단계로 돌아간다.

그림 1과 그림 2는 본 연구에서 적용한 강화 학습의 구조와 PI 제어 알고리즘을 결합한 개요도를 보여준다. 그림 1은 강화학습의 구조를, 그림 2는 결합도를 나타낸다. 위에서 살펴본 바와 같이 강화

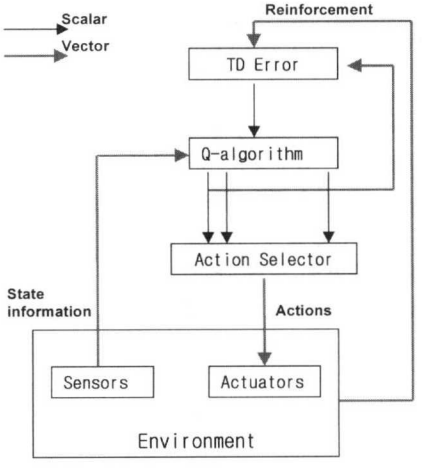


그림 1. Reinforcement learning architecture

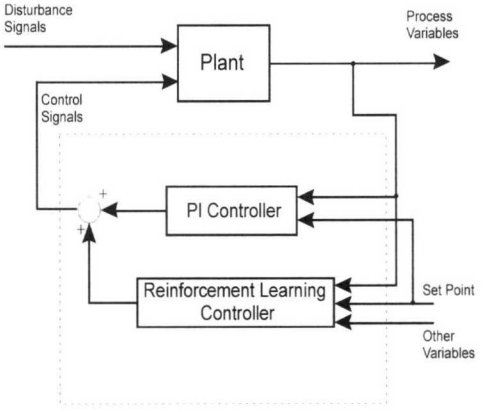


그림 2. Combination of RL and PI controller

학습은 두 가지의 구성 요소를 통해서 온라인 학습이 이루어진다. 행위자는 환경에서 주어진 상태에 따라 적절한 행위(action)를 환경에 행하며, 환경은 주어진 행위에 따라 변화된 상태와 행위가 적절했는가에 대한 판단을 강화신호로 PI 제어기의 출력을 보상하여 행위자에게 보내는데, 이와 같은 과정을 반복하며 학습이 이루어지게 된다.

Ⅲ. 실험 장치

1. 시험 주택

건물의 냉·난방 부하, 냉·난방 설비의 효율, 열환경, 에너지 절약, 건물 구조체의 heat transfer, Wall thermal mass effects, HVAC control, Access floor control 등에 관한 종합적인 실험을 수행할 수 있도록 인공 기후 실험동 내에 시험체 건물을 건립하였다.

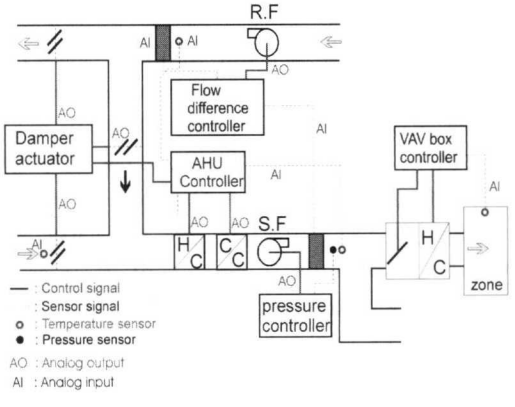


그림 3. Configuration of VAV AHU

그림 3은 시험 주택의 비온돌 실험실과 환경 실험실에 설치된 가변풍량(VAV)방식의 공조 자동제어시스템의 구성을 보여준다. 설치된 공조 시스템은 외기 및 실내 조건에 따라서 냉·난방을 1개의 공조기로 운전하도록 설계되었다. 급기 및 환기 송풍기는 가변속(VVVF) 제어가 가능하므로 경제성 및 에너지 절약 평가를 수행할 수 있고, 에너지 절약 효과와 건물의 내부 부하 변동에 따른 각 실내로 공급되는 공기량을 제어하기 위해서 VAV(Variable Air Volume) 박스를 설치하였다. 적용 건물의 모든 설비에 대한 운전 및 자료 수집은 자동 제어시스템에서 수행된다.

2. 시스템 구현

공조 시스템의 감시 운영 제어는 주컴퓨터에서의 감시 제어(Supervisory control)와 현장 제어(Local loop control)에서 이루어진다.

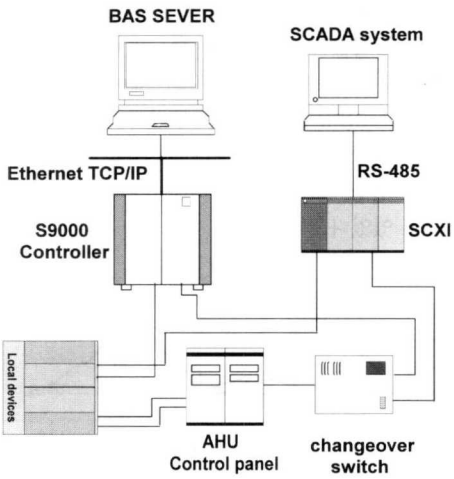


그림 4. System realization

그림 4는 시험 주택의 공조 시스템의 자동 운전을 위한 감시 운영 제어시스템의 구성을 보여준다. 그림에서 보듯이 시스템의 구성은 기존의 감시 운영 제어시스템과 제어 알고리즘 성능 실험용 감시 운영 제어시스템을 독립적으로 구성하였다. 기존의 감시 운영 제어시스템은 주컴퓨터의 감시 제어와 현장 제어기는 Ethernet TCP/IP를 이용한 데이터 인터페이스를 통해서 실시간 데이터 감시 및 운영 제어를 수행하지만, 실제 다양한 제어 알고리즘에 대한 성능 실험에는 한계가 있어 감시 운영 제어시스템과 독립적인 시스템의 데이터 인터페이스를 구성하여 자동 제어를 수행하여 제어 알고리즘 개발 및 적용 그리고 제어기 실증 실험을 통한 성능 특성을 비교·분석할 수 있는 감시 운영 제어시스템을 구현하였다.

Ⅳ. 실험 결과

강화학습(RL) 제어기의 실증 실험을 통한 제어 성능 특성을 비교·분석하기 위해서 시험체 건물 내의 가변풍량 공조기(VAV AHU)를 사용하여 기존의 PID 제어기와의 성능 실험을 수행하였다. 성능 실험에서는 전체 시스템을 대상으로 적용하기에 앞서 실제 시스템에 대한 적용 가능성을 조사하기 위해서 급기 온도 제어용 난방 코일의 제어 성능 실험을 수행하였다. 실험 조건은 외기 온도를 -1℃ ~ 0℃, 시험 건물의 혼합공기 온도 및 급기 온도의 변화 범위를 22℃ < Tma(혼합공기 온도) < 28℃, 33℃ < Tsa(급기 온도) <43℃ 범위로 하여 시스템을 운전하여 제어기 성능 실험을 수행하였다. 성능 실험을 하기 전에 시스템이 보다 안정되고 정밀한 제어를 수행하도록 Ziegler - Nichols[3]의 동조 방법을 이용하여 여러 형태의 루프 성능을 시험하여 최적화한 PI 제어기의 제어 변수 즉, 비례요소 Kp, 적분요소 Ki를 결정하여 사용하였다. PI 제어기에 강화학습 제어기를 추가로 연결하여 비례요소 Kp= 1.9, 적분요소 Ki = 7.5인 PI 제어기를 사용하여 난방 코일을 제어하고, 강화학습 제어기의 출력보상 제어 신호를 사용하여 제어 성능 실험을 수행하여 제어기를 설계하였다.

본 연구에서 사용한 강화학습 방법은 동적 프로그래밍으로 구체화시켜 환경에 대한 충분한 지식이 없어도 온라인으로 강화신호에 의하여 주어진 환경에 따른 시행착오를 통해서 최적의 행동을 학습할 수 있는 방법인 Q-learning 방법을 적용하였다. 강화학습의 행위자 입력변수는 혼합공기 온도(Tma), 급기 온도(Tsa) 그리고 설정치(SP)를 사용하였고, 강화학습의 출력보상제어 신호로 7개 이산출력신호 즉, [-2, -1, -0.2, 0, 0.2, 1, 2]로 정하여 PI 제어기의 출력 제어신호에 더하여 보낸다. 3개의 각 입력변수를 8개의 간격 범위로 분할하여 3차원 입력 공간을 7^3(343)개로 정하였다. 각 입력 공간에는 7개의 이산 출력 신호가 행위값(Q-value)에 저장된다. 강화학습 방정식은 다음과 같다.

$$R(t) = (T_{sa}(t)^* - T_{sa}(t))^2 \quad (14)$$

Tsa(t)* : 설정치, Tsa(t) : 관측치

그림 5는 강화학습(RL) 제어기의 온라인 학습을 위해서 PI 제어기에 강화학습 제어기를 추가하여 입력 상태에 따른 랜덤한 임의의 출력 신호를 더하여 보냈을 경우 난방 코일의 제어 성능 특성을 보여준다. 난방 코일이 운전되고 있을 때 이산 출력신호값 [-2, -1, -0.2, 0, 0.2, 1, 2]을 랜덤하게 발생시켜서 PI 제어기의 출력 제어신호와 더해지면서 일어나는 환경 변화에 대해 강화학습 제어기는 온라인 학습 과정을 통해서 행위값에 저장된다.

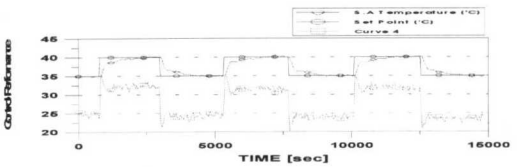


그림 5. Control performance when the selected action of RL agent is added to PI controller

그림 6, 그림 7은 비례요소 Kp, 적분요소 Ki가 서로 다른 PI 제어기에 추가로 학습이 완료된 강화학습 제어기를 적용하였을 경우 난방 코일의 제어 성능 특성을 보여준다.

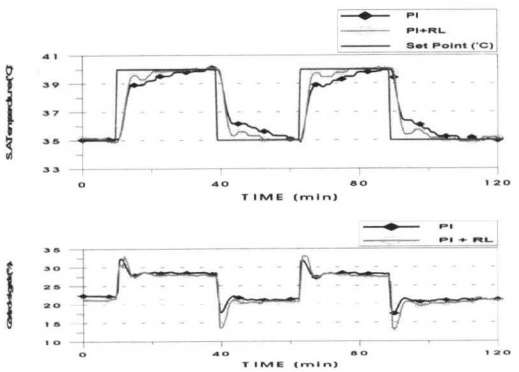


그림 6. Control performance with combined PI and RL controller(optimal gain)

그림 6은 동조 과정을 통해 오버슈트가 거의 없는 최적 제어 변수 즉, Kp= 1.9, Ki= 7.5인 PI 제어기를 사용하였을 경우와 PI 제어기에 추가로 학습이 완료된 강화학습 제어기를 적용하였을 경우, 난방 코일의 제어 성능 특성을 보여준다. 그림에서 보듯이 학습이 완료된 최적화 제어기 즉, 강화학습 제어기를 사용하였을 경우 외부 환경의 변화에 따른 급기온도 설정값 변화 시 PI 제어기보다 정상상태 오차를 상당히 감소시키고, 빠른 응답성을 가지는 등 제어기의 성능이 개선되는 것을 알 수 있다.

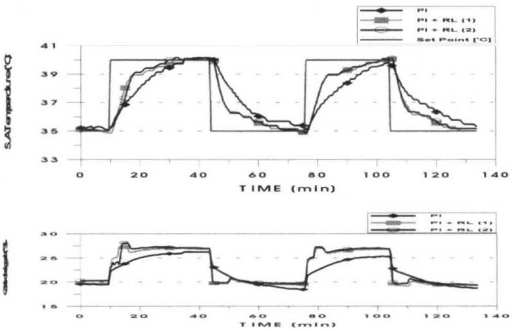


그림 7. Control performance with combined PI and RL controller(random gain)

그림 7은 상승 시간을 다소 증가시킨 제어 변수 Kp= 1, Ki= 4인 PI 제어기를 사용하였을 경우와 PI 및 강화학습 제어기를 적용하였을 경우 난방 코일의 제어 성능 특성을 보여준다. 환경 변화에 따른 급기온도 설정값 변화 시 PI 제어기보다 빠른 응답성 및 정상상태 오차 감소 등의 제어 성능이 개선됨을 알 수 있다. 또한 학습 루틴의 회수가 증가할수록 학습 오류에 대한 제어기의 성능 저하를 줄여줌으로서 난방 코일의 최적 제어가 가능하다는 것을 알 수 있었다.

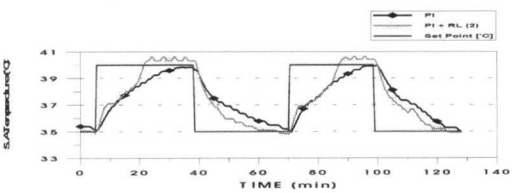


그림 8. Performance comparison of controller as learning error

그림 8은 학습 오류에 따른 강화학습 제어기에 의한 난방 코일의 제어 성능 특성을 보여준다. 학습 오류를 가진 강화학습 제어기와 PI 제어기만 사용하였을 경우 성능을 비교하면 응답성은 다소 개선되지만, 학습용 입력 변수의 경계치 부근에서의 수렴 문제로 인해 난방 코일의 출력 제어신호에 동요현상이 발생해서 정상상태 오차가 PI 제어기만을 사용하였을 경우보다 상당히 증가하는 것을 알 수 있다.

본 실증 실험을 통해 강화학습 제어기를 실제 시스템과 연계하여 제어기로 적용하기 위해서는 먼저 충분한 학습과 환경 변화에 따른 입력 변수들의 적

절한 선정과 변수별 경계 범위, 그리고 시스템별 출력 보상 신호값을 효율적으로 설정해야한다. 그렇지 않으면 변수의 경계 조건 혹은 학습이 이루어지지 않은 곳에서 학습 오류에 따른 동요 현상으로 인해 수렴이 되지 않아 제어기 성능이 저하될 수 있다. 그러나 적절한 입력 변수 및 경계 범위, 출력 신호 그리고 학습 회수가 정해진다면 시스템별 사용 목적에 따라 최적의 성능을 가진 제어기를 개발할 수 있을 것이라고 본다.

V. 결론 및 향후 연구

본 논문에서는 PI 제어기 그리고 강화학습(RL) 제어기를 사용하여 실제 공조시스템에 적용한 후 제어 성능 특성을 다음과 같은 조건 하에서 비교·분석하였다.

(1) 설정치(S.P)의 응답 시간 및 정상상태 오차,

(2) 피드백 제어변수와 다른 제어변수들을 포함하였을 경우의 제어 성능,

(3) 강화학습 진행에 따른 제어 성능의 차이 등이다.

그 결과로서 첫째, 학습 루틴 회수가 증가할수록 강화학습 제어기는 외부 환경의 변화에 따른 급기 온도 설정값 변화 시 PI 제어기의 출력을 증가 혹은 감소시켜 설정치와 급기 온도 사이의 정상상태 오차를 크게 감소시키지는 않지만, 다소 빠른 응답성을 가지고 설정치 추종 제어 시 난방 코일의 출력 제어신호가 상당히 개선되는 것을 알 수 있다.

둘째, 강화학습 제어기를 실제 시스템과 연계하여 제어기로 적용하기 위해서는 먼저 충분한 학습과 환경 변화에 따른 입력 변수들의 적절한 선정과 변수별 간격 범위 그리고 시스템별 출력 보상 제어 신호값을 효율적으로 설정해야 한다. 그렇지 않으면 변수의 경계 구간 조건 혹은 학습이 이루어지지 않은 곳이나 학습 오류에 따른 동요 현상의 발생으로 수렴되지 않아 제어기 성능이 저하될 수 있다. 그러나 적절한 입력 변수, 보상 출력 신호 그리고 학습 회수의 범위가 정해진다면 시스템별 사용 목적에 따라 최적의 성능을 가진 제어기가 될 수 있다.

셋째, 강화학습 제어 알고리즘을 적용한 제어기는 환경 변화에 빠른 적응성으로 제어기의 성능을 개선시킬 수 있다.

따라서 향후 연구에서는 계속적인 실증 실험으로 보다 많은 어플리케이션과 연계하여 범용성 있는 온라인 학습형 강화학습 제어기를 개발하고자 한다.

참 고 문 헌

[1] 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.

[2] Åström, K. J. and Hägglund, T., 1995, PID controllers: Theory, design and tuning, Research-Triangle Park, NC: Instrument Society of America.

[3] Hang, C. C. and Åström, K. J. and Ho, W. K., Ziegler-Nichols tuning formula., IEE Proc. D, Vol. 138, No.2, pp 111-118.

[4] Ziegler, J. G. and Nichols, N. B. "Optimum settings for automatic controllers", Trans. ASME, 1942, 65, pp. 433-444

[5] Hang, C. C., LIM, C. C. and Soon, S. H. "A new PID auto-tunner design based on correlation technique", Proc. 2nd Multinational Instrumentation Conf., China, 1986

[6] Hang, C. C. and Åström, K. J. "Refinements of the Ziegler Nichols tunning formula for PID auto-tunners", Proc. ISA Conf., USA

[7] Åström, K. J., Hang, C. C., Persson, P., Ho, W. K. "Towards Intelligent PID Control", 1991, International Federation of Automatic Control

[8] Åström, K. J. and C. C. Hang and P. Persson (1988), "Heuristics for assessment of PID control with Ziegler-Nichols tuning", Automatic Control, Lund Institute of Technology, Lund, Sweden

[9] Åström, K. J., and Hagglund, T, "Automatic tuning of simple regulators with specifications on phase and amplitude margins", Automatica, 1984, 20, pp. 645-651

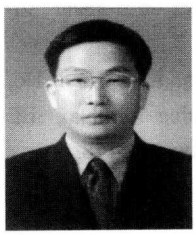
[10] Sutton, R. S. 1988. Learning to predict by the methods of TD(temporal differences). Machine Learn. 3, 9-44.

[11] Anderson, C. W., 1993, Q-learning with hidden-unit restarting, In Advances in Neural information processing system 5, pp81-88.

[12] Barto. A. G. and Bradtke, S. J. and Singh, S. P., 1995, Learning to act using real-time dynamic programming, Artificial Intelligence 72, pp 81-138.

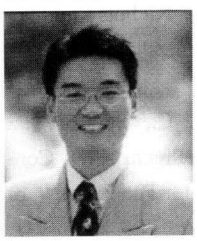
[13] Watkins, C. J. and DAYAN, P., 1992, Q-learning, Machine Learning 8,pp 279-292.

유 승 선(Seung-sun Yoo) 정회원



1988년 2월 : 한남대학교 전자계산학과 졸업
1994년 2월 : 한남대학교 전자계산공학과 석사
1997년 3월~현재 : 전북대학교 영상정보공학과 박사과정
1988년 6월 ~ 2001년 5월 : 한국에너지기술연구소 연구원
2001년 6월 ~ 현재 : (주)비맥 기술이사
<주관심 분야> Neural Network, Image Processing

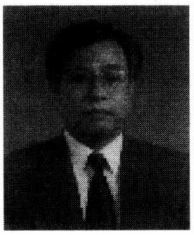
김 문 성(Moon-seong Kim) 정회원



1993년 2월 : 한남대학교 물리학과 졸업
1995년 2월 : 한남대학교 컴퓨터공학과 석사
1996년 3월~현재 : 전북대학교 컴퓨터공학과 박사과정
1997년 3월~현재 : 대원과학대학 컴퓨터정보통신과 조교수
<주관심 분야> Image Processing, Control Performance, Computer Simulation

소 정 훈(Jung-Hoon So) 정회원
1993년 2월 : 영남대학교 전기공학과 졸업
1995년 2월 : 영남대학교 전기공학과 석사
1995년 6월~현재 : 한국에너지기술연구소 연구원

곽 훈 성(Hoon-sung Kwak) 정회원



1971년 2월 : 전북대학교 전자공학과 졸업
1975년 2월 : 전북대학교 전자공학과 석사
1980년 2월 : 전북대학교 전자공학과 박사
1991년 1월~1993년 5월 : 미국 텍사스 대학교 교환교수
1993년 10월 ~ 1994년 10월 : 전북대학교 전자계산 소장
1996년 10월 ~ 1998년 11월 : 전북대학교 영상특성화 사업단 단장
1978년 4월 ~ 현재 : 전북대학교 전자정보공학부 정교수
<주관심 분야> 멀티미디어 통신, 영상처리, 영상압축