

시스템 출력의 퍼지추론결과를 이용한 제어기의 성능 개선

Performance Improvement of Controller using Fuzzy Inference Results of System Output

이 우 영*, 최 홍 문**
Woo-Young Lee*, Heung-Moon Choi**

요 약

퍼지 제어기에 신경회로망을 병렬로 연결시키므로 제어성능 향상을 위해 필요했던 소속함수의 미세조정 과정이 학습으로 대신되게하는 제어기 구조를 제안하였다. 신경회로망의 학습은 오차 역전파 알고리듬에 의해 수행되고 퍼지 제어기의 출력이 학습에 사용되는 오차량으로 사용된다.

따라서 본 제어기는 전문가의 경험과 지식을 제어기 설계에 이용할 수 있고, 별도의 학습과정 없이 제어과정 중에서 신경회로망 제어기가 학습되어 초기의 제어특성이 개선되어지는 특성이 있다. 그리고 본 구성에서 퍼지 제어기는 사용된 규칙에 의해 형성되는 위상평면상의 슬라이딩 면으로 필요한 제어특성과 신경회로망의 학습기준을 제시하는 한편 신경회로망이 학습되기전 제어 시스템의 제어특성이 안정되도록 하며, 신경회로망은 시스템의 상태궤적이 퍼지제어기에 의해 형성된 슬라이딩 면을 가능한한 균사하게 추종하도록 학습되어져 위상평면상 임의의 위치에 있는 시스템의 상태가 슬라이딩 면을 따라 안정점에 도달하도록 하게 한다.

ABSTRACT

The new architecture that fuzzy logic control(FLC) with difficulties for tuning membership function (MF) is parallel with neural networks(NN) to be learned from the output of FLC is proposed.

Therefore proposed scheme has the characteristics to utilize the expert knowledge in design process, to be learned during the operation without any learning mode.

In this architecture, the function of the FLC is to supply the sliding surface which is constructed on the phase plane by rule base for giving the desired control characteristics and learning criterion of NN and the stabilization of the control performance before NN is learned. The function of the NN is to let the system trajectory be tracked to the sliding surface and reached to the stable point.

I. 서 론

대상시스템이 비선형적이고 불명확하여 기존의 해석적 모델의 제어방식으로는 다루기 어려우나 전문가의 경험에 의해서는 용이하게 제어되고 있는 대상에 대해서 퍼지제어기가 잘 적용되고 있는 경우를 자주 볼 수 있다. 이

* 한국전기연구소 전력기기연구부

** 경북대학교 전자공학과

는 언어적 형태로 표현되는 정성적 제어규칙이 유용한 정보임에도 불구하고 적절히 사용되지 못하다가 퍼지시스템에 의해 정량화되어 효과적으로 제어에 사용되어짐을 나타낸다. 그러나 이러한 퍼지 제어기의 성능 향상을 위해서는 소속함수의 미세조정이 필요한데 이를 위한 체계적 방법이 없어서 퍼지제어기는 폭넓게 사용되지 못하고 있다.

관련된 많은 연구들이 체계화된 설계방법을 찾는데 관심을 보여오고 있으며, 그 중 대표적인 것으로 LMS 알고리듬[1][2][3]과 같은 최적화 방법들에 의해 소속함수의 미세조정을 위한 변수조정을 하는 것과 신경회로망[4][5][6]을 사용하는 방법들이 많이 발표되고 있다.

그리고 신경회로망 역시 임의의 비선형함수를 구현할 수 있는 특성으로 인해 제어기로서 많은 관심의 대상이되고 있으며 특히 학습기능을 통해 제어기로 사용되어지는 것이 특징이다. 그러나 신경회로망에서는 제어기 구현을 위한 학습정보를 어떻게 얻느냐가 큰 과제로 지금까지의 주된 방법은 기존의 제어기로부터 학습데이터를 얻는 방법, 제어에 사용되는 대상시스템의 역모델을 구현하기 위해 대상 시스템의 입출력 특성으로부터 학습데이터를 얻는 방법 그리고 대상시스템을 통해 시스템의 오차량이 역전파된 것으로 학습데이터를 얻는 방법등이 일반적으로 알려져 오고 있다.[7][8][9]

첫번째 방법은 이미 사용되고 있는 제어기를 신경회로망으로 구현하는 것으로 학습대상 제어기의 존재가 전제되어야하며 두번째 방법에는 일반적으로 역모델이 항상 존재하는 것은 아니라는 점과 또한 신경회로망의 오차량을 최소화하는 것이 시스템 오차량의 최소화를 보장하지는 못하는 문제점이 있다.[7] 그리고 세번째 방법은 대상 시스템의 Jacobian특성을 알아야하거나 신경회로망으로 시스템을 모델링하는 것이 필요하고 또 시간과 동작조건에 따라 시스템의 특성이 변화될 수 있기 때문에 시스템의 모델이 주기적으로 갱신되어져야하는데 제어과정중에서 이를 행하는 것은 용이한 일이 아니라는 문제점이 있다.

신경회로망과 퍼지시스템이 혼합된 형태의 제어기로는 퍼지추론과정을 신경회로망 형태로 표현하여 신경회로망의 학습기능을 퍼지시스템에 도입할 수 있도록 한 것으로 소속함수 조정에 학습을 이용코자하는 것이 주류를 이루고 있다.[5][10] 이는 전문가의 지식과 경험이 제어기 설계에 사용될 수 있어 신경회로망의 연결가중치의 초기치 설정에는 유리한 점이 있으나 구조적으로는 신경회로망에 속하게 되기 때문에 학습정보의 확보는 앞의 신경회로망과 동일한 문제로 남는다.

본 연구에서는 경험적 지식을 기반으로 주관적 방법에 의해 구성되는 퍼지 제어기와, 제어기 구성에 역모델이나 순방향 모델단계 혹은 제어대상 시스템의 Jacobian 특성 등이 필요했던 신경회로망을 혼합하여 구성하므로 전문가의 경험과 지식을 제어기 설계에 이용할 수 있게하고, 별도의 학습과정 없이 제어과정 중에서 신경회로망 제어기가 학습되게 할 수 있어 각각의 특성들을 상호보완할 수 있는 제어시스템을 제안하였다. 이 제안된 제어기는 퍼지제어기의 소속함수들 조정에 소요되는 많은 시간과 노력을 줄일 수 있고, 제어과정에서 퍼지 제어기의 출력을 신경회로망의 학습 오차량으로 받아들여 학습과 함께 제어성능이 개선되도록 하기 때문에 신경회로망의 학습을 위한 학습모드나 학습데이터를 구하기 위한 별도의 과정이 필요치않다. 또 상태공간 분할특성에 의존하는 기존 퍼지 제어기의 제어성능 한계성을 신경회로망의 도움으로 보완 할 수 있는 등의 특성을 지니게 된다.

II. 제어기 구성

2.1 제어기 구성과 동작

본 연구에서는 퍼지제어기의 제어성능 개선을 위한 소속함수의 조정과정을 신경회로망의 학습과정으로 대치할 수 있는 그림 1과 같은 제어기 구성을 제안하며 이는 기존의 퍼지 제어기 구성에서 퍼지 제어기와 병렬로 신경회로망을 연결한 형태이다.

제안된 제어기는 초기상태, 학습상태, 그리고 학습 완료상태의 세가지 상태로 구분할 수 있는데 초기상태는 신경회로망이 학습되기 전의 제어기 상태로서 기존의 경험적 방법에 의해 구성된 퍼지제어기가 제어 대상시스템에 대한 제어입력량을 전적으로 결정하게되는 주 제어기 역활을 하게되고 신경회로망의 출력량은 거의 영이 되는 상태이다. 이때 페루프 시스템의 동작은 상태변수들의 값이 허용된 영역안에 존재한다는 측면에서는 안정한 시스템으로 볼 수 있으나 퍼지제어기의 소속함수가 조정되지 않은 상태이기 때문에 성능면에서는 개선의 여지가 많은

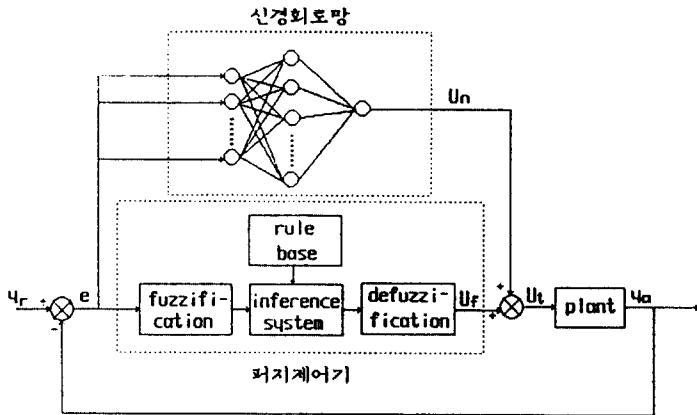


그림 1. 제안된 제어기 구성도

Fig. 1 Proposed controller schematic diagram

제어특성을 나타낸다. 학습상태에서는 신경회로망의 퍼지제어기의 출력에 의해 학습되어지는데 퍼지제어기의 출력 $u_f(t)$ 의 방향에 따라 신경회로망 출력 $u_n(t)$ 을 증가시킬 것인지 감소시킬 것인지를 결정하게 된다. 신경회로망의 학습이 진행되는 동안 제어입력량 u_i 에서 신경회로망 출력 u_n 이 차지하는 비중이 점점 커지게 되며 학습완료 상태에서는 신경회로망이 제어입력량을 주도적으로 결정하게 되어지게 된다. 그리고 학습된 신경회로망의 제어 특성은 퍼지제어기에 의해 주어지는 케적 즉 퍼지제어기의 출력량이 영인 케적을 따라 시스템 상태벡터의 천이케적이 상태공간을 이동하여 제어목적 지점까지 가게한다.

표 1. 퍼지 제어기의 제어규칙 테이블 예
Table 1. The example of rule table of FLC

ZR	PS	PM	PB	PB
NS	ZR	PS	PM	PB
NM	NS	ZR*	PS	PM
NB	NM	NS	ZR	PS
NB	NB	NM	NS	ZR

PB(NB) : Positive(Negative) Big

PM(NM) : Positive(Negative) Medium

PS(NS) : Positive(Negative) Small ZR : Zero

표1은 퍼지 제어기에 사용되는 일반적 형태의 규칙테이블로 주대각선이 ZR로 구성되어있으며 시스템의 상태 변수가 이 주대각선상에 위치하는 경우 퍼지 제어기의 출력이 영이 되며 대각선 위쪽으로 위치하면 양의 출력이 아래쪽으로 위치하면 음의 출력이 됨을 뜻하고 ZR*는 제어목표가되는 안정점(stable point)를 나타낸다. 규칙 테이블내 상태공간 임의의 곳에 위치된 상태변수가 안정점으로 이동하는 경로는 무수히 많이 존재하며 이를 경로는 소속함수의 조정에 따라 달라지게 된다.

신경회로망이 학습되어질수록 임의의 위치에 설정된 초기 상태변수는 가장 빠른시간 내에 퍼지 제어기의 출력이 영이되는 주 대각선상으로 이동하게 되고 주대각선을 따라 안정점으로 이동하게 된다.

이는 슬라이딩 모드 제어(sliding mode control)에서 상태 변수 오차가 슬라이딩 면(sliding surface)을 따라 안정점으로 이동하는 것과 유사하며[12] 규칙 테이블의 주 대각선이 슬라이딩 면의 역할을 하게 된다.

제안된 제어기 구성에서 신경회로망의 역할은 슬라이딩 면에 위치한 임의의 상태에서 안정점에 도달할 때까지의 퍼지 제어기 출력량 합이 최소가 되도록 하는 것으로 식 (1)와 같이 표시된 성능지수함수가 최소가 되도록 신경회로망의 연결 가중치를 조정해 나가게 된다.

$$J = \frac{\sum_i^N |u_f|}{N} \quad (1)$$

단 N은 안정점까지의 샘플링 갯수를 나타낸다.

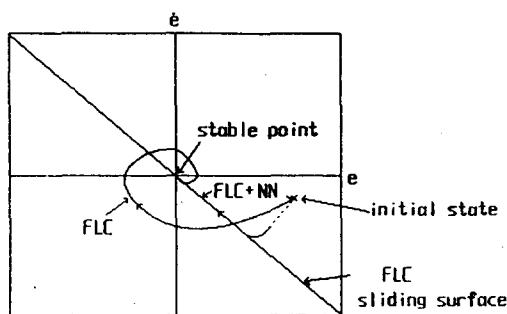


그림 2. 학습전과 학습후의 제어특성 비교

Fig. 2 Comparision of control results between before and after learning

그림 2에는 FLC로 표시된 신경회로망이 학습되기 전 퍼지제어기 만으로 제어된 특성과 신경회로망이 학습된 후의 FLC + NN으로 표시된 제어특성을 나타내었으며 제어대상 시스템으로 하여금 퍼지 제어기가 제시한 슬라이딩 면을 잘 추종해 가도록 하는 신경회로망의 역할을 보여준다.

퍼지 제어기의 역할은 학습 초기의 제어성능이 안정될 수 있게 하고 그림 2와 같이 신경회로망의 학습 기준에 가담하게 된다. 그림 3에서는 #1과 #2로 표기된 두 가지의 퍼지제어기에 의해 주어진 슬라이딩 면과 학습후의 제어결과를 통해 제안된 제어기 구성에서 퍼지제어기의 역할을 알 수 있게 한다.

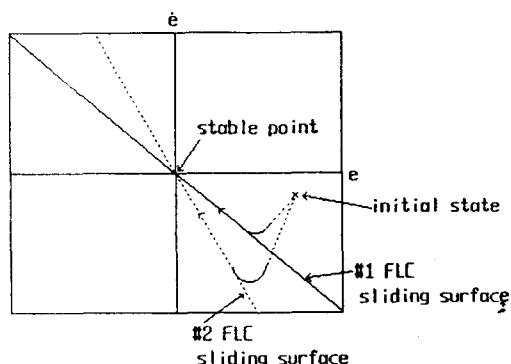


그림 3. FLC의 슬라이딩 면에 따른 제어결과

Fig. 3 Control results according to sliding surface of FLC

2.2 시스템 출력 오차에 의한 신경회로망의 학습

그림 4는 제안된 제어기의 학습시 고려될 수 있는 두 가지 경우의 오차량을 나타낸 것으로 신경회로망의 학습 기준치 u_d 가 존재하는 경우의 오차량 e_u 와 시스템 출력 오차량 e_y 에 의한 오차량 e_s 를 들 수 있다.

오차량 e_u 에 의한 학습은 일반적으로 신경회로망을 학습시키는데 어려움이 없으나 학습 기준치 u_d 를 알 수 있는 경우가 많지 않기 때문에 실제 측정이 가능한 시스템 출력오차 e_y 를 학습에 사용될 수 있는 방안이 필요하게 된다.

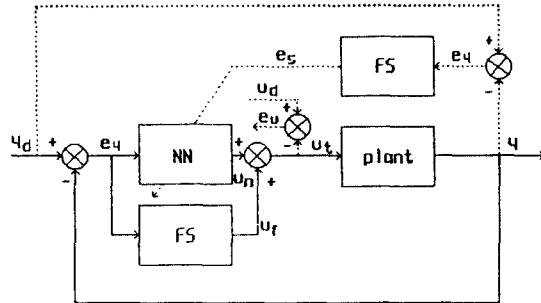


그림 4. 제안된 제어기의 학습 모드

Fig. 4 Learning mode for proposed controller

[13]에서는 식 (2)로 표현되는 등가오차 e_s 로부터 신경회로망이 학습되는 필요 충분조건이 식 (3)이 됨을 밝혔다.

$$e_s = \sum (y_d - y) \frac{\partial y}{\partial U} \quad (2)$$

$$\text{sign}(e_s) = \text{sign}(e_U) \quad (3)$$

그러나 $\frac{\partial y}{\partial U}$ 는 제어대상 시스템의 Jacobian특성을 나타내는 것으로 이를 아는 것은 용이한 일이 아니다. e_u 는 학습 기준치 u_d 로부터 알 수 있지만 이는 미지의 상태로 전제되었기 때문에 실제 시스템에 인가되는 u_t 가 u_d 를 기준으로 대소를 판단하는 것이 필요하다.

그림 5는 이차원의 오차 평면상에서 시각 ($t-m$)의 오차벡터 $E(t-m)$ 이 Δt 이후 시각($t-m+1$)에서 $E(t-m+1)$ 로 천이된 상태를 퍼지제어기에 의해 주어지는 슬라이딩 면을 기준으로 세가지의 경우에 대해 나타낸 것이다. 그리고 그림상 대각선을 중심으로 P, N으로 표시된 것은 퍼지제어기에 의해 주어진 슬라이딩 면을 경계로 위쪽은 양(P)의 값이 아래쪽은 음(N)의 값이 퍼지제어기의 출력으로 되어짐을 뜻하며 슬라이딩 면상의 상태는 안정점으로 향하는 시스템의 현상태가 퍼지제어기 규칙에 나타난 제어자의 제어특성에 적합함을 뜻한다. 천이경로 “1”은 천이전후의 퍼지제어기 출력이 동일한 부호를 가지게 되는 경우를 대표하며, “3”的 경우는 천이전과 다른 부호를 가지게 되는 경우이다. 천이경로 “2”는 시각($t-m+1$)에서의 오차벡터가 슬라이딩 면상에 존재하게 되는 경우로, 시각($t-m$)에서의 제어량이 제어대상 시스템의 오차벡터가 슬라이딩 면을 따라 가기를 원하는 제어목적에 적절함을 뜻한다.

앞서 제시된 학습가능여부를 가리기위한 부호판정의 기준은 시각($t-m+1$)의 퍼지제어기 출력이 시각($t-m$)의 값과 동일한 부호가 되는 “1”경로의 경우 이상적 제어출력 U_d 에 비해 실제 제어 출력 U_t 가 절대치로 작았던 것으로 판단하여 $e_u = (U_d - U_t)$ 의 부호가 $U_t(t-m)$ 의 절대치를 증가시키는 것으로 결정한다. 이와는 반대로 “3”的 경우는 U_d 에 비해 U_t 가 절대치로 큰 경우에 해당되며 e_u 의 부호가 $U_t(t-m)$ 의 절대치를 감소시키는 쪽으로 결정된다.

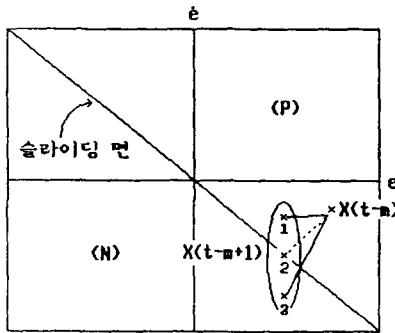


그림 5. 오차평면 상에서 오차벡터의 천이형태
Fig. 5 Transition pattern of error vector in the error space

이와같이 결정된 $sign(e_t)$ 는 시각(t-m+1)에서의 퍼지제어기 출력부호 즉 $sign(e_s)$ 와 동일하게되어 신경회로망 학습 가능을 위한 필요충분조건 식 (3)을 만족하게 된다. 따라서 퍼지제어기가 제시한 슬라이딩 면을 따라 대상 시스템의 제어궤적을 안정점에 도달시키기 위한 제어목적에 부합되는 기준 제어량이 없어도 퍼지제어기의 출력량을 신경회로망의 학습 오차량으로 삼아 신경회로망을 학습 시킴으로 초기 퍼지제어기의 성능보다 개선된 제어결과를 얻을 수 있다.

III. 수치 예

제안된 제어시스템의 성능을 확인하기 위해 퍼지 제어기 특성평가에 많이 사용되는 역진자 제어문제를 선택하였다. 역진자의 동특성은 식 (4), (5)와 같고 이는 Barto 등[14]이 사용한 것과 동일하며 마찰계수들은 무사하였다.

$$\ddot{\theta} = \frac{g\sin\theta + \cos\theta \frac{(-F - ml\dot{\theta}^2)}{(m_c + m)}}{l[\frac{4}{3} - \frac{m\cos^2\theta}{m_c + m}]} \quad (4)$$

$$\ddot{x} = \frac{F + ml[\theta^2 \sin\theta - \dot{\theta} \cos\theta]}{m_c + m} \quad (5)$$

$g = 9.8m/s^2$

중력가속도

$m_c = 1.0kg$

cart 무게

$m = 0.1kg$

pole의 무게

$l = 0.5m$

pole의 1/2길이

$-10N \leq F \leq 10N$

cart에 인가되는 힘

그리고 pole의 기울어진 각 θ , 각속도 $\dot{\theta}$, cart의 위치 x , cart의 속도 \dot{x} , 그리고 출력량 u의 변수들에 대한 퍼지 제어기의 소속함수는 그림 6의 간단한 삼각형태를 사용하였다.

사용한 신경회로망으로는 입력층 은닉층 그리고 출력층으로 구성하고 각층의 뉴런갯수는 5, 5, 1로 은닉층과 출력층이 식 (6)과 같은 비선형 특성을 가지게 하였으며 h는 기울기를 나타낸다.

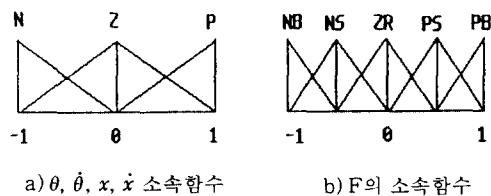


그림 6. 사용된 소속함수

Fig. 6 Membership function used in simulation

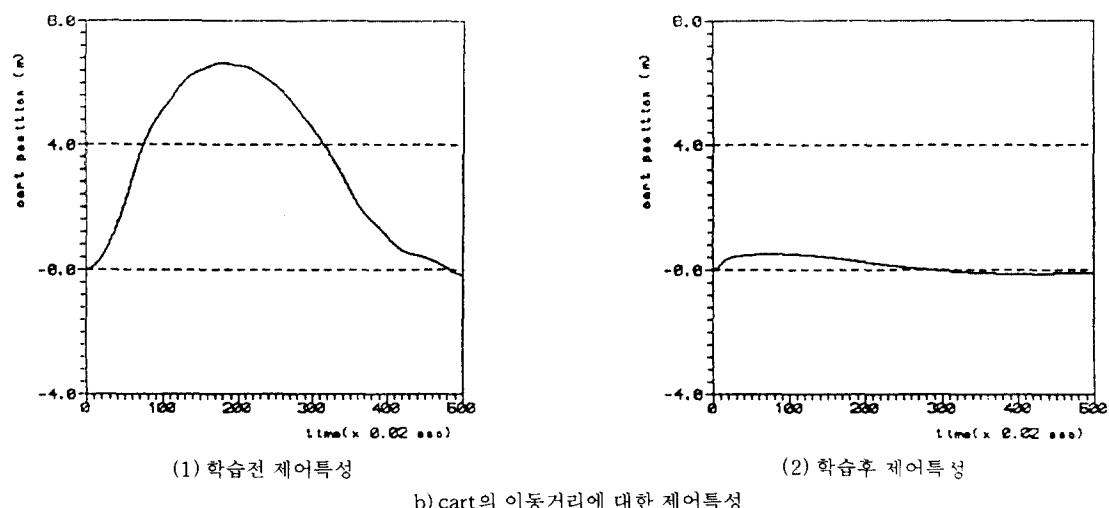
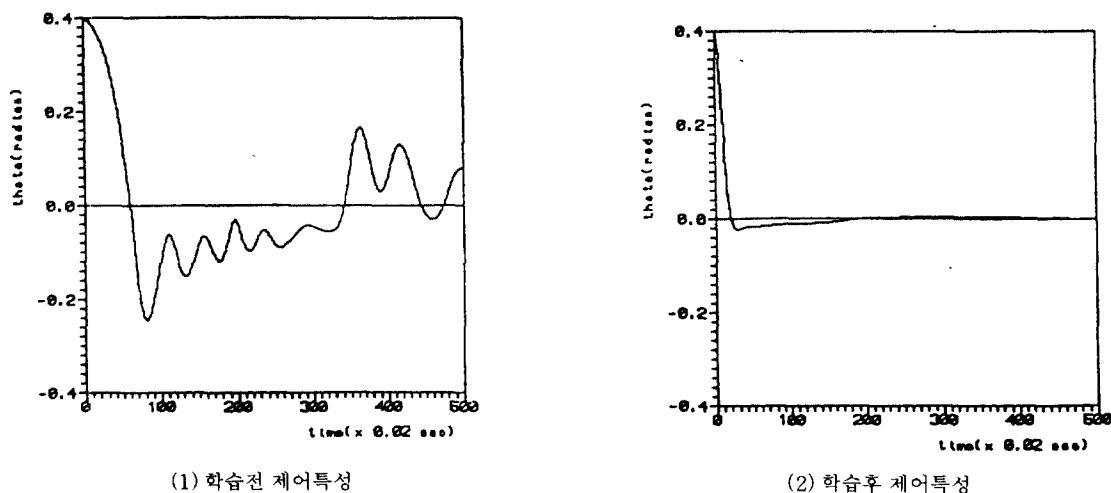


그림 7. 역진자의 제어특성 비교

Fig. 7 Comparision of control performance for inverted pendulum

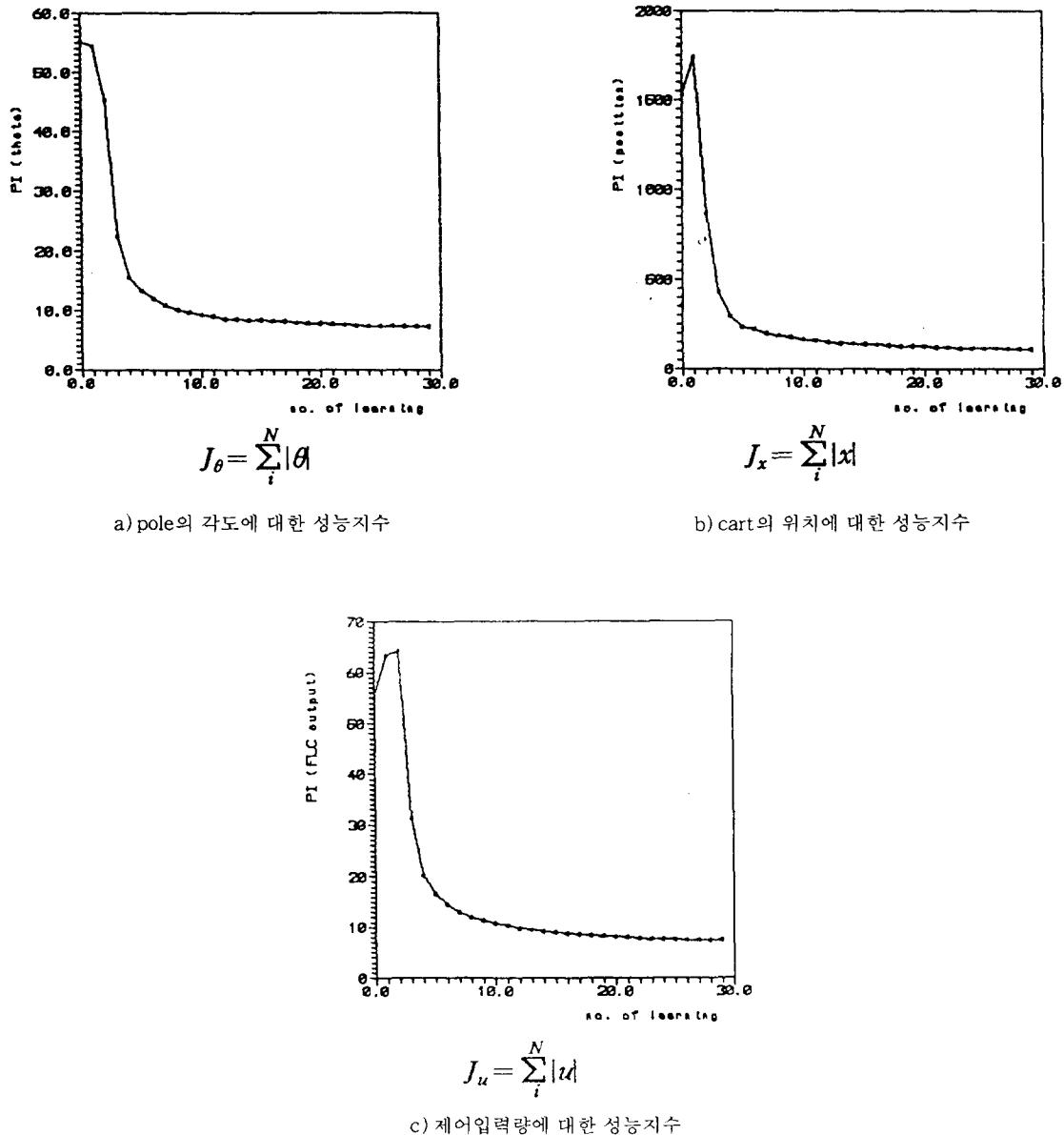


그림 8. 성능지수의 학습횟수에 따른 변화
Fig. 8 Performance index with learning number

$$f(x) = \frac{1 - \exp(-hx)}{1 + \exp(-hx)} \quad (6)$$

신경회로망의 학습에는 퍼지 제어기의 출력을 오차량으로 삼아 오차역전파 알고리듬으로 학습되게 하였다.

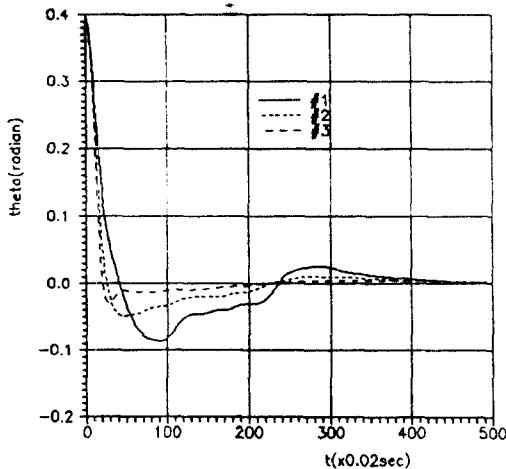
제어기 설계과정은 먼저 퍼지 제어기의 입·출력 변수와 제어규칙을 설정하고 제어성능은 좋지 않더라도 안정성이 있도록 퍼지 제어기의 입출력 변수들에 대한 환산계수(scaling factor)를 조정한 후 신경회로망을 사용하여

성능향상을 위한 학습을 수행한다. 그림 7의 학습전 상태는 폐지제어기만의 제어결과이며 제안된 제어시스템에서의 초기 제어상태에 해당하는 것으로 제어성능면에서는 향상되어야 할 여지가 많이 있음을 보여주고 있다. 그리고 그림 7에서 학습후 상태는 제안된 제어기를 300회의 학습을 시행한 후의 제어결과이며 pole의 각도와 cart의 위치로 보아 제어성능면에서 개선되었음을 볼 수 있다.

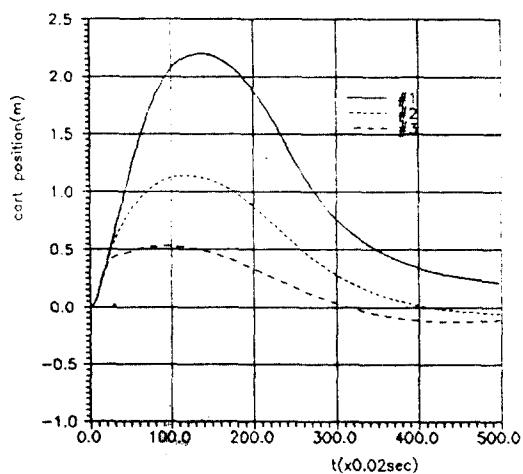
즉 그림 7에 나타난 학습전 상태는 폐지제어기만의 제어특성으로 시뮬레이션 조건을 샘플링 간격 0.02sec, 초기상태는 $\theta=0.4(\text{radian})$, $\dot{\theta}=0$, $\dot{x}=0$ 로 설정하였을 경우 10초(반복회수 500)가 경과해도 쓰러지지는 않았지만 역진자를 안정한 상태로 세우지 못하였고 또 cart의 이동거리도 5m가 넘는 제어특성이었으나 학습후는 제어시작 후 4초 정도에서 pole의 각도가 제어되고 cart의 이동거리도 0.5m내에서 이루어짐을 보였다.

그림 8은 성능지수 J의 학습회수에 따른 변화를 나타낸 것으로 학습횟수가 많아 질수록 감소되어짐을 보여준다. 즉 성능지수의 크기는 얼마만큼 빨리 안정상태에 도달하는가를 나타내는 속도로 성능지수가 작을수록 안정상태에 빨리 도달하게 됨을 뜻한다. 그러므로 초기 폐지제어기만의 제어특성으로부터 학습이 진행될수록 제어성능이 개선되는 것을 성능지수 변화특성으로부터 알 수 있다. 그리고 제어목적을 이루기위해 사용된 입력 제어량도 제어성능의 개선으로 감소되어짐을 그림 8 c)로부터 알 수 있다.

그림 9는 초기 폐지제어기에 의해 주어지는 슬라이딩 면의 사이에 대한 학습후 제어특성을 나타낸 것으로 각속도 $\dot{\theta}$ 의 소속함수의 전체집합 크기를 0.9, 1.8, 2.7(radian/sec)인 경우를 나타내었다.



a) pole의 각도에 대한 제어결과



b) cart의 위치에 대한 제어결과

그림 9. 초기 폐지제어기의 특성에 따른 학습후 제어특성

Fig. 9 Control performance according to the characteristics of FLC

IV. 결 론

신경회로망의 학습을 통해 초기 설계된 폐지제어기의 제어성능을 개선시킬 수 있는 제어기 구조를 제안하고 역진자 제어문제에 적용하여 이의 효용성을 확인하였다. 신경회로망이 학습되기전에는 전문가의 감각에 의해 설계된 폐지 제어기가 주된 제어역할을 담당하나 학습이 진행될수록 신경회로망의 참여비율이 증대되어 학습이 끝난 상태에서는 폐지제어기의 출력은 거의 영으로 되어진다.

이는 전문가의 제어에 관한 정성적 지식이 초기 제어기 구성에 적용되어져 언어적 형태로 표현되어지기 때문에

제어의 근간을 이루는 규칙이 쉽게 이해되어질 수 있으며 퍼지제어기의 성능개선에 필요한 소속함수의 조정효과가 신경회로망의 학습형태로 대체어지는 효과를 얻을 수 있어 기존의 퍼지제어기의 단점을 보완하게된다.

참 고 문 헌

1. S. M. Smith and D. J. Comer, "Automated Calibration of a FLC using a Cell State Space Algorithm," IEEE Control Systems, pp.18-28, Aug. 1991.
2. T. Takagi and M. Sugeno, "Fuzzy Identification of System and Its Applications to Modeling and Control," IEEE Trans. on Syst. Man and Cybern., SMC-15, pp.116-132, 1985.
3. H. Nomura, I. Hayashi and N. Wakami, "Self-Tuning Method of Fuzzy Control by Decent Method," Proceeding of 4th IFSA Congress, Engineering, pp.155-158, 1991.
4. C. C. Lee, "A Self-Learning Rule-Based Controller Employing Approximate Reasoning and Neural Net Concepts," International Journal of Intelligent Systems, vol.6, pp.71-93, 1991.
5. S. Horikawa and T. Furuhashi, "A Study on Fuzzy Modeling using Fuzzy Neural Networks," IFES, pp. 562-573, 1991.
6. C. T. Lin and C. S. George Lee, "Reinforcement Structure/Parameter Learning for Neural-Network-Based Fuzzy Logic Control Systems," IEEE Trans. on Fuzzy Systems, vol.2, no.1, pp.46-63, Feb. 1994.
7. H. Miyamoto, M. Kawato, "Feedback-Error-Learning Neural Network for Trajectory Control of a Robotic Manipulator," Neural Networks, vol.1, pp.251-265, 1988.
8. R. Jang and C. T. Sun, "Neuro-Fuzzy Modeling and Control," Proceedings of The IEEE, vol.83, no.3, pp.378-406, March 1995.
9. D. Psaltis, "A Multilayered Neural Network Controller," IEEE Control Systems Magazine, pp.17-21, April 1988.
10. J. R. Jang, "ANFIS : Adaptive-Network-Based Fuzzy Inference System" IEEE Trans. on Syst. Man Cybern., vol.23, no.3, pp.665-685, May/June 1993.
11. T. J. Procyk and E. H. Mamdani, "A Linguistic Self-Organizing Process Controller," Automatica vol. 15, pp.15-30, 1979.
12. R. Palm, "Robust Control by Fuzzy Sliding Mode," Automatica, vol.30, no.9, pp.1429-1437, 1994.
13. X. Cui, K. G. Shin, "Direct Control and Coordination Using Neural Networks," IEEE Trans. on Syst., Man and Cybern., vol.23, no.3, pp.686-697, 1993.
14. A. G. Barto, R. S. Sutton and C. W. Anderson, "Neuronlike Adaptive Elements That can Solve Difficult Learning Control Problems," IEEE Trans. on Syst. Man Cybern., SMC-13, no.5, pp.834-846, 1993.