'95 KACC (1995. 10. 23 ~ 25)

강인 학습 및 퍼지 추론에 의한 자율주행 캐치-볼 로보트에 관한 연구

Reinforcement Learning and Fuzzy Logic for An Autonomous Catch-Ball Robot

°조 상 규, 서 일 홍

한양대학교 전자공학과(Tel: 0345-408-5802; Fax: 0345-408-5803; E-mail: ihsuh@chollian.dacom.co.kr)

Abstracts A reinforcement learning system is shown to be enhanced by incorporating fuzzy inference system. To show the validity of the proposed method, numerical examples for catch-ball robot are illustrated.

Keywords Reinforcement learning, Fuzzy logic, Catch-ball robot

1. 서론

인공지능과 로보틱스에서의 궁극적인 목적은 주어진 작업을 수행하기에 적합한 내부구조를 스스로 조직할 수 있는, 학습능력을 갖춘 자율성을 갖는 로보트(autonomous agent)를 실현하는 것이다[5]. 강인 학습은 작업에 대한 추상적인 표현만으로도 작업과정(procedural program)을 스스로 만들어 내며 모델에 대한 정보가 없이도 예외처리 능력(reactive behavior) 및 적응성(adaptive behaviors)의 능력이 뛰어나[4] 많은 연구가 활발히 진행되고 있다. Asada는 시각정보에 의한 강인 학습에 대한 연구를 하고 공을 골문에 밀어 넣는 실험을 하였고[5], Connell은 상자를 물고 다니는 로보트를 구현하였다[4]. Aboaf는 로보트의 조인트의 변수를 모르고도 던져진 공의 결과에 대해 스스로 기구적 파라메터를 조절하여 목표물을 맞추는 실험을 하였으며[2] 이와 비슷한 방법이 저글링에서도 실험에 의해 적용되었다.

그러나, 강인 학습은 학습속도가 입력 상태의 수에 대해 기하급수적으로 증가하는 큰 단점이 있다. 실제 환경 하에서의 로보트는 복잡한 작업을 요구하는데, 작업이 복잡해짐에 따라 많은 입력 상태를 필요로 하게 된다. 따라서 학습 속도가 매우 커 실제의 로보트에 적용하기엔 어렵다. Asada는 LEM(Learning from Easy Mission)[5]을 사용하여 선형적으로 증가하도록 개선하였으나, 실제로 수행하고자 하는 작업의 경우 필요한 입력 상태의 수가 많아 여전히 많은 시간의 학습을 필요로 한다.

또한 시각정보의 특성상 로보트 가까이에서의 물체의 작은 움직임은 실제로 큰 변화를 주게 되고 또 카메라로부터 멀리 떨어진 곳에서의 대상물체의 움직임은 변화가 매우 작다. 이것을 상태-출력 편차(state-action deviation problem)[1]라 하는데 이것을 해결하기 위하여 입력 상태가 천이 되기 전까지는 출력(action)을 계속 유지하는, 다시 말해 하나의 입력 상태에 대해 한 출력만을 맵핑하는 것을 유지시켜 주도록 한다. 이러한 특성을 보정하기 위해 카메라 입력 상태에 대한 수준을 크게 나누고 학습속도의 향상도 꾀하였으나 그로 인하여 입력 상태가 이산적으로 존재하게 되고 따라서 출력 또한 이산적으로 나타나게 된다. 따라서 연속적으로 움직이는 물체에 대해서 연속적인 출력을 취할 수가 없어 로보트의 앞에서의 물체의 움직임을 놓치기가 쉬우며 멀리 있는 물체에 대해서도 그 추적이 매우 어려워지게 된다. 영상 자코비안이나 FMFNN[3]등의 방법에 의해 얻어지는 물체의 특징점과 같이 카메라의 입력 정보에 대한 분석적 정보를 센서 상태로 하는 것이 바람직하다. 그리고 입력 상태를 세밀하게 나누어 각각의 출력을 정의함으로써 위와 같은 문제를 해결할 수도 있겠으나 이는 학습시간을 현저하게 증가시키게 된다. 따라서 내부적으로는 적은 수의 입력 상태를 갖으면서도 연속적인 센서 정보로부터 연속적인 출력을 취할 수 있는 구조를 필요로 하게 된다. 본 논문에서는 입력 상태와 상태 출력을 각각 퍼지 변수로 하고 그 입출력 관계를 강인 학습에 의해 학습하는 Q-퍼지 규칙 학습기를 제안하여 위의 문제를 해결하였다. 강인 학습에 대해 퍼지 변수를 사용함으로써 학습에 의한 각각의 퍼지 관계로부터 임의의 입력값에 대한 출력을 추론해 낼수 있어 로보트는 연속적으로 움직이는 대상 물체에 대해, 또는 연속적으로 변화하는 입력값에 대해 연속적인 움직임을 가질 수 있게 된다.

캐치볼 로보트는 시각정보에 의해 공중으로 던져진 공을 추적하여 잡아내는 야구에서의 수비수와 같은 작업을 수행하는 AMR(Autonomous Mobile Robot)이다. 바퀴형(wheel-type) AMR은 기구적 제약(non-holonomic property)으로 인해 로보트 메니퓰레이터를 위한 기존의 시각 구동 이론을 그대로 적용하기가 어려우며 AMR을 위한 안정된 궤환 제어 법칙이 없어 제어가 어렵다.[7]

본 논문에서는 움직이는 공을 추적하여 잡아내는 캐치-볼 로보트를 모의 실험에 의해 구현하고 제안한 방법의 효용성을 보였다. 2장에서 Q-learning과 퍼지 추론 및 제안하고자 하는 방법에 대해 설명을 하였으며, 3장에서는 캐치-볼 로보트의 구현에 대해 기술하였고, 4장에서 모의 실험결과를 보였다.

2. Q-learing및 퍼지 추론

2.1 Q-learning

강인 학습에서는 로보트와 외부환경과의 상호작용을 마코프

의사결정 처리(Marcov decision process)로 모델링하고 시간적으로 축적된 총 보상(reward)의 식 (1)과 같은 측정치(measure)를 최대로 하는 제어 전략(control policy)을 학습한다. 이는 동적 프로그래밍(dynamic programming)에서의 최적값 이론(Optimal Theorem)[6]에 의해 최적의 해를 구할 수가 있으며 식(2)의 관계를 만족할 경우 식(3)으로부터 최적의 해를 갖게 된다.

$$\sum_{n=0}^{\infty}\gamma^{n}r_{t+n} \tag{1}$$

$$Q(x,f(x))=\max_{a'\in A}(Q_f(x,a')) \tag{2}$$

$$f^{*}=a \text{ such that } Q_f^{*}(x,a)=\max_{b'\in A}[\ Q_f^{*}(x,b')] \tag{3}$$

여기서, $Q^{*}(s,a)$는 입력 상태 s에 대해서 동작 a를 취하게 될 기대 리턴 값, 또는 상태치 함수(action-value function)이며 식 (4)와 같다.

$$Q^{*}(s,a)=r(s,a+\gamma\sum_{s'\in S}T(s,a,s')\max_{a'\in A}Q^{*}(s',a') \tag{4}$$

여기서 T는 state transition function이며 r은 environment로부터 받게 되는 reward이다.

Q value를 구하는 대표적인 방법으로 Watkins[1]가 제안한 Q-learning 알고리즘이 있으며, 식(5)에 의해 Q-value가 갱신되며 학습을 한다.

$$Q(s,a)\leftarrow(1-\alpha)Q(s,a)+\alpha(r+\gamma U(y))$$
$$U(y)=\max_{b'\in A}Q(x,b') \tag{5}$$

2.2 퍼지 추론

퍼지 이론은 Zadeh 교수에 의해 시작된 이래 국내에서도 이미 많은 연구사 된바가 있다. 퍼지 논리를 이용한 제어는 플랜트의 수학적 모델링이 필요 없고 설계자의 직관이나 경험에 의한 제어 이론으로써 모델링이 어렵거나 복잡한 플랜트를 제어하기에 적합하다. 이 이론은 다른 고전적 제어 이론에 비하여 쉽고 비교적 좋은 성능을 얻을 수 있는 반면 제어 시스템의 체계적인 설계절차가 없으며, 오랜 시간에 걸친 시행 착오에 의한 설계를 해야 하며, 소속함수의 형태를 결정하는 게 특별한 기준이 없다는 단점이 있다.

퍼지 추론은 다음과 같은 기본적 구조를 가진다.

만약 x가 A이고 y가 B이면 z는 C이다. (6)

제어 시스템의 퍼지 규칙이 주어진 후 퍼지 제어기는 임의의 입력에 해당하는 제어 입력을 계산하는데 이러한 제어 입력의 계산을 퍼지 추론(Fuzzy inference, Fuzzy composition)이라 한다. 퍼지 추론 방법에는 여러 가지가 제시되고 있지만, 가장 일반적인 방법으로 식(7)과 같은 SUP MIN 추론 방법이 있다.

$$u=x\cdot R \quad \text{또는}$$
$$\mu_u(u)=Sup_x[\ Min(\mu_x(x),\mu_R(x,u)] \tag{7}$$

3. 입력상태 변수와 출력변수의 구성

캐치-볼 로보트는 임의의 직선방향으로 움직이는 공을 그 움직이는 방향에 대해 일정한 자세를 유지하여 받도록 하는 것이다(그림 1.).

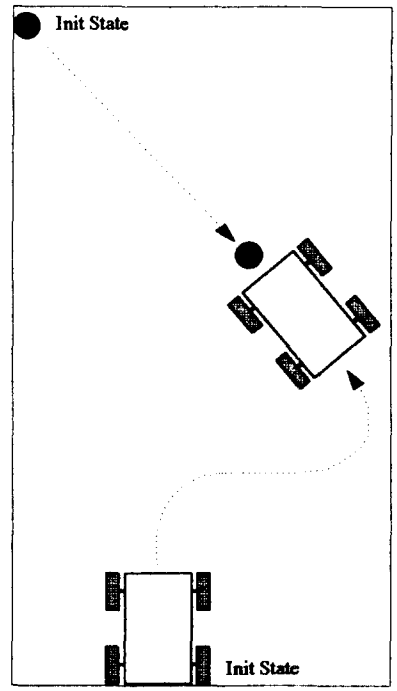


그림 1. 캐치-볼 로보트
Fig 1. The task of Catch-ball Robot

로보트는 카메라로부터 입력된 영상 신호를 전처리및 연산과정을 통해 로보트로부터의 상대적인 공의 위치만을 입력받는다. 로보트의 작업 제어기는 공의 위치를 입력 상태로 하고 좌우 각 바퀴의 동작을 출력으로 한다(그림 2.).

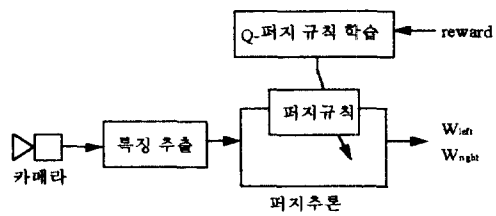


그림2. 제안된 로보트 학습 제어기의 구조
Fig 2. Block Diagram of Robot Task Controller

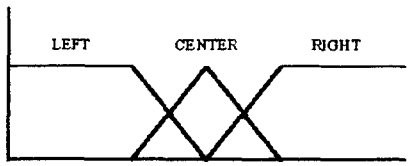
입력값과 출력값들을 그 구간에 대해 몇개의 상태로 분할하며 이것은 각각 강인 학습을 위한 입력 상태와 상태 출력으로 구성한다. 강인학습에 의해 만들어진 입력 상태와 상태 출력과의 관계는 다음과 같이 퍼지 규칙을 이루게 된다.

(*Rule*1)
if ($dist.=FAR$) and ($pos.=LEFT$) and ($dir.=LEFT$),
then $W_{left}=backward$ and $W_{right}=forward$.
(*Rule*2)
if ($dist.=FAR$) and ($pos.=LEFT$) and ($dir.=RIGHT$),
then $W_{left}=forward$ and $W_{right}=forward$.
.
.
.
(*Rule* n)

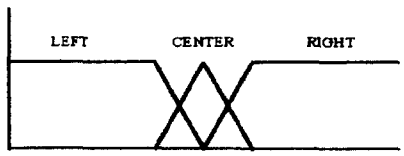
$$\text{if } (dist. = NEAR) \text{ and } (pos. = RIGHT) \text{ and } (dir. = RIGHT),$$
$$\text{then } W_{left} = backward \text{ and } W_{right} = forward.$$

따라서 강인 학습에 의해서 퍼지 규칙을 생성하는것과 같은 구조를 가지게 된다. 센서로부터 입력된 값은 퍼지화를 거친후 학습된 퍼지 규칙에 의해 새로운 퍼지 출력을 추론해 내고 이 값은 비퍼지화에 의해 연속적인 출력값을 갖게 된다.

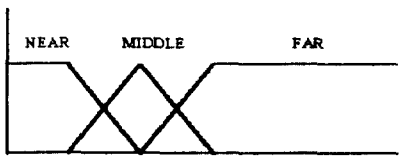
입력 센서의 상태 퍼지 집합과 출력 변수의 퍼지 집합은 다음과 같이 정의하였다. 입력 센서의 상태에 대해서는 로보트의 정면에서 본 공의 좌우 위치;왼쪽(LEFT), 가운데(CENTER), 오른쪽(RIGHT), 로보트에 대해 상대적으로 본, 공의 굴러가는 방향; 왼쪽(LEFT), 정지(STOP), 오른쪽(RIGHT), 로보트로부터 떨어진 공의 거리; 가깝다(NEAR), 멀다(FAR), 중간(MIDDLE). action에 대해서는 왼쪽 바퀴, 오른쪽 바퀴 각각에 대해 전방향 회전(F;Foward rotation), 후방향 회전(B; Backward rotation)으로 구분하였다. 그리고 퍼지 추론을 위한 퍼지 소속함수는 그림 3., 그림 4.과 같다.



(a) 공의 위치 Position of ball

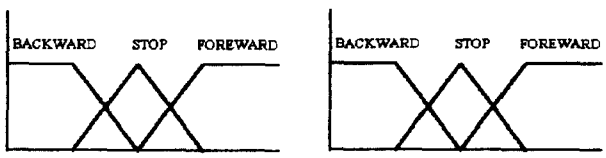


(b) 공이 움직이는 방향 Direction of ball



(c) 공과의 거리 Distance from ball

그림 3. 센서 상태 입력에 대한 퍼지 소속 함수
Fig 3. Fuzzy Membership Functions for Sensor State



(a) 왼쪽바퀴 Left Wheel (b) 오른쪽 바퀴 Right Wheel

그림 4. 출력 값에 대한 퍼지 소속 함수
Fig 4. Fuzzy Membership Functions for Output

4. 모의 실험

본 실험은 486 PC에서 10분간 약 6만회 학습을 하였다. 학습에 사용된 식(5)에서 감쇠인자 γ = 0.8, 학습속도 α = 0.3으로 하여 실험하였으며, 그림 5.와 같은 결과가 나타났다. 그림 5.의 우측 하단의 그래프는 학습 수렴 지수를 나타내는 그래프인데 2만회의 학습 후에 수렴이 되었다. 사용된 입력 상태의 수는 27개인데, 이것을 퍼지 추론을 사용하지 않고 유사한 성능을 얻기 위해서는 약 1000개의 입력 상태를 필요로 하며 약 500만회의 학습을 통해서도 40%밖에 학습이 되지 않았다. 같은 수의 입력 상태의 수에 대해 퍼지 추론을 사용하지 않았을 경우의 모의 실험결과를 그림 6.에 보인다. 그림에서 알 수 있듯이 입력상태의 수가 적은 경우 공의 움직임에 대해 AMR의 움직임이 둔해지며 공의 속도가 빠른 경우엔 대부분 놓치게 된다. 또한 학습률도 20%밖에 되지 않았다.

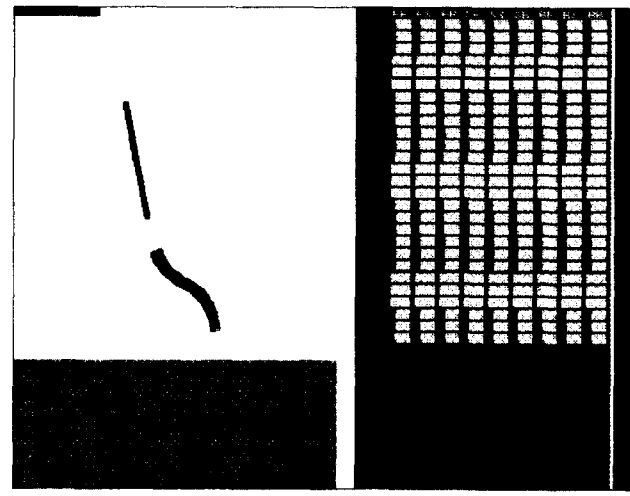


그림 5. 모의 실험 결과
Fig 5. The result of simulation

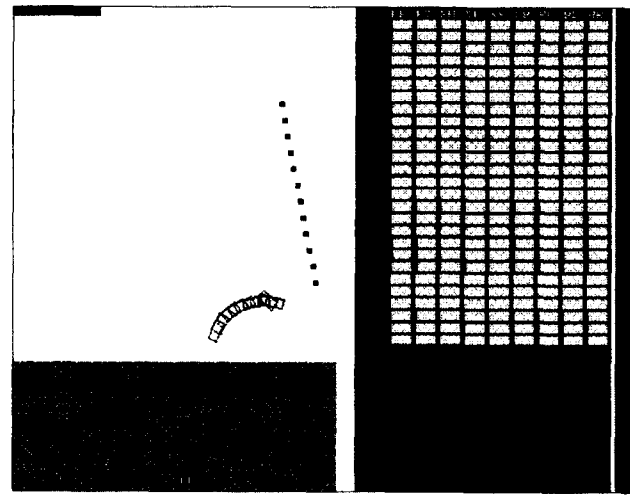


그림 6. 같은 수의 입력상태에 대한 강인 학습의 모의 실험 결과
Fig 6. The result of simulation for Reinforcement Learning in the same states set

5. 결론 및 추후 연구과제

본 논문에서는 강인 학습에 의하여 퍼지 규칙을 학습하며 퍼지 추론에 의해 연속적인 action을 할 수 있는 방법을 제안하였으며, 움직이는 대상을 추적하여 잡아내는 캐치-볼 로보트를 모의 실험에 의하여 구현하였다. 이로 인하여 고수준의 작업에 대

한 학습 속도를 현저하게 감소시킬 수가 있었고 로보트의 작업 성능을 향상시킬 수 있었다. 그러나 이러한 학습을 실제 로보트를 대상으로 하여 실험을 할 경우 같은 학습 반복횟수에 대해서도 상당한 시간이 소비된다. 따라서 모의 실험에 의해 학습된 결과를 실제 로보트의 학습의 초기 값으로 할 수 있는 연구가 필요하다.

6. 참고문헌

[1] C.J.C.H.Watkins. Learning from Delayed Rewards. PhD. Thesis, King's College, University of Cambridge, May 1989.

[2] E. W. Aboaf, et al, "Task-level robot learning," In *Proc. of 1988 IEEE Intr. Conf. on Robotics and Automation*, pp. 1309-1310, 1988

[3] I.H.Suh and T.W.Kim, "Fuzzy Membership Function Based Neural Networks with Application to the Visual Servoing of Robot Manipulators," *IEEE Trans. on Fuzzy Systems*, Vol.2, No.3, pp.203-220, 1994

[4] J. H. Connell, S. Mahadevan, "Rapid Task Learning For Real Robots," Robot Learning, Kluwer Academic Publishers, Ch 5.

[5] M. Asada, et al, "Vision-Based Reinforcement for Purposive Behavior Acquisition," *Proc. of Robotics and Automation*, pp. 146-153, 1995.

[6] R.E.Bellman. Dynamic Programming. Princeton University Press, Princeton, NJ, 1957.

[7] Y. Masutani, et al, "Visual Servoing for Non-Holonomic Mobile Robots," *Conf. on Intelligent Robot & Systems*, pp. 1133-1140, 1994