강화학습과 감정모델 기반의 지능적인 가상 캐릭터의 구현

Implementation of Intelligent Virtual Character Based on Reinforcement Learning and Emotion Model

우종하°, 박정은, 오경환
LG전자 MC사업본부 단말연구소°, 서강대학교 컴퓨터학과

Woo Jong Ha°, Jung-Eun Park, Kyung-Whan Oh
MOBILE HANDSET R&D CENTER°, LG Electronics Inc.,
Dept. of Computer Science and Engineering, Sogang University
E-mail : deepseas@lge.com°, jepark@sogang.ac.kr, kwoh@sogang.ac.kr

요 약

학습과 감정은 지능형 시스템을 구현하는데 있어 가장 중요한 요소이다. 본 논문에서는 강화학습을 이용하여 사용자와 상호작용을 하면서 학습을 수행하고 내부적인 감정모델을 가지고 있는 지능적인 가상 캐릭터를 구현하였다. 가상 캐릭터는 여러 가지 사물들로 이루어진 3D의 가상 환경 내에서 내부상태에 의해 자율적으로 동작하며, 또한 사용자는 가상 캐릭터에게 반복적인 명령을 통해 원하는 행동을 학습시킬 수 있다. 이러한 명령은 인공신경망을 사용하여 마우스의 제스처를 인식하여 수행할 수 있고 감정의 표현을 위해 Emotion-Mood-Personality 모델을 새로 제안하였다. 그리고 실험을 통해 사용자와 상호작용을 통한 감정의 변화를 살펴보았고 가상 캐릭터의 훈련에 따른 학습이 올바르게 수행되는 것을 확인하였다.

1. 서론

지능형 로봇은 21세기에 가장 유망한 산업중의 하나로 떠오르고 있다. 앞으로 로봇은 인간의 동반자로서 기존의 산업용 로봇에 국한되지 않고 가사지원, 오락, 교육, 의료, 군사 분야 등 매우 다양한 범위에서 활용될 것이다. 이러한 지능형 로봇을 구현하는데 있어 로봇공학, 제어공학 등 하드웨어적인 측면과 함께, 로봇 스스로 행동할 수 있는 인공지능, 자율제어와 같은 소프트웨어적인 측면 역시 핵심적인 역할을 담당한다.

지능 시스템을 구성하는데 있어 중요한 요소는 학습의 능력과 감정의 표현이다. 이와 같이 지능형 로봇에 학습과 감정의 기능을 구현하기 위한 연구가 지금까지 많이 수행되고 있으나, 아직까지 하지만 아직까지 학습과 감정 기능의 구현이 체계적으로 이루어지지는 않았다.

본 논문에서는 여러 학습 알고리즘 중 하나인 강화학습을 이용하여 사용자와 상호작용을 하면서 학습을 수행하고 내부적인 감정모델을 가지고 있는 지능적인 가상 캐릭터를 구현하였다. 가상 캐릭터는 여러 가지 사물들로 이루어진 3D의 가상 환경 내에서 내부상태에 의해 자율적으로 동작하며, 또한 사용자는 가상 캐릭터에게 반복적인 명령을 통해 원하는 행동을 학습시킬 수 있다. 이러한 명령은 인공신경망을 사용하여 마우스의 제스처를 인식하여 수행할 수 있다. 그리고 내부적인 행동과 사용자와의 상호작용에 의해 가상 캐릭터의 감정이 변하며, 이러한 감정은 학습이나 행동에 영향을 준다. 이와 같이 자율적으로 행동하고 사용자가 원하는 대로 행동을 수행하도록 훈련을 시키며 실제 애완동물과 같이 친근함을 느낄 수 있는 가상 캐릭터를 구현하는데 목적을 두었다. 본 논문은 다음과 같이 구성되어 있다. 2장에서는 가상 캐릭터의 내부 아키텍처와 인공신경망을 이용한 마우스의 제스처 인식, Emotion-Mood-Personality 기반의 인공감정 모델, 강화학습을 사용한 학습방법에 대해서 설명한다. 3장에서는 3D 환경에서 가상 캐릭터 구현방법, 제안한 감정모델과 학습방법이 적절히 수행되었는지 실험으

로 확인한다. 마지막으로 결론 및 앞으로의 연구방향에 대해 논의한다.

2. 강화학습 기반의 지능적 가상 캐릭터

2.1 가상 캐릭터 아키텍처

가상 캐릭터는 외부 환경과 사용자와 상호작용을 하며 내부적인 시스템들에 의해 행동을 수행한다. 크게 인지 시스템, 내부상태 시스템, 감정 시스템, 학습 시스템, 행동 시스템, 모터 시스템으로 이루어져 있으며, 다음 그림과 같다.

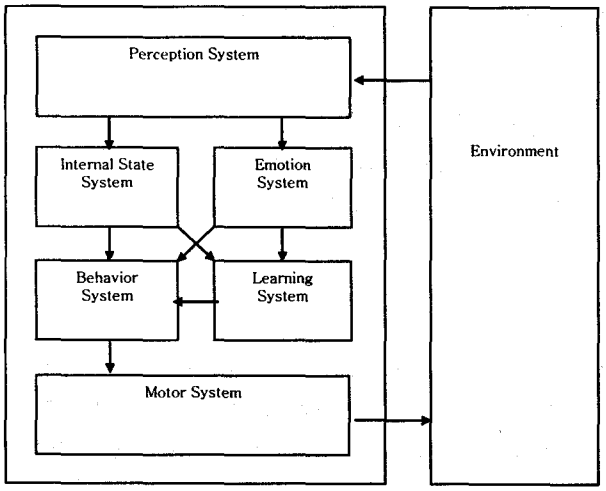


그림 1 가상 캐릭터의 아키텍처

■ 인지 시스템(perception system)

외부 환경과 사용자의 반응을 인식하는 부분이다. 자신의 현재 위치, 가상 환경내의 여러 가지 사물들을 파악하고 사용자의 명령과 자신의 행동에 대한 보상(reward)/벌칙(penalty)을 입력받을 수 있다. 사용자의 명령은 인공신경망을 사용해 마우스의 제스처를 인식한다.

■ 내부상태 시스템(internal state system)

가상 캐릭터의 내부적인 상태(배고픔, 피로도, 호기심 등)를 유지하는 부분이다. 사용자의 명령과 행동에 따라 상태값이 변하고, 각 상태들은 가상 캐릭터의 자율적인 행동양식에 영향을 준다.

■ 감정 시스템(emotion system)

가상 캐릭터의 감정을 나타내는 부분이다. 내부상태와 사용자의 명령, 보상/벌칙 등에 의해 영향을 받으며 자율적인 행동, 학습의 정도에 영향을 끼친다. 안정, 기쁨, 슬픔, 화남 등의 감정들이 존재한다.

■ 학습 시스템(learning system)

사용자의 반복적인 훈련을 통해 원하는 행동을 학습시키는 부분이다. 가상 캐릭터가 특정 행동을 수행하였을 때 보상/벌칙을 부여함으로써 임의의 상태에서 원하는 목표까지의 순차적인 행동을 강화학습을 통해 학습을 수행한다. 내부상태와 현재 감정에 의해 학습의 수행 정도에 영향을 받는다.

■ 행동 시스템(behavior system)

내부상태, 감정, 학습 모델들에 의해 현재 상태에서 가상 캐릭터가 수행해야할 행동을 결정하는 부분이다. 걷기, 앉기, 눕기, 일어나기, 공 줍기 등 각각의 상황에 따라 여러 행동들이 가능하다.

■ 모터 시스템(motor system)

행동 시스템에 의해 결정된 가상 캐릭터의 동작을 실제 3D 환경 내에서 수행하는 부분이다. 각 행동에 맞는 3D 모델이 애니메이션을 수행하고 행동에 따른 가상 캐릭터의 상태나 외부 환경을 변화시킨다.

2.2 인공신경망을 사용한 마우스의 제스처 인식

2.2.1 마우스 제스처의 벡터 변환

사용자는 가상 캐릭터에게 명령을 내리기 위해 마우스 제스처를 사용하지만, 이는 동일한 제스처를 입력하더라도 항상 조금씩 차이가 나기 때문에 정확히 제스처를 분류하기가 쉽지 않다. 그러므로 입력 데이터의 작은 변화에 상관없이 전체적인 패턴을 인식할 수 있는 인공신경망의 일반화 특성을 이용해 마우스 제스처를 판별하였다.

인공신경망 구성을 위한 첫단계인 입력데이터 설정을 위해 이 논문에서는 마우스의 제스처를 12개의 단위벡터로 변환하여 표현하였다.[1] 단위벡터로 변환하면 입력이 -1 ~ 1 사이로 표준화될 뿐만 아니라 제스처 패턴을 균등하게 나누어진 벡터로 표현하기 때문에 사용자가 입력한 제스처를 쉽게 인식할 수 있다.

마우스의 제스처를 12개의 벡터로 표현하려면 13개의 점이 필요하므로, 가공되지 않은 마우스 데이터를 13개의 점으로 변환되어야 한다. 이를 위해 우선 모든 점들을 검사하여 두 점 사이의 최소거리를 구한 후, 최소거리의 중간에 새로운 점을 삽입하고 양 끝점을 삭제한다. 이러한 방식으로 한 번에 한 점씩 줄여나가 점의 개수가 원하는 개수가 될 때까지 반복하여 13개의 점을 구한다.

2.2.2 마우스 제스처의 학습 및 인식

인공신경망을 구성하기 위하여 은닉층이 하나인 2층 퍼셉트론을 사용하였다. 인공신경망의 입력은 12개의 벡터를 사용하므로 모두 24이고 은닉층의 뉴런은 모두 6개를 사용하였다. 그리고 출력층의 뉴런은 5개를 사용하여 최대 5개의 제스처를 분류할 수 있도록 하였다. 분류할 수 있는 제스처의 개수가 많아지면 학습시간이 너무 오래 걸리기 때문에 제한을 두었다.

학습률은 1.0, 모멘텀은 0.9, 바이어스는 -1로 설정하였고, 출력값 판별시 0.96 이상일 경우 특정 제스처로 인식하도록 하였다. 또한 새로운 제스처가 입력되어 학습이 시작될 경우, 에러율은 0.003 이하일 때까지 실시간으로 반복 학습이 수행된다.

학습 중에는 새로운 제스처를 추가할 수 없다.

2.3 Emotion-Mood-Personality 기반의 감정 모델

사람들과 상호작용을 하며 단순히 기계적인 사물을 벗어나 친근한 느낌을 주기 위해 감정을 표현하는 기능은 필수적이다. 기존의 인공감정 모델에는 OCC 모델, Weiner 모델, FLAME 모델 등이 있는데, 대부분의 모델들은 감정요소가 너무 많고 이를 설계하거나 구현하기가 복잡해 실제로 사용하기가 힘든 점이 있다. 이러한 단점을 보완하여 Kubota는 Emotion-Mood 기반의 수치적이고 간단한 감정모델을 제안하였다.[2][3] 본 논문에서는 기존의 모델에 Personality라는 요소를 추가하고 Mood 요소에 감소인자(discount factor)를 도입, 발전시켜 Emotion-Mood- Personality 감정 모델을 설계하였다.

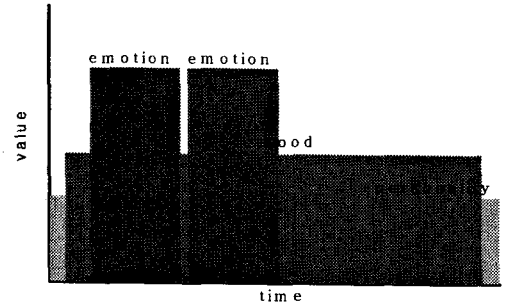


그림 2 감정의 계층

인간의 감정의 기본 바탕에는 크게 세 개의 계층으로 이루어져 있다. 가장 상위 레벨은 일시적인 감정이다. 이것은 어떤 이벤트에 대해 잠시 보이는 행동들을 말한다. 그 다음 레벨은 일시적인 감정들의 누적된 효과로 인해 좀 더 오래 지속되는 정서적인 상태를 말하는 기분(Mood)이다. 이 두 계층이 기반을 두고 있는 것이 성격(Personality)이다. 성격은 항상 나타나는 것으로, 일시적인 감정이나 감정이 없는 상황에서 일반적으로 보이는 태도이다. 감정의 계층은 그림 2와 같다. Emotion은 0 ~ 1 사이의 실수로 가상 캐릭터가 상황에 따라 입력받는 일시적 감정에 대한 수치값이다. Mood는 Emotion을 입력으로 받아 이전 Mood의 값에 더해지며 각 단계마다 일정한 수치만큼 감소가 되어 시간이 지날수록 현재 Mood 값이 떨어지게 된다.

$$Mood_m(n) = \frac{\sum_{i=0}^{a-1} \rho^i \{\sigma \cdot Emotion_m(n \quad i) + Mood_m(n-i-1)\}}{\sum_{i=0}^{a-1} \rho^i}$$

at each time step,
for all i (0 i a) {
$Mood_m(i) = Mood_m(i) - \varepsilon$
}

여기서 ρ, σ는 감소율(reduction rate), ε은 감소인자(discount factor), a는 특정 단위내의 감정의 개수이다. Mood는 다시 Personality의 입력으로 사용되며 이전 Personality에 더해진다. Personality는 단지 그 수치값보다 여러 Personality 사이의 비율이 중요하므로 정규화시킨다.

$$Personality_m(n) = \frac{\sum_{i=0}^{a-1} \lambda^i \{\mu \cdot Mood_m(n \quad i) + Personality_m(n-i-1)\}}{\sum_{i=0}^{a-1} \lambda^i}$$

$$NormP_m(n) = \frac{Personality_m(n)}{\sum_{i=0}^{e-1} Personality_i(n)}$$

if $Personality_m(n) \geq 1$,
then {
for all i (0 i e)
for all j (0 j a)
$Personality_i(j) = \frac{Personality_i(j)}{2}$
}

여기서 λ, μ는 감소율이고 NormP는 정규화 된 Personality 값이다. 최종적으로 외부로 표현되는 현재 감정인 Feeling은 Emotion, Mood, NormP를 특정 비율에 따라 조합하여 다음식과 같이 계산할 수 있다.

$$Feeling_m(n) = \alpha \cdot Emotion_m(n) \quad \beta \cdot Mood_m(n) \quad \gamma \cdot NormP_m(n)$$

2.4 강화학습을 통한 사용자의 명령 학습

2.4.1 Q학습을 사용한 가상 캐릭터의 학습

Q학습(Q-Learning)은 강화학습 방법 중 가장 널리 이용되는 대표적인 학습 방법이다.[4] 가상 캐릭터에 Q학습을 적용하기 위해서는 먼저 상태(states)와 행동(actions)을 결정해야 한다. 상태는 현재 환경 내에서 가상 캐릭터가 가질 수 있는 모든 경우의 수를 뜻한다. 여기서는 사용자의 명령 종류, 가상 캐릭터의 위치와 현재 동작상태, 여러 가지 사물들의 위치 등을 사용하였다. 행동은 가상 캐릭터가 현재 상태에서 수행할 수 있는 모든 움직임들을 나타낸다. 행동을 수행한 다음에는 그 행동에 따라 현재 상태가 변하게 된다. 가능한 행동에는 앉기, 눕기, 서기, 걷기, 물건을 줍기, 물건을 놓기 등이 있다. 다음으로 고려해야 될 점은 환경으로부터 받는 강화값을 결정하는 것이다. 강화값은 고정된 값이 아니라 사용자의 훈련목표에 따라서 동적으로 변한다. 그러므로 사용자가 가상 캐릭터에게 보상이나 벌칙을 줄때마다 현재상태와 강화값을 배열에 저장한 다음 Q학습에 사용한다.

가상 캐릭터를 학습시키기 위해서는 우선 제스처를 선택한 후 가상 캐릭터가 특정 행동을 취했을 때 보상이나 벌칙을 주면 Q학습에 의해 내부적으로 학습이 된다. 한 번에 학습할 수 있는 시간을 제한해 두었기 때문에 같은 동작을 반복하

여 학습시켜야 완벽하게 학습이 가능하다.

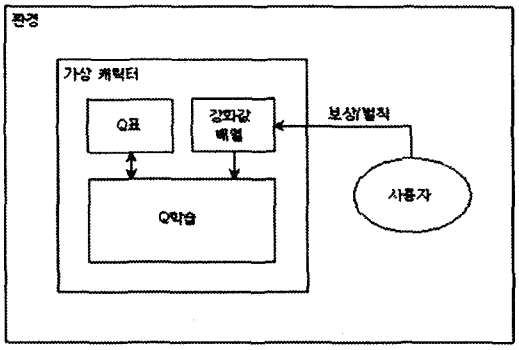


그림 3 Q학습을 사용한 가상 캐릭터의 학습 구조

3. 가상 캐릭터의 관찰 및 실험

가상 캐릭터의 구현 및 실험에 있어 Visual C++ 환경을 기반으로 DirectX 9.0 라이브러리를 이용하여 3D를 구현하였다. 가상 캐릭터는 3D 환경 내에서 자율적으로 동작하며, 현재 감정을 표현한다. 그리고 환경내의 여러 가지 사물들과 사용자의 명령에 따라 적절한 행동을 수행하도록 학습할 수 있다. 그림 4에서 화면 내에 가상 캐릭터와 여러 사물들이 존재하고 가상 캐릭터의 머리위로 감정을 나타내는 이모티콘이 있다. 화면 하단에는 내부상태와 메뉴아이콘, 마우스 제스처를 나타내고 있다.

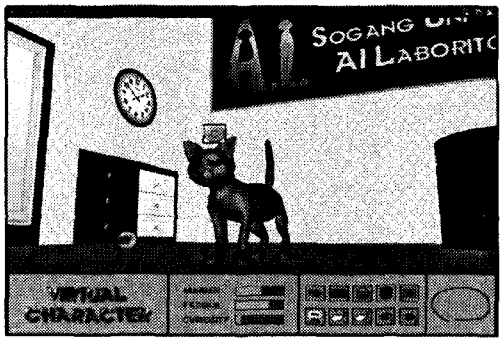


그림 4 가상 캐릭터의 실행화면

3.1 관찰 : 사용자와 상호작용을 통한 가상 캐릭터의 훈련

가상 캐릭터에게 훈련시킬 수 있는 다양한 행동 중에서 먼저 제일 간단한 행동인 앉는 동작을 학습시켰다. 제스처를 입력한 뒤에 가상 캐릭터가 앉는 동작을 취할 때 보상을 주어 학습시켰다. 일단 학습시키고 난 후에는 가상 캐릭터가 어느 위치에 있든지 제스처를 통해 명령을 내리면 가상 캐릭터는 앉는 동작을 수행하였다. 다음으로 가상 캐릭터에게 특정 위치로 이동하도록 훈련을 시켜보았다. 먼저 새로운 제스처를 입력한 다음 이동 명령을 사용해 원하는 위치로 움직이게 하였다. 그리고 목표한 위치에 도착했을 때 제스처를 입력한 후 가상 캐릭터에게 보상을 주었다. 학습을 시킨 후 제스처를 입력하자 가상 캐릭터는 장애물들을 피하며 원하는 위치로 올바르게 이동함을 볼 수 있었다.

마지막으로 공을 목표지점에 놓은 행동을 훈련시켜 보았다. 먼저 가상 캐릭터를 공이 위치한 곳으로 이동시켜 공을 집도록 하였다. 중간에 공을 다시 놓는 경우가 있으므로 공을 집는 행동을 별도의 명령으로 학습을 시켰다. 공을 집은 후, 다시 가상 캐릭터를 목표지점으로 이동시켜 공을 놓을 때 보상을 주었다. 그런 다음 제스처로 명령을 내렸으나, 복잡한 행동이기 때문에 모든 행동이 학습되지 않아 원하는 행동을 수행하지는 않았다. 여러 번 같은 동작을 반복해 학습시킨 후에야, 가상 캐릭터는 장애물 사이를 지나 공을 집은 뒤 다시 목표지점으로 가서 공을 내려놓는 동작을 완벽하게 수행하였다.

3.2 실험 1 : 가상 캐릭터의 감정변화

가상 캐릭터는 안정, 기쁨, 슬픔, 화남의 감정을 표현하였으며, 사용자와 환경으로부터 받는 Emotion 입력을 통해 감정모델의 Personality와 Mood가 변하게 된다. 그리고 Emotion, Mood, NormP를 사용해 실제 표현되는 감정인 Feeling을 계산할 수 있다.

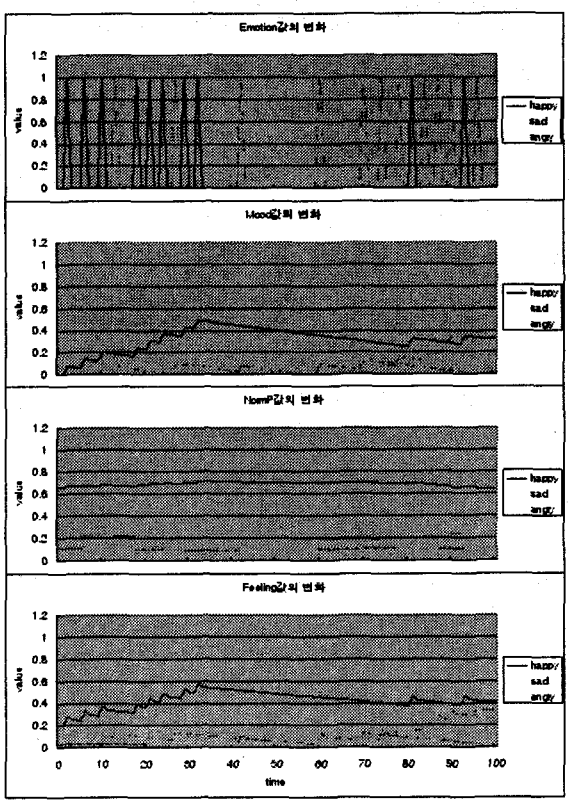


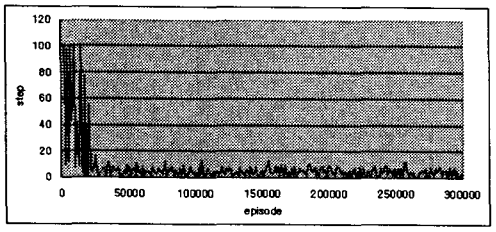
그림 5 감정상태의 변화

그림 5에서와 같이 Emotion의 입력에 따라 Mood가 변하며 시간이 지남에 따라 Mood가 조금씩 감소하며, Mood에 의해 NormP 역시 영향을 받

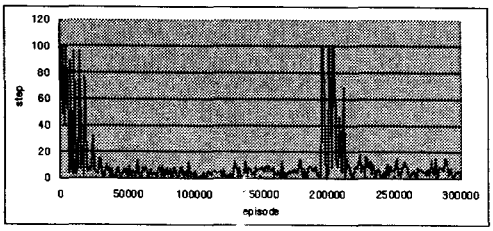
지만 그 변화의 정도는 Mood보다 훨씬 적음을 알 수 있다. 또한 특정 감정의 Mood가 높다 하더라도 NormP가 낮다면 최종적인 Feeling의 값이 달라져 성격이 감정에 영향을 주는 것을 확인할 수 있었다.

3.3 실험 2 : 가상 캐릭터의 학습결과

Q학습을 이용한 가상 캐릭터의 학습이 올바르게 수행되는지 살펴보았다. 그림 6은 특정 위치로 이동 훈련의 학습 결과이다. 처음에는 허용 가능한 행동의 수가 최대치를 넘도록 목표에 도달하지 못하지만 에피소드가 약 20000회 정도 반복되면 최소의 행동으로 수렴함을 볼 수 있다. 또한 중간에 목표의 위치를 다르게 하여 학습을 하는 경우에도 시간이 지남에 따라 다시 새로운 상황에 맞게 학습이 이루어졌다.

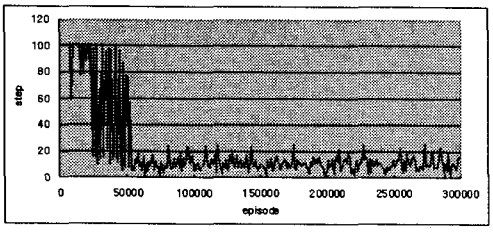


(a) 하나의 목표로 학습을 수행

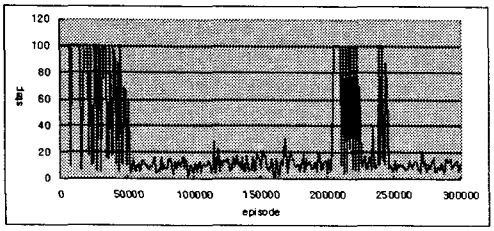


(b) 중간에 목표가 변경되어 학습을 수행

그림 6 특정 위치로 이동 훈련의 학습결과



(a) 환경이 변하지 않고 학습을 수행



(b) 중간에 환경이 변화되어 학습을 수행

그림 7 공을 집은 뒤 목표에 놓는 훈련의 학습결과

그림 7은 공을 집은 뒤 목표에 놓는 훈련의 학습결과이다. 특정 위치로 이동 훈련과 달리 공을 집은 다음 다시 목표까지 이동하여 공을 다시 내려놓아야 하기 때문에 필요한 행동의 수가 더 많다. 그래서 학습에 걸리는 시간이 더 늘어나 에피소드가 약 50000회 정도 되었을 때 수렴이 되었다. 또한 중간에 장애물의 위치가 바뀌어 환경이 변화되었을 때에도 시간이 지남에 따라 다시 학습이 수렴함을 볼 수 있다.

4. 결론 및 향후 연구과제

본 논문에서는 강화학습을 이용하여 사용자와 상호작용을 하면서 학습을 수행하고 내부적인 감정모델을 가지고 있는 지능적인 가상 캐릭터를 구현하였다. 가상 캐릭터는 여러 가지 사물들로 이루어진 3D의 가상 환경 내에서 내부상태에 의해 자율적으로 동작하며, 또한 사용자는 가상 캐릭터에게 반복적인 명령을 통해 원하는 행동을 학습시킬 수 있다. 이러한 명령은 인공신경망을 사용하여 마우스의 제스처를 인식하여 수행할 수 있고 감정의 표현을 위해 Emotion-Mood-Personality 모델을 새로 제안하였다. 그리고 실험을 통해 사용자와 상호작용을 통한 감정의 변화를 살펴보았고 가상 캐릭터의 훈련에 따른 학습이 올바르게 수행되는 것을 확인하였다.

향후 연구과제로는 기존 가상 캐릭터의 행동들과 가상 환경에서의 사물들의 종류를 좀더 다양화 시키고 보다 복잡한 행동양식을 수행할 수 있도록 확장할 필요가 있다. 또한 학습을 위한 상태들을 결정함에 있어 많은 상황들을 더 일반적이며 적은 상태를 유지할 수 있도록 구현해야 한다. 그리고 환경이 복잡해짐에 따라 학습시간이 증가하므로 이러한 학습이 실시간으로 처리될 수 있도록 학습 알고리즘을 개선해야 할 것이다.

감사의 글 : 본 연구는 2005년도 서강대학교 교내 연구비에 의해 지원되었음

5. 참고문헌

[1] Mat Buckland, "AI Techniques for Game Programming", Thomson Learning, 2002

[2] N. Kubota, Y. Nojima, N. Baba, F. Kojima, T. Fukuda, "Evolving Pet Robot with Emotional Model", Proceedings of the CEC, pp.1231-1237, 2000

[3] N. Kubota, F. Kojima, T. Fukuda, "Self-Consciousness and Emotion for A Pet Robot with Structured Intelligence", IFSA World Congress and 20th NAFIPS International Conference, pp.2786-2791, 2001

[4] C. J. Watkins, P. Dayan, "Technical Note: Q-Learning Machine Learning", pp.279-292, 1992