

강화학습을 이용한 지능형 게임캐릭터의 제어

Control of Intelligent Characters using Reinforcement Learning

신 용 우*

Yong-Woo Shin

요 약

과거에는 게임프로그램 제작이 단순히 3D, 온라인게임, 엔진프로그래밍 또는 게임프로그래밍으로 분류하여 제작하였다. 그러나 이제는 게임프로그래밍의 종류가 세분화되었고, 기존에 없던 인공지능 게임프로그래머의 역할이 게임을 좀 더 재미 있게 할 수 있는 시점이라 하겠다.

본 논문에서는 강화학습 알고리즘을 이용하여, 보상 값을 받아 게임캐릭터가 학습하여 지능적인 움직임을 나타나게 하였다. 구현된 게임캐릭터가 지능적으로 잘 움직이는지 확인하기위해, 슈팅게임을 제작하여 적 캐릭터와 전투를 하게 하였다. 실험결과 임의로 움직이는 캐릭터보다 월등히 방어함을 알 수 있었다.

Abstract

Game program had been classed by 3D or on-line game etc. and engine and game programming simply. But, game programmer's kind more classified now. Artificial Intelligence game programmer's role is important.

This paper makes game character study and moved by intelligence using reinforcement learning algorithm. Fought with character enemy using developed game. Confirmed whether embodied game character is facile by intelligence. As result of an experiment, we know, studied character defends excellently than randomly moved character.

☞ keywords : Reinforcement Learning, Game, Game Character, Artificial intelligence, Intelligence 강화학습 게임 게임캐릭터 인공지능 지능

1. 서 론

인터넷의 보급이 많아짐에 따라 온라인게임의 제작이 많아지고 또한 3D 게임의 제작 또한 많아지고 있다. 과거에는 엔진과 게임프로그램으로 나누어 개발하던 시점에서 이제는 좀 더 세분화하여 분야별 전문프로그래머가 필요하게 되었다.

인공지능 분야에서도 현재까지 캐릭터의 이동 처리를 위해서는 패턴(Pattern), A* 알고리즘이나 FSM, 퍼지(Fuzzy), FuSM 을 이용하여 캐릭터의 자동화를 하고 있다. 패턴이란 미리 주어진 방향으로 캐릭터가 이동하게 하는 단순한 논리이다. FSM 은 캐릭터의 여러 가지 행위를 상황에 따라

적용시키는 알고리즘이다. 또한 캐릭터의 움직임을 다양화 시키는 퍼지나 FuSM 그리고 길 찾기에 사용되는 A* 알고리즘 등은 캐릭터가 주어진 방향이나 목적지를 찾는 데에는 유용하게 사용할 수 있으나 캐릭터의 근본적인 지능을 높여주지는 못한다.

패턴이나 FSM 을 이용한 캐릭터의 이동방법은 경우의 수가 작게 제한되어 단순하고 보다 많은 경우의 수를 가지는 캐릭터의 전투상황에서 효율적인 전투를 벌이지는 못한다. 그러므로 학습 알고리즘을 적용하여야 한다.

강화학습은 온라인 학습기법으로 유사알고리즘 보다 비교적 짧은 시간에 효과적으로 캐릭터를 학습시킨다. 강화학습을 이용하여 특정목적을 달성하였을 때마다 보상을 하는 작업을 반복하다보면 캐릭터가 학습하게 된다.

* 정 회 원 : 동아방송예술대학 게임에니메이션계열 조교수
ywshin@dima.ac.kr

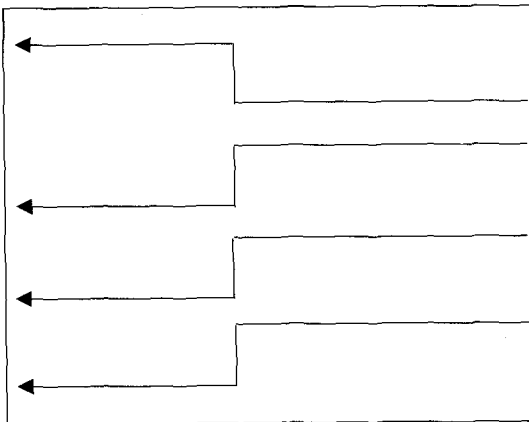
[2007/05/19 투고 - 2007/06/11 심사 - 2007/07/18 심사완료]

본 논문에서는 아군캐릭터와 적군캐릭터의 전투상황에서 아군캐릭터가 적군캐릭터의 공격을 효과적으로 피할 수 있도록 학습시킨다. 슈팅게임을 제작하여 일정한 패턴으로 움직이는 적군캐릭터들을 상대로 학습을 하여, 일정한 횟수의 학습에 다다랐을 때 스스로 적군캐릭터들을 지능적으로 상대하게 된다. 임의의 방향으로 움직이는 아군캐릭터와 비교할 때 월등하게 좋은 학습효과를 알 수 있다.

본 논문의 구성은 1장 서론에 이어 2장에서는 관련연구에 대해 살펴보고 3장에서는 지능형게임 캐릭터의 구현에 대해 알아보고 4장에서는 실험 및 결과에 대해 알아본다. 마지막으로 5장에서는 최종 결론을 맺도록 한다.

2. 관련연구

2.1 패턴 (Pattern)



(그림 1) (캐릭터의 움직임을 나타내는 패턴)

게임에서 자유자재로 움직이는 캐릭터들을 보았다면, 프로그래머의 입장에서 생각한다면 한번쯤은 그 캐릭터들이 어떻게 만들어져서 움직이는지를 생각해 보았을 것이다.

이것은 패턴이라는 미리 정해진 방법을 통하여 캐릭터들이 움직이는 것이고, 캐릭터마다 주어진

여러 가지 패턴으로 각각 다른 움직임을 움직이는 것처럼 보일 뿐이다. 패턴은 다소 단조로운 면이 있다. 그러나 한 가지 기본패턴을 만들었을 때, 여러 가지의 수치를 적용하여 파생패턴을 만들면 다양한 형태의 움직임을 나타낼 수 있다.

2.2 슈팅게임 (Shooting Game)

오늘날에는 다양한 장르의 게임이 존재하지만 원초적으로는 슈팅게임이 게임의 원초적인 형태라고 생각한다.

슈팅게임이란 단순한 키 조작이나 마우스, 조이스틱의 조작으로 적 캐릭터를 파괴하는 원초적인 형태의 게임으로 업소용 게임기나 가정용 게임기에서부터 최근에는 온라인게임에 이르기까지 여러 기기에서 구동되는 게임을 말한다.

게임을 하기 시작한 초창기에는 인베이더, 갤러그, 제비우스 등으로 시작하여 라이덴 등이 있다. 3차원그래픽을 사용한 이후에도 슈팅게임은 둠, 퀘이크 시리즈, 하프라이프 시리즈 등으로 이어지며 FPS 라는 독자적인 장르를 구축했다. 폴리곤을 사용한 3차원 환경에서 길을 찾으며 적 캐릭터와 전투를 벌인다. 최근에는 온라인 게임으로 인터넷을 이용하여 원격지의 사람과 대전을 하고 있다.

2.3 게임 인공지능의 개요

사람과 컴퓨터가 상호작용하여 플레이하는 게임에는 어디에나 인공지능이 존재하고 있다.

우리나라에서 아주 유명했던 스타크래프트라는 전략게임에서 캐릭터들의 길을 찾기 위해 사용되었던 A* 길 찾기 알고리즘은 장애물을 피해 원하는 목적지로 효율적인 방식으로 캐릭터들을 이동시키는 길을 생성하였다.

비교적 단순한 게임에도 인공지능은 사용되고 있다. 상대캐릭터의 이동이나 공격, 방어 등의 행위에 따라 아군캐릭터의 행위를 제어하는 유한상

태기계(FSM)는 모든 게임에서 광범위하게 사용되고 쉽게 구현가능하다.

인간의 일상생활을 게임화 하였던 심즈는 인공생명알고리즘을 게임에 사용하였다.

많은 몬스터들을 같은 방향으로 움직이는 경우에 플로킹 알고리즘을 사용한다.

게임의 재미를 배가하기 위해서 좀 더 다양한 값들을 선택할 수 있는 퍼지(Fuzzy) 나 퍼지와 유한상태기계의 장점을 결합한 퍼지상태기계(FuSM)를 사용할 수 있다.

이외에도 캐릭터 자체의 능력을 향상시킬 수 있도록 학습시키는 신경망이 있다. 그러나 신경망은 학습시키는 데에 많은 시간이 필요하다는 단점이 있다.

강화학습은 온라인 학습으로 캐릭터의 능력을 학습시키는 데에 적합하다. 또한 아직까지 강화학습을 게임에 적용한 논문은 발표되지 않고 있다.

본 논문에서는 강화학습을 이용하여 슈팅게임에서 캐릭터의 지능이 향상되도록 하였다.

3. 지능형게임캐릭터의 구현

본 논문에서는 적 캐릭터와의 전투를 통해 스스로 학습하는 아군 캐릭터를 구현한다.

3.1 캐릭터의 설계

캐릭터들은 크게 아군캐릭터와 적군캐릭터가 있다.

아군캐릭터가 하나 존재한다. 적군캐릭터의 경우 캐릭터 하나에서 부터 캐릭터 둘, 셋, 넷일 때까지의 네 가지 종류의 적군캐릭터가 존재한다.

아군캐릭터는 적군캐릭터의 공격을 방어하게 된다. 방어의 의미는 실질적으로는 적군캐릭터 또는 적군캐릭터의 총알을 아군캐릭터가 위아래로 움직여 피하는 것이다.

적군캐릭터의 공격은 일직선으로 한다. 적군캐릭터는 스스로의 몸을 움직이거나 총알을 발사하

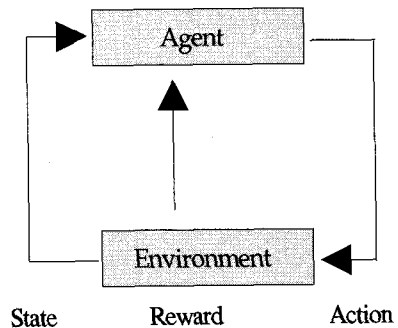
여 아군을 공격한다. 적군캐릭터나 적군캐릭터의 총알이 아군캐릭터와 충돌하지 않도록 아군캐릭터는 적절히 방어하여야 한다.



(그림 2) (아군캐릭터와 적군캐릭터의 움직임)

3.2 강화학습

강화학습이란 많은 상태들의 집합인 환경에서 목적달성을 위한 행위를 수행하여 보상을 받음으로 필요한 행위를 학습하는 것을 말한다.



(그림 3) (강화학습의 모델)

환경에는 목적달성에 필요하거나, 필요하지 않은 많은 상태들이 존재한다. 에이전트는 각각의 상태를 경험하여 목적달성의 경우 환경으로부터 보상을 받게 된다. 목적달성을 할 수 없는 상태인 경우 보상을 받을 수 없다. 그러므로 많은 시행착

오를 겪을 수 있으며, 모든 상태를 경험해 보아야 한다.

강화학습알고리즘으로서 일반적으로 많이 사용되는 Q-learning 에 대해 알아보자.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)) \quad [1]$$

특정상태 s 에서의 행위 a 에 대한 보상 r 을 얻을 수 있도록 식이 주어져 있다.

처음에는 모든 상태에 대하여 보상을 얻기 위해 무작위로 행위를 하여 경험을 쌓게 된다. 그러나 시간이 지나 경험한 상태가 많아질수록 보상을 토대로 불필요한 상태의 경험을 필요치 않는다. 거의 모든 상태를 경험함으로써 학습이 완료되어 지능적으로 동작하게 된다.

적군캐릭터의 경우 보스캐릭터는 고정된 위치에서 총알을 발사한다. 총알은 한발씩 발사하게 되고 목적지에 도달하였을 경우에 또 한발의 총알이 발사된다. 그러나 캐릭터의 애니메이션은 일반게임과 같다. 부하캐릭터의 경우 목적지인 좌측에 도달하였을 경우에 다시 우측에서 등장하게 된다. 이 경우에 무작위로 위치를 선정하여 등장한다.

화면의 크기는 일반게임과 같으나 캐릭터의 이동단위를 감안한다면 가로로 8칸 세로로 5칸의 배열과 같다고 할 수 있다.

아군캐릭터가 적군캐릭터의 총알을 피하기 위하여서는 상 또는 하로 이동하는 방법이 있을 뿐이다.

아군캐릭터, 각각의 적군캐릭터와 총알의 위치는 메모리에 각각 상황별로 저장되며 각각의 상황에서의 학습점수가 저장된다. 아군이 움직일 때에는 각 상황에서의 선택할 수 있는 점수 중 유리한 것을 선택하게 된다. 그러나 학습의 초창기에는 저장된 데이터가 많지 않아 좋은 선택을 못하게 된다.

아군캐릭터가 보스캐릭터의 총알이나 부하캐릭터와 충돌한 경우에는 -100 의 보상이 주어진다.

(표 1) (캐릭터 간 충돌의 경우의 보상점수)

공격주체	대 상	보상점수
적군캐릭터1	아군캐릭터	-100
적군캐릭터2	아군캐릭터	-100
적군캐릭터3	아군캐릭터	-100
적군총알	아군캐릭터	-100

처음에는 학습이 이루어지지 않았기 때문에 아군캐릭터와 적군캐릭터의 충돌은 빈번하게 이루어지게 된다.

그러나 학습이 이루어진 후에는 전혀 충돌이 일어나지 않게 된다.

4. 실험 및 결과

프로그래밍 언어는 C 언어를 이용하였다. Microsoft 사의 Visual C++6.0 을 이용하였고, 컴퓨터는 펜티엄 4 HT 2.60 GHz 이며 메모리는 512 MB 이다. 게임캐릭터를 표현하기 위하여 DirectX 를 이용하였다.

실험은 적군캐릭터가 하나, 둘, 셋, 넷일 때로 구분하여 실험하였다. 하나일 경우에는 보스캐릭터만이 존재하고 여럿일 경우는 부하와 보스캐릭터가 존재한다.

4.1 적군캐릭터가 최대 수 일 경우

하나의 상태공간은 8 × 5 이므로 40개의 격자셀로 구성된다. 아군캐릭터의 관점에서 볼 때 적군캐릭터가 넷일 경우에 고려하여야 할 요소들은 다음과 같다.

- ㉑ 아군캐릭터의 위치
- ㉒ 적군 보스캐릭터 총알의 위치
- ㉓ 적군캐릭터 셋의 위치
- ㉔ 아군캐릭터가 방어를 위해 이동 가능한 공간

보스캐릭터의 위치는 고정되어 있으므로 상태공간을 고려할 필요는 없다. 아군캐릭터의 상태공

간은 40 이므로 나머지 고려하여야 할 요소들의 경우 네 번째 경우를 제외하고는 모두 40 이 된다. 네 번째 경우는 상 또는 하로 이동하므로 2 가 된다. 그러므로 상태공간의 크기는 $40 \times 40 \times 40 \times 40 \times 2$ 로 204,800,000 이 된다.

아군캐릭터는 강화학습이 반영된 캐릭터와 무작위로 움직이는 아군캐릭터로 실험하였다. 적군 캐릭터인 보스캐릭터의 경우 일정한 위치에서 총알을 발사하게 된다.

(표 2) 제안하는 알고리즘

```

procedure Q_Learning
{
    Find Max from Q_Table
    Select character action from QTable

    Move player character
    Move enemy character

    if (collide)
        Generate reward

    Update QTable
}
    
```

(표 3) 적군캐릭터가 최대 수일 경우의 실험결과

	강화학습 후				Random 후			
	적1	적2	적3	적4	적1	적2	적3	적4
15,000	128	106	125	121	490	448	489	484
12,000	128	106	125	121	403	358	400	381
9,000	124	104	124	120	313	285	311	275
6,000	113	94	119	115	214	182	208	187
3,000	76	67	85	77	106	82	110	96

차트와 그림에서 보면 알 수 있듯이 캐릭터는 강화학습을 하는 캐릭터와 무작위로 움직이는 두 가지 형태의 캐릭터가 보스캐릭터와 대전함을 알 수 있다.

강화학습을 하는 경우에 방어율을 본다면 아군 캐릭터는 보스캐릭터의 총알을 106회 에서 128회 이상 맞은 후에는 더 이상 맞지 않는다는 것을 알 수 있다.

12,000 번 움직인 경우에 위와 같은 충돌횟수

가 생겼다. 그 이후로는 표와 같이 충돌회수는 그대로이므로 한 번도 충돌이 추가되지 않음을 알 수 있다.

무작위 캐릭터와 강화학습 된 캐릭터와의 충돌 횟수를 비교해 보면 초반 3,000 회전에서는 비슷한 양상을 보이거나 학습이 더 이루어지는 6,000 회에서는 이미 2배 가까운 차이를 보이고, 9,000 회, 12,000 회에서는 서너 배의 차이를 15,000 회에서 5배 가까이 차이를 벌리는 것을 알 수 있다.

더욱 중요한 것은 계속 실험을 할 경우에 무작위 캐릭터의 경우에 일정한 비율로 계속 충돌이 일어나지만 강화학습 된 캐릭터의 경우는 더 이상 단 한 번의 충돌도 일어나지 않는 데에 있다.

또한 적군캐릭터가 넷일 경우는 상하의 격자공간이 5 이므로 적군캐릭터를 제외한 피할 수 있는 공간이 한 군데 만이 존재한다. 그러므로 그곳으로 정확히 피해야만 충돌을 방지할 수 있다. 강화학습 된 캐릭터는 단 한군데만의 피할 수 있는 공간을 정확히 찾아 대피를 하는 것이다.

4.2 적군캐릭터가 최대 수가 아닐 경우

적군캐릭터 하나가 40 의 상태공간이므로 아군 캐릭터와 방어를 위한 이동공간을 고려한다면 적군캐릭터가 하나일 경우에는 $40 \times 40 \times 2$ 이며 적군캐릭터가 하나씩 늘어날 때마다 상태공간은 40 을 곱하면 된다.

고려하여야 할 요소는 적군캐릭터가 셋일 경우 다음과 같다.

- ㉠ 아군캐릭터의 위치
- ㉡ 적군 보스캐릭터 총알의 위치
- ㉢ 적군캐릭터 들의 위치
- ㉣ 아군캐릭터가 방어를 위해 이동 가능한 공간

아군캐릭터는 강화학습이 반영된 캐릭터와 무작위로 움직이는 아군캐릭터로 실험하였다. 적군 캐릭터가 하나일 경우에 다음과 같은 결과를 보인다.

(표 4) 적군캐릭터가 하나일 경우의 실험결과

	강화학습 후		Random 후	
	아군. 적 총알	아군. 부하캐릭터	아군. 적 총알	아군. 부하캐릭터
500	2		14	
400	2		13	
300	2		7	
200	2		5	
100	2		2	

상태공간이 $40 \times 40 \times 2$ 인 3,200 개에 불과하므로, 학습이 100회 이루어진 경우에 방어가 가능하다. 총돌이 2회 이루어진 다음에는 더 이상의 총돌이 없음을 알 수 있다.

무작위 캐릭터의 경우 지속적인 총돌이 발생한다.

(표 5) 적군캐릭터가 둘일 경우의 실험결과

	강화학습 후		Random 후	
	아군. 적 총알	아군. 부하캐릭터	아군. 적 총알	아군. 부하캐릭터
5,000	10	10	150	169
4,000	10	10	119	134
3,000	10	10	87	102
2,000	10	10	54	71
1,000	9	10	29	34

적군캐릭터가 둘일 경우에 위와 같은 결과를 보인다.

상태공간이 $40 \times 40 \times 40 \times 2$ 인 128,000 개이며, 학습이 2,000회 이루어진 경우에 완전한 방어가 가능하다. 총돌이 각 10회 이루어진 다음에는 더 이상의 총돌이 없음을 알 수 있다. 무작위 캐릭터의 경우 지속적인 총돌이 발생한다.

(표 6) (적군캐릭터가 셋일 경우의 실험결과)

	강화학습 후			Random 후		
	적1	적2	적3	적1	적2	적3
10,000	40	40	40	307	317	294
8,000	40	40	40	247	253	237
6,000	40	40	40	186	188	176
4,000	40	40	40	118	126	125
2,000	39	35	39	58	70	55

적군캐릭터가 셋일 경우에 위와 같은 결과를 보인다.

상태공간이 $40 \times 40 \times 40 \times 40 \times 2$ 인 5,120,000 개이며, 학습이 4,000회 이루어진 경우에 완전한 방어가 가능하다. 총돌이 각 40회 이루어진 다음에는 더 이상의 총돌이 없음을 알 수 있다. 무작위 캐릭터의 경우 지속적인 총돌이 발생한다.

상태공간이 작을수록 작은 횟수만큼의 총돌이 일어나는 것을 알 수 있다. 캐릭터의 움직임이 단순하므로 비교적 적은 횟수의 총돌로서 학습이 이루어지는 것을 알 수 있다.

5. 결 론

3D 게임과 온라인게임이 완전히 자리 잡은 요즘 게임시나리오가 게임의 흥미를 이끌어내지만 캐릭터의 자동화문제는 여전히 남은 숙제라고 할 수 있다. 캐릭터의 자동화로 인해 게임의 재미가 더할 수 있고, 그것은 게임프로그래머의 몫이다.

그 동안 여러 가지 인공지능 알고리즘이 연구되고, 사용되어왔지만 강화학습은 게임분야에서 연구되지 않았다.

본 논문에서는 강화학습 알고리즘을 이용하여 아군캐릭터와 적군캐릭터가 전투하는 상황에서 적군캐릭터가 아군캐릭터를 공격할 때에 음의 보상 값을 부여하여 게임캐릭터가 학습하게 하여 지능적으로 움직이게 하였다.

구현된 게임캐릭터가 지능적으로 잘 움직이는지 확인하기 위해, 슈팅게임을 제작하여 강화학습으로 움직이는 아군캐릭터와 적군캐릭터와 전투를 하게 하였다. 또한 무작위로 움직이는 아군캐릭터와 적군캐릭터와 전투를 하게 하였다.

실험결과 임의로 움직이는 캐릭터보다 강화학습한 게임캐릭터가 일정한 수의 총돌이 일어난 후, 더 이상의 총돌이 없는 완전한 방어가 이루어지는 것을 알 수 있다.

적군캐릭터를 하나에서 방어가 가능한 최대수의 적군캐릭터로 실험하였을 때, 상태공간이 늘어날 때마다 좀 더 많은 시행착오 후에 학습이 이루어지는 것을 확인하였다.

참 고 문 헌

- [1] Richard Sutton, Andrew G. Barto, "Reinforcement Learning :An Introduction", MIT Press, Cambridge, MA, 1998.
- [2] Steve Woodcock, "Game AI : The State of the Industry", Game Developer Magazine, 2000.
- [3] Steve Rabin, AI Game Programming Wisdom 2, Charles River Media, 2003
- [4] Steve Rabin, AI Game Programming Wisdom, Charles River Media, 2002
- [5] 신용우, 게임프로그래밍 길잡이, 대림출판, 2002
- [6] 신용우, 인공지능 게임프로그래밍, 대림출판, 2004
- [7] Andrew Kirmse, Game Programming Gems 4, Delmar Thomson Learning, 2004.
- [8] Mark Deloura, Game Programming Gems 3, Charles River Media, 2002.
- [9] Mark Deloura, Game Programming Gems 2, Charles River Media, 2001.

○ 저 자 소 개 ○



신 용 우(Yong-Woo Shin)

2004년 경희대학교 컴퓨터공학과 지능시스템전공 (박사수료)
1990년~1993년 (주)진도 전산부
1994년~1995년 프리랜서 게임프로그래머
1996년~2000년 LG데이콤 인터넷사업단
2000년~현재 동아방송예술대학 게임애니메이션계열 조교수
현재 한국게임학회 이사
현재 게임물등급위원회 자문위원
관심분야 : 게임프로그래밍, 데이터마이닝, 인공지능
E-mail : ywshin@dima.ac.kr