학습기법을 이용한 멀티 에이전트 시스템 자동 조정 모델

이 말 레[†] · 김 상 근[††]

요 약

멀티 에이전트 시스템은 분산적이고 개방적인 인터넷 환경에 잘 부합된다. 멀티 에이전트 시스템에서는 각 에이전트들이 자신의 목적을 위해 행동하기 때문에 에이전트간 충돌이 발생하는 경우에 조정을 통해 협력할 수 있어야 한다. 그러나 기존의 멀티 에이전트 시스템에서의 에이전트 간 협력 방법에 관한 연구 방법들은 동적 환경에서 서로 다른 목적을 갖는 에이전트간의 협동 문제를 올바로 해결할 수 없다는 문제가 있었다. 본 논문에서는 신경망과 강화학습을 이용하여 목적 패턴을 정확히 결정할 수 없는 복잡하고 동적인 환경하에서 멀티 에이전트의 자동 조정 모델을 제안한다. 이를 위해 복잡한 환경과 다양한 행동을 갖는 멀티 에이전트간의 경쟁 실험을 통해 멀티 에이전트들의 행동의 영향을 분석 평가하여 제안한 방법이 타당함을 보였다.

The Automatic Coordination Model for Multi-Agent System Using Learning Method

Mal Rey Lee[†] · Sang Geun Kim[††]

ABSTRACT

Multi-agent system fits to the distributed and open internet environments. In a multi-agent system, agents must cooperate with each other through a coordination procedure, when the conflicts between agents arise. Where those are caused by the point that each action acts for a purpose separately without coordination. But previous researches for coordination methods in multi-agent system have a deficiency that they cannot solve correctly the cooperation problem between agents, which have different goals in dynamic environment. In this paper, we suggest the automatic coordination model for multi-agent system using neural network and reinforcement learning in dynamic environment. We have competitive experiment between multi-agents that have complexity environment and diverse activity. And we analysis and evaluate effect of activity of multi-agents. The results show that the proposed method is proper.

키워드 : Multi-Agent, Automatic coordination, Learning method

1. 서 론

분산 개방 시스템인 인터넷은 다양한 종류의 정보가 서로 다른 기관이나 개인들에 의해 생성되어 지리적으로 광범위한 영역에 분산되어 있을 뿐 아니라, 정보자원이나 통신링크, 에이전트들의 생성과 소멸은 예측 할 수 없다. 이러한 상황에서는 지식과 컴퓨팅 자원, 능력이 제한되어 있는 단일 에이전트를 이용한 문제 해결 방법에 한계가 있다[1]. 이와 같은 환경에서의 문제를 해결하기 위하여 최근 널리 사용되는 방법이 멀티 에이전트 시스템이다. 멀티 에이전트 시스템에서는 에이전트가 대등하게 연결되어 있어 서로 정보를 주고받으며, 또한 서로간에 조정할 수 있는 능력을 가지고 있다.

멀티 에이전트 시스템에서 가장 중심적인 연구 과제는 에이전트간의 조정, 협력에 대한 것이다. 에이전트들은 자신에게 할당된 문제를 풀어 나가는 과정에서 부분 관찰(local view), 다중목표(multiple goal), 분산된 정보 등의 제약 때문에 혼란을 겪거나, 에이전트간의 충돌을 일으킬 수 있다. 또한, 에이전트들이 해결하고자 하는 문제에 대한 제약을 만나거나, 개개의 에이전트가 가지고 있는 각기 다른 능력과 특별한 지식들을 서로 공유할 필요가 있을 경우 에이전트간의 협력이나 조정이 필요하게 되며, 다른 에이전트의 행동에 따라 자신의 행동이 결정될 경우나, 전체 시스템의 효율을 높이기 위해서도 조정이나 협력이 필요하게 된다. 그러나 이를 위하여 제안된 여러 방법들은 각 에이전트들의 역할이 한 번 주어지면 고정되어 동적으로 변하는 개방 환경에의 적용에 적합하지 않다는 문제점을 가지고 있었다.

본 연구에서는 이런 문제점을 극복할 수 있는 에이전트

※ 본 논문은 여수대학교의 2000년 학술 연구지원비에 의하여 연구되었음.
† 정 회 원 : 국립여수대학교 멀티미디어학부 교수
†† 정 회 원 : 성결대학교 컴퓨터학부 교수
논문접수 : 2001년 10월 4일, 심사완료 : 2001년 12월 24일

사이의 역할 조정 모델을 제안한다. 이 모델에서는 에이전트들 사이에 역할 충돌이 일어나는 경우에 강화학습을 이용하여 자신의 역할을 수정한다. 본 논문의 구성은 다음과 같다. 먼저 2장에서는 기존의 에이전트간 협력 방법에 관한 연구 방법들과 그 제약점들을 살펴보고, 3장에서는 동적인 환경에 적용 가능한 역할 조정 모델을 제안한다. 4장에서는 제안한 모델을 인공생명 경쟁 문제에 적용한 실험 결과를 살펴보고 5장에서 결론을 내린다.

2. 기존의 에이전트간 협력방법

기존의 멀티 에이전트 시스템에 대한 연구에서 제안하고 있는 에이전트간의 조정 기술에는, 에이전트의 유기적인 구조를 설계하는 방법, 에이전트간의 계약을 이용하는 방법, 멀티 에이전트 계획을 이용하는 방법, 협상을 이용하는 방법 등이 있다. 우선 에이전트 유기적인 구조를 구성하는 방법은 가장 간단한 조정방법[5]으로, 에이전트들은 master/slave 혹은 client/server의 구조를 가지도록 설계한 다음, 에이전트간의 계층적인 관계를 통해 협력과 조정을 수행하게 된다. 이 방법에서 에이전트들은 주로 흑판 구조[6]를 이용해서 서로간의 통신을 하게 된다. 그러나 이 경우에는 에이전트들의 구조로 인한 추가적인 제어가 필요하며, 서로간의 통신을 위해서 흑판을 사용하기 때문에 병목 현상을 발생시키는 등, 멀티 에이전트 시스템의 이점을 저하시킬 수 있으며 에이전트들이 단순한 구조를 가져야 한다는 제약을 가지게 된다. 제약[7]을 통한 에이전트들간의 조정은 주로 contract net protocol(CNP)을 사용하며, 이는 분산된 환경에서 에이전트에 대한 자원 할당이나 문제의 분배에 주로 사용된다. 우선 관리자로 설정된 에이전트는 자신에게 할당된 문제를 작은 부분 문제들로 나누어 이를 수행할 다른 에이전트를 찾게 된다. 그럼 관리자가 요구하는 문제를 수행할 수 있는 에이전트는 계약자가 되어서 자신이 선택한 부분 문제를 해결하게 된다. 이런 과정이 재귀적으로 이루어지며 계약자가 되었던 에이전트가 다시 관리자가 되어 문제를 새로이 다른 에이전트에게 할당하게 된다. 에이전트간의 계약에 의한 조정에서는 에이전트간의 계층적인 관계를 자동으로 만들어 내며, 수행할 문제에 대한 자원의 동적 할당이 가능하고 자연스러운 부하조절이 가능하다는 장점을 지니고 있다. 그러나 에이전트간의 조정이 수동적이며, 조정을 위한 통신이 이루어지기 어려울 경우 전체적인 시스템 자체에 대한 문제를 불러일으킬 수 있다는 단점을 가지고 있다.

이 방법에는 게임이론을 기반으로 하는 협상[9], 계획에 기반을 둔 협상[10], 휴리스틱에 의한 협상[8, 11] 등의 방법이 있다. 게임이론을 이용하는 경우 수익 행렬을 사용하여, 에이전트 행동에 따른 이익에 따라 에이전트의 행동을 조정하는 방법을 사용한다. 이 경우에는 에이전트들이 공유해야 하는 서로에 대한 완전한 지식을 가정하기 때문에 현실세계의 문제에 적용하기 적합하지 않을 뿐만 아니라, 셋 이상의 에이전트의 경우에는 게임이론을 손쉽게 확장할 수 없다는 단점을 가지고 있다. 계획에 기반한 협상의 경우에는 계획에 대한 사전지식에 따라서 에이전트들이 자동적으로 협상을 수행하게 된다. 우선 에이전트들이 자동적으로 협상을 수행하게 된다. 우선 에이전트는 각자의 행동을 계획하고, 그 계획을 분리된 조정 에이전트가 에이전트의 상태나, 메시지 형태, 대화 방법을 이용해서 협상을 수행하게 된다.

그 밖에도 에이전트간 상호 메시지 전달 없이 사전 협약을 통한 협력 작업을 수행하는 방법[12]이나, 개미나 꿀벌 등 곤충들의 행동 유형을 분석하여 에이전트의 행동 결정에 적용하는 방법들이 있었으나, 모두 각 에이전트들에게 한번 할당된 목표는 고정되어 변하지 않았기 때문에 에이전트간의 역할에 충돌이 발생하는 경우 이를 해결할 수 없다는 문제를 가지고 있었다.

3. 자동 조정 모델

본 연구에서 제안한 조정 모델은 신경망의 입력으로 현재 에이전트가 위치한 곳을 중심으로 한 5*5 크기 격자 환경의 상태가 주어진다. 하나의 격자는 {아무도 없음, 자기편 있음, 상대편 있음}과 같은 3가지의 상태가 가능하므로 신경망이 받을 수 있는 서로 다른 입력 패턴의 수는 대략 3^{24}개 정도로 매우 많다. 또한, 신경망에 입력되는 패턴은 고정된 정적 패턴이 아닌 수시로 변하는 동적 패턴이므로 [입력패턴, 목적패턴] 형태의 학습 패턴을 정하는 것은 비 현실적이다. 따라서, 학습패턴을 이용한 감독자 학습을 수행할 수 없으며 본 조정모델에서는 대안으로 조정함수를 이용하여 강화 신호를 계산하는 강화 학습 모델을 사용한다. 입력 패턴에 대한 목적 패턴을 정해 놓지 않고 이전에 자신이 취했던 행동으로 인해 발생한 환경의 변화를 인지하고 행동의 유용성을 평가하여 조정 모델에 따른 강화 신호를 결정한다. 강화 신호는 인공 유기체 신경망 학습의 초기 오차값으로 입력되며, 은닉층과 입력층으로 전파된다.

목적 패턴의 역할을 대체하는 조정 함수는 멀티 에이전트 학습 행동의 방향과 효과를 좌우한다. 조정함수는 멀티 에이전트의 창발성 원리에 따라 시스템의 하부 구조인 멀티 에이전트의 시점에서 설계한다. 저 수준에서 조정 함수를 적용하더라도 창발성의 원리를 통해 복잡한 집단 학습의 총제적 양상이 나타난다. 본 조정 모델에서는 두 가지의 상이한 조정 모델을 적용하였다. 각 조정 모델은 행동의 강화와 억제의 두 가지 측면에서 설계되었다. 멀티 에이전트의 경쟁 환경에서 유리하다고 판단되는 행동은 (+)학습을 하여 강화했으며, 불리하다고 판단되는 행동은 (-)학습을 하여 억제했다. 멀티 에이전트의 행동 판단은 학습 행동 적용의 초기 단계로서 전문가의 경험에 의해 판단된다. 또한, 행동의 강화와 억제 효과에 있어 그 강도를 달리하기 위해

여러 가지 이산적인 값을 이용하였다. 멀티 에이전트가 환경으로부터 두 번의 공격을 받게 되는 행동을 하였다면, 한 번의 공격을 받게 되는 행동 보다 두 배 크기의 억제 학습을 수행하였다.

멀티 에이전트는 자신으로부터 시작한 "한 스텝"이 모두 끝났을 때 이전 스텝의 행동에 대한 강화 신호를 학습한다. 이를 통해 다른 유기체와 상호 작용하여 받게 되는 강화 신호를 모든 유기체에 대해 집단 내에서의 순서와 관계없이 평등하게 학습한다.

조정모델 1에서의 멀티 에이전트는 자신을 제외한 모든 것을 환경으로 인식한다. 따라서, 개체는 환경에 대한 공격 성공 시 자기편, 상대편의 구분 없이 (+2) 강화 신호를 받으며, 환경으로부터 공격을 받을 때는 (-)의 강화 신호를 받는다. 조정모델 1은 멀티 에이전트 유기체의 움직임을 활성화하고 멀티 에이전트의 공격성을 유도함으로써 보다 많은 경쟁 전략들이 출현할 수 있도록 하는 것이 목적이다. 또한 이 모델은 개체의 시점에서 자신에게 주어지는 손익을 따져 강화 신호를 계산한다. 조정모델 1의 기본 값은 (+1)로서 멀티 에이전트가 환경과의 상호 작용을 하지 않는 경우에는 행동을 강화시킨다.

이는 멀티 에이전트의 단순한 이동이나, 움직이지 않는 경우까지 행동 강화 학습을 수행한다. 반면에 환경으로부터 공격을 받는 경우에는 피공격 회수에 따라 행동 억제 학습 강도를 달리한다. 행동 강화 학습과 행동 억제 학습 사이의 균형을 위해 공격 성공시에는 (+2)(기본값 + 행동 판단 값)의 다소 높은 강화 신호를 제공한다. 또한, 공격 실패 행동은 기본값에 (-1)을 더하여 아무런 학습도 하지 않는다.

이동과 공격은 동시에 성립할 수 없지만 이동과 피공격, 공격과 피공격은 동시에 성립될 수 있다. 예를 들어, 멀티 에이전트가 앞으로 한번 이동한 후 상대편으로부터 두번의 공격을 받으면 최종적으로(-1)(= 1 + 0 - 2)의 강화 신호를 받는다. 조정모델 1의 알고리즘은 (그림 1)과 같다.

```
(단계 1) : 강화 신호를 기본값으로 초기화
(단계 2) : if (공격 성공) then 강화신호 = 강화신호 + 1
(단계 3) : if (공격 실패) then 강화신호 = 강화신호 - 1
(단계 4) : if (피공격)     then 강화신호 = 강화신호 - (피공격 회수)
```

(그림 1) 조정모델 1 알고리즘

조정모델 2에서의 멀티 에이전트는 조정모델 1에서 자신을 제외한 모든 것을 하나의 환경으로 인식하였던 데서 발전하여, 환경을 자기편과 상대편으로 분리한다. 따라서, 환경에 대한 공격 성공시 상대편인 경우는 (+3) 강화신호를, 자기편인 경우에는 (-1) 강화신호를 받는다. 이는 상대편 공격 성공 행동의 강화 신호와 피공격 행동의 행동 억제 신호 값 사이의 균형을 위한 것이다. 또한 조정모델 2에서는 불필요한 학습의 원인이 되었던 강화 신호의 기본값을 (0)으로 변경하였으며, 이동 실패나 공격 실패 행동의 억제 신호로서 (-1)의 강화 신호를 주었다. 또한, 환경의 공격을 유발하는 행동에 대해서는 (-피공격 회수) 크기의 억제 신호를 발생시켜 억제 효과의 강도를 조절한다.

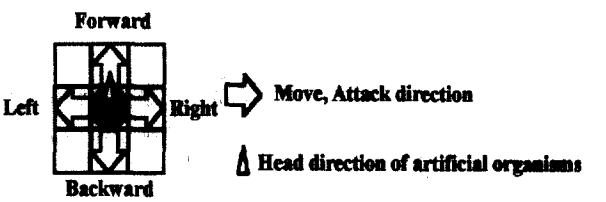
조정모델 2는 이동 실패 행동과 공격 실패 행동에 대한 억제 학습을 수행함으로써 에이전트가 좀더 합리적이고, 정확한 행동을 할 수 있도록 유도한다. 이와 아울러 단순한 이동이나 정지 행동에 대한 불필요한 강화학습을 배제함으로써 멀티 에이전트의 공격 행동을 유도한다. 조정모델 2 알고리즘은 (그림 2)와 같다

```
(단계 1) : 강화 신호를 기본값으로 초기화
(단계 2) : if (상대편 공격 성공) then 강화신호 = 강화신호 + 3
(단계 3) : if (자기편 공격 성공) then 강화신호 = 강화신호 - 1
(단계 4) : if (공격실패) then 강화신호 = 강화신호 - 1
(단계 5) : if (이동실패) then 강화신호 = 강화신호 - 1
(단계 6) : if (피공격) then 강화신호 = 강화신호 - (피공격 회수)
```

(그림 2) 조정모델 2 알고리즘

3.1 학습 기법을 이용한 자동 조정 방법

경쟁 환경은 12*5 크기의 격자 형태를 이루며 상하좌우의 경계가 없이 연속적이다. 경쟁하는 두 집단은 초기화 시에 경쟁 환경 내에서 임의로 왼쪽 혹은 오른쪽에 위치한다. 만일 멀티 에이전트의 움직임이 경쟁 환경 안에서 동서남북과 같은 절대적인 방향성을 갖는다면 집단의 초기 위치에 따라 서로 다른 결과를 보일 수 있다. 그러므로 각 에이전트는 머리 방향을 두고 (그림 3)와 같이 자신의 머리 방향을 기준으로 하여 전후좌우의 상대적인 방향성을 갖게 하였다. 경쟁 환경 내에는 경쟁을 위하여 선택된 두 집단만이 존재하며 장애물이나 먹이 같은 것은 존재하지 않는다.



(그림 3) 에이전트의 좌표계

3.2 경쟁 규칙

멀티 에이전트는 경쟁 환경 내에서 전후좌우로 이동하거나 공격하여 상대편 집단의 멀티 에이전트와 경쟁한다. 멀티 에이전트를 움직이는 순서는 두 집단에서 교대로 하나씩 움직인다. 경쟁의 초기 배치 상태에서는 A집단부터 움직이기 시작한다. 두 집단 A, B에 속한 모든 멀티 에이전트들이 한 번씩 움직이는 것을 '한 스텝'으로 정의하며 경쟁은 일정 스텝을 반복한 후 각 에이전트 집단의 적합도를 계산하여 승패를 결정한다. 멀티 에이전트를 단 한 번의 공격으로 죽게 하는 것보다 처음에 일정한 생명력을 주고 공격을 받을 때마다 일정한 생명력이 감소하여 생명력이 0이

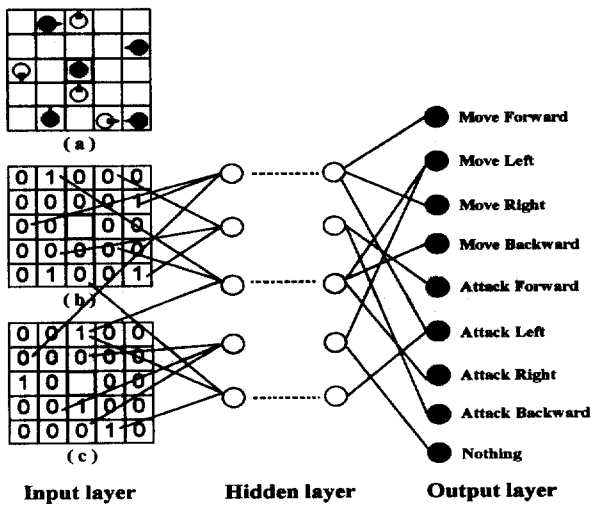
되었을 때 죽도록 하였다. 이러한 방법은 우연을 배제 하는 것 보다 현실적이고 합리적인 경쟁 매커니즘을 제공한다.

본 조정모델 에이전트의 유전 정보를 이용하여 신경망을 구성하고, 이를 멀티 에이전트의 조정 함수로 이용하였다. 멀티 에이전트의 신경망은 입력 노드 48개, 은닉 노드 71개, 출력 노드 9개로 전체 128개의 노드로 구성된다. 신경망의 입력으로는 현재 에이전트가 위치한 곳을 중심으로 5*5 크기의 격자 환경이 주어진다. 멀티 에이전트의 신경망은 자기 위치를 제외한 24곳에 대하여 상대편과 자기편의 유무를 1(존재함) 또는 0(존재하지 않음)의 형태로 입력받는다(그림 4). 이와 같이 비교적 넓은 범위의 주변 환경을 인식하는 것은 다른 응용 영역에 적용하는데 보다 일반적인 방법이 된다.

〈표 1〉 신경망의 구성 요소

Input node	24개의 주변 격자에 대한 자신편의 존재 여부 24개의 주변 격자에 대한 상대편의 존재 여부
Hidden node	71개로 내부적으로 개체의 기억장소로 사용
Output node	4개 방향에 대한 이동 행동 1개의 움직이지 않음 4개 방향에 대한 공격 행동

멀티 에이전트의 신경망은 9개의 출력 노드를 가지며 각 출력 노드는 나열된 멀티 에이전트가 취할 수 있는 기본 행동을 의미한다.



(그림 4) 멀티 에이전트 신경망 구성

3.3 멀티에이전트 학습

경쟁은 승자 결정, 패자 결정 두 번의 토너먼트를 통해 진행된다. 각 토너먼트마다 유전자 풀(Gene Pool)로부터 경쟁에 참여하는 에이전트 집단을 임의로 선택한 후 토너먼트를 거쳐 패자를 제거하고 승자의 자손으로 대체한다.

본 시뮬레이션은 진화를 위한 방법으로써 정상상태 유전 알고리즘(Steady-State Genetic Algorithm : SSGA) 기법을 적용한다. 정상상태 유전 알고리즘은 전체 개체군에 대한 적합도를 유지하지 않으며, 적합도로는 실제적인 경쟁 결과를 이용한다. 동적 경쟁 환경 하에서의 적합도는 일정한 수치 형태로 표현될 수 없으며, 오직 실제 경쟁의 결과만을 판단 기준으로 한다.

에이전트 집단 경쟁의 승패는 경쟁을 종료한 후 살아남은 에이전트의 수로 결정한다. 만일 A집단이 B집단에 비해 더 많은 에이전트가 살아 남았다면 A집단의 승리이다. 두 개의 집단에서 살아남은 에이전트의 수가 동일할 경우에는 각 집단 내 에이전트 생명력의 합이 큰 쪽을 승리로 결정한다.

4. 멀티에이전트 시뮬레이션

본 논문의 멀티에이전트 시뮬레이션에서 한 집단에 포함된 개체의 수는 1000개, 돌연변이율은 3%로 고정하였다. 시뮬레이션은 <표 2>와 같이 학습 기능의 유무, 학습률, 조정 모델을 달리하여 5가지의 실험 모델에 대해 수행하였다.

〈표 2〉 시뮬레이션 모델

실험 모델	학습 기능	조정 모델	학 습 률
A	없 음	-	-
B	있 음	1	2
C	있 음	1	4
D	있 음	2	2
E	있 음	2	4

A 실험 모델은 이전 실험에서 사용하였던 시뮬레이션을 다시 수행하여 얻은 결과로서, 본 논문에서 제안하는 시뮬레이션 방법의 비교 대상이 된다.

조정 모델은 에이전트의 신경망 학습을 위한 강화 신호를 계산하는 방법으로서, 조정 모델 1은 이동 행동과 공격 행동의 강화를 목표로 하였으며, 조정 모델 2는 이동 실패 행동과 공격 실패 행동에 대한 억제와 이를 통한 보다 정확한 행동의 강화를 목표로 하였다. 또한, 학습률은 신경망 학습시에 조정 모델을 통해 주어진 강화신호의 학습 강도를 조절한다. 학습률 4는 학습률 2에 비해 2배의 강도로 강화신호를 학습한다. 이를 통해 에이전트의 신경망은 강화신호의 종류와 크기의 영향을 보다 크게 받는다. 학습 률은 같은 조건의 실험 환경에 [1, 2, 4, 5. 7]의 학습률을 미리 적용하여 보고, 그 중 효과가 좋은 학습률 [2, 4]를 선택하여 사용하였다.

시뮬레이션 수행 중에 100세대 단위로 행동 자료를 누적, 파일로 기록한다. 행동 자료는 이동 성공, 이동 실패, 총공격, 상대편 공격, 자기편 공격, 공격 실패의 6가지 항목에 대한 행동 횟수이며, 이는 100세대 동안의 진화에 대한 분석 자료를 제공한다. 또한, 이 자료는 에이전트 집단 전체의 성격에 대한 추정을 가능케 한다. 각 실험 모델에서 얻은 행동 자료를 비교, 분석함으로써 해당 모델의 에이전트 집단이 진화한 방향과 속도를 비교한다. 이를 통해 에이전

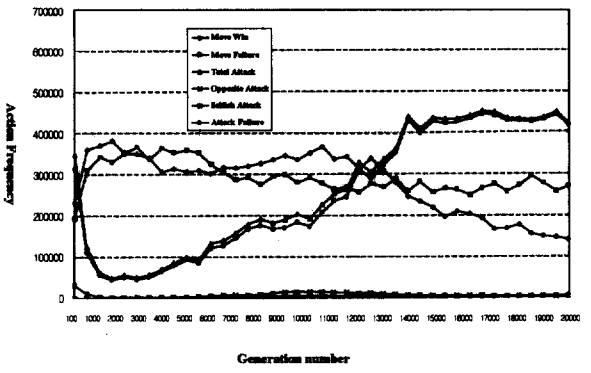
트 수준의 학습 기능이 전체 집단의 진화에 미친 영향을 평가한다.

100세대 단위로 기록된 행동 자료의 비교, 분석만으로는 에이전트의 학습 행동이 실제적으로 경쟁 전략을 향상시켰는지 확인할 수 없다. 이에 각 실험 모델에서 얻은 특정 세대의 에이전트 집단들 사이의 실제 경쟁을 통해 각 방법의 비교우위를 판정한다.

4.1 결과 분석 및 평가

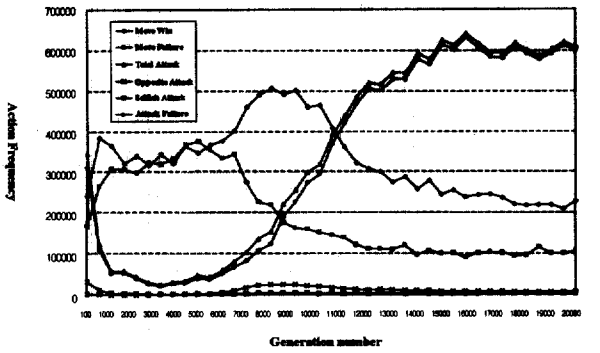
각 실험 모델 별로 전체적으로는 동일한 진화의 방향을 나타내었다. 진화가 계속 진행되면서 집단의 전체적인 이동성은 줄어들었으며, 그와 함께 빈자리에 대한 공격 회수가 증가하였다. 따라서, 집단 전체적으로 수비형 전략이 우세하게 되었다. 또한, 각 실험 모델들은 공통적으로 자기편 공격 행동의 억제 기능에 대해 빠른 학습 효과를 보였으며, 진화 초기에는 모든 집단에서 움직임이 빠른 속도로 증가하였다. 이는 진화의 초기에는 이동을 함으로써 자기편 공격 성공 회수를 줄일 수 있는 집단이 보다 유리하기 때문이다.

A, B, C실험 모델은 기본적으로 비슷한 양상의 그래프를 보여주었다. 반면에 D, E실험 모델은 앞의 세 가지 모델과는 좀 다른 형태의 그래프를 보여주었다.



(그림 5) 학습을 하지 않은 경우(A모델)

A실험 모델은 에이전트의 진화에 개체 수준의 학습 기능을 사용하지 않았다. 이는 이전 실험에서 사용하였던 시뮬레이션을 다시 수행하여 얻은 결과이며 비교 대상으로 사용된다. 진화의 초기에는 급격한 이동성이 증가하였고, 공격성이 감소하였다. 진화가 진행되어 3000세대 근처에 이르렀을 때부터 총공격의 횟수가 증가하기 시작하여 14000세대까지 지속적으로 증가하였다. 총공격의 횟수가 100000~400000구간을 통과하는데 걸린 시간은 약 8000세대이다. 진화 전반에 걸쳐 아무런 에이전트도 존재하지 않는 빈 공간에 대한 공격 행동인 공격 실패 행동의 횟수가 총공격의 횟수를 주도하였다. 이는 경쟁 전략의 전체적인 진화 방향이 공격 전략보다는 수비 전략임을 의미한다.



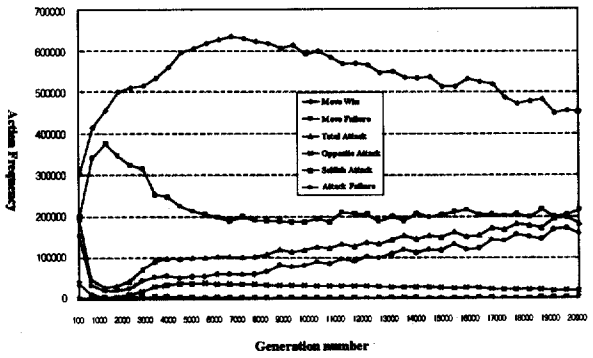
(그림 6) 조정모델 1 학습률 2인 경우(B모델)

B실험 모델의 경우 조정 모델 1, 학습률 2를 사용하였다. 학습 행동을 사용하지 않은 경우와 같이 상대편에 대한 공격이 어느 특정 세대 부근에서만 활성화되었을 뿐, 공격 성공 행동의 두드러진 향상은 눈에 띄지 않았다. 진화의 초기에서 이동성이 급격하게 증가하였으며, 그에 비례하여 공격성은 눈에 띄게 줄어들었다. 공격 행동은 5000세대 이후부터 증가하기 시작하며, 총공격 회수가 약 3000세대 만에 100000~400000빈도에 도달하였다. 이는 학습 행동을 포함하지 않는 A모델이 약 8000세대 정도 걸린 것과 비교하면 진화 속도가 약 267%정도 향상되었다. 이는 에이전트의 학습 행동을 통해 우성 인자가 집단 내에 빠르게 전파되고 있음을 의미한다. 세대는 시뮬레이션에서 사용하는 논리적 시간의 개념으로서 한 세대는 같은 시간의 계산 시간을 갖는다고 가정한다. 공격성이 증가하는 부분은 5000세대 정도로 A실험 모델의 3000세대에 비해 늦어진 것은 불필요한 학습으로 인한 진화의 지연 현상으로 생각된다. 진화의 후기에서는 급격히 증가한 공격 실패 행동이 일정한 수준을 유지하였으며, 에이전트의 이동성이 줄어들었다.

C실험 모델의 경우 조정모델 1,학습률 4를 이용하였다. A, B실험 모델에 비해 C실험 모델의 특징은 공격성이 12000세대 근처에서 매우 늦게 나타난다는 점이다. 이는 앞에서도 잠시 언급했던 바와 같이 불필요한 단순 이동 행동에 대해서 (+)학습을 하기 때문이다. C실험 모델 역시 다른 모델들과 마찬가지로 진화의 초기에는 이동성이 증가하고, 공격성이 감소하였다. 이는 자기편에 대한 공격이 감소하고, 주변 지역에 대한 탐색이 증가하기 때문이다. 진화가 진행됨에 따라 자신 편에 대한 공격 회수는 급격히 감소했으며, 이후 계속 낮은 수준을 유지하였다. C실험 모델 역시 B모델과 마찬가지로 총공격 회수가 100000~400000구간을 통과하는 데에 약 3000세대 정도가 소요되었다.

D실험 모델의 경우에는 앞의 A, B, C실험 모델과는 다른 형태의 행동 그래프를 보여준다. D실험 모델에서는 조정 모델 2를 사용하였다. 조정 모델 2는 이동 실패와 공격 실패 행동에 대해 (-)학습을 수행하며, 단순 이동 행동에 대한 학습을 하지 않는다. 또한, 자기편과 상대편에 대한 공격을 구분할 수 있는 지각 능력을 부여하였다. 결과적으로

이동 실패와 공격 실패 행동은 이전 모델들에 비해서 뚜렷이 억제되는 현상을 보였으며, 상대편에 대한 공격 회수가 높은 수준에서 일정하게 유지되었다. 공격 행동이 억제됨으로 해서 진화 후기까지 상대적으로 이동 행동이 높은 빈도를 차지한다. 하지만, 전체적인 진화의 방향은 이전 모델들과 같이 이동성이 점차 줄어들었으며, 빈자리에 대한 공격 회수가 조금씩 증가하여 수비형 전략이었다. 이동 실패 회수도 4000세대 이후부터 감소하기 시작하였다.



(그림 7) 조정 모델 2, 학습률 4인 경우(E모델)

E실험 모델의 경우는 조정 모델 2와 학습률 4를 이용한다. D모델과 마찬가지로 앞의 A, B, C모델과는 다른 행동 그래프를 보여주었다. 초기에는 다른 모델들과 같이 급격히 이동성이 증가하고, 공격성이 감소하였다. 하지만, 곧 이동 실패 행동에 대해 학습하기 시작하였으며, 공격 행동에 대해서도 학습하기 시작하였다. D모델과 마찬가지로 상대편에 대한 공격 회수를 어느 정도 수준에서 계속 유지하였다. 이는 조정 모델 2가 진화의 형태에 동일한 영향을 미쳤음을 의미한다. E모델 역시 진화의 후기에서는 이동성이 점차 감소하였으며, 빈자리에 대한 공격이 점차 증가하였다.

조정 모델 1이 적용된 B, C실험 모델은 다음과 같은 문제점이 관찰되었다. 첫째, 이동, 정지 행동의 경우에 무조건 1의 강화 신호를 받음으로 해서 불필요한 이동 행동이 강화되는 양상을 보였다. 이동 행동의 활성으로 인해 공격 행동에 대한 학습 시점이 A, D, E모델에 비해 늦어졌다. 둘째, 이동 실패나 공격 실패 행동을 억제할 수 있는 조정체제가 이루어지지 않았다. 이로 인해 진화를 거듭하여도 이동 실패와 공격 실패 행동에 대한 억제가 효과적이지 못했다. 또한 B, C실험 모델은 다른 모델들에 비해 전체 집단의 변화가 빠른 속도로 나타났다. A실험 모델이 총공격 횟수가 100000~400000구간을 통과하는 데에 걸린 세대수는 8000세대인 데 비해 B, D실험 모델은 3000세대 정도로 향상되었다.

조정 모델 2가 적용된 D, E실험 모델은 공격 실패에 대한 억제 효과가 눈에 띄게 나타났다. 이와 함께 이동 성공, 상대편에 대한 공격 성공 행동에 대한 학습도 이루어져 A, B, C실험 모델에 비해 높은 수준을 유지하였다. 이는 개체 수준의 조정 함수로서 전체 집단의 진화 방향을 조절할 수 있음을 제시한다.

에이전트 집단의 진화에 개체 수준의 학습 행동 개념을 첨가함으로써 전반적으로 집단의 변화 속도가 빨라지는 경향을 보였으나, 진화의 방향은 모든 실험 모델들이 동일하였다. 전반적인 진화의 방향은 경쟁 환경에 의해서 결정되며 에이전트 수준의 학습 행동을 추가함으로써 보다 빠른 탐색이 가능하였다.

4.2 집단의 우열 비교

집단간의 직접적인 경쟁을 통해 비교우위를 평가함으로써 그 집단을 진화시킨 방법에 대한 평가를 실시한다. 본 논문에서는 다음과 같은 세 가지 기준으로 비교 평가를 실시한다.

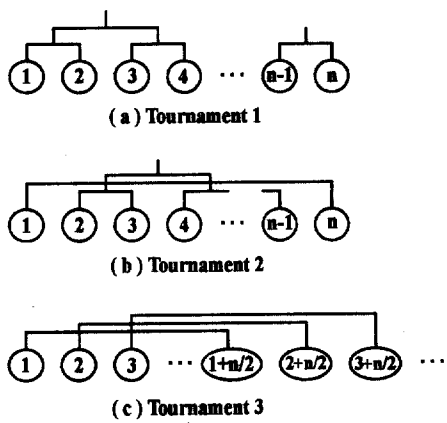
① 개체 수준의 학습 행동 유무
② 조정 모델
③ 학습률

집단의 우열 비교는 다음과 같은 두 가지 단계로 이루어지며, 각 비교 모델에 대해 5000, 10000, 15000, 20000세대의 4개 집단을 선택하여 실제 경쟁을 수행한다.

4.2.1 집단 내의 우수 에이전트 추출

집단 안에 포함된 모든 개체를 우열 비교 경쟁에 참여 시키면 너무 많은 시간을 소비하게 된다. 따라서, 본 논문에서는 집단 내에서 특정 개수의 우수 에이전트들을 추려내고 우수 에이전트들 사이의 우열 비교 경쟁을 통해 전체 집단의 우열을 판단한다.

집단 내에서 우수 에이전트는 다음 (그림 8)과 같은 세 번의 토너먼트를 통해 결정한다. 총 3번의 토너먼트를 수행하는 동안 에이전트가 얻은 승리 회수를 누적한다. 경쟁이 모두 끝나면 집단 내에서 누적 승리 회수가 많은 순서대로 정해진 수의 에이전트를 고른다. 만일 승리 회수가 같은 에이전트가 복수 개 존재한다면, 그 중에서 임의로 선택한다.

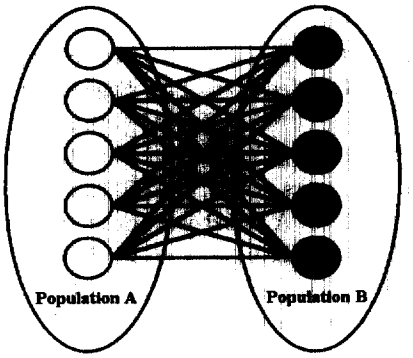


(그림 8) 토너먼트 방식

4.2.2 집단과 집단간의 직접 경쟁

집단 내에서 선택된 우수 에이전트들끼리는 완전(compl-

ete) 경쟁을 실시한다. (그림 9)은 각 집단에서 5개의 에이전트가 선택되었을 때의 경쟁을 나타낸다. 각 경쟁의 결과는 승패의 형태로 기록되며, 집단 내 유기체의 평균 승률로써 소속 집단의 우열을 판단한다. 예를 들어 A집단 내의 에이전트가 B집단 내의 에이전트에 대해 평균적으로 70%의 승률을 보인다면, A집단은 B집단보다 우수하다고 판단한다.



(그림 9) 우열 비교 경쟁

① 학습을 하지 않은 경우와 학습한 경우

에이전트의 학습 행동을 배제하고 진화시킨 집단과 학습 행동을 포함해 진화시킨 집단에 대한 상대적 우열을 비교한다. 직접적인 경쟁을 통한 우열 비교를 통해 학습 행동이 진화에 미친 영향을 평가한다. 우열 비교 경쟁은 여러 실험 모델과의 비교를 위해 {A-B, A-C, A-D, A-E}와 같이 네 번의 경쟁을 실시하였다. A실험 모델과 다른 실험 모델을 경쟁시킨 결과는 <표 3>와 같다. 표 항목은 각 비교 평가 집단이 A실험 모델에 대해 얻은 평균 승률을 나타낸다. B실험 모델의 5000세대의 개체가 A실험 모델 5000세대 개체에 대해 승리할 확률은 59.88%이다.

에이전트 학습 행동을 진화에 포함시킨 실험 모델이 본능 행동만을 통해 진화한 실험 모델에 대해 평균 75.09%의 승률을 보여 우세하였다. 이는 에이전트의 학습 행동이 에이전트진화의 선택 연산에 긍정적 영향을 주어 환경에 보다 적합한 에이전트를 선택할 수 있었기 때문이다. 특히, B실험 모델의 20000세대 집단은 94.6%의 높은 승률을 보여 같은 세대의 A실험 모델 집단에 대해 절대적인 우세를 보였다. B실험 모델의 20000세대 집단은 빈자리에 대한 공격 회수가 높고, 이동성이 낮은 전형적인 수비형 전략을 보인다. 따라서, 먼저 접근해오는 A실험 모델 에이전트를 한 스텝 먼저 공격함으로써 유리한 고지에 먼저 다다른다. 반면에 C실험 모델은 (그림 7)에서 와 같이 공격 성향이 나타나는 시점이 12000세대 부근으로 A실험 모델의 3000세대에 비해 현저히 늦다. 따라서 공격 성향이 나타나기 전인 5000, 10000세대에서의 승률은 A실험 모델에 비해 낮았으며, 공격 성향이 나타난 이후인 15000, 20000세대에서는 승률이 증가하였다.

〈표 3〉 우열 평가 결과 1

(단위 : %)

Model / Generation number	B	C	D	E	Average
5000	59.88	48.2	79.4	80.52	67.0
10000	80.8	45.17	76.0	65.4	66.84
15000	90.84	62.44	84.92	76.32	78.63
20000	94.6	87.93	84.72	84.32	87.89
Average	81.53	60.94	81.17	76.64	75.09

<표 3>에서 C, D, E실험 모델 10000세대의 승률이 낮아지는 것은 A실험 모델의 상대편에 대한 공격 성공 회수가 상대적으로 증가하기 때문이다. 이에 비해 B실험 모델의 경우에는 상대편에 대한 공격 성공 회수가 같이 증가하여 승률이 높아지는 결과를 보였다.

② 조정 모델

에이전트간의 직접 경쟁을 통해 조정 모델의 상대적 우열을 판단하며, 조정 모델이 에이전트집단의 진화에 미친 영향에 대해 평가한다. 이 실험은 학습률은 같고, 조정 모델이 서로 다른 실험 모델들에 대해 수행하였다.

조정 모델 2를 통해 진화한 에이전트들이 조정 모델 1을 통해 진화한 에이전트들에 대해 평균적으로 77.7%의 승률을 보임으로써 우세하였다<표 4>. 이는 조정 모델 2는 조정 모델 1이 가졌던 불필요한 행동에 대한 학습 요인을 제거하여 환경의 적응에 필요한 행동과 경쟁 전략을 일찍 학습할 수 있었기 때문이다. 조정 모델 2의 에이전트들은 조정 모델 1의 에이전트들에 비해 공격 성향이 빨리 나타났으며, 진화의 전반에 걸쳐 상대방 공격 성공 회수가 일정한 수준을 유지하였다.

〈표 4〉 우열 평가 결과 2

(단위 : %)

	학습률 2		학습률 4	
	조정 모델 1	조정 모델 2	조정 모델 1	조정 모델 2
5000	19.52	80.48	17.59	82.41
10000	36.44	63.56	18.39	81.61
15000	12.48	87.52	27.45	72.55
20000	14.56	85.44	31.96	68.04
평 균	20.75	79.25	23.85	76.15

③ 학습률

이 실험은 학습률이 진화의 속도에 미치는 영향을 평가하기 위한 것이다. 조정 모델의 종류가 같고, 학습률이 다른 실험 모델 사이의 실제 경쟁을 수행한다.

학습률에 따라서는 뚜렷한 비교 우위가 보이지 않았다<표 5>. 이는 학습률의 차이는 진화의 속도와 형태에 커다란 영향을 미치지 못하며, 진화는 조정 모델에만 영향을 받는 것으로 생각된다.

〈표 5〉 우열 평가 결과 3

(단위 : %)

	조정 모델 1		조정 모델 2	
	학습률 2	학습률 4	학습률 2	학습률 4
5000	69.72	30.28	35.08	64.92
10000	51.61	48.39	55.88	44.12
15000	39.05	60.95	58.04	41.96
20000	21.28	78.72	50.56	49.44
평 균	45.42	54.59	49.89	50.11

5. 결 론

본 논문에서는 집단간의 상호진화 시뮬레이션에 멀티에이전트 수준의 학습 행동을 도입하고, 그 영향을 평가하였다. 에이전트의 학습 행동을 도입하기 위해 신경망과, 강화학습을 적용하여 복잡하고 동적으로 결정되는 환경에서의 강화 학습 적용을 위한 조정 모델을 설계하였다.

에이전트의 학습 행동을 도입하였을 때 각 실험 모델 집단의 변화 속도는 빨라졌으며, 전체적인 진화의 성능이 향상되었다. 또한, 학습 행동을 추가하였을 때에 일정한 시간에 보다 우수한 경쟁 전략을 탐색할 수 있었다. 진화의 전체적인 방향은 학습의 영향보다는 환경의 영향을 받았으며, 에이전트 수준의 학습 행동을 통해 진화의 형태를 조절할 수 있었다. 에이전트의 학습 행동의 적용은 특정 목적을 가지는 에이전트 집단의 진화를 가능하게 하며, 이는 학습기법을 여러 분야의 특성에 맞게 적용할 수 있는 메커니즘을 제공한다. 예를 들면, 빨라진 집단의 변화 속도를 통해 환경에 대한 시스템의 적응성을 높일 수 있다. 또는 진화의 형태 조정을 통해 무인 탱크 시뮬레이션에서 공격 성향이 두드러진 경쟁 전략을 초기에 탐색하던지, 방어 성향이 강한 무인 탱크 행위자와 같은 특정 목적을 가지는 에이전트를 진화시킬 수 있다. 또한, 여러 개의 로봇이 공동 작업을 해야 하는 경우 그 행동 전략을 탐색할 수 있다.

본 논문의 시뮬레이션 결과는 강화 신호를 생성시키는 조정 모델에 따라 많은 영향을 받았다. 보다 적응적 시스템을 만들기 위해서는 본 논문에서 학습 행동 적용의 초기 단계로서 휴리스틱에 의거해 설계하였던 조정 모델을 인공진화 시뮬레이션을 통해 스스로 환경에 적합한 값을 찾아야 하며, 앞으로는 이에 관한 좀더 심도 있는 연구가 필요하다. 또한, 본 논문은 조직체가 환경과의 상호 작용에 의해 그 행동을 변경시키는 에이전트의 개별적 학습에 대해 연구하였다. 앞으로는 대규모의 에이전트들이 전체적인 제어 없이 지역적으로 상호 작용하여 에이전트가 형성되는 에이전트의 발상(morphogenesis)이나 진화의 과정에 의해 가능한 행동의 수가 늘어나는 종족의 발생(phylogenesis)에 관한 연구도 진행되어야 할 것이다.

참 고 문 헌

[1] 정보윤 외2인, "강화학습을 이용한 멀티에이전트 시스템의 자동 협력 조정 모델", 한국인지과학회논문지, 제10권 제1호, pp.1-11, 1999.

[2] Katia P. Sycara, "Multiagent System," AI MAGAZINE, Summer, 1998.

[3] H. S. Nwana, L. Lee. and N. R. Jennings, "Co-ordination in Multi-agent systems," Software Agents and Soft Computing, Towards Enhancing Machine Intelligence, Concepts and Applications, Springer, 1997.

[4] Peter Stone and Manuela Veloso, "Multi-agent Systems : A Survey from a Machine Learning Perspective," IEEE Trans. on Knowledge and Data Engineering, June, 1996.

[5] Durfee E. H., Lesser V. R., and Corkill D. D., "Coherent Cooperation among Communication Problem Solvers," IEEE Trans. Computers, Vol.11, pp.1275-1291, 1987.

[6] Hayes-Roth B, "A Blackboard Architecture for Control," Artificial Intelligence, No.25, pp.251-321, 1985.

[7] Smith R. G., "The Contract Net Protocol : High-Level Communication and control in a Distributed Problem Solver," IEEE Trans. on Computers, Vol.29. No.12, 1980.

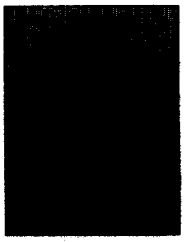
[8] Georgeff M., "A Theory of Action for Multi-Agent Planning," Proc. 1984 National Conf. Artificial Intelligence, pp. 121-125, August, 1984.

[9] David C. Parkes and Lyle H. Ungar, "Learning and Adaption in Multiagent Systems," AAAI workshop on Multiagent Learning Providence, June, 1997.

[10] Kreifelt T. and von Martial F., "A negotiation framework for autonomous agents," in Demazeau Y. and Muller J. P.(Eds) : Decentrilazed AI2, Elsevier Science, 1991.

[11] Werkman K. J., "Knowledge-based model of negotiation using shareable perspectives," Proc. Of the 10th International Workshop on DAI, Texas, 1990.

[12] Sandip Sen and Edmund H. Durfee, "The role of commitment in cooperative negotiation," International Journal on Intelligent & Cooperative Information System, Vol.3, No.1, pp.67-81, 1994.



이 말 례

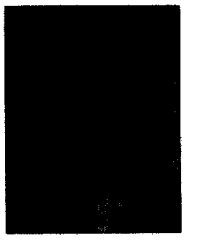
e-mail : mrlee@yosu.ac.kr

1998년 중앙대학교 컴퓨터공학과(공학박사)

1998년~1999년 조선이공대학 교수

1999년~현재 국립여수대학교 멀티미디어학부 교수

관심분야 : 인공지능, 가상현실, 멀티미디어, 전자상거래



김 상 근

email : sgkim@sungkyul.edu

1987년 중앙대학교 전자계산학과(이학사)

1989년 중앙대학교 전자계산학과(이학석사)

1996년 중앙대학교 컴퓨터공학과(공학박사)

1996년~현재 성결대학교 컴퓨터학부 조교수

관심분야 : 소프트웨어공학, CBR, 멀티에이전트, 전자상거래 기술