

# 진화학습을 이용한 다중에이전트의 일반화 성능향상을 위한 전략적 연합

## (Strategic Coalition for Improving Generalization Ability of Multi-agent with Evolutionary Learning)

양 승 룡<sup>†</sup>      조 성 배<sup>††</sup>  
(Seung-Ryong Yang)      (Sung-Bae Cho)

**요 약** 사회시스템이나 경제시스템 같이 동적으로 변하는 시스템에서는 그 구성원들 간에 복잡한 상호작용(행동)이 나타나게 되는데 구성원들의 행동은 변화하는 환경에 따라 적응하는 경향을 보인다. 그리고 이들의 행동양상은 흔히 생물학 분야의 조건반사에 비유되기도 한다. 본 논문에서는 복잡한 사회 현상을 모델링하고 분석하기 위하여 반복적 죄수의 딜레마 게임상에서 에이전트들의 전략적 연합을 통하여 변화하는 환경에 잘 적응하는 일반화 능력이 우수한 에이전트들을 자동 생성하는 방법을 제안한다. 또한 에이전트에 신뢰도를 부여하여 연합의 의사결정에 참가하게 함으로써 일반화 성능을 향상시키는 방법을 소개한다. 실험결과, 전략적 연합을 이용하여 진화된 에이전트들은 테스트 에이전트들에 비하여 일반화 성능이 우수함을 확인할 수 있었다.

**키워드** : 죄수의 딜레마 게임, 전략적 연합, 진화학습, 다중에이전트

**Abstract** In dynamic systems, such as social and economic systems, complex interactions emerge among its members. In that case, their behaviors become adaptive according to changing environment. In many cases, an individual's behaviors can be modeled by a stimulus-response system in a dynamic environment. In this paper, we use the Iterated Prisoner's Dilemma (IPD) game, which is simple yet capable of dealing with complex problems, to model the dynamic systems. We propose strategic coalition consisting of many agents and simulate their emergence in a co-evolutionary learning environment. Also we introduce the concept of confidence for agents in a coalition and show how such confidences help to improve the generalization ability of the whole coalition. Experimental results are presented to demonstrate that co-evolutionary learning with coalitions and confidence allows better performing strategies that generalize well.

**Key words** : IPD game, Strategic Coalition, Evolutionary Learning, Multi-agent

### 1. 서 론

진화연산(evolutionary computation)은 수학, 사회과학, 컴퓨터과학 등 많은 분야에서 동적으로 변하는 환경에 적응하는 시스템을 분석하거나 개발하는 데 이용되어져 왔다[1,2]. 그러나 최근에 와서 사회 및 경제 현상들이 매우 복잡한 양상을 보임에 따라 기존의 단순한 학습방법은 이들 시스템들을 모델링하는 데 한계를 보

이고 있다. 따라서 게임이론적인 접근방법과 같이 개선된 모델링 방법들이 연구되고 있다[3,4].

생물학에서 공진화(co-evolution)는 서로 영향을 미치는 다양한 개체들 사이의 진화과정을 말한다. 이 경우 어느 한 종이 진화하면 다른 종들도 따라서 영향을 받아 진화하게 된다. 게임이론과 진화게임에서, 반복적 죄수의 딜레마(Iterated Prisoner's Dilemma: IPD) 게임과 그 변형들은 공진화를 바탕으로 사회 또는 경제적인 현상들을 모델링하고 분석하는 데 널리 이용되어져 왔다[5,6]. 뿐만 아니라 공진화는 전략의 진화 학습에도 널리 사용되고 있는데, 특히 N명(N>2) 죄수의 딜레마 게임을 이용한 공진화 학습이 심도 깊게 연구되고 있다[7-9].

본 논문에서는 IPD 게임에서 공진화를 이용하여 에이

· 이 논문은 한국학술진흥재단의 연구과제(KRF-2002-005-H20002)에 의해 지원되었음

† 비회원 : 연세대학교 컴퓨터학과  
saddo@candy.yonsei.ac.kr

†† 종신회원 : 연세대학교 컴퓨터학과 교수  
sbddo@candy.yonsei.ac.kr

논문접수 : 2003년 8월 29일

심사완료 : 2003년 10월 30일

전트의 일반화 성능을 향상시키는 문제에 중점을 두었다. 이러한 문제는 고전적인 2IPD 게임에서부터 연구되어 왔으나, 여기에서는 에이전트들의 전략적 연합을 이용하여 향상된 성능을 얻고자 하였다.

연합은 집단으로부터 추출된 우수한 에이전트들이 더 많은 이득(payoff)을 얻거나 다음 세대에도 계속 살아남을 수 있도록 자동적으로 형성되게 하였다. 또한 연합에 소속된 에이전트들이 자신의 이득 값에 비례하여 신뢰도(가중치)를 가지게 함으로써 연합의 다음 행동을 결정하는데 영향을 미치도록 하였으며, 신뢰도는 고정하지 않고 진화 학습 과정에 따라 변화시켰다.

논문의 구성은 다음과 같다. 2장에서는 죄수의 딜레마 게임과 진화적인 접근방법에 대하여 소개하고, 전략의 표현방법에 대해 설명한다. 3장에서는 죄수의 딜레마 게임에서 적용하고자 하는 전략적 연합을 제안하며, 진화 과정 동안 연합이 어떻게 형성되는가를 설명한다. 4장에서는 실험결과를 제시하며, 마지막으로 5장에서는 논문의 결론을 제시한다.

2. 진화적 학습

2.1 반복적 죄수의 딜레마 게임

일반적인 죄수의 딜레마 게임에서 각 게임자는 “협동(cooperation)” 또는 “배반(defection)”을 선택할 수 있고 주어진 이득표에 따라 게임이 종료된 후 이득을 가져간다. 표 1은 Axelrod가 제안한 2인 죄수의 딜레마(2IPD) 게임의 이득표이다[10]. 이 게임은 비제로섬(non-zero sum) 게임이며 비협조 게임이다. 즉, 자신의 이득이 상대방의 손실과 같지 않으며 게임자간의 정보교환도 불가능하다. 이 게임은 무한히 반복되며 어느 게임자도 게임이 언제 끝날지 알 수 없다.

표 1 2IPD 게임의 이득표( $T > R > P > S$ ,  $2R > T + S$ )

		상대방의 행동	
		협동	배반
자신의 행동	협동	자신 : R (3점) 상대 : R (3점)	자신 : S (0점) 상대 : T (5점)
	배반	자신 : T (5점) 상대 : S (0점)	자신 : P (1점) 상대 : P (1점)

표 1에 따르면, 죄수의 딜레마 게임이 한 번만 진행될 경우 최선의 전략은 배반이 된다. 그러나 게임이 반복될 경우 상호협동(mutual cooperation)이 더 많은 이익을 가져다 줄 수 있기 때문에 배반은 최선의 전략이 될 수 없다. 이와 같이 2IPD 게임에서 사용할 수 있는 우수한 게임 전략을 개발하는 연구는 오랫동안 많이 진행되어 왔다[4].

반복적 죄수의 딜레마 게임은 수많은 내쉬 균형상태(Nash equilibrium)로 구성되어 있다[11]. 그림 1과 같이 단판 게임의 순수전략(pure strategy)들의 이득점(0,5), (3,3), (5,0), (1,1)을 연결하는 내부 영역이 게임자들의 이득 영역이 되며, 이득점 (1,1)이 내쉬 균형상태가 된다. 그러나 게임이 반복적으로 진행되는 동안 게임자들은 상호배반(mutual defection)이 결국 최악의 이득을 보장한다는 사실로부터 상호협동하는 전략을 선택하게 되어 내쉬 균형상태로부터 벗어나고자 한다.

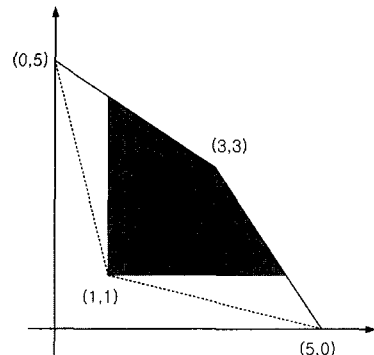


그림 1 죄수의 딜레마 게임에서 내쉬 균형점

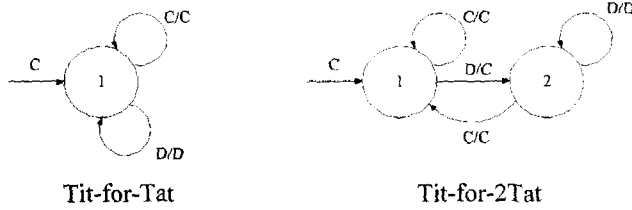
일반적인 반복적 죄수의 딜레마 게임은 두 명의 죄수 사이에서 일어나는 상호작용을 모델링한 게임이다. 반면에 이로부터 변형된 형태의 게임이 N명 반복적 죄수의 딜레마(NIPD) 게임이다. 실제계에 존재하는 많은 시스템들은 양자간이 아닌 다자간의 상호작용으로 이루어져 있다. 따라서 2IPD 게임과는 달리 NIPD 게임은 보다 사실적으로 사회 및 경제 시스템을 모델링할 수 있다는 장점이 있다[4]. Yao와 Darwen은 NIPD게임에 대한 다양한 실험적 결과를 통하여 N의 크기가 증가할수록 협동으로 진화하기가 어렵다는 사실을 밝혀내었다[7]. NIPD 게임은 2IPD 게임과는 다른 이득함수를 가지고 있어야 하는데, Yao와 Darwen은 표 2와 같은 이득함수를 사용하였다.

표 2 NIPD 게임의 이득함수( $D_x > C_x$ ,  $D_{x+1} > D_x$ ,  $C_{x+1} > C_x$ ,  $C_x > (D_x + C_{x-1})/2$ )

협동자의 수	0	1	...	x	...	n
협동	$C_0$	$C_1$	...	$C_x$	...	$C_n$
배반	$D_0$	$D_1$	...	$D_x$	...	$D_n$

2.2 전략의 표현

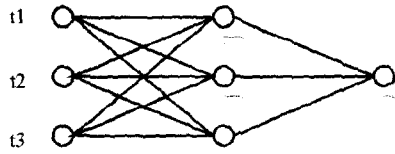
반복적 죄수의 딜레마 게임의 전략들을 진화적으로 학습시키기 위해서는 적절한 형태로 표현할 필요가 있



(a) 유한상태기계

	Strategy Table	History Table
Tit-for-Tat	0 0 1 1 1 1 1 0 1 1 0 0 1 1 0 0 1	1 0 0 1
CDCD	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	1 1 0 1
AllD	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 1 0 1

(b) 비트스트링



(c) 신경망

그림 2 전략의 표현방법 예

다. Axelrod은 유한상태기계(finite state machine), 논리트리(logic tree), 마코프체인(Markov chain) 등과 같은 여러 가지 전략의 표현방법들을 이용하여 실험하였다[12]. 그림 2는 이와 같은 다양한 전략의 표현방법들을 보여준다.

일반적으로는 간단한 비트스트링(bitstring) 표현방법을 주로 사용한다. 이 방법은 각 개체를 전략부와 히스토리부로 나누어 비트 형태로 표현한 것으로 그림 3과 같은 방법으로 나타낼 수 있다. 이 비트스트링 표현방법은 Axelrod에 의해 IPD 게임에 적용되었는데 본 논문에서도 이 표현방법을 이용하였다. 그림 3에서 히스토리 테이블은 자신과 상대방의 과거 행동을 기억하여 다음 전략을 결정하는데 사용된다. 예를 들어, 자신의 과거 행동이 협동이고 최근 행동이 배반이면 '01'로 표현되며, 상대방의 과거 행동이 배반이고 최근 행동이 협동이면 '10'이 되어 전체 히스토리리는 '0110'이 된다. 이를 십진수로 변환하면 '6'이 되고 전략테이블의 7번째 값(전략테이블이 '0'부터 시작하므로)을 다음 전략으로 선택하게 된다. 따라서 히스토리 크기가 2인 2IPD 게임을 비트로

표현하기 위해서는 20비트가 필요하다.  $l$ 을 히스토리의 길이,  $n$ 을 게임자의 수라 했을 때 이를 수식으로 표현하면 다음과 같다.

$$2^{(l+1 \log_2 n)} \tag{1}$$

수식 (1)은 게임이 진행될 때 필요한 비트의 수이므로 맨 처음 게임을 진행하기 위해서는 초기 히스토리 비트가 필요하다. 따라서 필요한 전체 비트 수는 다음과 같다.

$$2^{(l+1 \log_2 n)} + (l+1 \log_2 n) \tag{2}$$

### 2.3 전략의 진화

진화의 관점에서 IPD 게임은 그림 4와 같이 진행된다. 먼저 게임자(에이전트)들이 집단으로부터 무작위로 선택되어 서로 IPD 게임을 한다. 일반적으로 IPD 게임은 행동선택(select action), 행동(move), 그리고 전략갱신(update history) 등의 절차로 진행된다. 각 세대마다 충분한 게임이 진행된 다음, 각 게임자는 자신이 게임을 하는 동안 획득한 이득에 따라 적합도(fitness)를 평가 받는다.

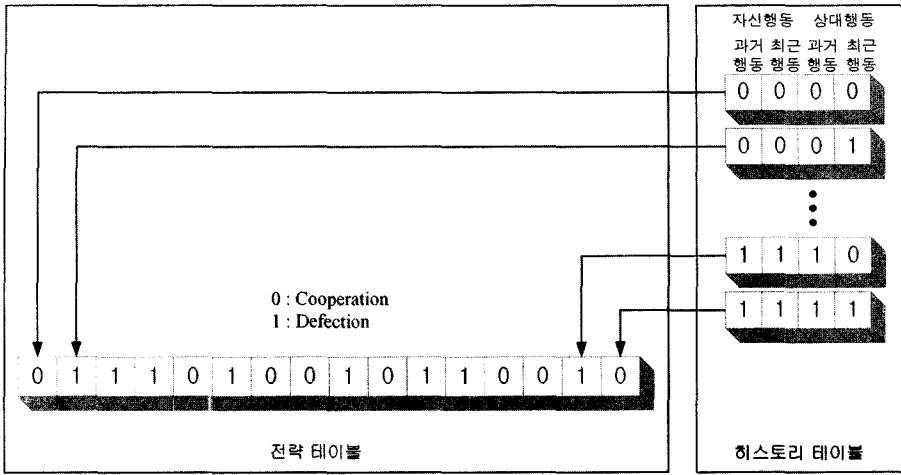


그림 3 2IPD 게임에서 전략의 비트스트링 표현 예(히스토리 길이=2)

### 3. 전략적 연합

공진화를 이용한 학습이 항상 일반화 능력이 우수한 에이전트를 생성하는 것은 아니다. 그 이유는 진화된 에이전트가 현재의 집단에 과적합(overfit)되어 새로운 외부 에이전트에 대해서 적응력이 떨어지기 때문이다[13]. 이러한 문제를 해결하기 위한 방법으로 공진화에 의해 학습된 에이전트의 일반화 능력을 향상시키기 위해 전략적 연합(strategic coalition)을 이용할 수 있다. 이는 다중 에이전트 환경에서 비록 에이전트들이 서로 이기적이라 할지라도 자율적인 에이전트들 사이에서 일어나는 연합은 상호간의 이익을 보장할 수도 있기 때문이다[14-16]. 전략적 연합에 속해 있는 에이전트들은 특정 형태의 상대 에이전트에 특화(specialization)되므로 전체적인 연합의 일반화 능력을 향상시키게 된다[13]. 여기에서는 IPD 게임에서 에이전트의 일반화 능력을 향상시킬 수 있는 전략적 연합을 제안한다.

#### 3.1. 전략적 연합

에이전트들 사이에서 전략적 연합을 형성하기 위해서는 두 가지 원칙이 먼저 선행되어야 한다. 첫 번째 원칙은 에이전트들이 어떠한 방법으로 자신들의 행동을 조절하고 협동하는가에 대한 것이다. 그리고 두 번째 원칙은 모든 가능한 연합구조 가운데에서 어떤 구조의 연합을 형성할 것인가이다[16]. IPD 게임에서 전략적 연합을 형성하기 위해서는 먼저 두 가지 가정(assumption)이 전제되어야 한다.

**가정 1.** 게임에 참가하는 에이전트들은 서로 연합 형성 의도를 파악하기 위하여 의사교환을 할 수 있어야 한다. 이 가정은 에이전트들이 IPD 게임을 끝낸 후 서로

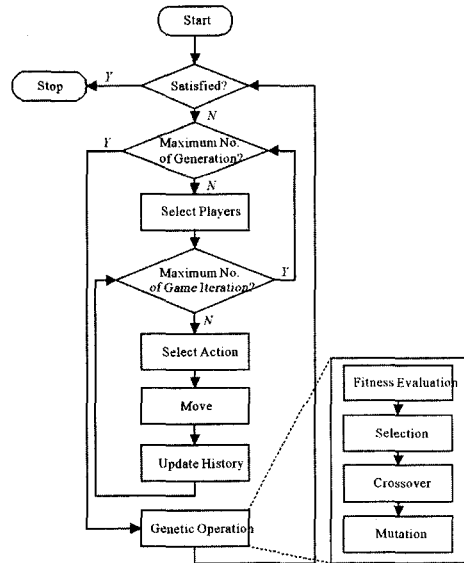


그림 4 IPD 게임의 진화 과정

로의 이득을 비교함으로써 달성될 수 있다.

**가정 2.** 에이전트들의 연합은 자율적으로 형성되어야 한다. 즉, 연합 형성에 관한 어떤 외압도 차단되어야 한다. 이와 같은 조건하에서 에이전트들은 자율적으로 연합을 형성하기 위하여 서로 협상할 수 있다.

IPD 게임은 전략적 연합의 형성과정을 알아볼 수 있는 좋은 예이다. 실제 사회에서 IPD 게임에서와 같은 개인과 개인, 개인과 그룹, 그룹과 그룹간의 전략적 연합은 쉽게 찾아볼 수 있다. IPD 게임에서 전략적 연합을 정의하기 위하여 다음과 같이 나타낸다.  $I = \{A_1, A_2,$

...,  $A_n$ )는 게임에 참가하는 에이전트의 집합이며,  $C = \{C_i, C_j, \dots, C_k\}$ ,  $|C| \geq 2$ 는 에이전트들이 형성하는 전략적 연합이다. 연합  $C$ 는 에이전트의 집합  $I: C \subseteq I, |C| \leq |I|$ 의 원소이다. 모든 에이전트는 게임을 통해 자신의 이득  $p_i$ 를 가진다. 따라서 연합  $C$ 는 벡터  $C = \langle S_c, N_c, f_p, D_c \rangle$ 를 가진다. 여기에서  $S_c, N_c, f_p, D_c$ 는 각각 연합의 이득 합, 연합에 참가한 에이전트의 수, 이득함수, 연합의 의사결정을 나타낸다. 전략적 연합은 다음과 같이 정의할 수 있다.

**정의 1. 연합값(Coalition Value):**  $\omega$ 를 각 게임자의 이득에 대한 가중치라고 하면, 연합의 이득  $C_p$ 는 연합에 참가하는 에이전트들의 신뢰도(가중치)에 의한 평균 이득이 된다.

$$S_C = \sum_{i=1}^{|C|} p_i \omega_i \quad (3)$$

$$\text{단, } \omega_i = \frac{p_i}{\sum_{j=1}^{|C|} p_j}, \quad C_p = \frac{S_C}{|C|}$$

**정의 2. 이득함수(Payoff Function):** 연합에 속한 에이전트들은 주어진 함수에 따라 이득을 얻는다. 일반적으로 2IPD 게임은 Axelrod의 이득함수를 따른다[10].

**정의 3. 연합식별(Coalition Identification):** 각각의 연합은 자신만의 고유 식별번호를 가진다. 이는 연합이 최초로 형성될 때 부여되며, 연합이 해체될 경우 함께 소멸한다. 이러한 절차는 진화과정에 따라 자율적으로 이루어진다.

**정의 4. 의사결정(Decision Making of Coalition):** 연합은 소속된 에이전트들의 의사를 모아 하나의 행동을 결정하여야 한다. 이 행동은 연합이 게임에 참가할 경우 다음 행동이 된다.

$$D_c = \begin{cases} 0 = \text{Cooperation,} & \text{if } 1 < \frac{\sum_{i=1}^{|C|} C_i^C \cdot \omega_i}{\sum_{i=1}^{|C|} C_i^D \cdot \omega_i} \\ 1 = \text{Defection,} & \text{if } 0 < \frac{\sum_{i=1}^{|C|} C_i^C \cdot \omega_i}{\sum_{i=1}^{|C|} C_i^D \cdot \omega_i} \leq 1 \end{cases} \quad (4)$$

**정의 5. 이득분배(Payoff Distribution):** 연합에 속한 에이전트들은 다른 에이전트나 연합과의 게임이 끝난 후, 그 결과에 따라 이득을 얻는다. 이 이득은 연합에 소속된 각 에이전트의 신뢰도에 따라 분배되어야 한다.

$$p_i = \omega_i \cdot \frac{S_c}{|C|} \quad (5)$$

위의 정의에서 연합은 2IPD 게임에 참가하는 에이전트들의 집합이다. 본 논문의 실험에서 연합은 게임이 진

행되는 동안 자율적으로 형성되며, 다음 세대로 변형없이 이어진다.

### 3.2 연합 형성

연합의 형성은 다음 세 가지 단계를 포함한다. (1) 연합 구조를 생성한다 (2) 최적의 연합 구조를 탐색한다 (3) 연합의 이득을 분배한다[17]. IPD 게임이 진행되는 동안, 에이전트들은 연합을 형성하기 위하여 상대방의 의사를 파악한다. 이때 에이전트들이 가지고 있는 이득이 연합 형성을 위한 주요 요소가 된다. 본 논문에서는 연합을 진화시키기 위하여 에이전트들의 신뢰도를 이용한다. 신뢰도는 고정된 것이 아니라 진화과정에 따라 동적으로 변한다. 에이전트는 대전하는 상대에 따라 다른 신뢰도를 가지게 되며, 자신이 가지는 신뢰도에 따라 연합의 의사결정에 참가한다. 다시 말해서, 연합의 다음 행동은 에이전트들이 가지는 신뢰도에 따라 결정된다.

연합을 이용한 IPD 게임은 그림 5와 같이 진행된다. 각 세대마다 무작위로 선택된 두 에이전트가 라운드로빈(round-robin) 방식으로 게임을 한다. 이들 에이전트는 다른 에이전트와 게임을 할 수도 있고, 그 에이전트가 소속된 연합과 게임을 할 수도 있다. 게임이 끝난 다음, 주어진 조건을 만족할 경우 에이전트들은 서로 연합을 형성할 수도 있으며, 이미 형성된 다른 연합에 참가할 수도 있다. 연합과 연합이 서로 게임을 할 경우 연합 구조에는 변함이 없으며, 적합도만 바뀐다.

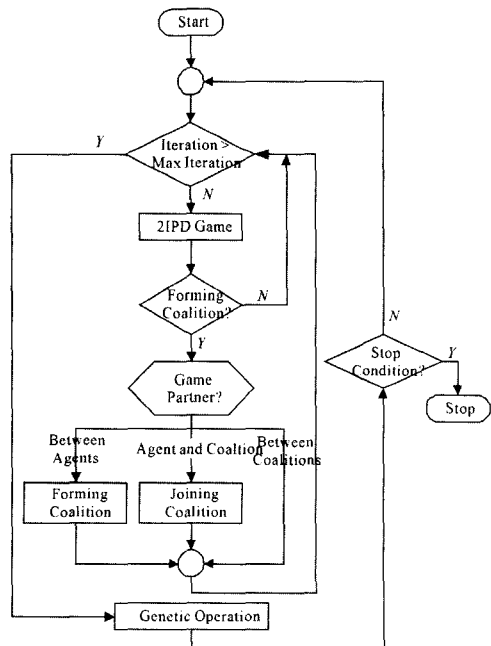


그림 5 전략적 연합을 이용한 2IPD

에이전트가 연합에 참가하면, 연합에 속해 있는 다른 에이전트들과 라운드로빈 방식으로 다시 게임을 하게 된다. 예를 들어 연합에  $k$ 명의 에이전트가 있다면, 모두  $k(k-1)$ 번의 게임이 일어난다. 연합에 참가한 에이전트의 수가 연합의 크기를 초과하면, 가장 약한 에이전트는 연합으로부터 제거된다. 연합에 속한 모든 에이전트들은 각자의 이득 값에 따라 다시 순위가 매겨지고 이 순위 에 따라 신뢰도가 주어진다. 여기에서 전체 신뢰도의 합 은 1이 된다. 이렇게 주어진 신뢰도는 연합이 다음 행동을 결정할 때 중요한 역할을 하게 된다.

그림 6은 연합에 속한 에이전트들이 다음 행동을 결정하기 위하여 자신의 신뢰도에 따라 의사를 반영하는 방법이다. 그림에서 *Others*는 상대방의 행동을 나타내며 *Self*는 이전 단계에서의 자신의 행동을 나타낸다.  $A, B, C, D$ 는 연합에 속한 에이전트들이며,  $A_d$ 는  $A$ 라는 에이전트의 다음 행동이 배반임을 나타낸다. 만약  $A$ 의 다음 행동이 협동이면  $A_c$ 는  $1 \times$  “ $A$ 의 신뢰도”가 되며  $A_d$ 는 ‘0’이 된다.

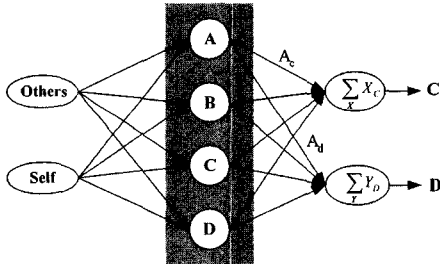


그림 6 연합에 소속된 에이전트들의 신뢰도를 이용한 연합의 의사결정

3.3 연합의 진화

연합은 진화 과정을 거치는 동안 새로이 형성되거나 제거될 수 있다. 실험에서 연합은 집단 전체의 평균 이

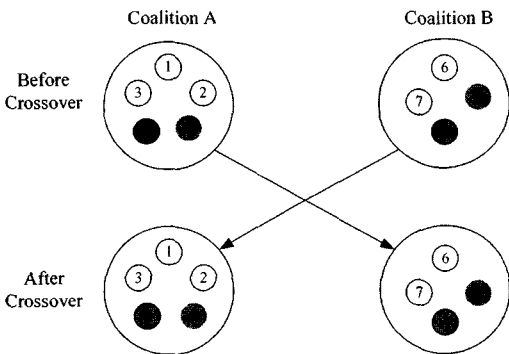


그림 7 연합간의 교차

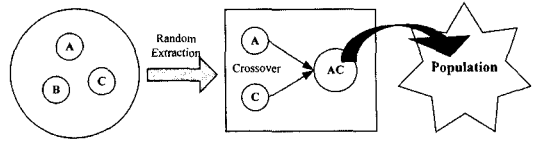


그림 8 연합에 속해 있는 에이전트를 교차하여 새로운 에이전트 생성

득보다 낮을 경우 해체된다. 연합은 연합끼리의 교차를 통해서도 생성될 수 있다. 연합간의 교차는 그림 7과 같 이 진행된다. 진화 과정이 계속되면 우수한 에이전트들은 연합에 소속되고 그 에이전트는 집단에서는 제거되므로 집단에 남아 있는 에이전트들은 대부분 약한 에이전트들이 될 수 있다. 따라서 이러한 문제를 해결하기 위해 연합에 속해 있는 에이전트들을 교차시켜 새로운 에이전트를 생성하여 이를 집단에 포함시킨다. 그림 8은 이러한 과정을 나타낸 것이다.

4. 실험결과

본 논문에서는 먼저 공진화 학습을 이용한 연합의 진화에 대해 실험한다. 그리고 연합이 다른 상대에 대하여 잘 적응하는 지를 알아보기 위하여 연합에 속한 에이전트들의 신뢰도를 진화시킨다.

4.1 연합의 진화

실험에 사용된 집단의 크기는 100으로 하였으며, 교차율은 0.6, 돌연변이율은 0.001로 하였다. 진화를 위한 개체 선택방법으로는 순위 기반(rank-based) 선택을 사용하였으며, 일점 교차를 이용한 유전자 알고리즘을 이용하였다. 전략을 표현하기 위하여 사용된 히스토리의 크기는 2이며, 연합에 포함될 수 있는 에이전트의 수는 10, 연합의 수는 집단 크기의 1/3로 하였다.

그림 9는 연합의 평균 적합도를 나타낸 것이다. 진화 초기에는 연합의 평균 적합도가 개별 에이전트의 평균 적합도 보다 높게 나타난다. 그러나 시간이 지날수록 연합의 평균 적합도가 낮아진다. 이는 집단에 속한 개별 에이전트들이 학습을 통하여 연합의 게임 방법을 터득하기 때문이다 (여기에서 개별 에이전트들은 연합에 속한 우수한 에이전트들을 이용하여 재생성된 에이전트들이다). 즉 집단에 속한 에이전트들은 점차 자신이 속한 환경에 적응하게 된다.

Bias는 집단의 행동양상을 분석하는데 사용되는 것으로 그림 10은 집단에 소속된 에이전트들이 어느 정도의 비율로 협동 또는 배반을 선택하는지를 보여준다. 예를 들어, Bias가 1이면 집단에 속한 에이전트들은 100% 협동을 선택하거나 100% 배반을 선택함을 의미한다. 그림 10에서와 같이 집단은 100세대에 이르러 수렴하는데

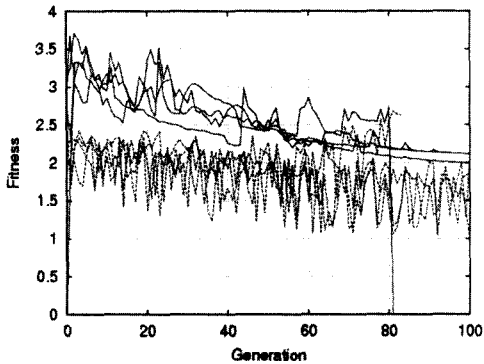


그림 9 연합에 속한 에이전트와 집단에 속한 에이전트의 평균 적합도. 직선은 연합 속의 에이전트, 점선은 집단 속의 에이전트

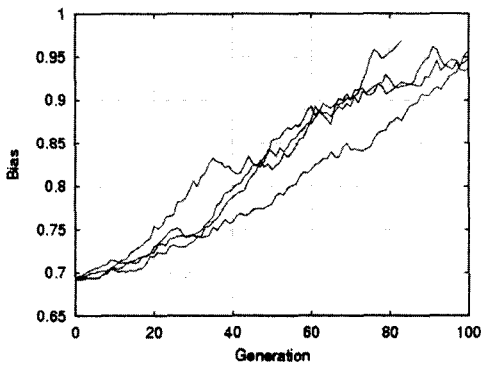


그림 10 집단의 수렴

이는 대부분의 에이전트들이 협동을 선택하기 때문이다. 이 실험에서 연합의 다음 행동은 연합에 참가한 에이전트들의 가중치 투표에 의해 결정되며, 이는 곧 연합의 전략이 하나의 에이전트 형태로 표현될 수 있음을 의미한다. 그림 11은 진화과정 동안 생성되는 연합의 수를 나타내고 있다.

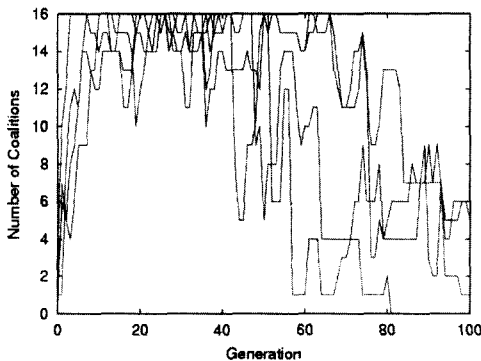


그림 11 진화과정 동안 생성되는 연합의 수

#### 4.2. 에이전트 신뢰도의 진화

진화된 신뢰도를 갖는 연합에서 소속된 에이전트들의 일반화 성능을 확인하기 위하여 표 3과 같이 널리 알려진 테스트 전략들을 사용하였다.

표 3 에이전트의 신뢰도 진화를 위한 테스트 전략

전략	설명
TFT	처음에는 협동, 그 다음부터는 상대방의 행동을 따른다
TF2T	처음에는 협동, 그 다음부터 상대방이 두 번 연속 배반하면 배반한다
Trigger	처음에는 협동, 그 이후 상대가 한번 배반하면 계속 배반한다
AID	항상 배반한다
CDCD	협동과 배반을 반복해서 한다
CCD	협동, 협동, 배반을 반복해서 한다
C10Dall	10회 협동 후 계속 배반한다
Random	무작위로 행동한다

그림 12와 13은 연합을 이용하여 진화된 에이전트들과 테스트 전략들과의 실험결과를 나타낸 것이다. 이 결과에서 연합을 이용하여 진화된 에이전트들이 협동을 주로 하는 우호적인 테스트 전략(TFT, Trigger, TF2T)에 대해서는 서로 협동하며, AID와 같은 배반 전략들에게는 복수하고 Random 전략에 대해서는 적절한 행동을 선택함으로써 성능이 우수함을 알 수 있다. 또한 C10Dall 전략과 CDCD 전략과의 게임에서도 높은 이득을 획득하고 있다. 이는 상대방이 어떤 전략인지에 관계없이 연합을 이용하여 진화된 에이전트들이 우수한 일반화 성능을 가지고 있음을 나타내는 것이다.

신뢰도를 이용하여 진화된 연합의 일반화 성능을 평가하기 위하여 본 논문에서는 표 4와 같이 300개의 무작위 에이전트들로 구성된 집단으로부터 일반적인 유전자 알고리즘에 의해 진화된 상위 30개의 테스트 에이전트를 추출하였다. 각 테스트 에이전트는 연합을 이용하

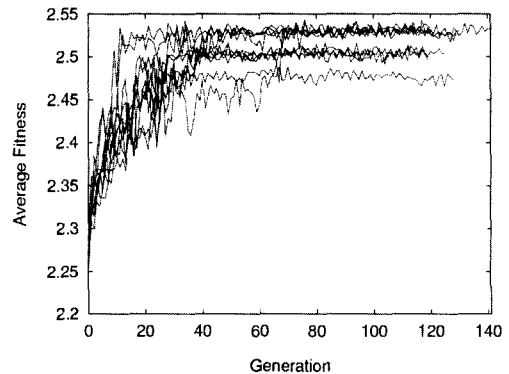


그림 12 연합을 이용하여 진화한 집단의 평균 이득

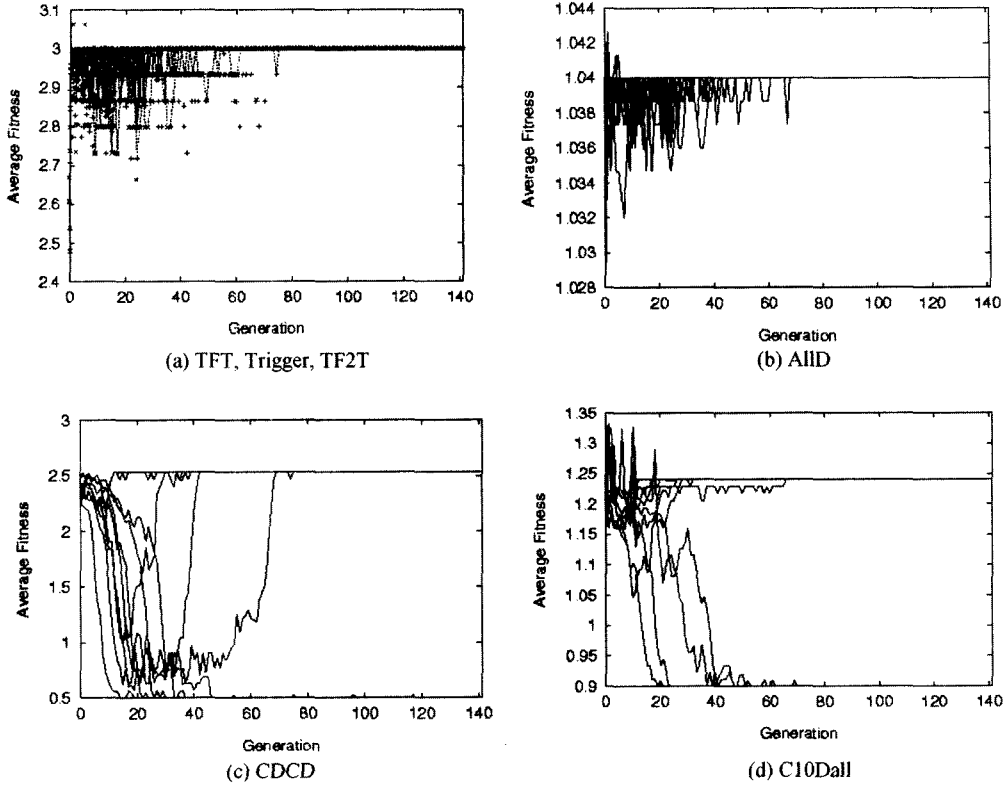


그림 13 테스트 전략의 평균 이득

표 4 300개의 무작위 집단으로부터 추출된 상위 30개의 테스트 에이전트

히스토리	전략표	히스토리	전략표
1000	0111101111101110	1111	1101111001110011
0100	1101111111110011	1100	0101100111010001
1000	1111111000000111	0111	0011111101010010
0111	0111111110110111	1111	0111011101010111
1100	0111001101010011	1111	0101101100111100
0001	1011111001011100	0111	1001110101010110
0001	0011010111111110	1101	0001000110011010
1100	0001110110110010	0010	1101110011111101
1000	1101110011011101	1001	1001100101011000
0010	1101101111000110	1000	1101000111101010
1111	1111101100011011	0110	1001011001110110
0110	0111010011011111	1011	0101111101110010
1010	1011010111111100	0001	0011110111011000
1101	1101000101011110	1110	0111010101110001
1100	1101011110111011	1011	0101011111110100

여 진화된 에이전트들과 라운드로빈 방식으로 IPD 게임을 하였다.

연합의 일반화 성능을 평가하는 실험에서 에이전트의 신뢰도는 상대에 따라 변화도록 하였다. 표 5는 300개의 무작위 집단으로부터 추출된 상위 30개의 테스트 에이

전트들에 대한 성능을 나타내고 있다. 표 5의 두 번째와 세 번째 열에서 보는 바와 같이 연합을 이용하여 진화된 에이전트들은 연합을 이용하지 않은 테스트 에이전트에 비해 승률과 평균이득이 크게 향상되었음을 알 수 있다. 표 5의 네 번째 열에서는 일반적으로 알려진 테스트



표 5 연합을 이용한 진화 에이전트와 테스트 에이전트의 성능 평가

전략	승	무	자신 평균이득	상대 평균이득
진화 전	8.64±4.9	6±2.19	1.84±0.28	1.75±0.59
진화 후	18.55±0.5	4±0.63	2.16±0.07	0.92±0.29
TFT	8	0	1.70	1.77
Trigger	30	0	2.13	0.80
TF2F	7	0	1.54	2.40
AllD	30	0	2.17	0.7
CDCD	0	0	1.05	2.75
CCD	0	0	0.91	3.34
C10Dall	27	0	1.97	1.12

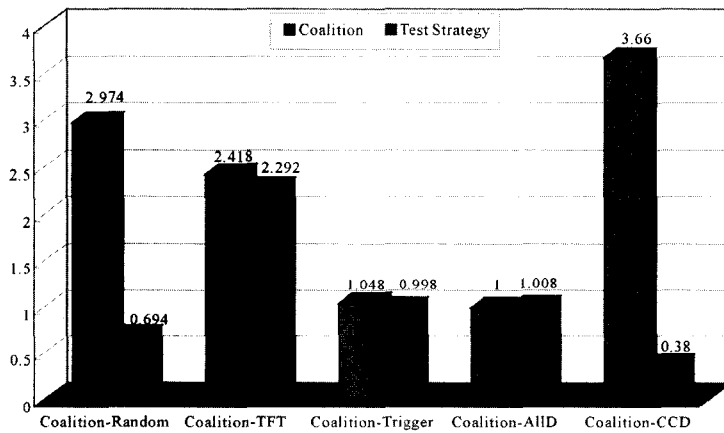


그림 14 연합을 이용한 진화 에이전트와 테스트 전략과의 평균 이득 비교

트 전략과 30개의 테스트 에이전트들과의 게임결과를 나타낸 것이다. 이 열에서 AllD 전략을 제외한 모든 테스트 전략은 연합을 이용하여 진화된 에이전트에 비하여 성능이 떨어졌다. AllD 전략은 일반적으로 전략이라고 평가할 수 없으므로 연합을 이용하여 진화된 에이전트는 성능이 우수하다고 말할 수 있다. 그림 14는 연합을 이용하여 진화된 에이전트들과 테스트 전략을 서로 대전시켜 평균 이득 값을 비교한 것이다. 이 실험에서도 연합을 이용한 에이전트는 일반적인 테스트 전략에 비하여 높은 이득을 얻음으로써 우수한 성능을 보였다.

### 5. 결론

다중에이전트 환경 하에서 에이전트들을 연합하는 것은 새로운 환경에 적응하는 학습 시스템을 만들기 위한 효과적인 방법이 될 수 있다. 본 논문에서는 죄수의 딜레마 게임에서 전략적 연합을 이용하여 에이전트들이 서로 연합함으로써 일반화 성능이 향상된다는 것을 실험적 결과로서 제시하였다. 연합에 소속된 각 에이전트들은 자신만의 신뢰도를 가지며, 이는 특정한 상대에 대하여 특화(specialized) 되었다.

본 논문에서의 실험은 전략적 연합을 이용하여 에이전트의 신뢰도를 진화하게 한 경우와 전략적 연합을 사용하지 않은 경우로 나누어 진행하고 그 결과를 상호 비교하였다. 결과적으로 전략적 연합을 이용하여 진화된 에이전트들이 상대가 협동전략인지 배반전략인지에 관계없이 일반화 성능이 우수하였다.

본 논문에서의 실험은 반복적 죄수의 딜레마 게임을 이용한 것으로 여기에서 얻은 실험결과는 좀더 복잡한 게임에도 적용 가능할 것이다. 특히 전략적 연합을 이용한 진화게임은 동적으로 변화하는 실제계 환경을 모델링하는데 매우 흥미롭게 응용될 수 있다. 이는 국가간의 협력관계, 기업들의 전략적 합병 등 다양한 활동들을 모델링하고 분석하는 데 유용하게 적용될 수 있을 것이다.

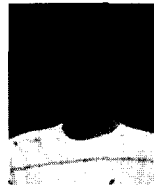
### 참고 문헌

[1] M. Patignani, "Stability of arbitrary genes: A new approach to cooperation," *Proceedings of Genetic and Evolutionary Computation Conference*, Morgan Kaufmann, San Francisco, California, USA, pp. 907, 2001.

[2] A. Francisco, "A computational evolutionary app-

- roach to evolving game strategy and cooperation," *IEEE Transactions on Systems, Man and Cybernetics*, Part B, Vol. 32, No. 5, pp. 498~502, 2002.
- [3] 서연규, 조성배, "진화방식을 이용한 N명 죄수 딜레마 게임의 협동연합에 관한 실험적 연구", *한국정보과학회 논문지 (B)*, Vol. 27(3), pp. 257~265, 2000.
- [4] A. M. Colman, *Game Theory and Experimental Games*, Oxford, England: Pergamon Press, 1982.
- [5] T. Ord and A. Blair, "Exploitation and peace-keeping: Introducing more sophisticated interactions to the iterated prisoner's dilemma," *Proceedings of 2002 Congress on Evolutionary Computation*, Vol. 2, pp. 1606~1611, 2002.
- [6] L. Tesfatsion, "Agent-based computational economics: Growing economics from the bottom up," *Artificial Life*, Vol. 8, pp. 55~82, 2002.
- [7] X. Yao and P. J. Darwen, "An experimental study of N-person iterated prisoner's dilemma games," *Informatica*, Vol. 18, pp. 435~450, 1994.
- [8] Y. G. Seo, S. B. Cho and X. Yao, "Exploiting coalition in co-evolutionary learning," *Proceedings of Congress on Evolutionary Computation 2000*, Vol. 2, pp. 1268~1275, 2000.
- [9] J. A. Fletcher and M. Zwick, "N-Player prisoner's dilemma in multiple groups: A model of multilevel selection," *Proceedings of Artificial Life VII Workshops*, Portland, Oregon, 2000.
- [10] R. Axelrod, *The Evolution of Cooperation*, New York: Basic Books, 1984.
- [11] J. Nash, "Equilibrium points in n-person games," *Proceedings of National Academy of Sciences*, Vol. 36, pp. 48~49, 1950.
- [12] D. Ashlock and M. Joenks, "ISAc lists, a different representation for program induction," *Genetic Programming 98, Proceedings of the Third Annual Genetic Programming Conference*, pp. 3~10, San Francisco, Morgan Kaufmann, 1998.
- [13] P. J. Darwen and X. Yao, "Speciation as automatic categorical modularization," *IEEE Transactions on Evolutionary Computation*, Vol. 1, No. 2, pp. 101~108, 1997.
- [14] K. P. Sycara, "Persuasive argumentation in negotiation," *Theory and Decision*, Vol. 28, pp. 203~242, 1990.
- [15] G. Zlotkin and J. Rosenschein, "Cooperation and conflict resolution via negotiation among autonomous agents in noncooperative domains," *IEEE Transactions on Systems, Man, and Cybernetics*, 21(6), pp. 1317~1324, 1991.
- [16] O. Shehory and S. Kraus, "Coalition formation among autonomous agents: Strategies and complexity," *Fifth European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, Springer-Verlag, Heidelberg, Germany, pp. 56~72, 1993.
- [17] T. W. Sandholm and V. R. Lesser, "Coalitions

among computationally bounded agents," *Artificial Intelligence*, Vol. 94, pp. 99~137, 1997.



양 승 룡

1992년 한국해양대학교 항해학과(학사)  
2000년 국방대학교 전산학과(석사). 2003  
년~현재 연세대학교 컴퓨터과학과 박사  
과정. 관심분야는 진화형 게임, 에이전트  
모델링, 생체인식



조 성 배

1988년 연세대학교 전산학과(학사)  
1990년 한국과학기술원 전산학과(석사)  
1993년 한국과학기술원 전산학과(박사)  
1993년~1995년 일본 ATR 인간정보통신  
연구소 객원연구원. 1998년 호주 Univ.  
of New South Wales 초청연구원. 1995  
년~현재 연세대학교 컴퓨터과학과 정교수. 관심분야는 신  
경망, 패턴인식, 지능정보처리