연합의 진화를 통한 IPD게임전략의 일반화 능력 개선

서 연규 °, 조 성배
연세대학교 컴퓨터과학과

Improving Generalization Ability of IPD Game Strategy by Evolution of Coalition

Yeon-Gyu Seo and Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

사회나 경제와 같은 동적 시스템에서 행동에 대한 적절성은 주위의 개체들에 의해 평가되고 일반적으로 동적 시스템에서 개체들의 행동은 주위 상황의 변화에 민감한 자극-반응의 형태로 나타난다. 본 논문에서는 그와 같은 동적 시스템을 간단한 반복적 죄수의 딜레마게임으로 모델링하고 에이전트들의 연합을 통해 일반화 능력을 향상시킴으로써 환경변화에 보다 적응적으로 반응하도록 한다. 이를 위해 반복적 죄수의 딜레마 게임에서 획득된 전략 연합에서 에이전트들의 신뢰도를 조정함으로써 일반화 능력이 향상되도록 하였다. 실험결과, 전략 연합에서 에이전트들의 신뢰도를 진화적으로 조정함으로써 일반화 능력을 크게 향상시킬 수 있음을 볼 수 있었다.

1. 서 론

최근 들어 동적인 환경에 대한 중요성이 인식되면서 이에 대한 연구가 활발히 진행되고 있다. 사회나 경제에서 발생하는 복잡한 현상에 대한 분석은 동적 환경에 대한 연구로 볼 수 있다. 생물학이나 경제학, 사회학, 수학 등 다양한 분야에서 연구되어온 경제 및 사회현상에 대한 분석은 주로 게임 이론적 측면에서 시도되어 왔지만[3], 한계에 도달하였으며 최근 들어 컴퓨터를 이용한 진화적 접근이 시도되고 있다. 그 중에서도 배반과 협동의 단순한 구조에서 협동의 진화를 통한 갈등의 해결과 분석을 위해 죄수의 딜레마 게임(Prisoner's Dilemma Game)[1, 2]이 많이 사용된다. 죄수의 딜레마 게임에서 협동의 발현요인을 분석하는 것은 복잡한 현상에 대한 이해를 제공하며 갈등에 대한 해결의 실마리를 제공할 수 있다.

죄수의 딜레마 게임은 복잡한 선택적 갈등문제를 '협동'(C: Cooperation)과 '배반'(D: Defection)이라는 두 가지 선택 구조로 다룬다. 게임자는 배반과 협동 중에서 반드시 하나를 선택해야 하는데 표1은 죄수가 2명인 경우의 선택조합에 대한 이득표를 보여주고 있다. 이 표에 의하면 게임자는 상대방의 선택에 관계없이 배반하는 경우의 이득이 더 크기 때문에 배반이 최선의 선택처럼 보인다. 그러나 이러한 논리는 게임이 무한히 반복되거나 유한하더라도 반복회수를 알 수 없는 경우에는 맞지 않는다. 왜냐하면 모두 배반하여 얻는 이득 P보다 모두 협동했을 때 얻을 수 있는 이득 R이 더 크며(R > P), 게임이 무한히 반복되고 집단 내에 조건적 협력전략들이 다수 존재하면 상시 배반전략 (All-D)보다 협동전략들의 점수가 높게 나타나서 배반전략들이 집단에서 사라지게 되기 때문이다[2].

일반적으로 경제나 사회와 같은 동적 시스템에서 개체들은 환경의 변화에 따라 다양한 행동을 보인다. 인공지능 분야에서도 생물체의 자연적 진화시스템에 대해 오랫동안 연구하여왔지만 대부분의 문제영역은 고정된 환경에서 이루어져 환경 변화에 대해 적절한 행동을 보이지 못하는 경우가 많다. 이는 주어진 자극에 대한 고정된 단일전략을 사용하기 때문으로 볼 수 있다. 동적인 환경에서 적응적 행동을 보이는 시스템은 여러 환경에서 적응적 행동을 보일 수 있는 행동전략들을 유지하고 주어진 환경에서 가장 적응적 행동을 보이는 전략으로 매핑되도록 함으로써 구현될 수 있다[3].

	협동	배반
협동	R	S
배반	T	P

표 1. 2IPD게임의 이득표. T > R > P > S, 2R > T + P

죄수의 딜레마 게임에서 적 전략의 변화에 따라 적응적으로 대응하도록 하기 위해서 다수의 우수전략들을 유지하고 주어진 상황에서 적응적 전략을 가진 에이전트로 매핑하기도 한다. 진화적 방식에서 다수의 우수 전략들을 얻기 위한 방법으로 적합도 분할(Fitness Sharing) 방법이 사용될 수 있는데 적합도가 여러 유사한 개체들에 분할됨으로써 하나의 해로 쏠리는 현상을 방지하는 방법이다[4]. 본 논문에서는 반복적 죄수의 딜레마 게임에서 다수 우수전략으로 구성된 전략 연합에서 에이전트들의 의사반영 비율(신뢰도)을 진화적으로 조정함으로써 상대가 바뀜에 따라 다르게 반응하도록 하여 연합의 일반화 능력[6]을 향상시키는 방법을 제안하고자 한다.

2장에서는 전략연합에서의 의사결정에 대해 설명하고 3장에서는 전략 연합의 일반화 능력을 향상시키기 위해 연합에 속한 에이전트들의 신뢰도를 조정하는 방법과 일반화 성능을 평가하기 위한 방법을

제시한다. 4장에서는 이러한 방법에 의한 실험결과를 제시하고 분석한 후 마지막으로 결론을 맺는다.

2. 전략 연합

반복적 죄수의 딜레마 게임에서 다수의 해를 얻기 위해 전략연합을 사용할 수 있다. 이 때 전략 연합은 다수의 우수한 에이전트들로 구성되며 연합의 의사결정을 위해 연합에 속한 여러 에이전트들의 순위에 의해 의사반영 비율이 결정된다. 이 때 순위는 에이전트들끼리 2IPD 게임을 통해 결정되며 전체 에이전트의 신뢰도가 1이 되도록 각 에이전트의 신뢰도를 부여한다[5].

그림 1은 연합의 의사결정과정을 보여주고 있는데 그림에서 A, B, C, D는 연합 내의 에이전트를 나타내고 A_C 와 A_D는 각각 에이전트 A의 협동과 배반 행동에 대한 가중치로서 에이전트의 신뢰도를 반영한다. 이러한 배반과 협동에 대한 에이전트들의 가중치 합 중에서 높은 것을 다음 행동으로 선택하는 방법(Weighted Voting)을 사용한다. 이러한 방법이 가지는 문제점은 전략 연합내에 속한 에이전트에 대한 신뢰도가 고정적일 경우 하나의 에이전트의 전략으로 표현될 수 있기 때문에 상대 게임자의 전략이 바뀌게 되면 적응하지 못하는 경우가 발생할 수 있다는 점이다. 이를 위해 에이전트의 신뢰도를 동적으로 조정하는 방법이 필요하다.

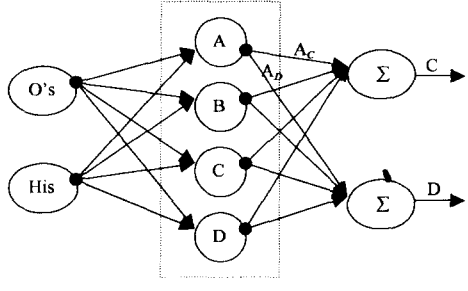

그림 1. 전략적 연합의 전략행동 결정. O's : 상대방의 이전 행동, His : 자신의 이전 행동

3. 전략적 연합을 이용한 일반화 능력의 향상

연합내의 에이전트들은 연합의사의 결정에 대한 자신의 의사 반영 비율(신뢰도)을 가지고 있는데 에이전트들에게 연합이득을 분배하기 위한 비율과도 일치한다. 만약 신뢰도가 고정적이라면 항상 똑같은 전략만 구사하게 되어 또 다른 하나의 전략으로 표현될 수 있으며 다른 하나의 전략연합이 진화 상에서 사라질 수도 있다. 이 경우에 각 상황에서 가장 잘 적응하는 에이전트들의 신뢰도를 높이는 방법으로 연합의 일반화 능력을 향상시킬 수 있다.

연합에는 여러 에이전트들이 있는데 에이전트들의 신뢰도 조정 방법은 상대방 행동 관찰가능 범위에 따라 크게 두 가지로 분류될 수 있다. 첫째, 상대방이 구사하는 전략이 무엇인지 안다면 상대방을 이길 수 있는 전략과 대전하게 하는 게이팅 방법이다[4]. 이 방법은 게임중에 상대방 구사전략을 파악하여야 하는데 상대 전략을 완전하게 알 수 없기 때문에 상대 전략의 근사적 모델링 방법을 사용하는 것이다. 그리고 또 다른 방법은 히스토리 테이블크기에서 최적 행동을 탐색하는 방법이다. 두 가지 방법 중에서 적 전략에 대한 완벽한 모델링이 가능하다면 전자가 더 나은 결과를 가져올 수 있지만, 구현상의 난점으로 상대방을 완전히 모델링할 수는 없다. 반면 후자의 경우 히스토리 테이블을 이용하기 때문에 구현이 쉽고 히스토리 크기 내에서는 최적 행동전략을 구사할 수 있다는 장점을 가지기 때문에 여기서는 후자를 이용하여 신뢰도를 조정하는 방법에 대해 살펴본다.

에이전트들로 구성된 연합의 일반화 능력을 향상시키기 위해 연합에서 잘 알려진 전략을 학습에이전트들로 선정하여 이들 전략에 잘 적응하도록 에이전트들의 신뢰도를 진화적으로 조정한다. 하나의 연합은 그림 2와 같이 상대방과 자신의 히스토리 테이블에 따른 모든 가능한 조합에 대한 에이전트들의 신뢰도를 테이블로 가지고 있다.

집단은 다른 신뢰도를 사용하는 여러 연합으로 구성되며 연합내의 에이전트들은 동일하다. 연합내의 각 에이전트에게 신뢰도가 주어지는데 히스토리 크기가 2이고 연합이 4개의 에이전트로 구성되는 경우에 히스토리가 1000이면 첫 번째 에이전트들의 신뢰도집합을 이용해 에이전트들에게 신뢰도를 부여한다. 예를 들어, 첫 번째 신뢰도 집합이 (0.4, 1.0, 1.2, 0.4)일 경우 에이전트1이 0.4, 에이전트2가 1.0, 에이전트3이 1.2, 그리고 에이전트4가 0.4의 신뢰도를 가지고 연합의 의사결정에 참여한다.

연합에서 에이전트들의 신뢰도를 진화시키기 위해 학습집합의 모든 에이전트와 라운드로빈 방식의 게임을 하고 적합도가 높은 연합의 신뢰도 조정전략을 선택과 교차, 돌연변이를 통해 진화시킨다. 이 때 돌연변이는 해당 에이전트의 신뢰도를 임의로 생성하는 방법을 사용한다.

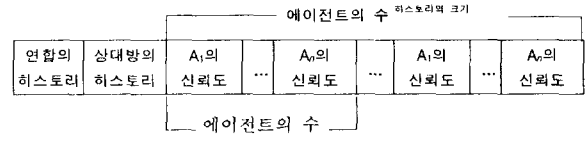

그림 2. 연합에서의 에이전트들에 대한 신뢰도테이블. A: 에이전트

4. 실험 결과

실험에서 집단은 50개의 전략연합들로 구성되며 하나의 연합은 진화과정에서 얻은 3개의 에이전트들로 구성된다. 에이전트의 신뢰도 조정을 위해 사용된 학습전략들은 TFT, Trigger, AllD, TF2T [2]로 하였다. 연합에서 에이전트의 신뢰도 진화를 위해 교차율은 0.6, 돌연변이율은 0.001, 그리고 $\mu-\lambda$선택방법을 사용하였다. 그리고 2IPD게임의 이득표에서 T, R, P, S는 전형적인 2IPD게임의 이득가운 5, 3, 1, 0으로 하였다.

그림 3에서 그림 5는 연합내 에이전트들의 신뢰도를 진화적으로 조정한 결과를 보여주고 있는데 각 실선은 실험회수를 의미한다. 그림을 보면 연합의 전체적인 적합도가 향상되는 것을 볼 수 있으며(그림 3), 조건적인 협동을 보이는 TFT나 Trigger 및 TF2T에 대해서는 서로 협동했을 때 얻을 수 있는 이득인 3으로 진화하고 있어 대부분 협동으로 진화하고 있음을 알 수 있다(그림 4). 또한, AllD과의 게임에서는 서로 배반했을 때 얻을 수 있는 이득인 1에 근사한 값으로 진화하고 있어 거의 대부분 배반함을 알 수 있다. 이는 배반전략에 대해서는 배반함으로써 배반전략의 적합도를 상대적으로 낮추고 있음을 알 수 있다(그림 5).

초기세대에서 300개의 에이전트를 임의생성하고 그 중에서 우수한 상위 30개 에이전트들을 적 전략으로 추출하여 연합전략의 일반화능

력을 평가하였다. 실험 결과, 표 2에서 볼 수 있듯이 신뢰도를 조정한 연합전략의 경우 신뢰도를 조정하지 않은 경우와 큰 차이를 보이고 있다. 평가결과 신뢰도가 조정된 연합전략의 경우에 일반적으로 상대방의 배반에 대해 관대하지 않은 Trigger나 상시배반전략을 제외하고 가장 우수한 성능을 보여주었다. 반면 Axelrod 토너먼트 환경에서 승자로 나타난 TFT와 그의 유사종 TF2T의 경우 좋지 않은 성적을 보여주고 있다. 이러한 결과는 성능평가에 사용된 적 전략에 배반전략이 많고 협동을 하는 전략이 상대적으로 적기 때문으로 보인다. 만약 적 전략에 조건적인 협력전략종이 다수 존재한다면 AllD와 같은 전략의 평균 적합도는 매우 낮아질 것이고 TFT나 연합전략 그리고 Trigger TF2T와 같은 전략은 적합도가 향상될 것이다.

5. 결 론

본 논문에서는 우수한 에이전트들로 구성된 전략 연합의 일반화 능력을 향상시키기 위해 잘 알려진 우수 전략을 학습 전략으로 하여 진화적으로 신뢰도를 조정함으로써 연합의 일반화 능력을 향상시키고자 하였다. 실험결과, 반복적 죄수의 딜레마 게임에서 사회현상에서 흔히 발생하는 전략연합을 이용하여 우수한 여러 전략들을 얻을 수 있었고 획득된 연합에서 에이전트들의 신뢰도를 조정함으로써 연합의 일반화 능력이 크게 향상됨을 알 수 있었다. 학습전략에 의한 연합의 신뢰도진화의 형태는 조건적인 협동을 구사하는 전략에 대해서는 대부분 협동으로 진화하였으며 배반하는 전략에 대해서는 계속하여 배반하는 전략으로 진화하였다.

본 논문에서는 연합의 일반화 능력 향상을 위해 학습전략으로 고정적인 전략을 사용하였다. 실제 환경에서는 적 전략도 같이 동적으로 변하는 것이 일반적이므로 적응적 적 전략에 대한 연합의 일반화 능력에 대한 검토가 이루어져야 할 것이다.

참고 문헌

[1] Axelrod, R., *The Evolution of Cooperation*, Basic Books, New York, 1984.

[2] Axelrod, R., "The evolution of strategies in the iterated prisoners dilemma," *Genetic Algorithms and Simulated Annealing*, pp. 32~41, 1987.

[3] Colman, A.M., *Game Theory and Experimental Games*, Oxford, England, 1982.

[4] Darwen, P.J., and Yao, X., "Speciation as automatic categorical modularization," *IEEE Transaction on Evolutionary Computation*, vol. 1, no. 2, pp. 101~108, 1997.

[5] 서 연규, 조 성배, "반복적 죄수의 딜레마 게임을 이용한 다중 에이전트의 전략적 연합에 관한 연구," *춘계 한국인지과학회 학술대회*, pp. 215~221, 1999.

[6] 서 연규, 조 성배, "N명 죄수딜레마 게임에서 지역화된 상호작용을 통해 진화된 전략의 일반화 능력," *춘계 한국정보과학회 학술대회*, pp. 230~232, 1999.

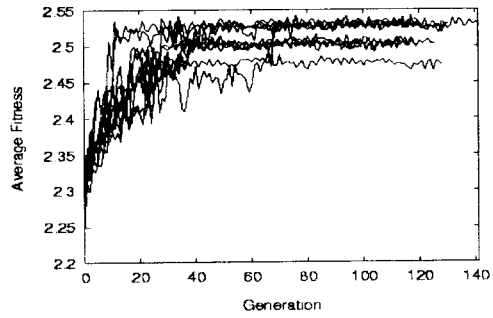

그림 3. 전략 연합의 신뢰도 진화

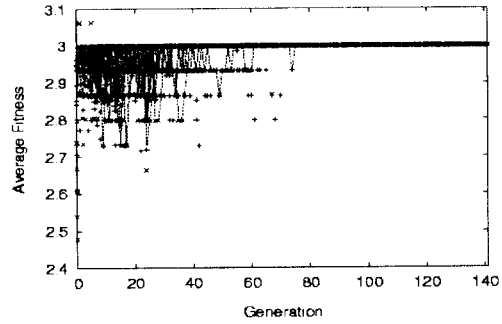

그림 4. 조건적 협동 전략들과의 진화결과

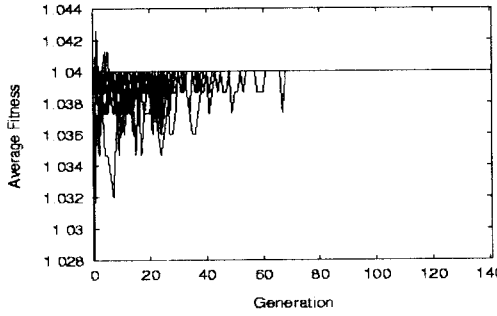

그림 5. 상시 배반전략과의 진화결과

전략		승리	비김	평균	적 전략의 평균
연합전략	신뢰도 조정전	8.64±4.9	6±2.19	1.84 ±0.28	1.75 ±0.59
	신뢰도 조정후	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
T2FT		7	0	1.54	2.40
AllD		30	0	2.17	0.7

표 2. 연합전략과 학습전략들의 적전략에 대한 일반화 성능 평가결과