자기 적응형 교배기법을 이용한 반복적 죄수 딜레마 게임의 진화적 협동 수렴 분석

김찬중, 이종현, 안창욱*
성균관대학교 정보통신공학부
e-mail : {rlackswnd, ljh08375, cwan}@skku.edu

Analysis on the a Self Adaptive Crossover for Iterated Prisoner's Dilemma Game of Evolutionary Convergence

Chan Joong Kim, Jong-Hyun Lee, Chang Wook Ahn*
School of Information and Communication Engineering
Sungkyunkwan University

요 약

본 논문에서는 경제학, 사회학, 수학 분야에서 수십년 전부터 연구해오던 죄수의 딜레마 게임의 협동진화에 대해 고찰해보고자 한다. 반복적 죄수의 딜레마 게임은 게임이론의 가장 기본적인 이론으로써, 사회적 상호작용, 경제활동, 국제관계 등 다양한 현상들을 모델링 하기 위한 하나의 방법이다. 그 중에 N명이 참가하는 반복적 죄수 딜레마 게임의 전략은 유전 알고리즘(Genetic Algorithms, GAs)을 통해 진화적으로 만들어 낼 수 있으며, 이 경우에 그 결과를 일반적인 내쉬 균형 이 아닌, 모든 개체들이 유전알고리즘을 통해 협동으로 수렴하도록 유도할 수 있다는 사실은 상당히 시사하는 바가 크다. 기존에 주로 연구되어오던 죄수의 딜레마 게임은 협동으로의 수렴과정에서 일반적으로 순위기반선택(Rank-based selection)과 1점 교배기법(1point crossover)을 사용한다. 그러나 순위기반선택은 모든 개체에 순위을 매겨야 하기 때문에, 개체수가 커질수록 성능이 저하되며, 1점 교배기법은 개체 값이 분산되어있을 경우, 최적해(Optimal solution)을 찾기 힘들다는 단점이 있어, 개체수가 많은 경우에 적용하기에는 비효율적이다. 본 논문에서는 토너먼트 선택기법(Tournament selection)과 자기 적응형 교배기법(Self-adaptive crossover)을 적용한 새로운 기법을 제안한다. 또한 기존 기법과 비교실험을 통해 제안기법이 기존기법에 비해 평균 수렴시간과 수렴 횟수에서 뛰어난 성능을 보이고 있음을 확인하였다.

1. 서론

유전 알고리즘은 자연세계의 생물의 진화 및 발생에 관계된 메커니즘을 바탕으로 컴퓨터상에서 시뮬레이션함으로써 복잡한 실세계의 문제를 해결하고자 하는 계산 모델이다. 유전 알고리즘(Genetic algorithms, GAs)은 구조가 간단하고 방법이 일반적이어서 응용범위가 매우 넓으며, 특히 적응적 탐색과 학습 및 최적화를 통한 공학적인 문제의 해결에 많이 이용되고 있다. 유전 알고리즘은 특히 다각도에서 동적으로 변화하는 환경에 적합한 시스템을 구축하는데 주로 적용되고 있으며, 최근 몇 년간의 사회적, 경제적 현상들의 급격한 변화 속도에 발맞추어 보다 향상된 시스템의 개발이 요구되고 있는 시점이다. 이런 문제 해결을 위해 과거 경제학에서 주로 사용되던 게임이론을 접목하는 사례가 늘고 있다 [3]~[9].

본 논문에서는 게임이론 중에서도 가장 널리 알려진 죄수의 딜레마, 그 중에서도 반복적인 죄수의 딜레마 게임(NIPD)에서 공진화를 통해 개체군들이 협동으로 수렴할 수 있음을 보여주는 기존 실험에서 사용된 순위기반선택 기법(Rank-based selection)과 1점 교배기법(1 point crossover)의 문제점을 분석하는 한편, 보다 효율적인 기법을 도출하기 위해 토너먼트 선택기법(Tournament selection)과 자기 적응형 교배기법(Self-Adative Crossover)을 고찰하고 이들 기법의 비교분석을 통해 보다 제안 기법의 성능향상을 얻고자 하였다.

2. NIPD게임과 GA

2.1. 죄수의 딜레마 게임

죄수의 딜레마 게임은 경찰에 체포된 두 피의자에 관한 이야기다. 경찰은 현재 이 두 사람이 1년씩 징역형을 받을 비교적 가벼운 범죄에 대해 확실한 물증을 가지고 있으나, 이들이 보다 심각한 은행강도 범죄를 저질렀다는 심증만 있을 뿐, 확실한 물증은 없다. 경찰은 두 사람을 독방에 감금하고 자백을 강요한다. 만일 두 사람이 모두 혐의를 인정하지 않으면(협동) 둘 다 1년간 징역을 살고, 둘 중에 한명이 다른 방에 있는 사람을 주범이라 증언(배반)하면, 그 사람은 석방되고, 대신 다른 사람은 20년형을 선고받는다. 그러나 둘 다 자백을 하면(배반) 공범으로 8년형을 살게 된다. 이 상황에서 두 용의자에게 가장 좋은 것은 둘 다 혐의를 부인하는 것이나, 서로 상대방의

전략을 알 수 없으므로, 자백하는 것이 각자의 우월전략이 되어, 결국 둘 다에게 불리한 결과를 초래하는 게임 상황을 "죄수의 딜레마"라 칭한다. 게임의 참가자들은 매 순간마다 "배반" 혹은 "협동"을 선택해야 하며 참가자들이 얻는 이득은 다음의 표와 같다 [7].

<표 1> 2IPD 게임의 이득표

	상대의 협동	상대의 배반
나의 협동	R(-1)	S(-20)
나의 배반	T(0)	P(-8)

$T > R > P > S$이고, $2R > T + P$
죄수 딜레마 게임은 N명의 참가자들에 대해서도 확장할 수 있으며, 이를 NIPD게임으로 약칭한다. 이 경우, 각 참가자가 얻는 이익은 다음과 같다.

<표 2> NIPD 게임의 이득표

나를 제외한 협동자수	0	1	....	N-2	N-1
나의 협동	C(0)	C(1)	....	C(N-2)	C(N-1)
나의 배반	D(0)	D(1)	....	D(N-2)	D(N-1)

단, x=0, 1, 2,...,N-1인 모든 x에 대해, 다음과 같은 조건을 갖는다.

<표 3> NIPD 게임의 조건

만약 $D(x) > C(x)$을 만족하면 $D(x+1) > D(x)$, $C(x+1) > C(x)$ $C(x) > D(x) + C(x-1)/2$

본 논문에서는 함수 C와 D를 수식 1과 같이 가정하였다.

$$C(x) = 2x,\ D(x) = 2x + 1 \qquad (1)$$

이 경우, N명의 참가자 k명이 협동했다면, 참가자의 전체 이익 S(k)과 참가자들의 평균 이익 P(n)은 표 4와 같다.

<표 4> 전체이익S(k) 과 평균이익P(n)의 정의

$$S(k) = 2(k-1)k + (2k+1)(N-k)$$
$$S(k+1) = 2k(k+1) + (2k+3)(N-k-1)$$

$$S(k)\text{의 계차수열 } B(k) = S(k+1) - S(k) = 2N-3$$

$$S(n) = S(0) + (B(0) + \ldots + B(n-1)) = (2N-3)n + N$$
$$P(n) = S(n)/N = 1 + (2N-3)n/N$$

따라서 P(n)값을 알면 협동하는 비율(n/N)을 알 수 있다[4].

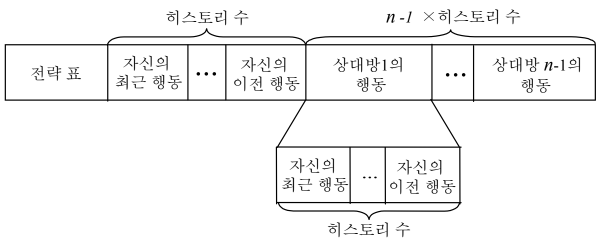
2.2 전략의 유전자형 표현

전략의 유전자형 표현을 위해 NIPD에서는 히스토리와 전략테이블을 이용한다. 죄수의 딜레마 게임에서 각 게임자가 다음 행동을 결정하는데 참조할 수 있는 정보는 바로 이전 단계의 상대방과 자신의 행동들이기 때문에 자신의 전략을 결정하는데 중요한 역할을 한다.

2.2.1 Axelrod의 방법

Axelrod는 전략의 표현에 있어 상대방의 이전 행동들에 대한 히스토리를 모두 저장하는 방법을 사용한다. 따라서, N명의 참가자가 게임을 할 때, 하나의 결과(round)는 다음과 같이 N-bit로 표현할 수 있다.

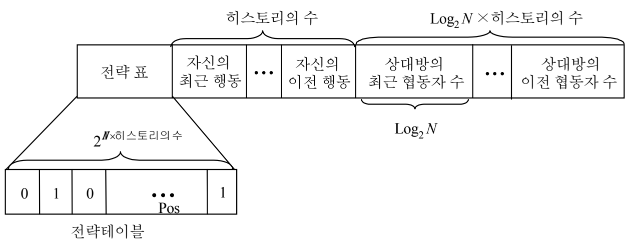
만일, 각 참가자가 최근 h개의 결과를 기억한다면, History는 H = hN개의 bit로 표현된다. 하나의 '전략'은 가능한 H-bit의 모든 조합에 대해 다음 행동을 결정할 수 있어야 하므로 2^H개의 bit로 표현된다. 그런데 게임을 처음 시작할 때는 히스토리가 없어, 히스토리를 임의로 설정해줘야 되므로, 하나의 전략은 $2^H + H$개의 bit로 표현할 수 있다 [6].



(그림 1) Axelrod전략의 유전자 표현

2.2.2 Yao와 Darwen의 표현방법

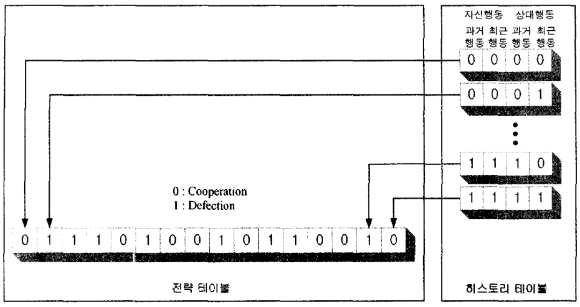
Axelrod의 표현방법은 참가자의 수에 따라 사용하는 메모리가 기하급수적으로 커진다. 또한, 각 참가자에게는 누가 협동했는가보다는 몇 명이 협동했는가가 중요하므로, Yao와 Darwen 표현방법에서는 라운드를 다음과 같이 표현할 수 있다 [8].



(그림 2) Yao와 Darwen의 유전자 표현에서 히스토리 테이블의 구성

따라서 Yao와 Darwen의 유전자형 표현방법을 활용하면 그림 3과 같은 형태로 나타낼 수 있다. 히스토리 테이블은 자신과 상대방의 과거 행동을 기억하여 다음 전략을 결정하는데 사용된다. 예를 들어 자신의 과거행동이 협동이고, 최근 행동이 배반이면 01로 표현되며, 상대방의 과거

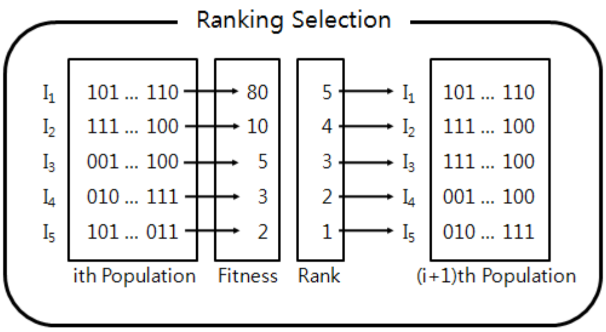
행동이 배반이고 최근 행동이 협동이면 10이 되어 전체 히스토리는 "0110"이 된다. 이를 십진수로 변환하면 "6"이 되고 전략테이블의 7번째 값을 다음 전략으로 선택하게 된다 [1].



(그림 3) 2IPD게임에서 전략의 비트스트링 표현

2.3 Selection 기법

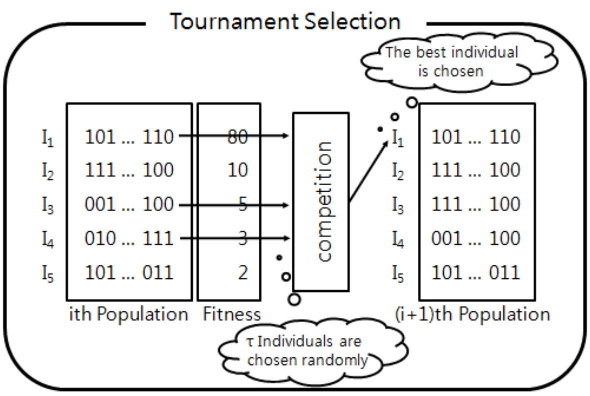
2.3.1 순위기반 선택기법(Ranking Selection)



(그림 4) 순위기반선택(Ranking Selection)의 형태

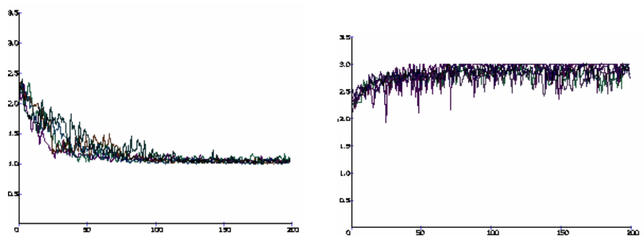
순위기반 선택기법은 각 개체들의 적합도(최적해에 근접한 정도를 나타내는 값)의 순서에 맞춰서 순위를 매기는 기법이다. 개체수가 적을 경우에는 효과적일 수 있으나, 개체수가 많아질 경우, 매 세대(Generation)마다 일일이 모든 개체들의 순위를 매기는 데 걸리는 시간이 지나치게 커져 비효율적인 기법이다.

2.3.2 토너먼트 선택기법(Tournament Selection)



(그림 5) 토너먼트 선택기법(Tournament Selection)의 형태

토너먼트 선택기법은 매번 2개 이상의 개체를 선택하여 가장 높은 값을 갖는 개체 값들만 다음 세대로 복사하는 형태로, 개체수가 아무리 증가하더라도, n(상수)배 이내이기 때문에 수렴 시간(Convergence time)이 상대적으로 적다.



(a)순위기반 선택기법 (b) 토너먼트 선택기법

(그림 6) 개체 선택 방법별 진화과정

위의 실험(개체수 100, 교차율 0.3, 돌연변이율 0.001, 2점 교차방식)은 순위기반 선택방법과 토너먼트 방법으로 2IPD 게임을 진행한 것으로, 협동율을 나타내는 세로축의 pay-off값이 순위기반 선택에 비해 토너먼트 기법일 때 더욱 높은 값으로 수렴하고 있음을 확인할 수 있다 [2].

2.4 자기적응형 교배 기법

교배의 교차점(Crossover Point) 수는 교배 기법을 결정 짓는 중요한 매개변수이다. 이 매개변수의 결정은 일반적으로 매개변수 조율에 의해 이루어지는데, 이는 시간과 노력을 소모하게 만들고 효율적이지 않다. 이 점을 개선하기 위해 본 논문에서는 자기 적응형 교배기법을 적용하였다. 이 기법은 속도엔 영향을 끼치지 않고 결과가 우수한 교배기법을 판단하기 위해 교배교차점 간의 경쟁기법을 적용한다. 즉, 전체 개체군을 여러 Group으로 나누어 각각 다른 교배를 수행하게 하고 이와 같은 교배 기법 사이의 경쟁을 각 세대마다 반복하게 되면 최종적으로 문제에 가장 적합한 교배 기법만이 남게 된다.

처음 알고리즘의 시작 단계에서 전체 개체군을 일정 크기의 그룹으로 나누고, 각 군집에 서로 다른 교배의 교차점 수를 지정한다. 각 그룹 내에서 적합성 평가와 선택 연산 및 교배(Crossover) 연산을 수행한 후, 다시 적합성 평가와 선택 연산을 한다. 이 때, 선택 연산에서 선택된 개체들이 가장 많이 포함되어 있는 그룹을 최적그룹, 가장 적게 포함되어 있는 그룹을 최악 그룹으로 지정하여, 다음 교배 연산 전에, 최적 그룹의 교차점 수를 최악 그룹의 교차점 수에 대입해 준다. 교배의 교차점 수는 그룹 간에 적자생존을 통해 상대적으로 우수한 매개변수가 더 많은 그룹들을 차지하게 되고, 최종적으로는 하나의 매개변수로 수렴하게 된다 [5].

3. 실험 결과 및 고찰

우선, 기존 기법을 소개한 논문을 바탕으로 본 논문에서 제안한 새로운 기법을 적용한 실험을 수행하였고 각 기법간의 수렴시간과 횟수를 측정하였다.

3.1 실험매개변수

- 개체수(Population) : 100
- 히스토리 수(History) : 2
- 세대당 게임 횟수 : 1000
- 게임 당 라운드 수(Round) : 100
- 최대 세대 전환횟수(Generation) : 1000
- 교배율(Crossover) : 0.6 (60%)
- 돌연변이율 (Mutation) : 0.001 (0.1%)

- 기존기법 : 순위기반 선택기법
  임의의 위치에서의 1점 교배기법
- 제안기법 : 토너먼트 선택기법
  자기적응형 교배기법

위의 매개변수를 활용하여 20번씩 실행하며 각 실행은 협동율이 5세대 연속 95%를 넘는 경우, 즉시 중지되며, 그렇지 않은 경우 최대 세대 전환횟수까지 실행된다 [4].

<표 5> 기존기법과 제안기법간의 성능비교

게임 크기	협동 수렴횟수 (Total 20 runs)		평균 수렴시간 (Generation)	
	기존기법	제안기법	기존기법	제안기법
2IPD	16	**20**	191.6	**48.85**
4IPD	9	**20**	134.8	**58.6**
8IPD	14	**20**	228.8	**24.55**
16IPD	8	**20**	188.2	**18.8**

3.2 실험결과분석

실험 결과, 기존 기법이 20번의 게임 횟수 중에 4IPD, 즉 4인이 참가하는 경우는 협동으로의 수렴율이 50%에도 못 미치는 9회에 불과했으며, 가장 높은 수렴율을 보인 2IPD게임의 경우에도 80%의 수렴율을 나타냈다. 그에 반해, 본 논문에서 제안한 기법을 활용한 경우, 모든 경우에서, 100%의 협동 수렴율을 나타냈다.

평균 수렴시간에서도 기존 기법에 비해 우수한 성능을 나타내었다. 총 1000세대 중에 기존기법은 130~230세대 정도의 평균수렴시간을 나타낸 반면, 제안 기법의 경우, 2인 게임이 액 49세대, 4인 게임이 58세대, 8인 게임이 약 24세대, 16인 게임이 약 19세대로 기본기법 대비 최소 10%에서 최대 43% 정도의 시간이 소모되었다.

위 실험결과를 통해 논문의 제안 기법이 기본 기법에 비해, 성능 우위와 더불어 시간적 효율성 측면에서도 우수함을 확인할 수 있었다.

4. 결론

반복적 죄수의 딜레마 게임에서 협동 수렴율 및 수렴시간을 단축하기 위해, 공진화 분야에서 많은 기법들이 연구되어 왔다. 본 논문에서는 죄수의 딜레마 게임에서 수렴시간(Convergence time)이 긴 순위기반 선택기법과 탐색효율이 낮은 1점 교배기법을 사용하는 대신, 토너먼트 선택기법(Tournament selection)과 자기적응형 교배기법(Self-adaptive crossover)을 적용함으로써 일반화 성능이 향상된다는 것을 실험 결과로서 확인하였다.

평균 협동 수렴율과 세대수(Generation)에 기초한 평균 수렴시간간의 직접적인 결과 값의 비교를 통해, 최적해로의 수렴율과 시간적 효율성 양쪽 측면에서 제안 기법이 기존 기법에 비해 훨씬 우수하였다.

본 논문에서 제안한 기법은 보다 복잡하고 다각도로 변화하는 사회적, 경제적 현상들에 NIPD 모형을 적용함에 있어 유용한 역할을 할 수 있을 것이다. 또한 기존 연구의 경우, 실제 현실에 접목하기에는 일부 한계성을 안고 있었으나, 본 연구를 통해 반복적 죄수의 딜레마 게임을 보다 현실적인 사회, 정치, 경제 현상들에 유용하게 접목 시킬 수 있을 것으로 예상된다.

참고문헌

[1] 양승룡, 조성배, “진화학습을 이용한 다중에이전트의 일반화 성능향상을 위한 전략적 연합”, 정보과학회논문지, Vol. 2. pp.101-110, 2004.

[2] 양승룡, 조성배, “다중에이전트를 이용한 IPD게임에서 전략적 연합의 체계적 성능 평가”, 한국정보처리학회 추계학술발표논문집 Vol. 1, pp. 315-318, 2002.

[3] 양승룡, 노현걸, 조성배, “NIPD 게임에서 게임자 수와 협동의 관계에 대한 진화적 연구”, 한국지능시스템학회, pp. 306-312, 1999.

[4] 이태양, 장병탁, “진화적 반복 죄수 딜레마 게임의 수렴 특성 분석”, 공학사 학위논문, 서울대학교 컴퓨터공학부, 2003.

[5] 이종현, 임동현, 안창욱, “유전 알고리즘의 성능 향상을 위한 자기-적응형 교배 기법”, 한국정보처리학회 학술발표논문집, Vol. 1, pp.130-133, 2010.

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

[7] N.Gregory Mankiw, *"The Principles of Economics 4th edition"*, Thomson, 2007.

[8] Yao, X. and Daewen, P. J. "The experimental study of N-person iterated prisoner's dillema.", Informatica, Vol. 18, pp. 435-450, 1994.

[9] 최정규, “게임이론과 진화 다이내믹스”, 이음사, 2009.