

유전자 알고리즘에서 Deceptive 문제에 대한 토너먼트 선택의 최적화

김예훈, 안창욱, 김기표, R.S.Ramakrishna
광주과학기술원 정보통신공학과
Email : {yehoon, cwan, kimkp, rsr}@kjist.ac.kr

An Optimization of Tournament Selection in Genetic Algorithms for Deceptive Problems

Yehoon Kim, Chang Wook Ahn, Kipyoo Kim, R. S. Ramakrishna
Dept. of Information and Communications,
Kwang-Ju Institute of Science and Technology (K-JIST)

요 약

본 논문은 토너먼트 선택을 사용한 Deceptive Problem 에서 최적 해(optimum)를 찾으면서 수렴(convergence)속도를 향상시키기 위한 최적의 조건을 찾고자 한다. 이를 위해 적합도가 높은 염색체(chromosome)를 다음 세대로 전달하면서 동시에 적합도가 낮은 염색체에 대해서도 적정 수준 전달되게 하였다. 또한 기존의 여러 선택기법 중 가장 일반적으로 사용되는 토너먼트 선택 기법의 문제점을 고찰하고, 제안 방법으로 최적도 및 수렴속도를 모의 실험을 통해 비교 및 분석한다. 실험 결과로부터 제안 방법에 대한 수렴속도의 경향을 고찰하였다.

1. 서 론

유전자 알고리즘은 최적해(optimal solution)를 찾는 확률적 탐색 모델로서 최적화 방법중의 하나인 계산 의존적 방법과 결정론적 방법에 비해 최적 해를 찾을 가능성이 높으며, 효율적이다[1]-[6]. 그러나 해결하고자 하는 문제가 선형이 아니라 비선형인 경우 대부분의 유전자 알고리즘은 비효율적 이거나 최적 해를 찾는데 실패하게 된다[7].

유전자 알고리즘에서 선택(selection)과 교배(crossover)는 유전자 알고리즘의 성능을 결정하는 매우 중요한 연산자이다[1]-[4]. 선택 과정[4]은 우수한 형질의 염색체가 다음 세대로 충분히 전달될 수 있도록 해야 하며 교배 과정에서는 선택과정에 의해 선택된 우수한 형질의 염색체를 이용하여 새로운 염색체를 생성할 수 있어야 한다[1],[3],[5].

유전자 알고리즘에서 일반적으로 사용되는 선택 메커니즘인 토너먼트 기법은 적합도(fitness)

비교우위의 염색체를 다음 세대의 개체군으로 전달하지만 상대적으로 적합도가 낮은 염색체는 다음 세대로 전달되기 어렵다는 단점이 있다. 이는 유전자 알고리즘의 다양성(diversity)을 감소시켜 최적해(global optimum)를 찾지 못하고 지역해(local optimum)를 찾게 하는 요인이 된다[5].

따라서 이 논문에서는 유전자 알고리즘의 다양성을 감소시키지 않으면서 최적 해를 찾을 수 있는 선택방법을 제안한다. 이를 위해 적합도가 높은 염색체에 대해서는 토너먼트 기법을 적용해 다음 세대로 전달하면서 동시에 적합도가 낮은 염색체에 대해서는 토너먼트 선택기법의 경쟁염색체 순위를 고려한 확률(동적 혹은 고정 방식)로 전달되는 것을 보장하였다.

2. 배경 지식

기존의 최적화 알고리즘은 최적화 하고자 하는 목적함수에서 기울기(gradient) 정보를 얻기 위해 미분 과정을 거치면서 탐색을 수행하지만, 유전자 알고리즘은 적합도와 유전 연산자 (선택, 교배, 돌연변이)를 이용해서 최적해를 탐색한다[1]. 한 세대의 개체군에 속해있는 염색체들은 세대를 거듭할수록 이전 세대까지 획득한 유전정보를 공유하며 진화해 나간다.

최적화하고자 하는 목적함수에 따라 적절한 유전연산자의 선택은 매우 중요하다. 예를 들어, 이 논문에서 다룬 그림 1-1 과 그림 1-2 와 같은 Deceptive Problem 의 경우 이전 세대에서 발견한 Building Block(BB)를 파괴시켜서는 안되므로, 이에대한 파괴확률이 적은 1-점 교배를 사용하였다. 한편, 본 논문에서는 선택조건만 검사하기 위해 돌연변이 연산은 수행하지 않았다.[3]

2.1 유전 연산자

2.1.1 선택 연산자(selection operator)

선택 연산자는 우수한 형질의 염색체가 다음 세대로 충분히 전달될 수 있도록 해야 한다. 이를 위해서 그림 1-1 과 같은 적합도 함수(fitness function)를 이용해 염색체를 평가한 후 우수한 염색체를 다음 세대로 전달을 한다. 하지만 전달될 염색체의 적합도만 가지고 전달할 경우 최적 해를 발견하지 못하고 지역 해만 탐색할 가능성이 있게 된다.

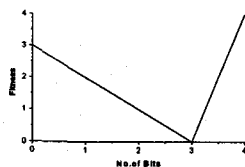
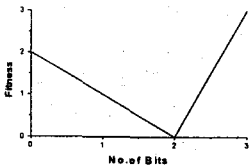


그림 1-1. 3bit deceptive function 그림 1-2. 4bit deceptive function

2.1.2 교배연산자(crossover operator)

선택연산자에 의해 전달된 우수한 형질의 염색체를 이용해 새로운 염색체를 만들 수 있어야 한다. 특히 최적화를 위한 목적함수가 비선형인 경우와 스키마(schema)의 길이가 2 이상일 경우 이전 세대에서의 얻은 유전 정보가 파괴되는 것을 막기 위하여 1-또는 2-점 교배가 주로 사용된다[3].

2.2. Deceptive Problem (DP)

그림 1-1 과 그림 1-2 는 각각 3bit/4bit Deceptive functions 을 나타낸다. x 축은 염색체에서 유전자의 값이 '1' 인 유전자 개수를 나타내며 y 축은 적합도를 뜻한다. 이러한 DP 들은 유전자 알고리즘의 성능을 평가하기 위한 수단으로 사용된다 [2] , [7].

3. 토너먼트 선택기법과 제안 선택기법

3.1 토너먼트 선택기법

토너먼트 선택기법은 개체군에 염색체가 n 개 있을 때, 2개 (혹은 그 이상)의 염색체를 무작위로 추출해 그 염색체들의 적합도를 비교해 가장 우수한 염색체를 다음 세대로 전달하는 방법이다. 하지만 이 방법은 적합도가 낮은 염색체는 전달되지 못한다는 단점을 갖고 있으며 이로 인해 다양성을 손실할 수 있다. 왜냐하면, 적합도가 낮은 염색체라 하더라도 Deceptive Problem 에서 그 염색체 내부에는 우수 구성요소(Building Block)가 포함될 수 있기 때문이다[1].

3.2 제안 선택방법

토너먼트 선택기법은 적합도가 낮은 염색체들은 다음 세대에 전달될 수 없다는 단점을 갖고 있다. 따라서 이 논문에서는 낮은 적합도를 갖고 있는 염색체들도 적절한 비율로 다음 세대로 전달했을 때 최적화된 조건이 무엇인지 알아보고 그에 따른 방법을 제안한다.

토너먼트 크기(size)가 커지면 그 중에서 적합도가 가장 좋은 염색체만을 다음세대에 넘겨주는 전통적인 토너먼트 선택의 문제점을 해결하기 위해 각 단계(level)별로 새로운 확률을 적용하여 다양성을 유지시킬 필요가 있다.

단계별로 확률을 적용시키는 방법에는 크게 두 가지가 있다. 토너먼트 크기를 s , 단계의 크기를 l 이라고 할 때 $s = l$ 인 경우와 $s > l$ 인 경우가 있다. s 가 커질 경우 단계가 늘어날수록 하위 단계에서의 확률간 차이로 인한 수렴속도와 해의 우수성 (quality) 비교는 의미가 크지 않기 때문에 s 를 비교적 적은 값으로 설정하고 그 안에서 l 의 변화에 따른 결과를 알아볼 필요가 있다.

한편, 크기(size)와 단계(level)가 간의 관계에 일정한 규칙을 적용 하기위해 확률을 적용시키는 방법을 다음 세가지로 분류할 수 있다.

1) 증가 방법 (Increasing probability way)

- 2) 감소 방법 (Decreasing probability way)
- 3) 고정 방법 (Constant probability way)

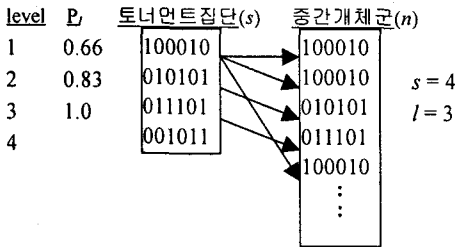


그림 2. 증가방법 토너먼트 선택의 예

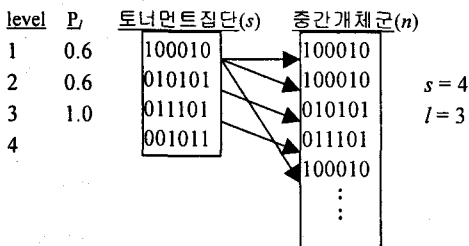


그림 3. 고정방법 토너먼트 선택의 예

각 단계별 확률을 P_i , 토너먼트 크기를 s , 단계의 길이(length)를 l 이라고 하면 증가방법의 경우 i 번째 단계의 토너먼트 확률은 다음과 같다.

$$P_i = 0.5 + (1/2l) * i, (1 < i \leq s) \quad (1)$$

단계별로 확률이 위식에 의해 적용되고 하위 모든 단계중에서 가장 좋은해가 다음세대로 전달되므로 l 번째 확률은 1.0 이다. (그림 2) 그런데 감소방법으로 확률을 적용시키면 이미 1, 2 단계에서 1.0 에 가까운 확률로 좋은 해가 선택되기 때문에 레벨의 크기가 결과에 영향을 거의 미치지 않게 되므로 고려할 필요가 없다. 고정방법의 경우 토너먼트 확률은 다음과 같다.(그림.3)

$$P_i = 0.5 + \Delta c, (0 < c \leq 0.5) \quad (2)$$

따라서, 본 논문에서는 증가방법, 고정방법 두 가지 경우에 따른 단계의 크기별 수렴속도 및 최적도를 실험을 통해 고찰 해보도록 한다.

4. 모의 실험 결과 및 분석

이 장에서는 3 비트 Deceptive Problem 에 대한 제안된 세가지 방법의 성능을 비교 및 검증한다. 모든 실험에서는 1-점 교배를 사용하고 선택기법의 성능만을 비교, 분석하기 위해 돌연변이 연산은

사용하지 않는다. 단계의 크기 l 에 따른 성능을 비교하기 위해 개체군의 크기(population size)는 100, 염색체의 길이는 3bit 의 배수인 99, 토너먼트 크기 s 는 4 로 고정시켰다. 그리고 증가방법의 경우 단계의 크기는 2, 3, 4 를 선택하여 비교하였다. 고정방법의 경우 $\Delta c = 0.6$ 으로 설정하여 적정 수준으로 좋은 해와 나쁜 해가 선택되도록 하였다. 선택 및 교배가 일어날 확률을 1.0 으로 설정하였고, Tournament selection without replacement 방식을 사용하였다. 그리고 각 방법에 대해 모든 개체군이 하나의 해에 수렴할 때까지 실험을 수행하였다. 그림에 나타난 값들은 50 번의 실험을 통한 평균값이다.

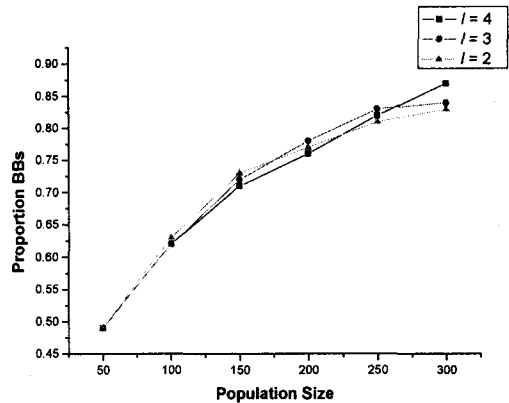


그림 4-1. 증가방법(Increasing way)에서의 최적성

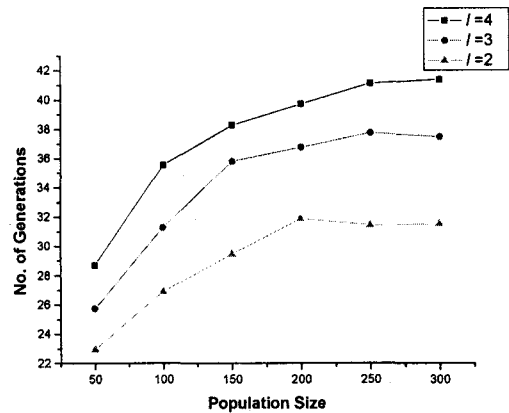


그림 4-2. 증가방법(Increasing way)에서의 수렴 세대수

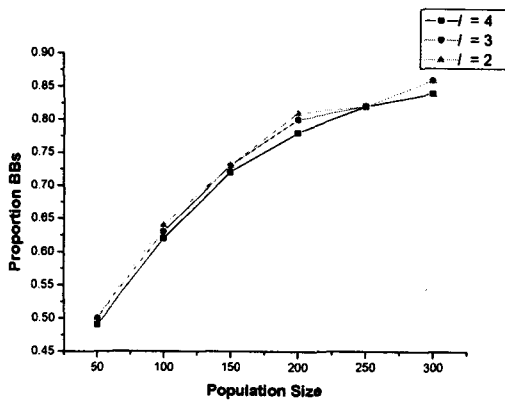


그림 5-1. 고정방법(Constant way)에서의 최적성

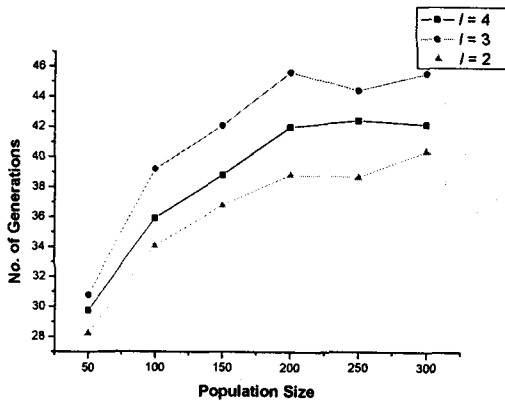


그림 5-2. 고정방법(Constant way)에서의 수렴 세대수

그림 4-1 과 그림 5-1 을 보면 같은 토너먼트 크기 안에서 단계의 크기(l)를 바꿔주었을 때 거의 비슷한 최적도를 나타낸다. 그러나 그림 4-2 와 보여지는 것과 같이 증가 방법에서의 세대 수는 단계의 크기가 작을수록 적게 나타나서 빠른 수렴을 하는 것을 알 수 있다.

한편 그림 5-2 와 같이 고정 방법에서의 세대 수는 $l = 4$ 일 때 오히려 $l = 3$ 일 때 보다 빠른 수렴이 일어나 증가 방법과는 다른 경향을 보였다.

5. 결 론

본 논문에서는 기존의 토너먼트 선택 기법의 문제점을 고찰하고 성능을 향상시킬 수 있는 최적의 조건을 찾기 위해 단계별 경향을 알아보았다. 토너먼트 선택 메커니즘은 우수 염색체를 다음 세대로 전달한다는 점에서 효과적이지만, 열성 염색체는 전달되기 힘들다는 단점을 갖고 있어서 해-

표면을 충분히 탐색하지 못했다. 그러나 제안된 두 방법에서는 열성 염색체 집단에 대해서도 여러 번 확률을 부여하여 다음 세대로 전달될 수 있도록 함으로써 다양성을 유지함과 동시에 토너먼트 선택기법의 우수 염색체 전달특성을 동시에 갖게 하였다.

6. 참고문헌

- [1] D.E.Goldberg - *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, MA: Addison Wesley, 1989.
- [2] D.E.Goldberg - *Simple genetic algorithms and the minimal, deceptive problem*. In L. Davis, editor, *Genetic Algorithms and simulated annealing*, pages 74-88. Morgan Kaufmann, Los Altos, CA., 1987.
- [3] G. Harik, E. Cantü-Paz, D.E. Goldberg, and B.L. Miller, *The Gambler's Ruin Problem, Genetic Algorithms, and the Sizing of Populations*, *Evolutionary Computation*, Vol. 7, No. 3, pp. 231-253, 1999.
- [4] C.W. Ahn and R.S. Ramakrishna, *A Genetic Algorithm for Shortest Path Routing Algorithm and the Sizing of Populations*, Vol. 6, No. 6, pp. 566-579, 2002.
- [5] X. Hue, *Genetic Algorithms for Optimization: Background and Applications*, Edinburgh Parallel Computing Centre The University of Ediburgh, Ver 1.0 February 1997.
- [6] E. Cantü -Paz, *Efficient and Accurate Parallel Genetic Algorithm*, Kluwer Academic Publishers, 2000.
- [7] A. Agapie , H. Dediu , *GA for Deceptive Problems : Inverting Schemata by a Statistical Approach*, IEEE , ICEC International Conference on Evolutionary Computation, (ICEC'96)