

유전자 알고리즘의 다양성과 수렴성을 고려한 새로운 선택기법

김기표⁰, 안창욱, R.S.Ramakrishna
광주과학기술원 정보통신공학과
{kimkp⁰, cwan, rsr}@kjist.ac.kr

A New Selection Mechanism of Genetic Algorithms for Diversity Maintenance and Fast Convergence

Kipyoo Kim⁰, Chang Wook Ahn, R.S.Ramakrishna
Dept. of Information and Communications, Kwang-Ju Institute of Science and Technology (K-JIST)

요 약

본 논문은 유전자 알고리즘의 다양성(diversity)을 유지하면서 동시에 수렴(convergence) 속도를 향상시키기 위한 새로운 선택기법을 제안한다. 이를 위해 적합도가 높은 염색체를 다음 세대로 전달하면서 동시에 적합도가 낮은 염색체에 대해서도 일정 수준 전달되게 하였다. 또한 기존의 여러 선택기법 중 가장 일반적으로 사용되는 토너먼트 선택 기법의 문제점을 고찰하고, 제안 알고리즘의 최적도 및 수렴속도를 모의 실험을 통해 비교 및 분석한다. 실험 결과로부터 제안 알고리즘은 기존의 토너먼트 선택기법에 비해 우수함을 확인 하였다.

1. 서 론

유전자 알고리즘은 최적해(optimal solution)를 찾는 확률적 탐색 모델로서 최적화 방법중의 하나인 계산 의존적 방법과 결정론적 방법에 비해 최적해를 찾을 가능성이 높으며, 효율적이다[1]-[6]. 그러나 해결하고자 하는 문제가 선형이 아니라 비선형인 경우 대부분의 유전자 알고리즘은 비효율적이거나 최적해를 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 만들 수 있어야 한다[1],[3],[5].

유전자 알고리즘에서 일반적으로 사용되는 선택 메커니즘인 토너먼트기법은 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 토너먼트 기법의 장점을 그대로 수용하는 새로운 형태의 선택기법을 제안한다. 이를 위해 적합도가 높은 염색체에 대해서는 토너먼트 기법을 적용해 다음 세대로 전달하면서 동시에 적합도가 낮은 염색체에 대해서는 일정 수준으로 전달되는 것을 보장하였다.

2. 배경 지식

기존의 최적화 알고리즘은 최적화 하고자 하는 목적함수를 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 다중 그림 1-1 과 그림 1-2와 같은 Deceptive Problem의 경우 이전 세대대에서 발견한 유전정보를 파괴시켜서는 안되므로, 유전정보를 파괴할 확률이 적은 1-점 교배, 2-점 교배 중에서 2-점 교배를 사용하였다.[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 1-1과 같은 적합도 함수(fitness function)를 이용해 염색체를 평가한 후 우수한 염색체를 다음 세대로 전달을 한다. 하지만 전달될 염색체의 적합도만 가지고 전달할 경우 최적해를 발견하지 못하고 지역해만 탐색할 가능성이 있게 된다.

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

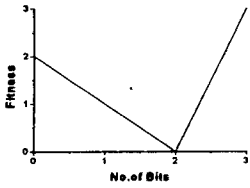


그림 1-1. 3bit deceptive function

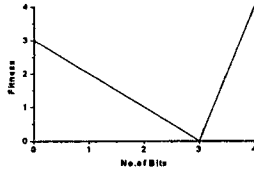


그림 1-2. 4bit deceptive function

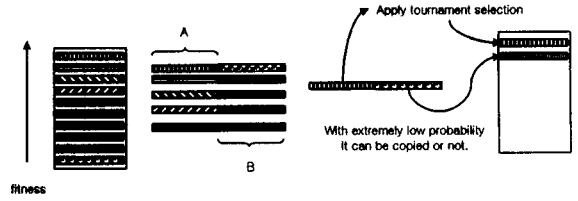


그림 2. 제안 알고리즘

경우 이전 세대에서 얻은 유전 정보가 파괴되는 것을 막기 위하여 1- 또는 2-점 교배가 주로 사용된다[3].

2.2 Deceptive Problem (DP)

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

3. 토너먼트 선택기법과 제안 선택기법

3.1 토너먼트 선택기법

토너먼트 선택기법은 개체군에 영색체가 n 개 있을 때, 2개 (혹은 그 이상)의 영색체를 무작위로 추출해 그 영색체들의 적합도를 비교해 가장 우수한 영색체를 다음 세대로 전달하는 방법이다. 하지만 이 방법은 적합도가 낮은 영색체는 전달되지 못한다는 단점을 갖고 있으며 이로 인해 다양성을 손실할 수 있다. 왜냐하면, 적합도가 낮은 영색체라 하더라도 그 영색체 내부에는 우수 구성요소(Building Block)가 포함될 수 있기 때문이다[1].

3.2 제안 선택기법

토너먼트 선택기법은 적합도가 낮은 영색체들은 다음 세대에 전달될 수 없다는 단점을 갖고 있다. 따라서 이 논문에서는 낮은 적합도를 갖고 있는 영색체들을 다음 세대로 전달할 수 있도록 하는 방법을 제안한다(그림 2, 그림 3).

n 개의 영색체가 개체군을 이룬다고 가정할 때, 그림 2처럼 초기 개체군을 적합도 순서대로 정렬한 후 첫 번째 영색체와 n 번째 영색체를 결합하여 하나의 쌍을 만들고, 두 번째 영색체와 $n-1$ 번째 영색체를 결합하여 또 다른 하나의 쌍을 만든다. 이러한 방식으로 모든 영색체에 대해 쌍을 만들면 1부터 $2/n$ 번째까지의 영색체는 그림 2의 A 부분에 속하게 되고 $2/(n+1)$ 번째 영색체부터 n 번째 영색체는 B 부분에 속하게 된다. 이때, $(1, n), (2, n-1), \dots, (2/n, 2/(n+1))$ 이러한 방식으로 순서쌍을 구성함으로써 n 번째 영색체는 $2/(n+1)$ 번째 영색체보다 다음 세대로 전달될 확률이 낮지만, A 부분에서 1번째 영색체는 A부분의 $2/n$ 번째 영색체보다 선택될 확률을 높기 때문에 A부분의 영향을 받아 B부분의 적합도가 낮은 영색체 또한 다음 세대로 전달될 여지가 남아있게 된다. 예를 들어, 3비트 DP(그림 1-1)에서 적합도가 결정되는 경우를 고려하자.

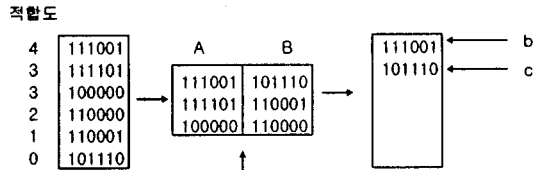


그림 3. 제안 알고리즘의 예

- 1) 적합도를 기준으로 정렬한 후, 2개의 부분으로 그룹 짓는다. (그림 3-a)
- 2) A 부분은 토너먼트 선택방법을 그대로 적용한다.(그림 3-b)
- 3) A 부분에서 선택된 유전자에 따라 B부분의 유전자는 다음 세대로 전달 여부를 확률로서 결정한다. (그림 3-c)

그림 3에서 그림 3의 A, B 부분에서 첫 번째(111001 101110)와 세 번째(100000 110000) 영색체 쌍이 무작위 방법으로 선택 되었을 경우, A 부분의 적합도만 고려하면 '111001'의 적합도가 '100000' 보다 높기 때문에 '111001'은 토너먼트 선택방법에 의하여 그대로 다음 세대에 전달된다. 그리고 '111001'과 그룹 지어진 '101110'은 적절한 확률에 따라 다음 세대로 전달될 것인지 결정하게 된다. 1) ~ 3)의 순서를 통해 얻은 두 개의 영색체(111001 과 101110)는 교배를 거치면 전체해인 '111111'을 찾을 수 있는 확률이 존재한다.

같은 문제에 대해 토너먼트 선택방법으로 최적해를 찾으려 할 경우 적합도가 낮은 '101110'은 적합도가 높은 영색체들이 갖지 못한 구성요소 (4, 5번째 비트의 위치의 유전자 값이 '1')를 갖고 있음에도 불구하고 적합도가 낮다는 이유로 다음 세대로 상속되지 못하고 도태될 가능성이 높다. 이는 다양성을 감소 시키거나 최적해를 찾기 위해 많은 수렴 세대수를 거쳐야 하는 단점을 초래하게 된다. 따라서 이 논문에서 제안한 새로운 선택방법은 토너먼트 선택 방법이 갖고 있는 장점은 그대로 수용하면서 기존의 선택방법과는 달리 선택된 영색체에 따라 그룹 지어진 우수 영색체들도 다음 단계로 넘겨줄 가능성을 남겨줌으로써 다양성도 유지하는 장점을 갖는다.

4. 모의 실험 결과 및 분석

이 장에서는 3비트 4비트 DP에 대한 제안된 알고리즘의 성능을 검증한다. 3, 4비트 DP의 경우 스키마의 길이가 각각 3, 4 이므로 이전세대에서 획득한 유전정보의 손실을 막기 위하여 2 -

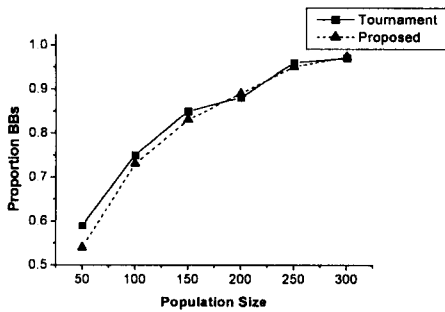


그림3-1. 3비트 디셉티브 문제에서의 수렴성

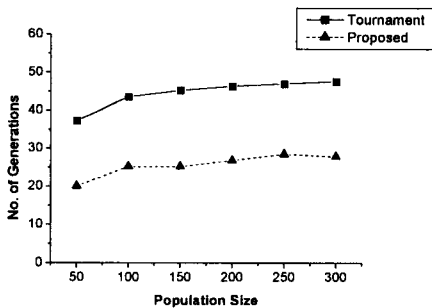


그림3-2. 3비트 디셉티브 문제에서의 수렴 세대수

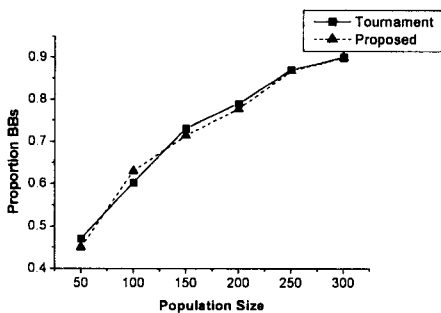


그림4-1. 4비트 디셉티브 문제에서의 수렴성

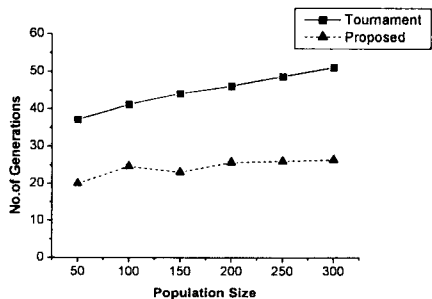


그림4-2. 4비트 디셉티브 문제에서의 수렴 세대수

정 교배를 사용하였다. 각각 선택 및 교배가 일어날 확률을 '1.0'으로 설정하였고, B부분의 영색체가 다음 세대로 전달될 확률은 '0.1'로 놓았다. 또한, BB 정보가 파괴되는 현상을 방지하기 위하여 돌연변이 연산자는 사용하지 않았다. 한편, 각 기법에 대해 공정한 비교를 위해 모든 개체군이 하나의 해에 수렴할 때까지 실험을 수행하였다[6]. 그림3-1, 3-2, 4-1, 4-2에 나타난 값들은 50번의 실험을 통한 평균값이다.

그림3-1과 그림4-1을 보면 제안 알고리즘은 기존의 선택 기법과 비슷한 최적도를 나타내지만 그림3-2와 4-2에서 보여진 것과 같이 제안 알고리즘은 기존 알고리즘에 비해 단지 50%의 세대수만 진화하더라도 충분히 수렴할 수 있음을 알 수 있었다. 이는 제안 알고리즘이 기존 알고리즘과 유사한 품질의 해를 발견하면서 동시에 50%의 수렴 성능의 향상을 의미한다.

5. 결론

본 논문에서는 기존의 토너먼트 선택 메커니즘의 문제점을 고찰하고 이를 보완하기 위한 방안을 제시했다. 토너먼트 선택 메커니즘은 우수 영색체를 다음 세대로 전달한다는 점에서 효과적이지만, 열성 영색체는 전달되기 힘들다는 단점을 갖고 있어서 해-표면을 충분히 탐색하지 못했다. 그러나 제안 선택 메커니즘에서는 열성 영색체에 대해 일정수준 다음 세대로 전달될 수 있도록 함으로써 다양성을 유지함과 동시에 토너먼트 선택기법의 우수 영색체 전달특성을 동시에 갖게 하였다. 따라서, 해의 최적도를 감소시키지 않으면서 50%의 수렴성능을 향상할 수 있었다.

본 논문은 국제화 캠퍼스 사업(계정번호 : GK00280) 외국인 교수 지원프로젝트에 의해 수행된 연구결과와 일부임

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 Edinburgh, 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)