

유전자 알고리즘에서 수렴속도 향상을 위한 새로운 토너먼트 선택 기법

이용채^o 손진곤

한국방송통신대학교 평생대학원 정보학과

lguana@bytegram.co.kr, jgshon@knou.ac.kr

A New Tournament Selection Technique

for Fast Convergence in Genetic Algorithms

Yong Chae Lee^o, Jin Gon Shon

Dept. of Computer Science, Graduate School, Korea National Open University

요약

유전자 알고리즘에서 좋은 염색체(chromosome)를 선택하는 방법은 알고리즘의 성능을 향상시키는데 매우 중요한 핵심 요소이다. 이러한 선택 기법 중에는 비례 선택 기법, 순위기반 선택 기법, 토너먼트 선택 기법 등이 잘 알려져 있다. 이 중 가장 성능이 좋은 토너먼트 선택 기법은 열성 염색체중 우성인 유전자를 포함하는 열성 염색체가 선택에서 배제되어 지역적 최적해(local minima)를 구할 가능성, 열성 염색체가 다음 세대 진화를 방해할 가능성 등의 문제점을 가지고 있다. 본 논문에서는 토너먼트 선택 기법의 문제점을 해결하기 위해서 토너먼트-교배 선택 기법을 제안하였다. 이 방법은 토너먼트 선택 기법을 기반으로 하되 열성 염색체가 선택되었을 경우 그 안에 들어 있는 우성 유전자를 알고리즘 진화에 반영시키고자 교배 단계를 추가한 기법이다. 제안된 토너먼트-교배 선택 기법을 이용하면 기존의 토너먼트 선택 기법보다 평균수행시간이 짧아져 해에 수렴하는 속도가 향상된다.

1. 서론

유전자 알고리즘은 적자생존과 유전의 메커니즘을 바탕으로 하는 일종의 탐색 알고리즘이다. 즉 해의 초기 추정치를 문자, 이진 스트링, 실수 등으로 표현해 유전자(염색체)를 표현하는 최소단위: gene) 형태로 코딩 작업한 초기 염색체 집단을 만들고 이 염색체 집단 내에서 선택, 교배, 돌연변이, 재생산 등의 과정을 반복 수행한다. 이렇게 반복 수행하는 과정 중에 적자생존에 따른 일정한 영역으로 염색체들이 접근하는 모습이 나타나 구하고자 하는 염색체 집단을 추정할 수 있다. 추정된 염색체를 디코딩 과정을 통해 구하고자 하는 실제의 해를 구할 수 있다. 이때 일정한 영역으로 접근하는 것을 수렴이라고 하고 어떤 경우에는 수렴을 하지 않는 문제도 있다. 본 논문에서는 수렴하는 문제만을 다룬다.

유전자 알고리즘의 성능에 매우 중요한 염색체 선택(selection)이란 집단 내에 존재하는 염색체 중 문제의 해에 가까운 염색체를 다음 세대에 남기기 위한 과정이다. 선택이 중요한 이유는 적자생존원리에 따라서 해에 가까지 않은 염색체는 버리고 해에 가까운 염색체의 수는 많이 늘림으로써 적당한 해 집단을 탐색하는 과정이기 때문이다. 즉 선택이 잘못되었을 경우 탐색하는 방법이 잘못되어 원하는 해 집단을 얻을 수가 없다.

선택 기법은 해에 적합한 염색체(우성)를 다음 세대에 많이 남기고 해에 적합하지 않은 염색체(열성)를 적게 남겨서 해에 접근하는 방법을 말하는데 많이 알려진 비례 선택 기법, 순위기반 선택 기법, 토너먼트 선택 기법 등이 있다. 본 논문의 제2절에서 자세히 논하기로 한다.

본 논문에서 개선하고자 하는 토너먼트 선택기법은 다른 기법보다 비교적 간단해 많은 수의 염색체를 가지는 경우, 즉 집단의 크기(인구수)가 큰 경우 주로 사용한다. 토너먼트 기법은 운동경기에서 우승자를 고를 때 무작위로 경기를 치른 후 최종 우승자를 가리는 것과 유사한 절차를 거친다. 즉 임의로 n개의 유전자를 선택한 후 우성인 유전자를 남기고 원하는 인구수가 될 때까지 반복 수행해서 최종적으로 다음 세대의 염색체 집단을 형성한다.

이러한 토너먼트 기법은 단순한 반면 다음과 같은 단점이 있다.

첫째, 우성 염색체를 다음 세대로 넘기는 확률이 높아서 열성 중 우성 유전자를 가진 경우는 선택에서 배제될 확률이 많아 지역적 최적해를 구하게 될 수 있다는 문제점이 있다. 이 문제를 Deceptive Problem이라고 부르는데 주로 유전자의 다양성을 높이는 방법을 찾아서 해결하고 있다[1][2][11].

둘째, 확률에 의해서 열성이 선택되었을 경우, 다음 세대에서도 열성 염색체와 다른 염색체를 교배해서 생성된 새로운 염색체가 수렴 속도 향상에 항상 기여한다고 보장할 수 없다는 문제점이 있다.

본 논문에서는, 기술된 문제점 중 두 번째 문제점을 해결하기 위해 새로운 토너먼트-교배 선택 기법을 제안한다.

2. 선택 기법

2.1 선택 기법의 분류

적자생존의 원리에 따른 선택을 하기 위해서는 기준이 되는 수치 값이 있어야 한다. 이것은 선택된 염색체가 주어진 문제에 적합한 정도를 판별하기 위한 적합도 함수(fitness function)에 따라서 산출된 값을 말한다[9]. 이러한 적합도 함수를 비롯하여 여러 가지 사용되는 기술에 따라 선택 기법들은 다음과 같이 분류될 수 있다.

2.1.1 비례 선택 기법(proportionate selection)

전체 적합도의 합에 대한 각 염색체의 적합도의 비를 염색체 선택의 확률로 이용하는 방법으로서 유전자 알고리즘에서 가장 기본적인 선택 기법이다. 하지만 집단내의 모든 염색체들의 적합도를 구한 후 각각 확률에 따라 선택되기 때문에 인구수가 많은 집단에는 적용하기 어렵다. 특히 각각의 선택 확률이 연속적이지 않고 계단 형식으로 나타나서 인구수가 너무 적을 경우 다양성을 해치는 결과를 가져와 사용하기 어렵다[7].

2.1.2 토너먼트 선택 기법(tournament selection)

일정 수의 염색체를 임의로 선택하고 그 중 최고의 적합도를 갖는 염색체를 남기고 원하는 인구수까지 반복 선택한다. 인구

수가 많을 경우 주로 사용하는 방법으로서 알고리즘의 구조는 간단하지만 서론에서 언급했던 문제점이 있다. 이에 관해서는 제2.2절에서 보다 자세히 설명한다.

2.1.3 순위기반 선택 기법(rank selection)

적합도의 크기 순서에 따라 염색체의 순위를 정하고 그에 따라 염색체를 남길 확률을 결정하는 선택 방법이다. 비례 선택 기법에서 나타나는 염색체의 선택 확률이 연속적이지 못해 발생하는 단점을 보완하기 위해서, 적합도에 따른 순위를 매긴 유전자 집단에 직선형 선택 확률을 적용한다.[8].

2.1.4 엘리트 보존 선택 기법(elitist preserving selection)

특히 좋은 해가 선택에서 배제되는 것을 방지하기 위해 적합도가 가장 높은 염색체는 우선적으로 남기는 방법으로서 일반적으로 다른 선택 기법과 함께 사용한다[8].

2.2 토너먼트 선택 기법

이 기법은 일정 수의 염색체를 단순히 추출해 적합도를 계산한 것 중에서 가장 좋은 것을 선택하기 때문에 다른 선택 기법들에 비해 빠르게 다음 세대에 남길 염색체를 결정할 수 있다.

토너먼트 선택 기법은 우선 서론에서 설명한 첫 번째 문제점을 가지고 있다. 즉 현재 집단 내에 있는 염색체들이 우성으로 모두 통일되어 집단 내에 존재하지 않은 더 좋은 해로 수렴하는 것을 방해한다. 물론, 몇 번을 반복 후 돌연변이 연산으로 새로운 유전자를 얻어 알고리즘이 진행은 되지만 성능을 저하시키는 결과를 가져온다.

이러한 단점을 해결하기 위해서 선택압(selection pressure)을 이용하여 적당히 열성을 남기는 토너먼트 선택 기법을 아래 그림 1을 이용하여 설명한다[12][7].

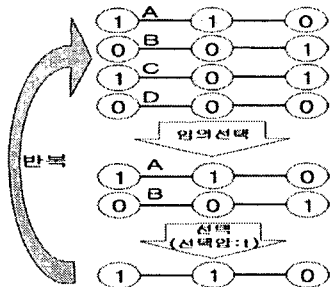


그림 1. 선택압을 이용한 토너먼트 선택 기법

그림 1에서 염색체 집단의 크기는 4이고 염색체 각각은 검은 원으로 표시되는 유전자를 3개씩 포함하고 있다(예를 들어 염색체 A는 '1', '1', '0'의 유전자로 구성되어 있음). 염색체 A, B, C, D 중 무작위로 A, D가 선택되었을 때 적합도에 따라 우성과 열성으로 나누어지게 된다. 그런 다음 0~1사이의 난수를 발생해 미리 정해 놓은 선택압과 비교해서 적은 값이면 우성, 큰 값이면 열성을 최종적으로 선택한다. 이 과정을 집단의 크기와 동일하게 될 때까지 반복한다. 이러한 선택 기법은 서론에서 설명한 첫 번째의 문제점인 다양성의 결여를 적당한 분량의 열성 염색체를 다음 세대로 남기도록 함으로써 해결한다.

그러나 이때 남겨지는 열성 염색체가 항상 해 집단으로의 수렴 속도를 향상시킨다고 보장할 수는 없다는 문제점이 있다.

3. 개선된 토너먼트 선택 기법

3.1 수렴속도 및 제어 파라미터

3.1.1 수렴속도

주어진 문제가 수렴하는 문제일 때 알고리즘의 수행에 따라

서 해 집단이 특정 해 영역내로 접근하는 것을 볼 수 있다. 이러한 수렴의 속도는 좋은 해를 구하는 척도 중의 하나이다. 너무 빠른 수렴은 해 집단의 다양성을 해쳐서 지역적 최적 해를 얻거나 해의 정밀도를 낮춘다. 따라서 유전자 알고리즘은 이러한 수렴 속도와 관련있는 제어 파라미터가 존재한다. 이러한 파라미터의 조정으로 더욱 더 정밀한 해를 구할 수 있다.

3.1.2 제어 파라미터

다음 절에서 토너먼트-교배 선택 기법 구현할 때 사용하는 제어 파라미터에 대해서 설명한다.

① **염색체길이(l)**는 여러 개의 유전자로 표현된 해가 있을 때 우리는 염색체라고 하며 이때 표현된 유전자연결 고리의 크기를 말한다. 보통 문자열의 길이로 측정된다. 길이를 길게 하면 성능의 저하를 가져오기 때문에 적당한 길이를 취해야 한다. 만약 2진 스트링 유전자를 사용한다면 $l \geq \log_2[(\text{탐색영역최대} - \text{탐색영역최소})10^d + 1]$ 으로 정의 할 수 있다. 단, d는 소수점 이하 자리수로서 정밀도를 나타낸다[6].

② **집단의크기(p)**는 염색체 선택이 수행될 집단내의 염색체 수를 말한다. 이것은 성능과 해의 정밀도에 영향을 준다. 즉 집단의 크기가 너무 작으면 해를 빠르게 구할 수는 있지만 해의 정밀도가 낮고 집단의 크기가 크면 연산 속도는 느려지나 해의 정밀도가 높아진다. 적당한 집단의 크기는 30~200이다[3].

③ **선택압(f)**은 우성과 열성을 남기는 비율을 조정하기 위한 확률을 정의한 상수이다. 높이면 집단의 다양성의 결여로 빠른 수렴은 하지만 지역적 최적해에 도달할 가능성이 높다. 낮으면 집단의 다양성이 높아져서 수렴의 속도는 늦어지나 최적해에 도달할 가능성이 높다. 적당한 선택압은 0.6이상을 추천하고 있다[5].

④ **돌연변이율(m)**은 염색체에 새로운 유전형질로 변화될 확률을 정의하는 상수이다. 너무 크면 무작위로 검색하는 결과를 초래해 성능을 저하 시키고 너무 작으면 집단에 새로운 유전자를 얻을 기회를 잃어서 다양성을 유지하는데 실패한다. 적당한 돌연변이율은 0.5~1사이로 추천하고 있다[4].

⑤ **교배율(c)**은 집단의 염색체가 교배에 참여할 확률을 정의하는 상수이다. 크게 하면 탐색공간을 빠르게 검색하는 결과를 초래한다. 작게 하면 새로운 형태의 유전자의 출현을 방해해 탐색이 침체되는 경우가 발생된다[5].

3.2 토너먼트-교배 선택 기법

제2.2절에서 언급한 토너먼트 선택 기법은 서론의 첫 번째 문제점을 해결하였으나 두 번째 문제점을 해결하기 위해서는 열성 염색체 내에 우성인 유전자가 포함되어 있는가를 검증할 절차가 필요하다. 이런 검증 절차를 정리하고 그림 2로 상세히 도식화했다.

현재 토너먼트 선택 기법에 따라 두 개의 염색체 A와 B를 임의의 추출하여 적합도를 검사하였더니 A가 우성, B가 열성이라고 하고, 선택압에 따라 열성 염색체의 B를 남겨야 하는 경우라고 가정하자. 이때 검증 절차는 다음과 같다.

- ① 현재 선택된 두 염색체(A, B)를 교배한다.
- ② 교배하여 얻은 염색체(A')와 열성 염색체(B)를 비교한다.
- ③ ②단계에서 우성인 염색체(A' 혹은 B)를 선택한다.

이러한 절차를 거치게 되면, 열성 염색체내에 우성 염색체보다 더 좋은 유전자가 존재하였을 경우, 이 유전자는 교배로 인해 새로 생긴 염색체에 반영된다.

그림 2를 보면 우성 염색체가 선택되는 경우 우성 영역이 수행되고, 열성 염색체가 선택되는 경우 열성 영역이 수행된다. 우성 영역에서는 기존 토너먼트 기법에서 사용한 방식을 그대로 사용한다. 열성 염색체가 선택되는 경우 실행되는 열성 영

역에서는, 열성 염색체 B, 그리고 A와 B를 교배시켜 얻은 A에 대해 적합도 함수 값을 서로 비교하여 우성의 염색체를 남긴다. 이러한 과정을 원하는 인구수에 도달할 때까지 반복한다.

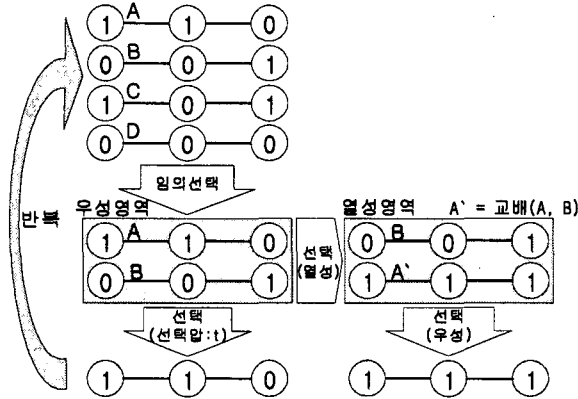


그림 2. 토너먼트-교배 선택 기법

4. 구현 및 성능 평가

토너먼트-교배 선택 기법의 성능을 평가하기 위해서 우선 Java 프로그래밍 언어를 사용해 토너먼트 선택 기법을 프로그램으로 구현하고, 그 프로그램을 기반으로 본 논문에서 제시한 토너먼트-교배 선택 기법도 함수부분만 다르게 만들어서 구현하였다. 그런 다음, 두 개의 프로그램을 동일 환경에서 수행시켜 해에 수렴하는 시간을 측정했다. 이때, 초기 해 집단을 동일하게 하고, 적합도 함수는 스트링 비교함수를 사용하였으며, 정지 조건은 모든 염색체내의 유전자 '1'로 바뀔 때로 정하였다. 프로그램의 중요 부분은 다음 그림 3과 같다.

```

population = initPopulation(population);
do {
    // 토너먼트선택기법
    //population = tournament(population, rand);
    // 제안 교배선택기법
    population = crossoverTournament(population, rand);
    for( int i = 0 ; i < MaxPopulation ; i++ ) {
        choosen = choosenTwo(population, rand);
        offspring[i] = crossover(choosen[0],choosen[1]);
        offspring[i] = mutation(offspring[i], rand);
    }
    moveToPop(population, offspring );
} while(isStop(population));
    
```

그림 3. 성능 평가를 위한 프로그램

성능 평가를 위한 초기 집단은 2개의 초기 염색체 A형 ('101010')과 B형('100010')을 7:3의 비율(A형 염색체 35개, B형 염색체 15개)로 복제해서 만들었다. 또한 초기 파라미터로 집단의 크기(p)를 50, 선택압(t)을 0.7, 돌연변이율(m)을 0.005, 교배율(c)을 100%, 염색체 길이(l)는 6으로 하였으며, 교배는 무작위 일점교배 방식으로 하였다. 그런 다음, 각각의 선택 기법을 10번씩 수행한 것을 1회로 잡고 총 5회를 반복 수행했다.

그림 4는 X축을 수행회수, Y축을 수행평균시간(ms)으로 하여 두 선택 기법의 프로그램을 수행시켰을 때 해에 수렴할 때까지 걸린 시간의 평균치를 그래프로 그린 것이다. 이 성능 평가에 따르면, 토너먼트 선택 기법의 평균 수행시간은 15.42(ms)이고,

토너먼트-교배 선택 기법의 평균 수행시간은 8.8(ms)이다. 이 결과를 가지고 효율을 구하면 약 43%의 성능이 높아진 것을 알 수 있다. 그리고 집단의 크기(p)를 100, 200, 500으로 각각 증가시키면 효율이 약 44%, 40%, 35%씩 높아졌다.

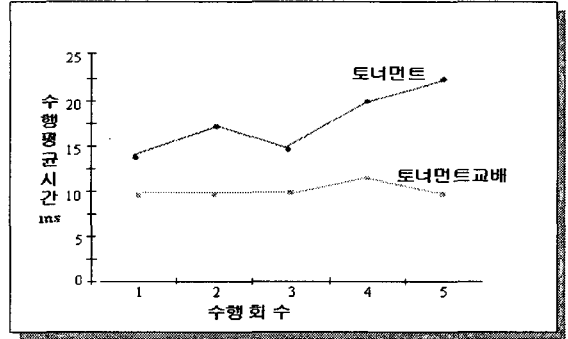


그림 4. 평균 수행시간

5. 결론

본 논문에서는 열성 염색체가 다음 세대 진화를 방해할 가능성의 문제점을 해결하기 위하여 새로운 토너먼트-교배 선택 기법을 제안했으며, 이는 유전자 알고리즘에서 해 집단으로의 수렴 속도를 향상시키는 결과를 얻었다. 성능 평가에 따르면 제안된 토너먼트-교배 선택 기법이 기존의 토너먼트 선택 기법에 비해 최저 35%의 수렴 속도가 향상되었다.

참고문헌

- [1] X. Hue, "Genetic Algorithms for Optimization: Background and Applications," Edinburgh Parallel Computing Centre, The University of Edinburgh, Ver 1.0, February 1997.
- [2] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley Pub. Co. Inc., N.Y., 1989.
- [3] D. Goldberg, "Optimal Initial Population Size for Binary-Coded Genetic Algorithms," TCGA Report No.85001, University of Alabama, 1985.
- [4] K. DeJong, "An Analysis of the Behavior of a Class of Genetic Adaptive System," Doctoral Dissertation, The University of Michigan, Ann Arbor, Michigan, 1975.
- [5] J. Schaffer, R. Caruana, L. Eshelman and R. Das, "A Study of Control Parameters Affecting Online Performance of Genetic Algorithms and Their Function Optimization," Proc. 3th Int. Conf. on Genetic Algorithms and Their Applications, Arlington, VA, pp.51-60, 1989.
- [6] 진강규, *유전자알고리즘과 그 응용*, pp.377, 2004.2
- [7] 문병로, *유전알고리즘*, 두양사, pp.247, 2003.3
- [8] 공성근의 4인, *유전자알고리즘입문*, 진영사, pp.301, 2004.3
- [9] 이수정, 이병일, "효과적인 군집화를 위한 유전자알고리즘의 적합도 함수에 관한 연구", 한국정보과학회 2001년 CVPR 추계워크샵, vol.##, no.##, pp.0191-0192, 2001.11.
- [10] 김예훈, 안창욱, 김기표, R. Ramakrishna, "유전자 알고리즘에서 Deceptive 문제에 대한 토너먼트 선택의 최적화", 한국정보처리학회 2003년 추계학술대회 발표논문집, vol.10, no.02, pp.0305-0308, 2003.11.
- [11] 김기표, 안창욱, R. Ramakrishna, "유전자 알고리즘의 다양성과 수렴성을 고려한 새로운 선택기법", 한국정보과학회 2003년 춘계학술대회 발표논문집, vol.30, no.02, pp.0353-0355, 2003.04.
- [12] S. Mahfound, "Niching method," *Evolutionary Computation 2: Advanced Algorithms and Operators*, Institute of Physics Publishing, pp.87-92, 2000.