

GA의 교차점 위치의 영향 분석

An Analysis of Crossover Point's Effect in GA

고명숙*
Myung-Sook Ko

요 약

유전자 알고리즘은 탐색 공간에서 새로운 샘플을 찾기 위하여 선택과 재조합의 연산을 사용하는 모 집단 기반 모델이다. 교차 연산은 최적화 알고리즘인 유전자 알고리즘(GA)에서 가장 중요한 부분을 차지한다. 이 논문에서는 교차 연산 수행 시 필요한 교차점 위치 조절이 수렴에 어떠한 영향을 주는지 분석해 보고자 한다. 교차점 위치를 세대 진행수와 수렴 정도에 따라 특정 위치로 조절한 방법이 일반적인 확률로 교차점 위치를 정하여 최적화를 수행한 경우보다 전역 해를 더 빨리 찾아감을 볼 수 있었다.

1. 서 론

유전자 알고리즘(GA: Genetic Algorithm)은 실제 여러 분야에서 응용할 수 있는 견고한 탐색 및 최적화 기법이다. 유전자 알고리즘은 임의적이지만 방향성 있는 탐색을 함으로써 전체적인 최적 해를 찾아 나가는 방법이다. 전형적으로 유전자 알고리즘은 다음과 같은 요소들로 구성되어 있다[1,3,4].

- 최적화 문제에 대한 가능한 해를 나타내기 위한 유전적 표현(또는 인코딩)
- 인코딩 된 해들의 모집단
- 각 해의 최적성을 평가하는 적합도 함수
- 존재하는 모집단으로부터 새로운 모집단을 만들어가는 유전 연산자
- 제어 파라미터

유전자 알고리즘은 개체(individual)가 동시에 병렬적으로 주어진 환경에 따라 생태계의 자연 진화의 원리에 의해 진화해 가면서 최적 해를 찾아가는 알고리즘으로서 홀랜드에 의해 최초로 도입되었다[4]. 모집단은 최적화 문제에서 후보 해를 나타내는 개체들의 모임이며, 이 개체들은 교차 연산과 돌연 변이와 같은 유전 연산(genetic operation)을 통해 진화한다. 모집단이 진화할 때 각 세대가 얻은 환경에 대한 정보는 개체에 저장되어 다음 세대로 전달된다. 각 세대에서 한 개체가 선택되어 다음 세대로 자신의 유전 형질을 물려줄 수 있는 확률은 적자 생존의 원리를 기반으로 하며, 그 개

체가 주위 환경과 나머지 개체와의 상호 작용에 의해 변하는 적합도 값(fitness value)에 따라 변한다. 유전자 알고리즘의 기본 구조는 다음 그림 1과 같다[2,3].

```

Genetic Algorithm()
{
  initialize population;
  evaluate population;
  while convergence not achieved
  {
    scale population fitness;
    select solutions for next population;
    perform crossover and mutation;
    evaluate population;
  }
}

```

그림 1. 유전자 알고리즘의 기본 구조
Fig 1. Basic Structure of GA

이 논문에서는 교차 연산 수행 시 필요한 교차점 위치 조절이 수렴에 어떠한 영향을 주는지 분석함으로써 유전자 알고리즘에서의 교차 연산이 최적화에 어떤 영향을 미치는지를 알아보고자 한다. 2장에서는 유전자 알고리즘의 교차 연산에서 최적 해를 더 빨리 탐색할 수 있는 교차점 위치 설정 방법과 탐색 효과에 대하여 기술한다. 3장에서는 테스트 함수에 대하여 교차점 위치를 조절했을 때의 교차 연산의 성능을 비교 평가해 본다. 마지막으로 4장에서는 결론을 맺고 앞으로의 연구

* 부천대학

방향을 제시한다.

2. 교차 연산에서의 교차점 결정

일반적으로 교차 연산은 임의의 점을 선택하여 그 위치를 교차점으로 정한 후 그 위치를 기준으로 두 개체의 유전 형질을 바꿈으로써 완성된다 [4]. 본 논문은 임의의 위치를 교차점으로 정했을 경우 수렴 정도나 최적 해에 근사했을 경우에도 여전히 확률적인 방법으로 교차 연산이 수행되기 때문에 그 결과가 교차 연산이 수행되었음에도 불구하고 전혀 변화가 없는 경우가 생길 수 있다는 점을 지적하고자 하는 것이다. 즉, 최적 해에 수렴해 갈수록 모집단에 속해있는 개체들의 최 상위 비트 부분은 거의 같아지므로 최 상위 비트에 가까운 위치에서의 교차점 선택은 의미가 없어지게 된다.

함수 최적화의 경우 유전자 알고리즘이 임의의 값으로 수렴해 가고 있을 때 개체를 표현하는 비트 위치 중 최 상위 비트와 최하위 비트간의 이진 값들은 그 위치가 중요한 의미를 가질 수 있다. 왜냐하면 최 상위 비트에 가까운 이진 값들은 모집단이 수렴해 갈 때 거의 변화를 보이지 않으며 그 값들이 고정되어 있다. 또한, 전체 모집단 내에서 각 개체를 비교해 보았을 때 비교적 높은 적합도 함수 값을 갖는 개체들의 최하위 비트에 가까운 이진 값들이 서로 다를 수 있다. 본 논문에서는 세대가 진행되어감에 따라서 최 하위 비트에 가까운 부분들로 교차점 위치를 제한함으로써 유전자 알고리즘의 수행 횟수를 줄이고, 무의미한 교차 연산을 줄임으로써 최적 해를 더 빨리 찾을 수 있는 방법을 제안하고자 한다.

2.1 Fmax - Favg 값의 변화에 따른 교차점 위치 설정

함수 최적화 문제에서는 개체의 특정 국소 부분의 변화가 전체 성능(최적 점을 찾는)에 영향을 미칠 수 있으므로 먼저 탐색이 국부 최적 점에서 빠져나와 전역 해를 찾을 수 있도록 유도하기 위하여 다음과 같은 방법을 제안한다.

탐색 공간에 대해 최적 해를 찾는 과정에서 알고리즘의 현재 상태가 최적 해에 수렴하고 있는지를 알 수 있어야 하는데, 그 수렴 정도를 측정하는 방법 중의 하나가 모집단 내의 평균 적합도 함수 값(Fmax)과 최대 적합도 함수 값(Favg) 간의 관

계를 고려하는 것이다. Fmax - Favg의 값은 모집단내의 개체들이 전체 탐색 공간에 산재되어 있을 때보다 최적해로 수렴해 갈수록 더 작아진다. 다음 그림 2는 세대가 진행됨에 따라 Fmax - Favg값의 전형적인 변화를 보여준다. 그림 2에서 Fmax - Favg의 값은 적합도 함수 값이 0.25인 국부 최소 점으로 수렴할 때 감소하고 있음을 볼 수 있다. 따라서 적합도 함수의 평균값과 최대 값간의 차를 GA의 수렴 정도를 측정하는 기준으로 사용할 수 있다.

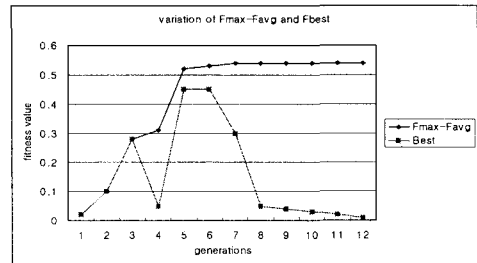


그림 2. Fmax-Favg, Fbest 값의 변화
Fig 2. The variation of Fmax-Favg & Fbest

전역해가 존재할 가능성이 높은 구간에 대하여 수렴 속도를 높이기 위하여 모집단의 수렴 정도와 세대 반복 횟수 및 개체의 목적 함수 값의 향상 정도를 기준으로 교차점의 위치를 정한다. 즉, 세대가 진행되어 갈수록 교차점의 위치가 왼쪽에서 오른쪽으로 이동하게 되는 것이다. 즉, 개체의 비유사 부분(non-matching alleles)에서 교차 연산이 적용되도록 교차점의 위치를 조정하는 것이다. 그림 2에서 세대 반복 횟수가 0~2.5구간을 A구간, 2.5~5.5 구간을 B구간, 그 이하를 C구간이라 나누었을 때 최 상위 비트에서 출발한 교차점 위치의 범위는 다음 식 1의 변화량 y_i 값에 의해 증가된다. y_i 값은 적합도 함수 값의 변화량에 의해 양 또는 음의 값을 얻게 되는데, 이 값이 양의 값일 경우 교차점이 적용될 범위를 감소시키며, 초기 세대에서는 증가량을 크게 하며 나중 세대에서는 증가량이 작아지도록 조절하는 변수 값이다. 반대로, 적합도 함수 값의 변화가 음일 경우는 함수의 최적 해에 도달하였거나 또는 국부 최적 점에 빠진 경우에 해당하므로 교차점이 적용될 범위를 증가시킨다(식1 참고).

$$y_i = \begin{cases} \sqrt{(\Delta f)} & \text{if } (\Delta(f_{\max} - f_{\text{avg}})) > 0 \\ -\sqrt{(\Delta f)} / f(\Delta f) & \text{if } \Delta(f_{\max} - f_{\text{avg}}) \leq \Delta(f_{\max} - f_{\text{avg}})^{-1} < 0 \\ e^{\Delta f} & \text{if } \Delta(f_{\max} - f_{\text{avg}})^2 \times \Delta(f_{\max} - f_{\text{avg}})^{-1} > 0 \end{cases} \quad \text{식 1.}$$

2.2 탐색 효과

새로운 교차점 범위 설정 방법을 이용하여 새로 생성된 개체의 적합도 함수 값의 향상이 없을 경우는 이전 세대에서 결정되었던 교차점 범위가 MSB쪽으로 이동하도록 하여 교차 연산이 갖는 특성이 원거리 이동(long jump)의 가능성이 높아 지도록 하였다. 전체 개체에 대하여 교차점 위치를 설정하는 것은 임의 탐색의 장점을 활용할 수는 있으나, 수렴이 어느 정도 진행되어진 다음에는 성능 향상을 보장할 수 없다는 단점을 지닌다. 또한 모집단 내의 개체 다양성의 감소는 최적해로의 수렴율을 낮출 뿐만 아니라 상태 공간 탐색에 대한 초기 수렴 현상을 보인다. 교차점 위치를 고려하지 않은 전체 개체에 대한 교차 연산의 적용은 교차 연산의 적용 후 새롭게 생성된 자식 개체가 그 부모 개체와 비슷해질 확률을 높이므로 국부 최적 점에 빠졌을 경우에는 더 이상 성능 향상을 보장할 수 없다. 따라서 상태 공간 탐색에 대한 전체 탐색을 가능하게 해 줌으로써 국부 최적 점에서 벗어날 수 있도록 하여야 한다. 이를 위해서 생성된 새로운 개체의 적합도 함수 값의 향상이 없을 경우, 개체의 특정 부분(두 부모 개체에 대해 비 유사 부분이 존재하는 구간이 선택되도록 함)에 대해서만 교차 연산이 일어날 수 있도록 교차점 위치를 조절하는 것이 매우 중요하다.

교차 연산에서의 교차점 위치가 세대가 진행되어 감에 따라 조절되게 함으로써, 개체의 최하위 비트 부분의 변화에 가중치를 부여함으로써 교차 연산이 행하는 원거리 이동 효과를 볼 수 있다.

3. 실험 및 평가

함수 최적화 문제에서 교차 연산의 교차점 위치 조절을 이용한 개체의 특정 국소 부분의 변화가 전체 성능에 영향을 미칠 수 있음을 평가하기 위하여 다음 표1의 테스트 베드 함수를 이용하여 실험하였다[1].

표 1. 테스트 함수

Table 1. Test functions

함수	함수	변수 범위	참고
f1	$f_1(x_1, x_2) = 21.5 + x_1 \sin(4\pi x_1) + x_2(20\pi x_2)$	$-3.0 \leq x_1 \leq 12$ $4.1 \leq x_2 \leq 5.$	[6]
f2	$f_2(x) = x(\sin 10\pi x) + 1.0$	$-1.0 \leq x \leq 2.0$	[6]
f3	$f_3(x_i) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2$	$-2.048 \leq x_i \leq 2.$	[7]

다음 그림 3과 그림 4는 함수 f1에 대하여 탐색 방향을 나타내는 Fmax - Favg값의 세대 진행에 따른 변화와 세대 진행에 따른 교차 연산의 교차점의 위치를 보여준다. 그림 3에서 볼 수 있듯이 세대가 진행해 감에 따라서 Fmax-Favg의 값은 증가에서 감소 곡선을 그리고 있다. 특히 교차점 위치를 고려하지 않은 경우는 여전히 높은 Fmax-Favg의 값을 나타냄을 볼 수 있는데, 이러한 현상은 탐색 과정에서 모집단 내의 개체들의 적합도 값이 거의 유사해지는 현상(유전자 표류)으로 인해 세대가 진행 되어도 더 이상 성능 향상에 영향을 끼치지 못함을 보여주는 것이다.

그림 4에서는 세대 진행에 따른 교차점 범위 변화를 보여주고 있다. 교차점 위치가 초기 최 상위 비트 쪽에서 LSB쪽으로 진행되어 감을 볼 수 있는데, 이는 두 개체 간의 비 유사 부분이 교차 연산에 적용될 수 있도록 하는 효과를 나타낸다.

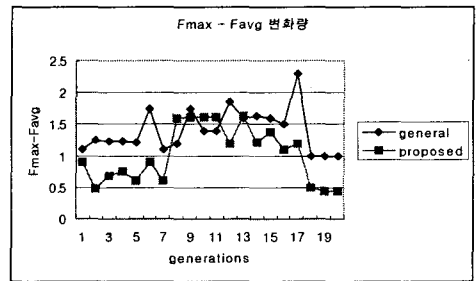


그림 3. Fmx-Favg의 변화(f1)

Fig 3. Variation of Fmax-Favg(f1)

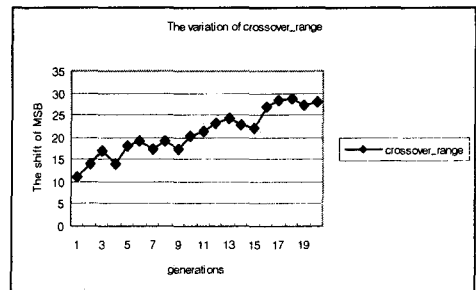


그림 4. 세대 진행에 따른 교차점 범위 변화(f1)

fig 4. Variation of Crossover Range in growing generation(f1)

또한 표 2에서 볼 수 있는 것처럼 교차점 위치를 조절한 경우가 그렇지 않은 알고리즘 보다 더

적은 세대 반복회수로 최적 해를 더 잘 찾을 수 있음을 알 수 있다.

표 2. 제안 방법과 기존방법의 수렴 횟수 비교
Table2. Convergence comparison between general and proposed method

Fn.	General	Proposed	Optimum(G/P)
f1	20	12	X/O
f2	15	10	X/O
f3	20	9	O/O

4. 결론

이 논문에서는 유전자 알고리즘에서 기본적인 교차 연산 방법과 교차점 위치를 조절한 교차 연산을 실제 최적화 문제에 적용시켜 봄으로써 교차 연산이 탐색에 어떤 영향을 미치는지를 비교 분석하여 보았다.

각 세대마다 2.2 절에서 기술한 바와같이 유전자 알고리즘이 최적해로 수렴해가고 있는 정도를 측정하기 위한 기준으로 $F_{max} - F_{avg}$ 값을 이용하였으며 이 값의 변화량에 따라 각 구간에 대한 교차점 위치를 조절하였다. 이 논문에서 제안하는 교차 연산의 교차점 조절 방법은 세대가 진행되어 감에 따라 식 1에서 제시한 변수 값의 변화량에 따라 교차점 위치를 조절함으로써 함수 최적화에서 최적 값을 생성해낼 가능성이 높은 변수 값의 의미 있는 부분(desired portion)만을 향상시킬 수 있도록 한 방법이다. 앞으로의 연구 방향은 교차 연산과 돌연 변이 확률의 최적 값을 교차 연산의 교차점 조절 방법과 결합시켜 다양한 최적화 문제에 적용시키는 방법에 대한 연구이다.

참 고 문 헌

- [1] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag Pub., 1992.
- [2] M. F. Bramlette, "Initialization, Mutation and Selection Methods in Genetic Algorithms for Function Optimization," Proc. 4th Int. Conf. on Genetic Algorithms, Morgan Kaufmann Pub., pp. 128-135, 1991
- [3] M. Srinivas and L. M. Patnaik, "Genetic Algorithms: A Survey," IEEE Trans. Computers, June, pp. 17-26, 1994
- [4] M. Mitchell, An Introduction to Genetic Algorithms, MIT Press, 1996.
- [5] M. Lipsitch, "Adaptation on Rugged Landscapes generated by Iterated Local Interactions of Neighboring Genes," Proc. 4th Int. Conf. on Genetic Algorithms, Morgan Kaufmann Pub., pp. 128-135, 1991
- [6] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag Pub., 1992.
- [7] J. M. Yang, C. Y. Kao, "A Combined Evolutionary Algorithm for Real Parameters Optimization", IEEE Int. Conf. on Evolutionary Computation, Vol. 2, pp.732-737, 1995