

최적해 탐색 문제의 효율적인 해결을 위한 유전 알고리즘의 집중화 집단과 다각화 집단의 분리

박태진^a, 류광렬^b

^a 부산대학교 컴퓨터공학과
부산 금정구 장전동 산 30번지

Tel: +82-51-510-3531, Fax: +82-51-510-2431, E-mail: parktj@pusan.ac.kr

^b 부산대학교 컴퓨터공학과
부산 금정구 장전동 산 30번지

Tel: +82-51-510-2453, Fax: +82-51-510-2431, E-mail: krroyu@pusan.ac.kr

초 록

본 논문에서 우리는 최적화 문제를 효율적으로 해결하기 위한 새로운 유전 알고리즘을 제안하였다. 새로운 유전 알고리즘의 가장 큰 특징은 서로 다른 목적을 가진 2개의 집단을 가지고 진화를 수행한다는 점이다. 하나의 집단은 일반적인 유전 알고리즘의 그것과 비슷하지만, 좋은 영역의 탐색에 집중하는 경향을 보이며 다른 하나의 집단은 이러한 좋은 영역에서의 집중 시 나타날 수 있는 조기 수렴 문제를 보완하기 위해 탐색을 다각화 하는 방향으로 진화한다. 몇 가지 최적화 문제에 대한 실험 결과 이러한 특성이 다른 2개 집단에 의한 진화 방법이 문제 해결에 도움이 됨을 확인할 수 있었다.

Keywords:

유전 알고리즘, 최적화, 탐색 다각화

1. 서 론

탐색 알고리즘이 제한된 시간 내에 좋은 결과를 찾기 위해서는 정보의 유효한 영역에 대한 이용(exploitation)과 미지의 영역에 대한 탐험(exploration)의 균형이 잘 유지되어야 한다. 유효한 영역에 대한 이용에 너무 치우친 언덕 오르기 탐색(hill climbing) 탐색 같은 경우, 지역 최적해는 빠른 시간 내에 찾아낼 수 있지만, 전체 탐색 공간을 고루 살펴볼 수 없어서 실제 좋은 결과를 얻기는 힘들다. 반대로 미지의 영역에 대한 탐험에 너무 치우치는 경우에는 여러 탐색 공간을 고루 살펴볼 수는 있겠지만 그 찾아낸 영역들

중에서 실제 좋은 해를 찾아내지는 못할 것이다. 이에 담금질 탐색(simulated annealing), 타부 탐색(Tabu search)과 같은 이웃해 탐색 알고리즘은 언덕 오르기 탐색을 기반으로 탐험을 좀 더 잘하기 위한 방법들을 추가하였다.

유전 알고리즘의 경우 집단을 기반으로 한 탐색 방법으로서 이웃해 탐색 방법들에 비해 탐험 능력은 뛰어나지만 이용능력은 부족한 것으로 알려져 있다. 이에 기존에는 유전 알고리즘의 탐험 능력과 이웃해 탐색 알고리즘의 이용 능력을 결합하여 보다 좋은 결과를 얻기 위한 연구가 많이 진행되었다. 또한 유전 알고리즘의 선발압(selection pressure)을 적절하게 조절하여 이용과 탐험의 균형을 유지하려는 연구도 많이 진행되었다. 허나 본 논문에서는 기존의 유전 알고리즘이 하나의 집단 내에서 탐험과 이용 능력의 균형을 유지하려 노력한 것과 달리 2개의 서로 다른 집단을 두고 하나의 집단에서는 유효한 영역에 대한 이용에 집중하고, 다른 하나의 집단에서는 미지의 영역에 대한 탐험에 집중하는 방법으로 균형을 유지하기 위해 노력하였다. 몇 가지 최적화 문제에 대한 실험 결과는 이러한 역할 분리에 의한 진화 방법이 문제 해결에 큰 도움이 될 수 있음을 보여주고 있다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 본 새로운 유전 알고리즘에 대해서 소개하고, 3장에서는 우리가 제안한 유전 알고리즘과 일반적인 유전 알고리즘의 비교 실험을 위해 사용한 문제에 대해 설명하며, 4장에서는 비교 실험한 결과를 보여준다. 마지막으로 5장에서는 결론을 맺고 향후 연구 방향에 대해 논의 한다.

2. 이중집단 유전 알고리즘

본 논문에서 제안한 유전 알고리즘은 서로 다른 목적을 가진 2개의 집단을 가지고 탐색을 수행한다. 본 논문에서는 이들 중 정보의 유효한 영역에 대한 이용(exploitation)이 강조된 집단을 집중화 집단(intensification population)이라 부르며, 미지의 영역에 대한 탐색(exploration)이 강조된 집단을 다각화 집단(diversification population)이라 부르기로 한다.

집중화 집단의 이용 능력을 높이기 위해 본 논문에서는 그리디 교차 연산자(greedy crossover operator)를 사용하였다. 일반적인 유전 알고리즘의 교차 연산 과정은 이웃해 탐색 방법으로 보면 이웃해 생성 과정과 유사하여 선택된 부모 개체로부터 새로운 개체를 생성하는 것에만 집중한다. 그러나 본 논문에서 사용한 그리디 교차 연산자는 생성된 자손 중 우수한 자손만이 다음 세대로 이어지며 선택되지 못한 자손은 그 자리에서 도태되어 버린다. 일반적인 유전 알고리즘이 생성된 모든 자손을 다음 세대에 넘겨준 후 그 자손의 이후 생존에 대해서는 확실적인 선택(selection)에 맡기는 것과는 달리 이용 집단은 교차 연산 직후 바로 우수한 자손만을 선택하는 추가 과정을 가지는 것이다. 마치 언덕 오리기 탐색 방법에서 생성된 이웃해 중 가장 좋은 해만이 다음 해가 되고 나머지는 버려지는 것과 유사한 과정을 거치게 되는 것이다.

그리디 교차 연산자를 사용할 경우 빠른 수렴을 통해 비교적 좋은 해를 빨리 찾을 수 있지만, 동시에 탐색의 균형을 잃게 하여 집단이 다양성을 빠르게 잃으며 좋지 못한 영역으로 수렴하는 조기 수렴(premature convergence) 문제를 나타나게 한다. 다각화 집단은 이러한 그리디 교차 연산자가 가져다 줄 수 있는 조기 수렴 문제를 보완하여 수렴 속도를 느리게 만들어주고, 집중화 집단이 수렴된 후에도 탐색이 가능하도록 충분한 다양성을 유지할 수 있게 하는 역할을 한다. 이를 위해 다각화 집단의 개체들은 진화과정에서 집중화 집단에 속해 있는 개체들과 가능한 다른 형태로 진화해가기 위해 노력한다. 본 논문에서는 이러한 형태로 다각화 집단을 진화시키기 위해 앞서 언급한 그리디 교차 과정에서 집중화 집단을 위해 선택되지 못한 개체들 중에서 집중화 집단을 위해 선택된 개체와 가장 다른 개체, 즉 탐색 공간 상에서 가장 거리가 먼 개체를 집중화 집단의 개체로 선택하였다. 개체간의 거리는 유클리드 거리(Euclid distance), 해밍 거리(Hamming distance) 등 염색체를 벡터(vector)로 보고 거리를 측정하는 방법이라도 적용할 수 있다.

기존의 유전 알고리즘의 경우 교차 연산을 위해 2개의 염색체를 선택하는 반면 우리의 그리디 교차 연산은 3개 이상의 염색체를 교차 연산을 위해

선택한다. 선택 방법은 집중화 집단의 경우 룰렛 선택 (roulette wheel selection), 순위 선택(ranking selection) 등 기존 유전 알고리즘을 위한 어떤 방법의 선택 방법이라도 사용할 수 있다. 다각화 집단의 경우 각 개체의 평가에 대한 부담으로 현재는 개체를 평가 하지 않은 상태에서 완전 임의 선택 (fully random selection) 방법을 사용하고 있다.

그리디 교차 연산 과정에서는 선택된 염색체들 간에 일반적인 교차 연산을 사용하여 자손 염색체들을 생성한 후, 가장 좋은 자손 염색체를 집중화 집단을 위해 선택하고, 집중화 집단을 위해 선택된 염색체와 다른 모든 염색체와의 거리를 측정한 후 가장 거리가 가장 먼 염색체를 다각화 집단을 위해 선택한다. 본 논문에서는 그리디 교차 연산 과정에서 선택되지 못하고 버려지는 염색체의 수를 줄이기 위해 집중화 집단으로부터 2개 다각화 집단으로부터 1개의 염색체를 그리디 교차 연산을 위해 선택하였으며, 그리디 교차 연산 과정에서는 다각화 집단으로부터 선택한 2개의 염색체 중 1개를 주 염색체로 선택한 후 이 염색체와 남은 2개의 염색체간에만 교차 연산을 수행한다. 즉, 2번의 교차 연산을 통해 총 4개의 자손 염색체를 생성하여, 이 중 2개는 집중화 집단과 다각화 집단 어느 쪽에도 선택되지 못한 채 버려지게 된다.

본 논문에서 제안한 유전 알고리즘의 자세한 알고리즘은 아래와 같다.

1. 집중화 집단내의 개체들을 평가한다.
2. 다음 세대를 위한 모든 염색체가 생성될 때까지 다음 과정을 반복한다.
 - A. 집중화 집단에서 선택 알고리즘을 사용하여 s_i 개의 염색체를 선택한다.
 - B. 다각화 집단에서 s_d 개의 염색체를 임의로 선택한다.
 - C. 집중화 집단으로부터 1개의 염색체를 임의로 선택 하여 pp 라 한다.
 - D. 선택된 pp 와 다른 모든 선택된 염색체와 교차 연산을 수행한다.
 - E. 생성된 모든 자손 염색체를 평가하여 가장 좋은 것을 다음 집중화 집단을 위해 선택한다.
 - F. 집중화 집단을 위해 선택된 자손 염색체와 다른 모든 자손 염색체를 비교하여 가장 거리가 먼 것을 다각화 집단을 위해 선택한다.
 - G. E와 F 과정에서 선택된 염색체 각 각에 돌연변이 연산을 적용한다.
3. 현재 집중화 집단과 다각화 집단을 새로운 집단으로 교체 한다.
4. 2에서부터 종료 조건이 만족할 때까지 반복 수행한다.

3. 실험 대상

제안한 새로운 유전 알고리즘을 평가하기 위해 기존의 일반적인 유전 알고리즘과 비교하였다. 비교 실험은 속임수 함수, 왕도 함수, 4정상 문제와 외판원 문제에 대해 수행되었다.

속임수 함수 (Deceptive Function)

유전 알고리즘이 정상적으로 작동하지 않고, 평가치가 높은 개체의 생성이 방해되는 문제를 속임수 문제라 한다. [1] 속임수 함수는 일반적으로 최적해가 탐색 공간상에서 진화가 진행되리라 기대되는 반대 방향에 존재 하게 만들어져 있어, 일반적인 유전 알고리즘으로는 최적해에 도달하기 힘들다고 기대되는 문제이다.

본 연구에 사용된 속임수 문제는 해의 총 길이가 40이며 길이가 4인 10개의 부분 해로 나누어서 평가한다. 10개의 길이 4인 조각에 대해 아래 와 같은 함수를 이용하여 평가한 후 그 합을 통해 전체 해를 평가한다. 최적해는 전체가 다 '1' 인 경우이며 그 때의 평가 값은 300 (= 30 * 10) 인 된다.

f(1111) = 30	f(0110) = 14
f(0000) = 28	f(1001) = 12
f(0001) = 26	f(1010) = 10
f(0010) = 24	f(1100) = 08
f(0100) = 22	f(1110) = 06
f(1000) = 20	f(1101) = 04
f(0011) = 18	f(1011) = 02
f(0101) = 16	f(0111) = 00

왕도 함수 (Royal road function)

왕도 함수는 이론적으로 유전 알고리즘의 최적화 메커니즘을 규명하기 위해 제안된 빌딩 블록 이론(building block theory)에 의해 매우 잘 해결할 수 있게끔 설계된 문제로서, 앞서 설명한 속임수 문제와 함께 유전 알고리즘에 대한 이론적인 분석과정에서 제안된 문제이다[2]. 이 함수의 해는 길이가 총 64이며 속임수 함수와 유사하게 길이가 8인 8개의 부분으로 나누어 각 각에 대한 평가값을 합쳐서 전체 해를 평가한다. 왕도 함수의 부분 해는 부분 해 내의 8개 값이 모두 '1' 인 경우 만 8의 평가 값을 받으며 그렇지 않은 경우 0의 평가값을 받는다. 왕도 함수의 최적해는 전체가 다 '1' 인 경우이며 이 때 최적값을 64 (= 8 * 8)이 된다.

4-정상 문제 (4 peaks problem)

4-정상 문제는 탐색 공간 상에서 4개의 정상지점(peak point)을 가지며 이들 중 2개는 지역

최적해이며, 나머지 2개는 전역 최적해이다. 이 문제는 아래와 같이 정의 되며, 해 안의 모든 값이 1이나 0인 경우 평가 값이 100이 되어 지역 최적해가 되며, 좌측으로부터 T+1개의 '0' 이 이어지고 나머지가 전부 '1' 이거나 우측으로부터 T+1개의 '1' 이 이어지고 나머지가 전부 '0' 인 경우 평가 값이 199가 되어 최적해가 된다.

$z(x)$ = Number of contiguous Zeros ending in Position 100
 $o(x)$ = Number of contiguous Ones starting in Position 1

$$REWARD = \begin{cases} 100+T & \text{if } o(x) > T \wedge z(x) > T \\ 0 & \text{else} \end{cases}$$

$$f(x) = MAX(o(x), z(x)) + REWARD$$

탐색 초기에 좌측이나 우측으로부터 연속되는 '0' 이나 '1' 을 T개 이상 가지고 시작하지 않는 경우, 자연스럽게 지역 최적해로 수렴되게 설계되어 있어, T 값이 커질 수록 지역 최적 해에 쉽게 빠져서 전역 최적 해에 도달하기 어렵게 된다.

외판원 문제 (Traveling Salesman Problem: TSP)

외판원 문제는 가장 잘 알려진 NP-complete 문제 중 하나이다. TSP 문제에서는 N개의 도시와 주어지고 도시 간의 거리가 주어지며, 전체 도시를 단 한번씩만 방문하여 전체 거리를 최단 거리로 방문하는 것이 목적이다. TSP를 유전 알고리즘으로 해결하는 방법은 여러 가지로 연구되었지만 본 논문에서는 도시를 방문 순서대로 나열하여 염색체를 구성하고 순서 교차(Ordered Crossover: OX)를 이용하여 문제를 해결하였다.

4. 실험 결과

실험을 위한 알고리즘의 설정은 다음과 같다. 일반적인 유전 알고리즘(SGA)의 경우 집단 크기는 200, 돌연 변이 확률은 0.001로 설정하였으며, 본 논문에서 제안한 유전 알고리즘(DPGA)의 경우 그리디 교차 연산 과정에서 생성된 자손 중 절반이 버려지는 관계로 공정한 비교를 위해 집중화 집단과 다각화 집단의 크기를 각 50, 두 집단 합쳐서 총 100의 으로 설정하였으며 돌연 변이 확률의 경우 집중화 집단은 일반적인 유전 알고리즘과 동일한 0.001로 설정하였으며, 다각화 집단의 경우 0.1로 설정하였다. 실험 결과 다각화 집단의 돌연 변이 확률은 다소 높게 설정하는 것이 좋다는 것이 밝혀져 이와 같이 설정하였다. 일반적인 유전 알고리즘의 경우에는 이와 같이 돌연 변이 확률을 무척 높게 설정할 경우 탐색의 임의성이 높아져서 잘 수렴되지 않아 좋은 결과를 얻을 수 없는 것으로

알려져 있다.

아래 그림 1은 속임수 문제에 대한 비교 실험 결과이며, 두 알고리즘에 30회씩 실험을 것을 각 세대까지 발견된 가장 좋은 해의 평균을 낸 것이다. 실험 결과 기존의 유전 알고리즘은 200세대 이내에 최적해에 도달하는 경우가 단 한번도 발견되지 않았으며, 새로 제안한 유전 알고리즘의 경우 모든 경우에 대해 150세대 경에 최적해에 도달하는 것을 확인할 수 있었다.

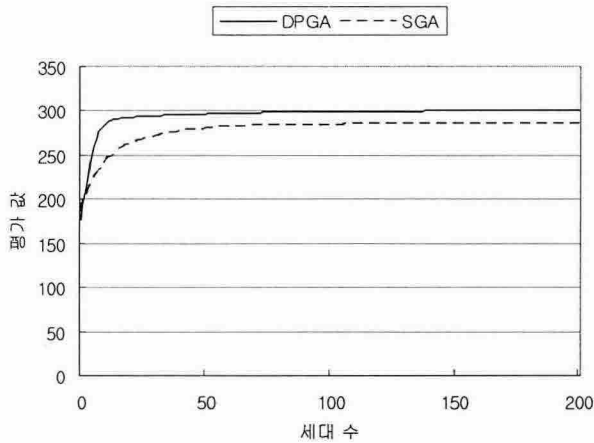


그림 1. 속임수 함수에 대한 비교 실험 결과

그림 2는 왕도 함수를 30회 수행한 후 각 세대까지 발견된 가장 좋은 해의 평균 값을 평균 낸 것이다. 제안한 DPGA가 더 좋은 결과를 보임을 확인할 수 있다. 표 1은 최적해에 도달할 때까지 평가한 해의 수를 비교한 것이며, RMHC(Random Mutation Hill Climbing)의 실험 결과는 [3]에서 인용해온 것이다. 실험 결과에 의하면 새로 제안한 DPGA의 경우에도 왕도 함수에서 언덕 오르기 탐색보다 최적해 발견을 위해 훨씬 시간이 걸리는 것은 마찬가지이지만, SGA에 비해서는 절반만의 시간만으로 최적해를 발견해 낼 수 있다는 것을 확인할 수 있다.

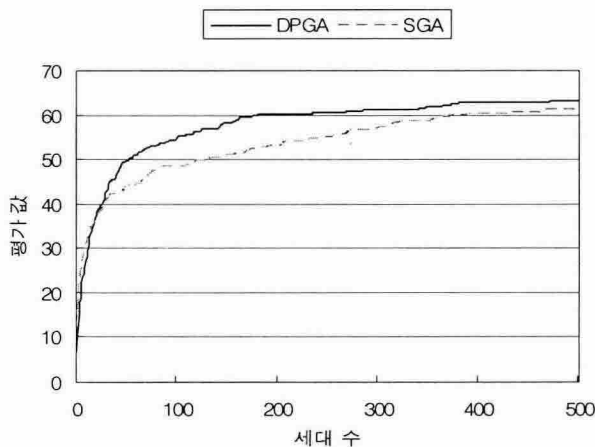


그림 2. 왕도 함수에 대한 비교 실험 결과

표 1. 최적해 발견까지 평가한 평균 해의 수 비교

	SGA	DPGA	RMHC
평가횟수	61,327	33,582	6,179

아래 표 2는 4정상 문제의 T값을 11에서 25까지 변화 시키면서 수행한 결과이며, 앞서 언급한 바와 같이 T가 커질수록 최적해에 도달하는 것이 어려운 문제이다. 실험에서 GA-1~5와 PBIL의 실험 결과는 기존 연구 [4]에서 인용해온 것이며 GA1~5 각각의 설정은 아래와 같다. PBIL 알고리즘의 경우에는 집단 기반의 탐색에 경쟁 학습 기법을 결합한 탐색 방법으로 성공적인 유전 알고리즘의 변형 중에 하나로 알려져 있다. 자세한 알고리즘은 기존 연구 []에서 자세히 확인할 수 있다.

- GA1:** Pop.: 500, One Point Crossover, Crossover rate = 80%, Mut. Rate=0.001, Elitist On
- GA2:** Pop.: 500, One Point Crossover, Crossover rate = 100%, Mut. Rate=0.001, Elitist On
- GA3:** Pop.: 500, One Point Crossover, Crossover rate = 60%, Mut. Rate=0.010, Elitist On
- GA4:** Pop.: 500, Uniform Crossover, Crossover rate = 100%, Mut. Rate=0.001, Elitist On
- GA5:** Pop.: 500, One Point Crossover, Crossover rate = 80%, Mut. Rate=0.010, Elitist On

인용해온 결과와 동일한 상황에서의 실험을 위해 각 T값에 대해 25회의 실험을 수행하였으며, 테이블의 각 숫자는 30,000번의 해를 평가하는 동안 진화 시킨 후 그때까지 발견된 가장 좋은 평가 값이 100을 넘은 횟수를 기록한 것이다. 실험 결과에서 보면 T 값이 작은 경우에는 일반적인 유전 알고리즘도 좋은 결과를 보여 주지만 T값이 커져감에 따라 그 성능이 급속도로 나빠짐을 확인할 수 있다. PBIL과의 비교에서도 모든 T값에 대해 제안한 알고리즘이 보다 좋은 결과를 보여 주는 것을 확인할 수 있다.

표 2. 4-정상 문제에 대한 비교 실험 결과

	11	13	15	17	19	21	23	25
GA-1	25	23	22	14	11	3	6	2
GA-2	25	25	22	19	10	5	3	1
GA-3	25	25	23	15	13	3	2	3
GA-4	25	21	19	12	2	4	2	0
GA-5	25	22	15	7	4	1	0	1
PBIL	25	25	25	24	23	15	12	5
DPGA	25	25	25	25	23	17	13	12

표 3은 도시 수를 10개에서 50개까지 변화시키면서 TSP 문제에 대해 비교 실험한 결과이다. 이 실험의 경우 SGA의 집단 크기는 800으로 DPGA의 집단 크기는 200 + 200으로 설정하였다. 각 문제에 대해 1000세대의 제한을 가지고 30회 수행한 결과의 평균을 비교한 것이다. 실험 결과 도시의 수가 10개로 비교적 적은 경우에는 DPGA와 SGA가 동일한 결과를 보이지만 도시의 수가 많은 경우 DPGA가 더 좋은 결과를 보여줌을 확인할 수 있다.

표 3. 외판원 문제(TSP)에 대한 비교 실험 결과

도시 수	DPGA	SGA
10	262	262
20	550	570
30	1005	1040
40	1440	1477
50	1824	1857

4. 결론 및 향후 연구 과제

본 논문에서 우리는 하나의 집단에서는 탐색의 집중화를 다른 하나의 집단에서는 탐색의 다각화를 주로 수행하는 다른 두 개의 집단을 이용하여 탐색을 수행하는 새로운 유전 알고리즘을 소개하였다. 몇 가지 최적화 문제에 대한 비교 실험 결과 제안한 알고리즘이 전통적인 유전 알고리즘에 비해 최적화 문제 해결에 우수함을 확인할 수 있었다.

본 논문에서 우리는 최적화 집단과 다각화 집단을 분리하여 진화를 수행하는 방법의 가능성을 확인하였으나, 각 각의 집단을 가장 잘 진화시키는 방법에 대한 연구는 아직 추가로 필요하다고 생각된다. 예를 들어, 다각화 집단의 경우 교배를 위한 선택을 현재는 완전 랜덤 선택을 수행하고 있으나, 집중화 집단을 기준으로 각 개체를 평가하여 진화시킬 경우 현재보다 더 좋은 결과를 얻을 수 있을 것이라 생각된다.

또한, 우리는 향후 보다 다양한 문제에 대한 실험과, 다른 변형된 유전 알고리즘과의 비교 실험 등 보다 많은 비교 실험을 수행해 나갈 예정이다.

참고 문헌

- [1] Goldberg, D.E., and Richardson, J. (1987). "Genetic Algorithms with sharing for multimodal function optimization." *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*.
- [2] Mitchell, M., Forrest, S., and Holland, J.H. (1992). "The royal road for genetic algorithms: Fitness landscapes and GA performance." *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*. MIT Press.
- [3] Mitchell, M. (1996) *An Introduction to Genetic Algorithm*, MIT Press.
- [4] Balusa, S., and Caruana, R., (1996). "Removing the Genetics from the Standard Genetic Algorithm", *Proceedings of the International Conference on Machine Learning*, pp. 38-46