n차원 구면상에서의 실수 코딩 유전 알고리즘

김진현^{O}, 문병로
서울대학교 컴퓨터공학부
jh@soar.snu.ac.kr, moon@snu.ac.kr

Real Coded Genetic Algorithm On n-Dimensional Sphere

Jin-Hyun Kim^{O}, Byung-Ro Moon
School of Computer Science & Engineering
Seoul National University

요 약

본 연구에서는 실수 코딩을 사용하는 유전 알고리즘의 문제공간이 n차원 구면으로 제한된 경우에 사용할 교배 연산자와 변이 연산자를 제안하고, 이를 실제로 사용한 실험 결과를 제시한다. n차원 실수 공간에서 일반적으로 사용되는 연산자를 n차원 구면에 사영하는 방법을 사용하였으며, 해의 범위가 제한된 경우에 사용할 해의 수선 방법도 제안하였다. 제안된 연산자를 사용하며 몇 가지 최적화 문제를 푸는 실험을 한 결과 평균 오차율 2.0%내에서 최적해를 구함을 확인하였다.

1. 서론

유전 알고리즘(Genetic Algorithm)은 진화의 원리를 문제 해결에 이용하는 방법의 하나다. 이 때 문제의 해들은 알고리즘 상에서 하나의 염색체로 표현되며 이를 표현하는 방법을 코딩이라 한다. 대표적인 코딩 방법으로는 이진수 형태로 표현하는 이진 코딩(Binary Coding)과 실수 형태로 표현하는 실수 코딩(Real Coding) 등이 있다.

초창기 유전 알고리즘에서는 교차 연산의 편의 등을 이유로 이진 코딩을 많이 사용하였다. 이후 독일의 진화전략(Evolutionary Strategy) 그룹이 사용하던 실수 코딩이 유전 알고리즘에도 시도되며 다양한 연구가 있어 왔다. 실수 코딩을 사용하면 n차원 실수 공간 등의 연속적인 공간에서의 문제를 풀 수 있으며 해의 속성이 실수인 경우 직관적으로 해를 표현할 수 있다는 장점이 있다[1, 2].

그동안 문제공간이 n차원 실수 공간 전체인 경우, 혹은 범위가 제한된 부분 공간인 경우에 유전 알고리즘을 사용한 연구가 꾸준히 있어 왔다. 그러나 해의 각 변수들의 제곱 합이 1인 n차원 구면상에서의 연구는 별로 없었다. 이러한 n차원 구면 문제공간은 변수들의 절대적인 크기보다 상대적인 크기가 중요한 경우에 응용될 수 있으며 해를 정규화(Normalization)하는 효과가 있다는 장점이 있다.

본 연구에서는 이러한 n차원 구면상에서의 실수 코딩 유전 알고리즘을 위한 연산자를 제안하고 이를 이용한 실험 결과를 제시한다. 이를 위하여 기존에 n차원 실수 공간에서 사용되던 연산자를 n차원 구면에 적용한다. 문제에 따라서는 해의 변수들이 양수여야 하는 등의 제약이 있을 수 있는데, 이처럼 범위가 제한된 경우에 대한 방법도 소개한다. 실험 결과 이러한 방법을 활용할 경우 평균적으로 오차율 2% 수준에서 최적해를 찾음을 확인할 수 있었다.

이 논문의 2장에서는 먼저 문제공간이 되는 n차원 구면에 대해서 설명한다. 3장에서는 이러한 문제공간에서 사용할 교배 연산자와 변이 연산자를 제안하고, 범위가 제한된 경우에 사용할 해의 수선 방법도 제안한다. 4장에서는 제안된 연산자를 사용하여 실험한 결과를 정리한 후 5장에서 결론을 맺는다.

2. 문제공간

유전 알고리즘에서 실수 코딩을 사용하여 문제를 푸는 경우 n개의 변수들은 임의의 실수값을 가질 수 있다. 따라서 문제의 해는 n차원 실수 공간을 문제공간으로 갖게 된다. 이러한 문제공간에서는 해의 적합도 함수(Fitness Function)가 최소값이나 최대값을 갖지 않을 수도 있으며, 따라서 문제의 요구사항에 따라 범위가 제한된 문제를 다룬다. 즉, 각각의 변수가 일정한 범위 제약식을 만족하는 문제를 다룬다.

본 연구에서는 n차원 구면을 문제공간으로 갖는 유전 알고리즘의 연산자를 제안한다. n차원 구면은 해를 이루는 n개의 변수의 제곱합이 1인 공간이다. 즉, n개의 변수가 $x_1, x_2, \cdots, x_n$일 때 제약식

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1$$

를 만족하는 공간이다. 유전 알고리즘의 연산자에 따라 해를 생성한 후, 이를 범위 제약식에 맞춰 수선(Repair)하는 작업은 간단하다. 그러나 n차원 구면의 경우에는 제곱합에 대한 제약식도 만족해야 하므로 해를 생성하는 것이 간단하지 않다.

n차원 구면의 경우에도 제약식과 실수의 성질로부터

자연스럽게 범위 제한이 도출된다. 즉, 각각의 변수 x_i ($1 \le i \le n$)는 $-1 \le x_i \le 1$의 범위 제한을 만족해야 한다. 특별한 경우에는 이보다 축소된 범위를 다룰 수도 있다. 예를 들어, 각각의 변수가 거리나 비용 등을 의미하는 경우 그 값을 $0 \le x_i \le 1$처럼 양수로 제한하는 것이 자연스러울 수 있다. 이처럼 추가로 범위 제한이 있는 경우에는 유전 알고리즘에서 해를 생성하는 것이 보다 까다롭다.

해의 적합도 함수가 좋은 성질을 만족하는 경우에는 라그랑즈 승수(Lagrange Multiplier)를 이용하여 최적해를 구할 수 있다[3]. 적합도 함수가 $f(x_1, x_2, \cdots, x_n)$일 때, 최적해는 제곱합에 대한 제약식과 미분에 대한 제약식

$$\frac{\partial f(x_1, x_2, \cdots, x_n)}{\partial x_i} - 2\lambda x_i = 0$$

을 만족해야 한다. 따라서 이를 x_i와 λ에 대한 방정식으로 보고 이를 풀어서 해를 구할 수 있다. 그러나 이러한 방법을 사용할 수 있는 경우는 제한적인데, 이는 적합도 함수 f가 미분 가능해야 하며 f를 미분한 결과로부터 얻어지는 방정식이 풀 수 있는 형태여야 하기 때문이다. 따라서 유전 알고리즘을 사용하여 이러한 문제를 푸는 것은 자연스러운 시도다.

이러한 n차원 구면상에서 문제를 풀기 위한 시도는 다양한 활용 예를 가질 수 있다. 특별히 변수들이 제곱합에 대한 제약식을 만족하는 경우 해를 정규화하는 효과가 있다. 이를 이용하면 각 변수들의 절대적인 수치보다도 상대적인 크기가 중요한 문제를 모델링 하는 데에 활용할 수 있다.

3. 연산자 제안

3.1. 교배 연산

유전 알고리즘에서의 교배 연산(Crossover Operator)은 두 개의 부모 해를 통해서 새로운 자식 해를 만들어내는 연산이다. 이 때 자식 해는 부모 해의 매력적인 형질을 상속받으며, 유전 알고리즘의 세대가 진행되면 이러한 형질들이 쌓여서 최적해를 탐색하게 된다.

n차원 실수 공간 전체에서 실수 코딩을 사용하는 경우 가장 직관적인 교배 연산은 부모 해를 나타내는 두 점의 중점, 혹은 내분점을 자식 해로 삼는 것이다. 이러한 교배 연산은 초기에 해들이 이루는 영역의 내부만을 계속 탐색하므로, 채굴(exploitation)의 경향은 강하지만 탐험(exploration)의 경향은 약하다. 따라서 최적해가 문제공간의 바깥쪽에 위치하는 경우에 약점을 보인다.

이를 보완하기 위해 제안된 교배 연산이 BLX-α다[1]. 이 교배 연산에서는 두 변수의 내분점 외에도 α만큼의 외분점도 고려 대상이 된다. 두 부모 해의 i번째 변수가 각각 x_i, y_i라 하자. 먼저 분할 비율 β가 $[-\alpha, 1+\alpha]$의 범위에서 무작위로 선택된다. 그러면 자식 해의 i번째 변수는 $x_i \times \beta + y_i \times (1-\beta)$의 값을 갖는다. β가 0보다 작은 경우에는 y_i쪽의 외분점이 선택되는 것이고, 1보다 큰 경우에는 x_i쪽의 외분점이 선택되는 것이다. 그 외의 경우에는 내분점이 선택된다. α는 문제의 특성에 따라 다양하게 선택될 수 있는데, 일반적으로 0.25가 많이 사용된다. α가 0인 경우는 내분점만을 사용하는 경우다.

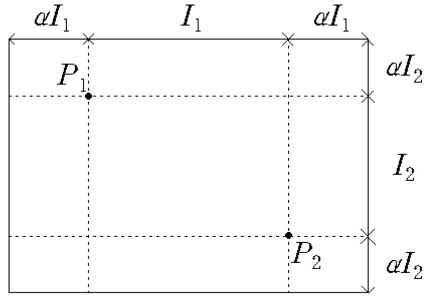

그림 1. 2차원 BLX-α

그림 1은 이러한 BLX-α를 2차원에 적용한 예다. 두 부모 해는 2차원 평면상의 두 점 P_1, P_2로 나타나며, 자식 해는 교배 연산의 결과 전체 사각형 내부의 영역에서 선택되게 된다. n차원의 경우에는 두 점을 대각선으로 마주보는 n차원 박스를 각 방향으로 α의 비율로 확장하여 그 내부에서 자식 해를 선택하게 된다.

이러한 BLX-α는 문제공간이 n차원 실수 공간 전체인 경우, 혹은 각 변수가 범위 제한을 갖는 경우에 사용될 수 있는 방법이다. 이를 n차원 구면상에서 활용하기 위해서는 자식 해를 구면상으로 사영(Projection)하면 된다. 이는 n차원 벡터를 길이가 1인 벡터로 정규화하는 것으로 i번째 변수 i를 벡터의 길이로 나눈

$$\frac{1}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}} x_i$$

로 치환하는 것이다.

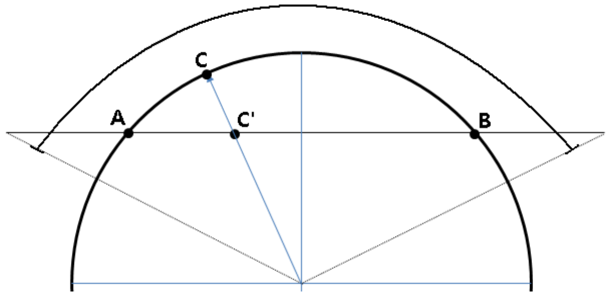

그림 2. 2차원 구면상의 BLX-α

그림 2는 이러한 구면상의 BLX-α를 2차원 구면에 적용한 예다. 두 부모 해는 구면상의 점 A, B로 나타난다. 먼저 2차면 구면을 고려하지 않은 상태에서 BLX-α를 수행하여 점 C를 얻는다. 그 후, 이를 구면상으로 사영한 점 C를 자식 해로 사용한다.

부모 해와 분할 비율에 따라서는 자식 해를 정규화할 수 없는 경우도 있다. 이는 벡터의 길이가 0인 경우로, 예를 들어 $x_1 = 1$인 해와 $y_1 = -1$인 해를 부모 해로 선택하여 BLX-α를 수행하면 원점이 자식 해가 될 수도 있다. 이를 막기 위해서 길이가 0인 벡터가 생성되는 경우에는 교배 연산을 다시 수행해야 한다.

n차원 실수 공간에서 각 변수가 범위를 갖는 경우를 다뤘던 것처럼 n차원 구면에서도 이러한 경우를 다뤄야 할 때가 있다. 이러한 경우에는 교배 연산을 수행한 후 해의 수선 과정을 거쳐야 하는데 이에 대해서는 3절에서 설명한다.

3.2. 변이 연산

유전 알고리즘에서의 변이 연산(Mutation Operator)은 생성된 자식 해에 추가적인 변형을 가하는 연산이다. 자식 해는 부모 해들로부터 매력적인 형질을 상속받지만 이들에게서 나타나지 않는 매력적인 형질도 있다. 이처럼 완전히 새로운 형질을 갖기 위해서는 변형이 필요하며 이러한 변형이 매력적인 경우에는 유전 알고리즘의 세대가 진행되어도 계속 보존된다.

n차원 실수 공간 전체에서 실수 코딩을 사용하는 경우 가장 직관적인 변이 연산은 특정 변수의 값을 임의의 값으로 치환하는 것이다. 이를 사용할 경우 n차원 실수 공간의 한 축 방향으로 큰 탐험의 효과를 기대할 수 있지만 부모 해와 지나치게 동떨어진 공간을 탐색할 수도 있다. 이를 보완한 방법으로는 난수를 발생시켜 기존 변수에 더하는 방법이 있다.

일반적으로 한 변수 x_i에 변이를 가할 때는 가우시안 노이즈(Gaussian Noise)를 가하는 방법이 사용된다[4]. 즉, 평균이 0이고 표준편차가 σ_i인 가우시안 분포 $N(0, \sigma_i)$를 따르는 난수를 발생시켜 x_i에 더한다. 표준편차 σ_i의 값은 문제에 따라 다양하게 사용할 수 있는데, 일반적으로 $(\max_i - \min_i)/4$의 값이 널리 사용된다. $\min_i$와 $\max_i$는 각각 i번째 변수가 가질 수 있는 최소값과 최대값이다.

n차원 구면에서 변이 연산을 수행할 때에도 변이를 가한 해를 구면상에 사영하는 방법을 사용한다. 먼저 자식 해를 n차원 실수 공간상의 점으로 보고 변이를 가하고, 얻어진 점을 구면상에 사영한 것을 해로 사용한다. 변이 연산에서도 결과로 얻어진 벡터의 길이가 0일 수 있다. 이때에는 하나 이상의 유전자에 대한 변이 연산을 취소하면 된다. 본 연구에서는 전체 연산을 취소하고 변이 연산을 처음부터 다시 시작하는 방법을 택하였다.

변이 연산도 교배 연산과 마찬가지로 범위가 제한된 n차원 구면에서는 추가적인 수선 작업이 필요할 수 있다. 이에 대해서는 3절에서 설명한다.

3.3. 해의 수선

n차원 실수 공간은 원점을 제외하고 모두 n차원 구면상에 사영 가능하다. 따라서 범위 제약 없이 n차원 구면 전체를 문제공간으로 삼는 경우에는 원점이 해로 생성되지 않는 이상 수선 과정이 필요하지 않다. 그러나 범위 제한이 있는 경우에는 각 변수마다 범위 제한을 맞춰야 하며, 이를 n차원 구면상에 사영한 결과도 범위 제한을 맞춰야 하므로 해의 수선이 단순하지 않다.

교배 연산 및 변이 연산이 수행된 후, 먼저 각 변수의 범위 제한을 조절한다. 변수 x_i가 $\min_i$보다 작은 값을 갖는 경우에는 $\min_i$로 설정하며 $\max_i$보다 큰 값을 갖는 경우에는 $\max_i$로 설정한다. 만일 이와 같은 범위 조절을 수행한 결과가 원점이라면 n차원 구면상에 사영이 불가능하므로 연산을 처음부터 다시 수행한다.

범위 조절이 완료된 후에는 결과로 얻어진 점을 n차원 구면상에 사영한다. 각 변수가 가질 수 있는 값이 범위가 좁은 경우에는 사영한 결과가 다시 범위 제한을 벗어날 수 있다. 이때에는 n차원 구면상에 사영한 점에 대해서 범위를 조절하는 과정과 n차원 구면상에 사영하는 과정을 반복적으로 수행한다.

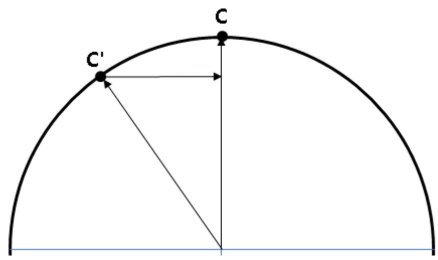

그림 3. 2차원 구면상의 수선

그림 3은 2차원 구면상에서 해를 수선하는 과정의 예다. 해의 범위는 두 변수가 모두 0 이상인 제 1사분면이라 하자. 교배 연산 및 변이 연산을 통해서 범위 제한을 벗어난 C가 먼저 생성된다. 범위를 벗어난 변수 x_1을 0으로 맞추고, 이를 2차원 구면상에 사영한 점 C'를 얻게 된다. 이때에는 C'가 범위 제약을 만족하므로 수선 과정이 완료된다.

4. 실험

4.1. 대상 함수

실수 코딩을 사용하는 유전 알고리즘의 성능을 평가할 수 있는 다양한 모델들이 이미 제시되어 있다[1, 4, 5]. 그러나 이러한 모델들은 범위가 제한된 n차원 실수 공간을 대상으로 한다. 이러한 함수들을 n차원 구면상에 제한할 경우 원래 모델의 의미를 충분히 살리지 못할 수 있다. 본 연구에서는 이러한 모델을 참고로 한 함수와 몇 가지 간단한 함수, 그리고 수치 해석적으로 의미를 갖는 함수[6, 7]를 사용하여 실험하였다. 범위가 제한된 n차원 구면상에서의 성능을 확인하기 위해서 모든 변수의 범위를 $0 \le x_i \le 1$로 제한하였다. 각 함수는 다음과 같으며 모두 차원의 크기로는 $n = 100$을 사용하였다.

- $f_1(x) = -\sum_{i=1}^{100} x_i$
- $f_2(x) = \sum_{i=1}^{50} x_i^2 + \sum_{i=51}^{100} (x_i - y)^2, \quad y = \frac{1}{\sqrt{50}}$
- $f_3(x) = x^T Qx, \quad Q_{i,i} = 2, \quad Q_{i,i-1} = Q_{i,i+1} = -1$
- $f_4(x) = x^T Qx, \quad Q_{i,i} = n \quad Q_{i,j(i \neq j)} = -1$
- $f_5(x) = \sum_{i=1}^{100} \cos(i\pi x_i)$

4.2. 실험 설정

본 연구에서는 순수 안정 상태 유전 알고리즘(Pure Steady State Genetic Algorithm)을 사용하여 실험하였다. 교배 연산과 변이 연산은 3장에서 제안한 것들을 사용하였으며, 나머지 설정은 다음과 같다.

- 초기화: n차원 구면상의 점을 임의로 선택하였다.
- 해집단의 크기: 100을 사용하였다.
- 선택 연산: 2원 토너먼트 선택을 사용하였다. 임의로 두 해를 고른 후, 80%의 확률로 보다 좋은 해가 부모로 선택되며 20%의 확률로 보다 나쁜 해가 부모로 선택된다.
- 변이 확률: 0.5%를 사용하였다. 이는 각 변수에 대해서 서로 독립적으로 적용된다.
- 대치 연산: 해집단에서 품질이 가장 나쁜 해를 대치한다. 자식 해의 품질이 보다 나쁘다 하더라도 무조건 대치하도록 하였다.
- 종료조건: 5만 세대가 진행된 후 종료하도록 하였다.

4.3. 결과 및 분석

각 함수에 대해 유전 알고리즘을 천 번씩 수행한 후, 유전 알고리즘이 구한 해의 평균을 계산하고 이를 최적해와 비교하였다. 또한 이를 바탕으로 오차율을 계산하였으며, 최적해가 0인 경우에는 상대 오차를 구할 수 없으므로 절대 오차를 사용하여 오차율을 계산하였다.

<표 1>

	평균해	최적해	오차율
1	-9.9987	-10	0.0013%
2	0.0027482	0	0.27%
3	0.0068459	0	0.68%
4	1.0251	1	2.5%
5	-93.534	-100	6.5%

실험 결과 다섯 가지 경우 모두 최적해와 유사한 값을 구하였다. 가장 간단한 함수인 첫 번째 함수의 경우 실제 최적해와 대단히 가까운 값을 구할 수 있었으며, 오차율이 가장 큰 경우도 6.5%로 충분히 작았다. 평균적인 오차율은 2.0%였다. 다섯 번째 함수의 경우에는 최적해를 명시적으로 알 수 없어서 삼각 함수의 범위에 따라 -100으로 가정하였다. 따라서 실제 최적해는 이보다 약간 큰 값일 수 있으며, 이 경우 오차율이 약간 과대 산정되었다.

실험상에서 실제로 최적해를 정확하게 구한 경우는 없었다. 그러나 일반적으로 n차원 실수 공간에서 유전 알고리즘을 사용할 때에는 일정 오차 범위(0.1, 0.01, 등) 내에서 해를 구한 경우에 모두 최적해로 간주한다[1, 4]. 따라서 모든 실험에서 유의한 수준의 최적해를 구하였으며, 이를 통해서 제안한 교배 연산과 변이 연산이 유의함을 확인할 수 있었다.

5. 결론

지금까지 n차원 구면상에서 실수 코딩을 사용하는 유전 알고리즘을 위한 교배 연산과 변이 연산을 제안하고 그 성능을 실험해 보았다. 기존의 n차원 실수 공간에서 사용되던 연산자를 자연스럽게 n차원 구면에 적용하였으며, 특별히 범위가 제한된 n차원 구면의 경우에 발생할 수 있는 문제를 해결하는 방안도 제시하였다. 이를 바탕으로 실제 최적화 문제를 풀어 본 결과 평균 오차율 2.0% 내에서 유의한 결과를 냄을 확인하였다.

본 연구에서는 실험 결과 비교상의 편의를 위하여 유전 알고리즘의 설정을 동일하게 유지하였다. 이러한 설정을 문제에 맞춰 조절할 경우 유전 알고리즘의 성능을 향상시킬 수 있다. 또한 순수 유전 알고리즘은 미세 조정(Fine Tuning) 능력이 떨어지므로 지역 최적화 알고리즘(Local Optimization Algorithm)과 결합하여 사용하는 것이 일반적이다. 이러한 점들을 고려하여 문제에 최적화 된 유전 알고리즘을 사용할 경우 제안된 연산자들의 성능이 보다 향상될 것이다.

Acknowledgement

이 논문은 두뇌한국21사업에 의하여 지원되었으며, 서울대학교 BK21 정보기술사업단의 도움으로 작성되었습니다. 또한 이 연구를 위해 연구 장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에 감사드립니다.

참고 문헌

[1] F. Herrera, M. Lozano, and A. M. Sanchez, Hybrid crossover operator for real-coded genetic algorithms: an experimental study, Soft Computing - A Fusion of Foundations, Methodologies and Applications, Volume 9, Issue 4, 280-298, 2005.
[2] 문병로, “쉽게 배우는 유전 알고리즘: 진화적 접근법”, 한빛미디어, 2008.
[3] David G. Luenberger, "Linear and Nonlinear Programming", Addison-Wesely, 1984.

[4] Shigeyoshi Tsutsui, Masayuki Yamamura, and Takahide Higuchi, Multi-parent Recombination with Simplex Crossover in Real Coded Genetic Algorithms, Proceedings of the Genetic and Evolutionary Computation Conference 1:657-664, 1999.
[5] F. Seredynski, A. Y. Zomaya, and P. Bouvry, Function Optimization with Coevolutionary Algorithms, "Intelligent Information Processing and Web Mining", 13-22, Springer, 2003.
[6] 김홍종, “미적분학 2”, 서울대학교출판부, 2009.
[7] Erwin Kreyszig, "Advanced Engineering Mathematics", John Wiley & Sons Inc., 2005.