

캐시풀을 이용한 유성생식 유전알고리즘의 새로운 탐색전략

*류근배 김창업 이학성

효성중공업(주) 기술연구소

New Search Strategy in Sexual Reproduction Genetic Algorithms Using Cache Pool

*Keun Bae Ryu Chang Eob Kim Hahk Sung Lee

R&D Institute, Hyosung Industries Co., Ltd.

Abstract

A new method is proposed for tracking the optimum points in nonstationary problem via genetic search. Cache Pool to save the past genetic informations is added to population in search using Sexual Reproduction Genetic Algorithm(SRGA). In Cache Pool, elite chromosomes from population are accumulated. A best Individual is made up from these chromosomes in varying environment and inserted into the newly reproduced population every generation.. Experimental results indicate changing global optima are accurately identified and followed.

서론

유전알고리즘(Genetic Algorithm; GA)은 강력한 최적점 탐색방법으로서 넓은 응용영역을 갖는다. 특히, 다극점문제(multimodal problem)나 미분불가능한 문제등에 대해서도 전역해(global solution)를 찾는 데 효과적이다. 그러나 이러한 장점에도 불구하고 시간에 따라 변하는 최적점을 추적해야 하는 변동문제(nonstationary problem)에 대해서는 성능이 크게 약화된다. 그 이유는 정확한 해를 찾아야 하는 국부탐색(local search)의 과정에서 개체들의 유전적 다양성(genetic diversity)을 잃기 때문이다. 이에 대한 대응으로 개체를 2개의 염색체로 하여 이배체(diploid)로 표현하고 우성(dominance)관계를 부가하는 방법이 제시되었다.[1][2] 이러한 방법은 정보의 이중저장으로 장기간기억(long term memory)의 효과를 갖고 다양성의 상실을 최소화할 수 있지만 많은 개체를 필요로 하고 최적점이 둘 이상의 상태로 크게 변동할 때에는 한계가 있는 단점이 있다. 또한 다극점문제에 있어서의 성능은 아직 검증된 바 없다.

본 논문에서는 변동문제에 대한 대응으로 유성생식 유전알고리즘(Sexual Reproduction Genetic Algorithm; SRGA)[3]의 강력한 탐색 성능을 활용하면서 환경변화에 대한 직접적인 해결방법을 검토한다. SRGA에서도 이배체 개체를 도입하고 있으며, 감수분열(meiosis)과정을 통해 배우자(gamete)를 선택하는 다중선택(multi-selection)의 구조를 갖는다. 본 논문에서는 SRGA의 탐색성능을 활용하면서 변동문제에 대한 직접적인 해결방법으로 캐시풀(Cache Pool)의 도입을 제시한다. 캐시풀은 랜덤으로 초기화되는 일반적인 개체군(population)과는 별도로 각 세대에서 최우량 개체의 염색체를 저장하도록 하여 형성한다. 캐시풀에 저장된 염색체들은 생식(reproduction)과정을 통해 새로운 세대가 형성될 때마다 저장중인 염색체중 가장 높은 적합도를 갖는 것만을 합성하여 새로운 개체를 만든 후 개체군에 추가된다. 본 논문은 캐시풀 도입의 효용성을 보이기 위해 함수최적화(function optimization)에 적용하고 그 결과를 보인다.

함수최적화의 변동문제

테스트 함수

GA의 탐색성능을 평가하기 위한 여러 가지 테스트함수가 제기되어 왔다. F6 함수는 대표적인 다극점문제로서 GA의 탐색성

능을 평가하는데 많이 쓰인다. 변수가 2개일 경우 F6는 다음과 같다.

$$0.5 - \frac{\sin^2 \sqrt{(x-x_0)^2 + (y-y_0)^2} - 0.5}{[1 + 0.001 \{(x-x_0)^2 + (y-y_0)^2\}]^2} \quad (1)$$

여기서, (x_0, y_0) 는 최대값을 갖는 X, Y 변수의 최적점 좌표이다. 그림 1은 Y 변수값을 고정하고 X변수에 따라 2차원 상으로 F6함수를 나타낸 것이다.

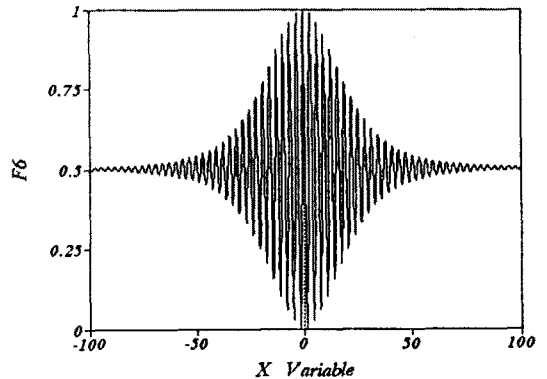


그림 1. F6 함수

환경의 변화 - 최적점의 이동

GA를 이용한 일반적인 최적점 탐색에서는, 랜덤 초기화로부터 일단 최적점을 찾은 이후에는 대부분의 개체가 최적점으로 모이게 되어 다양성을 잃게 된다. 이러한 상태에서 최적점이 급격히 변화할 경우 이전의 최적값을 가졌던 개체들은 대부분 낮은 적합도 값을 가져 다음 세대로 존속하지 못하게 되며 다양성을 잃은 상태에서 최적화를 다시 진행하게 되는 상황을 초래한다. 따라서 그동안의 탐색과정에서 얻었던 탐색결과는 무용하게 되며, 변화가 계속되면 될수록 안정되지 못하는 문제가 발생한다. 이러한 점은 병렬탐색의 잇점에도 불구하고 변동문제에 대해서는 전역최적해를 추적하지 못하는 기존 GA의 근본적인 단점이다. 본 논문에서는 최적점이 이동하는 상황을 다음과 같이 가정하였다. 즉, F6함수의 전역최적점이 그림 2에 보이는 바와 같이 (-50, -50)에서부터 원점 (0,0)를 거쳐 (50,50)까지 이동하고 다시 처음부터 주기적으로 반복하는 형태로 설정하였다. 이러한 형태는 Smith[1]가 예로 들었던 배낭문제(knapsack problem)가 이동되는 최적점을 2개로 한정되었던 것보다 더욱 복잡한 형태이며, 테스트 함수 자체가 많은 국부최적점(local optima)을 포함하는 다극점 문제라는 점이 다른 점이다.

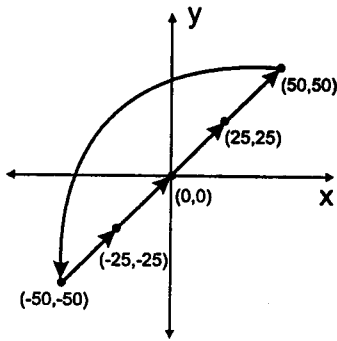


그림 2. 최적점 좌표의 이동

캐시풀(Cache Pool)의 도입

유성생식 유전알고리즘

유성생식 유전알고리즘(Sexual Reproduction Genetic Algorithm; SRGA)[3]은 자연계의 생식과정에 기초한 유전알고리즘 모델로서 생태계와 유사한 이배체(diploid) 개체를 사용한다. 개체의 생식(reproduction)을 위해서 기존의 유전알고리즘에서 도입되지 않은 감수분열(meiosis)을 인공적으로 모델링하여 이용한다. 감수분열을 통해 생식세포(gamete)가 형성되고, 자손(offspring)으로 이어질 배우자는 경쟁원리를 도입하여 선택할 수 있다. 이러한 배우자 선택(gamete selection)과정이 진화연산자로서 기능한다. 따라서 SRGA에서는 교배(mating)할 개체를 선택하는 개체선택(individual selection)과 생식과정에서의 배우자선택이 다중선택(multi-selection)의 구조를 이루게 되며 이배체와 다중선택의 구조에 의해 최적화 성능향상과 다양한 응용이 가능하다.[3][4]

캐시풀의 활용 방법

캐시풀은 최적화를 위한 개체군(population)과는 별도로 형성되며, 각 세대에서 가장 높은 적합도(fitness)를 갖는 염색체(elite chromosome)가 저장된다. 저장되는 염색체는 최적점을 찾기 위한 과도과정에는 거의 매세대마다 캐시풀에 저장된다. 저장할 때에는 이미 저장된 염색체와 중복되지 않도록 하므로, 일단 최적해를 찾은 이후의 정상상태(steady state)에서는 새로 추가되는 염색체는 없게 된다. 저장된 염색체들은 다음 세대 개체군과 함께 동시에 평가과정을 거친 후 환경에 가장 적합한 염색체가 개체군에 추가된다. SRGA는 이배체 모델이므로 최우량 개체와 두 번째로 큰 적합도를 갖는 염색체를 조합하여 새로운 개체를 만든다. 최적화 과정은 다음과 같다.

- 과정(1) N개의 개체를 갖는 개체군(population)을 랜덤 초기화한다.
- 과정(2) 개체군내 개체들의 적합도 값을 구한다. 캐시풀(cache pool)이 비어있는 초기에는 과정(4)로 넘어간다.
- 과정(3) 캐시풀에 있는 과거의 정보를 활용하기 위해 다음 과정을 행한다.
 - a) 캐시풀내의 염색체들의 적합도(fitness)를 구한다.
 - b) 캐시풀내의 염색체들 중 가장 높은 적합도를 갖는 염색체 2개를 조합하여 하나의 개체로 합성한다.
 - c) 합성된 개체를 개체군에 넣는다. 개체수는 N개
- 과정(4) 개체군내에서 가장 높은 적합도를 갖는 개체의 우성염색체를 캐시풀에 복사하여 저장한다.
- 과정(5) 다음세대의 개체군을 생성하기 위한 생식(reproduction)을 행한다.
 - a) 두 개의 개체를 적합도에 비례하는 확률로 선택한다.
 - b) 각각 감수분열(meiosis)을 시행한다. 이 때 염색체간에 교차(crossover)와 돌연변이(mutation)를 시행한다.
 - c) 감수분열중 발생한 배우자를 각각 선택하고 결합하여 개체를 발생시킨다.
- N-1 개의 개체가 발생할 때까지 과정(5)를 반복한다.
- 과정(6) 계속해서 과정(2)~과정(5)를 반복한다.

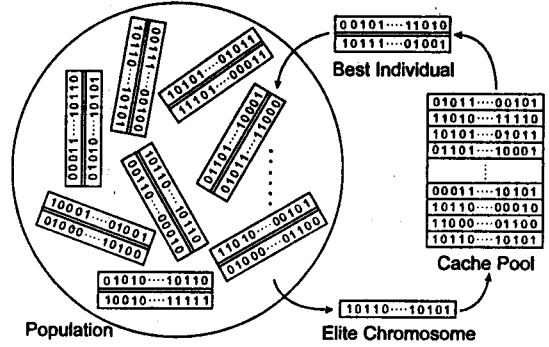


그림 3. 개체군과 캐시풀간의 염색체/개체 교환

그림 3은 세대마다 발생하는 개체군과 캐시풀간의 유전정보 교환과정을 나타낸 것이다. 최적점이 고정된 상태일 때의 캐시풀의 기능은 바로 이전세대의 최우량 개체(best individual)가 다음 세대에 반드시 존속하도록 하는 엘리트전략(elitism)의 효과와 같다. 따라서, 기존의 GA에서 많이 쓰이는 엘리트전략은 캐시풀의 크기를 하나로 제한한 특별한 경우로 볼 수 있다. 최적점이 이동하는 상황에서는 바로 이전세대의 정보만을 이용하는 엘리트전략만으로는 효과가 없으며, 장기간 저장되었던 정보의 활용이 요구된다. 과거의 최적화 과정에서 캐시풀에 저장되었던 염색체들 중 현재 환경에 가장 적합한 염색체를 재활용함으로써 재탐색에 의한 시간소모를 최소화하고 성능저하를 방지할 수 있다.

캐시풀 도입의 의미

캐시(cache)란 개념은 컴퓨터구조에 도입되면서 컴퓨터의 속도향상에 큰 기여를 하였다. 예를 들어, 사용빈도가 높은 데이터는 원래위치와 별도로 빠른 접근이 가능한 특별한 곳, 즉 캐시에 저장하여 재활용도를 높이고 전반적인 속도가 개선되는 효과를 가져온다. 이러한 원리를 최적화과정에도 적용할 수 있다. GA를 이용하여 광범위한 해공간(solution space)을 탐색하는 과정에서 최적점 이동이 반복되는 경우 개체들이 갖고 있는 불확실한 기억에 의존하는 것보다는 과거의 정보를 별도로 저장하여 재활용하는 것이 더 효율적이다.

캐시풀을 이용하는 이러한 방법은 인공적인 의미를 갖고 있어서 자연계에서 유사성을 찾을 수는 없다. 그러나, 자연계에서 수많은 개체들이 각각의 환경에 적응하면서 오랜 기간 진화하는 과정을 시뮬레이션하는 것은 현실적으로 어려운 일이다. 제한된 개체수만으로 짧은 시간내에 가장 높은 적합도를 갖는 개체를 얻기 위한다는 목적하에서는 이러한 인위적 장치의 도입이 결과적으로는 자연계의 한 변형모델로 다루어질 수 있을 것이다. 또한, 유전공학(Genetic Engineering)의 발전으로 염색체나 생식세포들을 저장하였다가 오랜 시간이 지난 후 새로운 개체로 복원하는 일들이 단순히 먼 훗날의 일만은 아닐 것이라는 점을 생각해보면 실험실적 모델로도 간주할 수 있다.

최적화 결과

환경 설정

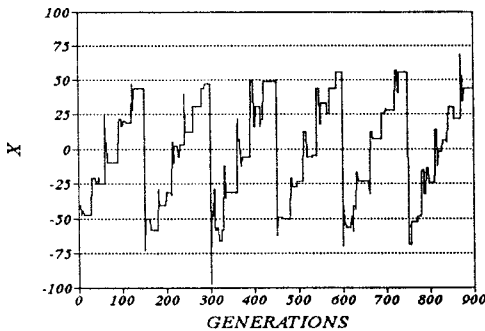
테스트에서 이진스트링(binary string)으로 구성되는 염색체 길이는 16으로 하고, 전반부 8비트는 변수 X, 후반부 8비트는 변수 Y로 할당하였다. 이진수 표현을 십진수로 변환할 때 각 변수는 -128~127의 범위를 갖게 되며, $2^{16}(=65536)$ 크기의 해공간에서 전역최적해를 찾는 문제가 된다.

SRGA에서는 감수분열과 다중선택의 방법들을 조합하여 여러 가지 탐색전략중 선택이 가능하나, [3]에서와 마찬가지로 (M-III, S-II)를 사용한다. 이 방법은 개체들의 교배횟수를 최적화하여 다양한 유전자형을 얻을 수 있고, 극부탐색에 강한 방법이다. 개체군의 크기 N은 10으로 설정하였다. SRGA가 이배체이므로 염색체의 실제개수는 20개가 된다. (M-III, S-II)의 전략을 채택할 때 한 번 교배로 하나의 개체만이 얻어지지만 교차는 두 개의 어버이 개체에서 작용하므로 총 10개의 개체발생을 위해서는 20번의 교차가 이루어지며, 돌연변이는 모든 배우자에

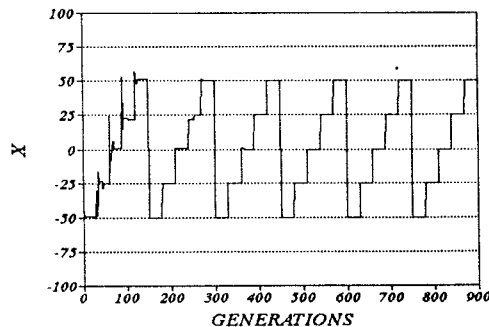
작용하므로 적은 수의 염색체만으로도 다양성을 얻을 수 있다. 캐시풀을 이용할 때에는 N-1개 즉, 9개의 개체만을 증식시키고 1개의 개체는 캐시풀로부터 얻어진다. 성능을 비교하기 위한 단순유전알고리즘(Simple Genetic Algorithm; SGA)[5]의 개체군 크기는 50으로 하였다. 이는 캐시풀 크기까지 포함한 SRGA의 염색체수와 가급적 동등하게 하기 위한 것이며, 다른 조건도 모두 동일하게 적용하였다. 개체군의 랜덤초기화는 유전자 균일교차법(Genic Uniform Distribution Method)[3]에 따라 시행하고, 교차방법은 균일교차법(uniform crossover), 돌연변이확률 P_m 은 0.125로 하였다. 돌연변이확률이 다소 크게 적용된 것은 환경변화가 이루어질 때 개체들이 다양성을 잃은 상태이고, 이런 상태에서는 교차보다는 돌연변이가 주된 진화요인이 되기 때문이다.

결과 및 분석

테스트한 결과를 그림 4, 5에 제시한다. 변수 X만을 보았을 때, SGA는 이동되는 최적점을 정확히 추적하지 못하고 있으나, 캐시풀을 이용한 SRGA는 빠른 환경적응 능력을 보이고 있다. 그림 6은 캐시풀 크기의 증가추이를 보이고 있다. 실제 5개의 최적점을 찾은 데 35개의 변이만이 필요했음을 보이고 있다. 이는 전역최적점 근방에 가파른 진동이 많고 최저점과 근사최적점이 존재하고 있어 최적점으로서의 접근적인 접근이 쉽지 않은 상태에서의 Hill Climbing의 결과로 볼 수 있다. 또한 실현수값이 정수이므로 비교적 적은 수만으로도 최적점 추적이 가능했던 것으로 해석된다. 염색체 길이가 더욱 길어지고 해공간의 분해능(resolution)이 커지면 변이가 많이 발생하므로 캐시풀에 저장되는 염색체의 수도 많아질 것이다. 실제로 본 테스트에서는 SGA와 메모리(염색체) 사용을 동등하게 하기 위해 캐시풀 크기를 30으로 제한하였다. 이 때 캐시풀 크기를 일정하게 유지하기 위해서는 저장기간이 길고 재활용율이 낮은 염색체를 FIFO(First In First Out)방식으로 제거하는 것이 필요하다. 환경변화가 많을 때에는 저장할 과거의 정보도 많아지므로 캐시풀 크기도 성능에 영향을 미칠 수 있다.



(a) SGA



(b) 캐시풀을 이용한 SRGA

그림 4. 최적점 추적성능

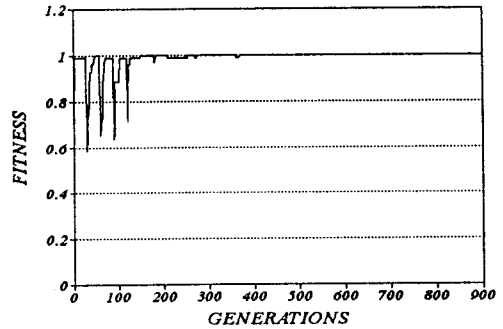


그림 5. 캐시풀을 이용한 SRGA의 적합도 복구성능

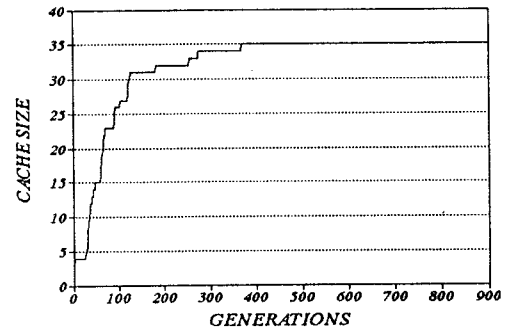


그림 6. 캐시풀 크기의 변화

결론

본 논문에서는 SRGA의 탐색성능을 활용하면서 변동문제에 대한 직접적인 해결방법으로 캐시풀(Cache Pool)의 도입을 제안하였다. 결과에서 볼 수 있는 바와 같이 캐시풀의 사용으로 최적점 추적성능이 크게 향상되었음을 알 수 있다.

캐시풀의 효율성은 두가지로 요약할 수 있다. 먼저 국부탐색에 있어서 수렴성을 향상시킨다. 즉, 환경변화가 없는 기간에는 캐시풀에 저장된 최우량개체가 개체군에 되먹임(feedback)됨으로써 전세대의 최우량 개체가 다음 세대로 반드시 존속되게 되므로 엘리트전략(elitism)으로서 역할한다. 엘리트전략은 최적해로의 수렴성을 보장하는데 중요한 방법이다. 두 번째는 원래의 목적인 장기간 기억(long term memory)의 역할이다. 최적화 과정에서 환경변화가 이루어질 경우 과거 거처왔던 탐색해들 중 이동된 최적점에 가장 가까운 또는 이에 준하는 염색체를 캐시풀에서 추출하게 하므로써 과거정보를 재활용한다는 의미에서 효율성을 제고할 수 있다.

참고문헌

- [1] Robert E. Smith and David E. Goldberg, "Diploidy and Dominance in Artificial Genetic Search," *Complex System*, 6, pp.251~285, 1992.
- [2] F. Greene, "A Method for Utilizing Diploid/Dominance in Genetic Search," *Proceedings of the First IEEE Conference on Evolution Computation*, IEEE Press, pp.439~444, 1994.
- [3] 류근배, 최영준, 김창업의, "유성생식 유전알고리즘 : 다중선택과 이배성이 탐색성능에 미치는 영향," 대한전기학회 하계학술대회 논문집 B, pp.1006~1010, 1995.
- [4] 류근배, 최영준, 김창업의, "유전알고리즘을 이용한 선형유도전동기의 다중목적 최적설계," 대한전기학회 하계학술대회 논문집 A, pp.165~167, 1994.
- [5] David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.