

유성생식 유전알고리즘 : 다중선택과 이배성이 탐색성능에 미치는 영향

'류 균 배' '최 영 준' '김 창업' '이 학 성' '정 창 기'[†]

효성중공업(주) 기술연구소^{*} 한국전력공사 기술연구원[†]

Sexual Reproduction Genetic Algorithms : The Effects of Multi-Selection & Diploidy on Search Performances

'K.B. RYU' 'Y.J. CHOI' 'C.E. KIM' 'H.S. LEE' 'C.K. JUNG'[†]

R&D Institute, Hyosung Industries Co., Ltd.^{*} Research Center, Korea Electric Power Co.[†]

Abstract

This paper describes Sexual Reproduction Genetic Algorithm(SRGA) for function optimization. In SRGA, each individual utilize a diploid chromosome structure. Sex cells(gametes) are produced through artificial meiosis in which crossover and mutation occur. The proposed method has two selection operators, one, individual selection which selects the individual to fertilize, and the other, gamete selection which makes zygote for offspring production. We consider the effects of multi-selection and diploidy on search performance. SRGA improves local and global search(exploitation and exploration) and show optimum tracking performance in nonstationary environments. Gray coding is incorporated to transforming the search space and Genic uniform distribution method is proposed to alleviate the problem of premature convergence.

서 론

유전알고리즘(Genetic Algorithm; GA)은 자연의 유전학(natural genetics)과 자연선택(natural selection)의 원리[1]에 근거한 최적점 탐색방법(optimum search method)으로 모의진화(simulated evolution)의 한 방법이다. 60년대부터 지속적으로 연구가 진행되어온 GA는 확률적인 방법과 개체간의 정보교환을 통해 탐색공간(search space)을 병렬로 조사해나감으로써 주위환경(environment)에 알맞는 가장 적절한 해를 얻고자하는 방법이다. GA가 기존의 최적점 탐색방법과 다른 점은 기존방법이 탐색공간의 한 점에서 그 다음 한점으로 순차적으로 해를 구해 진행해가는 국부적 탐색(local search)인대 비해 유전알고리즘은 여러 점을 동시에 병렬로 탐색하는 전역적 탐색(global search)을 행한다는 점이다. 따라서 국부수렴(local convergence)할 가능성이 적고, 파라미터 자체가 아닌 유전자로 코딩(coding)하여 탐색하므로 탐색공간에 대한 해석적 제한이 없으며, 목적함수의 도함수와 같은 정보를 요하지 않으므로 비선형 최적화(nonlinear optimization)에 특히 유용하다.

모의진화를 이용한 최적화방법은 GA외에도 진화전략(evolution strategy; ES)이나 진화적 프로그래밍(evolutionary programming; EP)이 있다.[2] 자연의 진화원리를 적용한다는 점은 이러한 진화적 최적화방법에 모두 동일하게 도입된다. 그러나, GA에는 ES나 EP와 구별되는 특징이 있다. 그것은 ES나 EP는 선택과 돌연변이에 의해 무성생식(asexual reproduction)에 의해 진화하는 모델인 반면, GA는 염색체간의 유전정보를 교환하는 교차(crossover)가 중요한 역할을 수행하며, 이는 유성생식(sexual reproduction)의 원리에 기반한다는 점이다. 또한, GA는 유전자단위로 정보를 세분화하여 정보를 다루므로 생물학적 원리에 좀 더 접근된 방법으로 볼 수 있다. 이러한 점이 GA가 공학적인 최적화 문제에만 한정하지 않고 넓은 응용범위를 갖게 한다. 이러한 점이 주목되어, GA는 인공생명(Artificial Life)[3,4]의 핵심적 기술로 연구되고 있으며, 유전공학(Genetic Engineering)[5]의 발달 및 사회생물학(Sociobiology)[6,7]의

확산과 더불어 이 분야의 연구가 더욱 많은 관심을 끌게 될 것으로 보인다.

본 논문에서 설명하는 유성생식 유전알고리즘(Sexual Reproduction Genetic Algorithm; SRGA)은 자연계의 생식과정에 기초한 모델로서 반수성(haploidy)을 갖는 기존 방식의 단일염색체 개체(one-chromosome individual)방법을 개선하여 생태계와 유사한 이체제(diploid) 개체를 도입한다. 이러한 이체 형태는 최적화단계에서 이중적 정보저장능력으로 환경변화요인에 우수한 적응성을 보여준다. 또한, 개체의 생식을 위해서 기존의 GA에서 도입되지 않은 감수분열(meiosis)을 인공적으로 모델링한다. 감수분열(meiosis)을 통해 생식세포(sex cell; gamete)가 형성되고, 이 과정에서 염색체간의 교차와 돌연변이(mutation)를 통해 다양한 형태의 형질발생이 이루어진다. 개체내부에 생산된 생식세포포자손(offspring)으로 이어질 배우자는 경쟁원리 등을 도입하여 선택할 수 있으며, 이러한 배우자 선택(gamete selection)과정이 진화연산자로 추가된다. 따라서, SRGA에서는 교배(mating)할 개체를 선택하는 개체선택(individual selection)과 생식과정에서의 배우자선택이 다중선택(multi-selection)의 구조를 이루게 된다. 이배성과 다중선택의 구조에 의해 최적화 성능향상과 다양한 용용이 가능하다. 본 논문은 이배성과 다중선택이 탐색성능에 미치는 영향을 살펴보기 위해 함수최적화(function optimization)에 적용하여 그 결과를 보인다. 실험결과는 GA의 일반적 문제인 전역탐색과 국부탐색에 있어 개선된 결과를 보인다. 특히, 최적점이 이동하는 변동문제(nonstationary problem)에 적용할 때, 기존의 GA는 최적점탐색이 어려우나 SRGA는 우수한 추적성능을 보인다. 또한, 조기수렴(premature convergence)과 기만문제(deception problem)에 대한 대책으로, 탐색공간 변환을 위해 그레이코드(gray code)를 도입하고, 유전자 균일분포법(Genic Uniform Distribution method)을 제안한다.

단순유전알고리즘

Holland가 자연의 유전·진화이론을 최적화에 도입한 단순유전알고리즘(Simple Genetic Algorithm; SGA)[8]의 최적화 개념은 선택(selection)과 유전자 교차(crossover), 돌연변이(mutation)의 작용에 기초한다. SGA에 의한 유전적 진화는 자연의 경우와 마찬가지로 염색체의 변화에 의해 진행된다. 염색체는 0과 1의 이진스트링(binary string)으로 부호화되며, 이 부호값에 따라 개체가 환경에 적응하는 정도가 결정된다. 선택은 환경에 더욱 잘 적응하는 개체를 다음 세대에 더 많은 확률로 재생산 되도록 하는 과정이다. 선택된 어버이 개체(또는 염색체)는 다른 개체와 유전정보를 교환하여 자손을 만드는 과정을 거친다. 이것이 교차이다. 돌연변이는 유전자의 돌연한 형태변화에 의한 새로운 형질이 발생할 근거로 작용한다. 여러 개체들이 임재적인 해로서 개체군(population)을 이루게 되고, 이러한 과정이 개체군에서 반복되어 점차 환경에 적합한 개체만이 살아남게 되고, 결국 적자생존(the survival of the fittest)방식으로 최적의 해가 일어지게 된다. 이러한 원리는 이후 GA의 개선 연구에 크게 변하지 않고 적용되면서, GA의 탐색성능을 향상시키기 위한 시도가 꾸준히 지속되었다.

GA를 최적화방법에서 문제점으로 지적되어온 것은 크게 두 가지로 요약된다. 첫번째는 해공간(solution space)를 충분히 탐색하여 전역적 최적해(globally optimal solution)을 얻고자 하는 문제(exploration)이며, 두번째는 정확한 최적해를 얻고자 하는 국부탐색의 문제(exploitation)이다. 이러한 문제는 GA가 병렬탐색의 강점을 가짐에도 불구하고, 개체군의 불균일한 초기화와 상대적으로 높은 적합도(fitness)를 갖는 개체의 금속한 확산으로 인해 국부최적해(locally optimal solution)로 초기수렴할 가능성이 있기 때문이다. 또한, 2진수로 코딩함에 따라 제기되는 기만문제(deception problem)도 전역적 해를 찾는데 어려움을 준다. 더구나 GA의 해석모델인 스키마(schema)가 진화연산자인 교차나 돌연변이에 의해 상실될 수 있다는 점도 역설적으로 GA가 갖는 문제점으로 지적된다.

유성생식 유전알고리즘

생물학적 기초

생물의 유전정보는 그 생물의 세세포(somatic cell) 속에 있는 염색체(chromosome) 속에 존재하며, 이는 염색체안에 개체의 형질(trait)을 결정하는 유전자(gene)가 존재하기 때문이다. 이를 염색체들은 모양과 크기가 비슷한 염색체끼리 쌍을 이루는 이배성(diploidy)을 갖는다. 이와 같은 한 쌍의 염색체를 상동염색체(homologous chromosomes)라 하는데 이러한 구조는 하나의 형질에 대한 유전정보가 2중적임을 알려주는 사실이다. 19세기 중반의 Mendel과 그 뒤의 유전학자들의 연구결과에 따르면, 상동염색체의 같은 좌위(locus)상에 자리하고 있는 대립유전자는 한쌍의 대립형질(alleles)을 나타내며, 이중에서 한가지 형질만이 발현된다는 사실을 발견하였고 이로부터 형질에 우열관계가 있음이 알려졌다. 또한 유전형질의 외적발현과 관련한 우성(dominant)인자의 유전과 염색체내에 잠재해 있는 열성(recessive)인자의 유전은 각각 독립적으로 이루어진다는 사실도 관찰되었다. 개체의 형질발현과 적절적인 관계가 없는 열성인자가 후손에 와서 발현될 가능성이 일정비율 존재한다는 사실로부터 유전정보는 장기간기억(long term memory)의 효과를 가지게 되며, 생태환경의 변화에 따른 생존여부와 관련하여 선별의 다양성이 주어진다. 실제로 유전자의 우열관계 자체가 진화한다는 사실이 생물학자들에 의해 증명되어 왔으며, 이러한 사실은 형질의 우열관계가 개체에게 주어진 환경변화와 밀접한 관계가 있다는 것을 의미한다.

자연생태계에서 이배성을 갖는 종족들이 번식을 하기 위한 유용한 생식방식이 바로 유성생식(sexual reproduction)이다. 유성생식을 통해 자손의 염색체 구성은 아버지로부터 각각 50% 비율로 이루어지며 유전자의 새로운 조합이 생겨남으로써 새로운 형질이 나타난다. 새로운 형질의 발생으로 환경변화에 대하여 적응능력이 우수한 자손이 나타날 수 있다. 유성생식과정에서 아버지로부터 염색체의 수가 증가하지 않고 일정하게 유지되는 것은 생식세포(sexual cell; gamete) 발생을 위한 감수분열(meiosis)과정에서 상동염색체가 분리되어 염색체의 수가 반감(n) 되기 때문인데, 그 결과 발생된 자웅 양쪽의 생식세포가 교배(mating)에 의한 수정(fertilization)과정을 통해 이배성의 단일세포(2n) 즉 접합체(zygote)를 이룬다. 이 접합체로부터 새로운 개체가 형성된다.[1,9]

개체의 구성 ~ 이배성

일반적으로 유전알고리즘에서 염색체상의 유전자(gene)는 이진값을 갖는 비트(bit)단위로 구성되어 1 또는 0의 값을 갖게되며, 염색체는 유한길이의 2진스트링(binary string)으로 표현된다. 이러한 이진값들의 조합으로 형질(trait)이 결정된다.



그림 1. 개체의 구성

SRGA에서는 강인한 최적점 탐색성능을 얻기 위해 그림 1과 같이 한 개체에 염색체가 쌍으로 존재하는 이배체(diploid)구성을 갖는다.

적합도와 우성결정

개체와 환경의 상호관계는 그 개체가 갖는 적합도(fitness)로 표현되는데, 개체가 다음 세대(generation)로 자신의 유전자형(genotype)을 존속시키는 것과 깊은 관련을 맺는다. 이배체의 경우 개체의 우성형질만이 발현되므로, 우성과 열성을 결정하는 우성결정(mapping)이 필요하다. 가장 간단한 방법으로 두 스트링의 적합도를 각각 구하고 그 값이 상대적으로 큰 스트링을 우성으로 결정한다. 따라서, 우성인 스트링은 그 개체의 형질들을 나타내게 되며, 개체의 적합도도 우성의 적합도를 따른다. A, B 두개의 스트링을 갖는 i번째 개체의 적합도 F는 다음과 같다.

$$F(A, B, i) = \text{MAX} [F(A, i), F(B, i)] \quad (1)$$

환경이 변화하지 않으면 염색체쌍에 대한 우열관계는 일정하나, 변화하는 환경, 즉 적합도값이 변화할 때에는 우열관계도 그에 따라 변할 수 있다. 따라서, 이전 세대에서 열성인자로 잡제해 있었던 형질도 상대적 우성이 되어 환경변화에 대처하게 되고, 종(species)의 적응성이 유지될 수 있다.

감수분열(meiosis)

생태계의 감수분열 단계에서 상동염색체는 각각 2가의 염색분체(chromatid)가 되었다가 이중 하나씩이 일부 유전자를 서로 교환하는 교차현상이 발생한다. 교차를 통해 유전자가 교환되는 과정은 유한갯수의 염색체를 갖는 개체가 생식을 통해 유전적 다양성(genetic diversity)을 얻는 것과 관련하여 중요한 현상이다. 탐색과정에서 염색체간의 국부적 동작(local action)으로 어버이보다 더욱 나은 전역적 성능(global performance)을 얻을 수 있는 점이 유전적 탐색방법의 장점중의 하나이며, 교차는 그 핵심적 역할을 수행한다. 유전자의 돌연변이 새로운 형질이 발생하게 되는 현상이 돌연변이(mutation)이다. 돌연변이는 생물의 진화에서 교차와 마찬가지로 중요한 요인이다. 그러나, 한편으로 돌연변이는 환경에 대한 적응성과는 무관하게 랜덤하게 발생하므로 그 발생률을 너무 크게 하면 중요한 형질을 잃을 가능성이 있다.

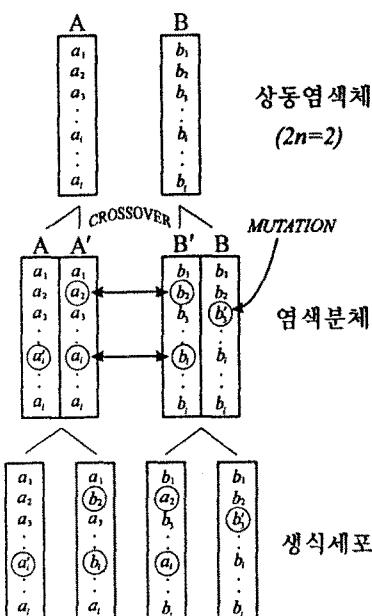


그림 2. 인공감수분열

그림 2는 스트링 길이가 1인 한쌍의 상동염색체로부터 갑수분열을 교차 및 돌연변이가 일어나고 생식세포가 발생하는 과정을 나타낸 것이다. 여기서, a_i, b_i ($= \{0, 1\}$)는 각각 A, B 스트링의 i번째 유전자를 가리키며, 염색분체 A'와 B'의 2번 째와 i번째 비트에서 교차가 발생하고, A의 i번째 비트와 B의 3 번째 비트에서 돌연변이가 발생할 경우를 예시하고 있다. a'_i, b'_i 는 각각 a_i, b_i 의 보수이다.

SGA에서는 선택, 교차, 돌연변이가 별도의 과정으로 분리되어 개체군레벨에서 이루어진다. 그러므로, 반드시 교차와 돌연변이는 다음세대에 영향을 미친다. 그러나, SRGA에서는 갑수분열의 도입으로 교차와 돌연변이가 개체레벨로 그 내부에서 진행되며, 발생하는 반수성의 배우자는 모두 또는 일부만 수정된다. 그러므로, 개체선택→갑수분열→배우자선택→수정으로 이어지는 과정에서 교차와 돌연변이는 갑수분열의 하위과정으로 온다. 배우자선택이 도입되므로 변이의 결과가 반드시 자손으로 전달되지는 않는다. 즉, 해로운 변이가 자손에서 나타날 가능성이 사전단원 수 있는 것이다.

SRGA에서는 교차가 반드시 발생하더라도 스키마가 손상되지 않는 여벌의 염색체(A, B)가 있으므로, 교차확률 $P_c = 1$ 의 경우도 가능하다. 또한, 돌연변이는 부분적으로(A와 B 또는 A'와 B'), 또는 배우자 모두에 적용 가능하다. 확률적으로 교차나 돌연변이가 일어나지 않은 경우에는 어버이의 형질이 자손에 그대로 전달될 수 있다. 교차확률과 돌연변이가 발생하는 염색체를 지정하는 각각의 경우는 표 1과 같다.

표 1. 갑수분열 전략

Strategies	Crossover	Mutation	
		A, B	A', B'
M- I	$P_c = 1$	○	×
M- II	$P_c = 1$	×	○
M- III	$P_c = 1$	○	○
M- IV	$P_c < 1$	○	×
M- V	$P_c < 1$	×	○
M- VI	$P_c < 1$	○	○

표에서 다양성이 극대가 되는 M-III가 전역탐색에 이로우나 형질상실 가능성이 가장 크므로 국부탐색이 어려우며, 반대로 다양한 변이가 어려운 M-V의 경우가 상대적으로 국부탐색에서 이로우나 전역탐색에 있어 불리하다는 것을 알 수 있다.

다중선택(Multi-Selection)

생태계에서는 여러 개체가 모여 하나의 개체군을 이루고 교배에 의한 생식과정을 통해 자손을 증식하고 여러 세대에 걸쳐 종족을 보존한다. 개체군에 주어지는 생태환경의 변화에 따라 적응성이 높은 개체만이 적자생존 방식으로 선택되어 자손을 번식시킬 수 있다. 일반적으로 GA에서는 개체의 적합도에 비례하여 그 개체의 교배할 수 있는 확률이 주어진다. 가장 많이 사용되는 확률적 샘플링(Stochastic Sampling)방법은 weighted roulette wheel[8]을 이용하는 방법이다. SRGA에서는 1회의 교배에 반드시 2개의 개체가 선택되어야 하며, 배우자선택은 모두를 선택(4개)하는 방법과 적합도 값에 따라, 반(2개) 또는 엘리트(1개)를 선택하는 방법 등 표 2와 같이 여러 가지 조합이 가능하다.

표 2. 다중선택 전략

Strategies	Individual Selection	Gamete Selection	Offsprings/Mating
S- I	All	Half(2)	2 ~ 4
S- II	Stochastic	Elite(1)	1
S- III	Stochastic	Half(2)	2 ~ 4
S- IV	Stochastic	All(4)	4 ~ 16

선택된 배우자는 짝을 이루어 수정(fertilization)하면, 새로운 자손이 발생한다. S-I, S-III, S-IV와 같이 복수의 배우자선택전략일 때는 두 개체 배우자의 수정방법에 따라 교배 1회당 2 ~ 16의 자손수 발생하는 경우의 수가 가능하다.

탐색전략(Search Strategies)

S-I의 경우는 확률적 선택을 배제하고 모든 개체가 1번씩 군집하게 교배가 되며, M-II나 M-V와 조합할 경우 어버이세대의 최우량 개체(Best Individual)가 자손세대로의 보존이 보장되므로 별도의 엘리트전략(elitism)이 필요 없다. 그 외의 경우는 지속적인 탐색성능의 유지하고 최적해로의 수렴을 보장하기 위해서는 엘리트전략의 채용이 바람직하다. [11]

S-II의 경우는 1회 교배로 가장 적합도가 큰 배우자를 각각 선택하여 하나의 개체가 생성되므로 국부탐색에 가장 효과적이다. 그러나, 교배 1회당 자손수가 1이므로 교배횟수가 많아지는 단점이 있다.

이와 같이 갑수분열과 다중선택의 방법들을 조합하여 여러 가지 탐색전략이 가능하다. 일반적으로, 국부탐색을 강화하면 전역탐색이 약화되고, 반대로 전역탐색을 강화하면 국부탐색이 약화되는 trade-off가 있으므로 적절한 탐색전략을 선택할 필요가 있다. [9]에서는 (M-II, S-IV)를 조합하여 기기설계에 응용하였으며, [10]에서는 (M-III, S-II)를 조합하고 각각의 선택에 다른 목적함수(objective function)를 적용하는 전략으로 다중목적 최적화(Multiobjective Optimization)를 구현하였다.

SRGA를 이용하여 단일목적 최적화에 적용할 경우, 최적화 절차는 그림 3과 같다.

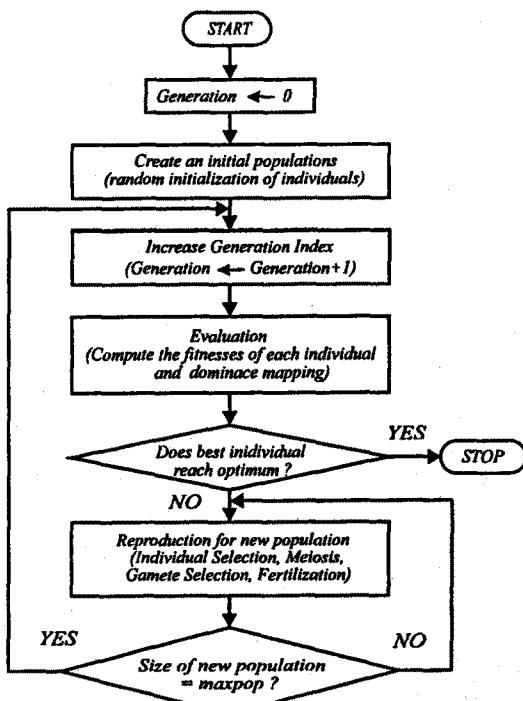


그림 3. 최적화 절차

SRGA를 이용한 함수최적화

테스트 함수

GA의 탐색성능을 평가하기 위한 여러 가지 테스트함수가 제기되어 왔다. SRGA의 일반적인 성능을 보기 위해 FG함수에 적

용한다. F6는 대표적인 다극점 함수로 GA의 테스트에 많이 쓰인다.[12] 변수가 2개일 경우 F6는 다음과 같다.

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

여기서, (x_0, y_0) 은 최대값을 갖는 X, Y 변수의 최적점좌표이다. 그림 4는 Y변수를 0으로 하고 2차원상으로 F6함수를 나타낸 것이다. 3차원의 경우는 그림 5와 같이 표현된다.

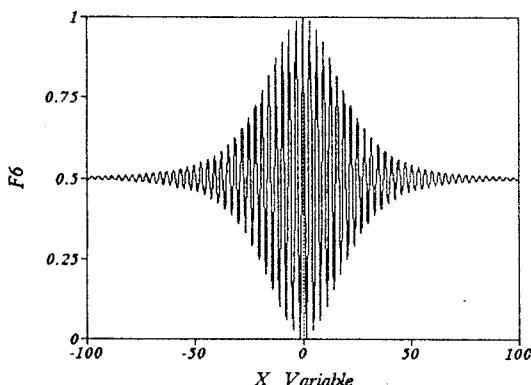


그림 4. F6 함수 (2차원 표현)

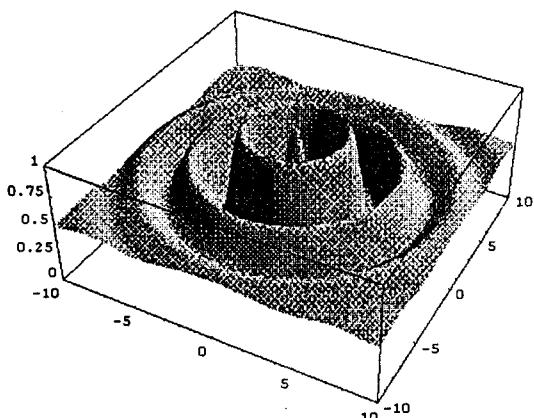


그림 5. F6 합수 (3차원 표현)

그럼에서 볼 수 있는 바와 같이 FG함수에는 1개의 전역최적점이 있으며, 이를 근사최적해가 겹친다. 둘러싸고 있다. 또한, 전역해와 근사해 사이에는 최저값이 놓여 있다. 하나의 최적해를 무수한 근사해가 둘러싸고 있고 가파른 진동이 많으로 최저값을 찾기가 쉽지 않을을 알 수 있다.

SRGa의 테스트에서 염색체 길이는 16으로 하고, 전반부 8비트는 변수 X, 후반부 8비트는 변수 Y로 할당하였다. 이진수 표현을 실현수로 변환할 때 각 변수는 -128~127의 범위를 갖도록 하였다. 이러한 할당으로 $2^{16} (=65536)$ 크기의 해공간에서 전역 최적해를 찾는 문제가 된다.

탐색공간의 변화 - 그레이코드

GA의 약점이 되는 것 중에 기만문제(deception problem)가 있다. 기만문제는 GA의 교차작용으로는 전역최적해에 도달하기 어려운 문제를 말한다. 가장 간단한 예로 최소기만문제(minimal deception problem; MDP)[13]의 예를 들 수 있다. 전역최적해가 11이고 00이 국부최적해라 할 때, 개체군이 00으로 조기수렴

하고 01, 10은 적합도 같아 낫다면, 들연변이에 의하지 않고서는 교차로는 11에 도달하기 어려울 것이다. 이와 유사한 문제가 위의 F6테스트에도 발생할 수 있다. 위의 경우처럼 8비트 스트링을 이진수로 엔코딩할 경우, 최적해 (0,0)을 갖는 유전자형은 (1000000010000000)이 된다. 그러나 초기개체군의 웨임초기화가 불균일하여 ($1*****1*****$)의 스끼마가 부족하고, ($0*****0*****$) 또는 ($0*****1*****$)의 스끼마가 우세하여 최적화 진행으로 급속히 확산되며, 나중에는 유전자형 (0111111101111111)이 되거나 (0111111100000000)으로 수렴하게 된다. 이것은 표현형으로는 (-1, -1)이나 (-1, 0)이 되는 경우로 결국 최적해로 수렴하지 못한다. 이와 같은 문제는 GA의 국부탐색성능을 약화시키는 문제가 된다. 이는 이진코드의 구조적 문제로서 이진스트링간의 해밍거리(Hamming distance)가 일정치 않아 해밍절벽(Hamming cliff)이 많기 때문이다. 이에 대한 대안으로 그레이코드(gray code)의 사용을 도입한다. 그레이코드는 인접스트링간 해밍거리가 1로 고정되어, 해밍절벽으로 인한 국부최적점의 발생을 억제할 수 있다. 2번수 F6합수의 경우, 그레이코드를 썼을 때와 이진코드를 쓸 때의 해밍공간(Hamming Space)상의 국부최적점의 수는 25 : 361이며, 이는 번수의 증가에 따라 지수함수적으로 더욱 큰 차이가 나는 것으로 알려졌다.[14] 따라서, 그레이코드의 사용은 이진수 표현의 단점을 줄이는 데 효과적이라고 볼 수 있다.

유전자 균일분포법

랜덤 초기화의 불균일성을 억제하여 전역 탐색을 강화하기 위해 유전자 군일 분포법(Genic Uniform Distribution Method)을 제안한다. 이 방법은 전체 개체군중 반수의 개체(염색체)를 랜덤 한 이진값으로 먼저 초기화한 후, 이 이진값들의 보수(complement)로 나머지 반수의 개체군을 초기화한다. 이 방법으로 개체군내에서 각 유전자 좌위(locus)상에서 유전자 0과 1의 개체군내 분포는 1:1로 균일하게 된다. 따라서, 제한된 개체군 크기에서 초기 개체군의 편향된 분포로 인한 국부최적점으로의 조기 수렴을 억제하는 효과를 기대할 수 있다.

교차와 물연변이

교차가 일어나는 교차점 즉 키아스마(chiasma)의 개수에 따라 일점교차(one-point crossover), 다점교차(multi-point crossover), 그리고 균일교차(uniform crossover) 등 여러 가지 일반적인 방법이 적용될 수 있다. 본 텍스트에서는 유전자 하나하나에 대해 일정 확률로 교차여부를 결정하는 균일교차를 이용한다. 이 방법은 복수의 유전자가 한 염색체 좌우상에서 갖는 상호연관성이 파괴되거나 쉬우나, 다양한 표현형(phenotype)을 얻을 수 있는 장점을 있고 구현이 쉬우므로 일반적인 디버깅 최적화 문제에서 많이 쓰이는 방법이다.

교차이후 돌연변이를 행한다. 돌연변이 확률 P_m 은 0.0625로 하였다. 여러 가지 돌연변이 방법 중 역위(inversion)에 의한 방법도 사용한다. 역위는 염색체의 일정구간을 확률적으로 선정, 사외비트와 차외비트의 값을 서로 염수으로 교환하는 방법이다.

최저점 범위에 의한 환경변화

SRGA의 개체군의 크기는 30으로 설정하였고, 탐색전략은 (M-III, S-II)를 선택하였다. 성능을 비교하기 위해 SGA는 개체군 크기를 60으로 하였다. 이는 아래에 모델인 SRGA와 염색체 수를 동등하게 하기 위한 것이며, 교차 및 몰연번이 방법도 모두 같은 방식을 사용한다. 염색체의 초기값도 같은 값을 갖도록 하였다.

이제 성에 의한 환경적응과 다중탐색의 신속한 탐색성능을 테스트하기 위해 FG함수의 최적점((x_0, y_0))를 주기적으로 변화시켜 환경변화를 유도한다. 테스트에서는 최적점을 $(-10, -10)$ 과 $(10, 10)$ 로 30세대마다 급격히 절환하여 알고리즘의 적응성을 비교하였다.

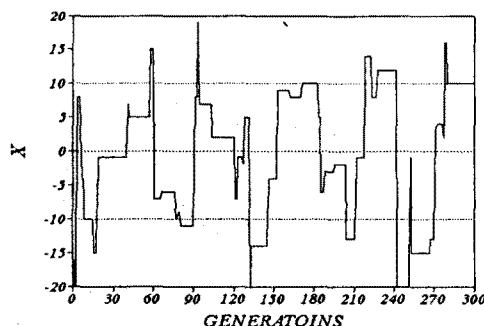
테스트한 결과를 그림 6, 7에 제시한다. 변수 X만을 보았을 때, SGA는 최적점 이동에 적응을 못하는 현상을 보이고 있으나, SRGA는 좋은 최적점 추적 성능과 적합도 복구능력을 보인다. 이는 이해해보기 위한 형질순서의 방지와 다중선택에 의한 우수한 탐색성능의 결과로 볼 수 있다.

결 론

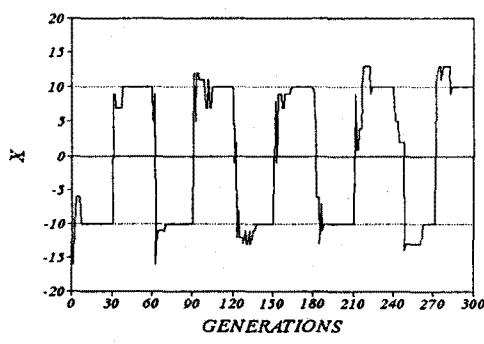
유성생식 유전알고리즘(SRGA)은 이배체 개체의 도입과 다중선택의 구조로 다양한 탐색전략을 가능하게 한다. 교배가 이루어질 개체를 선택하는 개체선택(individual selection)과 감수분율후 수정이 이루어질 배우자의 선택(gamete)의 다중선택은 각각의 선택전략을 조합하여 여러 가지 탐색전략이 가능하다. 특히, 최적점이 이동하는 경우에 우수한 최적점 추적성능을 보인다. 또한, 2진스트링의 단점을 보완하기 위해 그레이코드를 채택하고, 유전자 군일분포법을 제안하여 사용하였다. SRGA는 기존 방법에 대한 구조적 개선이므로, GA의 실현수 유전자 방법[15]에도 그대로 사용될 수 있다. 후속 연구과제로 탐색전략에 따른 성능비교 및 장단점 분석이 필요하며, 적절한 전략선택에 따라 더 나은 성능을 기대할 수 있을 것이다.

참고문헌

- [1] Francisco J. Ayala and John A. Kiger, Jr., *Modern Genetics*, Benjamin/Cummings, 1980.
- [2] David B. Fogel, "An Introduction to Simulated Evolution Optimization," *IEEE Trans. Neural Networks*, special issue on Evolutionary Computation, Vol.5, No.1, pp.3~14, 1994.
- [3] Steven Levy, *Artificial Life*, Pantheon Books, 1992.
- [4] 이인식, *사람과 컴퓨터*, 카치, 1992.
- [5] Robert Cook-Deegan, *THE GENE WARS: Science, Politics and the Human Genome*, W.W.Norton & Company, 1994.
- [6] Edward O. Wilson, *Sociobiology: The new synthesis*, Harvard Univ. Press, 1975.
- [7] Richard Dawkins, *The Selfish Gene*, Oxford Univ. Press, 1976.
- [8] David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.
- [9] 류근배, 최영준, 김창업 외, "유전알고리즘을 이용한 편측식 선형유도전동기의 최적설계," 대한전기학회 하계학술대회 논문집 B, pp.923~928, 1993.
- [10] 류근배, 최영준, 김창업 외, "유전알고리즘을 이용한 선형유도전동기의 다중목적 최적설계," 대한전기학회 하계학술대회 논문집 A, 165~167, 1994.
- [11] Günter Rudolph, "Convergence Analysis of Canonical Genetic Algorithm," *IEEE Trans. Neural Networks*, special issue on EP, Vol.5, No.1, pp.96~101, 1994.
- [12] L. Davis, Ed., *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, 1991.
- [13] 北野宏明編, *遺傳的アルゴリズム*, 菲葉圖書, 1993.
- [14] Keith E. Mathias and L. Darrell Whitley, "Transforming the Search Space with Gray Coding," *Proceedings of the First IEEE Conference on Evolution Computation*, IEEE Press, pp.513~518, 1994.
- [15] Zbigniew Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, 1994.

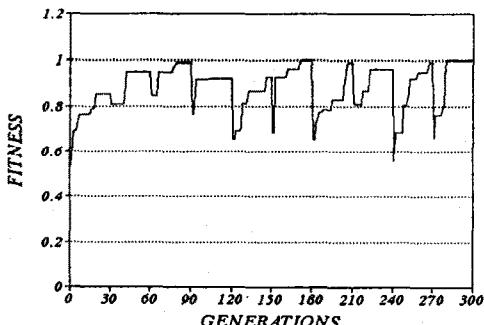


(a) SGA (Haploidy & Single Selection)

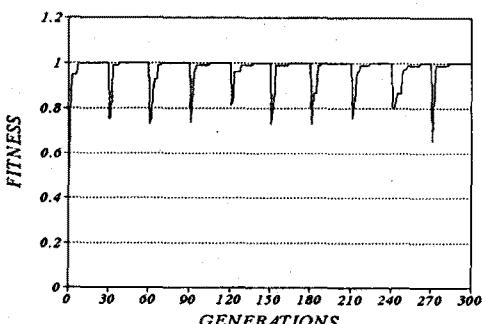


(b) SRGA (Diploidy & Multi-Selection)

그림 6. 최적점 추적성능



(a) SGA (Haploid & Single Selection)



(b) SRGA (Diploidy & Multi-Selection)

그림 7. 적합도 복구성능