

HFC 기반 유전자알고리즘에 관한 연구

A study on HFC-based GA

김길성¹, 최정내², 오성권³, 김현기⁴

^{1,2,3,4} 경기도 화성시 수원대학교 전기공학과
E-mail: ohsk@suwon.ac.kr

요 약

본 논문에서는 계층적 공정 경쟁 개념을 병렬 유전자 알고리즘에 적용하여 계층적 공정 경쟁 기반 병렬유전자 알고리즘 (Hierarchical Fair Competition Genetic Algorithm: HFCGA)을 구현하였을 뿐만 아니라 실수코딩 유전자 알고리즘(Real-Coded Genetic Algorithm: RCGA)에서 좋은 성능을 갖는 산술교배(Arithmetic crossover), 수정된 단순교배(modified simple crossover) 그리고 UNDX(unimodal normal distribution crossover)등의 다양한 교배연산자들을 적용, 분석함으로써 개선된 병렬 유전자 알고리즘을 제안하였다. UNDX연산자는 다수의 부모(multiple parents)를 이용하여 부모들의 기하학적 중심(geometric center)에 근접하게 정규분포를 이루며 생성된다. 본 논문은 UNDX를 이용한 HFCGA모델을 구현하고 함수파라미터 최적화 문제에 많이 쓰이는 함수들에 적용시킴으로써 그 성능의 우수성을 증명 한다.

Key Words : Parallel Genetic Algorithm (PGA), Hierarchical Fair Competition-based Genetic Algorithm(HFCGA), migration, crossover, parameter optimization

1. 서 론

단순 유전자 알고리즘(simple genetic algorithm: SGA)은 유전학과 자연환경에서 일어나는 진화원리를 흉내 낸 적응 탐색법으로써 1975년 Holland에 의해 개발되었으며[1], 초기의 알고리즘임에도 불구하고 해를 찾는 능력이 강인하여 아직도 많은 최적화 문제에 응용되고 있다. SGA를 개선하고 변형한 여러 유전알고리즘이 끊임없이 연구되어지고 있지만 이들이 채용하고 있는 기본적인 메커니즘은 거의 유사하며 해공간의 표현방식과 선택, 교배, 돌연변이 등의 유전 연산자에 대한 연구위주로 진행되어져 왔다[2-5].

GA는 염색체의 표현방식에 따라 이진코딩(binary coded GA; BCGA), 기호코딩(symbolic coded GA; SCGA) 그리고 실수코딩 유전자알고리즘(real coded GA ; RCGA)이 있다. 실세계의 문제들은 고정밀도의 해를 요구하고, 해에 대한 사전지식이 결여되거나 복잡한 제약 조건이 존재하는 경우가 많이 있다.

GA에서 가장 고려되어야 할 요소는 개체들의 다양성 유지이다. GA는 전역해를 구할 수 있다는 강점으로 인하여 최근 많은 응용분야에 적용되고 있지만 실질적으로는 높은 선택압으로 인하여 초기 생성된 개체 중에서 열등한 개체는 초기에 도태되고 우수한 개체들을 통하여 진화하게 되며 세대가 진화 할수록 대부분의 개체는 비슷한 값을 갖게 되어 다양성이 결여되고 모든 개체들이 일정지역 안으로 수렴하게 되는 조기수렴 문제를 내포하고 있다.

병렬 유전자 알고리즘(parallel genetic algorithm; PGA)는 이러한 문제를 해결하기 위한 방법으로 제시되었으며 많은 모델들이 연구되어져왔다[6-8]. 본 논문에서는 유전자 프로그램(genetic program)에서 제시되었던 계층적 공정 경쟁 개념을 병렬 유전자 알고리즘에 적용하여 계층적 공정 경쟁 기반 병렬유전자 알고리즘 (HFCGA)을 제시하였을 뿐만 아니라 RCGA에서 좋은 성능을 갖는 산술교배(Arithmetic crossover:AX), 수정된 단순교배(modified simple crossover:MSX) 그리고

UNDX(unimodal normal distribution crossover:UNDX)등의 다양한 교배연산자들을 적용 분석함으로써 개선된 유전자 알고리즘을 제안하였다.

2. 실수코딩 유전자 알고리즘과 교배 연산자의 개관

RCGA는 하나의 염색체가 실수값을 갖도록 코딩하는 방식으로 변수와 유전자간에 일대일 대응하므로 프로그램이 간편해지고 BCGA에서 요구되는 부호화와 복호화 과정이 필요 없게 되어 탐색 속도를 높일 수 있고 더 적은 메모리 공간을 요구한다. 또한 같은 고정밀도의 해를 구할 수 있으며 해에 관한 사전지식이 없는 경우에는 매우 큰 탐색 범위를 지정하는 것도 가능하기 때문에 최근 많이 사용되고 있다. 그렇지만 재생산연산자는 기존의 BCGA와 동일하게 적용할 수 있지만 교배와 돌연변이 연산자는 실수코딩방식에 적합하도록 고안되어야 한다. RCGA에서 교배연산자는 다른 연산자에 비해 성능에 많은 영향을 미치게 되며 교배연산자에 대한 많은 연구가 진행되었다. 본 논문에서는 선택연산자는 룰렛 휠 방식을, 돌연변이 연산자는 균등돌연변이를 사용하였으며, 교배연산자는 산술적 교배연산자(Arithmetic crossover), 수정된 단순교배(modified simple crossover) 그리고 UNDX 알고리즘을 사용하였다. UNDX는 최근에 제안된 알고리즘으로 다른 연산자보다 성능이 우수함이 알려져 있다. 본 장에서는 UNDX 알고리즘만을 소개한다.

• Unimodal Normal Distribution Crossover

기존의 다른 연산자는 두개의 부모로부터 자손이 생성되는 반면에 UNDX는 다수의 부모(multiple parents)를 이용하여 부모들의 기하학적 중심(geometric center)에 근접하게 정규분포를 이루며 생성된다. 정규분포를 따르므로 작은 확률로 부모들의 중심으로부터 먼 자손들도 생성된다. 그림 1에서 정규분포의 중심은 parent1과 parent2의 중심 점이 되고 그 중심으로부터 표준편차를 따르는 랜덤변수(Gaussian random variable)로서 거리가 구해진다. parent3으로부터 parent1과 parent2를 잇는 선분에 수직인 벡터를 구하고 이 벡터와 또 다른 가우시안 랜덤변수와 비례하는 점에 자손이 위치하게 된다. 표 1은 UNDX의 전체 알고리즘을 보여준다.

w_i 와 v_i 는 평균이 0이고 분산이 각각 σ_1^2 , σ_2^2 인 가우시안 랜덤변수이며, $\sigma_1^2 = 1/\sqrt{\mu-2}$, $\sigma_2^2 = 0.35/\sqrt{\mu-\mu-2}$ 일 때, μ 가 3에서 7일 때

좋은 성능을 보인다. 식 (1)의 세 번째 항을 계산하는 데 있어서 Gram-Schmidt방법을 이용하여 $O(\mu^2)$ 의 비용으로 자손을 생산한다.

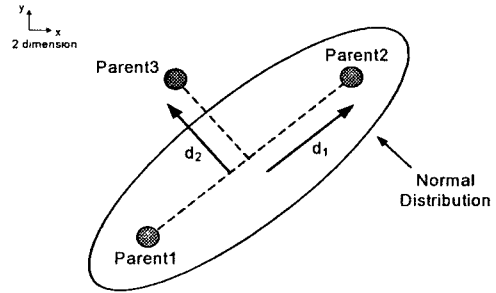


그림 1. UNDX의 기하학적 표현

표 1. UNDX 알고리즘

<p>Step1) select μ parent $x_i(i=1 \text{ to } \mu, x \in R^n)$ are randomly chosen from population P.</p> <p>Step2) Taking $\mu-1$ parents, geometric center c is computed.</p> <p>Step3) Calculate $d_i = x_i - c (i=1 \text{ to } \mu)$.</p> <p>Step4) The orthogonal component of d_μ to all d_i has length D.</p> <p>Step5) The orthonormal basis of the complementary subspace to space spanned by $d_i(i=1 \text{ to } \mu-1)$ is $e_i(i=\mu \text{ to } n)$</p> <p>Step6) The offspring y is given by (1)</p> $y = c + \sum_{i=1}^{\mu-1} w_i d_i + \sum_{i=\mu}^n v_i D e_i \quad (1)$
--

3. 계층적 공정 경쟁 병렬 유전자 알고리즘

진화알고리즘에서 중심적인 문제 중의 하나는 조기 수렴을 방지하고 균형 있는 탐색을 시도하는 것이다. 진화과정이 진행하면서 집단 평균 적합도는 점점 더 높아지게 되며, 단지 높은 적합도를 갖는 유사한 새로운 개체들이 생존하는 경향이 있다. 공정하게 탐색 되어진 다른 영역에 있는 새로운 개체는 그들의 유익한 특성이 발생되지 않는 한 일반적으로 낮은 적합도를 갖게 된다. 그래서 RCGA는 점점 더 발견된 지역 극점 근방에서 중점적으로 탐색을 수행하는 경향을 갖게 되며, 지역극점에 빠지게 된다. 이러한 현상은 높은 차수 문제

(high-dimensionality problems)와 다봉성을 가진 문제(multi-modal problems)에 있어서 더욱 많이 발생한다[8].

PGA는 이런 문제를 해결하기 위해 고안되었으며 RCGA에서는 단일 집단을 사용하는 반면 PGA 여러 개의 소집단으로 구성되었으며 각각의 집단은 다른 집단과 독립적으로 진화하면서 일정세대 간격으로 특정 개체는 다른 집단으로 이주하게 된다. 그림2에서는 구현된 병렬유전자 알고리즘의 흐름도를 보인다.

진화 과정은 RCGA와 거의 유사하며, 단지 이주 알고리즘을 더 포함하고 있다. 각각의 집단은 서로 다른 환경에서 진화하도록 집단크기, 진화연산자(선택, 교배, 그리고 돌연변이) 그리고 다른 파라미터 (교배율, 돌연변이율) 등을 집단 별로 다르게 설정할 수 있도록 구현되었다. Coarse-grained 모델에서 의 가장 큰 논점은 이주방법론이다. : 얼마나 많은 소집단을 사용할 것인가?, 어떻게 (얼마나 많이) 개체가 이주하는가?. 이주 방법에 따라 완전네트워크 구조, 링구조, 이웃구조가 있으며 최근에 계층적 공정 경쟁(HFC) 구조가 제안되었다[8].

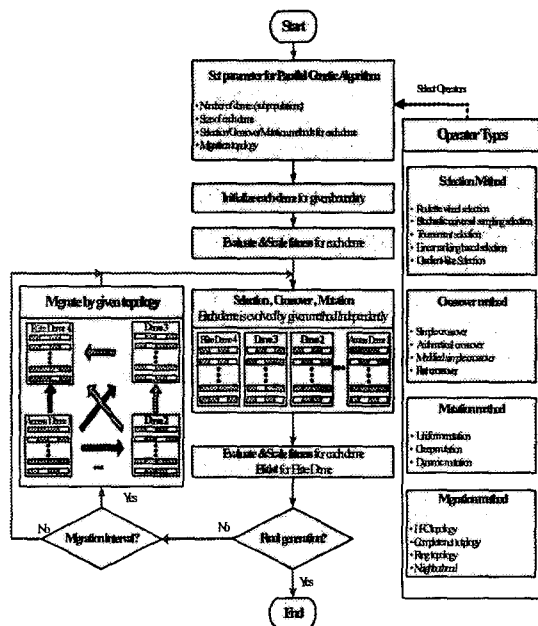


그림 2. 병렬유전자 알고리즘의 전체 흐름도

• HFC 구조 및 이주 방법

HFCGA에서 여러 개의 소집단은 계층적으로 구성되고, 각 집단은 특정범위의 적합도안에서 개체들을 수용할 수 있다. 각 서브 집단은 각 집단의 등급의 범위를 의미하는 진입문턱(admission threshold)을 갖고 있으며 이주할 개체를 저장할 수 있는 진입버퍼(admission

buffer)가 존재한다. 초기 생성된 개체들은 평가를 통하여 적합도가 구해지며 각 등급별로 소집단으로 나뉘게 된다. 그 후 각 집단의 개체들은 다른 집단과 독립적으로 주어진 환경(진화연산자, 교배율 및 돌연변이율)에 진화하게 되며 일정 세대동안 진화된 후 집단들 사이에 개체들의 이주가 이루어진다. 이주 시에 모든 집단 내에 있는 개체들의 적합도를 검사하여 상위집단의 허용 문턱값보다 높은 적합도를 갖는 개체는 상위 집단에 허용되는 진입버퍼에 저장된다. 이때 어떠한 개체도 하위 집단의 버퍼로 저장되지 않고 상위로만 올라가게 되는 단방향성을 갖는다. 이는 대학생이 초등학생이 될 수 없음을 뜻한다. 그 후 진입버퍼에 있는 모든 개체들은 해당 집단의 가장 열등한 개체와 비교를 통하여 교체가 이루어진다. 또 최하위 집단은 진입버퍼가 없으며 개체들의 다양성 유지를 위하여 모든 개체는 랜덤하게 초기화된다.

4. 실험 및 결과고찰

본 절에서는 본 연구의 평가로서 HFCGA, RCGA에 대하여 UNDX, 산술적 교배(Arithmetical Crossover: AX), 수정단순교배(Modified Simple Crossover: MSX)의 3가지 교배 연산자를 채택하여 4가지 테스트 함수에 대하여 비교 하였다. 본 연구의 테스트 함수로서 function parameter optimization, neural network training, fuzzy set optimization 등의 여러 분야에서 테스트 함수로 사용되는 Rosenbrock Function(FROS), Rastrigin Function(FRAS), Schwefel Function(FSCH) 그리고 Griewank Function(FGRI)을 사용하여 비교 분석을 수행하였다. 표 2에서는 네 가지 함수에 대한 특성을 보여주며 이들 함수는 최소값으로 0을 가짐을 미리 알 수 있다. 주어진 문제 해결을 위한 각 유전자 알고리즘에 사용된 파라미터는 표3와 같다. 표 4는 테스트 함수 4가지에 HFCGA와 RCGA를 3가지 교배 연산자를 이용하여 구해진 최소값을 보인다. 실험 결과에 따르면 HFCGA가 RCGA와 비교하여 보다 좋은 결과를 보이며 UNDX 교배연산자가 다른 교배 연산자보다 우수한 성능을 보임을 볼 수 있다.

5. 결론

본 논문에서는 계층적 공정 경쟁(HFC) 개념을 유전자 알고리즘에 적용하여 HFCGA를 구현하고 다양한 교배연산자를 적용하여 비교 분석을 수행하였다. 제안된 HFCGA는 서브집단이 계층적으로 이루어져 전역해를 찾기에 용이하며 또한 UNDX 연산자는 부모 개체로부터 생성될 수 있는 개체의 종류가 다른 알고리즘

보다 다양하여 전역해를 찾을 확률이 더욱 높다. 특히 다봉성이 크거나 고차원 문제에서는 HFCGA가 우수한 특성을 보임을 알 수 있었다. 테스트 함수뿐 아니라 많은 최적화 문제에서도 제안된 알고리즘은 더 우수한 성능을 보일 것으로 사료된다.

표 2 : 테스트 함수 요약

Functions	Range of x_i	Global minimum	N
$F_{ROS} = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2) + (1 - x_i)^2$	[-30, 30]	$x^*=(1, \dots, 1)$ $f(x^*) = 0$	10
$F_{RAS} = 10 \cdot n + \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i))$	[-5.12, 5.12]	$x^*=(0, \dots, 0)$ $f(x^*) = 0$	20
$F_{SCH} = 418.9829 \cdot n - \sum_{i=1}^n (x_i \sin \sqrt{ x_i })$	[-500, 500]	$x^*=(420.9687, \dots, 420.9687)$ $f(x^*) = 0$	2
$F_{CRI} = 1 + \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos(\frac{x_i}{\sqrt{i}})$	[-600, 600]	$x^*=(0, \dots, 0)$ $f(x^*) = 0$	10

N: the number of variables(dimension)

표 3. 유전자 알고리즘에서 사용된 파라미터

Parameters	HFCGA	RCGA
Generation	1000	1000
Number of demes	5	1
Number of individuals	[80, 80, 80, 80, 80]	400
Crossover rate	[0.7, 0.7, 0.7, 0.7, 0.7]	0.7
Mutation rate	[0.1, 0.1, 0.1, 0.1, 0.1]	0.1
Mutation	uniform for each deme	uniform
Selection	Roulette wheel for each deme	Roulette wheel
Calibration Stage	5 generation	X
Migration Interval	10 generation	X
Admission threshold	[0.1 0.3 0.5 0.85]	X

표 4 : 각 테스트 함수에 대한 실험 결과

Function	HFCGA			RCGA		
	UNDX	AX	MSX	UNDX	AX	MSX
F_{ROS}	0.00017	0.00016	0.06879	0.0249	0.0512	0.2451
F_{RAS}	26.9	13.4	45.8	58.3	23.8	24.5
F_{SCH}	2.6e-5	2.6e-5	2.7e-5	3.6e-5	5.7e-5	3.1e-5
F_{CRI}	0.04500	0.05330	0.06224	0.0469	0.10842	0.0723

감사의 글

이 논문은 2006년도 교육인적자원부의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2006-311-D00194)

참 고 문 헌

[1] J. H. Holland, Adaptation in Natural and Artificial Systems, The University of Michigan Press, Michigan, 1975.
 [2] D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Publishing Co. Inc., N.Y., 1989.
 [3] 진강규, "Genetic Algorithms and Their Applications" (2000).
 [4] Ono, I and Kobayashi, S : A Real-coded Genetic Algorithm for Function Optimization Using Unimodal Normal Distribution Crossover, Proceedings of the Seventh International Conference on Genetic Algorithms, pp. 246-253.
 [5] A. Wright, Genetic Algorithms for Real Parameter Optimization, Foundation of Genetic Algorithms 1, G. J. E. Rawlin(Ed.), Morgan Kaufmann Publishers, San Mateo, CA, 1991.
 [6] M. Nowostawski and R. Poil. Parallel genetic algorithm taxonomy. In L. Jain, editor, Proceedings, pages 88-92, Adelaide, August 1999. International conference on knowledge-based intelligent information engineering systems (KES'99), 3, Adelaide(AU), IEEE.
 [7] R. Tanese, "Distributed Genetic Algorithms" Proc. 3rd Int. Conf. on Genetic Algorithms, J.D.Schaffer(ed.), Morgan Kaufmann Publishers, San Mateo, pp. 434-439, 1989.
 [8] R. Lohmann, "Application of Evolution Stragy in Parallel Populations" Parallel Problem Solving from Nature, Lecture Notes in Computer Science, H.-P. Schwefel and R. Männer(Eds), Springer-Verlag, Vol. 496, pp. 198-208, 1991.