

유전자알고리즘에서 여왕벌 진화를 위한 강돌연변이 비율 및 확률의 자체조정

정성훈*

*한성대학교 정보통신공학과

e-mail:shjung@hansung.ac.kr

Self-tuning of Strong Mutation Rate and Probability for Queen-Bee Evolution in Genetic Algorithms

Sung Hoon Jung*

*Dept of Information and Communications Engineering, Hansung University

요 약

본 논문에서는 여왕벌 진화를 모방하여 개발한 유전자알고리즘에서 강돌연변이 수행비율 및 강돌연변이 확률을 자체적으로 조정하는 방법을 제안한다. 이렇게 함으로써 적절한 강돌연변이 수행비율 및 강돌연변이 확률을 여러 번의 실험을 통하여 경험적으로 선택하는 문제를 완화하여 여왕벌 진화의 적용을 보다 쉽게 할 수 있다. 3개의 최적화문제에 제안한 방법을 적용해 본 결과 비교적 우수한 성능을 보였다. 하지만 다수의 실험을 통하여 얻은 최고의 성능보다는 우수하지는 못했는데 추후 성능을 보다 더 개선하여 이에 근접한 성능을 얻을 수 있는 알고리즘의 개발이 필요하다.

1. 서론

유전자알고리즘은 다윈의 적자생존을 기반으로 만들어진 최적화알고리즘으로 많은 공학문제에 다양하게 적용되어 성공적으로 활용되어 왔다 [1-3]. 유전자알고리즘의 장점으로는 여러 개의 개체가 동시에 탐색을 수행함으로써 지역 최적해에 빠질 가능성이 적다는 것과 문제영역과 상관없이 비트열로 변환된 영역에서 탐색을 수행함으로써 다양한 문제에 적용할 수 있다는 것이다 [1]. 그러나 단점으로는 전역 최적해를 보장하지 못한다는 것과 또한 문제에 따라서 진화가 더더 만족할만한 해를 찾지 못할 수 있다는 것이다. 이러한 이유로 빠르게 전역 최적해를 찾을 수 있는 유전자알고리즘을 개발하려는 노력들이 진행되어 왔으며 우리는 여왕벌의 진화방법을 모사한 여왕벌진화 알고리즘을 개발하였다 [4].

여왕벌진화 알고리즘에서는 자연계에서의 여왕벌의생식과 유사하게 여왕벌(한 세대에서 최고의 적합도를 갖는 개체)과 적합도에 비례하여 선택된 다른 개체가 생식을 수행한다. 이러한 생식은 적합도에 비례하여 선택된 개체들끼리 생식하는 기존의 유전자알고리즘의 생식방법보다 더 기존의 찾은 정보를 주로 이용하는 개발(exploitation)이 강화된 것으로 새로운 영역을 찾아보는 개척(exploration)이 약화된다. 개발이 강화되면 최적해 쪽으로 빨리 이동하는 장점이 있지만 해당 최적해가 지역최적해일 경우 반대로 지역 최적해에 빠져서 빠져나오기 어려운 문제를 야기한다. 개체들이 여러 영역을 찾아보지 않고 초기세대에서 찾은 지역해로 급격하게 이동하는 경우 최종해의 질이 나빠지는데, 이러한 문제를 조속수렴(premature

convergence)문제라고 하며 이를 해결하기 위해서는 개체의 다양성을 유지해 주어야한다. 특히 여왕벌 진화같이 개발이 강화된 알고리즘에서는 개체를 더욱 강화시켜 주어야한다. 유전자알고리즘의 연산은 교배(crossover)와 돌연변이(mutation)로 이루어져 있는데 교배는 개발중심 연산이며 돌연변이는 개척중심 연산이다. 그러므로 여왕벌 진화에서는 개체를 강화시키기 위하여 돌연변이를 강화시킨 강돌연변이 연산을 이용하였다. 좀 더 자세히 설명하면 선택된 개체를 이용하여 여왕벌과 교배연산을 수행한 후 돌연변이 연산에서 일정부분의 개체를 정상 돌연변이 확률보다 더 높은 확률로 돌연변이 시키는 강돌연변이 수행한다. 이 때 정상돌연변이 연산을 수행할 개체와 강돌연변이 연산을 수행할 개체의 비와 강돌연변이 확률 값이 미리 설정되어 있어야한다. 논문 [4]에서 여러 개의 값을 이용하여 실험해 본 결과 이 비율과 확률 값이 성능에 많은 영향을 주는 것을 확인하였다. 그러나 아직까지 이러한 값들을 시스템적으로 설정하는 방법이 개발되지 않아 여왕벌진화 방법이 성능이 우수함에도 실제 문제에 적용하는데 문제점으로 작용하였다.

본 논문에서는 강돌연변이 연산을 수행할 개체 비율과 강돌연변이 확률 값을 유전자알고리즘 스스로 자체 조정해 나가며 여왕벌 진화를 하는 새로운 방법을 제안한다. 제안한 방법을 3가지 최적화 문제에 적용해 본 결과 비교적 좋은 성능을 보였다. 그러나 논문 [4]에서 선택하여 실험한 결과 중 가장 좋은 실험결과보다는 성능이 좋지 못하였다. 이는 본 논문에서 제안한 방법은 계속적으로 두 값을 자체 조정하는 방법으로 논문 [4]에서처럼 처음 정한

값이 유지되어 진행되는 것과의 차이에서 올 수 있다. 그렇다하더라도 더 좋은 자체 조정하는 방법이 있을 수 있기 때문에 향후 추가적인 연구가 필요하다.

2. 제안한 유전자알고리즘

다음 Algorithm 1은 강돌연변이 비율 및 확률을 자체 조정하는 유전자알고리즘을 보여준다. Algorithm 1에서 라인 11이 강돌연변이 비율 및 확률을 자체 조정하기 위하여 추가해준 부분이다. 실제 조정하는 부분은 라인 12~18에 있다.

Algorithm 1 Proposed genetic algorithm

```

// t : time //
// n : population size //
// P : populations //
//  $\xi$  (1- $\xi$ ) : normal (strong) mutation rate //
//  $p_m$  : normal mutation probability //
//  $p'_m$  : strong mutation probability //
//  $I_q$  : a queen-bee individual //
//  $I_m$  : selected individuals //
//  $f_{avg}^{(t-1)}$ : average fitness at time  $t-1$  //
//  $f_{avg}^t$  : average fitness at time  $t$  //
1 t  $\leftarrow$  0
2 initialize  $P(t)$ 
3 evaluate  $P(t)$ 
4 while (not termination-condition)
5 do
6   t  $\leftarrow$  t + 1
7   select  $P(t)$  from  $P(t-1)$ 
8    $P(t) = \{(I_q(t-1), I_m(t-1))\}$ 
9   recombine  $P(t)$ 
10  do crossover
11  do self-tuning of  $\xi$  and  $p'_m$  (♠)
12    if  $f_{avg}^t \leq f_{avg}^{(t-1)}$ 
13       $\xi = \xi - 0.01$ , where  $\xi \geq 0.5$ 
14       $p'_m = p'_m + 0.05$ , where  $p'_m \leq 1$ 
15    else
16       $\xi = \xi + 0.01$ , where  $\xi \leq 1$ 
17       $p'_m = p'_m - 0.05$ , where  $p'_m \geq 0.05$ 
18    end if
19  do mutation
20    for  $i = 1$  to  $n$ 
21      if  $i \leq (\xi \times n)$ 
22        do mutation with  $p_m$ 
23      else
24        do mutation with  $p'_m$ 

```

```

25      end if
26    end for
27  evaluate  $P(t)$ 
28 end

```

강돌연변이 비율과 확률은 유전자알고리즘이 지역 최적해에 빠졌을 때 높여주어 지역 최적해에서 빠져나오게 해야 한다. 그렇게 하기위해 유전자알고리즘이 지역 최적해에 빠졌는지를 평가하는 척도가 있어야한다. 유전자알고리즘이 지역 최적해에 빠졌는지를 평가하는 척도는 어떤 파라미터를 동원하는가, 해당 파라미터를 어떻게 이용하여 판단하는가, 판단 기준이 되는 값은 어떤 값을 사용하는가, 파라미터 값을 어떻게 조정하는가 등에 따라서 매우 다양하게 고안될 수 있다. 본 논문에서는 비교적 단순하게 이전 세대에서의 평균적합도 $f_{avg}^{(t-1)}$ 과 현재 세대에서의 평균적합도 f_{avg}^t 를 이용하여 판단한다. Algorithm 1의 라인 12에 있듯이 만약 현재 세대의 평균적합도 f_{avg}^t 가 이전 세대의 평균적합도 $f_{avg}^{(t-1)}$ 보다 작거나 같다면 지역 최적해에 빠진 것으로 판단하여 정상 돌연변이를 수행할 비율을 0.01만큼 줄이고 (다시 말하면 강돌연변이를 수행할 비율을 늘리고) 강돌연변이 확률 p'_m 를 0.05만큼 크게 한다. 다만, 최대 강돌연변이 비율(1- ξ)은 0.5로 한정한다. 만약 강돌연변이에 의하여 지역 최적해 영역을 빠져나오면 다시 현재 평균적합도는 이전의 평균적합도보다 좋아질 것이고 이 경우에는 정상 돌연변이 비율을 늘리고 강돌연변이 확률을 줄인다. 다만 강돌연변이 확률은 정상돌연변이 확률인 0.05보다 작아지지 않게 한다.

3. 실험결과

제안한 알고리즘의 성능을 측정하기 위하여 3개의 최적화 문제에 적용하여 실험하였다. 3개의 최적화문제는 다음 식 (1)과 같으며 각 문제의 적합도가 그림 1에 그려져 있다.

$$\begin{aligned}
 f_1 &= \sum_{j=1}^h m_j \begin{cases} m_j = 1 & \text{if } T_j = I_j \\ m_j = 0 & \text{if } T_j \neq I_j \end{cases} \\
 f_2 &= 100(x_1^2 - x_2)^2 + (1 - x_1)^2, \\
 &\quad \text{where } -2.048 \leq x_i \leq 2.048 \\
 f_3 &= 0.5 - \frac{\sin(\sqrt{x_1^2 + x_2^2})\sin(\sqrt{x_1^2 + x_2^2}) - 0.5}{(1.0 + 0.001(x_1^2 + x_2^2))(1.0 + 0.001(x_1^2 + x_2^2))}, \\
 &\quad \text{where } -10 \leq x_i \leq -10
 \end{aligned} \tag{1}$$

첫 번째 문제는 패턴매칭 문제로서 임의의 목표패턴에 대하여 유전자알고리즘 개체의 0과 1의 패턴을 일치시키는 문제이다. 그러므로 최대 적합도는 목표패턴의 비트수인 h 가 된다. 두 번째 문제는 DeJong 함수 2로서 크게 보면 한 개의 지역 최적해 영역과 한 개의 전역 최적해 영역으로 이루어져 있다. 이 문제의 경우 지역 최적해 영역과 전

역 최적해 영역이 넓게 분포하고 있어서 개체가 지역 최적해 영역에 빠진 경우 쉽게 지역 최적해 영역을 빠져 나

진화가 일어난 경우에는 단봉함수와 유사해서 비교적 쉽게 전역 최적해를 찾을 수 있다. 세 번째 문제는 멕시코모자 문제로서 가운데에 있는 좁은 영역이 전역 최적해 영역이며 그 이웃하는 영역은 모두 지역 최적해 영역으로 지역 최적해 영역이 전역 최적해 영역을 둘러싸고 있어서 전역 최적해를 찾기 매우 어려운 문제이다. 지역 최적해 영역을 빠져나오는 효과적인 방법이 없는 알고리즘의 경우에는 지역 최적해 영역에서 오래 동안 머물러 성능이 매우 좋지 않게 된다.

위와 같은 세 개의 문제에 기존의 방법과 본 논문에서 제안한 방법을 적용하여 실험하였다. 표 1은 유전자알고리즘의 설정된 파라미터로서 논문 [4]와 똑같이 설정하였다.

표 1: 유전자알고리즘의 파라미터

파라미터	값
교배확률 (p_c)	0.6
정상 돌연변이 확률 (p_m)	0.05
개체 수	10
개체 비트	24 bits
정상 돌연변이 비율 (ξ)	0.4, 0.6, 0.8, 1.0
강돌연변이 확률 (p'_m)	0.6, 0.8, 1.0

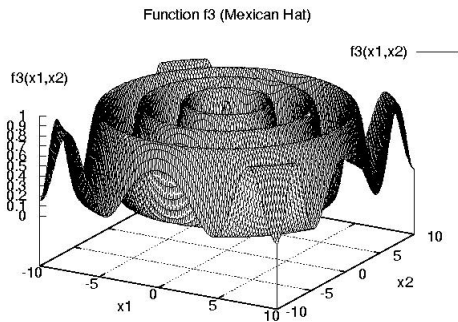
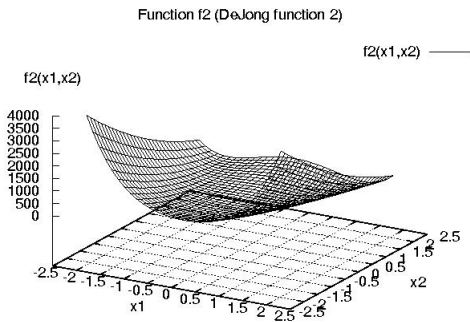
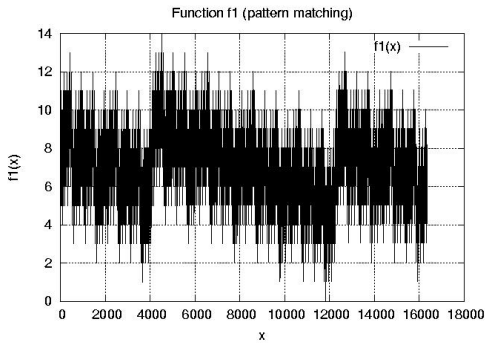


그림 1: 최적화 문제

오기 어려운 문제에 해당한다. 물론 전역 최적해 영역으로

실험은 주어진 파라미터와 무작위 초기개체를 이용하여 전역 최적해를 찾는 실험을 하였다. 전역 최적해를 찾았을 때의 세대 수를 기록하고 이 기록을 평균하여 실험결과로 제시하였다. 각 파라미터에 대하여 다른 무작위 초기 개체로 10번의 실험을 하였으며 그 값을 평균한 실험결과가 표 2에 있다 (QGA의 실험결과는 숫자의 정밀도 차이로 인하여 논문 [4]의 결과와 약간 다르다). 표 2에서 SGA는 simple GA로서 초기에 제안된 방법이며 QGA는 여왕벌 진화를 적용한 방법 그리고 PGA는 본 논문에서 제안한 방법을 표시한다. 표 2에서 보듯이 기존의 방법 QGA는 정상 돌연변이 비율 ξ 와 강돌연변이 확률 p'_m 에 따라서 성능이 크게 좌우된다. 본 논문에서 제안한 방법인 PGA는 정상 돌연변이 비율과 강돌연변이 확률을 자체 조정하면서 동작하기 때문에 초기 값으로 1과 0.05만 설정하고 따

표 2: 실험결과

SGA		64201.3		59078.5		776085.3	
ξ	p'_m	f_1		f_2		f_3	
		QGA	PGA	QGA	PGA	QGA	PGA
0.4	0.6	40039.6	59.3	50147.8	995.9	1368656.2	13194.6
	0.8	56686.6		24768.4		549066.7	
	1.0	6933.1		10347.2		7701.1	
0.6	0.6	1115.3		1560.8		355116.1	
	0.8	654.9		2354.5		295721.0	
	1.0	351.0		962.6		4170.3	
0.8	0.6	84.2		732.4		65066.5	
	0.8	97.5		349.2		37867.0	
	1.0	77.5		(*) 298.0		(*) 3767.3	
1.0	(*)	58.4	1840432.7	17779.8			

로 설정할 필요가 없다. 표 2에서 보듯이 첫 번째 문제 f_1 에서는 거의 유사한 성능을 보였다. 두 번째 문제 f_2 에서는 최고 성능인 298.0보다는 좋지 않은 995.9를 보였으나 비교적 크게 나쁘지 않은 성능을 보였다. 세 번째 문제에서는 이전 두 문제에서보다는 비교적 나쁜 성능을 보였으나 SGA 보다는 월등히 좋았다. 세 번째 문제에서 성능이 좋지 않은 이유는 세 번째 문제는 이전에서 언급한 것처럼 전역 최적해 영역을 지역 최적해 영역이 몇 겹으로 둘러싸고 있기 때문인 것으로 보인다. 이 경우 개체가 지역 최적해 영역을 벗어나서 전역 최적해 영역으로 이동하기가 어려우며 정교한 제어가 필요하기 때문이다.

4. 결론

본 논문에서는 유전자알고리즘의 성능향상을 위하여 제안된 여왕벌진화방법에서 강돌연변이 비율 및 강돌연변이 확률을 자체 조정하는 방법을 제안하였다. 성능에 많은 영향을 미치는 강돌연변이 비율 및 강돌연변이 확률을 경험적으로 선정하는 것은 많은 시행착오를 거칠 수밖에 없기 때문에 효율적인 알고리즘의 개발이 매우 중요하다. 제안한 알고리즘을 적용해서 실험한 결과 비교적 단순한 문제에서는 비교적 좋은 성능을 보였으나 지역 최적해가 많아 최적화가 어려운 문제에서는 성능이 그렇게 좋지는 않았다. 추후 보다 더 정교한 방법을 개발할 필요가 있다.

참고문헌

- [1] D. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning," Addison-Wesley.
- [2] C. Xudong, Q. Jingen, N. Guangzheng, Y. Shiyong, and Z. Mingliu, "An Improved Genetic Algorithm for Global Optimization of Electromagnetic Problems," IEEE Trans. on Magnetics, vol. 37, pp. 3579-3583, Sep. 2001.
- [3] Zhihua Tang, Youtuan Zhu, Guo Wei, and Jinkang Zhu, "An Elitist Selection Adaptive Genetic Algorithm for Resource Allocation in Multiuser Packet-based OFDM Systems," Journal of Communications, vol. 3, no. 3, pp. 27-32, Jul 2008.
- [4] S. H. Jung, "Queen-bee Evolution for Genetic Algorithms," Electronics Letters, vol. 39, no. 6, pp. 575-576, Mar 2003.