

유전자 알고리즘에서 bias에 의한 adaptive한  
개체군 크기의 설정

Design of Adaptive Population-size on Bias  
in Genetic Algorithms

\* 김 용 범      \*\* 오 충 환

< Abstract >

One of the problems brought up in the effective execution of genetic algorithms is that if they come under any influences according as the population size is large or small. In the case of small population size the opportunities of premature convergence are increased when the greatly powerful or no good individual is generated during search of the solution space. And searching the solution space in the case of large population size, the difficulties under the execution cause to searching all for one by one individual in every generation applied is limited, this gives the many interruptions to the convergence of final solution. Now this paper gives a suggestion to set up the adaptive population size which could compute the more correct solution and simplify the development of computation performance.

---

\* 명지 대학교 산업공학과 박사과정

\*\* 수원전문대학 교수

## 1. 서 론

지금까지 생물의 발달에 대해 창조론과 진화론의 양대주장이 팽팽히 맞서고 있다. 다윈의 이론은 많은 생물 개체들간의 경쟁에 의한 환경의 적응으로 인해 좀 더 나은 개체들이 종족을 이어간다는 소위 적자 생존이론이 그 배경을 이루고 있다. 이러한 원리를 컴퓨터에 적용하여 보다 지능적인 판단을 가능케하는 노력이 진행되고 있는데 이것이 곧 진화연산(evolutionary computation)이다. 진화연산의 중요한 연구과제는 robustness와 서로다른 환경에서 생존에 필요한 효율(efficiency)과 효능(efficacy)이 균형을 이루어야 한다는 것이다[3].

이러한 측면에서 robustness를 갖는 기법에 대한 연구가 최근에 대단히 활발하다. robustness를 갖는 진화연산은 유전자 알고리즘(genetic algorithm), 진화기법(evolutionary strategy), 진화 프로그래밍(evolutionary programming) 등이 있다. 이 가운데서 Holland(1975)에 의하여 제시된 유전자 알고리즘은 실용성 면에서 매우 우수하다고 알려져 있다.

유전자 알고리즘을 수행하는데 있어 대부분의 의문사항들 중의 하나는 그것이 개체군의 크기에 어떻게 영향을 받느냐 하는 것이다. 개체군의 크기가 적은 경우, 해 공간을 탐색하는데 있어 너무 강한 개체가 발생되거나, 또는 좋지 못한 개체의 발생으로 인해 사전수렴(premature convergence)의 기회가 증가하게 된다[1]. 한편 개체군의 크기가 큰 경우, 해공간을 탐색하는데 있어 세대마다 하나 하나의 개체를 모두 탐색함으로써 인한 수행상의 어려움이 존재한다. 만약 활용되는 세대수가 제한되어 있다면, 무엇보다도 해의 수렴에 많은 방해가 받게 된다. 따라서 adaptive한 개체군의 선정이 무엇보다도 중요하게 대두된다.

대부분의 실제 유전자 알고리즘이 적용된 문제에 있어서는 극단적으로 큰 개체군의 크기가 적용되어 겪게되는 수행상의 문제점으로 인해 알고리즘을 수행하는 것이 simulated annealing과 TABU search와 같은 최적화 기법과는 경쟁이 되지 않는다 [1]. 개체군의 크기가 큰 경우, 유전자 알고리즘을 실행하는데 많은 문제점이 대두되므로 효과적인 유전자 알고리즘을 수행하기 위해서는 적합한 개체군의 크기로 시작하는 것이 불가피하다. 다행히도, 대부분의 실험 결과에서는 30개 정도의 개체군의 크기가 적합하다고 제시되었으나, 이와같은 경우 무엇보다도 선정된 초기 개체군에 의해 해가 매우 민감하게 반응하게 된다[4]. 따라서 본 연구에서는 보다 정확한 해를 계산하고, 계산의 수행을 보다 간편화시킬 수 있는 adaptive한 개체군의 크기를 설정하고자 한다.

한편 유전자 알고리즘은 해의 수렴 단계에서 미리 정해진 세대에 도달되면 종결한다. 이러한 종결방식은 대상문제자체에 이용되는 유전자 알고리즘의 parameter의 특성을 반영시키지 못하므로 이러한 측면에서 parameter의 특성을 자동적으로 반영하는 계

산 종결조건을 도입 할 필요가 있다. 이러한 조건은 최적해를 얻는데 필요한 시간과 최적해 자체의 신뢰도와 밀접한 관계가 있을 뿐만 아니라 대상문제에 적합하다고 가정되는 세대수를 제한할 수 있다. 즉 이와같은 종결조건을 도입함으로써 이 기법의 효율 및 효능 이 크게 증가될 수 있음을 의미한다. 이로써 대상문제에 적합하다고 가정되는 제한된 세대수를 이용하여 최적해를 구하기 위한 적응적인 개체군의 크기의 특성을 연구할 필요성이 크다고 볼 수 있다.

## 2. 유전자 알고리즘의 개요

자연적 현상에서는 정교한 정보의 과정과 adaptation에 관한 많은 방식이 존재한다. 이러한 많은 방식중에서 특히 필적할 만한 정교함을 갖는 인공시스템을 설계하기 위해서는 자연적인 현상을 탐색기법으로 모델화한 확률적 알고리즘인 유전자 알고리즘이 매우 효과적이다[5].

유전자 알고리즘을 시작하기 위해서는 대상문제를 해결하기 위해 유전자 알고리즘과의 연결구조를 고려하여야 한다. 이러한 구조는 염색체(chromosome)로 대상문제의 해(solution)를 code화하는 방식과, 대상문제와 관련된 염색체의 가치를 측정하기 위한 평가함수(evaluation function)이다. 해를 code화하기 위한 기법은 매우 다양하다. Holland(1975)는 대상문제의 coding기법을 비트(bit)단위로 이루어지는 string들을 이용하여 수행하였다. 한편, 평가함수는 유전자 알고리즘과 대상문제간의 연결고리 역할을 한다. 평가함수는 투입요소로서 염색체를 취하고 대상문제를 해결하기 위한 염색체의 수행도(performance)를 측정하는 것으로, 환경(environment)이 자연상태에서의 진화역할을 하는 것처럼 유전자 알고리즘에서 같은 역할을 한다. 그와같은 환경을 갖는 개체(individual)들간의 상호작용은 적합도(fitness)를 측정하는 것이고, 평가함수를 갖는 염색체 간의 상호작용은 유전자 알고리즘이 복제(reproduction)를 수행할 때 이용되는 적합도의 측정치를 제공한다. 대상문제의 해를 탐색(search)하기 위해서는 이러한 초기 요소들이 주어졌을 때, 개체군(population)을 이용한 모의진화과정(simulated evolution process)을 단계별로 수행하는 유전자 알고리즘을 이용한다.

기본적인 복제계획은 두 가지의 유전 연산자인 교차와 돌연변이에 의해 이루어진다. 이와같은 연산자를 더욱 자세하게 명시하기 위해, 개체군 A로부터 원소(element)  $s_i$ 는 string  $a_1 a_2 a_3 \dots a_{ii}$ 로써 표현된다고 하자. 여기서  $a_{ii}$ 는 유전인자의 값을 표현한 것이다.

교차는  $s_i$ 로부터 교차점 앞의 유전인자 부분과  $s_j$ 로부터 교차점(crossover point) 뒤의 유전인자 부분을 연결하여 이미 존재하는 개체  $s_i$ 와  $s_j$ 로부터 새로운 개체인  $s_i'$

과  $s_j'$ 를 생성하는 것이다. 이러한 유전인자의 부분들은 유전인자의  $l-1$  위치들에서 균등분포(uniform distribution)로부터 랜덤하게 선정된 교차점에 의해 정의한다. 예를 들어, 만약 교차가 3번째와 4번째 유전인자 사이에서 발생되었다면, 개체  $s_i'$ 과  $s_j'$ 는  $s_i$ 와  $s_j$ 로부터 다음과 같이 생성된다.

$$\begin{array}{r}
 \text{Parent 1 : } s_i = a_{i1} \ a_{i2} \ a_{i3} \ | \ a_{i4} \dots\dots\dots a_{in} \\
 \text{Parent 2 : } s_j = a_{j1} \ a_{j2} \ a_{j3} \ | \ a_{j4} \dots\dots\dots a_{jn} \\
 \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \uparrow \\
 \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \text{교차점(pos)} \\
 \\
 \text{Offspring 1 : } s_i' = a_{i1} \ a_{i2} \ a_{i3} \ a_{j4} \dots\dots\dots a_{in} \\
 \text{Offspring 2 : } s_j' = a_{j1} \ a_{j2} \ a_{j3} \ a_{i4} \dots\dots\dots a_{jn}
 \end{array}$$

교차는 다음과 같은 방식으로 유전자 알고리즘에 이용한다. 자손을 생성하기 위해 개체군 A(N)으로부터 선정된 개체  $s_{iN}$ 이 주어지면, 또 다른 개체  $s_{jN}$ 는 선택확률을 이용하여 개체군 A(N)로부터 선정한다. 그러면 새로운 자손은 교차에 의해 생성된다. 따라서 교차는 기본적인 알고리즘의 적합도에 의해 선정된 점들의 샘플로부터 새로운 점들의 샘플인 더 좋은 해들을 생성하기 위해 공간A의 탐색에 적용된 전략이다. 그러나 만약 특정한 유전인자  $a_{ij}$ 가 개체군 A(N)에 존재하지 않는다면, 교차에 의해 생성되는 어떠한 자손도  $a_{ij}$ 를 포함하지 않는다. 또한 하나의 유전인자는 여러가지 이유로 인하여 개체군 A(N)으로부터 소멸될 수 있다. 이것은 좋지 못한 적합도 때문에 선택에 의해 소멸되거나, 개체군 A(N)의 제한된 크기 때문에 단순하게 소멸된다.

단순 유전자 알고리즘(simple genetic algorithms:sGA)에서 새로운 유전인자들은 두 번째 유전 연산자인 돌연변이에 의해 개체군 A(N)에서 생성된다. 돌연변이는 개체에서 하나 또는 그 이상인 유전인자의 값을 독립적(independent)으로 수정하여 새로운 개체를 생성하는 연산자이다. 하나의 유전인자는 l개의 유전인자의 위치들에서 균등분포로부터 수정되기 위해 선택된다. 즉, 새로운 유전인자의 값은 기존의 유전인자의 값에 의해 랜덤하게 선정되는 것으로 1은 0으로 0은 1로 바뀐다. 돌연변이는 전형적으로 매우 낮은 확률값을 갖는다. 예를 들어 돌연변이 확률이 0.01로 가정되었다 하자. 그러면, 돌연변이가 하나의 유전인자에 적용되었을 때, 염색체내의 각 비트는 랜덤하게 선정될 기회를 100개중 1개를 가지게 된다. [그림 1]은 돌연변이  $P_m=0.01$ 이 적용된 예제를 표현한 것이다. 일반적으로 자연계에서 한 유전인자에 돌연변이가 진행될 확률은 0.001

보다 적은 값으로 적용한다[3]. 이것은 어떠한 유전인자도 개체군 A( N)으로부터 영구히 소멸되지 않는다는 것을 보장하는 배경 연산자(background operator)로서 적용된다.

기존의 염색체	난 수				새로운 비트	새로운 염색체
1 0 1 0	0.801	0.102	0.266	0.377	-	1 0 1 0
1 1 0 0	0.120	0.196	<u>0.005</u>	0.840	1	1 1 <u>1</u> 0
0 0 1 0	0.760	0.473	0.894	<u>0.001</u>	1	0 0 1 <u>1</u>

[그림 1] 돌연변이의 적용예제 ( $P_m = 0.01$ )

모의진화 과정에서 이와같은 구성요소들을 이용한 유전자 알고리즘의 적용단계는 다음과 같다.

#### 적용단계

- 단계 1. [염색체들의 초기 개체군을 랜덤(random)하게 생성한다.]
- 단계 2. [개체군 내에 존재하는 각 염색체의 값을 구한다.]
- 단계 3. [선정된 현재의 염색체들에 교차(crossover)와 돌연변이(mutation)를 적용하여 새로운 염색체를 생성한다.]
- 단계 4. [새롭게 생성된 염색체들을 위해, 개체군에서 해당되는 조상(parents)의 염색체를 삭제한다.]
- 단계 5. [새롭게 생성된 염색체들의 값을 구하고 개체군 내에 그것들을 삽입한다.]
- 단계 6. [만약 정해진 세대(generation)에 도달되면 알고리즘을 끝내고 그때의 최상의 염색체가 최적해가 된다. 만약 그렇지 않다면 단계 3으로간다.]

### 3. 개체군(Population)에 대한 모델

개체군의 유전자는 그 개체군이 갖고있는 수준에서 유전자 형질(heredity)과 진화(evolution)의 특성에 관련된다. 이것은 유전자 형질의 구조에 대한 Mendelian의 견해를 가정한 것으로, 즉 유전자의 유형은 개체군을 형성하는 개체들에서 준수해야 할 가

정들을 조정하고 있는 유전인자(gene)들로 구성된 염색체들의 결합으로 표현한다[2]. 하나의 개체군은 환경에서 존재하는 인자(factor)들에 대한 한 세대로부터 그 다음 세대까지 변화하게 되는 특성을 갖는 유전자 정보(genetic information)의 유동적인 집합소(pool)로써 고찰된다. 주어진 세대에서 그 환경의 복잡한 변화에 적합하게 대응하는 개체군에 대한 많은 예제가 제시되었다[1].

유전자의 개체군 모델에서 개체(individual)들은 유전자를 형성하는 형태로서 단순하게 표현한다. 이와같은 유전형질은 2진수(binary)로 표현되며, 단순한 하나의 염색체로 이루어진 개체들로부터 복잡한 다중 염색체(multi chromosome)로 이루어진 개체들까지 매우 다양하다[3]. 유전형질에 대한 특정한 표현에서 개체들의 특성은 2진수로 이루어진 유전인자들의 목록(list)으로 구성된 염색체의 함수로서 정의한다.

세대가 진행되어가면서 그에 따른 개체군의 개선(development)은 적합도(fitness), 생존과 소멸의 주기(life-death cycle), 교배 규칙(mating rule), 변동성(mobility), 성별(sex), 종(species) 등의 형태에 의해 이루어진다[2]. 이와 같은 개선을 위해 제공되는 adaptation의 구조적인 측면에서 흥미로운 점은 매우 단순한 인공적인 체계에서도 현저한 adaptive의 가능성이 존재한다는 것이다.

#### 4. 실험적 연구

본 연구에서는 bias에 의한 종결조건을 이용한 유전자 알고리즘을 적용한 adaptive한 개체군의 크기를 설정하기위해 프로그램을 작성하고 최적해를 구하기위한 실험을 수행하였다. 실험에 사용된 컴퓨터는 Pentium 100, 2MB VGAcad, 16MB Memory의 환경에서 실행하였으며, 사용된 프로그램은 Mathematica 2.2.3으로 적성하였다.

또한 해의 종결조건은 개체군들간의 일치정도를 측정하는 bias의 개념을 이용하여 측정하였고, 현 세대에서 우성인 개체를 선택하는 방식으로 roulette wheel 선택방법을 사용하였다.

본 연구에서 실험에 이용된 대상문제는 다음과 같다.

$$\begin{aligned} \max \quad & f_1(x, y) = 21.5 + x\sin(4\pi x) + y\sin(20\pi y) \\ & -3.0 \leq x \leq 12.1 \\ & 4.1 \leq y \leq 5.8 \end{aligned}$$

[그림 2]는 개체군의 크기 50, 100, 300 적용시 유전인자의 평균 수렴비율을 도식화

한 것이다. 기대되었던 것처럼, 개체군 크기의 증가는 제한된 구간에서 유전인자의 수렴비율을 현저하게 감소시킨다. 이와같은 효과는 세대수에 의한 것 보다 시행수에 의한 형태로 표현하게 되면 더욱 분명하게 나타난다. 즉, 유전인자의 수렴도는 제한된 세대에서 큰 개체군의 경우가 작은 개체군의 경우보다 유전인자의 수렴비율이 떨어진다. 이것은 보다 크게 적용된 개체군에서는 작은 개체군에 비해 유전인자의 손실비율이 그만큼 적게 발생되므로 인해 나타나는 낮은 확률적인 효과로 감소되는 것 같다. 따라서 유전인자의 사전수렴 문제는 개체군의 크기를 증가시킴으로 인해 효과적으로 감소시킬 수 있다. 그렇지만, 이와같은 경우는 알고리즘의 수행도에서 많은 문제가 발생된다.

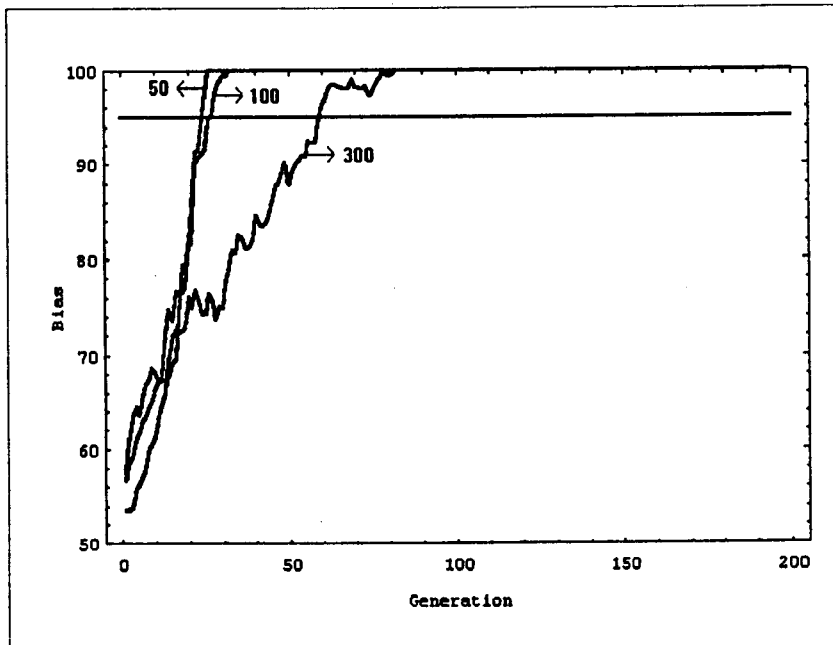
주어진 환경(roulette wheel 선택기법)에서  $f_1$ 의 adaptive한 수행도인 적합도를 측정하기 위해 평가함수의 최대치인  $f_1$ 와 평균치인  $\bar{f}_1$ 에 다양한 개체군의 크기를 적용하였다. [그림 3]은 개체군의 크기에 따른  $f_1(N)$ 의 최대 적합도를 제시한 그래프이다. 여기에서  $n = 50$ 인 개체군은  $n = 100$ 과  $n = 300$ 인 개체군에 비해 해의 수렴도는 빠르게 얻을 수 있으나, 최적이지 아닌 상태인 사전수렴의 결과가 나타난다. 그러나 개체군의 크기가 큰 경우, 작은 개체군의 크기에 비해 해의 수렴도는 느릴지라도 보다 좋은 해의 수행결과를 얻을 수 있다. [그림 4]는 개체군의 크기에 따른  $f_1(N)$ 의 평균 적합도를 제시한 그래프이다. 여기에서는 [그림 3]의 결과와는 반대로  $n = 50$ 에서 주목할만한 결과가 제시된다. 즉 상대적으로 작은 개체군이 적용된  $n = 50$ 인 경우  $n = 100$ 과  $n = 300$ 에 비해 평균적으로 우성인자가 사전에 생성된다는 것을 의미한다. 따라서 함수  $f_1$ 에 알고리즘 적용시 상대적으로 개체군의 크기가 적을수록 해의 수렴도는 빠르게 얻을 수 있으나, 우성인자의 발생으로 인해 사전수렴의 원인이된다. 그렇지만 제한된 상태에서는 우성인자의 사전 생성으로 인해 전반적으로 개체군의 크기가 작은 경우에서 보다 adaptive한 결과가 제시된다.

## 5. 결 론

본 연구에서는 유전자 알고리즘에서 bias에 의한 adaptive한 개체군의 크기에 대해 연구하였다. 그 결과 최적해에 근사한 값을 구하기 위한 adaptive한 개체군의 크기를 설정 할 수 있었다. 최근에는 유전자 알고리즘의 여러 가지 조작을 이용한 변형된 알고리즘들이 많이 고안되었다. 향후 변형된 알고리즘들을 적용한 새로운 문제해결 방안도 연구과제가 될 것이다.

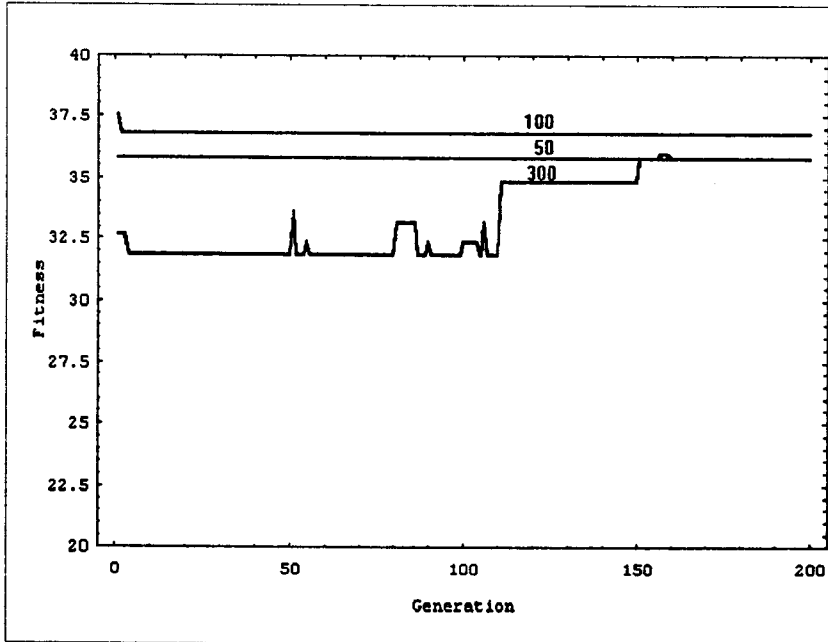
## 참 고 문 헌

1. Colin, R. R. (1993), "Using Genetic Algorithms with Small Populations", Proceedings of the fifth international conference on Genetic Algorithms, Morgan kaufmann Publishers, San Mateo, CA, pp 92 - 99.
2. De Jong, K. A., (1992), "Are Genetic Algorithms Function Optimizers?", Proceedings of the Second Conference on Parallel Problem Solving from Nature, North-Holland, pp 3 - 14.
3. Goldberg, D. E. (1989), "genetic Algorithms in search", Optimization and Machine learning, Addison-Wesley, Reading, MA.
4. Goldberg, D. E. (1985), "Optimal initial population size for binary-coded genetic algorithms", TCGA Report 85001, University of Alabama, Tuscaloosa.
5. Holland, J. H. (1975), "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor.

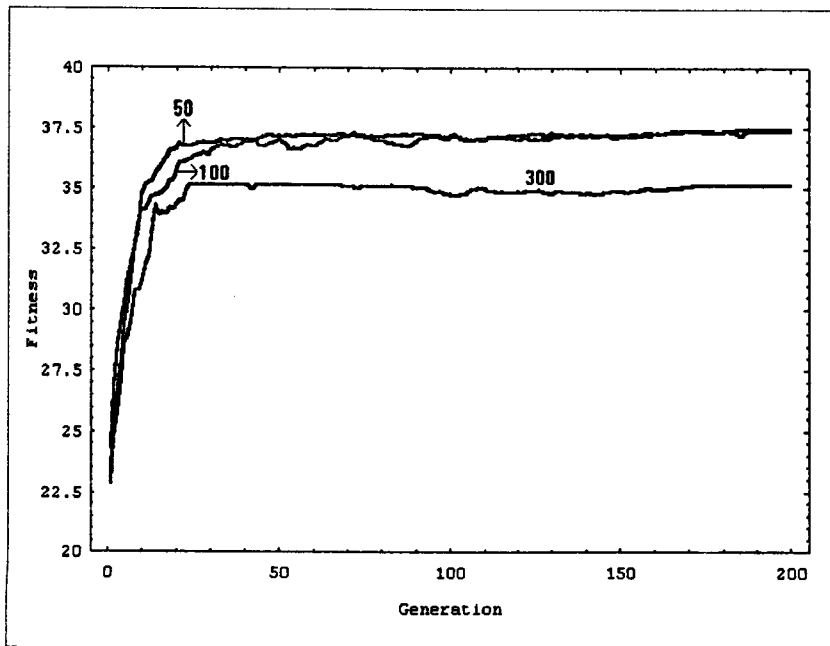


[그림 2] 개체군의 크기 50, 100, 300에 대한 수렴도 평가





[그림 3] 개체군의 크기 50, 100, 300에 대한 최대 적합도 평가



[그림 4] 개체군의 크기 50, 100, 300에 대한 평균 적합도 평가