

Genetic Algorithm

<http://raic.kunsan.ac.kr>

로보틱스 및 인공지능제어 연구실

Robotics & Artificial Intelligent Control Laboratory

2002 / Designed by RAIC LAB. All rights reserved.

목 차

1. 서론
2. 유전자 알고리즘의 기본 개념 및 용어
3. 이론적 기반
(스키마 이론, 스키마 정리)
4. 유전자 알고리즘의 동작 및 특성
5. 유전자 알고리즘의 구성 요소
(개체 표현 방법, 유전 연산자, 적합도 함수,
선택 메커니즘, 제어 파라미터)
6. 유전 알고리즘 응용분야
7. 모의실험(2)

서론

진화 알고리즘:

- 자연 세계의 진화 과정을 컴퓨터 상에서 시뮬레이션 함으로써 복잡한 실세계의 문제를 해결하고자 하는 계산모델.
- 구조가 간단하고 방법이 일반적이어서 응용범위가 매우 넓으며 적응적 탐색과 학습 및 최적화를 통한 공학적 문제 해결에 사용.
- 신경망, 퍼지 로직과의 결합으로 응용 범위가 넓어짐.
- 진화 알고리즘의 종류:
유전자 알고리즘(GA), 진화전략(ES), 진화프로그래밍(EP), 유전자 프로그래밍(GP)

Darwin's 진화론

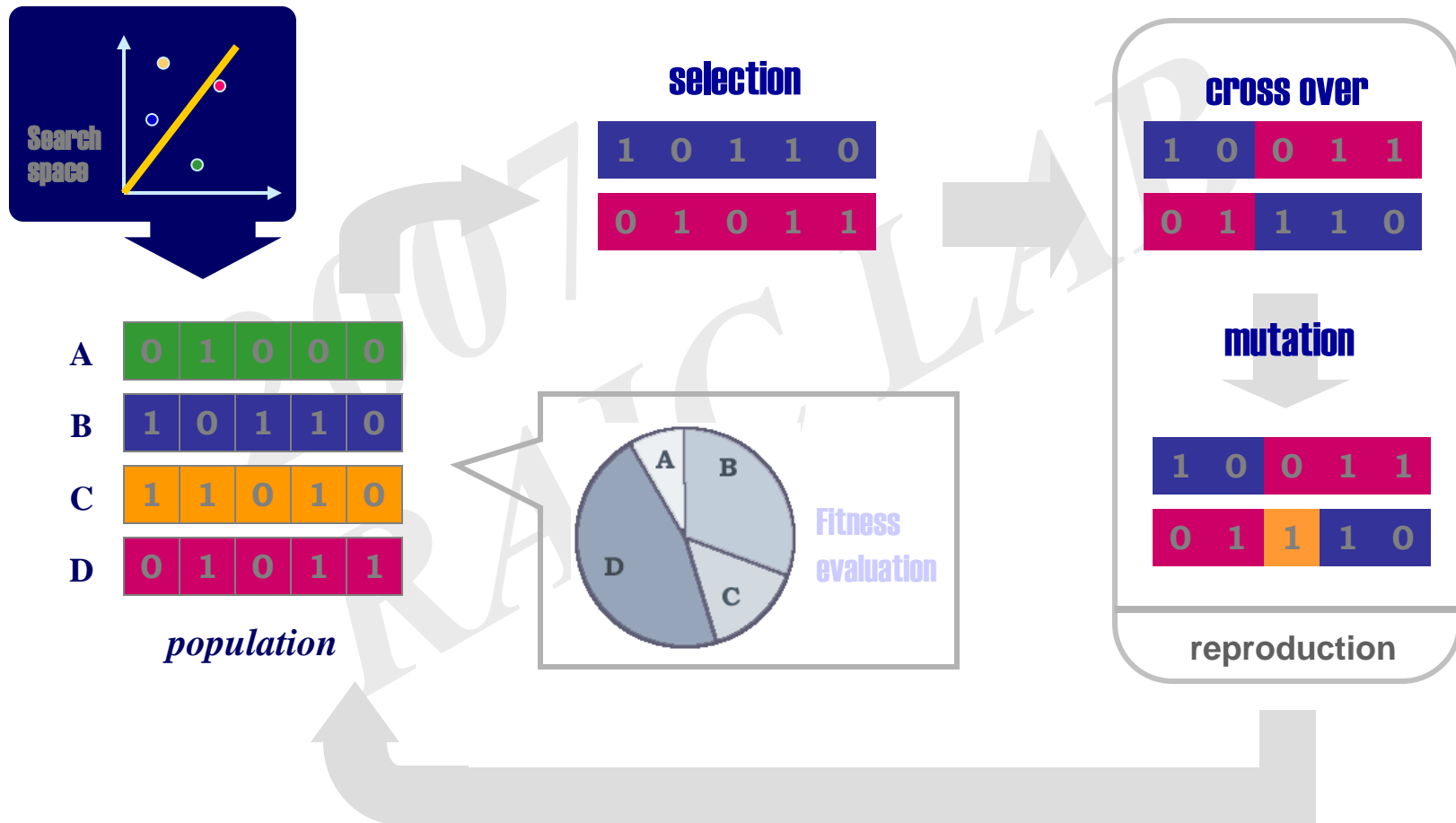
다산 → 생존경쟁 → 변이 → 자연선택 → 진화
e.g.) 똑똑한 토끼가 살아 남는다?



유전자 알고리즘의 기본 개념

- 유전적 계승과 다윈의 생존 경쟁(자연 도태설)이라는 자연 현상을 모델링한 확률적인 탐색 방법
- 1960년대 초 미시간 대학의 John Holland에 의해 연구, 제안됨
- 무작위 탐색 방법+ 방향성을 지닌 탐색 방법
- 진화과정: 어떤 세대(generation)를 형성하는 개체(individual)들의 집합, 즉 개체군(population)중에서 환경에 대한 적합도(fitness)가 높은 개체가 높은 확률로 살아남아 재생(reproduction)할수 있게 되며, 이때 교배(crossover) 및 돌연변이(mutation)로서 다음 세대를 형성

유전자 알고리즘의 구조



유전자 알고리즘

function GENETIC-ALGORITHM(*population*, FITNESS-FN) **returns** an individual

input: *population*, a set of individuals

FITNESS-FN, a function that measures the fitness of an individual

repeat

parents \leftarrow SELECTION(*population*, FITNESS-FN)

population \leftarrow REPRODUCTION(*parents*)

until some individual is fit enough

return the best individual in *population*, according to FITNESS-FN

% REPRODUCTION
= cross-over + mutation

유전자 알고리즘의 구성요소

개체 표현방법 (Encoding Scheme)

문제의 해가 될 가능성이 있는 것의 유전자적 표현방법

적합도 함수 (Fitness Function)

환경의 역할을 수행하는, 즉 “적합도”에 의해 해를 평가하는 평가함수

선택 메커니즘

해가 될 가능성이 있는 것들의 초기 개체집단을 만들어 내는 방법

유전 연산자 (Genetic Operators)

자손의 합성을 변화시키는 유전 연산자들

알고리즘 제어 파라메타 (Parameter Setting)

유전자 알고리즘이 사용하는 여러 가지 매개변수의 값
개체집단의 크기, 유전 연산자를 적용시키는 확률 등

개체표현방법 [Encoding Scheme]

여러가지 표현형태들

Binary Encoding

e.g.)

1	1	0	0	1	1	1	1
---	---	---	---	---	---	---	---

Permutation Encoding

e.g.)

1	5	3	2	6	8	4	7
---	---	---	---	---	---	---	---

Value Encoding

e.g.)

1.2	5.3	0.4	2.3	5	3.1	06	7.2
A	B	D	J	E	D	I	B

주로 0과 1의 Binary encoding을 사용

적합도 함수 [Fitness Function]

염색체의 해(solution)로서의 적합도를 평가

e.g.)



선택 메커니즘 [1/2]

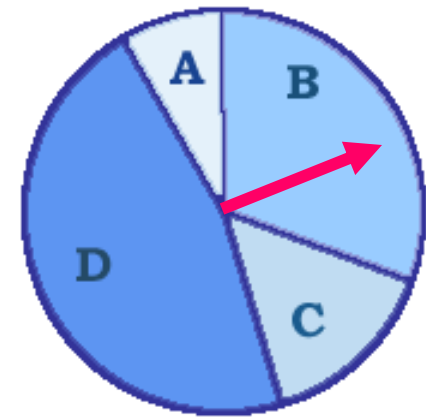
Roulette wheel selection

각 염색체의 적합도에 비례하는 만큼 roulette의 영역을 할당한 다음, roulette을 돌려 화살표가 가리키는 영역의 염색체를 선택

적합도가 높은 것은 선택될 확률이 그만큼 많고 적합도가 낮은 것은 선택될 확률이 상대적으로 낮다

e.g.)

	염색체	적합도	% of total
A	01000	64	5.5
B	10011	361	30.9
C	01101	169	14.4
D	11000	576	49.2



Roulette Wheel

선택 메커니즘 [2/2]

Expected-value selection

: 적합도에 대한 각 개체의 확률적인 재생 개체수를 구하여 선택

적합도	6	1	10	11	17	32	4	12	5	2
기대치	0.6	0.1	1	1.1	1.7	3.2	0.4	1.2	0.5	0.2
재생수	1	0	1	1	2	3	0	1	1	0

Ranking selection

적합도	32	17	12	11	10	6	5	4	2	1
순위	1	2	3	4	5	6	7	8	9	10
재생수	2	2	2	1	1	1	0	0	0	0

Tournament selection

Elitist preserving selection

유전 연산자 : 교배 연산자

두 부모의 염색체를 부분적으로 서로 바꾸어 자식의 염색체를 생성

Single point crossover



Two point crossover



Uniform crossover

Arithmetic crossover

유전 연산자 : 돌연변이 연산자

유전자를 일정한 확률로 변화시키는 조작



개체군의 다양성 유지

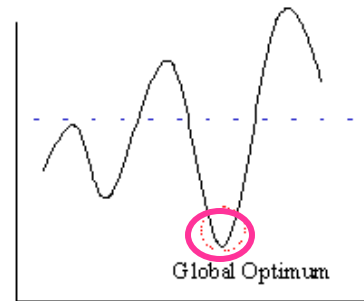
돌연변이가 없는 경우 초기 유전자 조합 이외의 공간을 탐색할 수 없어 초기 조합에 적절한 해가 없을 경우 원하는 해를 구할 수 없다.

→ local optimum 방지



Hill-climbing Method

cf.



GA Search Method

알고리즘 제어 파라미터

개체군의 크기 (Population size)

How many chromosomes are in population

Too few chromosome → small part of search space

Too many chromosome → GA slow down

Recommendation : 20–30, 50–100

교배 확률 (Probability of crossover)

How often will be crossover performed

Recommendation : 80% –95%

돌연변이 확률 (Probability of mutation)

How often will be parts of chromosome mutated

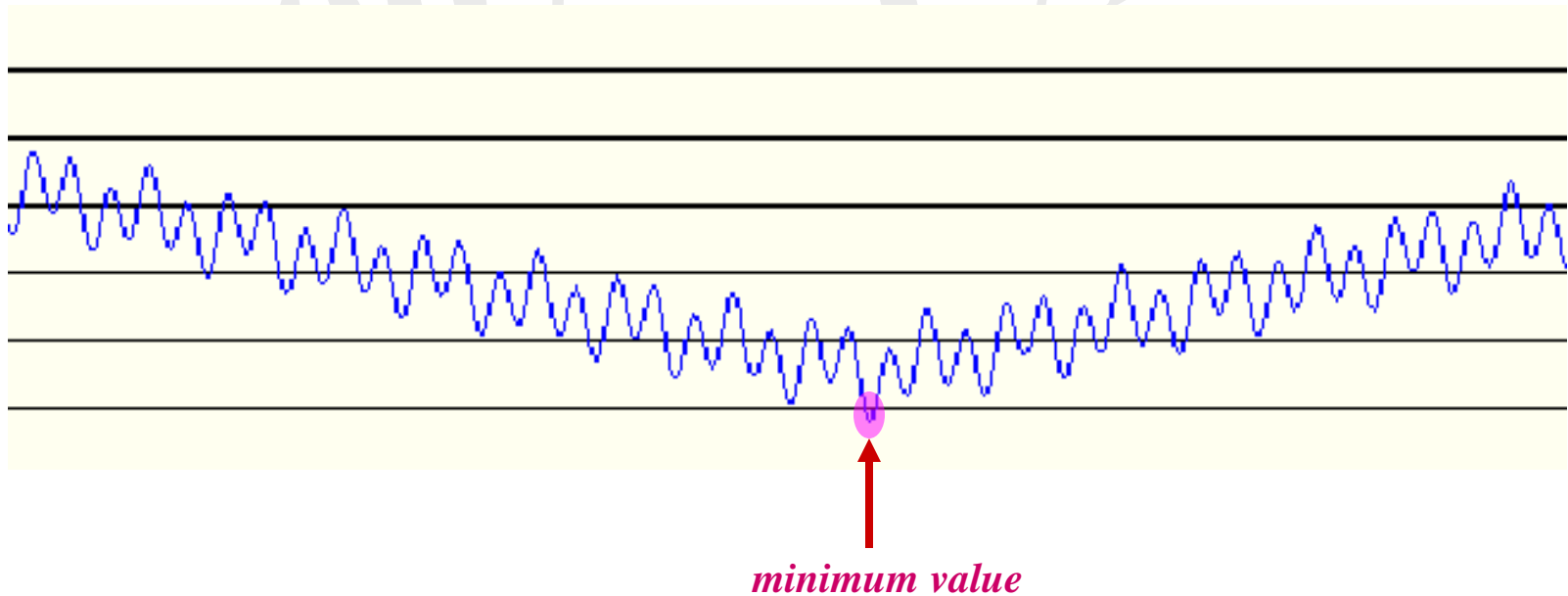
Recommendation : 0.5% – 1%

유전자 알고리즘 적용 예: Minimum of Function

Minimum of Function

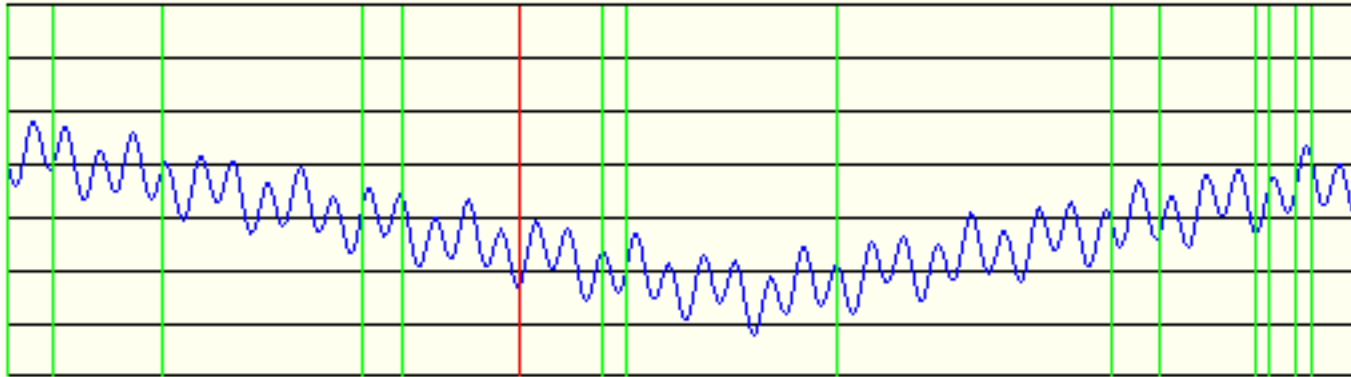
: 함수의 최소값을 찾는 문제

Example Function

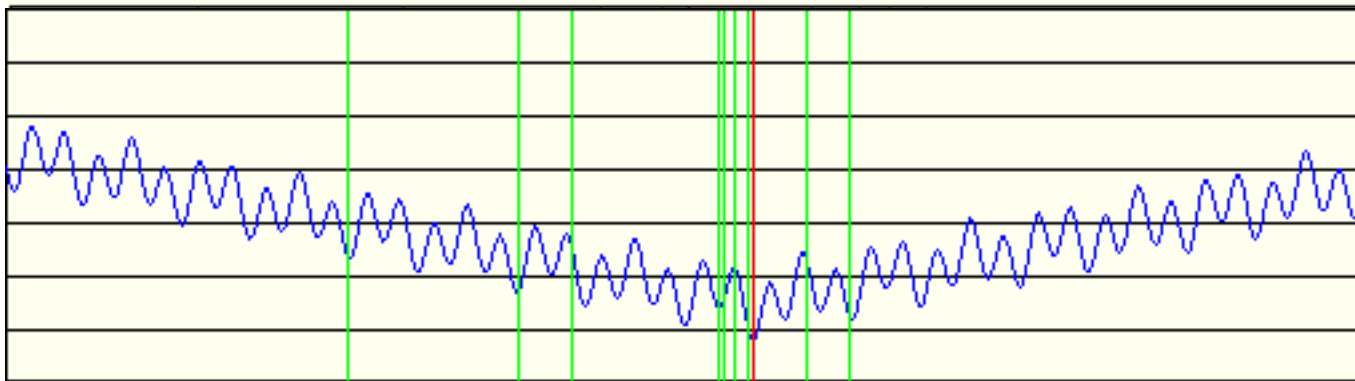


유전자 알고리즘 적용 예: Minimum of Function

- **Initial state**



- **state after evolutions**



유전자 알고리즘의 특성

- 어려운 비선형 문제에서 최적해를 찾는데 적합하다.
 - 선로 라우팅, 적응제어, 게임놀이, 인지 모델링,
 - 운송문제, 순회판매원문제, 최적제어문제, ...
- Local optima를 피해갈 능력이 있는 메커니즘을 가지고 있다.
- 해를 나타내는 파라미터를 염색체 형태로 코드화하여 이용한다.
- 점(point)이 아닌 다점(multi points) 탐색방법이다.
- 탐색에 적합도 함수(fitness function)을 이용하며 blind search를 한다.
- 결정론적인 규칙이 없고 확률적 연산자를 사용하여 수행된다.

유전자 알고리즘 용어

생물학	유전자 알고리즘
염색체(<i>chromosome</i>)	문자열(<i>string</i>)
유전자(<i>gene</i>)	특성(<i>feature</i>), 형질(<i>character</i>)
대립 유전자(<i>allele</i>)	특성 치(<i>feature value</i>)
유전자 좌(<i>locus</i>)	문자열의 위치(<i>string position</i>)
유전형(<i>genotype</i>)	구조체(<i>structure</i>)
표현형(<i>phenotype</i>)	파라미터 집합, 대체해, 디코드화를 위한 구조체
에피스타시스(<i>epistasis</i>)	비선형성(<i>nonlinearity</i>)

이론적 기반(1)

스키마 이론(schema):

- 정해진 스트링위치(유전자좌)에 같은 비트 값을 가진 모든 가능한 스트링들의 부분 집합
ex) $H = 1**0 = \{1000, 1010, 1100, 1110\}$
- 인스턴스(instance): 스키마에 의해 표현되는 각각의 스트링, 즉 1000, 1010, 1100, 1110은 $1**0$ 의 인스턴스이다
- 고정위채(fixed position): 0또는 1의 구체적인 값을 가진 유전자의 위치
- 스키마의 차수(order: $o(H)$): 고정위치의 개수
- 스키마의 길이(defining length: $\delta(H)$): 가장 왼쪽 고정위치와 가장 오른쪽 고정 위치 사이의 길이

이론적 기반(2)

스키마 정리(schema theorem):

- 스키마 성질:

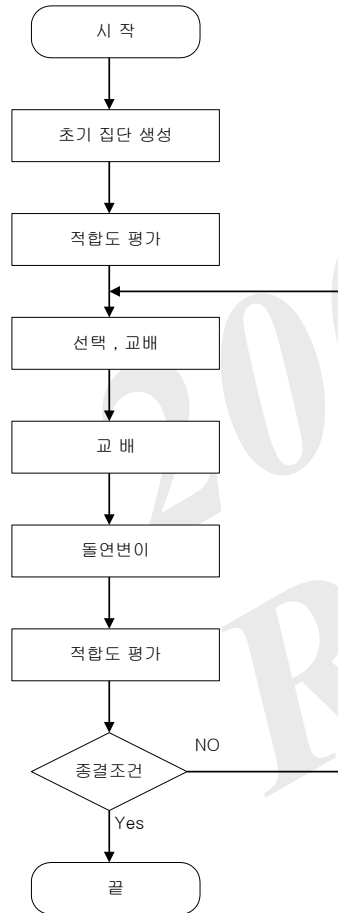
1개의 특정한 개체가 속하는 스키마의 개수는 2^l 개 (l :염색체 비트수)
크기가 n 인 개체군 내에서 얻을 수 있는 스키마의 최대 개수는 $n \cdot 2^l$
비트수가 l 인 코드에 대한 스키마의 수는 총 3^l 개 이다.

- 정리: 스키마의 길이가 짧고 차수가 낮으면 평균치 이상의 적합도를 가진 스키마(빌딩블록)의 다음세대에 선택될 인스턴스의 수는 지수 함수적으로 증가한다.

빌딩블록(building blocks hypothesis):

- 길이가 짧고 적합도가 높은 스키마를 가르킨다.

유전 알고리즘의 동작 및 특징(1)



SGA의 흐름도

```
Procedure SGA()
  initialize(Population);
  evaluate(Population);
  while not(terminal condition satisfied) do
    MatingPool = reproduce(Population);
    MutationPool = crossover(Matingpool);
    Population = mutation(MutationPool);
    evaluate(Population);
  end while
End Procedure
```

SGA의 알고리즘

유전자 알고리즘의 동작 및 특징(2)

특징:

- 파라미터를 코딩한 것을 직접 이용
 - 점(point)이 아닌 군(population) 탐색 방법
 - 탐색에 비용정보(fitness function)을 이용
 - Blind search를 함(미분값과 같은 추가적인 지식 불필요)
 - 결정론적인 규칙이 없고 확률적 연산자를 사용하여 수행
- 최적화 방법중 계산 의존 방법에 비해 전역적 해 구할 가능성이 높고, 효율적이다.

유전자 알고리즘 구성 요소(1)

개체 표현 방법:

- 이진 스트링으로 표현
- 코드화의 유효성 판단 조건:
 - 완전성(Completeness):해 후보는 모두 스트링으로 표현됨
 - 건전성(Soundness):GA공간상과 문제공간의 해 후보 대응
 - 비용장성(nonredundancy):염색체와 해 후보의 1:1 대응

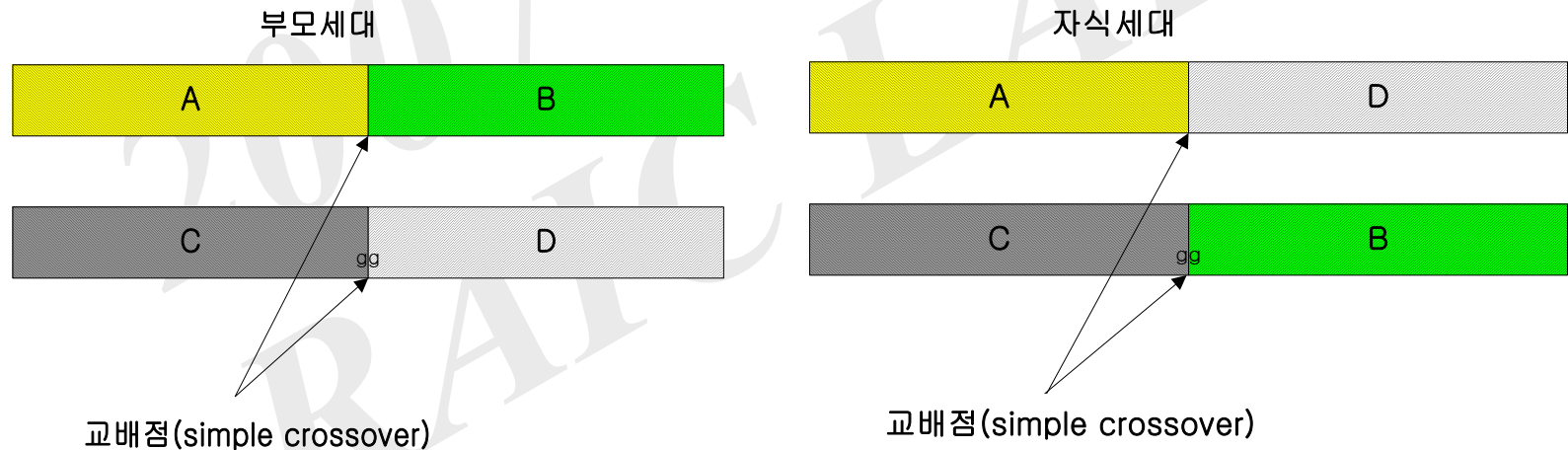
유전자 알고리즘 구성 요소(2)

유전 연산자(Genetic operator):

- 주 연산자:
 - 교배(crossover)
 - 돌연변이(mutation)
- 부 연산자: 문제에 따라서 사용됨
 - 역위(inversion)
 - 치환(displacement)
 - 중복(duplication)
 - 추가(addition)
 - 제거(deletion)

유전자 알고리즘 구성 요소(3)

교배(crossover): 단순, 복수점, 균일, 부분일치, 순서, 주기 교배

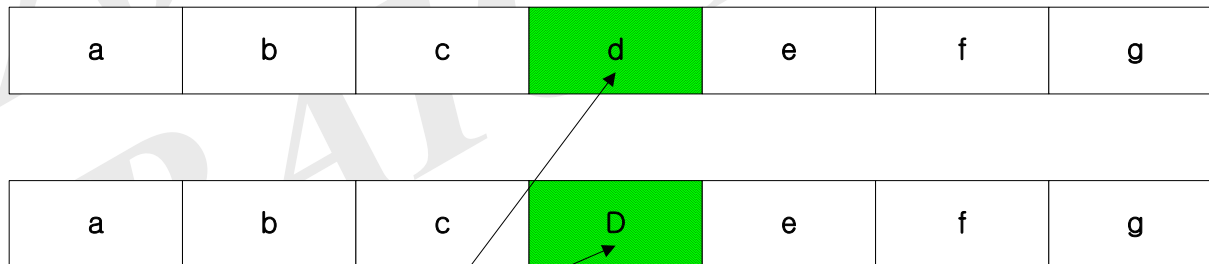


단순 교배 연산 전과 후

유전자 알고리즘 구성 요소(4)

돌연변이(mutation): 유전자좌의 유전자에 대하여 일정한 돌연변이 확률(P_m)을 적용하여 대립 유전자의 값으로 바꾼다.

- 극소적 랜덤 탐색의 일종
- 개체군의 잃어버린 유전형질 복구



돌연변이 적용 (4번째 비트)

돌연변이 적용

유전자 알고리즘 구성 요소(5)

적합도 함수(fitness function):

- 최적화 하고자 하는 함수는 각 개체의 적합도를 평가하는 기반
- raw fitness:표준화 되기 이전의 적합도 값.

raw fitness 표준화 방법:

- 선형 표준화(linear scaling)
- σ 절단(σ truncation)
- 거듭 제곱 표준화(power law scaling)

유전자 알고리즘 구성 요소(6)

선택(selection): 자연 선택 현상을 모델링

- 적합도 비례 선택(proportionate selection)
- 룰렛 선택(roulette selection)
- 기대치 선택(expected-value selection)
- 순위 선택(ranking selection)
- 토너먼트 선택(tournament selection)
- 엘리트 보존 선택(elitist preserving selection)

유전자 알고리즘 구성 요소(7)

- 적합도 비례 선택: 각 개체 s_i 의 적합도 $f(s_i)(>0), i=1,...,N$ 의 총합을 구해, 각 개체 s_i 의 선택 확률을 구한다

$$P_i = \frac{f(s_i)}{\sum_{j=1}^N f(s_j)}$$

- 토너먼트 선택: 개체군 중에서 일정한 개수의 개체를 임의로 선택하여 그중에 최고의 적합도를 가지는 개체를 다음 세대에 남기는 방법
다음 세대의 개체수가 모두 찰 때까지 반복적으로 계속된다.

유전자 알고리즘 구성 요소(8)

제어 파라미터:

- 미지 영역에 대한 탐색 : 랜덤 탐색(random search)
- 획득한 정보 : 기존의 등고선법(hill-climbing)과 유사함
- 제어 파라미터 3가지 :
개체군의 크기(M), 교배확률(P_c), 돌연변이 확률(P_m)
- 교배확률과 돌연변이 확률이 클 경우: 진화 초기 적합도가 높은 탐색공간을 찾는데 유리(exploration 능력 향상)
- M의 크기가 작으면 적합도 계산에 필요한 시간을 절약할 수 있는 반면 개체간의 다양성의 빠른 손실로 인한 최적의 해를 구하기 전에 수렴할 위험이 있다.

응용 분야

- ORP(operational Research Problem)
- 집적회로에서의 회로설계
- 통신망 관리의 최적화를 위한 설계
- 로봇팔의 PID 제어와 팔의 충돌 방지를 위한 설계
- 생산 라인에서 주어진 다수의 작업을 어떻게 효율적으로
분배하는가 하는 결정 문제
- 이산 이벤트 시스템 방법을 통한 승강기 제어
- 가스 파이프 라인의 최적화
- 순회 판매원 문제
- 신경망의 학습
- 퍼지 멤버쉽 함수의 최적화등등.

응 용 (모의실험1)

최고 피크(peak)점 문제:

- 비선형 시스템의 수학적 모델:

$$z = 3(1-x)^2 e^{-x^2-(y+1)^2} - 10\left(\frac{x}{5} - x^3 - y^5\right) e^{-x^2-y^2} - \frac{1}{3} e^{-(x+1)^2-y^2} \quad (1 \leq x, y \leq 5)$$

- 초기값:

generation = 30;

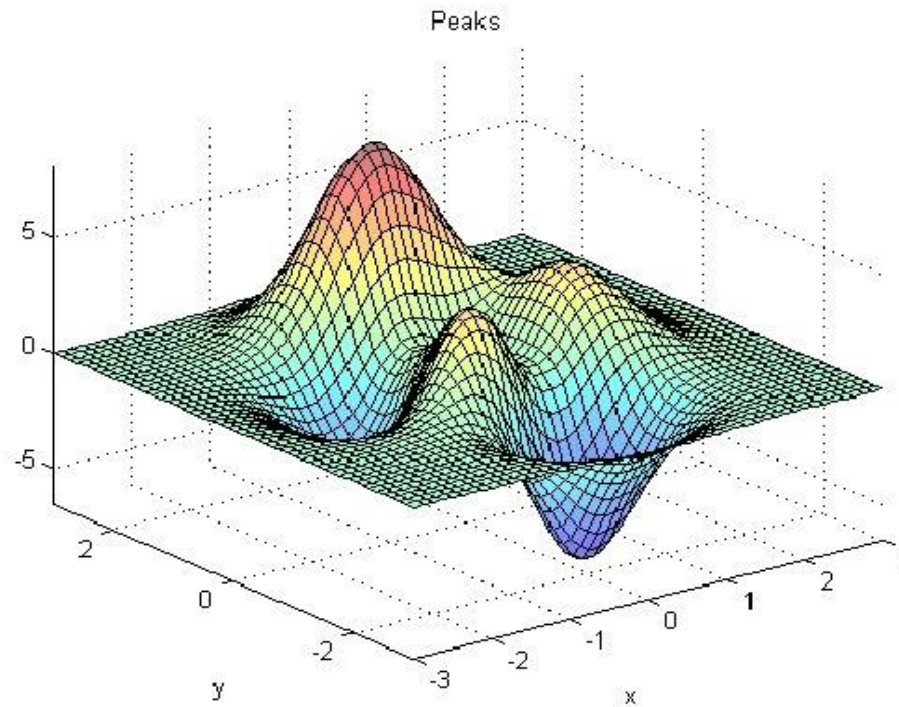
population size = 20;

crossover rate = 1.0;

mutation rate = 0.01;

응용 (모의실험1)

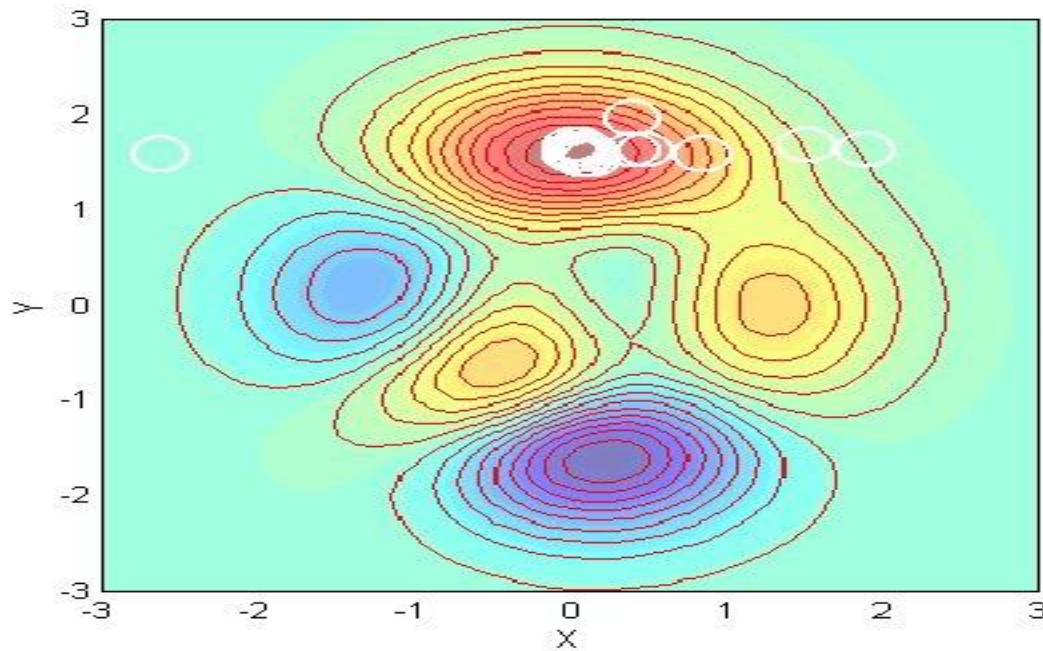
제어평면(control surface):



The Peaks function

응 용 (모의실험1)

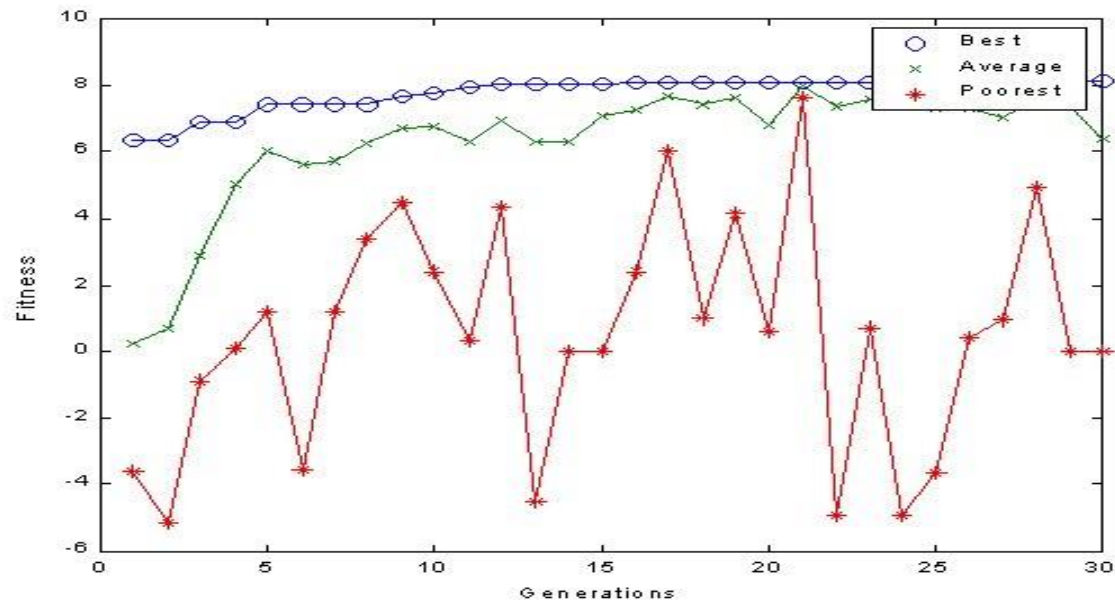
Using Gas to find the maximum of the peaks function



20 세대후 최고점 동정

응용 (모의실험1)

결과



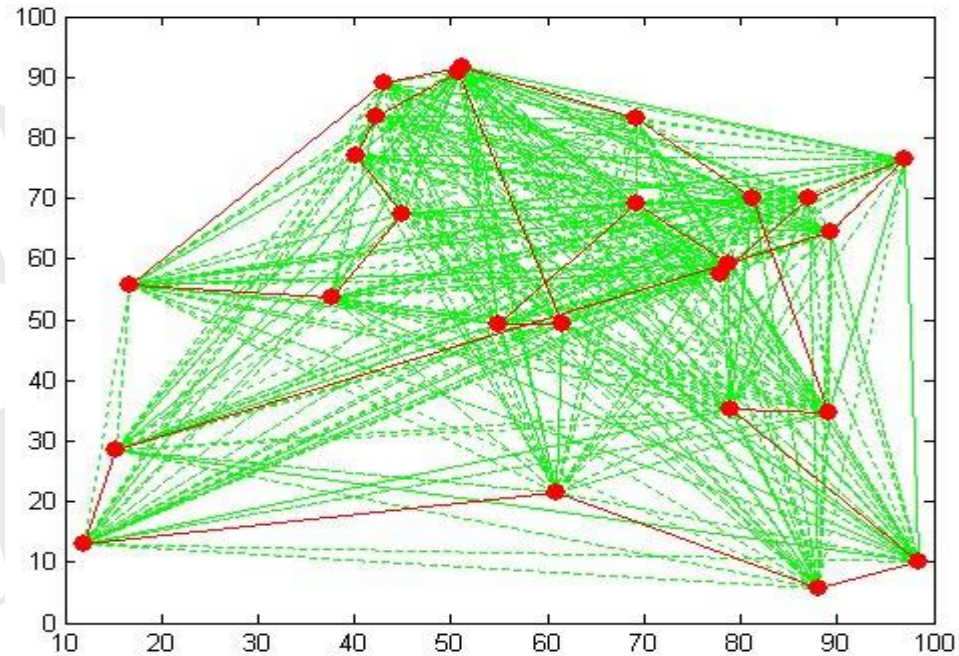
응 용 (모의실험2)

순회 판매원 문제:

- 25개 도시의 순회 최단 거리를 구하는 실험
- Generating function: Gaussian probability function
- Population size = 100;
- Crossover = cyclicxover;
- Mutation = inversionmutation;
- Select = 0.08;

응 용 (모의실험2)

각 위치에서의 최 단거리 탐색:



응 용 (모의실험2)

25개 도시의 최 단거리:

