Genetic Algorithms and Their Applications

Susmita Ghosh
Department of Computer Science & Engineering
Jadavpur University
Kolkata
INDIA
susmitaghoshju@gmail.com

What are EAs?

- Evolutionary Computing (EC) refers to the study of the foundations and applications of certain heuristic techniques based on the principles of natural evolution.

- The aim of designing evolutionary algorithms (EAs) is to mimic some of the processes taking place in natural evolution in algorithmic way.

- Four major categories of EAs (depending more on historical development rather than major functional differences).

  EC = GA $\cup$ GP $\cup$ ES $\cup$ EP

  GA = Genetic Algorithms, Holland

  GP=Genetic Programming, Koza

  ES = Evolution Strategies, , Rechenberg, Schweffel

  EP = Evolutionary Programming, Fogel

Basic metaphor

Although, biological evolution are not yet completely understood; strong experimental evidence is there to support the following points.

- Evolution is a process operating on ***chromosomes*** rather than on organisms.
- Natural selection is the mechanism that selects organisms which are ***well-adapted*** to the environment to reproduce more often than those which are not.
- Evolutionary process takes place during the reproduction stage that includes mutation (causing offspring to be different from parents) and recombination (combines chromosome segments of the parents to produce offspring).

Skeleton of an EA

Generate [P(0)] (initial population)

t $\leftarrow$ 0

WHILE NOT Termination-Criteria

DO

  Evaluate [P(t)] (population at time t)

  P' (t) $\leftarrow$ Select [P(t)]

  P''(t) $\leftarrow$ Apply Reproduction-Operators on [P'(t)]

  [P(t+1)] $\leftarrow$ Replace by [P(t), P''(t)]

  t $\leftarrow$ t + 1

END

RETURN Best-Solution

When should an EA be used?

- The search space is large
- The search space is known not to be perfectly smooth
- The search space is not unimodal / not well understood
- Fitness function is noisy
- Search time should be minimum

How GA Evolved?

- Genetic Algorithms have been developed by John Holland in 1960 and his students at the University of Michigan.
- He got inspiration from the Charles Darwin's (1859) "On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the struggle for Life".

Their goal was :

- To abstract and rigorously explain the adaptive process of natural systems.
- To design artificial systems software that retains the important mechanics of natural systems science.

What are genetic algorithms?

- GAs are adaptive computational procedures modeled on the mechanics of natural genetic systems. They act as biological metaphor and try to emulate some of the processes observed in natural evolution.

- Natural evolution operates on encoding of biological entities in the form of a collection of genes called a chromosome. Similarly, GAs operate on string representation of possible solutions (individuals/ chromosomes) containing the features.

- Selection obeys Darwinian survival of the fittest (determined by the objective function) strategy. Nature acts as environment, objective function plays the same role.

- Variation is introduced mainly through genetic operations like recombination (crossover) and mutation.

Similarities of natural evolution and GA terminologies

Natural evolution	**GA**
Chromosome	String
Gene	Feature
Allele	Feature value
Genotype	String structure
Phenotype	Decoded structure

Components of a GA

- Population of individuals
- Encoding/decoding (of individuals) mechanism
- Objective function & associated fitness evaluation criterion
- Selection procedure
- Genetic operators (recombination/crossover, mutation)
- Probabilities to perform genetic operations
- Replacement technique
- Termination conditions

Population

- A set of individuals (chromosomes) representing the parameter set

$x_1, x_2, \ldots., x_p$

$x_1 \rightarrow 0\ 0\ 0\ 0$

$x_2 \rightarrow 0\ 1\ 0\ 0$

$\ldots \qquad \ldots..$

$x_p \rightarrow 1\ 1\ 0\ 0$

chromosome: 0 0 0 0 0 10 0 … 1 1 0 0

- Each member refers to a coded *possible* solution
- Fixed/variable size
- Generally, initial population is chosen randomly

Encoding/decoding mechanism

Coding

- Converts parameter values into chromosome representation
- For the *continuous* valued parameters decimal to the binary conversion used.

  For example 13 == 01101 (for 5 bit representation)
- For a parameter having *categorical* values a particular bit position in the chromosome representation is set to 1 if it comes from that category.

  For example the parameter *marital status* can have values from {married, unmarried, divorced, widow}. So, unmarried == 0100 widow==0001
- These strings (representing the parameters of a problem) are concatenated to form a chromosome.

Decoding

- Decoding is the reverse of encoding.
- For continuous valued parameter the binary representation is converted to continuous value by the following formula

$$\text{Lower bound} + \frac{\sum_{i=0}^{\#bits\ -1} bit_i * 2^i}{2^{(\#bits)} - 1} * (\text{Upper bound - Lower bound})$$

01101==40+(13/31)*(60-40)=48.387

- For categorical valued parameters the value is found by consulting the range of the parameter.

  0001== widow

  0100== unmarried

Evaluation and selection

- A measure of chromosome's performance. More suitable strings should get high fitness values.

- Selection gives more chance to better fitted individuals (Mimics natural selection procedure)
- Popular selection techniques
  - Roulette wheel selection
  - Linear normalization selection
  - Tournament selection
  - Stochastic Universal Sampling

Roulette wheel selection

- Sum the fitness of all the chromosomes of the population. Call it ***total-fitness***.
- Generate a random number ***n*** in [**0**, ***total-fitness***]
- Return the first chromosome whose fitness when added to the fitness of the preceding population member is greater or equal to ***n***.

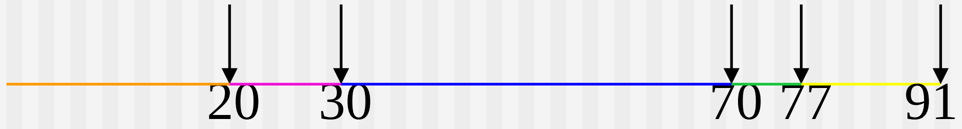
Example:

Let there be five chromosomes with fitness 20, 10, 40, 7, 14

Then ***total-fitness***=91.

Say, the random number drawn (***n***) is 45.

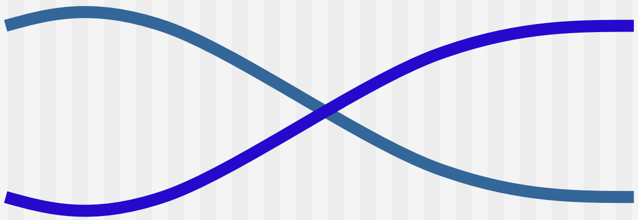
Select the 3rd chromosome (since 20+10+40 > 45).



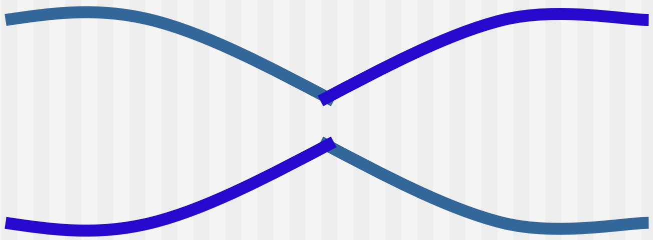
Recombination/crossover

- Exchange of information; exploitation
- Choose mating pairs (from the selected chromosomes).
- Check (using p_c) whether this pair should go for recombination or not. If yes, interchange chromosome segments using cross-sites.
  $\rightarrow$ one point, two point, multi point, uniform,...

One point crossover



before crossover



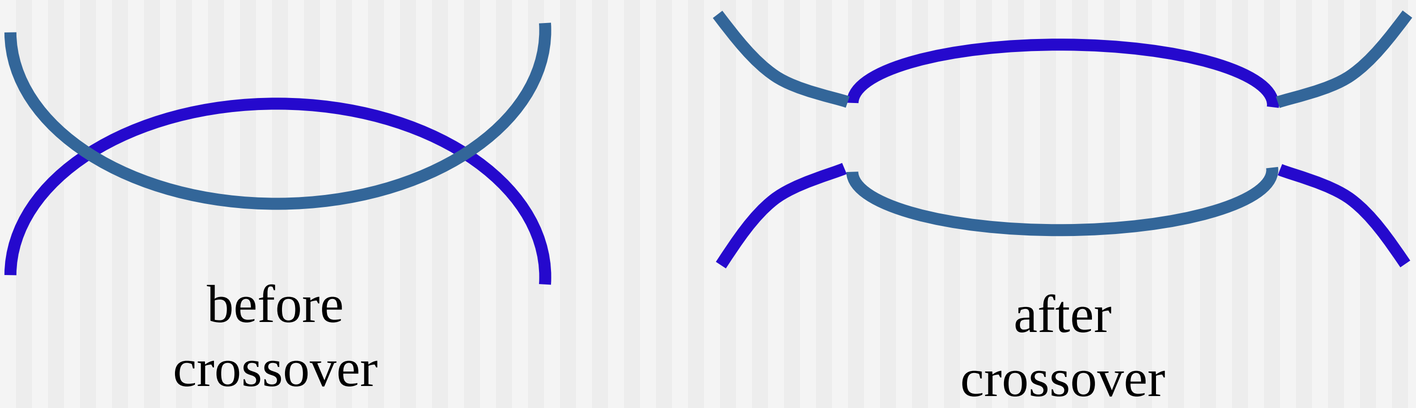
after crossover

- Parent1: x y x y x y x y child1: x y x y x b a b

  Parent2: a b a b a b a b child2: a b a b a y x y

Recombination (contd.)

Two point crossover



- parent1: xy xy xy xy child1: xy ab a y xy

  parent2: ab ab ab ab child2: ab xy x b ab

Mutation

- Introduces diversity, helps to regain lost genetic material
- Exploration
- Bit mutation.
- Check (using p_m) whether this bit should be mutated or not. If yes, flip the bit.
- 00**1**000 $\rightarrow$ 00**0**000

Probabilities to perform genetic operations

- May be fixed or made variable.
- p_c : 0.6 to 0.9 p_m : 0.001 to 0.01

Replacement techniques

- **Generational -** replaces all the individuals at a time
  - Create **N** children through reproduction
  - Replace the old population with these new individuals
- **Steady state -** replaces a few individuals at a time
  - Create **m** (<**N)** children through reproduction
  - Delete **m** members of the population to make room for them
  - Insert the children into the population

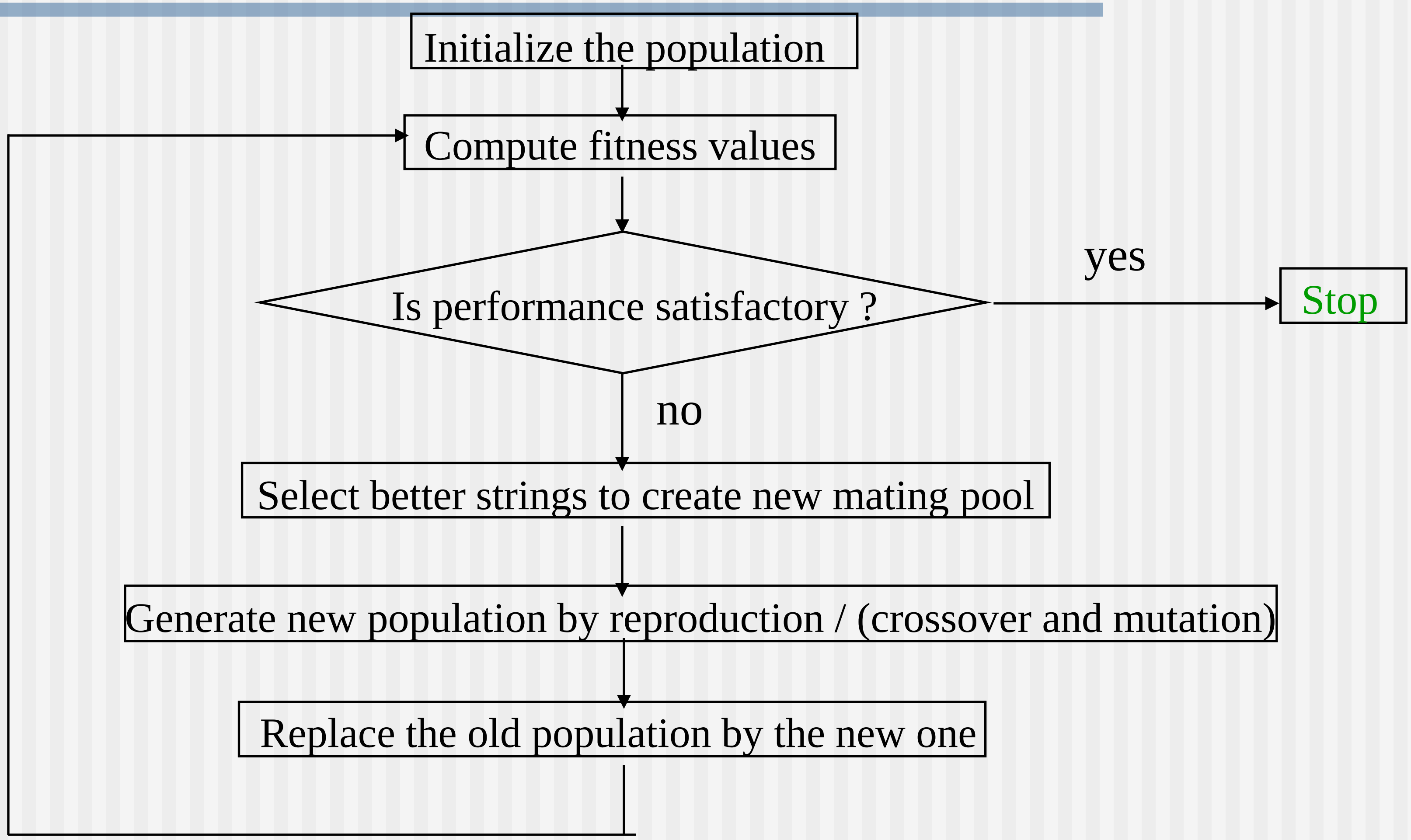
Terminating criterion

- Execute for a fixed number of generations/iterations.
- Until a string with a certain fitness value is located.
- Until the population attains a certain degree of homogeneity (most of the individuals become similar).

Elitism (*optional*)

Keeps track of /store the best solution obtained so far.

Flow diagram of a GA



Example 1

Maximize $f(x) = x^2$

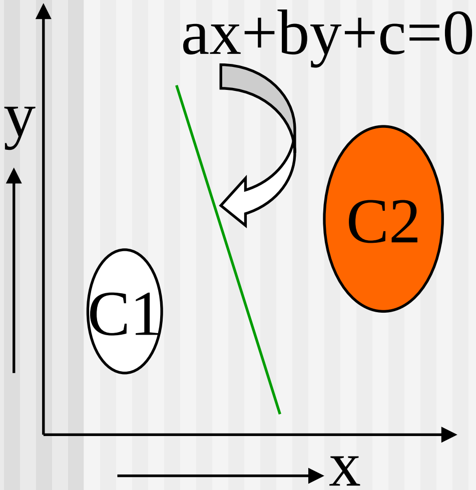
String#	Initial Pop	x-value	$f(x)=x^2$	Expected count (fi/f-av)	Actual count from r-wheel
1	01101	13	169	0.58	1
2	11000	24	576	1.97	2
3	01000	8	64	0.22	0
4	10011	19	361	1.23	1

sum = 1170; average = 293; max = 576

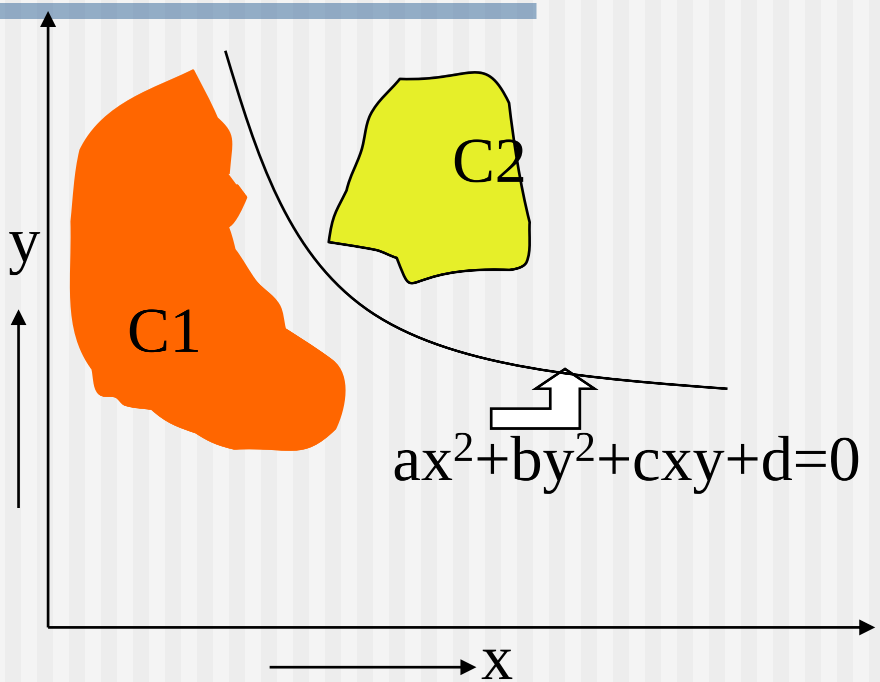
Mating pool (after reproduction)	Mate	cross-site (random select)	New Pop	x-value	$f(x)=x^2$
01101	2	4	01100	12	144
11000	1	4	11001	25	625
11000	4	2	11011	27	729
10011	3	2	10000	16	256

sum = 1754; average = 439; max = 729

Example 2 (classification)



Parameters : a,b & c



Parameters : a,b,c & d

class	P length	S length				
c1						p1
c2						p2
c1						p3

Distinguishing characteristics of GAs

- Multiple point searching (population based)
  - implicit parallelism
  - sometimes helps to prevent getting stuck to local optima
- Works on coded parameter set
  - resolution of the solutions can be controlled
- Search space may be discontinuous
- Uses probabilistic state transition rules
- Does not require any auxiliary information

Deviation from conventional GAs

- Distributed GA
- Parallel GA
- Structured GA
- Hybridization with neural networks and fuzzy logic
  - Fuzzy-GA
  - Neuro-fuzzy GA
  - Neuro-GA
- Hybridization with simulated annealing

Applications areas by domain

- Numerical, combinatorial and constrained optimization
- Scheduling, TSP, Graph Coloring
- Industrial design by parameterization
- Network design by construction, routing
- Automatic programming - evolves computer programs for specific tasks (Genetic Programming)
- Pattern recognition - classification, clustering, prediction
- Image processing --- segmentation, enhancement
- Data mining --- rule mining, clustering
- Bioinformatics – docking, prediction of structure of protein
- Economics - financial prediction
- Molecular biology – molecular conformation

Thank you