Fitness Evaluation and Selection

Debasis Samanta

Indian Institute of Technology Kharagpur

dsamanta@iitkgp.ac.in

12.03.2024

Important GA Operations

1. Encoding
2. Fitness Evaluation and Selection
3. Mating pool
4. Crossover
5. Mutation
6. Inversion
7. Convergence test

Important GA Operations

1. Encoding
2. **Fitness evaluation and Selection**
3. Mating pool
4. Crossover
5. Mutation
6. Inversion
7. Convergence test

GA Selection

- After deciding an encoding scheme, the second important things is how to perform selection from a set of population, that is, how to choose the individuals in the population that will create offspring for the next generation and how many offspring each will create.
- The purpose of selection is, of course, to emphasize **fittest** individuals in the population in hopes that their offspring will in turn have even higher fitness.

Selection operation in GAs

Selection is the process for creating the population for next generation from the current generation

To generate new population: Breeding in GA

- Create a mating pool
- Select a pair
- Reproduce

Fitness evaluation

- **In GA, there is a need to create next generation**
  - The next generation should be such that it is toward the (global) optimum solution
  - Random population generation may not be a wiser strategy
  - Better strategy follows the biological process: **Selection**
- **Selection involves:**
  - Survival of the fittest
  - Struggle for the existence
- **Fitness evaluation is to evaluate the survivability of each individual in the current population**

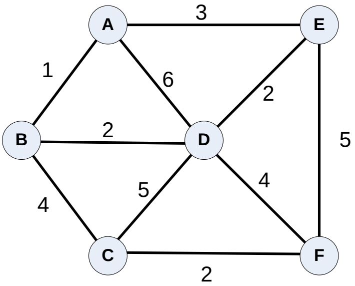
Fitness evaluation

How to evaluate the fitness of an individual?

- A simplest strategy could be to take the confidence of the value(s) of the objective function(s)
  - Simple, if there is a single objective function
  - But, needs a different treatment if there are two or more objective functions
    - They may be in different scales
    - All of them may not be same significant level in the fitness calculation

  . . . etc.

An example



P1:	C	B	A	D	F	E	11
P2:	A	B	D	C	E	F	19
P3:	A	C	B	F	E	D	16
P4:	F	C	D	B	E	A	12
P5:	C	F	D	A	B	E	10

Selection Schemes in GAs

Different strategies are known for the selection:

- **Canonical selection** (also called proportionate-based selection)
- **Roulette Wheel selection** (also called proportionate-based selection)
- **Rank-based selection** (also called as ordinal-based selection)
- **Tournament selection**
- **Steady-state selection**
- Boltzman selection

Canonical Selection

Canonical selection

- In this techniques, fitness is defined for the $i-th$ individual as follows.

  $$fitness(i) = \frac{f_i}{\bar{F}}$$

  where f_i is the evaluation associated with the $i-th$ individual in the population.

- $\bar{F}$ is the average evaluation of all individuals in the population size N and is defined as follows.

  $$\bar{F} = \frac{\sum_{i=1}^{N} f_i}{N}$$

Canonical selection

- In an iteration, we calculate $\frac{f_i}{\bar{F}}$ for all individuals in the current population.
- In Canonical selection, the probability that individuals in the current population are copied and placed in the mating pool is proportional to their fitness.

Note :

- Here, the size of the mating pool is $p\% \times N$, for some p.
- Convergence rate depends on p.

Roulette-Wheel Selection

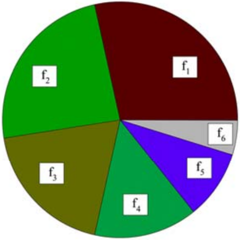
Roulette-Wheel selection

- In this scheme, the probability for an individual being selected in the mating pool is considered to be proportional to its fitness.
- It is implemented with the help of a wheel as shown.

$i \longrightarrow f_i$

$j \longrightarrow f_j$

fi > fj ⇒ i (Preferable than j)



[Wheel Game (Rotate and See the Pointer)]

Roulette-Wheel selection mechanism

- The top surface area of the wheel is divided into N parts in proportion to the fitness values $f_1, f_2, f_3 \cdots f_N$.
- The wheel is rotated in a particular direction (either clockwise or anticlockwise) and a fixed pointer is used to indicate the winning area, when it stops rotation.
- A particular sub-area representing a GA-Solution is selected to be winner probabilistically and the probability that the $i-th$ area will be declared as

$$p_i = \frac{f_i}{\sum_{i=1}^{N} f_i}$$

- In other words, the individual having higher fitness value is likely to be selected more.

Roulette-Wheel selection mechanism

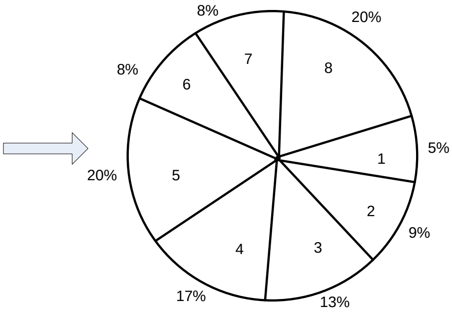
The wheel is rotated for N_p times (where $N_p = p\%N$, for some p) and each time, only one area is identified by the pointer to be the winner.

Note :

- Here, an individual may be selected more than once.
- Convergence rate is fast.

Roulette-Wheel selection mechanism: An Example

Individual	Fitness value	p_i
1	1.01	**0.05**
2	2.11	**0.09**
3	3.11	**0.13**
4	4.01	**0.17**
5	4.66	**0.20**
6	1.91	**0.08**
7	1.93	**0.08**
8	4.51	**0.20**



Roulette-Wheel selection : Implementation

Input: A Population of size N with their fitness values
Output: A mating pool of size N_p

Steps:

1. Compute $p_i = \frac{f_i}{\sum_{i=1}^{N} f_i}, \forall i = 1, 2 \cdots N$
2. Calculate the cumulative probability for each of the individual starting from the top of the list, that is
$P_i = \sum_{j=1}^{i} p_j$, for all $j = 1, 2 \cdots N$
3. Generate a random number say r between 0 and 1.
4. Select the j-th individual such that $P_{j-1} < r \leq P_j$
5. Repeat Step 3-4 to select N_p individuals.
6. End

Roulette-Wheel selection: Example

The probability that i-th individual will be pointed is

$$p_i = \frac{f_i}{\sum_{i=1}^{N} f_i}$$

Example:

Individual	p_i	P_i	r	T
1	0.05	0.05	0.26	I
2	0.09	0.14	0.04	I
3	0.13	0.27	0.48	II
4	0.17	0.44	0.43	I
5	0.20	0.64	0.09	II
6	0.08	0.72	0.30	
7	0.08	0.80	0.61	
8	0.20	1.0	0.89	I

p_i = Probability of an individual

r = Random Number between 0..1

P_i = Cumulative Probability

T=Tally count of selection

Roulette-Wheel selection

Following are the point to be noted:

1. The bottom-most individual in the population has a cumulative probability $P_N = 1$
2. Cumulative probability of any individual lies between 0 and 1
3. The i-th individual in the population represents the cumulative probability from P_{i-1} to P_i
4. The top-most individual represents the cumulative probability values between 0 and p_1
5. It may be checked that the selection is consistent with the expected count $E_i = N \times p_i$ for the i-th individual.

Does the selection is sensitive to ordering, say in ascending order of their fitness values?

Drawback in Roulette-Wheel selection

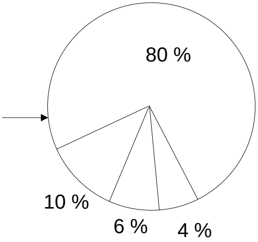
- Suppose, there are only four binary string in a population, whose fitness values are f_1, f_2, f_3 and f_4.
- Their values 80%, 10%, 6% and 4%, respectively.

What is the expected count of selecting f_3, f_4, f_2 or f_1?

Problem with Roulette-Wheel selection scheme

The limitations in the Roulette-Wheel selection scheme can be better illustrated with the following figure.

Individual (i)	Fitness (fi)	RW (Area)
1	0.4	80 %
2	0.05	10 %
3	0.03	6 %
4	0.02	4 %



The observation is that the individual with higher fitness values will guard the other to be selected for mating. This leads to a lesser diversity and hence fewer scope toward exploring the alternative solution and also premature convergence or early convergence with local optimal solution.

Rank-based Selection

Rank-based selection

- To overcome the problem with Roulette-Wheel selection, a rank-based selection scheme has been proposed.
- The process of ranking selection consists of two steps.
  1. Individuals are arranged in an ascending order of their fitness values. The individual, which has the lowest value of fitness is assigned rank 1, and other individuals are ranked accordingly.
  2. The proportionate based selection scheme is then followed based on the assigned rank.

Note:

- The % area to be occupied by a particular individual i, is given by

$$\frac{r_i}{\sum_{i=1}^{N} r_i} \times 100$$

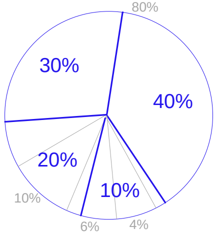
where r_i indicates the rank of $i - th$ individual.

- Two or more individuals with the same fitness values should have the same rank.

Rank-based selection: Example

- Continuing with the population of 4 individuals with fitness values: $f_1 = 0.40$, $f_2 = 0.05$, $f_3 = 0.03$ and $f_4 = 0.02$.
- Their proportionate area on the wheel are: 80%, 10%, 6% and 4%
- Their ranks are shown in the following figure.

Individual (i)	Fitness (fi)	RW (Area)	Rank	RS (Area)
1	0.4	80 %	4	40 %
2	0.05	10 %	3	30 %
3	0.03	6 %	2	20 %
4	0.02	4 %	1	10 %



It is evident that expectation counts have been improved compared to Routlette-Wheel selection.

Rank-based selection: Implementation

Input: A population of size N with their fitness values
Output: A mating pool of size N_p.

Steps:

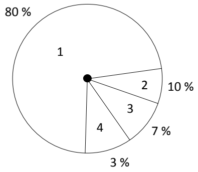
1. Arrange all individuals in ascending order of their fitness value.
2. Rank the individuals according to their position in the order, that is, the worst will have rank 1, the next rank 2 and best will have rank N.
3. Apply the Roulette-Wheel selection but based on their assigned ranks. For example, the probability p_i of the i-th individual would be

$$p_i = \frac{r_i}{\sum_{j=1}^{i} r_j}$$

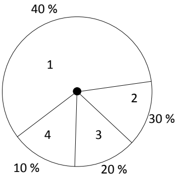
4. Stop

Comparing Rank-based selection with Roulette-Wheel selection

Individual	% Area	f_i	Rank (r_i)	% Area
1	80 %	0.4	4	40 %
2	10 %	0.05	3	30 %
3	7 %	0.03	2	20 %
4	4 %	0.02	1	10 %



Roulette-Wheel based on proportionate-based selection



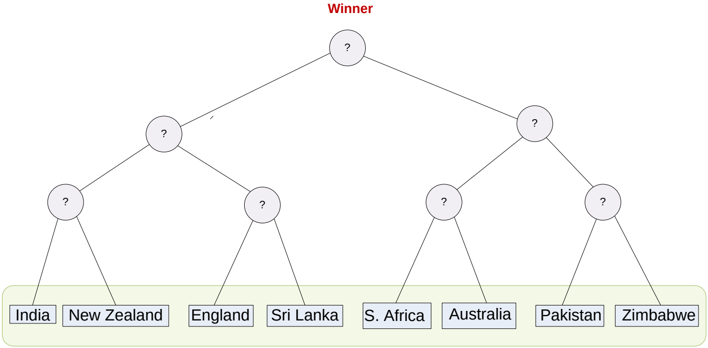
Roulette-Wheel based on ordinal-based selection

A rank-based selection is expected to performs better than the Roulette-Wheel selection, in general.

Tournament Selection

Basic concept of tournament selection

Who will win the match in this tournament?



Tournament selection

1. In this scheme, we select the tournament size n (say 2 or 3) at random.
2. We pick n individuals from the population, at random and determine the best one in terms of their fitness values.
3. The best individual is copied into the mating pool.
4. Thus, in this scheme only one individual is selected per tournament and N_p tournaments are to be played to make the size of mating pool equals to N_p.

Note :

- Here, there is a chance for a good individual to be copied into the mating pool more than once.
- This techniques founds to be computationally more faster than both Roulette-Wheel and Rank-based selection scheme.

Tournament selection : Implementation

The tournament selection scheme can be stated as follows.

Input : A Population of size N with their fitness values
Output : A mating pool of size $N_p (N_p \leq N)$

Steps:

1. Select N_U individuals at random ($N_U \leq N$).
2. Out of N_U individuals, choose the individual with highest fitness value as the winner.
3. Add the winner to the mating pool, which is initially empty.
4. Repeat Steps 1-3 until the mating pool contains N_p individuals
5. Stop

Tournament selection : Example

$$N = 8, N_U = 2, N_p = 8$$

Input :

Individual	1	2	3	4	5	6	7	8
Fintess	1.0	2.1	3.1	4.0	4.6	1.9	1.8	4.5

Output :

Trial	Individuals	Selected
1	2, 4	4
2	3, 8	8
3	1, 3	3
4	4, 5	5
5	1, 6	6
6	1, 2	2
7	4, 2	4
8	8, 3	8

If the fitness values of two individuals are same, than there is a tie in the match!! So, what to do????

Tournament selection

Note :

There are different twists can be made into the basic Tournament selection scheme:

1. Frequency of N_U= small value $(2, 3)$, moderate 50 % of N and large $N_U \approx N$.
2. Once an individual is selected for a mating pool, it can be discarded from the current population, thus disallowing the repetition in selecting an individual more than once.
3. Replace the worst individual in the mating pool with those are not winners in any trials.

Steady-State Selection

Steady-State selection algorithm

Given a sample of size, say N and decide a value $N_U < N$.

Steps :

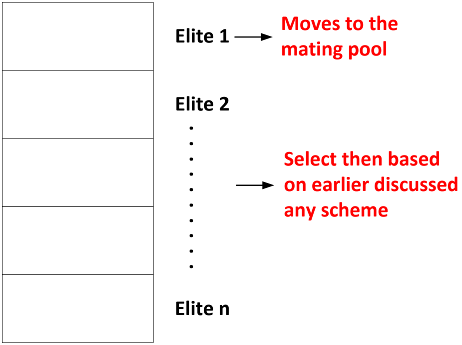
1. N_U individuals with highest fitness values are selected.
2. N_U individuals selected in Step 1 are added into the mating pool.
3. N_U Individuals with worst fitness values are removed from the sample.

This completes the selection procedure for one iteration. Repeat the iteration until the mating pool of desired size is obtained.

Elitism

Elitisms

- In this scheme, an elite class (in terms of fitness) is identified first in a population of strings.
- It is then directly copied into the next generation to ensure their presence.



Survey on GA selection strategies

- **Reference:**
  D. D. Goldberg and K. Deb,*”A comparison of selection schemes in foundation of GA”*, Vol. 1, 1991, Pg. 69-93
  Web link : K. Deb Website, IIT Kanpur

Comparing selection schemes

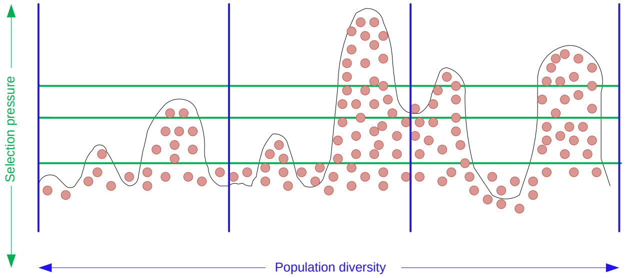
- Usually, a selection scheme follows Darwin's principle of ”Survival of the fittest”.
- In other words, a selection strategy in GA is a process that favours the selection of better individuals in the population for the matting pool (so that better genes are inherited to the new offspring) and hence search leads to the global optima.

There are two issues to decide the effectiveness of any selection scheme.

- Population diversity
- Selection pressure

Analyzing a selection schemes

- More population diversity means more exploration
- Higher selection pressure means lesser exploitation



Effectiveness of any selection scheme

Population diversity

- This is similar to the concept of exploration.
- The population diversity means that the genes from the already discovered good individuals are exploited while permitting the new area of search space continue to be explored.

Selection pressure

- This is similar to the concept of exploitation.
- It is defined as the degree to which the better individuals are favoured.

Effectiveness of any selection schemes

These two factors are inversely related to each other in the sense that if the selection pressure increases, the population diversity decrease and vice-versa. Thus,

1. **If selection pressure is HIGH**
   - The search focuses only on good individuals (in terms of fitness) at the moment.
   - It loses the population diversity.
   - Higher rate of convergence. Often leads to pre-mature convergence of the solution to a sub-optimal solution.
2. **If the selection pressure is LOW**
   - May not be able to drive the search properly and consequently the stagnation may occurs.
   - The convergence rate is low and GA takes unnecessary long time to find optimal solution.
   - Accuracy of solution increases (as more genes are usually explored in the search).

Analysis of different selection strategies

Selection Scheme	Population Diversity	Selection Pressure
Roulette-wheel selection (It works fine when fitness values are informally distributed)	• Low Population Diversity - Pre-mature convergence - Less Accuracy in solution	• It is with high selection pressure - Stagnation of Search
Rank Selection (It works fine when fitness values are not necessarily uniformly distributed)	• Favors a high population diversity - Slow rate of convergence	• Selection pressure is low - Explore more solutions
Tournament Selection (It works fine when population are with very diversified fitness values)	• Population diversity is moderate - Ends up with a moderate rate of convergence	• It provides very high selection pressure - better exploration of search space
Steady-state Selection	• Population diversity is decreases gradually as the generation advances	• Selection pressure is too low. - Convergence rate is too slow

Fine tuning a selection operator : Generation Gap

The generation gap is defined as the proportion of individuals in the population, which are replaced in each generation, that is,

$$G_p = \frac{p}{N}$$

Where N is the population size and p is the number of individuals that will be replaced.

Note that in steady-state GA, $p = 2$ and hence $G_p \approx 0$ for a large population whereas with the majority of the selection schemes $G_p \approx 1$

Fine tuning a selection operator : Generation Gap

To make the G_p a large value, several strategies may be adopted.

1. Selection of individuals according to their fitness and replacement at random
2. Selection of individuals at random and replacement according to the inverse of their fitness values.
3. Selection of both parents and replacement of according to fitness or inverse fitness.

Any question??