Study of the Various Selection Techniques in Genetic Algorithms

Parminder Kaur

Department of Computer Science, Dr. B.R. Ambedkar Government College Kaithal, Haryana, India

---***---

Abstract - This study examines a variety of selection strategies typically employed in contemporary genetic algorithms. This paper examines the classification of various selection procedures and their merits and downsides.

Key Words*: *Genetic Algorithms, Genetic operators, diversity

1. INTRODUCTION

Genetic algorithms are optimization strategies that resemble natural evolution and natural genetics. The notion of genetic algorithms (GAs) was created by John Holland in 1960, and was completely explored in his 1975 book titled "Adaptation in Natural and Artificial Systems" [1]. The objective of their research was to create software for artificial systems that retains the essential mechanics of natural systems. There are numerous uses for genetic algorithms in science, the business, and research and development at now.

In genetic algorithms, the possible solutions to the issue to be optimized consist of a population of individuals selected from the search space, typically at random. Individuals in this group are rated based on their adaption function ("fitness"). Afterwards, a selection method is employed to determine which individuals will serve as parents to the following generation. The new offspring will be produced by crossing and mutating these individuals. Finally, the next generation is being developed. This procedure is repeated until a specified condition is met.

The genetic algorithm notion can be explained as follows:

1. In genetic algorithms, a population of candidate solutions is chosen at random to optimize the problem; this is known as initialization.

2. Using simulations or computations, each prospective solution is evaluated to determine its utility.

3. Once the candidate's usefulness has been determined, they are assigned a fitness value.

4. Using the selection operator, only those individuals with a high probability, i.e., the suitable individuals, are eliminated. The candidates picked are the most likely to reproduce. This selection of prospects constitutes the mating pool.

5. During the reproduction phase, the offspring (new candidate solutions) are generated by applying the crossover operator to the parent solutions.

6. The group of newborns constitutes the new population.

7. If certain termination requirements are met, the evolution process is terminated; otherwise, it proceeds to step 2.

Objective function, genetic representation, and genetic operators are the three most crucial parts of a genetic algorithm. After these three are described, the generic genetic algorithm should function adequately. [2] The optimization of objective functions is a consideration. Different genetic operators include selection, crossover, and mutation.

2. Genetic Operators

The following three operators are used in genetic algorithms:

1. Selection This operator selects chromosomes for reproduction within a population. The more times a chromosome is likely to be picked for reproduction, the more fit it is. [3]

2. Crossover This operator arbitrarily selects a location and swaps the subsequences preceding and following it between two chromosomes to produce two children. [3]. These strings are chosen at random from the mating pool, and the crossover point is chosen at random as well.

3. Mutation This operator is used to occasionally introduce new features into the population's solution strings. This operator's purpose is to preserve the diversity.

3. Selection

Techniques of selection determine which solutions are to be retained and permitted to procreate and which ones must perish. The primary purpose of the selection operator is to highlight the good solutions and eliminate the undesirable ones from a population while maintaining the population size. This is what a selection operator does: "selects the best and discards the rest".

3.1 Classification of selection mechanism

The implementation of selection may be either deterministic or random, depending on the application in question. There are two types of selection algorithms: those with and those without replacement. In a selection process without replacement, each member of the population is considered for reproduction no more than once, and consequently can only become a parent once. The mating pool list provided by algorithms with replacement may contain numerous instances of the same person. The mating pool is always expressed as a list and never as a set for this reason. The selection algorithm's parameter n corresponds to the desired size of the mating pool. [4]

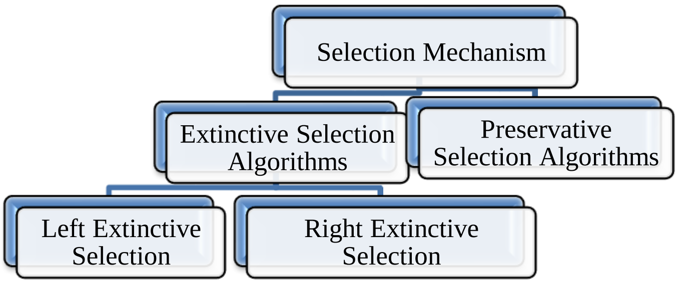

Fig 1 classification of selection mechanism

In genetic algorithms, algorithms that are extinct are referred to as generational. In generational algorithms, the future generation will consist exclusively of the offspring of the present generation, with no parent persons being retained.
Moreover, the extinctive algorithms are divided into left and right categories. In left extinctive selects, the best individuals are not permitted to reproduce in order to avoid the optimization process from converging too quickly. And in the correct extinctive algorithms, the least fit individuals are not let to reproduce in order to lessen selective pressure, as their offspring would disperse fitness too much.
In algorithms that use a preservative selection strategy, the population consists of the following population and its children [4]. The biological analogy for such algorithms is the fact that the lifespan of numerous species transcends one generation. Hence, parent and kid individuals compete for survival.

Let us adopt the following notation to further illustrate the concept of extinctive and preservation selection.

- λ denotes the number of offspring created and
- μ is the number of parent individuals.

Extinctive selection strategy is denoted as (μ, λ)-strategy and will create $\lambda > \mu$ child individuals from the μ patterns and only keep the μ best offspring while discarding the μ parents and the $\lambda - \mu$ worst children. [4]

In (μ + λ)-strategy, again λ children are generated from μ parents with also often $\lambda > \mu$. Then the parent and offspring populations are united (to a population of the size $\lambda+\mu$) and from this unison, the μ best individuals will survive. (μ + λ)-strategies are thus preservative. [4]

4. Selection techniques

The various selections used in modern genetic algorithm are as following:

4.1 Truncation selection

Truncation selection is the simplest and, arguably, least effective selection technique. Truncation selection merely keeps the x% of the population that is the fittest. These fittest individuals are reproduced in order to sustain the population number. For instance, if the population size is 80, we may select 25% of the fittest individuals, or 20 individuals, even though the population number is 80. Therefore, in order to maintain the population size, we must duplicate the four healthiest individuals.

Advantages

- This is an easy selection strategy to implement.

Disadvantage

• It can lead to precocious convergence as less suited candidates are eliminated and denied the chance to reproduce.

• It does not protect biodiversity.

•It cannot be utilized in situations when no person is superior to the others; it indicates that the set is optimal.

Individuals are ranked according to their fitness in truncation selection. Only the most qualified persons are chosen as parents. The truncation threshold Trunc is the truncation selection parameter. Trunc specifies the proportion of the population to be chosen as parents and can take values between 50% and 10%. People below the truncation threshold are unable to reproduce. [5]

4.2 Fitness proportionate selection

Every individual should have a chance of being picked for reproduction, but fitter candidates should have a greater chance of being chosen than weaker people. This is accomplished by making the survival probability of an individual a function of its fitness score. These methods are referred to as fitness-proportionate selection. [6]

4.3 Random selection

A random selection chooses elements at random. The possible antecedent fitness assignment processes and the individuals' objective values play no effect whatsoever. This effectively transforms the optimization method into a random walk, preventing it from following any gradient in the fitness landscape. Hence, random selection is not exclusively utilized, but can also act as a mating selection scheme alongside a different environmental selection. [4]

Advantages

• It is a straightforward method because it is based on uniformly distributed random numbers. • It maintains diversity.
• In cases when individuals must be selected from an optimal set, random selection is a viable option.

4.4 Roulette wheel selection

The most prevalent fitness-proportional selection method is known as Roulette Wheel Selection. Each member of the population is conceptually assigned a sector of a fictitious roulette wheel. In contrast to a true roulette wheel, the parts are proportionate to the individual's fitness, with the fittest candidate receiving the largest section and the weakest candidate receiving the smallest. The winner is then determined by spinning the wheel and selecting the occupant of the winning area. As many times as required, the wheel is spun to determine the complete set of parents for the next generation. If there is a particularly fit member of the population, we would anticipate it to have greater reproductive success than its weaker competitor. [6]

This selection approach is distinguished by the fact that it assigns each member of the present population a probability of selection proportionate to their fitness (x)

$$p(x) = \frac{f(x)}{\sum_{y-1}^{n} f(y)} \quad (1)$$

Where *n* denotes the population size.

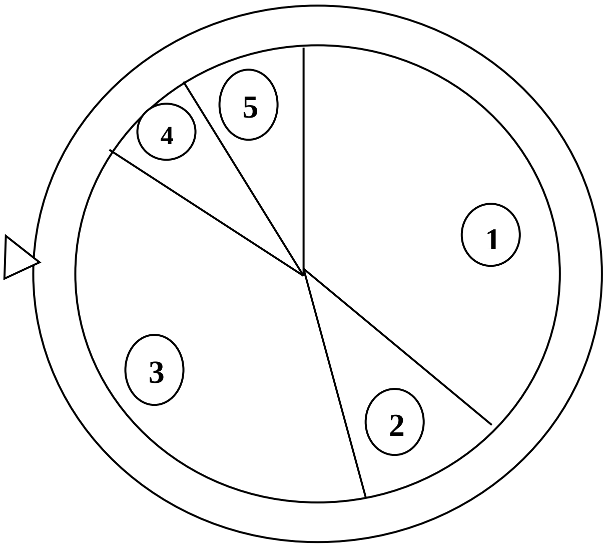

Figure 2 Roulette wheel

The normal method used is the roulette wheel (as shown in Figure 2 above). The following table lists a sample population of 5 individuals (a typical population of 400 would be difficult to illustrate).

No.	Chromosome	Fitness f(x)	%age
1	0001101011	6.82	31
2	1111011000	1.11	5
3	0100000101	8.48	38
4	1110100000	2.57	12
5	1110001011	3.08	14
		22.05	**100**

Table 1

$$f(x) = -\frac{1}{4}x^2 + 2x + 5 \quad (2)$$

These individuals consist of 10 bit chromosomes and are being used to optimize a simple mathematical function, *f(x)*. The fitness values are then taken as the function of x. We can see from the table (column Fitness f(x)) that individual No. 3 is the fittest and No. 2 is the weakest.. This gives the strongest individual a value of 38% and the weakest 5%.

These percentage fitness values can then be used to configure the roulette wheel. Figure 2 highlights that individual No. 3 has a segment equal to 38% of the area.

Advantages

- It is possible that one or more individuals are selected multiple times.

Disadvantages

- The risk of premature convergence of the GA to a local optimum, due to the possible presence of a dominant individual that always wins the competition and is selected as a parent.
- The fittest individual occupies the largest circumference on the wheel; hence the chances of its selection increases and the other individuals have few chances.

4.5 Stochastic Universal Sampling

Stochastic Universal Sampling is an alternative to roulette wheel sampling. Stochastic Universal Sampling assures that the observed selection frequencies of each individual correspond to the anticipated selection frequencies. If an individual occupies 4.5% of the wheel and 100 individuals are selected, we would anticipate that individual to be chosen between four and five times on average. Probabilistic Universal Sampling ensures this. The individual will be chosen either four or five times, but not three, zero, or one hundred times. The standard selection of roulette wheels does not provide this guarantee.

The Stochastic Universal Sampling technique consists of a single spin of the roulette wheel. This provides a starting position and the first individual to be chosen. The selection process then continues by advancing around the wheel in stages of equal size, where the size of each step is decided by the number of individuals to be chosen. Hence, if we are selecting 30 people, we will move 1/30 x 360 degrees with each selection. This does not necessarily imply that every candidate on the wheel will be chosen. Some weak individuals will get very thin slices of the wheel, and depending on their random beginning position, these may be walked over entirely. [6]

4.6 Rank selection

Rank Selection is comparable to fitness-proportionate selection, with the exception that the likelihood of selection is proportional to relative fitness rather than absolute fitness. In other words, it makes no difference if the fittest applicant is ten times or 0.001% fitter than the next fittest candidate. In both instances, the probability of selection would be same; all that matters is an individual's standing relative to others [6]. The choosing of ranks is a two-step process. Secondly, the list of individuals must be sorted, followed by some form of proportional selection based on the assignment values.

Advantages

- It preserves the diversity.

Disadvantages

- This method can lead to slower convergence, because there is not much difference between the best individuals.

4.7 Tournament selection

Tournament Selection is one of the most widely employed selection procedures in evolutionary algorithms. It is effective for a wide variety of issues.

In tournament selection, each member of the population is randomly paired with another. Each pair's fitness levels are compared. The fitter competitor advances to the next round, while the other is eliminated. This procedure is repeated until the number of winners equals the required number of parents. This final batch of victors is then paired to produce new individuals. [7]

Advantages

- It can be implemented very efficiently as no sorting of the population is required.

4.8 Sigma scaling

When all individuals in a population are extremely similar (the fitness variance is low), there are no significant fitness differences for natural selection to exploit, and evolution comes to an almost complete standstill. Consequently, the rate of evolution is dependent on the fitness variance within a population.

Similar to rank selection, Sigma Scaling aims to adjust selection pressure over time such that it is neither overly strong in early generations nor too weak once the population has stabilized and fitness disparities have diminished. In statistics, the Greek symbol Sigma denotes standard deviation, and the same is true here. The standard deviation of the population's fitness is utilized to scale the fitness scores such that the selection pressure remains roughly constant over the course of the evolutionary program's lifetime [6]. Under sigma scaling, the expected value of an individual is proportional to its fitness, the population mean, and the population standard deviation.

4.9 Steady State Selection

Some of the GAs are "generational, while in other schemes, subsequent generations overlap to some extent—a component of the previous generation is maintained in the new population. The proportion of new people in each generation is known as the "generation gap." In steadystate selection, just a small number of individuals are replaced in each generation: typically, a small number of the least suited individuals are replaced by offspring resulting from crossing and mutation of the fittest individuals. Steadystate GAs are frequently employed in evolving rulebased systems where incremental learning (and remembering what has already been learnt) is crucial and population members (rather than individuals) tackle the problem at hand collectively (rather than individually). [3]

5. Conclusion

This work examines, implements, and compares the relative performance of six well-known GA selection strategies using a set of four benchmark functions. There are two classifications for these methods: proportional and elitist. The first category employs a probability of selection proportionate to an individual's fitness. This family of selection strategies allows the population of candidate solutions to maintain genetic diversity throughout generations, which is a desirable trait that prevents the GA from convergent to local optima. On the other hand, these strategies tend to lengthen the convergence time. In contrast, selection methods in the second category choose only the best individual, which boosts convergence speed but carries the danger of convergence to local optima due to the loss of genetic diversity in the candidate solution population.

References

[1] John H. Holland, *Adaptation in Natural and Artifical Systems*.: The University of Michigan Press, 1975.
[2] Sailing Matthew Wall. Introduction to Genetic algorithms. [Online]. http://lancet.mit.edu/mbwall/presentations/IntroToGAs/
[3] Melanie Mitchell, *An Introduction to Genetic Algorithms*. London England: MIT Press, 1996.
[4] Thomas Weise, *Global Optimization Algoritms- Theory and Application*., 2007.
[5] Hartmut Pohlheim. (2006) Genetic and Evolutionary Algorithms Toolbox for use with matlab. [Online]. http://www.geatbx.com/docu/algindex-02.html
[6] Daniel W. Dyer. (2010) Evolutionary Computation in Java. [Online]. http://watchmaker.uncommons.org/manual/ch03.html#d0e717
[7] S. M. Shah, Mahesh Panchal Chetan Chudasama, "Comparison of Parents Selection Methods of Genetic Algorithm for TSP," in *International Journal of Computer Applications*, 2011, pp. 85-87.