International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-2, Issue-4, September 2012

A Comparative Analysis of Selection Scheme

Sonali Gandhi, Deeba Khan, Vikram Singh Solanki

Abstract — Selection scheme is an important part of genetic algorithms, which chooses a chromosome from the current generation's population for inclusion in the next generation's population, is the main subject of this paper. A selection operator selects the best chromosome using fitness function. Selection scheme is used to improve chances of the survivals of the fittest individuals. This paper recommends a number of selection (reproduction) methods most commonly used in genetic algorithms and analyzes them. These methods are: roulette wheel, rank selection, Boltzmann selection, tournament selection, steady state selection and elitism are compared on the basis of performance and takeover time computations .the analysis provides approximate or exact solutions. The paper recommends practical application and analyses a number of ways for more detailed analytical investigation of selection schemes.

Index Terms— Roulette Wheel, Rank selection, Boltzmann Selection, Tournament selection, Steady State Selection, Elitism.

I. INTRODUCTION

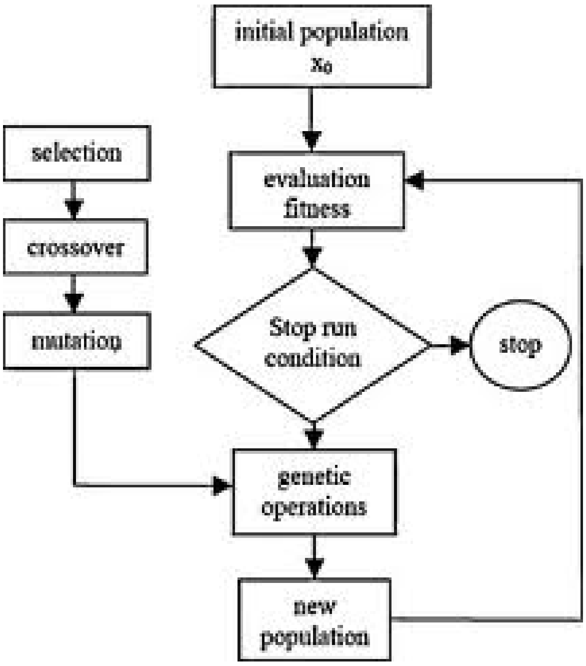
In the computer science field of artificial intelligence, a genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

Population is necessary. For better individuals, these should be from the fittest individuals of the previous population. There exist a number of selection operators in GA literature, but the essential idea in all of them is that the above average strings are picked from the current population and their multiple copies are inserted in the mating pool in a probabilistic manner.

II. ROLES OF SELECTION OPERATOR IN GA

Selection (or reproduction) is an operator that makes more copies of better strings in a new population. Selection is usually the first operator applied on population. Selection operator selects good strings in a population and forms a mating pool. This is one of the reasons for the selection operation to be sometimes known as reproduction operator. Thus, in selection operation the process of natural selection cause those individual that encode successful structures to produce copies more frequently.

To sustain the generation of a new population, the selection of individual in the current.



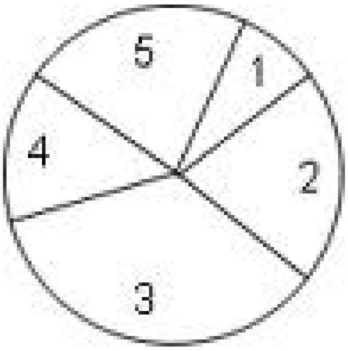
Flow Chart of Genetic Algorithm

III. METHODS OF SELECTION OF CHROMOSOMES

A. Roulette wheel selection
B. Rank selection
C. Steady state selection
D. Elitism
E. Boltzmann Selection
F. Tournament Selection

A. Roulette Wheel Selection

Parents are selected according to their fitness. The better the chromosomes are, the more chances to be selected they have. Imagine a roulette wheel where are placed all chromosomes in the population, every has its place big accordingly to its fitness function, like on the following picture.



Manuscript received on September, 2012.

Sonali Gandhi, Computer Science Engineering, Indore Institute of Science and technology, Indore, India.
Deeba Khan, Computer Science Engineering, Indore Institute of Science and Technology ,Indore, India.
Vikram Singh Solanki, Computer Science Engineering, Indore Institute of Science and Technology, Indore, India.

Then a marble is thrown there and selects the chromosome. Chromosome with bigger fitness will be selected more times.

This can be simulated by following algorithm.

Step 1: Find the fitness value (fv) for each chromosome in the population using Fitness Function.

Step 2: Calculate sum fitness (Sf) for all the chromosomes in the population:

$$\mathbf{Sf} = \sum_{i=1}^{n} \mathbf{fv}_i \quad (1)$$

Step 3: Calculate average fitness (Af) in the Population:

$$\mathbf{Af} = \frac{\mathbf{Sf}}{\mathbf{n}} \quad (2)$$

Step 4: Find the expected fitness (Ef) for each chromosome in the population:

$$\mathbf{Ef}_i = \frac{\mathbf{fv}_i}{\mathbf{Af}} \quad (3)$$

Step 5: Calculate sum expected fitness (Sum Ef) for all the chromosomes in the population:

$$\mathbf{SumEf} = \sum_{i=1}^{n} \mathbf{Ef}_i \quad (4)$$

Step 6: Generate random number (G) in the range [0,SumEf]

$$\mathbf{G} = \mathbf{Rnd}()\,mod\,\mathbf{SumEf} \quad (5)$$

Step 8: Go to step 6, repeat n times, where n is a population size.

i. Time Complexity Of Roulette Wheel

The simplest implementation (and one of the earliest to be used) is to simulate the spin of a weighted roulette wheel (Goldberg, 1989a). If the search for the location of the chosen slot is performed via linear search from the beginning of the list, each selection requires O(n) steps, because on average half the list will be searched. Overall, roulette wheel selection performed in this method requires 0(n2) steps, because in a generation n spins are required to fill the population. Roulette wheel selection can be hurried somewhat, if a binary search (like the bisection method in numerical methods) is used to locate the correct slot. This requires additional memory locations and an O(n) sweep through the list to calculate cumulative slot totals, but overall the complexity reduces to O(n log n), because binary search requires O(log n) steps per spin and n spins.

B. Rank Selection

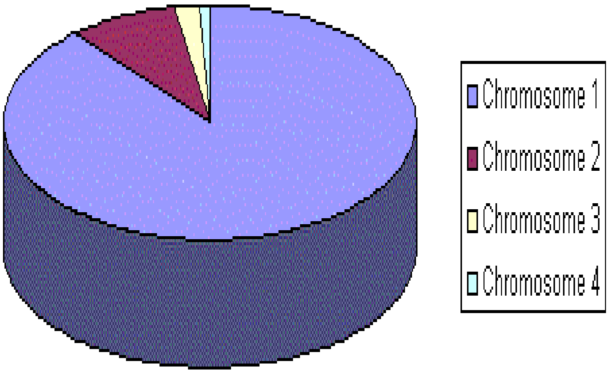
The previous selection will have problems when the finesses differ very much. For example, if the best chromosome fitness is 90% of the entire roulette wheel then the other chromosomes will have very few chances to be selected.

Rank selection first ranks the population and then every chromosome receives fitness from this ranking. The worst will have fitness 1, second worst 2 etc. and the best will have fitness N (number of chromosomes in population).

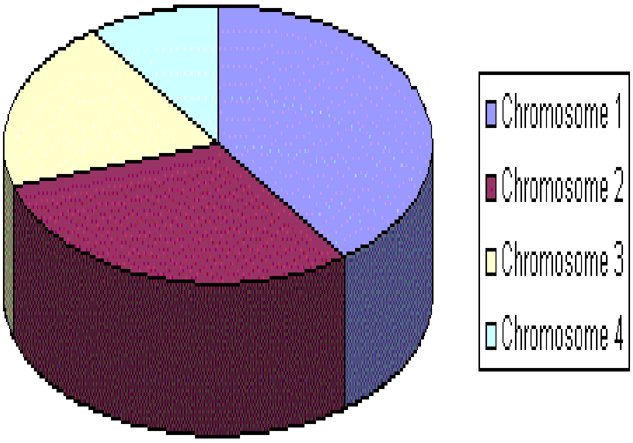
Now probability of population Pi can be calculated in Rank selection through following formula:

Pi=(Rank of Pi) / (Total sum of individual Rank)

You can see in following picture, how the situation changes after changing fitness to order number.



(a)Situation before ranking (graph of finesses)



(b) Situation after ranking (graph of order numbers)

After this all the chromosomes have a chance to be selected. But this method can lead to slower convergence, because the best chromosomes do not differ so much from other ones.

i. Time Complexity Of Rank Selection

Ranking is a two-step process. First the list of individuals must be sorted, and next the assignment values must be used in some form of proportionate selection. The calculation of the time complexity of ranking requires the consideration of these separate steps. Sorting can be performed in O(n log n) steps, using standard techniques. Thereafter, we know from previous results that proportionate selection can be performed in something between O(n) and O(n2). Here, we will assume that a method no worse than O(n logn) is adopted, concluding that ranking has time complexity O(n logn).

C. Steady State Selection

This is not particular method of selecting parents. Main idea of this selection is that big part of chromosomes should survive to next generation. GA then works in a following way. In every generation are selected a few (good - with high fitness) chromosomes for creating a new offspring. Then some (bad - with low fitness) chromosomes are removed and the new offspring is placed in their place. The rest of population survives to new generation.

i. Time Complexity Of Steady State Selection

The time Complexity of steady State Selection is O(n logn).

D. Elitism

Idea of elitism has been already introduced. When creating new population by crossover and mutation, we have a big chance, that we will loose the best chromosome.

The idea here is to arrange the chromosomes in the decreasing order according to their fitness values. Then apply the selection with each two chromosomes in the arranged set. In this way, Genetic Algorithm will be applied between strong chromosomes or between weak chromosomes. This means there is no chance to apply Genetic Algorithm between weak and strong chromosomes.

Elitism is name of method, which first copies the best chromosome (or a few best chromosomes) to new population. The rest is done in classical way. Elitism can very rapidly increase performance of GA, because it prevents losing the best found solution.

i. Time Complexity Of Elitism

The time Complexity of Elitism is O(MN2), where M is a number of objectives and N is population size.

E. Boltzmann Selection

Simulated annealing is a method of function minimization or maximization .This method simulates the process of slow cooling of molten metal to achieve the minimum function value in a minimization problem. The cooling phenomenon is simulated by controlling a temperature like parameter introduced with the concept of Boltzmann probability distribution so that a system in thermal equilibrium at a temperature T has its energy distributed probability according to $P(E)=\exp(-E/kT)$

Where 'k' is Boltzmann constant. This expression suggests that a system at a high temperature has almost uniform probability of being at any energy state, but at a low temperature it has a small probability of being at a high energy state. Therefore, by controlling the temperature T and assuming search process follows Boltzmann probability distribution, the convergence of the algorithm is controlled.

i. Simulated Annealing And Genetic Algorithm

The connections between genetic algorithms and simulated annealing have been explored elsewhere, but only loosely. Davis and Steenstrup acknowledge the procedures' common natural roots, but it is odd that this lead essay from a collection of papers ent itled *Genetic Algorithms and Simulated Annealing* chooses only to describe each method separately, largely ignoring connections between them. Sirag and Weisser get somewhat closer to the heart of the matter in their study of "thermally" motivated adaptation of mutation and crossover probabilities; however, tighter connections may b drawn in a straightforward manner.

Simulated annealing contains three basic elements:

1. Probabilistic acceptance (Metropolis or logistic forms),
2. Neighborhood exploration, and
3. A cooling schedule that respects thermal equilibrium.

On the other hand, genetic algorithms may consist of any number of different operators:

1. Probabilistic selection (via reproduction or replacement),
2. Recombination,
3. Recoding
4. Expression,
5. Diversity generation (mutation or other)

ii. Time Complexity Of Boltzmann Selection

The time Complexity of Boltzmann Selection is O(n2).

F. Tournament Selection

GA uses a strategy to select the individuals from population and insert them into a mating pool. Individuals from the mating pool are used to generate new offspring, which are the basic for the next generation. As the individuals in the mating pool are the ones whose genes will be inherited by the next generation, it is desirable that the mating pool consists of good individuals. A selection strategy in GA is simply a process that favors the selection of better individuals in the population for the mating pool.

Conditions for tournament selection are:

1. The total no of matches will be equal to number of teams.
2. No team can play more than two matches.

i. Time Complexity Of Tournament Selection

The time Complexity of Tournament Selection is O(n).

IV. CONCLUSION

This paper has compared the expected behavior of six selection schemes on the basis of their performance and takeover time computation. Roulette wheel is found to be extremely slower than other five schemes. Tournament Selection and Rank Selection are found to be best in different situations. Also Tournament Selection has the better convergence criteria. Although the actual performance is very much dependent on the different criteria at the time of selection of chromosomes, these criteria could be anything like type of population, type of encoding used, number of chromosomes in population etc.

REFERENCES

[1] Goldberg, D. E. (1990). A note on Boltzmann tournament selection forgenetic algorithms and population-oriented simulated annealing. Complex Systems, .

[2] Ankenbrandt, C. A. (1990). An extension to the theory of convergence and a proof of the time complexity of genetic algorithms (Technical Report CS/CIAKS- 90-0010/TU) New Orleans: Center for Intelligent and Knowledge-based Systems, Tulane University.

[3] Introduction to Genetic Algorithms: http://www.obitko.com/tutorials/genetic-algorithms/selection.php

[4] VOL. 3, NO.7, July 2012 ISSN 2079-8407 Journal of Emerging Trends in Computing and Information Sciences**:** Comparison of Selection Methods and Crossover Operations using Steady State Genetic Based Intrusion Detection System Firas Alabsi,Reyadh Naoum Middle East University(2012)

[5] Baker, J. E. (1987). Reducing bias and inefficiency in the selection algorithm. Proceedings of the Second International Conference on Genetic Algorithms, 14-21.

[6] David E. Goldberg and Kalyanmoy Deb Department of General Engineering University of Illinois at Urbana-Champaign 117 Transportation Building 104 South Mathews Urbana, IL 61801-2996.A Comparative Analysis of Selection Schemes used in Genetic algorithm.

[7] Www2.chemie.uni-erlang... Genetic Algorithm : Flowchart of the hybrid algorithm

[8] Fundamentals of Genetic Algorithms www.myreaders.info09_Genetic_Algorithms

[9] Introduction to Soft Computing (Neural Network , Fuzzy Logic and Genetic Algorithm):Ashish Kumar Sharma (Dhanpat Rai & co.)

[10] A Note on Boltzmann Tournament Selection for Genetic Algorithms and Population-Oriented Simulated Annealing .David E. Gold berg *University* of *Illinois at Urbana-Champaign, Urbana IL 61801 USA*

[11] Davis and M. Steenstrup, "Genetic algorithms and simulat ed annealing: An overview." In *Genetic Algorithms* and *Simulated Annealing,* 1. Davis, ed. (Pitman, London, 1989)

IJSCE International Journal of Soft Computing and Engineering www.ijsce.org Exploring Innovation