Evolution of Fitness-value Distribution: A Visual and Quantitative Investigation in Genetic Algorithm

Sajalendu Dey[1]

[1]Lindenwood University, St. Charles, Missouri

Abstract

This paper attempts to describe the results of exploration into the nature of evolution of the fitness –value distribution of the non-binary chromosomes over the generations of the population. Both qualitative and quantitative methods have been used to describe the results.

Keywords: Genetic Algorithm, Fitness Function, Frequency Distribution, Function Optimization.

1. Introduction

Originally developed by John Holland [1], genetic algorithms (GAs) have come to be recognized as powerful search algorithms. The workings of these algorithms are based on the principle of natural genetics that emphasizes survival of the fittest. In these algorithms, a potential solution is structured as a function whose parameters are represented as a population of individuals. These individuals are made to undergo a sequence of unary transformations via mutations, and higher order transformations via cross-over. The survival of these individuals are processed via a selection scheme which is, naturally, biased towards individuals with superior fitness attributes. Once the population passes through a number of generations, the algorithm is designed to converge and extract optimum solution.

Although many papers have been reported in the literature on the validity of the technique of genetic algorithm in function optimization and control application, significance of the evolution of the fitness –value distribution, over the generations of the population, remains to be explored and understood. Any function optimization technique is studied not only on the basis of how well the optimization of the objective function is achieved but also on the basis of how the search process moves over the solution space. The first aspect may be viewed as a macro aspect of the optimization technique whereas the latter aspect may be viewed as a micro aspect of it. Since a better understanding of the micro process has the potential of helping our understanding of the macro process in a more systematic and structured way , we have undertaken this research project with the motivation that a systematic study of the evolution of the fitness-value distribution over the generations of the population will enhance our understanding of the search process of the genetic algorithm.

The objective of this project is two fold: (i.) to see if we can visualize how the distribution of the fitness values of the chromosomes evolve with generations, and (ii.) to see if we can identify and capture some interesting patterns in the evolution process.

2. Computational Details

Since the objective of our study is not to optimize a new function but to investigate the nature of evolution of the fitness-value distribution over the generations, we have used a function that is already well-reported, the Function – 2 of De Jong [2]. A detailed description of this function is available in the literature [3]. We have used a simple continuous genetic algorithm with Selection / Select Mates / Reproduction, Crossover, and Mutation. Indeed, we have followed the flow-chart that is already available [4]. We have used a population size of 500, and used generation number 250 as the stopping condition. We have run our algorithm for a total of five runs (for each run, using 250 generations). The initial population, in each run, was generated randomly and independently.

For the selection process, we have used a roulette wheel where each chromosome in the current population has a roulette wheel slot that is sized in proportion to the chromosome’s fitness-value. For crossover, we have used a simple uniform crossover with random mating. For mutation, we have used a mutation probability of 0.001, and the process of mutation being just the creation of a new chromosome randomly.

3. Data Analysis

For each of the 250 generations of each of the five runs, our program generated an output file of fitness values vs. the chromosomes. This file was sorted in the ascending order of the fitness values. The fitness value for each chromosome was then plotted to generate the fitness-value distribution plot for each generation. All our results were extracted from these processed data files.

4. Results

For each of the five runs, the randomly generated initial population was found to have a fitness-value distribution with familiar exponential shape (an example is shown in Fig. 1).

From generation to generation, the fitness-value distribution curve evolves in two distinct patterns: (i.) smooth pattern: when the shape of the curve is mostly smooth (an example is shown in Fig. 2), and (ii.) stepped pattern: when the shape of the curve is characteristically "stepped-shaped" (an example is shown in Fig. 3).

Although higher generations tend to produce "converged population" generating population with higher fitness values, it was found that in generation 250, the shape of the fitness-value distribution curve was different in different runs. For example, in run #2, the shape of the fitness-value distribution curve for generation 250 was a horizontal line, meaning that the generation was completely occupied by the best fitness value (Fig. 4); but for run #3, the curve for generation 250 displayed a highly stepped pattern (Fig. 5).

Also, the best fitness value and its fraction of occupation (number of chromosomes with this fitness value divided by the total number of chromosomes in that generation) for different runs were found to be not identical, but different (Table-1).

5. Conclusions

On the basis of our given observations, we can draw conclusions at least on two aspects of our GA-based computations.

Specifying the termination condition (for each run over the generations) on the basis of the generation number may not be the best way to find the globally optimum fitness value. If this approach is to be used in any computation, then it is better to run the algorithm over multiple runs, extract the optimum value for each run, and then select the best value from all the extracted best values.

While the chromosomes evolve from one generation to the next, at certain points in time, certain ranges of fitness values become forbidden giving rise to the "stepped patterns" of the fitness-value distribution curve

6. References

[1] J. H. Holland (1975). Adaptation in Natural and Artificial Systems. Ann Arbor: University of Michigan Press.

[2] K. A. De Jong (1975). An analysis of the behavior of a class of genetic adaptive systems. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International 36 (10), 5140B.* (University Microfilms No. 76-9381)

[3] D. E. Goldberg (1989). Genetic Algorithms in Search, Optimization & Machine Learning. Addison – Wesley, pp. 106-120.

[4] R. L. Haupt and S. E. Haupt (2004). Practical Genetic Algorithms. John wiley & Sons, p. 52.

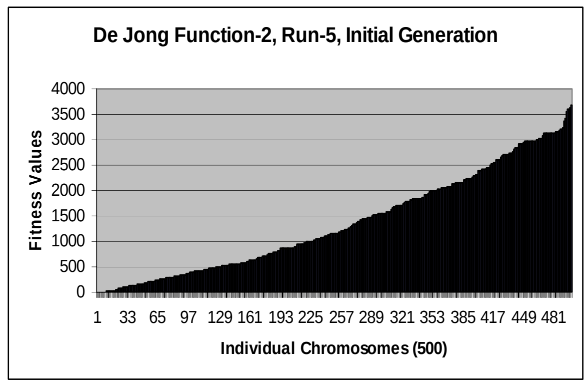

Fig.1. Fitness-value distribution of the initial population, run-5, displays exponential shape.

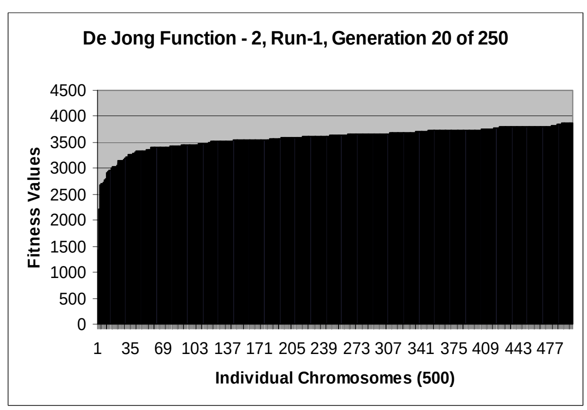

Fig.2. Fitness-value distribution of an evolved generation, displaying smooth shape.

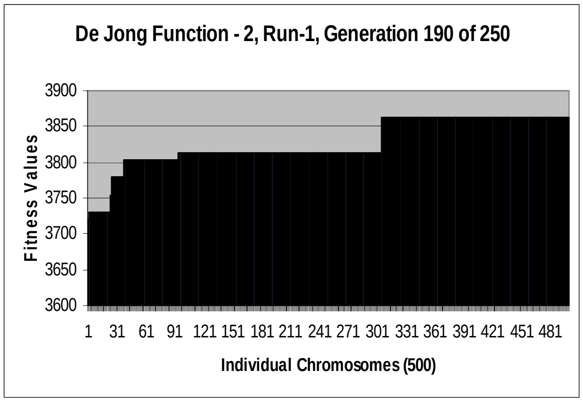

Fig.3. Fitness-value distribution of an evolved generation, displaying stepped pattern.

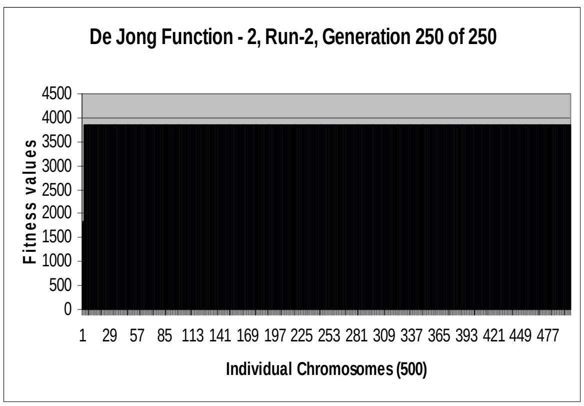

Fig.4. Fitness-value distribution of generation-250 (of 250, run-2) displaying an occupied generation.

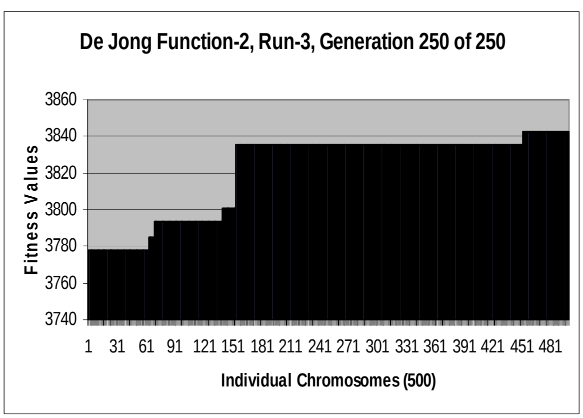

Fig.5. Fitness-value distribution of generation-250 (of 250, run-3) displaying a highly stepped pattern.

Run Number	Optimum Fitness Value	Fraction of Occupation
1	3862.99	0.518
2	3839.02	0.878
3	3842.88	0.096
4	3870.80	0.470
5	3864.84	0.766

Table-1. Optimum fitness value and its occupation fraction in generation 250 for various runs.