Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 2, Number 4; April-June, 2015 pp. 329-333

http://www.krishisanskriti.org/acsit.html

Genetic Algorithms-Overview, Limitations and Solutions

Yoshita Sharma[1], Neha Gupta[2] and Dayanand[3]

[1, 2]B.Tech (CSE) HMR Institute of Technology & Management (G.G.S.I.P.U) New Delhi, India
[3]Department of CSE HMR Institute of Tech. & Mgt., New Delhi, India
E-mail: [1]yoshita.a_sharma@yahoo.com, [2]ngupta647@gmail.com, [3]dayanand1003.11@bitmesra.ac.in

Abstract—*This paper presents the basic concepts of genetic algorithms (GA) which is a field of artificial intelligence used to find solution to problems which cannot be found by procedural programming or traditional algorithms. The paper includes the results and observations made on the basis of tool that has been constructed to analyze the performance of GA to find the best optimum solution. The problem of premature (local) convergence has been focused upon and the research done to solve this problem over the years has been reviewed. The implementation of chaos theory is in order to incorporate the 'randomness' factor in GA to avoid local convergence has been studied. The observations show how chaos theory supports GA to find the global optimum solution and overcome its limitation.*

Keywords: *Genetic Algorithm, Analysis tool, Tournament selection, Mutation factor, Uniform Crossover, Premature Convergence, Chaos theory,*

1. INTRODUCTION

1.1 Basics of GA

Genetic Algorithm is a part of Evolutionary Algorithms. It is a population-based search optimization algorithm which applies basic concepts of biology, namely-genetic behavior, selection,crossover and mutation, to generate the optimal solution for any complex problem which cannot be solved by traditional procedural programming.

Genetic Algorithms (GAs) were first invented by John Holland in the 1960s. Along with his students and colleagues, he worked on the development of GA during 1970s.[1].It is based on the Darwinian principle of evolution:"Survival of the Fittest". Among an occurring population of a particular generation, the fittest individuals should survive to be the parents for the population of next generation. This is determined by natural selection. The GA steps used in our tool, depicting the basic algorithm are represented in Fig.1. There is a variety of selection, mutation and crossover methods used for implementation of GA.

The technique of 'Tournament Selection' has been used in our tool to analyze the performance of GA with specified parameters. The solutions found using GA is termed as chromosomes or strings.A chromosome contains a group of numbers that completely specifies a candidate solution during the optimization process.[4]The problems in real life are of non-linear nature for which a global optimum solution has to be found. Hyper-heuristic techniques and Evolutionary algorithms are two solutions for nonlinear optimization problems.[2]

The solution found using GA is optimum as compared to those found by heuristic search techniques in terms of time complexity and optimization standards.

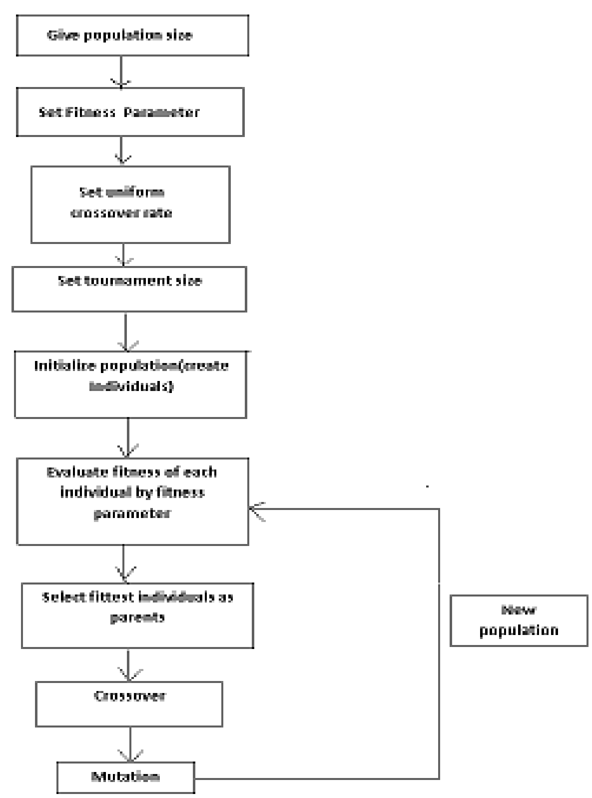


Fig. 1 : GA steps used in tool

1.2 Chaos Theory

Chaos theory is a phenomenon that derives its conclusions by applications of branches like meteorology, sociology, physics, engineering, economics, biology, and philosophy. The term ‘chaos’ itself depicts randomness. This theory is applied to find a solution to nonlinear problems where factors are known but conclusions are unpredictable. Most of the real world problems like weather turbulence, stock market behavior, etc. are nonlinear problems whose state at a particular point of time is unpredictable. The theory states that a small disturbance or the smallest changes in the input domain can lead to drastic and variant changes in the final solution of a problem. Based on this principle of unpredictability and mixing of factors, this theory is used to calculate global optimal solutions for genetic algorithms, which is explained later in the paper.

2. GA ANALYSIS TOOL

The GA Analysis Tool has been designed to implement the GA graphically using applets and user interactive controls. The GUI has been provided so that the users and learners can understand the basic procedure and applications of GA. Suitable parameters; namely-population size, fitness factor, crossover rate, mutation rate and tour size are set dynamically through the tools, using which GA is implemented and the results are found. The population size is initiated. The crossover technique to be used determines the value of crossover rate. In our tool, we apply uniform crossover type, which means that the child will have the genes of both the parents in equal proportion. So the rate has been set to 0.5. The selection technique used is tournament selection. The tour size, i.e. the number of individuals competing for survival is set by the user.The fitness parameter is provided as a string of 0s and 1s. This determines the gene sequence or the structure of the individual’s chromosome. The screenshot of the tool and the result is as shown below:

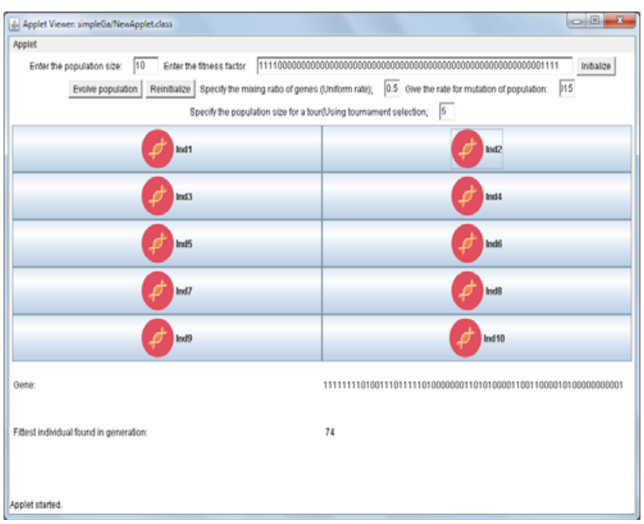


Fig. 2: GA Tool1

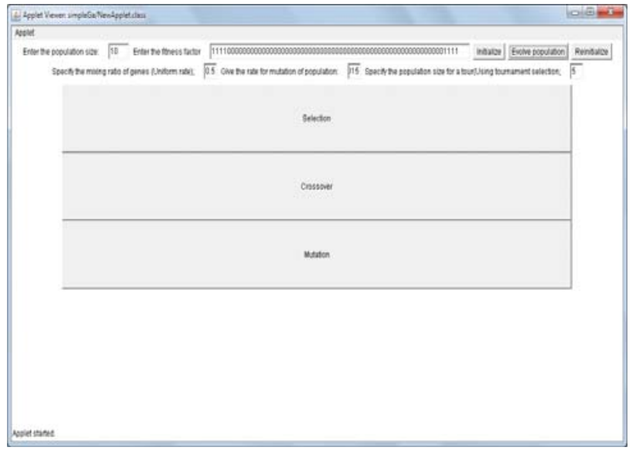


Fig. 3: GA Tool 2

2.1 Selection Technique – Tournament Selection

Tournament selection provides selection pressure by holding a tournament among “s“ competitors, with “s” being the tournament size [4]. The winner of the tournament is the individual with the highest fitness of the s tournament competitors. The winners become parents for the next generation and the process continues till the fittest individual is found. The winner is always the fittest individual in the population. Steps of the technique are:

1. Initialize the population with the specified tournament size.
2. Save every individual for the tournament to find the fittest.
3. Find the fittest individual according the given fitness parameter.

2.2 Mutation Rate and Crossover

The mutation rate is the determinant factor for convergence of the population to find the ultimate fittest individual. The crossover technique is used to form a new individual using the chromosome configuration of parents. The two activities to find the resultant solution are- exploration and exploitation. Exploration is used to analyze every possible solution in the search space. The best solution is then exploited via crossover to find the fittest individual from parent individuals that were earlier obtained by exploration.

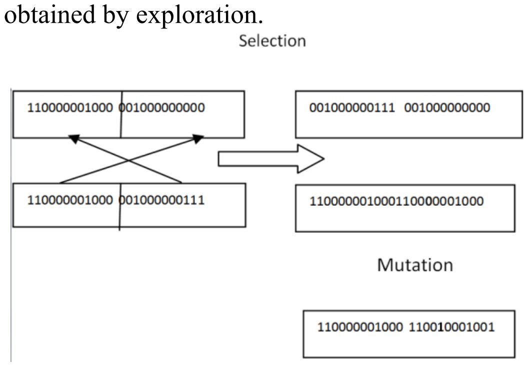
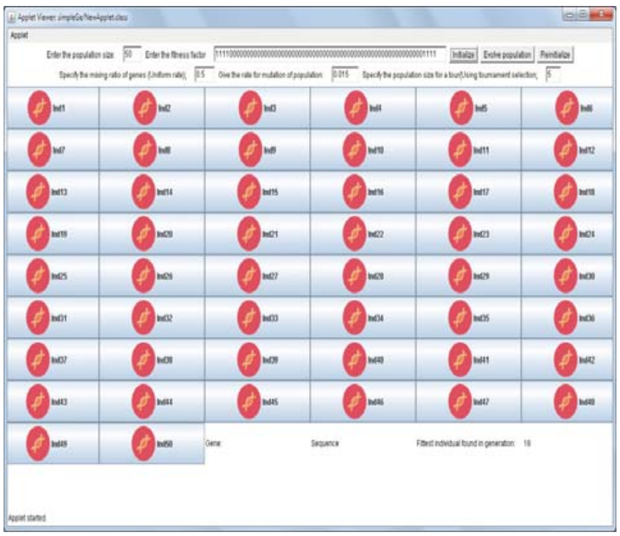


Fig. 4: Crossover and Mutation

3. PREMATURE CONVERGENCEIN GA

Premature convergence is the term that refers to the phenomenon of reaching a local optimum solution rather than the global optimal solution, when GA is applied to the real world problem. It occurs when genes of high rated individuals dominates the population. It is a suboptimal state where GA operators can't produce offsprings with a better performance than their parents. A stage comes when the variety in gene sequences of the individuals in a population reduces. This demonstrates convergence as most of the population is now identical. Following are the screenshots that describe the variation in the final result after running the tool more than once with the same parameters:



As observed, the fittest individual in the first attempt was found in the 16th generation whereas in the second attempt it was found in the 15th generation. So we can conclude that the global optimum solution has not been achieved. Over the years, many experiments have been carried out to find a way to modify the GA to prevent premature convergence so that it gives a globally optimum solution. Chaos Theory has played an important role as a solution to this problem.

4. SOLVING PREMATURE CONVERGENCE

Various techniques have been adopted and implemented by applying different algorithms and proposing different models. Chaos theory has also been used for the same. The significant solutions have been described here.

4.1 DGCA and Elitist Technique

To reduce early convergence, some control factors need to be monitored to maintain the diversity and fitness of the population. The factors are:

1. Genetic Diversity
2. Selective Pressure [3]

Two techniques of selection, namely, Roulette Wheel Selection and Rank Selection were blended and used as an operator. The former technique works on the principle of exploitation and the later involves exploration. This was kept in mind while implementing Dynamic Genetic Clustering Algorithm (DGCA).

As shown in our tool, the technique of tournament selection has been applied on the entire population. So it may be possible that the best solutions in the form of fittest individuals have been ignored on the way. Thus, it converged towards local optimum. This limitation is overcome by DGCA as in this technique the entire population is divided into clusters. The blended operator is applied to select the fittest individual out of each cluster. Suppose two clusters are considered and after applying the blended selection operator and performing crossover, a child is generated. Now the comparison is made among the fittest individuals of the two clusters and their child. The fittest out of the three becomes the candidate for parenting the next generation. This technique is called the 'elitist technique' as it follows the principle of elitism. Elitism emphasizes on the fact that only the individuals of a certain ancestry, in this case the fittest individuals should participate in the evolution of next generation. The following equation was used to decide the number of clusters on which the selection operator was applied:

x=round(log2(psize));

Where x is no. of clusters formed and psize is population size. [3]

Each cluster is further divided into two clusters to perform crossover. The algorithm was implemented on the Training Salesman Problem to find the shortest path that covers all the cities at least once. The comparison was made by applied normal GA and DGCA. The number of iterations made in DGCA to calculate the best path was 1/4th that of the number of iterations using normal GA and the best path had lesser distance using DGCA as compared to GA.

4.2 Using Average Hamming Distance

Using Hamming distance, the time convergence has been focused upon. This refers to the time it takes to obtain the configuration of a particular allele through the entire process of selection, crossover and mutation, using fitness ratios to obtain bounds on time complexity. The average hamming distance of a population is the average distance between all pairs of strings in a population of size N[3]. The size used to calculate the average is

N(N-1)/2.

In this technique, the reason attributed for a faster convergence is based on two offsprings having the same alleles or similar gene sequence if one parent has contributed those alleles to their production. Lesser the hamming average , faster the convergence. Random search is done on a flat function i.e.

$f(x_i)$=constant.

Some lemmas have been proposed to explain that what exactly leads to faster convergence. It has been found that selection factor determines the time convergence. For crossover, the lemma stated is:

"Traditional crossover operators do no change average hamming distance of given population."

Letting $h_{i,t}$ stand for the hamming average of the ith locus we have:

$h_{i,t} = h_{i;t}+1$[4]

i.e. the average hamming distance of the population in one generation is equal to that of the next generation. The diversity of a population is maintained by mutation. Here, instead of altering the mutation rate, the gene sequence of an individual is added to its bit complement to the population. Combining fitness prediction with hamming average prediction give an idea how much progress is possible and how much work has been done.

4.3 Using Iterated Prisoner's Dilemma (IPD) and Chaos Theory

This problem has been used as it provides a dynamic environment with feedback. Special mutation operator called 'λ' mutation is used. IPD was first implemented by Axelorod in the 1970s, in his book 'The Evolution of Cooperation' (1984). In this problem, a tournament is organized between N prisoners and their mutual strategies are recorded. In the first iteration of the game, mutual strategy is implemented among players to get scores A better strategy is 'tit for tat with cooperation'. A high score for the player in a game using better strategies in the game represents a light sentence and a successful pair of decisions. The dilemma is whether the prisoners should 'cooperate' or 'defect' for the appropriate move. A set of moves is used to create a 64 bit string which represents each player in the algorithm. The best strategies are clustered together and patterns are recognized in the data. The rules are generated on the basis of the fittest chromosome string formed by clustering the best strategies. These rules were tested using "Travelling Salesman Problem". By analysis of these rules, a simple chaotic function was generated:

$x_{n+1}=\lambda x_n(1-x_n)$ [6]

Following conclusions were made:

1. $0<\lambda<3$- convergent pattern
2. $3<\lambda<3.56$- bifurcating
3. $3.56<\lambda<4$- chaotic

The chromosomes are generated by the IPD strategy. It was found that an individual with convergent 'λ' produced an offspring with rigid uniform crossover masks.[4] Individuals with non-convergent 'λ' have a high degree of variability over generations. This shows chaotic behavior which is favorable for diversity. In this process, masked uniform crossover and bit flipping mutation was used. Using the equation above, the test was repeated 10 times with different mutation probabilities. Increase in the rate of 'λ' mutation operator improved the average IPD score, i.e. confirmed the strategy to be used to ensure the diversity of gene sequences of the chromosome.

5. CONCLUSION

The tool has been designed to understand the implementation of GA and the reason for the occurrence of the phenomena called premature convergence. Many techniques including chaos theory have been used to solve the conclusion to this problem. The main objective of all the solutions has been to increase the efficiency of result set by modifying the selection operator used to find the fittest individual to produce the next generation. The three solutions by implementing DGCA, Hamming Distance and Chaos Theory have been effectively explained In case of DGCA, clustering plays a major role in preventing early convergence. In the application of Hamming Distance, combining fitness prediction with hamming average prediction gives an idea about how much progress is possible in the production of the next generation and how much work needs to be done to maintain its efficiency. The role of chaos theory in solving the premature convergence problem has been explained using 'Iterated Prisoner's Dilemma', by explaining the involvement of mutation factor 'λ' and how it effects the crossover process for the production of offspring. The techniques discussed above can help to obtain a global optimum solution to a non-linear, real world problem.

6. FUTURE SCOPE

A lot of work is being done on combining genetic algorithm with some other technique to find a global optimum solution. The combination is termed as 'Hybrid Genetic Algorithm'. On combining GA with another technique, a hybrid procedure is created which brings out the best characteristics in both GA and the technique. The changes are made to the GA by combining it with either a local search method or by making some changes in the operators like selection, crossover and mutation by using specified rules. Work is being done to modify the GA to fulfill the objective of finding the best possible solution.

REFERENCES

[1] Anuj Mehta, Anshul Sharma, "Review Paper of Various Selection Methods in Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Issue7, Vol.3, July 2013, Page(s):1, 3

[2] Reza Ebrahimzadeh, Mahdi Jampour, "Chaotic Genetic Algorithm based on Lorenz Chaotic System for Optimization Problems", I.J. Intelligent Systems and Applications, Modern Education and Computer Science Press , April 2013, Page(s):1-3

[3] Ms.Shikha Malik, Mr. SumitWadhwa, “Preventing Premature Convergence in Genetic Algorithm Using DGCA and Elitist Technique”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6**,** June 2014, Page(s): 1, 4 , 9

[4] Mr. Sushil J. Louis, Mr. Gregory J.E. Rawlins, Indiana University, “Predicting Convergence time for Genetic Agorithms”, Page(s):2-5

[5] Data electronically available on “www.obitko.com/tutorials/” (Introduction to Genetic Algorithms), accessed on 11th March 2015.

[6] John Determan, James A.Foster, “Using Chaos in Genetic Algorithms”, Data electronically available on people.ibest.uidaho.edu/~foster/Papers/11805.pdf, Page(s) 1, 2, 5

[7] Data electronically available on “http://www.abarim-publications.com/” (Chaos theory), accessed on 16th March, 2015.

[8] Jennifer Golbeck, University of Maryland, College Park, MD, USA, “Evolving Strategies for the Prisoner’s Dilemma”, Data electronically available on: https://www.cs.umd.edu/~*golbeck/downloads/JGolbeck_prison.pdf*, Page 1

[9] Adaptation of Robert Axelrod, “The Evolution of Strategies in the Iterated Prisoner'sDilemma”, in Lawrence Davis (ed.), Genetic Algorithms and Simulated Annealing, 1987, Page(s):1-3

[10] Miguel Rocha and Jos_eNeves, “Preventing Premature Convergence to Local Optima in Genetic Algorithms via Random Offspring Generation”, Data electronically available at http://repositorium.sdum.uminho.pt/bitstream/1822/4291/1/P078.pdfPage(s):1-6.

[11] Jyoti Sharma, Ravi Shankar Singhal, “Genetic Algorithm and Hybrid Genetic Algorithm for Space Allocation Problems- A Review”, International Journal of Computer Applications (0975 – 8887) Volume 95– No.4, June 2014, Page(s):1-3

[12] Manoj Kumar, Mohammad Husian, Naveen Upreti&Deepti Gupta, “Genetic Algorithm: Reviewand Application”, International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, Page(s): 1, 2

[13] Gopesh Joshi**, “**Review of Genetic Algorithm: An Optimization Technique”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, April 2014

[14] Tarek A. El-Mihoub, Adrian A. Hopgood, Lars Nolle, Alan Battersby, “Hybrid Genetic Algorithms: A Review”, Engineering Letters, 13:2, EL_13_2_11, August 2006, Page(s):1-4

[15] Shashi Mittal, “Machine Learning inIterated Prisoner's Dilemmausing Evolutionary Algorithms”, Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur.