Role and Working of Genetic Algorithm in Computer Science

Manik Sharma
Assistant Professor and Head,
Department of Computer Science and Applications.
Sewa Devi SD College Tarn Taran.

Abstract:

Genetic algorithm is an evolutionary computing approach which is recurrently used in different areas of Computer Science, Management, Physics, and Engineering etc. In this paper, an effort has been made to demonstrate the role of genetic algorithm in the different domains of Computer Science. The study reveal that genetic algorithm is effectively used to solve the different research problems viz. query optimization, task scheduling, data mining, part of speech tagger, phrase chunker, image segmentation, inventory management etc.

Keywords: Genetic Algorithm, Software Testing, Task Scheduling, Query Optimization etc.

1. Introduction.

'Genetic Algorithm' commonly abbreviated as 'GA' is one of the prominent evolutionary approaches. The general conception of 'Genetic Algorithm' was proposed by John Holland. He is considered as the father of 'Genetic Algorithms'. These are search algorithms specifically designed to simulate the principle of the natural biological evolution process. 'GA' borrows its essential features from natural genetics. In other words, 'Genetic Algorithms' are stochastic techniques that stipulate good-quality solution with low time complexity.

It permits a population composed of many individual chromosomes to evolve under delineated selection rules to generate a state that optimize the objective function. These types of algorithms successfully operate on a population of solutions rather than a single solution. It generally employs some heuristics like 'Selection', 'Crossover', and 'Mutation' to develop better solutions. [Goldberg 1999][Sevinc and Cosar 2011].

'Genetic Algorithms' are capable of being applied to an enormously wide range of problems. Some of the major applications of these algorithms are as given below:

- Query Optimization
- Parallel Processing
- Software Engineering
- Game Theory
- Image Processing
- Data Mining
- Machine Learning
- Disease Diagnosis
- Agriculture
- Inventory Management etc.

2. Working of Genetic Algorithm

In general, these types of algorithms aim for searching better solution from a number of available solutions. As stated, GA starts its working from a set of solutions rather than a single solution. The set of solutions is called population. The initial population is generated randomly. Each solution of the problem is adequately represented by encoding a string (Chromosome) of bits or characters (Genes). Every chromosome has a fitness value associated with it.

The collection of chromosomes with their corresponding fitness values is called population. The population at a particular instance is called generation.

Fitness function is one of the major decisive parameters of 'Genetic Algorithm'. It defines the objective of the problem to be optimized. A pair of chromosomes based on their fitness values is used to reproduce offsprings.

The genetic properties of both the chromosomes are intermixed to generate better offspring, such a mechanism is called crossover. After crossover operation, the genetic characteristics of the generated offspring are further modified. Mutation is a procedure to modify the characteristics of the generated offspring to make it more effective. The algorithm terminates when the required condition is fulfilled [Goldberg 1999] [Man et al. 1996].

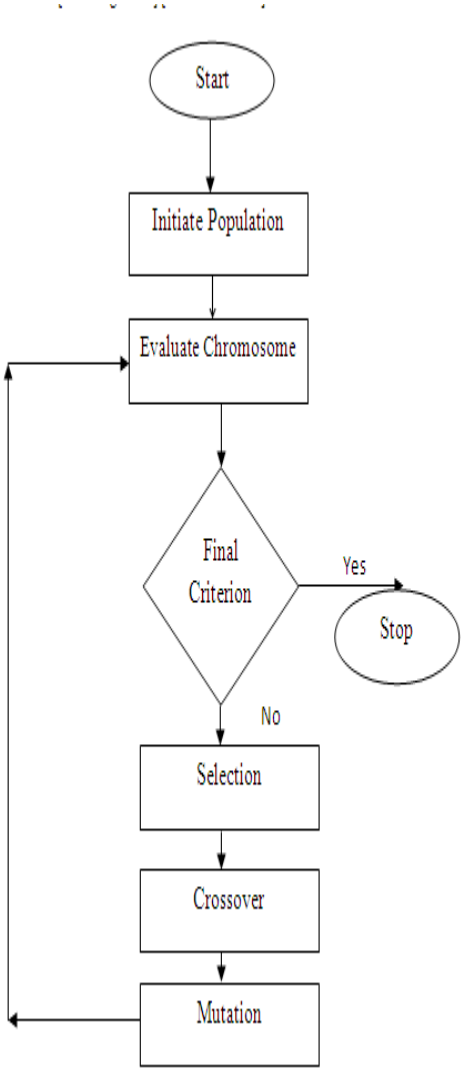

Figure 1: Flow Chart of Genetic Algorithm

The pseudo-code for the above flow chart is as given below:

Gen=0 ; // Generation Counter
Initiate_Poulation POP(Gen);
// Initial population generation
Fitness_EvaluationPOP(Gen); // Evaluates Individual's Fitness
While Gen<Max_Gen // Terminating Criterion
Begin
Gen=Gen+1 // Increase Generation Counter
Select Parents
// Select Parents from the population
Crossover ;
// Apply crossover to selected parents
Mutation
// Apply Mutation to offsprings
Evaluate_Fitness;
// Evaluates Fitness of offspring
Select best Offspring for next
// Select Survivor Individuals
Generation
End

Start and end of 'Genetic Algorithm' are very much analogous to all other optimization techniques. 'GA' starts working with the assertions of optimization variables, their allied costs and end up with an optimum solution. However, the complete process from beginning to ending is quite different from other optimization techniques [Haupt&Haupt 2004].

One of the foremost objectives of 'Genetic Algorithm' is to stipulate best possible solution from fixed and finite sized population in a minimum amount of time. It works in two phases. In the first phase, selection is performed from the population and in the second phase selected generation is manipulated to create new generation. One of the outstanding advantages of these algorithms is that purely non-heuristic search throughout the solution space is performed. Therefore, no specific knowledge of the problem space is required in advance. These are flexible and robust in solving the complex problems.

GA Operators

After initial population, the whole process of 'GA' revolves around three operations called 'Selection', 'Crossover' and 'Mutation'.

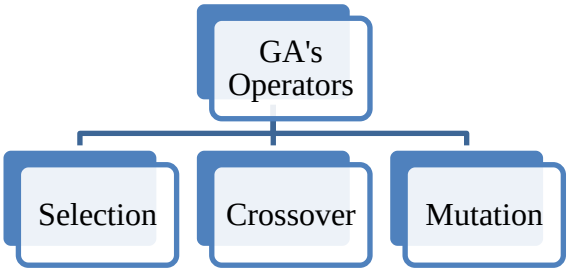

Figure 2: Operators of Genetic Algorithm

Selection: Selection is one of the important processes of 'GA'. It is used to select an individual on the basis of its fitness value. The individual chromosomes that have higher fitness values are most probable candidates to be selected for reproduction. The individuals having low values will get little chance of their selection. In simple words

the probability of a chromosome to be selected for reproduction is proportional to its fitness value.

Initial Population: It is the first phase of 'Genetic Algorithm'. In general, initial population is generated arbitrarily. The important aspect of initial population is that the design of initial chromosome must match with the design of final chromosome. It is one of the important phases because the member of this population encodes the optimal solution of the problem. While generating initial population, one has to focus on the search space of solution, fitness function and number of individuals. The past research showed that the good initial population always leads to better possible solution and vice versa.

Crossover: It is a technique used to combine the individual chromosomes which produces a new chromosome. The offspring generated by crossover is supposed to have the features of both of the parent's chromosomes. The objective of crossover operation is to find better solution from good solution. It is also called recombination operator. It selects a pair of individual chromosomes (Parents) for mating. First of all, a location for crossover is selected, and then bits or characters of the selected parents following marked location are swapped. Crossover operation is used to avoid duplication of parents used in recombination. The offsprings generated by crossover operation may have either higher fitness or in some cases lower fitness than their parents. However a high fitness offspring is desirable from crossover operation. Crossover operator can be implemented by using several techniques.

3. Characteristics of Genetic Algorithm: The various features of Genetic Algorithm are summarized as below (Goldberg 1999):

- Genetic Algorithm is based on the concept of natural selection and natural genetics.
- These are easy to understand and implement.
- These are extensively used in optimization problems.
- These work on population of points rather than an individual point.
- These use probabilistic transition rule instead of deterministic rules.
- Genetic Algorithms can effectively deal with large number of variables.
- These algorithms do not require any derivative information.
- These are more effective for complex problems than simple problems.
- These provide solution quickly as compared to other traditional optimization techniques.

4. Applications of Genetic Approach

Query Optimization: Query is an indispensable component of a distributed database system. In general, it is used to successfully accomplish different operations of the database objects. Query optimization is a process that generates the different operation site allocation plans to execute the query. Operation site allocation plan represents the mechanism of query execution. The objective of operation site allocation problem is to select an optimal query execution plan which optimizes the *Total Costs* of the distributed decision support system query. Number of researchers have optimizes the query using genetic algorithm and its variation. They found that GA works better than other traditional techniques as it gives the optimal solution very swiftly [1][2][3][4][5].

Parallel Processing: Parallel processing is an efficient approach to meet the computational requirements of the large number of problems. Lots of researchers are coming up with innovations and bringing new ideas for its development. In parallel processing, one of the major problems is the scheduling of tasks to the different machines or nodes. Numbers of research have tried to explore the usage of parallel processing to minimize the makespan of the task. Authors elucidated that the processors and communication links are vital resources in parallel computing systems and their efficient management through proper scheduling is essential for obtaining high performance. Parallel computing algorithms are normally classified as UNC (unbounded number of clusters) scheduling, the BNP (bounded number of processors) scheduling, the TDB (task duplication based) scheduling, and APN arbitrary processor network) scheduling. To get the maximum benefit from the parallel processing some of the authors have used the concept of clustering. It is observed that usage of genetic algorithm has significantly reduced the makespan of the task as compared to above said algorithms.

Software Engineering: Software testing is one of the important phase of software engineering that is used to reveal the error in the code. Software is examined by using different types of testing viz. White box, Black box, stress, load etc. Genetic algorithm is best suited to test the complex software as in large and complex software exhaustive testing is not possible. The research reveals that number of authors have used GA in different types of software

testing. Different authors have varied the rate of crossover and mutation to get the better results. In addition, authors also varied size of initial population, number of generations and encoding schemes[24][25][26].

Game Theory: It is a formal study of decision making process used in multidisciplinary studies viz. Economics, Political Science, Biology, Computer Science, Psychology etc. Several authors have solved game theory problem by using genetic algorithm. Authors found that GA gives optimal strategy as compared to other techniques viz. linear and integer programming [16][17].

Image Segmentation: It is another important research areas of Computer Science. The objective of image segmentation is to decompose an image into non overlapping segments. Lucia Ballerni and Cagnoni have done an extensive research to analyze snake's algorithm using GA. In their study, authors focused on optimizing the energy function of Snake's algorithm [6][7][8][9].GA is also effectively used in implementing the edge detection in image processing. Edge detection is a two stage process. The first stage is responsible for edge enhancement process. Second stage uses boundary detection and edge linking to select and combine edge map pixels [10][11].

Data Mining: It is one of the important concepts which is used to explore some meaningful information or pattern from the huge amount of data. GA is effectively used in classification and clustering of data in different domains. It was found that number of authors have used genetic algorithm to predict the different medical diseases like cancer, heart problems, tumor etc. [18][19][20]. In addition significant work has been done in the field of sentiment analysis also known as opinion mining, text and web mining. In text and web mining, the focus was given on reducing the effort to extract meaningful information from momentous amount of text.

Machine Learning: It is a process in which a machine is made intelligence by incorporating some sort of learning mechanism. GA is effectively used in solving the chess problem. In addition, it is used in solving the POS Tagging problem, in which every word is assigned its grammatical information. The problem of phrase chunking is also effectively solved by genetic algorithm. Phrase chunking is a process to extract certain portion of text from a sentence [20][21][22][23].

Inventory Management: Genetic algorithm is also effectively implemented in inventory management. It was found that in inventory and supply chain management, one of the major issues is to maintain the optimal level of stock i.e. it should neither be in excess or in short. Research reveals that with the use of GA, one can easily detect the optimal level of stock as compared to other traditional approaches like linear and integer programming[27][28][29].

Figure 3 represents the number of papers published from 1950 to 2012 that employs Genetic Algorithm in different domains of Computer Sciences. The data has been collected from Google Scholar by executing different queries.

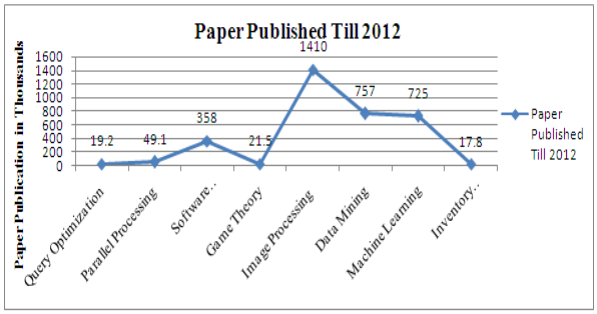

Figure 3: Number of Paper Published using GA during 1950-2012

5. Conclusion

Genetic Algorithm is an assertive optimization procedure, which is competently used to resolve optimization problems in different research areas like Task Scheduling, Query Optimization, Image Processing, Data Mining, Game Theory, Software Engineering, Inventory Management, Machine Learning etc. The working of GA is based upon three major operators known as selection, crossover and mutation. It was observed from the past research that GA is a dominant research approach in Computer Science. Authors varied different parameters like number of generation, crossover and mutation rate, size of initial population, encoding scheme etc. The data collected from Google Scholar reveals that Genetic Algorithm has been widely used in Software Engineering and Image Processing. However, the areas like Data Mining and Query Optimization may still be the candidates for Genetic Algorithm. In addition, to get the better results. However, to get the better results, one may combine GA with other techniques like Neural Network, Automata, Information Theory, Fuzzy Logic etc.

6. Acknowledgement: Author is greatly thankful to Dr. Gurvinder Singh, Dr.Rajinder Virk and Dr. Gurdev Singh for their precious guidance and support.

7. References

1.Apers Peter M.G., Hevner Alan N., Yao Bing S. 1983. "Optimization Algorithms for Distributed Queries". IEEE Transaction on Software Engineering. SE-9.1.pp.57-68.
2. Cornell Douglas W., Philip S.Yu.1989. "On Optimal Site Assignment for Relations in the Distributed Database Environment". IEEE Transactions on Software Engineering. 15.8. pp.1004-1009.
3. Kumar,T.V.,Singh,V.,Verma,A.K. 2011. "Distributed Query Processing Plan Generation using Genetic Algorithm". International Journal of Computer Theory and Engineering.3.1.pp.38-45.
4.March.S.T, Rho. 1995. "Allocating Data and Operations to Nodes in distributed Database Design". IEEE Transactions on Knowledge and Data Engineering. 7.2.pp. 305- 317
5. Martin T.P., Lam K.H., Russel Judy I. 1990. "An Evaluation of Site Selection Algorithm For Distributed Query Processing". The Computer Journal. 33.1.pp. 61-70.
6. L. Ballerini. Genetic snakes for medical images segmentation. Evolutionary image analysis, signal processing and telecomunications, 1999, 69 – 73.
7. L. Ballerini. Segmentation of ocular fundus images using genetic snakes. 1999. Proceedings European medical and biological engineering conference, 1040 – 1041.
8. L. Ballerini. Multiple genetic snakes for bone segmentation. 2003. Applications of evolutionary computing, , 346 – 356.
9. S. Cagnoni, A.B. Dobrzeniecki, R. Poli, J.C. Yanch. 1999. Genetic algorithm based interactive segmentation of 3D medical images. Image and vision computing, ,17.12.pp.881 – 895.
10. A. Rosenfeld, A.C. Kak. Digital picture processing. Morgan Kaufmann, 1982, 435.
11. D. Williams, M. Shah. 1990. Edge contours using multiple scales. Computer vision, graphics, and image Processing. 51.3.pp.256 – 274.
12. Kaur Kamaljit, Chhabra Amit, Singh Gurvinder. 2010. "Heuristics Based Genetic Algorithm for Scheduling Static Tasks in Homogenous Parallel System". International Journal of Computer Science and Security. 4.2.pp.182-198.
13. Kwok Yu-Kwong, Ahmad Ishfaq. 1999. "Benchmarking and Comparison of Task Graph Scheduling Algorithms". Journal of Parallel and Distributed Computing. 59.3.pp.381-422.
14.Yi-Hsuan Lee, Cheng Chen. 1999. "A Modified Genetic Algorithm for Task Scheduling in Multiprocessor Systems", Proc. Of 6th International Conference Systems and Applications..
15.Tzung-Pei Hong, Sheng-Shin Jou, and Pei-Chen Sun, "Finding the Nearly Optimal Makespan on Identical Machines with Mold Constraints Based on Genetic Algorithms", Proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational Science.
16. .I.A.Ismail ,N.A.El_Ramly , M.M.El_Kafrawy , and M.M.Nasef. 2007. "Game Theory Using Genetic Algorithms". Proceedings of the World Congress on Engineering 2007 Vol I WCE 2007, July 2 - 4, London, U.K.
17. C. Birchenhall. 1995. "Modular technical change and genetic algorithm", Computational Economics Journal, vol. 8, pp. 233-253.
18. GunjanVerma, VineetaVerma. 2012. "Role and Applications of Genetic Algorithm in Data Mining". International Journal of Computer Applications (0975 – 888) Volume 48– No.17
19. Jain, A. K.; Zongker, D. 1997. "Feature Selection: Evaluation, Application, and Small Sample Performance", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 19, No. 2.
20. Ian W. Flockhart" and Nicholas J. Radcliffe. 1996. "A Genetic Algorithm based Approach to Data Mining". KDD-96 Proceedings.
21. Amanpreet Kaur, Gagandeep Jindal.2013. "Overview of Tumor Detection using Genetic Algorithm". International Journal of Innovation in Engineering and Technology. Vol2. Issue 2.
22. Ana Paula Silva, Arlindo Silva, Irene Rodrigues. 2013. "A New Approach to the POS Tagging Problem Using Evolutionary Computation". Proceedings of Recent Advances in Natural Language Processing, pages 619–625, Hissar, Bulgaria
23. Claudia Borg, Mike Rosner, Gordon Pace. 2009. "Evolutionary Algorithms for Definition Extraction". Workshop On Definition Extraction 2009 - Borovets, Bulgaria, pages.
24. Bo Zhang, Chen Wang, 2011. "Automatic generation of test data for path testing by adaptive genetic simulated annealing algorithm", IEEE, 2011, pp. 38 – 42.
25. Praveen Ranjan Srivastava and Tai-hoon Kim, "Application of genetic algorithm in software testing" International Journal of software Engineering and its Applications, 3(4), 2009, pp.87 – 96.
26. Girgis. 2005. "Automatic Test Generation for Data Flow Testing using a Genetic Algorithm", Journal of computer science, 11 (6), pp. 898 – 915.
27. Leena Thakur, Aaditya A. Desai. Inventory Analysis Using Genetic Algorithm in Supply Chain Management. International Journal of Engineering Research & Technology (IJERT). 2.7.pp.1281-1285.
28. P.Radhakrishnan, V.M.Prasad, M. R. Gopalan. 2009. "Inventory Optimization in Supply Chain Management using

Genetic Algorithm". International Journal of Computer Science and Network Security. 9.1:33-40.
29. K. Wang, Y. Wang, "Applying Genetic Algorithms to Optimize the Cost of Multiple Sourcing Supply Chain Systems – An Industry Case Study", Volume 92, Page: 355-372, 2008.