

Applications of Genetic Algorithm in Distributed Computing, Machine Learning and Software Engineering

Faheem Ahmad*, Dr. Shafiqul Abidin**

*(Research Scholar, Shri Venkateshwara University Gajraula, Uttar Pradesh India)

**(Head, Department of IT, HMR Institute of Technology & Management, New Delhi India)

Corresponding Author: Faheem Ahmad

ABSTRACT: *There are many types of computational techniques like deterministic, evolutionary and random. Evolutionary techniques are also recognized as the technique which has been inspired by nature as these kinds of techniques have taken the concept from nature.*

Genetic algorithm (GA) is one of the most frequently utilized evolutionary techniques which is utilized to solve different NP-hard computational problems. The base of Genetic Algorithm is the principle of human genetic. Earlier research presented that it was one of the most successfully used technique to answer the different problems from the domain of Computer Science viz. task scheduling, software cost estimation, image processing and query optimization etc. In this paper, an effort is used to study the application of Genetic Algorithm in Distributed Computing, Query Optimization, Software Engineering, and Machine Learning.

KEY WORDS: *Software Engineering, Database, Genetic algorithm Evolutionary algorithm*

Date of Submission: 11-05-2018

Date of acceptance: 28-06-2018

I. INTRODUCTION:

In last few years, a substantial development has been found in software and research industry. Firstly, all the NP-Hard problems and complex system were created using deterministic techniques. However, these days, evolutionary techniques are most commonly applied to solve the difficult problems. The evolutionary Algorithms (EA) are basically developed on Charles Darwin's theory of Evolution. In Evolutionary Algorithm, the individuals are symbolized by fixed length strings called chromosomes.

The key concept behind all the techniques is same as firstly the population of individuals is randomly selected then the function which is called as fitness function is applied to choose the most suitable participants. The new offspring is generated by the application of Crossover and Mutation operators. We can say that, evolutionary computation applies iterative growth, such as growth or upgrading in a population. This population is then selected in a steered random explore by applying parallel handling to get the wanted result. Such processes are often encouraged by biological mechanisms of evolution. As evolution can generate extremely optimized processes and networks, it has many functions in computer science.

Few of the important evolutionary techniques are given below:

- Genetic algorithm
- Honey bee
- Harmony search
- Genetic Programming
- Cultural algorithm
- Ant colony optimization Particle swarm optimization
- Differential evolution

II. GENETIC ALGORITHM:

Based upon intricacy, a problem can be categorized as P, NP-hard and NP-complete. NP hard problem are complicated and complex and utilize substantial volume of CPU time for their execution.

In Computer Science, the problems like task scheduling, query optimization, software cost estimation, and data mining all are NP-hard in character. It is challenging to explain these problems using traditional techniques. Genetics is a heuristic approach in which assumptions can be made. The implementation of the problem can be swifter by using genetics approach. The implementation of Genetic Algorithm starts with randomly generated chromosomes. The Fitness function is use to select the fittest ones. In it, two different types

of operators called the Crossover and Mutation are applied to the better candidates to produce new offspring. The algorithm ends when either the candidate with sufficient quality is generated or the highest figures of generations have been yielded. The genetic algorithms are very useful in answering the problems like optimization problems, search problems etc. The structure and functioning of Genetic Algorithm is started upon some concepts like chromosome, fitness function, selection, crossover and mutation. In addition, size of population and number of generation also plays an important role.

III. ROLE OF GENETIC ALGORITHM IN SOFTWARE ENGINEERING:

Today, one of the most important and desired research area is Software Engineering. In simple words, software engineering is a systematic approach to develop and maintain the automated system. Some of the important aspects of software engineering are software metric, software testing and software quality assurance. Software metrics help us to measure the performance of the system. The objective of software testing is to reveal the bugs involved in a software module. Additionally, to judge and enhance the value of a software, software quality assurance is used.

Many authors stated that the estimation of cost is a vital factor in software development. Different authors have applied and shown different techniques to estimate the cost of software module. Software cost estimation is one of the most difficult jobs in software engineering. Isa Maleki [1] have proposed a hybrid solution for cost estimation of the software. Authors stated that cost estimation is a significant factor of software development. Authors first apply genetic algorithm to create the early population. After applying crossover and mutation operations, Ant Colony Optimization is applied to guide the system to compute software cost estimate and assess the result.

Authors [3] have developed a binary genetic solution for the same. Authors tested the performance of their algorithm on NASA software and found it more effective as compared to current software cost estimation models.

Tai-hoon and Praveen Ranjan Srivastava have applied genetic algorithm for enhancing the testing capacity of the software. Authors mentioned that using genetic approach; one is capable to effectively search the most critical path in the software. In software testing, Genetic Algorithm is better than even exhaustive testing for small scale difficulties.

S. Keshavarz and Reza Javidan have proposed a genetic based answer to regulate software quality assurance. The aim was to create optimal test data. Authors have shown their tests and shown that genetic approach is best in comparison to others.

IV. ROLE OF GA IN DISTRIBUTED COMPUTING:

For last couple of decades, different types of approaches have been founded to make the speedy execution of process of a job. In the beginning, the attempts were to enhance the design of hardware. However, these days, stress is focused on the software methodology so that the job can be effectively expected. In general, the distributed job can be classified as distributed application and distributed queries. In distributed application, the aim is to minimize the make span on the task so that the resources can be efficiently employed. The objective of the distributed query is to reduce the total time or time to response of a query.

Genetic Algorithm has been fully employed in optimization of distributed jobs. Different authors have used GA to optimize distributed task scheduling. The following pieces of this section discuss the tasks done to optimize make span in parallel task scheduling and distributed query optimization.

Tzung-Pei Hong, Pei-Chen Sun and Sheng-Shin Jou have projected a GA-based approach to solve the scheduling problem with the minimum make span on identical machines with mold constraints. This GA-based algorithm used an adjustment operator and they presented that the adjustment operator does enhance the functionality of the scheduling.

Kamaljit Kaur, Gurvinder Singh, Amit Chhabra mentioned that the usage of multiprocessors has appeared as an efficient computing mean for running real-time applications, especially those tasks which a uni-processor system can't execute. For this purpose, a successful algorithm is required which will optimally find out the time and schedule of the tasks. One can break a task into number of smaller sub tasks. Authors denoted a cluster as a DAG (Directed Acyclic Graph). In multiprocessor scheduling challenges, a fixed program is to be scheduled in a system which is a multiprocessor system that the program's execution time should be reduced. They used Genetic Algorithm method as it is the heuristic approaches which have the high capability to resolve the complex tasks like the task scheduling. These people developed a new Genetic Algorithm, which is called heuristics based Genetic Algorithm which can be used for scheduling of static tasks in homogeneous and parallel system in which its population magnitude and the number of generations hang on the number of tasks. This algorithm tends to reduce the completion time and surges the throughput of the system. Then they related the combined approaches which are heuristics based genetic algorithm (HGA) that is based on MET (Minimum execution time)/Min-Min heuristics, MET heuristic and First Come First Serve (FCFS) and b-level or t-level

precedence resolution and min-min heuristic with a pure Genetic Algorithm approach. They explained that the heuristics based method produces much better results in terms of quality of solutions. Their performance study is based on the best randomly generated schedule of the suggested GA.

Yi-Hsuan Lee, Cheng Chen proposed an improved Genetic Algorithm to schedule parallel program on multiprocessor system. They also constructed a simulation and evaluation environment to evaluate the execution of tasks in parallel and find a schedule that diminishes the completion time. They used the heuristic methods to obtain near-optimal solutions. Genetic Algorithms are used to solve this problem. Genetic algorithms are powerful but usually endure from lengthier scheduling time. But the suggested algorithm outweighs this drawback. The proposed algorithm combines the concept of Divide-and-Conquer mechanism to partition the entire problem into subgroups and solve them separately. It was shown that the recommended algorithm can not only decrease the scheduling time, but also get same type of performances as original genetic algorithms, sometimes it is even healthier. This feature makes the proposed algorithm more scalable and stretch out its practicability.

Fatos Xhafa and Javier Carretero gave Genetic Algorithms based schedulers for efficiently assigning jobs to resources in a Grid system. Two encoding schemes were used and most of Genetic Algorithm operators are employed on them. They showed that those schedulers which were based on GA are very rapid and hence can be employed to those jobs which are dynamically schedule and reached in the grid system by running in batch mode for a little time.

Suad Kasapovic and Marin Golub innovated an efficient method which is based on Genetic Algorithms to answer the multiprocessor scheduling of task to find out the tasks to the processors and the execution order of the tasks so that the execution time is diminished. They assumed the certain number of processors and tasks are shown by a directed acyclic graph (DAG) which is known as called task graph.

Optimization of Distributed Query: Two majors building blocks of an effective distributed system are query execution plans and design of database. Past research discovered that GA are also successfully and effectively used in optimizing distributed and centralized queries. Numbers of researcher have utilized different type of algorithms like ACO, PSO, Genetic Algorithm, Honey Bee etc. It was found that in the deterministic approaches, using these techniques, one is able to obtain an optimal query execution plan in seconds as compared to minutes, hours or weeks. Authors have applied many type of Genetic Algorithm which are restricted genetic approach and hybrid genetic approach to optimize OLTP and DSS queries, simple genetic approach, novel genetic approach and stochastic approach. In comparison to other variants of Genetic Algorithm, hybrid genetic approach gives better result. Navid Khlilzadeh Sourati and Farhad Ramezni have applied genetic algorithm to minimize the cost of shipping of the fragments in a distributed database system. Authors observed that their method is helpful in minimizing the cost of shipping by effectively allocating the fragments.

Machine Learning: Data Mining is also known as Knowledge data discovery. It is used to reveal or extract some meaningful information from the raw data. It is normally used to handle large or high volume data sets. Data mining is a broad subject that deals with association, prediction, correlation, classification and clustering.

Vineta Verma and Gunjan Verma have examined that Genetic Algorithm is efficiently applied in business optimization, pattern recognition and stock exchange data mining. To identify, one of the killer disease called lung cancer they also used artificial neural network to diagnose it. They observed that once the system is used to with application of Artificial Neural Network, the functionality in terms of accuracy enhance sharply.

William F. Punch, Behrouz Minaei-Bidgoli have applied GA for the development of web based educational application system. Authors have used GA to forecast and categorize the students who are based upon their logged in features. Using GA based support vector machines, Lean Yu has assessed the tendency of stock market. Authors have combined the features of genetic algorithm and support vector machine to explore the stock market policies. GA is used to assist in choosing the parameters and to enhance the speed of SVM in envisaging the exact position of stock market. Authors have empirically test different cases and found their proposed approach as best.

Natural language processing deals with the interaction among computers and natural (human) language, as spoken and written language bodies are being processed for many type of objectives. Authors discussed that the evolutionary computing can be efficiently applied in many phases of natural language processing like machine translation, machine text summarization, part of speech tagging etc. Enrique Alba has attempted to solve the different type of optimization techniques to solve the problem of word categorization. It was discussed that the coding of integer produces better results as compared to binary coding in GA. Authors found that the rate of accuracy of Genetic Algorithm is very closely similar to one of the method of word categorization i.e. Viterbi.

V. CONCLUSION:

The application of Genetic Algorithm has grown up exponentially in last few years. GA is used in different fields like Mechanical Engineering, Chemistry, Astronomy, Agriculture, Physics, Computer Science, Medical Science etc. Genetic Algorithm helps us to find the answer for NP-hard problems in few seconds. At the

same time the accuracy rate of genetic algorithm was also found very high. In this paper, an attempt was made to discuss briefly some of the important researches of different researchers who have tried and employed Genetic Algorithm to solve the complex problems which are related to Database, Software Engineering, and Machine Learning. The review shows that GA has been efficiently has been used in these areas for solving different and complex problems like quality improvement design optimization of database, software cost estimation, software testing, query optimization, data mining, grammar checking etc.

REFERENCES:

- [1]. Manik Sharma, Gurvinder Singh --. Stochastic Analysis of DSS Queries for a Distributed Database Design. International Journal of Computer Applications (IJCA), Vol. 82, No. 5.
- [2]. Chande Swati V, Sinha Madhvi. —Genetic algorithm: a versatile optimization tool. BVICAM's Int. J. Inf. Technol. Volume 1, Issue 1. 2008
- [3]. Sevinc Ender, Cosar Ahmat. —An evolutionary genetic algorithm for optimization of distributed database queries. The Computer Journal. Volume 54, Issue 5.2011
- [4]. Gunjan verma and Vineta Verma. —Role and Applications of Genetic Algorithm in Data Mining. International Journal of Computer Applications (0975 – 888) Volume 48– No.17, June 2012
- [5]. K. Venkatesh,N. Ganesan, . —Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data. International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 26. 2010
- [6]. Lars Bungum, Bjorn Gambäck. 2010. —Evolutionary Algorithms in Natural Language Processing. Norwegian Artificial Intelligence Symposium, Gjøvik, 22 November 2010.
- [7]. Enrique Alba a, Gabriel Luque and Lourdes Araujo. —Natural language tagging with genetic algorithms. Information Processing Letters 100 (2006)
- [8]. Praveen Ranjan Srivastava¹ and Tai-hoon Kim². —Application of Genetic Algorithm in Software Testing. International Journal of Software Engineering and Its Applications Vol. 3, No.4, October 2009.
- [9]. Behrouz Minaei-Bidgoli , William F. Punch.. —Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System. Genetic and Evolutionary Computation — GECCO 2003
- [10]. Lean Yu¹, Shouyang Wan², and Kin Keung Lai³. . —Mining Stock Market Tendency Using GA-Based Support Vector Machines. LNCS 3828, pp. 336 – 345, 2005.
- [11]. Navid Khilzadeh Sourati¹, Farhad Ramezani.. —Reducing Transfer Costs Of Fragments Allocation In Replicated Distributed Database Using Genetic Algorithms. Advances in Science and Technology Research Journal Volume 9, No. 25, March 2015, pages 1–6 DOI: 10.12913/22998624/1917
- [12]. S. Keshavarz and Reza Javidan. —Software Quality Control Based on Genetic Algorithm. International Journal of Computer Theory and Engineering, Vol. 3, No. 4, August 2011
- [13]. Yi-Hsuan Lee, Cheng Chen, —A Modified Genetic Algorithm for Task Scheduling in Multiprocessor Systems, Proc. Of 6th International Conference Systems and Applications, 1999.
- [14]. Javier Carretero, Fatos Xhafa, —Genetic Algorithm Based Schedulers for Grid Computing Systems, International Journal of Innovative Computing, Information and Control, ICIC International, Volume 3, Number 6, December 2007.
- [15].

Faheem Ahmad" Applications of Genetic Algorithm in Distributed Computing, Machine Learning and Software Engineering" International Refereed Journal of Engineering and Science (IRJES), vol. 08, no. 02, 2019, pp 09-12