Improved Distributed Genetic Algorithms Based on Their Methodologies and Processes

ARWA AL-EDAILY*
AEDAILY@KSU.EDU.SA

NADA AL-ZABEN*
NZABEN@KSU.EDU.SA

SHAREFAH AL-GHAMDI*
SHARFAH@KSU.EDU.SA

HATIM ABOALSAMH
HATIM@KSU.EDU.SA

*DEPT. OF COMPUTER SCIENCE
KING SAUD UNIVERSITY
RIYADH, KSA

Abstract— In this paper we evaluates the effectiveness of three different distributed genetic algorithms (DGAs). The first one is DGA with Diversity Guided Migration, second one is DGA with Automated Adaptive Migration and the last one is DGA with Bi-coded chromosomes and confidence rates. All these algorithms were investigated to improve the overall quality of solutions in the distributed genetic algorithm for different problems. Our comparison between those algorithms depended on some important factors ;like, achieving diversity in selecting individuals, process of replacing the individuals between subpopulations, computational time and memory space. As a result, DGA with Diversity Guided Migration (DGM), was nominated to be better than the other DGAs. It improves the performance for many problems and search spaces.

Key-words:- Genetic Algorithm, Distributed Genetic Algorithm , Distributed Genetic Algorithm performance

1 Introduction

Genetic Algorithms are search algorithms that are based on natural selection and natural genetic concepts. The GA aims to find global optimization solutions in a complex search space through specific procedures based on genetic models. Moreover, they are acknowledged as efficient and effective search in problem domains. For that, nowadays they are widely applied in many businesses, scientific and engineering applications [6] due to their proven features in improving solutions through powerful and computationally less complex searches. But due to the expansion of distributed systems [5], the growing in problems complexity and the need to decentralize the processes, this had led to the study of Distributed Genetic Algorithms (DGA).

Distributed genetic algorithms (DGAs) are based on the idea of dividing the population into a number of sub-populations [2]. Every sub-population evolves over a few generations and applies the GA's operations (selection, mutation, and crossover) on their own individuals (chromosomes), which are isolated before one or more individuals are exchanged between the sub-populations [3] [7]. DGAs aim to reduce the execution time and increase the numerical performance in complex applications [7].

After DGA proved itself in many problems, various approaches provide practical strategies to improve the DGA performance in execution speed, finding the optimal solution...etc [1] [7]. They became different in the implementation of DGA operations (crossover, mutation, migration) [1] [2] [3] or in the usage of specific techniques before the first generation or during the process of generating the sub-population [1].

Three DGA algorithms [2] [3] [4] were investigated in this paper to improve the overall quality of solutions in different problems. The first one is DGA with Diversity Guided that uses a replacement policy for migration captures a diverse range of individuals from the population instead of selecting individuals based on their fitness [2]. The second one is DGA with Automated Adaptive Migration that defines a set of relevant parameters and adjusts their values automatically depending on the environments of the individuals [3]. And the third one is DGA with Bi-coded chromosomes using new Evaluation function that represents each individual by two chromosomes and evaluate the effectiveness of the chromosomes by new fitness function [4].

This paper will highlight these three DGAs used to improve the overall quality of solutions in several problems and will produce comparison between them. The outline is as follows: Section 2 presents a background on our main concepts (Genetic Algorithm and Distributed Genetic Algorithm). Then section 3 will provide an overview of the three DGAs and how they act to find optimal solutions. In section 4 Comparison and contrast of different aspects of

those methods will be presented. Section 5 is discussion of our results, Section 6 concludes this paper.

2 Background on GA And DGA

Genetic Algorithm (GA) is search algorithm that are based on natural selection and natural genetic concepts which are nowadays acknowledged as efficient and effective search in problem domains. It aims to find global optimization solution in a complex search space and in order to apprehend its goal it is recommended to define its process in detail. The GA copies the laws of nature that govern natural selection. Conventionally it comprises a number of so called individuals as a possible solution of the optimization problem where each individual is represented (coded) in a string of bits. In each generation all the individuals are tested or their fitness value is valuated.

Subsequently the individuals are allowed to reproduce by mating. In nature the next generation is produced using a mating process. This is performed by two parents creating some offspring. The offspring will consist of the genetic material of both parents. There are three options regarding to the fitness of the offspring, they can be weaker, the same or fitter than their parents. If they are weaker they will tend to die out – if they are stronger their chances of survival are better. The mating process is done through the probability of a crossover operation which is done in three steps. Firstly, select a crossover point randomly then develop part one which consists of the first parent starting from first bit until the crossover point and part two will consist of the second parent starting from first bit until the crossover point. Then exchange part one with part two to produce the offspring.

This reproduction process results in two offspring sharing both parents' genetic characteristics. The variation caused by this process allows the offspring to search out different available features, find better fitness values and subsequently better solutions. This may be followed by mutation operation which changes a randomly selected string bit in the offspring. Then a new pair of parents is chosen and the reproduction process is repeated until the number of offspring reaches the original number of individuals then the process continues with the testing of the new generation. Optimization stops if the performance is sufficiently good or after a given number of populations.

Distributed Genetic Algorithms (DGA) was innovated due to the expansion of distributed systems along with the growing in problems complexity and the need to decentralize the process of GA. The DGA are based on the idea of dividing the population into a number of subpopulations. Then every subpopulation evolves over a few generations by applying the GAs operations (selecting-crossover- mutation) on their own individuals who are later on isolated before one or more individuals are exchanged between the subpopulations. The communication paths between subpopulations are determined by the topology of a DGA. The concept of exchanging individuals between the

subpopulations is called Migration. It is the key concept of most DGAs which their process is based on sending selected individuals from one population to another to replace individuals within the population in accordance with the DGAs replacement strategy in a phase called migration phase. This phase can occur asynchronously or synchronously when all subpopulations are synchronized [2]. After that, each sub-population resume as before until the next migration phase.

Note that this movement of individuals between subpopulations is the glue that holds subpopulations together and its importance rises in the improvement of the solutions quality in DGA and increases diversity within population and decrease premature convergence. The individuals in migration are called migrants and their number is migration rate. The importance of DGA is clear in reducing the execution time and increasing the numerical performance in complex applications [2][3].

2.1   Overview Of Selected DGAs

In this section we will provide a general description of the three different DGAs [2],[3]and [4] that we selected to compare in this paper. Each of these algorithms proposes different methods and strategies for some of basics in the DGA, such as migration process, fitness evaluation and representation of the chromosomes. In our overview we will explain each of the DGAs; what is the main idea of the algorithm?; how it acts?; what is its objective?; what the results obtained when it was tested?. In this section we provide all these algorithms from the view of their researchers.

2.2   DGA with Diversity Guided Migration

This algorithm proposed by [D. Power, C. Ryan and R. Azad, 2005]. The main idea of their algorithm was using the migration phase of DGA to inject the populations with a diverse range of individuals rather than just a cluster of highly fit individuals. This strategy is an alternative to the send-best-replace-worst standard migration scheme that sends the individual with the best fitness to the connected subpopulation and replaces the least fit one by this best individual. But in the Diversity Guided Migration Strategy the selection of individuals to be in the sending list it depends on the similarity between the individuals then on the fitness score.

How does this algorithm capture a diverse range of the individuals from the population? The individuals are selected based on their location within the population. So D. Power, C. Ryan and R. Azad proposed a population representative which is an individual that have the lowest mean distance score from all other individuals in the population. They used equation (1) to calculate the individuals mean distance:

$$\text{mean-distance } (ind_i) = \frac{\sum_{j=1}^{|p|} distance \ (ind_i , ind_j)}{|p| - 1} \quad (1)$$

Where |P| is the population size and distance is the hamming distance between two individuals, taken from equation (2).

$$distance \ (ind_i , ind_j) = \quad \sum_{k=1}^{n} \quad (ind_{ik} \otimes ind_{jk}) \quad (2)$$

Where $\otimes$ indicates the XOR-operator, k refers to the Gene position on the chromosome of individuals i and j.and n represents the problem length.

By knowing the population reprehensive, the individuals that are not located around the cluster of individuals at the centre of the population it can be selected and then add them to the migration send list. The population representative will be selected first to be in the sending list. After that, a number of the other individuals will be selected based on their distance from this population representative and whose fitness is greater than of the population representative. Then the same number of individuals will be selected based on their distance from this population representative and whose fitness is less than of the population representative. The number of individuals that could be selected as previous strategy it depends on the migration rate. The fittest individual in the population will be in the end of the sending list.

The replacement method in the migration that was proposed in this algorithm will replace clones instead of low scoring individuals. So the individuals with the same genetic information as other individuals (clones) will be in the top of the sending list, by order of their fitness the least fit will be replaced before clones with higher fitness scores. If there are no clones, those individuals with lowest fitness values will be replaced.

This replacement method aims to rescue the individuals from local maxima. And the motivation behind it was to try and prolong the existence of individuals that contain unique information, because the unique genetic information could potentially rescue the individuals from local maxima. Where In the event of a local maximum being found the similar of individuals may not contain enough genetic information for the receiving population to climb out of the local peak and find the overall global maximum.

TABLE 1: Fitness Results for different problems with 40% migration rate

Problem	Migration policy	Optimal found	Mean best fitness
Five Hills and Four Valleys	Standard	0	2.02455
	Diverse	**19**	**2.48811**
Waves	Standard	0	7.31735
	Diverse	**21**	**8.68075**
Two Hills	Standard	0	4.73849
	Diverse	**28**	**4.81496**
One Centre and Four Neighbors	Standard	0	1.29936
	Diverse	**42**	**1.49123**
Six Hump Camel Back	Standard	1	0.941727
	Diverse	**27**	**1.02874**

D. Power, C. Ryan and R. Azad tested their algorithm on a number of problems: One-max problem, Multi-peaked Three-Dimensional Landscapes problems. We present in Table 1 a summary of their fitness results in different five problems with 40% migration rate. Those results clarify that the optimal solution was found far more times than when the diversity guided migration strategy was used for all migration rates tested. Also results produced best individuals that are located around the global maximum.

This algorithm provided higher of the fitness score than standard migration strategy and improved solution quality with the mean best fitness values for all migration rates that was tested. In addition, when the migration rates are high and worst individuals are being replaced by similar highly fit individuals in every generation, then this would causes the population to converge quickly.

2.3 DGA With Automated Adaptive Migration

This algorithm is proposed by [H.Lee, B. Oh, J. Yang, and S. Kim, 2009] based on determining how many and which individuals move between subpopulations at each site adaptively. It helps the individuals from one subpopulation not be weeded out but adapted to the new subpopulation. For that they define set of relevant parameters their values adjusted automatically with regard to the environment. This helps to find a good solution faster.

Automated migration process and selection mechanism using aging was produced in this algorithm. The automated migration process aimed to avoid the conquest problem. However, if the best individual at each subpopulation was sent to another subpopulation, the global best individual may dominate in all subpopulations. This was called "the conquest problem" so this automated migration process could guarantee the avoidance of the conquest problem as well as the survival of migrants at their new places.

How the individual in this migration strategy will be chosen to be migrated? The individuals are chosen automatically according to the average quality of the

subpopulations. The individual that has fitness higher than the average fitness of the subpopulation it belongs to, will migrate. Then the individual with the lowest fitness will be replaced with the immigrated one. There for, H.Lee, B. Oh, J. Yang, and S. Kim calculate the average fitness of the subpopulation by equation 3:

$$Avg\_F(S_i) = \frac{F(ivd_{ij})}{|S_i|}$$

(3)

where $F(ivd_{ij})$ is the fitness of the ivd_{ij}.

Individuals migrate in two directions; the first direction was from subpopulation with lower average fitness to the subpopulation with higher average fitness. And the second direction is from subpopulation with higher average fitness to the subpopulation with lower average fitness. In the first direction the individual with higher fitness than the average fitness of the subpopulation it is joining, can migrate without any restriction. Where in the second direction emigrant must has higher fitness than the average fitness of the joining subpopulation but lower fitness than the average of the leaving subpopulation. This bi-directional migration enables both subpopulations to evolve and guarantees the average fitness of each subpopulation does not decrease. This is to decrease the possibility of conquest problem.

For new immigrants, their fitness (in terms of the classification accuracy against the data at the new site) can be low since they have been generated with respect to the data at the previous sites. This may cause their extermination after the migration. In order to avoid this, the researchers proposed selection mechanism using aging technique that is adopted in the fitness evaluation. This selection mechanism is used to select individuals for the evolutionary process. They considered two factors in terms of biological age and chronological age. The first one is replaced by the fitness of each individual for the subpopulation they belong to. Where the second one calculated by considering four cases: selection, crossover, mutation, and migration. At first, all individual's chronological ages set to zero. Then will change in the following three cases:

- If the individuals are selected and survive in the next generation their ages increase by one.
- The offspring generated by the crossover and the mutation are zero year old.
- When an individual emigrates to other subpopulation, its age is also initialized to zero in the new environment.

In this algorithm with aging technique the fitter and the younger individuals are more probable to be selected in the evolutionary process. So H.Lee, B. Oh, J. Yang, and S. Kim define the following selection probability:

$$SP(ivd_{ij}) = \alpha \times \frac{I\text{-}A(ivd_{ii})}{\Sigma_k(I\text{-}A(ivd_{ik}))} + (1-\alpha) \times \frac{F(ivd_{ij})}{\Sigma_k F(ivd_{ik})}$$

(4)

where $SP(ivd_{ij})$ is the selection probability of ivd_{ij}, $A(ivd_{ij})$ is chronological age of ivd_{ij}, I is total number of iteration in DGA, and is the constant weight for aging.

By this mechanism each subpopulation is able to produce improved performance without losing the diversity after migration. Because even though migrants may have low fitness in the new subpopulation they join due to the different data, their young age can relieve this weakness and participate in the evolutionary process successfully.

The algorithm was applied to the feature subset selection task. In the flow of feature subset selection using this DGA algorithm, at each site, GA was run independently. The feature subsets generated by GA were fed to the learning algorithm for fitness evaluation. They used for experiments six data sets from UCI machine learning. Each data set had different parameter values such as number of features, number of classes, number of samples, number of subpopulations, and total population size.

They compared the results of applying their DGA algorithm with the performance of the single GA in terms of the classification accuracy. The results stated that the DGA guides subpopulations into the right direction during the evolution. At first, they applied their DGA and used the automated migration but its selection probability was determined by the fitness only like the case of centralized GA then applied the DGA. It used both their migration process and the selection mechanism. The latter case outperforms the DGA without aging for most datasets.

2.4 DGA with Bi-Coded Chromosomes and a New Evaluation Function

This algorithm was proposed by [T. Hamdani, A. Alimi, and F. Karray, 2006]. It is a solution for feature extraction problem. It uses DGA based on bi-coded chromosomes and new fitness function. In the bi-coded chromosomes representing; a binary chromosome denoting the presence or absence of a feature in the selected vector; a real one representing the pertinence rate if the pointed feature is selected and the probability of selection in a genetic operation elsewhere. The confidence rates help in estimating the contribution of each feature.

Each chromosome is evaluated in its fitness through an optimization function in order to survive to the next generation. This fitness function takes into account two major parameters to compute the evaluation of a chromosome; the classification performance and the proportion of the used features relatively to the number of all defined features. They applied uniform method for crossover for the bi-coded chromosome.

This solution that they proposed is composed of heterogeneous genetic algorithm which is used to identify the

interesting sizes and several homogenous genetic algorithms that will do local optimization. That is to minimize the complexity and to accelerate the algorithm convergence. Because the Genetic algorithms with heterogeneous population are more efficient but have more complex task and spend much more time to converge. So the results can be approved with making a first selection with the heterogeneous algorithm to identify the interesting sizes to compute. The homogeneous algorithms will take the results of the fast exploration to make fine selection in a specific size to reinforce the precision.

The algorithm was applied on six different data sets. Each data set had different parameter values such as the number of Training points, number of Testing points, Number of features, number of classes and recognition rate. They implemented and tested some of classical algorithms to compare these algorithms performance with their solution. They found that their DGA increase the global performance which cost more computational time to achieve this high performance.

3 Comparison

As we proposed in the previous sections, the three DGAs have been explained with their results. To compare between those algorithms, a number of different factors will be considered in this section depending on the criteria, processes and calculations in each algorithm.

Each DGA achieved its aim in improving the performance as shown in the different experiments applied for the three algorithms. But it is difficult to compare the improving percentage for the performance after applying them because they were tested on different problems with different datasets and different DGA parameters and conditions. [2] [3] [4].

One of the factors that can be compared between the three algorithms is based on achieving the diversity in selecting chromosomes for the migration process in each generation. Diversity is very important to reduce premature convergence and give more search space [3].

To compare between the three DGAs, we assumed applying them under the same conditions and calculated the diversity Percentage between chromosomes chosen for the migration process (equation 1). Fig. 1 shows the diversity Percentage comparison for one generation containing five individual and need to choose three of them for migration process. The individuals have these fitness's and distances respectively: *Points:* A,B,C,D,E,F. *Fitness:* 12,11,10,8,6 *Distances:* 2 from A to B, 3 from A to C, 6 from A to E, 4 from B to E, 5 from C to E, 5 from C to F and finally, 12 from E to F.

Replacement process when sending the individuals between subpopulation is very important effect. DGA with Diversity Guided Migration chooses a special individual to replaced (clones) that is very useful in helping escape from local maxima. DGA with Automated Adaptive Migration and DGA with Bi-coded chromosomes and new evaluation function replace the worst individuals only which may handicap a subpopulation by removing unique genetic algorithm.

The computational time and memory space for fitness functions are important factors for comparison but they are different from one algorithm to another because they are related to the problem procedures and the size of the problem space. Ignoring the fitness calculations – because they are different from problem to another problem – we just compared between the complexities for the extra calculations in each strategy as shown in Table 2.

From Table 2.and equations 1, 2, 3 and 4 it can be seen that the time complexity is O(1) but the difference in the extra calculations must be considered in comparison because it affects the processor execution time especially for huge problems and deep search spaces.
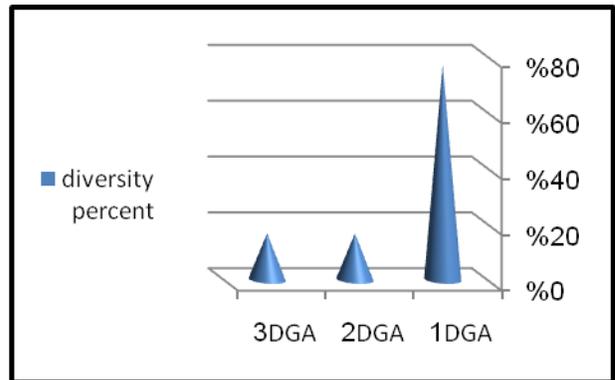

Figure 1. A comparison between individuals diversity chosen for migration process in DGA with Diversity Guided Migration , DGA With Automated Adaptive Migration, And DGA with Bi-Coded Chromosomes

TABLE 2: Extra calculations comparison

Algorithm	Computational time and memory
DGA with Diversity Guided Migration	- In each generation, *mean distance* and *distance* between each individual must be calculated (equation 1 and 2). - In equation (2) XOR operator may take extra time depending on the machine hardware.
DGA with Automated Adaptive Migration	- In each generation, the *selection probability* (equation 3) and the *average fitness* of each subpopulation (equation 4) must be calculated. - Performs bi-directional migration process which take more time in the network communication time.
DGA with Bi-coded chromosomes and new evaluation function	- No extra calculations in addition with the fitness. - Each individual needs two different chromosome coding representations which consume more memory space.

4 Results and Discussions

From previous sections it is clear that all the three DGAs improved the performance but depending on other factors discussed in the comparison section, we can nominate one algorithm to be the best then followed by the others.

DGA with Diversity Guided Migration can be chosen as the best strategy because of its aim to achieve a high percent of diversity in each subpopulation in each generation (Fig. 1 as an example) which is a strong advantage for its strategy and very helpful in different problems with different sizes. Also, this algorithm avoided losing unique individuals in the replacement process. Although XOR may consume extra time in execution, this can be avoided by perform the XOR in the software than perform it by the hardware (and that always happens). So, DGA with Diversity Guided Migration still is considered the best between these three algorithms.

DGA with Automated Adaptive Migration competes the DGA with Diversity Guided Migration but its complexity is higher than DGA with Diversity Guided Migration (see Table 2) because it changed the evolution process itself using bi-direction migration process in each generation. And unfortunately, it is not concerned with the diversity and a locked searching in the local search space will have higher probability to happen. [2]

DGA with bi-coded chromosomes and new evaluation function comes at last because it very complex and it needs high memory space for big problems that have a lot of chromosomes and many generations which seems that this algorithm may lack under reality conditions. Also, it does not care about the diversity in choosing chromosomes as DGA with Automated Adaptive Migration.

As we have discussed and analyzed above, DGA with Diversity Guided Migration becomes the best and more efficient compared with DGA with Automated Adaptive Migration and DGA with bi-coded chromosomes and new evaluation function depending on some factors.

Applying these algorithms on one problem under same conditions and parameters will show the best algorithm clearly. But still, analyzing the criteria of each algorithm can make good idea about the best one.

5 Conclusion

This paper presented three different DGAs that improve the overall quality of solutions in many problems. Some comparisons are studied to find the more efficient algorithm through some factors except fitness complexity and performance improving percentage because they differ from problem to another. We established that DGA with Diversity Guided Migration is the best and the more efficient algorithm regarding to avoid some local search lacking by achieving the diversity in each subpopulation and applying new process to replace the individuals at the end of the

migration process rather than replace the worst one only to avoid destroying unique individuals with low fitness.

A suggested future work could be testing the three algorithms into same problems and with same parameters and datasets. That gives strict judgment about these strategies and help in categorizing DGAs depends on the problem.

References

[1] M. Miki, Y. Yamamoto, S. Wake, and T. Hiroyasu. "Global Asynchronous Distributed Interactive Genetic Algorithm" in *Systems, Man and Cybernetics, SMC '06. IEEE International Conference*, 2006, pp. 3481–3485.

[2] D. Power, C. Ryan and R. Azad. "Promoting diversity using migration strategies in distributed genetic algorithms" *in the Evolutionary Computation, 2005 the 2005 IEEE Congress on*, 2005, pp. 1831-1838

[3] H.Lee, B. Oh, J. Yang, and S. Kim. "Distributed Genetic Algorithm using Automated Adaptive Migration" in *the Evolutionary Computation, 2009 the 2009 IEEE Congress on*, 2009, pp. 1835-1840

[4] T. Hamdani, A. Alimi, and F. Karray. "Distributed Genetic Algorithm with Bi-Coded Chromosomes and a New Evaluation Function for Features Selection" in *2006 IEEE Congress on Evolutionary Computation*, 2006, pp. 581-588

[5] P. Whigham and G. Dick ,"AVoronoi-based distributed genetic algorithm" in *proceedings of the 15th annual colloquium of the Spatial Information Research center* ,2003, pp.133-138

[6] S. Sivanandam, and S. Deepa , (2008) ,*Introduction to genetic algorithms*. [online book], available at: http://books.google.com/books?id=wonrLjj2GagC&pg=PA106&lpg=PA106&dq=distributed+genetic+algorithm&source=bl&ots=A8eFhxRP6W&sig=ZygANBOQcWMVYdHzz-__gvYTz8c&hl=ar&ei=PTbDTPHjKYXf4gayo8m5Aw&sa=X&oi=book_result&ct=result&resnum=6&ved=0CDIQ6AEwBThG#v=onepage&q=distributed%20genetic%20algorithm&f=false

[7] H. Huang, " Efficient hybrid distributed genetic algorithms for wind turbine positioning in large wind farms" in *presented in Industrial Electronics,2009. ISIE 2009. IEEE International Symposium*, 2009, pp. 2196-2201