



GLOBAL JOURNAL OF ADVANCED RESEARCH
(Scholarly Peer Review Publishing System)

An Analysis of Migration in the Island Model on Feed Forward Neural Network

Htet Thazin Tike Thein

Student

University of Computer Studies in
Yangon, Myanmar

Khin Mo Mo Tun

Department of Computational
Mathematics

University of Computer Studies in
Yangon, Myanmar

ABSTRACT

In this paper an analysis of migration in island model on feed forward neural network is presented. Feed forward neural network is one of the classification algorithms which, is currently a 'hot' research area in medicine. Constructing a classification model is important in machine learning for a particular task. In this paper, island-based parallel differential evolution algorithm is chosen and applied to feed forward neural network to enhance the learning process and the network learning is validated in terms of convergence rate and classification accuracy. The proposed system implements the island-based training method to be better accuracy and less training time by using and analyzing between four different migration topologies. The proposed method can be used to classify medical datasets with higher accuracy and less training time.

Keywords

Artificial Neural Network, Differential Evolution Algorithm, Island Model, Migration Topology, Medical Diagnosis

1. INTRODUCTION

Classification is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. Many problems in business, science, industry, and medicine can be treated as classification problems. Neural networks have emerged as an important tool for classification. Medicine advances on all fronts to improve the care of patients and defeat this disease of the century. Because of this, it is essential that several disciplines continue to make their contribution and particularly data mining or artificial Intelligence. To provide assistance to the medical, robust and reliable diagnosis, neural networks can be a powerful tool for distributed diagnosis [3]. Traditional algorithm uses the gradient based approach which either trains slowly or get stuck with local minimum. Instead of using gradient-based learning techniques, one many apply the commonly used optimization methods such as Genetic Algorithm (GAs), Particle Swarm Optimization (PSO), Ant Colony Optimization, Differential Evolution Algorithm (DE) to find the network weights. A variant of evolutionary computing namely the differential evolution [1] is a population based stochastic optimization method similar to genetic algorithm [2] that finds an increasing interest in the



recent year as an optimization technique. DE is a stochastic general search method, capable of effectively exploring large search spaces.

Increasing pressure to solve real world complex problems has led to the development of Parallel Evolutionary Algorithms (PEAs) which exploit the intrinsically parallel nature of EAs. An extensive review of parallel evolutionary models, parallel implementations, and pressing theoretical issues can be found in [7]. Parallelization of an evolutionary algorithm can be done at any of the following levels: objective function evaluation level (master-slave model), population level (island model or migration model) and elements level (cellular level). The first two lead to coarse grained parallelization while the cellular model leads to fine grained parallelization. Over the past few years, considerable amount of work has been done on parallelization using island model (IM) strategy [8], [9], [10], [11], [12]. [9] gives a detailed study on the effects of number and size of the population in IMs. Another study on the various parameters of the island models have been done in [10], [11].

In this paper, we tested the performance of the neural networks by using island-based parallel differential evolution algorithm (IBPDEA) based on medical datasets. In this paper, a parallel approach, which uses neural network technique, is proposed to help is trained with medical datasets by using feed forward neural network model. The performance of the network is evaluated. This paper presents a detailed study on the various migration topologies of an island model based differential evolution learning scheme in feed forward neural network.

2. BACKGROUND

2.1 Neural Network

Neural network (NN) have been widely used in various fields as an intelligent tool in recent years, such as artificial intelligence, pattern recognition, medical diagnosis, machine learning and so on [4]. The major steps of using neural network learning algorithm can be summarized as follows: to begin with, through the provision of training samples and the class of sample, the network prediction of each sample is compared with the actual known class label, and then the weight of each training sample is adjusted to achieve the purpose of classifying other sample data.

2.2 Parallel Differential Evolution

The Differential Evolution (DE) algorithm was proposed by Price, Storn and Lampinen in 1995 [13]. It is a population based optimizer or metaheuristic that starts generating a population of D-dimensional vectors whose initial values are randomly obtained based on the limits defined by the inputs of the algorithm. The total of individuals in the population is a known constant value.

Each individual belongs to a generation g , i.e., let $X_{i,g} = (x_{i,g}^1, \dots, x_{i,g}^D)$ an individual of the population, with $i = 1, \dots, N$ where the index i denotes i -th population individual, g determines the generation to which the individual belongs and N is the total number of individuals in the population.

The main idea of the method is to use vector difference in order to modify the population vector. This idea has been integrated into a recombination operator of two or more solutions, consisting of two phases (mutation and crossover), with the aim of guiding the search towards "good" solutions.

In the following, we explain the three classical main operators of DE.

Mutation: After initialization, DE mutates and recombines the current population to produce another one constituted by N individuals. The mutation process begins in each generation selecting random individuals $X_{r_1,g}$, $X_{r_2,g}$. The i -th individual is perturbed using the strategy of the formula (1), where the indexes i , r_1 , r_2 and r_3 are integers numbers different from each other, randomly generated in the range $[1, N]$.



$$V_{i,g+1} = X_{r1,g} + (X_{r2,g} - X_{r3,g}) F \quad (1)$$

The constant F represents a scaling factor and controls the difference amplification between individual r_1 and r_2 , and it is used to avoid stagnation in the search process.

Crossover : After the mutation phase, each perturbed individual $V_{i,g+1} = (v_{i,g+1}^1, \dots, v_{i,g+1}^D)$ and the individual $X_{i,g} = (x_{i,g}^1, \dots, x_{i,g}^D)$ are involved in the crossover operation, generating a new vector $U_{i,g+1} = (u_{i,g+1}^1, \dots, u_{i,g+1}^D)$, denominated "trial vector", and obtained using the expression (2).

$$U_{i,g+1}^j = \begin{cases} u_{i,g+1}^j & \text{if } rand_j \leq Cr \text{ or } j = k \\ x_{i,g}^j & \text{in other case} \end{cases} \quad (2)$$

where $j = 1, \dots, D$, and $k \in \{1, \dots, D\}$. The constant $Cr \in [0; 1]$, denominated crossover factor, is a parameter of the algorithm defined by the user. Cr is used to control the values fraction that is copied from the mutant vector V . $rand_j$ is the output of a uniformly distributed random number generator, and is generated in each comparison made on the vector components. The value k is a randomly generated index chosen for each individual. The vector component for that index is taken from the mutated vector to ensure that the trial vector is not exactly equal to its source vector $X_{i,g}$.

There are two crossing operators that can be applied: binomial or exponential. Both types use the expression (2), but differ in the way it is applied. The binomial crossover operator copies the j th parameter value from the mutant vector $V_{i,g+1}$ to the corresponding element in the trial vector $U_{i,g+1}$ if $rand_j \leq Cr$ or $j = k$. Otherwise, it is copied from the corresponding target (or parent) vector $X_{i,g}$. Instead, the exponential crossover operator inherits the parameters of trial vector $U_{i,g+1}$ from the corresponding mutant vector $V_{i,g}$ starting from a randomly chosen parameter index. Then, it continues copying the parameter values from the mutant vector $V_{i,g}$ till the j th parameter value satisfying $rand_j > Cr$. The remaining parameters of the trial vector $U_{i,g+1}$ are copied from the corresponding target vector $X_{i,g}$. To complete the notation, when the crossover applied is binomial the method is named "DE/rand/1/bin". If an exponential crossover is used, it is referred as "DE/rand/1/exp".

Selection: This phase determines which element will be part of the next generation. The objective function of each trial vector $U_{i,g+1}$ is evaluated and compared with the objective function value for its counterpart $X_{i,g}$ in the current population. If the trial vector has less or equal objective function target value (for minimization problems) it will replace the vector $X_{i,g}$ in the next generation population. The scheme followed is presented in the expression (3).

$$X_{i,g+1} = \begin{cases} U_{i,g+1} & \text{if } f(U_{i,g+1}) \leq f(X_{i,g}) \\ X_{i,g} & \text{in other case} \end{cases} \quad (3)$$

The three stages mentioned above are repeated from generation to generation until the specified termination criterion is satisfied. This criterion could be finding a predefined minimal error or reaching a certain number of iterations.

Due to the potentialities of DE for solving optimization problems, in recent years, numerous variations and methods have been proposed with the aim of improving the performance of the classic technique. Among them are those trying to adapt



DE parameters such as self-adjusting [14], [15]; others using different mechanisms to optimize the individuals selection for the mutation and selection phases [16], and some combining both methods [17]. In several studies DE have been used in combination with other metaheuristics, obtaining hybrid approaches [18], [19], [20], in order to improve the solutions quality, or to accelerate the speed of convergence. In addition, there are several studies that incorporate parallelism in order to improve the quality of the solutions obtained and/or diminish the execution time of this metaheuristic.

2.3 Island Models

Independent runs suffer from obvious drawbacks: once a run reaches a situation where its population has become stuck in a difficult local optimum, it will most likely remain stuck forever. This is unfortunate since other runs might reach more promising regions of the search space at the same time. It makes more sense to establish some form of communication between the different runs to coordinate search, so that runs that have reached low-quality solutions can join in on the search in more promising regions.

In island models, also called distributed EAs, coarse-grained model, or multi-deme model, the population of each run is regarded as an island. One often speaks of islands as subpopulations that together form the population of the whole island model. Island evolves independently as in the independent run model, for most of the time. But periodically solutions are exchanged between islands in a process called migration.

The idea is to have a migration topology, a directed graph with islands as its nodes and directed edges connecting two islands. At certain points of time selected individuals from each island are sent off to neighboring islands, i.e., islands that can be reached by a directed edge in the topology. These individuals are called migrants and they are included in the target island after a further selection process. This way, islands can communicate and compete with one another. Islands that got stuck in low-fitness regions of the search space can be taken over by individuals from more successful islands. This helps to coordinate search, focus on the most promising regions of the search space, and use the available resources effectively.

In the island model approach, each island executes a standard sequential evolutionary algorithm. The communication between sub-population is assured by a migration process. Some randomly selected individuals (migration size) migrate from one island to another after every certain number of generations (migration interval) depending upon a communication topology (migration topology). The two basic and most sensitive parameters of island model strategy are: migration size, which indicates the number of individuals migrating and controls the quantitative aspect of migration; and migration interval denoting the frequency of migration. Although different aspects of migration size and interval were studied in the past, we are unaware of any work studying directly the influence of these parameters on the behavior of island model based differential evolution, though [6] presents a similar study on a set of 8 standard functions.

3. THE PROPOSED MODEL

The proposed system can be used to classify the medical datasets using four different migration topologies based parallel differential evolution algorithm and compare the results of four different topologies. The overview of the proposed system is presented in Figure 1.

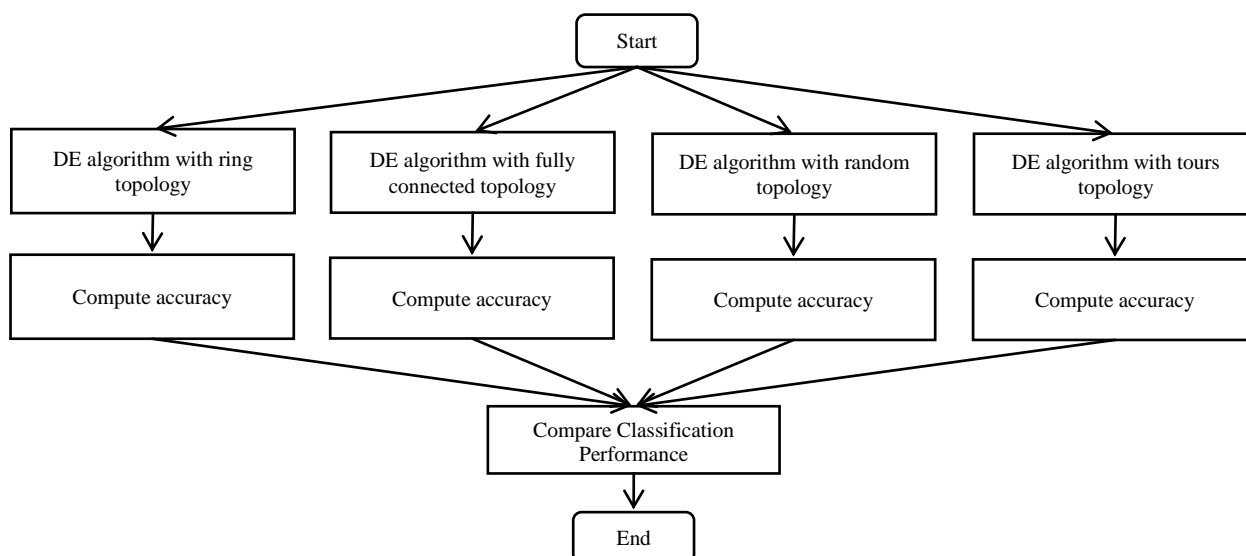


Figure 1: IBPDEA System flow

4. EXPERIMENTS AND PERFORMANCE EVALUATION

4.1 Data Set Description

This paper used three medical datasets for testing.

- 1) Heart: There are 270 instances, 13 continuous and 2 classes. The attribute are age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood pressure, serum cholesterol, fasting blood sugar > 120 mg/dl, resting electrocardiographic results, maximum heart rate achieved, exercise induced angina, oldpeak, the slope of the peak exercise ST segment, number of major vessels and thal. The classes are absent (1) and present (2).
- 2) Liver: There are 345 instances, 6 continuous attributes and 2 classes. The attributes are mean corpuscular volume, alkaline phosphatase, alamine aminotransferase, aspartate, aminotransferase, gamma-glutamyl transpeptidase and number of halfpint equivalents of alcoholic beverages drunk per day. The classes are absents (1) and present (2). The first 5 attributes are all blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption. Each record is a single male individual.
- 3) Breast Cancer: The cancer dataset requires the decision maker to correctly diagnose breast lumps as either benign or malignant based on data from automated microscopic examination of cells collected by needle aspiration. The dataset includes 8 inputs and 2 outputs. A total of 683 instances are available in breast cancer data set. 456 instances are used for training and 227 instances are used for testing.

All the medical datasets are from University of California Irvine (UCI) machine learning repository.

4.2 Data Normalization

The data normalization is considered as the most important pre-processing step using neural networks. To improve the performance of multilayer neural networks, it is better to normalize the data entry such that will be found in the interval of [0 1]. To transform the data into digital form, and use it as inputs of the neural network, scaling or normalization should be realized for each attribute. The nine numerical attributes, in the analog form, are scaled with a range of 0 and 1. There are many types of normalization that are found in the literature. The new values obtained after normalization, follow this equation:



$$\text{New value (after normalization)} = \frac{\text{current} - \text{min}}{\text{max} - \text{min}} \quad (4)$$

4.3 Classifier Accuracy

Estimating classifier accuracy is important since it determines to evaluate how accurately a given classifier will label future data, data on which the classifier has not been trained. Accuracy estimates also help in the comparison of different classifiers. The following classification features are used to train and test the classifier.

Given: A collection of labeled records (training set). Each record contains a set of features (attributes), and the true class (label).

Find: A model for the class as a function of the values of the features.

Gold: Previously unseen records should be assigned a class as accurately as possible. A test set is used to determine the accuracy of the model. Usually, the given data set is divided into training and test sets, with training set used to build the model and test set used to validate it. The Sensitivity and Specificity measures can be used to determine the accuracy measures.

Precision may also be used to access the percentage of samples labeled as for example, “cancer” that actually are “cancer” samples. These measures are defined as

$$\text{Specificity} = \frac{t_neg}{neg} \quad (5)$$

$$\text{Precision} = \frac{t_pos}{t_pos + f_pos} \quad (6)$$

$$\text{Sensitivity} = \frac{t_pos}{pos} \quad (7)$$

Where,

t_pos = the number of true positives (“medical dataset class” samples that were correctly classified as such class),

pos = the number of positive (“medical dataset class”) samples

t_neg = the number of true negative (“not medical dataset class” samples that were correctly classified as such class)

neg = the number of negative samples

f_pos = number of false positive (“not medical dataset class” samples that were incorrectly labeled as such class)

$$\text{accuracy} = \text{sensitivity} \frac{pos}{pos + neg} + \text{specificity} \frac{neg}{pos + neg} \quad (8)$$

4.4 Experimental Details

This paper presents the classification results of island-based parallel differential evolution algorithm on three datasets namely Heart, Liver and Breast Cancer. These datasets are downloaded from the UCI machine learning repository [5]. Java programming language, which is the platform independent and a general-purpose development language, is



used to implement the proposed system. First of all, the medical datasets are accessed. And then, normalization is made for pre-processing steps according to the different medical data sets. Experiments are performed with best-worst migration policies. Currently the system experiment the island model with best-worst migration policy and four different migration topologies. In the experiments we used identical islands, i.e islands with same parameters. We used four different topologies for our experiments. Simulations were run with setups of five, seven and nine islands. As can be observed that the error value of test data are nearly same for the five and seven island models, it goes a little high for nine islands, but no significant change in the number of islands. Some experiments were also conducted with four islands and it had no significant influence on the algorithmic behavior. We therefore chose a modest five island setup for our rest of the experiments. The policy of migration used was best-worst policy in which best string from an island replaces other worst string of another island based on four different topologies. The island model used the iteration as the migration interval and one-third of the old population is used to migrate and replace. The results for each dataset are compared and analyzed based on the convergence rate and classification performance.

Table 1. Dataset Information

Dataset	Number of records	Number of attributes	Attributes Values
Heart Diseases	303	14	Numerical
Liver Disorders	345	7	Numerical
Breast Cancer	683	8	Numerical

Table 2. Comparison on four different topologies

Datasets	Ring Topology		Fully Connected Topology		Torus Topology		Random Topology	
	Accuracy (%)	Time (sec)	Accuracy (%)	Time (sec)	Accuracy (%)	Time (sec)	Accuracy (%)	Time (sec)
Heart Diseases	100	9	100	11	100	39	98.889	13
Liver Disorders	84.4156	10	83.9827	10	82.684	43	83.1169	14
Breast Cancer	99.342	10	100.0	12	100	46	99.5614	16

5. PERFORMANCE COMPARISON

This analysis is carried out to compare the results of four different migration topologies. To do this, the learning patterns for the proposed system is compared using all three medical datasets. The comparative correct classification percentage for all datasets is shown in Figure 2.

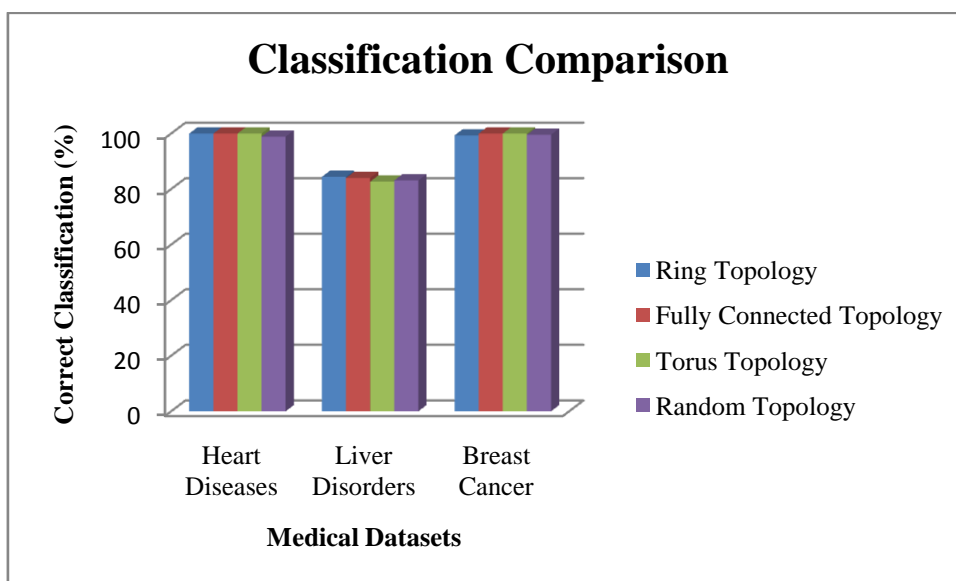


Figure 2: Comparison of Four Different Migration Topologies

In this paper, we have proposed four different migration topologies with best-worst policy and compared their results. Figure 2 shows the accuracy comparison with four migration topologies by using medical datasets. Medical datasets are used to implement this proposed system which shows better experiments with higher accuracy. The proposed system also reduces the computing time. For heart diseases ring, fully connected and torus topology get 100% correct classification percentage. For liver disorders dataset, the results show that ring topology has better results on convergence time and correct classification percentage. For breast cancer dataset, fully connected and torus topology classification results are better than ring and random topology. For all medical datasets, the experiments show that torus migration topology torus topology is more require computing time than other topologies. The proposed algorithm converges in a short time with high correct classification percentage.

6. CONCLUSION

This paper focuses on the analysis of various migration topologies on feed forward neural network which employs DE as the learning strategy. In this paper, application of island-based differential evolution algorithm has been experimented for classification of medical dataset. The proposed system implements the island-based training method to be better accuracy and less training time by using and analyzing between four different migration topologies with best-worst migration policy. The optimal network connection weights can be obtained by using island-based parallel differential evolution algorithm is a stochastic general search method, capable of effectively exploring large search spaces. The proposed system shows substantial improvement in classification accuracy. Through the results analysis it was found that the subpopulation model reduces significantly the computing time and the solutions quality is improved significantly. According to experiments, the torus topology is more require long training time than other three topologies but it solution quality results are similar to other topologies. Based on the analyses, it shows that torus topology is better results compared to other topologies on some medical datasets.



7. REFERENCES

- [1] Storn, R., and Price, K., "Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces", Journal of Global Optimization, vol.11, 341-359, 1997.
- [2] Vavak, F., Fogarty, T. and Jukes, K "A genetic algorithm with variable range of local search for tracking changing environments", In Proceedings of the 4th Conference on Parallel Problem Solving from Nature, 1996.
- [3] Arun George Eapen, master thesis, Application of Data Mining in Medical Applications, Waterloo, Ontario, Canada, 2004.
- [4] D. E. Goldberg. "Genetic Algorithms in Search, Optimization and Machine Learning," Addison Wesley, Reading, MA, 1989.
- [5] <https://archive.ics.uci.edu/ml/datasets.html>
- [6] Z. Skolicki and K. De Jong, "The influence of migration sizes and intervals on island models," in Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2005), ACM Press, 2001.
- [7] E.Alba and M. Tomassini, "Parallelism and Evolutionary Algorithms, IEEE Transactions on Evolutionary Computation," vol. 6, no. 3, pp 373-388, Aug. 1998.
- [8] Z. Skolicki and K. De Jong, "Improving evolutionary algorithms with multi-representation island models," in Parallel Problem Solving from Nature - PPSN VIII 8th International Conference, Springer-Verlag, 2004.
- [9] D. Whitley, S. Rana and R. B. Heckendorn, "The island model genetic Algorithm: On separability, population size and convergence, " Jnl.of Computing and Information Technology, vol. 7, no. 1, pp. 33-47, 1999.
- [10] E. Cantu-Paz, "Migration policies, selection pressure, and parallel evolutionary algorithms," Jnl. of Heuristics, vol. 7, no. 4, pp. 311-334, 2001.
- [11] R. Storn, "On the usage of differential evolution for function optimization," in Proc. Of the 1996 Biennial Conference of the North American Fuzzy Information processing society – NAFIPS, Edited by: M. H. Smith, M. A. Lee, J. Keller and J. Yen, June 19-22, Berkeley, CA, USA, IEEE Press, New York, pp 519-523.
- [12] L. Singh and S. Kumar, "Parallel Evolutionary Asymmetric Subsethood Product Fuzzy-Neural Inference System: An Island Model Approach," in Proceedings of the International Conference on Computing: Theory and Applications (ICCTA-2007), pp. 282-286, 2007.
- [13] Price, K., Storn R., Lampinen J.: Differential Evolution: A Practical Approach to Global Optimization. Springer. New York , 2005.
- [14] Yang, Z., Tang, K., Yao, X.: Self-adaptive Differential Evolution with Neighborhood Search. In: Proc. of the IEEE Congr. on Evol. Comp., pp. 1110-1116 2008.
- [15] Brest, J., Zamuda, A., Bokovie, B., Maucec M., Zumer, V.: High-Dimensional RealParameter Optimization using Self-Adaptive Differential Evolution Algorithm with Population Size Reduction. In: Proc. of the IEEE Congr. On Evol. Comp., pp 2032-2039 2008.
- [16] Martinez, C., Rodriguez, F., Lozano, M.: Role differentiation and malleable mating for differential evolution: an analysis on large-scale optimisation. In: Soft Computing, vol. 15, issue 11, pp. 2109-2126 2011.
- [17] Yang, Z., Tang, K., Yao, X.: Scalability of generalized adaptive differential evolution for large-scale continuous optimization. In: Soft Computing, vol. 15, issue 11, pp. 2141-2155 2011.
- [18] Ali, M., Pant, M., Nagar, A.: Two local Search Strategies for Differential Evolution. In: 5th IEEE Inter. Conf. on Bio-Inspired Computing: Theories and Appl., vol. 2, pp 1429-1435 2010.



GLOBAL JOURNAL OF ADVANCED RESEARCH
(Scholarly Peer Review Publishing System)

- [19] Noman, N., Iba, H.: Accelerating Differential Evolution Using an Adaptive Local Search. In: Proc. of the IEEE Trans. on Evol. Comp., vol. 12, pp 107-125 2008.
- [20] Tardivo, L., Cagnina, L., Leguizamon, G.: A Hybrid Metaheuristic Based on Differential Evolution and Local Search with Quadratic Interpolation. In: XVIII Congreso Argentino de Ciencias de la Computacion 2012.