

A Novel Immune Optimization with Shuffled Frog Leaping Algorithm - A Parallel Approach for Unsupervised Data Clustering

Suresh Chittineni
Professor, Dept. IT
ANITS
Visakhapatnam

P.V.G.D. Prasad Reddy
Professor, Dept. CS&SE
AU Engineering College
Visakhapatnam

Suresh Chandra
Satapathy
Professor, Dept. CSE
ANITS
Visakhapatnam

ABSTRACT

Data clustering is one of the data mining task, it is used to group the data objects according to their similarity. It is an optimization problem to find optimal results apply the proposed parallel approach called P-AISFLA. This hybrid algorithm is developed by utilizing the benefits of both social and immune mechanisms. The social algorithm Shuffled Frog Leaping Algorithm is a new parameter free population based algorithm combined with Clonal selection algorithm CSA.

This hybrid algorithm performs the parallel computation of social behavior based SFLA and Immune behavior based CSA to improve the ability to reach the global optimal solution with a faster and a rapid convergence rate.

The proposed algorithm PAISFL is applied for the data clustering applications and proved that it produces optimal results than SFLA and PSO.

Keywords

Data clustering, Shuffled Frog Leaping Algorithm (SFLA), CLONALG and P-AISFLA.

1. INTRODUCTION

Data Clustering is the process of dividing the data objects into clusters in such a way that there is a high similarity between the data within the cluster and have dissimilar data when compared to the other clusters. There are different types of cluster methods, which are mainly partitioning and hierarchical.

Partitioning clustering methods are popular due to its simple nature. These algorithms include K-means, ISODATA and Learning Vector Quantizers (LVQ). [1][2] In order to analyze the data many statistical techniques have been proposed such as analysis of variance(AOV), linear regression(LR), discriminate analysis(DA), canonical correlation analysis(CCL) multi-dimensional scaling, factor analysis, principal component analysis, and cluster analysis.

K-means Algorithm:

K-means algorithm is a method of vector quantization, which is popular for data clustering [3]. It is a popular partition method for cluster the objects into K number of clusters where K value has to be known in prior [4]. In this algorithm cluster membership is being determined with the help of calculating the centroid for each group and assigning such each object to the group with the closest centroid [5]. This approach minimizes all the overall within-cluster dispersion by iterative reallocation of cluster members.

In general K-means takes set S of objects and an integer K as input and gives the partition of S into subsets S_1, S_2, \dots, S_n . Here

the optimisation criterion is sum of squares. Let us consider x_r^i as the r^{th} element of S_i and $|S_i|$ as the number of elements in S_i and the distance between the elements x_r^i and x_s^i as $d(x_r^i, x_s^i)$. Here the cost function is considered as the sum of the squares and it is determined as follows:

$$C(S_i) = \sum_{r=1}^{|S_i|} \sum_{s=1}^{|S_i|} (d(x_r^i, x_s^i))^2 \quad (1)$$

In regular K-means works by the calculation of centroid for each cluster S_i , and it is denoted as x^i and the cost function for optimizing is as follows:

$$C(S_i) = \sum_{r=1}^{|S_i|} (d(x^i, x_r^i))^2 \quad (2)$$

The main aim of this algorithm is to minimize the cost function:

$$C(S_1) + \dots + C(S_k) \quad (3)$$

Clustering is an optimization problem, to find the optimal results of clustering with the help of the swarm intelligence in particular Shuffled Frog Leaping Algorithm (SFLA), Clonal Selection Algorithm (CSA) and hybridization of SFLA and CSA

The objective function of the clustering is based on squared error, which is mentioned in the equation (2).

Swarm intelligence mainly simulates the population behaviors of the complicate system in nature or society. Shuffled Frog Leaping Algorithm (SFLA), designed from the food hunting behavior of the frog and Clonal Selection Algorithm (CSA), designed from the immune system of the human body are such examples of this field.

SFLA is a population based memetic algorithm that carries the method of memetic evolution in the form of infection of ideas from one individual to another in a local search. A shuffling methodology allows for the exchange of information between local searches to move toward a global optimum [6]. This algorithm is used to calculate the global optimal of complex optimization verified [7].

On the other side Clonal Selection Algorithm (CSA) is an Artificial Immune System Technique (AIS) that is formulated by the functioning of Clonal selection theory of acquired immunity. The algorithm provides two mechanisms for searching for the desired final pool of memory antibodies. The first is a local search provided via affinity maturation (hyper mutation) of cloned antibodies [8]. The second search mechanism provides a global scope and absorbs the insertion of randomly generated antibodies to be inserted into the population to further increase the diversity and provide a means for potentially escaping from local optimal [9].

Both these algorithms are most powerful and more efficient in finding the global optimal solution. SFLA attain this with the help of social behavior and local search exploitation. CLONALG achieves this by performing the process of cloning and Hyper mutation. Both algorithms ensure powerful search capability in different ways. So, in order to improve the global search capability, to speed up the rate of convergence and to achieve a stable solution, these two algorithms are combined to form a hybrid evolutionary algorithm. A hybrid algorithm is developed by making use of the benefits of both social and immune mechanisms [10]. This hybrid algorithm performs the computation of both SFLA and CLONALG in a parallel fashion. This Hybrid evolutionary algorithm conserve the features of both SFLA and CLONALGSA and helps to ensure the global optimal in a more rapidly way with a better preserved mean solution. This hybrid algorithm is tested on various standard and well known bench mark optimization methods are found that, the proposed hybrid algorithm P-AISFLA outperforms both SFLA and CLONALG.

The residue of the paper is organized as follows: section II describes the Shuffled Frog Leaping algorithm (SFLA) subsequently a brief description of Clonal Selection Algorithm (CSA) in section III. The section IV described hybrid algorithm of SFLA and CLONALG. Section V covers experiment results over clustering and followed by conclusion of the work.

2. PROPOSED WORK

2.1 Shuffled Frog Leaping Algorithm (Sfla)

SFLA is a memetic algorithm inspired by the research of food hunting behavior of frogs. It is based on evolution of memes carried by the interactive frogs and by the global exchange of information among themselves. SFLA is a combination of both deterministic and random approaches [7].

The SFL algorithm progresses by transforming “frogs” in a memetic evolution. In this algorithm, frogs are seen as hosts for memes and described as a memetic vector. Each meme consists of a number of memo types[13]. The memo types represent an idea in a manner similar to a gene representing a trait in a chromosome in a genetic algorithm. The frogs can communicate with each other, and can improve their memes by infecting (passing information) each other [10]. Improvement of memes results in changing an individual frog’s position by adjusting its leaping step size.

2.2 (Clonalg)Immune Optimization Algorithm

It is a population based Meta heuristic algorithm and its core search power relies on cloning operator and mutation operator [14]. The Clonal Selection Algorithm (CSA) replicates only those individuals with high affinities and selects their improved and matures progenies. This strategy suggests that the algorithm performs a greedy search, where single members will be locally optimized and the new comers yield a broader exploration of the search-space [9]. This characteristic makes the CSA very suitable for solving optimization tasks. The basic steps mentioned below.

- Step1. Antibody Pool (AB) Initialization
- Step2. Antibody Selection
- Step3. Cloning
- Step4. Affinity Maturation
- Step5. Antibody Restraint
- Step6. Convergence Check

3. HYBRID ALGORITHM COMBINATION OF SFLA AND CLONALG

Both SFLA and CLONALG are evolutionary Algorithms which have their own advantages. Both have different kinds of features and also different implementation procedures. [18] Let us examine the advantages of SFLA and AIS.

a. Importance of SFLA

- A memetic algorithm which has the property of local search.
- Shuffling operator allows for Infection and exchange of ideas to take place.
- Exhibits social behavior
- Few parameters are involved i.e. Number of memplexes, Number of frogs, Iterations in each memplex.

b. Importance of CLONALG

- Cloning operator allows for replication of best fit antibodies from the antibody pool.
- Mutation operator brings diversification to the antibody population.
- Exhibits the Immune Behavior.
- Antibody Diversity Maintenance further increases the diversity and prevents from reaching to local optimal.

c. Hybrid Evolution of SFLA and CLONALG (P-AISFLA):

Both the advantages from Social and immune behaviors are incorporated in the proposed Hybrid approach. [18]Social behavior based SFLA and Immune behavior based CLONALG are combined together to evaluate a Hybrid Algorithm that has the advantages of both the conventional Algorithms.

According to this algorithm, initially consider 2N population was randomly generated within the given range. Then determine their Corresponding fitness or affinity values. Then sort all the population in the ascending order basing on their Computed fitness or affinity values. Then divide the entire 2N population in to two groups each with N population. Dividing should be done in such a way that, the best N Sorted population should go to one group and the remaining N sorted population should go to another group.

Since SFLA consists of social behavior component, the worst group of population should be further processed by SFLA and the best group of population should be processed by CSA since CSA has immunity mechanisms.

After processing each group by the algorithm, the result of the two algorithms is merged again to become 2N population and the obtained 2N population should be repeated by the same above process. [11][12]This whole of a process is repeated for specific number of iterations or until Convergence value is attained as the stopping criterion. The entire working of this Hybrid Evolutionary algorithm can be shown in Fig.1.

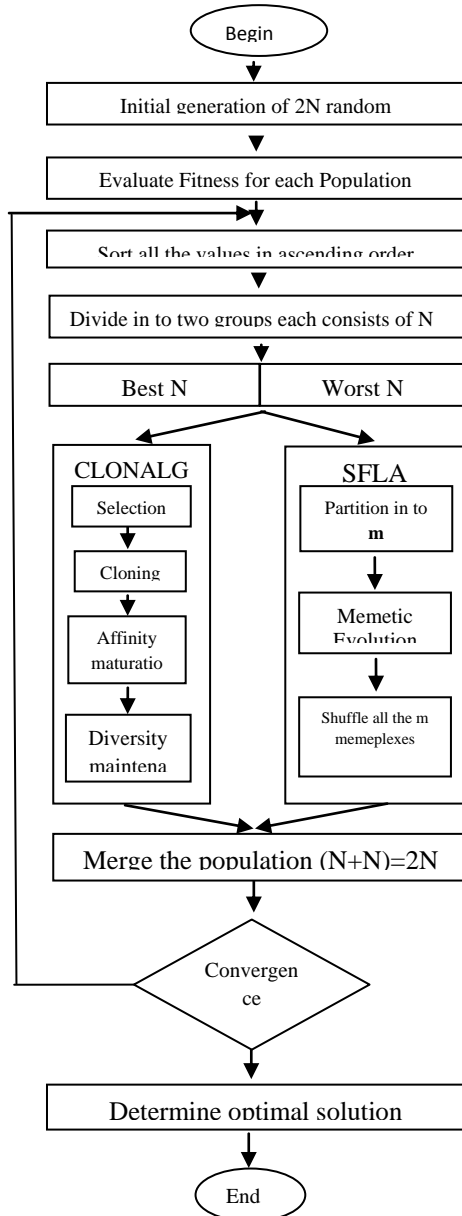


Fig.1. Basic Flow chart of P-AISFLA

4. EXPERIMENTS AND RESULTS

4.1 Experiments And Results Over Swarm Based Clustering:

A. Benchmark Functions for Simulation

A suite of standard benchmark functions [15] [16] [17] are taken into consideration to test the effectiveness and efficiency of the proposed approach P-AISFLA with the conventional CLONALG and SFLA for cluster the standard dataset optimally. All these benchmark functions are meant to be minimized. The first function is unimodal with only one global minimum. The others are multimodal with a considerable number of local minima.

B. Experimental Setup:

All the three algorithms were written in Matlab scripting language. The matlab codes were run on a computing Laptop machine at 1.8 GHz Intel processor. Each algorithm optimized each function 30 times independently.

The following simulation conditions are used for CLONALG:

- Initial Antibody Pool Size, $AB = 20$
- Clonal Multiplying factor's Range $\beta = [0.5-1]$
- Mutation type used: Gaussian
- Gaussian mutation probability $P_{mg} = 0.01$
- Percentage of Suppression $P_{sup} = 0.2$
- Number of Iterations=1000
- Number of Dimensions taken for each Benchmark Function=1

The following simulation conditions are used for SFLA Algorithm

- Initial Population, $N_{pop} = 20$
- Number of Memeplexes, $M = 2$
- Max. Local search iterations 10
- Number of Main Iterations=1000
- Number of Dimensions taken for each Benchmark Function=15

The number of generations used for CLONALG, SFLA and the proposed hybrid approach is P-AISFLA was 30. In each of the 30 computations, CLONALG, SFLA and P-AISFLA employed different initial populations randomly generated in the problem's search space.

A Successful Optimal Solution represents that it converges to the global optimum. Otherwise, the algorithms were regarded as getting stuck or trapped at the local optimal

Due to Cloning, mutation and local search operators available in P-AISFLA, the global search capability of it is dominant in comparison with CLONALG and SFLA and its convergence speed is predominantly faster than CLONALG and SFLA in terms of function evaluations for majority of the functions.

It can be observed that the proposed approach possesses a relatively higher success rate than both CLONALG and SFLA implying that P-AISFLA has nice global search performance to avoid local optimal and the optimal solution gained by it is the most accurate.

The results are shown in the Table.1 get from the clustering methods with 20 simulation runs with standard deviation denotes the convergence. *Hybrid SFLA & CLONALG* performs somehow better fitness in terms of quantization error and inter_cluster. Benchmark data set is taken from the UCI machine repository.

Table 1: Quantization error and Standard deviation of clustering methods with using evolutionary methods.

Data Sets	Algorithm	Quantization error, std	Inter – cluster distance, std
Iris	<i>K means</i>	0.834, 0.22076	9.3238, 1.6821
	SFLA	0.026209, 0.017972	16.41, 2.4693
	CLONALG	0.023201, 0.09345	21.079, 3.8171
	<i>Hybrid SFLA & CLONALG</i>	0.011477, 0.01947	20.278, 0.2204

Hayes Roth	<i>K means</i>	11.961, 1.573	8.9353, 1.2419
	<i>SFLA</i>	0.77086, 0.0408	324.25, 5.7895
	<i>CLONALG</i>	3.99, 3.3429	313.1, 3.4562
	<i>Hybrid SFLA & CLONALG</i>	0.55914, 1.6488	325.51, 66.738
Wine	<i>K means</i>	116.29, 0.83715	2019.2, 0.234
	<i>SFLA</i>	10.765, 3.7278	3272.8, 292.89
	<i>CLONALG</i>	33.622, 7.6328	2859.7, 339.91
	<i>Hybrid SFLA & CLONALG</i>	4.9763, 3.3043	3598.7, 483.11
Diabetes	<i>K means</i>	78.995, 7.6753	21.12, 3.521
	<i>SFLA</i>	69.234, 2.4840	30.22, 2.820
	<i>CLONALG</i>	36.98, 2.397	35.109, 2.564
	<i>Hybrid SFLA & CLONALG</i>	30.204, 2.6501	38.0173, 2.4714

The results are simulated and compared with K-means and swam based clustering algorithm. The quality of the clustering is measured based on the quantization error, inter-cluster distance. The intra-cluster distance is the distance between the centroid and the existing objects in the cluster.

The results are reported and averaged over 20 run of simulations. The algorithm used 20 particles in hybrid SFLA&CLONALG the seed is derived from the resultant centroid of K-means which ran initially. The derived seed from K-means acts like one of the particle in the initial swarm of particles in swam based methods. In other words while implementation only n-1 particles are randomly generated. In the implementation process of the algorithm starting weight is used at 0.9 and the final weight used at 0.4. To get a better convergence the acceleration coefficients are used at 1.042. Swam based algorithms in particular hybrid SFLA&CLONALG performs better than K-means based on quantization error.

Table 2: Result showing for the wine data with 20 runs

Method	Function		
	F _{best case}	F _{average case}	F _{worst case}
SFLA	51.5133	51.9344	52.9962
CLONALG	51.7634	51.5645	51.1353
<i>Hybrid SFLA & CLONALG</i>	48.0271	47.4692	48.1309

Graphical representation for table - 2 data

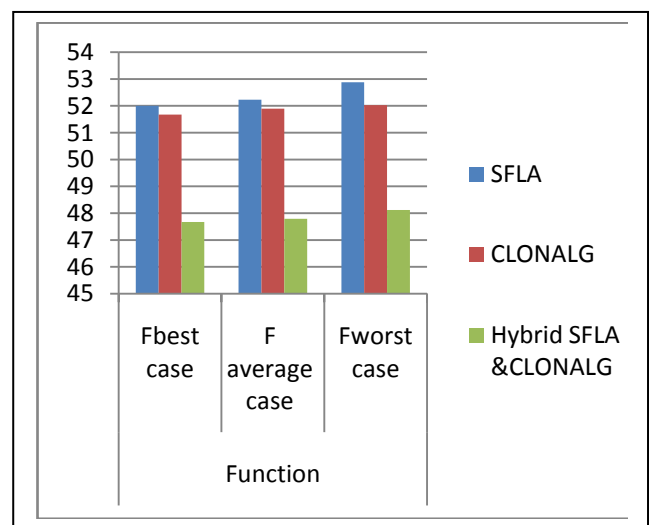


Table 3: Results showing for the IRIS data set with 20 runs

Method	Function		
	F _{best case}	F _{average case}	F _{worst case}
SFLA	52.1204	52.4534	52.5467
CLONALG	52.0011	51.5698	52.2378
<i>Hybrid SFLA & CLONALG</i>	46.6323	47.8894	48.34520

Graphical representation for table - 3 data

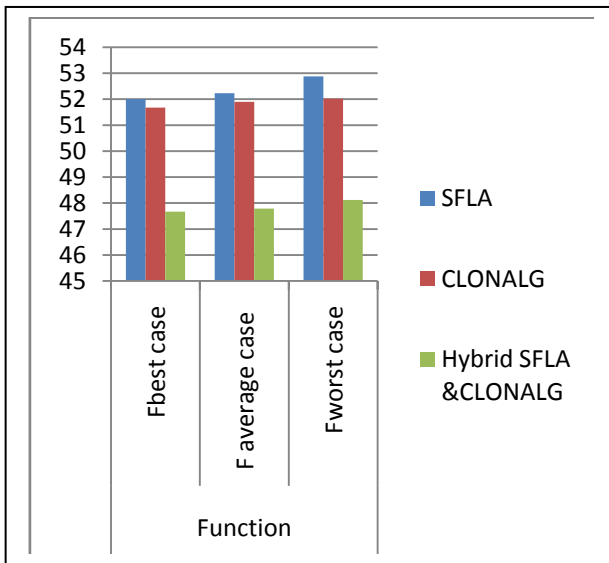


Table 4: Results showing for the Hayes Roth with 20 runs

Method	Function		
	F _{best case}	F _{average case}	F _{worst case}
SFLA	52.2334	52.2314	52.8765
CLONALG	50.96773	51.8974	51.0324
Hybrid SFLA & CLONALG	48.37453	48.3594	49.0504

Graphical representation for table - 4 data

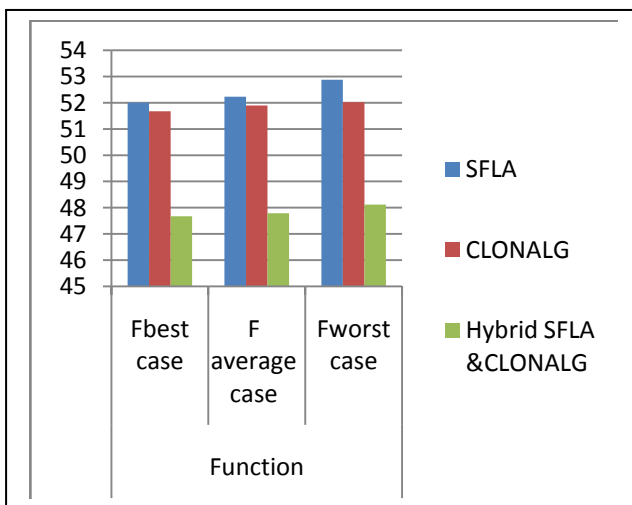
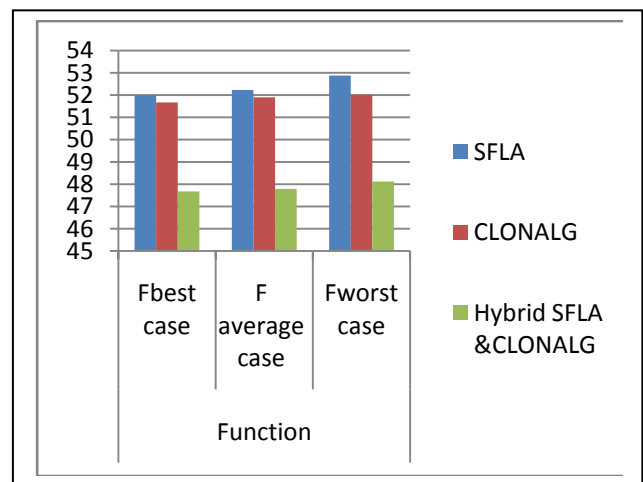


Table 5: Results showing for the diabetes with 20 runs

Method	Function		
	F _{best case}	F _{average case}	F _{worst case}
SFLA	52.0034	52.2314	52.8765
CLONALG	51.6733	51.8974	52.0234
Hybrid SFLA & CLONALG	47.6721	47.7893	48.1205

Graphical representation for table – 5



The result shows that the proposed Hybrid method optimization approach can be considered as a viable and an efficient heuristic to find optimal or near-optimal solutions to clustering problems of allocating N objects to k clusters.

The result shows that the strong point in favor of swarm intelligence based clustering. Due to this reason many machine learning, data mining researchers prefer to use swarm based clustering.

5. CONCLUSIONS

In this work a novel and hybrid approach is used, P-AISFLA that utilizes the benefits of both cloning operator and local search operator in a parallel fashion for the clustering technique to get optimal cluster results. The local search operator and Mutation operators are used for improving local and global search capabilities. The parallel search mechanism is utilized to improve the performance of search efficiency of the algorithm. On solving a suite of benchmark functions, P-AISFLA performs better than both CLONALG and SFLA in terms of Mean optimal solution values, stability of the optimal solution, number of function evaluations and success rate. Swarm intelligence techniques are applied on data clustering methods to get optimized results in terms of quantization error and standard deviations..

6. REFERENCES

- [1] A. Jain, M. Murty, and P. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, 1999, pp. 264–323.
- [2] M. Halkidi and M. Vazirgiannis, "Clustering validity assessment: Finding the optimal partitioning of a data set," in *Proc. IEEE ICDM*, San Jose, CA, 2001, pp. 187–194.
- [3] Tapas Kanungo, Nathan S. Netanyahu, Angela Y. Wu "An efficient K-means clustering algorithm: analysis and implementation" *IEEE Transaction on pattern analysis and machine intelligence*, vol 24, No-27,pp 881-892,july 2002.
- [4] K. Krishna and M. Murty, "Genetic K-means algorithm," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 3, Jun. 1999,pp. 433–439.
- [5] Raghavan,V. V., Birchand, K., Paterlinia, S. &Krink, T. (2006). Differential evolution and particle swarm optimization in partitional clustering. *Comput. Stat. Data Anal.*50, no. 5, 1220–1247.
- [6] Emad Elbeltagi, Tarek Hegazy, and Donald Grierson, "Comparison among five evolutionary-based optimization algorithms", *Advanced Engineering Informatics*, 2005 (19), pp. 43-53.
- [7] M. M. Eusuff, K. E. Lansey, F. Pasha, "Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization," *Engineering Optimization*, 2006, vol. 38, no. 2, pp. 129-154.
- [8] Leandro N. de Castro and Fernando J. Von Zuben,"Learning and Optimization Using the Clonal Selection Principle", *IEEE Trans. on Evolutionary Computation*, Vol. 6, No. 3, JUNE 2002, pp. 239-251.
- [9] De Castro and Jonathan Timmis. *An Introduction to Artificial Immune Systems:A New Computational Intelligence Paradigm*, Springer Verlag, 2002.
- [10] Liong S-Y, Atiquzzaman Md. Optimal design of water distribution network using shuffled complex evolution. *J Inst Eng, Singapore* 2004;44(1):93–107.
- [11] Xuncaizhang^{1,2}, Xuemei Hu³, Guangzhao Cui², Yanfeng Wang², Ying Niu², "An Improved Shuffled Frog Leaping Algorithm with Cognitive Behavior", *Proceedings of the 7th World Congress on Intelligent Control and Automation June 25 - 27, 2008, Chongqing, China.*
- [12] Nareli Cruz-Cortes, Daniel Trejo-perez,A.Coello Coello, "Handling Constraints in Global Optimization using an Artificial Immune System".*ICARIS-2005, Springer LNCS*,pp 234-247,2005.
- [13] Jonathan Timmis, C.Edmonds and Kelsey, "Assessing the Performance of Two Immune Inspired Algorithms and a Hybrid Genetic Algorithm for Function Optimisation," *Proceedings of the Congress on Evolutionary Computation*, 2004, pp.1044-1 051.
- [14] Lijun Pan and Z Fu, *A Clonal Selection Algorithm for Open Vehicle Routing Problem: Proceedings of third International Conference on Genetic and Evolutionary Computing*, 2009.
- [15] XQ Zuo and YS Fan,"A chaos search immune algorithm with its application to neuro-fuzzy controller design", *Chaos, Solitons and Fractals ;30 (2006) 94–109.*
- [16] Suganthan and S. Baskar ,"Comprehensive Learning Particle Swarm Optimizer for Global Optimization of Multimodal Functions", *IEEE Trans. on Evolutionary Computation*, , Vol. 10, No. 3, June 2006.
- [17] S. H. Ling and C. Iu," Hybrid Particle Swarm Optimization With Wavelet Mutation and Its Industrial Applications", *IEEE Tran. On Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 38, No. 3, June 2008.