International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 01, Issue – 09, December – 2016, PP – 64-70

An Enhancement for the optimization of feature selection to perform classification Using Meta Heuristic Algorithms

T Sai Durga[1] , V.Yasaswini[2]
[1]Assistant Professor, Department of Computer Science Engineering,
[2]Final Year, Department of Computer Science Engineering,
Kakinada Institute of Engineering & Technology For Women,
Korangi, India.

Abstract : The dimensionality of the feature space when being high affects the classification accuracies and the computational complexity due to redundant, irrelevant and noisy features present in the dataset. Feature Selection extracts the more informative and distinctive features from any dataset to improve the classification accuracy. Nature Inspired Algorithms are famous meta-heuristic search algorithm used in solving combinatorial optimization problems. Previously, we have proposed FS algorithms based on ACO, ABC, EABC and by the convincing results produced by these algorithms we have proposed Firefly Algorithm(FA),Cuckoo Search(CA) ,Harmony Search(HAS) for feature selection procedure. This paper proposes a new method of feature selection by using FA to optimize the selection of features. Ten UCI datasets have been used for evaluating the proposed algorithm. Experimental results show that, FA-Feature Selection has resulted in optimal feature subset configuration and increased classification accuracies.

1. INTRODUCTION AND BACKGROUND

Pattern classification is the task of assigning or classifying an input pattern to predefined categories or classes and the algorithm that does the task of classification is called the classifier. The task of the classifier is to partition the feature space into class-labeled decision regions based on their category. Classification is a twofold process: learning and deciding. In learning, the classifier identifies the invariant and common properties of a set of samples (subset) that characterize a class and then constructs a complete trained model. In deciding, the classifier uses the previously trained model to assign a category of class to any of the new instance that arrives based on the commonness of their properties to those of the set of samples [1].

To classify any particular domain, the classifier takes the entire set of attributes or features in the dataset and each feature in the feature space has different effect on the performance of the classifier. Some features might be irrelevant to classification and may lack in increasing the discriminative power of the classifier whereas some others are relevant and highly correlated to specific classification [2, 3 and 4]. Extracting such features relevant to classification is very important which is achieved through Feature Selection (FS). Feature Selection is a process of finding a minimal feature subset while retaining a suitably high accuracy in representing the original features.

FS when seen as an optimization problem, obtaining the optimal subset of features is very important. Evolutionary and swarm intelligent algorithms like Genetic Algorithm, Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Artificial Bee Colony algorithm (ABC) and Enhanced Artificial Bee Colony algorithm (EABC) have been widely used in optimizing FS [5 -12 and 27]. Though many works on optimization of FS has been carried out, none of them has been proved to be a consistent performer for the entire application domain. In this manner, our previously proposed algorithm to optimize FS using Ant Colony Optimization, Artificial Bee Colony and Enhanced ABC algorithm had been proved to be successful in optimizing the feature subset.

In Literature, Feature Selection optimization in Ant Colony and Artificial Bee Colony have been overthrown by the recently evolving nature inspired algorithms such as Firefly Algorithm, Cuckoo Search Algorithm, Harmony Search Algorithm and Bat Algorithm [20]. Firefly Algorithm (FA) is a meta-heuristic algorithm, which works by the flashing behavior of the fireflies that guides the FS procedure. The primary purpose for a firefly's flash is to act as a signal system to attract the other fireflies [19]. Cuckoo Search Algorithm (CSA) is developed on the brooding parasitism of the cuckoo species and simulates by the egg laying behavior of the species in another host nest that guides the FS procedure [25]. Harmony Search Algorithm (HSA) is a music-based meta-heuristic optimization algorithm which is conceptualized using the musical process to search for the perfect state of harmony [26].

CSA and HSA will be implemented and later while Bat Algorithm (BA) when analyzed for the FS procedure it doesn't show positive results and thus it won't be applicable for the feature selection procedure. This study proposes a novel feature selection method in which FA is used to generate the feature subsets and a classifier is used to evaluate the feature subsets generated by FA. In this method, each firefly is allocated to a feature (food source) and the less bright firefly tries to make all possible combinations of finding the brightest firefly and this configures the feature subset. The proposed algorithm has shown competitive performance compared to ACO, ABC and EABC based on feature selection.

This paper is organized as follows: Section 2 gives a brief description about feature selection and classification related to feature selection. The concept of Firefly algorithm is explained in section 3. The proposed method is discussed in section 4. The concept of Firefly algorithm is explained in section 5.The concept of Firefly algorithm is explained in section 6. Computations and results are discussed in section 7. Section 8 concludes the paper.

2. FEATURE SELECTION AND CLASSIFICATION

2.1 FEATURE SELECTION

Feature Selection (FS) is a commonly used pre processing step in data mining, especially when dealing with high dimensional space of features. The process of feature selection requires two important components: Evaluation function and Generation procedure. Evaluation function is to evaluate the candidate feature subset and Generation procedure is to generate the candidate feature subsets. When the evaluation function makes use of a classifier to evaluate the generated feature subsets, it is called as wrapper approach. When a classifier is not involved and feature subsets are evaluated by looking into the intrinsic properties of data, it is known as Filter approach [24].

The main objective of FS is to choose a subset of features from the original set of features. Feature selection is extensive and it spreads throughout many fields which includes text categorization, machine learning, pattern recognition, and signal processing. Considering the entire features may slowdown the learning process and may reduce the performance of the classifier because of redundant and irrelevant features. Thus it is essential to reduce the number of features by selecting the most relevant features to represent a dataset. FS allows the reduction of feature space, which is crucial in reducing the training time and improving the prediction accuracy. This is achieved by removing irrelevant, redundant, and noisy features. (i.e.,selecting the subset of features that can achieve the best performance in terms of accuracy and computational time).

2.2 CLASSIFICATION

A classifier takes a set of features as input and these features have different effect on the performance of classifier. Some features are irrelevant and have no ability to increase the discriminative power of the classifier. Some features are relevant and highly correlated to a specific classification [1 and 6].

For classification, sometimes obtaining extra irrelevant features is very unsafe and risky [17]. A reduced feature subset, containing only the relevant features helps in increasing the classification accuracy and reducing the time required for training.

3. FIREFLY ALGORITHM

FA depends on the variation of light intensity and the formulation of attractiveness [19]. The FA is a nature-inspired, population-based meta-heuristic that employs three idealized rules: (i) All fireflies within a population are unisex, so that one firefly will be attracted to other fireflies irrespective of their sex; (ii) Attractiveness between fireflies is proportional to their brightness, implying that for any two flashing fireflies, the less bright one will move towards the brighter one. Attractiveness and brightness both decrease as the distance between fireflies increases. If there is no brighter firefly within its visible vicinity, then a particular firefly will move randomly; and (iii) The brightness of a firefly is determined by the landscape of the objective function. Based upon these three rules, the basic operational steps of the FA are summarized within the pseudo-code of Figure 1.

Objective Function F(X), X = $(x_1, x_2, \ldots, x_d\ )$
Generate the initial population of n fireflies, X_i, i = 1, 2, … , n
Light intensity I_i at X_i is determined by F(X_i)
Define the light absorption coefficient γ
while (t < maxgeneration)
for i = 1: n , all n fireflies
for j = 1: n , all n fireflies (inner loop)

if ($I_i < I_j$)
Move firefly i towards j;
end if
Vary attractiveness with distance r via e- γr
end for j
end for i
Rank the fireflies and find the current global best solution G*
end while
Postprocess the results

FIGURE.1 PSEUDO CODE OF FA

4. FIREFLY ALGORITHM BASED FEATURE SELECTION (FA-FS)

Based on the original Firefly Algorithm we have proposed two more enhancements on the existing FA that is choosing the brightest firefly and choosing the comparatively brightest firefly. Features are considered as fireflies and hence the number of brightest and comparatively brightest fireflies is equal to the number of features in the dataset.

The attractiveness of a firefly is determined by its brightness which in turn is associated with the encoded objective function. In the simplest case, the brightness of a firefly at a particular location X would be its calculated objective value F(X). However, the attractiveness β between the fireflies is relative and will vary with the distance r_{ij} between firefly i and firefly j. In addition to this, light intensity decreases with the distance from its source, and light is also absorbed in the medium, so that the attractiveness should be allowed to vary with the degree of absorption. If the distance r_{ij} between any two fireflies *i* and *j* located at x_i and x_j, respectively is calculated using the Euclidean norm, then the movement of a firefly *i* that is attracted to another more attractive (i.e. brighter) firefly *j* is determined by equation 1.

$$X_j = X_i + \beta_0 \exp(-\gamma\, r_{ij})^2)(X_i - X_j) + \alpha\, \varepsilon_I \quad (1)$$

Where, x_i is the solution pointed by the current firefly (Classification Accuracy) and x_j is the solution pointed by the brightest firefly and comparatively brighter firefly (First method and Second method). β_0 is the attractiveness measure between 0 and 1. γ is the variation of attractiveness whose value is chosen between 0.1 and 10. Distance r_{ij} is set to 1. α is a randomization **parameter normally selected within the range [0,1]** and ε_i is a vector of random numbers drawn from either a Gaussian or uniform (generally [-0.5,0.5]) distribution.

In choosing the brightest firefly method, initially, the classifier Decision Tree J48 evaluates the discriminating ability of each individual feature in the dataset. The classification accuracy x_i of each feature i is calculated by employing 10-fold cross validation [6 and 22]. Then, the objective function F_i is calculated for each feature from its indiscriminability relation. Then the firefly computes the new solution x_j by using the classification accuracies of the feature and the less bright firefly points to the feature with the brightest accuracy. If the new solution x_j is greater than x_i , the less brightest firefly will be pointing to the feature subset that consists of the newly selected feature. If x_j is not greater than x_i then, the less bright firefly feature will be retained and the newly selected feature is neglected. The new solution x_j is computed by using equation (1).

In choosing the comparatively brightest firefly method, the same method of choosing the classifier is done using the J48 classifier to evaluate the discriminating ability of each individual feature in the dataset and the classification accuracy is obtained using 10-fold cross validation. Also, the objective function F_i is calculated for each feature. Then the firefly computes the new solution x_j by using the classification accuracies of the feature and the already existing brighter firefly points to the feature with a still brightest firefly. In this method, the firefly goes about randomly in the large search space to choose the brightest firefly and forms the feature subset with the newly paired up firefly. The new solution is computed by using equation (1).

In any given optimization problem, for a very large number of fireflies *n* >> *k* where *k* is the number of local optima, the initial locations of the *n* fireflies should be distributed relatively uniformly throughout the entire search space. As the FA proceeds, the fireflies would converge into all of these local optima (including the global ones). By comparing the best solutions among all these optima, the global optima can easily be determined. [19] demonstrates that the FA will approach the global optima when n $\to \infty$ and the number of iterations *t*, is set so that *t* >>1. In reality, the FA has been found to converge extremely quickly.

5. CUCKOO SEARCH ALGORITHM

The Cuckoo Search Algorithm (CSA) being idealized by its breeding behavior was tested on engineering optimization and embedded design problems. The results being obtained by this algorithm for engineering optimization problems are quite convincing in its results. Hence forth, we have implemented Cuckoo Search Algorithm for FS. Since the ABC algorithm does not consider the initial population to be feasible, in this Cuckoo Search Algorithm we have decided to add three constraints such as Eviction, Abandon and Survival. This type of method uses its historical memories for the location and status of the eggs being laid by the cuckoos.

The aim is to use the new and potentially better solutions (cuckoos) to replace a not-so-good solution in the nests. The position of the egg replaces the position of the random new eggs in another nest in three cases: if the eggs have been evicted, if the cuckoos abandon the nest or if the eggs have been hatched resulting to its survival.

Hatching of eggs – Survival of the Cuckoo
Abandoning the nest – Host bird abandons its nest and migrates to some other place to build another nest.
Evicting the eggs – The host bird throws the cuckoo bird's eggs.

```
Pseudo-code of Cuckoo Search (CS) is given below.
Begin
define objective function.
generate initial population of host nests.
while (criteria not met)
{
get a cuckoo randomly;
evaluate the fitness of it;
choose a nest from the population randomly;
if(fitness of selected nest is high)
end
Abandon a fraction of worse nests and build new
ones at new locations;
keep the best nests (solutions);
rank the nests and find the current best;
}
post process results and visualization;
End
```

---

6. HARMONY SEARCH ALGORITHM

The underlying principle behind Harmony Search Algorithm (HSA) has been that this algorithm might face the search on the grounds of Pitch, Amplitude and Timbre producing a perfect harmony. With successful results, this concept was applied to Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) in among other attempts. The initial random solutions that are considered may be far away from the feasible solutions. To get closer to the feasible and promising solutions, the exact solutions can be obtained by choosing only the amplitude of the tone. Musical performances seek to find to be a pleasing harmony determined by the aesthetic standard, in our proposed ground the aesthetic standard is based on amplitude. In this method, the best accuracy of Tunes is found with the frequency value of Tunes which varies with time(t) and computation is done for the new solution.In HSA, each musician (decision variable) plays a note (value) for finding a best harmony (global optimum) all together. Based on the Lambda value two constraints will be considered: Noise (high pitch) and Melody (low pitch). Equation (6.3) has been used as the main formula for the computation of Amplitude in Harmony Search Algorithm.

$$\lambda = \frac{C}{h_i[next] * h_i[music]} \text{ ----------------------- (6.3)}$$

Where,

Table 6.1 Parameter Settings for Harmony Search Algorithm

λ	Lambda Value (Amplitude)
C	Velocity of Light

h_i[next]	Accuracy of newly selected tune
h_i[music]	Accuracy of the existing tune that needs to be compared with the new tune

7. RESULTS AND DISCUSSION

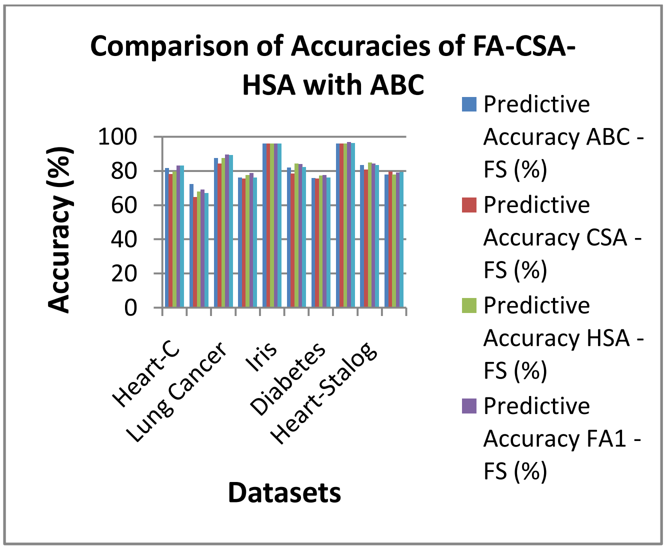
The performance of the FA-FS, CSA-FS, HS-FS proposed in this paper, has been tested with 10 different medical datasets and their details are given in Table I. The datasets are taken from UCI (University of California, Irvine) machine learning repository [21], a popular repository containing classification and regression datasets. The UCI datasets that have been used to evaluate FA-FS, CSA-FS, HS-FS are the same 10 datasets that EABC-FS had been experimented with so that, the performance of FA-FS, CSA-FS, HS-FS can be compared to that of EABC-FS.

TABLE I DATASET DESCRIPTION

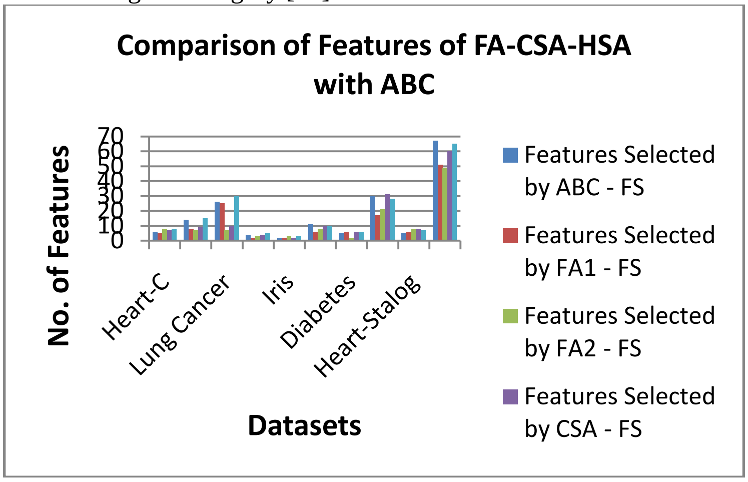
Datasets	Instances	Features	Class
Heart C	303	13	2
Dermatology	366	34	6
Hepatitis	155	19	2
Lung Cancer	32	56	2
Pima	768	8	2
Iris	150	4	3
Wisconsin	699	10	2
Lymph	148	18	4
Diabetes	768	8	2
Heart Stalog	270	13	2

In FA-FS, CSA-FS, HS-FS, FS plays the role of generation procedure and generate the feature subsets along with a classifier. The Decision Tree is used to evaluate each of the generated feature subsets. Decision Tree is implemented using J48 algorithm in WEKA (Waikato Environment of Knowledge Analysis) tool [23]. The generation procedure of FA-FS, CSA-FS, HS-FS makes use of the parameter settings given in Table II and is implemented using Net Beans IDE.

GRAPH I PARAMETER COMPARISON FOR 10 DATASET'S ACCURACIES FOR ALL THE ALGORITHMS



The results obtained from FA-FS, CSA-FS, HS-FS are Shown in Graph II. In Graph II, for each dataset, the actual numbers of features present in the dataset and the size of the feature subset obtained through FA-FS are presented for comparison. The Predictive Accuracy (PA) obtained for each dataset with reduced features have also been presented in Table III. PA is the percentage of instances that have been correctly classified as instances of their original category [22].



GRAPH II FEATURE SELECTION AND CLASSIFICATON ACCURACY BY FA-FS

Finally analyzing the three algorithms, we find that Firefly Algorithm gives better accuracy and Cuckoo Search and Harmony Search Algorithm gives accuracies equal to and greater than ABC-FS in earlier iterations. Also the integration has been carried out since the individual accuracies have found to be giving promising results and the integration of Firefly Algorithm with Harmony Search Algorithm performs better than the other integration due to their individual FS process.

8. CONCLUSION

With the emerging trends of Information Technology, Data Mining throws out its purpose to each individual spotting out only the relevant features by the Feature Selection process along with Classification from eons of database to reach the final outcome. This specifically defines our purpose of study for Feature Selection to perform the Classification. However, Bee Colony Algorithm and Ant Colony algorithm, etc has a drawback of not memorizing the historical instances and global optimum solution. This study has been attempted on other Nature Inspired Algorithms such as Firefly Algorithm, Cuckoo Search Algorithm and Harmony Search Algorithm which has been carried out previously only for mathematical optimization to the Feature Selection process with the objective of producing better results.

Firefly Algorithm is proposed in such a way the two sub algorithms works in a manner that the Feature Selection process performs Classification and all the previous instances of the Brightest Firefly are kept in record with its particular Beta value till the final iteration offering a diversified performance in the search space producing global optimum solution.

By performing Cuckoo Search Algorithm, the Feature Selection process is done based on the Egg Laying Behavior of the Cuckoo Birds' with the existence of the host birds' eggs by defining three sub algorithms of eviction, abandoning and survival of the cuckoo birds' eggs. This algorithm promotes the local optimum searches towards the global optimum in a short period of time because of the Alpha value of choosing the eggs is based on Classification and hence producing better accuracy with less number of features.

Furthermore, Harmony Search Algorithm is attempted where the Feature Selection process is done based on Rhythm by selecting the new tones from the existing tones in the dataset. Though considering the Harmony Memory considering Rate (HMCR), it has been extended on taking into account the wavelength of the new tone with their pitch and timbre based on their Lambda value producing the global solution with the previous instances in order to produce a perfect harmony.

After being performed all the above three algorithms, we further moved for the integration of each algorithm since the individual accuracies have found to be giving promising results and at last integrated all the three algorithms and the results have been recorded.

Thus from the study carried out Firefly Algorithm, Cuckoo Search Algorithm, Harmony Search Algorithm and all the Integration have been giving better results than the ABC-FS which would be visualized by their Predictive Accuracy and F-measure. Finally, comparing all the individual algorithm results, with that of our prework, we have inferred that Firefly Algorithm gives better accuracy and Cuckoo Search Algorithm and Harmony Search Algorithm gives accuracies greater than and equal to that of ABC-FS but results are obtained in earlier iterations thus reducing the execution time. Also integration has been carried out and the integration of Firefly Algorithm with Harmony Search Algorithm performs better than the other integration due to their individual FS process.

REFERENCES

[1]. R.O. Duda, P.E. Hart and D.G. Stork, Pattern Recognition, John Wiley & Sons, Inc 2nd edition, 2001.
[2]. Dash.M and Liu.H., "Feature Selection for Classification", Intelligent Data Analysis, Vol.39, No.1, pp.131–156, 1997.
[3]. L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp.123-140, 1996.
[4]. Y. Freund, R.E. Schapire, "Experiments with a new boosting algorithm", Proceeding of the Thirteenth International conference on Machine Learning, 148-156, 1996.
[5]. Ahmed Al-Ani: Ant Colony Optimization for Feature Subset Selection. WEC (2) 2005: 35-38.
[6]. L.I.Kuncheva, Combining Pattern Classifiers, Methods and Algorithms, Wiley Interscience, 2005.
[7]. Nadia Abd-Alsabour and Marcus Randall, Feature Selection for Classification Using an Ant Colony System, Proc. Sixth IEEE International Conference on e–Science Workshops, 2010, pp 86- 91.
[8]. Kabir M.M., Shahjahan, M. and Murase, K. (2008) 'A new hybrid ant colony optimization algorithm for feature selection', Journal of Applied Soft Computing, Vol. 8, pp. 687–697.
[9]. Tu, C.J., Chuang, L., Chang, J. and Yang, C. (2007) 'Feature Selection using PSO-SVM', IAENG International Journal of Computer Science, Vol. 33, No. 1, pp. 18 – 23.
[10]. N.Suguna and K.G.Thanushkodi, "An Independent Rough Set Approach Hybrid with Artificial Bee Colony Algorithm for Dimensionality Reduction", American Journal of Applied Sciences 8 (3): 261 – 266, 2011.
[11]. N.Suguna and K.G.Thanushkodi, "A novel Rough Set Reduct Algorithm for Medical Domain based on Bee Colony Optimization", Journal of Computing, Vol. 2(6), 2010, 49 -54
[12]. Shunmugapriya .P and Kanmani S, "Artificial Bee Colony Approach for Optimizing Feature Selection", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012, pp.432-438.
[13]. D.Karaboga. An idea based on honey bee swarm for numerical optimization. Technical Report TR06, Engineering Faculty, Computer Engineering Department, Erciyes University, 2005.
[14]. A.R. Yildiz, A new hybrid bee colony optimization approach for global optimization in design and manufacturing, Appl. Soft Comput. J. (2012),
[15]. Weifeng Gao, Sanyang Liu, Lingling Huang, A global best artificial bee colony algorithm for global optimization, Journal of Computational and Applied Mathematics 236 (2012) 2741–2753.
[16]. Milan Tuba, Nebojsa Bacanin and Nadezda Stanarevic, "Guided artificial bee colony algorithm", ACM Proc. European Computing Conference,2011,Pg.398-403.
[17]. ShiXin Yu, Feature Selection and Classifier Ensembles: A Study on Hyperspectral Remote Sensing Data, Ph.D. Thesis, The University of Antwerp, 2003.
[18]. D. Karaboga, Bahriye Akay, A comparative study of Artificial Bee Colony algorithm, Applied Mathematics and Computation 214 (2009) 108–132
[19]. RahaImanirad, Xin-She Yang and Julian Scott Yeomans, "Modelling To Generate Alternatives Via the Firefly Algorithm", Journal Of Applied Operational Research, 2013.
[20]. Xin-She Yang, "Nature Inspired Meta-Heuristic Algorithm", 2nd Edition, Luniver Press, 2010.
[21]. A.Frank, A. Asuncion, UCI Machine Learning Repository, (http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science (2010))
[22]. Han, J., and Kamber, M.: Data mining concepts and techniques, Academic Press, 2001.
[23]. WEKA: A Java Machine Learning Package, http://www.cs.waikato.ac.nz/˜m l/weka/.
[24]. Dash.M and Liu.H., "Feature Selection for Classification", Intelligent Data Analysis, Vol.39, No. 1, pp. 131 – 156, 1997.
[25]. Xin-She Yang, Suash Deb, "Cuckoo Search Via Levy Flights", World Congress On Nature and Biologically Inspired Computing (NaBIC 2009)
[26]. Xin-She Yang, "Harmony Search As A Metaheuristic Algorithm", Music-Inspired Harmony Search Algorithm: Theory and Applications (Editor Z. W. Geem), Studies in Computational Intelligence, Springer Berlin, vol. 191, pp. 1-14 (2009).
[27]. Shunmugapriya .P and Kanmani .S, "Feature Selection Optimization through Enhanced Artificial Bee Colony Algorithm".