Artificial Bee Colony Approach for Optimizing Feature Selection

Shunmugapriya Palanisamy[1] and Kanmani S[2]

**Research Scholar, Department of Computer Science and Engineering,
Pondicherry Engineering College, Puducherry, India. 605014**

**Professor, Department of Information Technology,
Pondicherry Engineering College, Puducherry, India. 605014**

Abstract

High dimensionality of the feature space affects the classification accuracies and the computational complexity due to redundant, irrelevant and noisy features present in the dataset. Feature Selection extracts the relevant and most useful information and helps to speed up the task of classification. Feature selection is seen as an optimization problem because selecting the appropriate, optimal feature subset is very important. The Artificial Bee Colony algorithm is a famous meta-heuristic search algorithm used in solving combinatorial optimization problems. This paper proposes a new method of feature selection, which uses the ABC algorithm to optimize the selection of features. Ten UCI datasets have been used for evaluating the proposed algorithm. Experimental results show that, ABC-Feature Selection has resulted in optimal feature subset configuration and increased classification accuracies up to 12% compared to the classifier and standard ensembles.

Keywords: *Classification, Feature Selection, Ant Colony Optimization, Bee Colony Optimization, Artificial Bee Colony Algorithm.*

1. Introduction

Feature Selection is an important preprocessing step for most machine learning algorithms especially pattern classification [1 and 7]. Feature Selection aims in determining the most relevant and useful subset of features from the dataset representing any application domain, without compromising the predictive accuracy represented by the actual set of features [16, 17 and 21].

Recently, evolutionary and swarm intelligent algorithms are employed to optimize feature selection because obtaining the optimal feature combination is very important. With optimal feature subset, it is possible to gain good prediction accuracies with low computational complexity. Artificial bee colony algorithm is a stochastic search algorithm inspired by the minimal modal of foraging behavior in bees [2, 3 and 4]. ABC is simple in concept, easy to implement and has only fewer control parameters [14]. It serves as a powerful optimization tool and has been successfully used in solving many combinatorial optimization problems [5 and 13].

ACO and ABC algorithms have successfully solved the problem of feature selection [11, 15, 23 and 24]. Though feature selection has been optimized in numerous works, no method has been proved as a consistent performer in optimization of feature selection. In this regard, we have proposed a new technique for optimizing feature selection using ABC.

In literature, ABC has been successfully used for feature selection optimization [23 and 24]. Before applying ABC for feature selection, the core reducts are obtained with Rough set theory. The feature subset important for identification of each class is listed separately and the features in common for the decision attributes are chosen as the core reduct. Discarding the core reduct from the feature set, ABC is applied to rest of the features for selection. The resultant feature subset will be containing the core reduct plus ABC selected features [23 and 24].

This study proposes a novel feature selection method in which ABC is used to generate the feature subsets and a classifier is used to evaluate the feature subsets generated by ABC. In this method, each employed bee is allocated a feature (food source) and the onlooker tries to make all possible combinations with other features to configure the feature subset. The proposed algorithm has shown competitive performance compared to ACO based feature selection.

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

2. Feature Selection and Classification

2.1 Feature Selection

Feature selection is an important pre-requisite for classification [7]. Feature selection is the process of extracting the relevant and useful features from the dataset by removing the redundant, irrelevant and noisy features. The process of feature selection requires two important components: Evaluation function to evaluate the candidate feature subset; Generation procedure 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 [21].

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, 7 and 18].

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. Artificial Bee Colony Algorithm (ABC)

The Artificial Bee Colony (ABC) algorithm is a relatively new technique proposed by Karaboga [2]. ABC is inspired by the foraging behavior of honey bee swarms. In the ABC algorithm, the colony consists of three kinds of bees: employed bees, onlooker bees and scouts. The population of the colony is double the size of food sources. The number of food sources represents the position of possible solutions of optimization problem and the nectar amount of a food source represents the quality (fitness) of the associated solution [5]. Employed bees are responsible for exploiting the food sources and they pass the information to the onlookers about the nectar quality of the food sources they are exploiting. The number of onlookers is equal to the number of employed bees. The onlookers decide a food source to exploit based on the information from the employed bees. The employers of exhausted food sources become scouts and they keep randomly searching for new food sources. Increased amount of nectar increases the probability of selection of a particular food source by the onlooker bees [2, 3 and 4]. The ABC algorithm is given in Fig.1 [2].

1. Initialize the food source positions
2. Evaluate the food sources
3. Produce new food sources(solutions) for the employed bees
4. Apply greedy selection
5. Calculate the fitness and probability values
6. Produce new food sources for onlookers
7. Apply greedy selection
8. Determine the food source to be abandoned and allocate its employed bee as a
   Scout for searching the new food sources
9. Memorize the best food source found
10. Repeat steps 3-9 for a pre-determined number of iterations

Fig. 1 Steps of the ABC algorithm

4. The Proposed Method

In the proposed feature selection approach, ABC algorithm optimizes the process of feature selection and yields the best optimal feature subset which increases the predictive accuracy of the classifier. Fig.2 represents the block diagram of performing feature selection using ABC. ABC is used as a feature selector and generates the feature subsets and a classifier is used to evaluate each feature subset produced by the onlookers; hence the proposed system is a wrapper based system.

The steps of the proposed feature selection algorithm are given in Fig.3. Initially, the classifier Decision Tree J48 evaluates the discriminating ability of each individual feature in the dataset. The predictive accuracy (x_i) of each feature i is calculated by employing 10-fold cross validation [6] and [7]. Then, the objective (f_i) is calculated for each feature from its indiscernibility relation.

A binary bit string (of length equal to the number of features in the dataset [11]) is assigned to each employed bee to represent its feature selection. In the binary string, if n^{th} bit is a '1' it means the associated feature is selected and a '0' means the feature is not selected. Features are considered as food sources and hence the population of employed and onlooker bees are equal to the number of features (m) in the dataset.

Each employed bee is allocated a feature and it evaluates the fitness of the feature from the objective (f_i) by using equation (1).

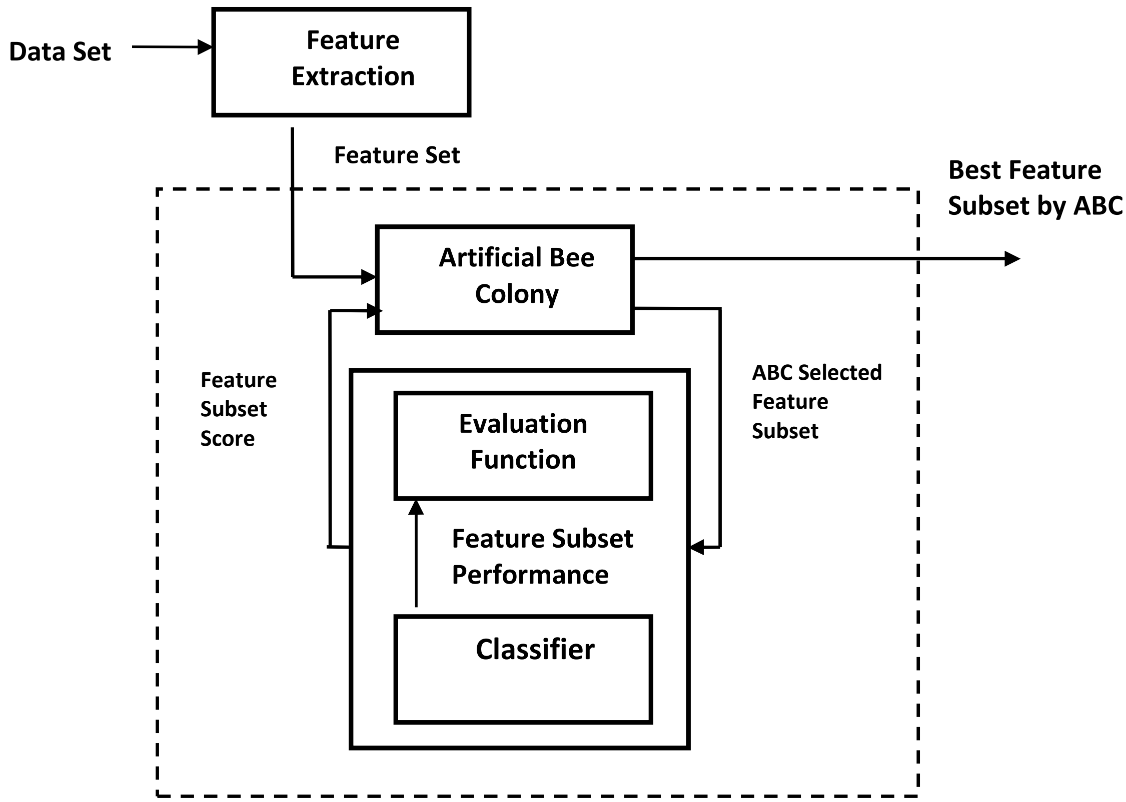

Fig. 2 Block Diagram of ABC Based Feature Selection

$$fit_i = 1/(1+f_i) \qquad (1)$$

The onlooker bee gains information from the employed bee and calculates the probability of selecting a feature using equation (2).

$$p_i = \frac{fit_i}{\sum_{i=1}^{m} fit_i} \qquad (2)$$

Then the onlooker computes the new solution v_i using the predictive accuracies of the feature the employed bee is pointing to and the feature the onlooker bee has selected. If the new solution v_i is greater than x_i , the employed bee will be pointing to feature subset consisting of the feature it was previously pointing and the newly selected feature. If v_i is not greater than x_i then, the employed bees feature will be retained and the

1.Cycle =1
2. Initialize ABC parameters
3. Evaluate the fitness of each individual feature
4. Repeat
5. Construct solutions by the employed bees
- Assign feature subset configurations (binary bit string) to each employed bee
- Produce new feature subsets v_i
- Pass the produced feature subset to the classifier
- Evaluate the fitness (fit_i) of the feature subset
- Calculate the probability p_i of feature subset solution

6. Construct solutions by the onlookers
- Select a feature based on the probability p_i
- Compute v_i using x_i and x_j
- Apply greedy selection between v_i and x_i

7. Determine the scout bee and the abandoned solution
8. Calculate the best feature subset of the cycle
9. Memorize the best optimal feature subset
10. Cycle = Cycle + 1
11. Until pre-determined number of cycles is reached
12. Employ the same searching procedure of bees to generate the optimal feature subset configurations

Fig. 3 The ABC-Feature Selection Algorithm newly selected feature is neglected. The new solution v_i is computed by using equation (3).

$$v_i = x_i + \varphi_i(x_i - x_j) \qquad (3)$$

where, x_i is the predictive accuracy of the feature allocated to the employed bee and x_j is the predictive accuracy of the feature the onlooker has selected. φ_i is a uniformly distributed real random number in the range [0, 1]. This way, each time the employed bee is assigned a new feature subset, the onlooker exploits and tries to produce new feature subset configuration.

After all possible features are exploited for forming the feature subset, the nectar content gets accumulated towards better feature subset configuration. If any employed bee has not improved, then the employed bee becomes a scout. The scout is assigned a new binary feature set based on the equation (4).

$$x_i^j = x_{\min}^j + rand[0,1](x_{\max}^j - x_{\min}^j) \qquad (4)$$

Where x_{max}^j and x_{min}^j represents the lower and upper bounds of the dimension of the population. The bees keep executing the same procedure for a number of runs until best feature subset is formed.

5. Experiments and Discussions

The datasets used, the implementation and the results of ABC-Feature Selection are discussed in this section.

5.1 Datasets

The proposed feature selection algorithm has been implemented and tested using 10 datasets from different fields of medicine. The datasets are taken from University of California (UCI), Irvine data repository and Table 1 briefs the description of the datasets [8]. The UCI datasets have been widely used in classification, feature selection and classifier ensemble works for experimental proofs and hence we have adapted them for our work. The datasets are also selected such that they have varied number of classes and features (attributes): the effect of feature selection is easily visible and the performance of the feature subset selected can also be visualized for varying class range.

5.2 Implementation of ABC-Feature Selection

The classifier employed for evaluating the feature subsets generated is Decision Tree. Decision tree is implemented using J48 algorithm in WEKA from Waikato University [9]. Feature subset generation and selection by ABC has been implemented using Net Beans IDE.

Table 1: Datasets Description

Dataset	Instances	Features	Classes
Heart-C	303	14	2
Dermatology	366	34	6
Hepatitis	155	19	2
Lung Cancer	32	56	2
Pima Indian Diabetes	768	8	2
Iris	150	4	3
Wisconsin	699	9	2
Lymphography	148	18	4
Diabetes	768	9	2
Heart-Stalog	270	13	2

The parameters initializations of ABC- Feature selection are given below:

Population Size p : 2 * No. of features in the data set
Dimension of the population : p × N
Lower Bound : 1
Upper Bound : N
No. of runs : 10
Maximum Number of iterations : Equal to the number of features
φ : 0.3

With these parameter settings, the artificial bees search and the feature subset generated through the cycles with highest predictive accuracy is recorded as the best optimal feature subset.

ABC takes up the role of generation procedure for generating candidate feature subset and J48 plays the role of evaluation function to evaluate the candidate feature subsets. Accuracy of the classifier is used as the fitness function to evaluate the feature subsets. Accuracy of the classifier is defined as how correctly a given classifier identifies and labels instances on which it is not trained [22]. So, feature subset yielding highest accuracy of classification over the iterations is recorded as the best optimal feature subset. Classification accuracy is computed by using 10-fold cross validation. In *k*-fold cross validation, the dataset is divided into *k* equally sized folds and the learning algorithm is executed for *k* times [6], [7] and [22].

When the algorithm is implemented, the onlooker bees make use of equations (1) and (2) to select a feature for inclusion into the feature subset. The decision about including the feature into the subset is based on the value of v_i computed using equation (3). Each time a feature is added to the binary string, the employed bee computes the accuracy of the new feature subset and this becomes the new x_i value. This way features are added to the binary string representing feature selection and the bees explore all possible combinations of features during the iterations. The feature subset and the classification accuracies obtained by the proposed method are shown in Table 2. ABC-Feature Selection has resulted in reduced feature size compared to the actual features and good classification accuracies.

Table 2: ABC based Feature Selection- Classification Accuracy and Feature Subset

Dataset	**No. of Features**	**Features Selected by ABC-FS**	**Predictive Accuracy (ABC-FS)(%)**
Heart-C	14	6-7	86.92
Dermatology	34	24	98.55
Hepatitis	19	11	81.26
Lung Cancer	56	27	89.25
Pima	8	6	90.08
Iris	4	2	96.00
Wisconsin	9	4	96.99
Lymph	18	9	96.69
Diabetes	9	5	83.12
Heart-Stalog	13	6	84.07

In our previous work, we have implemented the optimization of feature selection using ACO technique [15]. In ACO, ants select a feature based on the move probability associated with each feature. The move probability depends on the heuristic information of pheromone value associated with each feature and the proportion of ants that have selected particular feature at that instant [11] and [15]. Table 3 and Table 4 show the comparative performance of ACO based feature selection and ABC based feature selection. In Table 3, we have listed the reduced feature size obtained through ACO [15] and ABC optimization techniques. It can be inferred from Table 3 that ABC has resulted in reduced feature subset equivalent to ACO in some cases and comparatively better optimization for other datasets.

In Table 4, the classification accuracies achieved through the proposed method is compared with the accuracies of ACO based FS [15], classifier J48 (used in this experiment) and standard ensembles bagging [7], [18] and [19] and boosting [7], [18] and [20].

Table 3: Comparison of Features Selected by ACO and ABC

Dataset	**No. of Features**	**ACO - FS**	**ABC-FS**
Heart-C	14	8	6-7
Dermatology	34	28	24
Hepatitis	19	13	11
Lung Cancer	56	33	27
Pima	8	6	6
Iris	4	3	2
Wisconsin	9	4	4
Lymphography	18	16	9
Diabetes	9	6	5
Heart-Stalog	13	8	6

When the ABC algorithm has been applied to the datasets, it has resulted in feature subset (consisting relevant and optimal features) and increased recognition rates for all datasets.

From the data represented in Table 2, Table 3 and Table 4, it can be inferred that:

i. Feature selection definitely increases the classification accuracy and speeds up the process of classification
ii. For all datasets except Hepatitis and Diabetes, ABC-FS has given the highest recognition rates
iii. For Hepatitis, Boosting has given the highest accuracy and even ACO has lost in case of hepatitis
iv. For diabetes, ACO has the leading performance and accuracy of ABC-FS is marginally low compared to ACO-FS
v. For Heart-c, Iris, Pima and Wisconsin, feature subset obtained is almost of same size as in ACO-FS
vi. For Lung Cancer, Lymph and Stalog, size of the feature set is minimized to a greater level with good prediction accuracies. This very well explains the effectiveness of the proposed method
vii. For Lung Cancer, Lymph and Stalog, size of the feature set is minimized to a greater level with good prediction accuracies. This very well explains the effectiveness of the proposed method
viii. Convergence of the search space is achieved quickly
ix. In the histogram, the bar tends to raise on the ABCE methods for most of the datasets, which shows the betterment of the proposed method.

Table 4: Comparison of Prediction Accuracies of 10 Datasets through 10-Fold Cross Validation

Dataset	**J48**	**Bagging (C4.5)**	**Boosting (C4.5)**	**ACO Based FS**	**ABC Based FS**
Heart-C	81.19	78.88	76.9	86.85	**86.92**
Dermatology	95.90	95.90	95.90	98.35	**98.55**
Hepatitis	63.22	83.23	**85.81**	77.45	81.26
Lung Cancer	71.87	78.12	75	87.37	**89.25**
Pima	72.00	74.09	72.4	89.82	**90.08**
Iris	94.67	95.33	93.33	93.35	**96.00**
Lymph	82.43	95.14	96.42	78.35	**96.99**
Wisconsin	95.57	74.09	72.4	87.65	**96.69**
Diabetes	73.95	79.05	83.11	**84.11**	83.12
Heart Stalog	81.85	80	80.37	82.12	84.07

(All numbers are in percentage)

6. Conclusion

ABC is a powerful optimization technique and has been widely used in solving combinatorial optimization problems. In this work, ABC has been employed to select optimal feature subset. The onlooker bee evaluates all possible combinations of features before yielding the optimal feature subset. Experimental results show the superiority of ABC- FS. The proposed algorithm has resulted in reduced feature size of the feature subset, increased classification accuracies, low computational complexity

References

[1] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Recognition*, John Wiley & Sons, Inc 2nd edition, 2001.

[2] Karaboga and B. Basturk, “A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm”, Journal of Global Optimization, Springer Netherlands, vol.39, no.3, 2007, pp.459-471

[3] Karaboga, “An Idea Based on Honey Bee Swarm for Numerical Optimization”, Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.

[4] D. Karaboga, B. Basturk, “Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems”, LNCS: Advances in Soft Computing: Foundations of Fuzzy Logic and Soft Computing, Springer – Verlag, IFSA 2007, vol. 4529/2007, pp.789-798.

[5] Fei Kang, Junjie Li, Haojin Li, Zhenyue Ma and Qing Xu, An Improved Artificial Bee Colony Algorithm, IEEE 2010.

[6] Han, J., and Kamber, M.: Data mining concepts and techniques, Academic Press, 2001.

[7] L.I.Kuncheva, Combining Pattern Classifiers, Methods and Algorithms, Wiley Interscience, 2005.

[8] 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))

[9] WEKA: A Java Machine Learning Package, http://www.cs.waikato.ac.nz/˜ml/weka/.

[10] M. Dorigo and T. Stutzle, Ant Colony Optimization, The MIT Press, Massachusetts, USA, 2004.

[11] 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.

[12] Wenping Zou, Yunlong Zhu, Hanning Chen and Zhu Zhu, Cooperative Approaches to Artificial Bee Colony Algorithm, Proc. IEEE International Conference on Computer Application and System Modeling, Vol. 9, 2010, 44-48.

[13] Mohammed El-Abd, A Cooperative Approach to the Artificial Bee Colony Algorithm, *IEEE*, 2010.

[14] Li Bao and Jian-chao Zeng, Comparison and Analysis of the Selection Mechanism in the Artificial Bee Colony Algorithm, Proc. IEEE Ninth International Conference on Hybrid Intelligent Systems, 2009, 411-416.

[15] Shunmugapriya P, Kanmani S, Devipriya S, Archana J and Pushpa J “Investigation on the Effects of ACO Parameters for Feature Selection and Classification”, Proc. Springer - The Third International Conference on Advances in Communications, Networks and Computing, LNICST pp. 136–145, 2012.

[16] Ahmed, E.F., W.J. Yang and M.Y. Abdullah, “Novel method of the combination of forecasts based on rough sets”, J. Comput. Sci., 5: 440-444.2009.

[17] ShiXin Yu, Feature Selection and Classifier Ensembles: A Study on Hyperspectral Remote Sensing Data, Ph.D. Thesis, The University of Antwerp, 2003.

[18] Polikar R., “Ensemble based Systems in decision making” IEEE Circuits and Systems Mag., vol. 6, no. 3, pp. 21-45, 2006.

[19] L. Breiman, “Bagging predictors,” Machine Learning, vol. 24, no. 2, pp.123-140, 1996.

[20] Y. Freund, R.E. Schapire, “Experiments with a new boosting algorithm”, Proceeding of the Thirteenth

International conference on Machine Learning, 148-156, 1996.

[21] Dash.M and Liu.H., "Feature Selection for Classification", Intelligent Data Analysis, Vol.39, No. 1, pp. 131 – 156, 1997.

[22] Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques. San Francisco: Morgan Kaufmann.

[23] 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.

[24] 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.

P.Shunmugapriya received her M.E. degree in 2006 from Department of Computer Science and Engineering, FEAT, Annamalai University, Chidambaram. She had been working as a Senior Lecturer for the past 7 years in the Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Affiliated to Pondicherry University, Puducherry. Currently she is working towards her Ph.D degree in Optimal Design of Classifier Ensembles using Swarm Intelligent, Meta-Heuristic Search Algorithms. Her areas of interest are Artificial Intelligence, Ontology based Software Engineering, Classifier Ensembles and Swarm Intelligence.

Dr. S. Kanmani received her B.E and M.E in Computer Science and Engineering from Bharathiyar University and Ph.D in Anna University, Chennai. She had been the faculty of Department of Computer Science and Engineering, Pondicherry Engineering College from 1992 onwards. Presently she is a Professor in the Department of Information Technology, Pondicherry Engineering College, Puducherry. Her research interests are Software Engineering, Software testing, Object oriented system, and Data Mining. She is Member of Computer Society of India, ISTE and Institute of Engineers, India. She has published about 50 papers in various international conferences and journals.