Proceedings of the International Conference on Machine Vision and Machine Learning
Prague, Czech Republic, August 14-15, 2014
Paper No. 131

Performance Comparison of ABC and A-ABC Algorithms on Clustering Problems

Ahmet Ozkis, Ahmet Babalik
Selcuk Universit, Department of Computer Engineering
Selçuklu, Konya, Turkey
ahmetozkis@selcuk.edu.tr; ababalik@selcuk.edu.tr

Abstract **-** Clustering analysis, used in many science area and applications, is a considerable tool and a descriptive task trying to identify similar groups of objects based on the values of their attributes. To solve clustering problems there are different ways including machine learning techniques, statistics and metaheuristic algorithms. Artificial Bee Colony (ABC) algorithm is one of the most popular swarm intelligence based algorithms which simulates the clever foraging behaviour of honey bees. In this study, we used ABC algorithm and Accelerated ABC (A-ABC) algorithm, which is a modified version of the ABC algorithm, for data clustering. Four datasets (Vertebral-2 Coloumn, Iris, Wine, Dermatology) taken from University of California Irvine (UCI) Machine Learning Repository are used to compare the results of ABC and A-ABC algorithms. Obtained results show that our proposed A-ABC algorithm has generally better performance than standard ABC algorithm on the used datasets.

Keywords**:** Artificial bee colony, Clustering, Metaheuristic algorithm, Swarm intelligence.

1. Introduction

Clustering means the act of dividing a whole dataset into groups according to their similarities. Each instance called as an object and each group of data called as a cluster (Abraham et al., 2008). The target of clustering is to group data into clusters such that the similarities among data members in same cluster are maximal, while similarities among data members from other clusters are minimal (Karaboga and Ozturk, 2011).

Clustering has an essential role in a variety of fields including engineering (mechanical, electrical, industrial engineering), computer sciences (machine learning, artificial intelligence, pattern recognition, web mining, spatial database analysis, textual document collection, image segmentation), life and medical sciences (genetics, biology, microbiology, paleontology, psychiatry, pathology), earth sciences (geography. geology, remote sensing), social sciences (sociology, psychology, archeology, education), and economics (marketing, business) (Abraham et al., 2008).

Clustering algorithms can be grouped into two main classes of algorithms called as supervised and unsupervised. In supervised clustering, learning algorithm has a guide which indicates the goal class to that a data vector must belong. In unsupervised clustering, there is no guide, and data vectors are grouped according to distance from each other (V.Merwe and Engelbrecht, 2003).

Variety of unsupervised clustering algorithms developed such as K-Means, ISODATA, learning vector quantizers (LVQ). Especially K-means is a well-known successfull clustering algorithms which is a center based, simple and fast algorithm. However, because of the K-means algorithm converges to the nearest local optimum from the initial position, success of the algorithm is highly depend on the initial state of the center of the cluster. Different techniques such as statistics, graph theory, expectation maximization algorithms, artificial neural networks, evolutionary computing and swarm intelligence algorithms are used to overcome local optima problem on clustering (Karaboga and Ozturk, 2011).

The Artificial Bee Colony algorithm which is a member of family of swarm intelligence algorithms is developed by Karaboga (2005) by simulating clever foraging behaviour of honey bees. Firstly, the ABC algorithm used on the numerical test functions (Karaboga, 2005; Karaboga and Basturk, 2007; Karaboga and Akay, 2009) and its promising performance on the test functions encourages the researchers to study on it. In literature, there are many modification studies to develop the performance of the ABC algorithm. Banharnsakun et al. (2011) proposed a modification by using the best so far food source for onlooker bees and greedy selection mechanism is applied by using objective value instead of fitness value. Akay and Karaboga (2012) added a new parameter called as modification rate (MR), which controls number of modified dimension of the test function. Qingxian and Haijun (2008) modified the ABC algorithm by making the initial group symmetrical in the launching scheme and by using Boltzman selection mechanism instead of roulette wheel to boost the convergence ability of the ABC algorithm. Baykasoglu et al. (2007) modified the ABC algorithm with shift neighborhood searches and Greedy Randomized Adaptive Search Heuristic (GRAH).

Except modification studies, there are lots of application studies on the ABC. The ABC algorithm is used for signal processing to design digital IIR filters (Karaboga N., 2009), training the artificial neural networks (Karaboga and Akay, 2007; Karaboga et al. 2007), synthetic aperture radar (SAR) image segmentation (M.Liang et all, 2011), designing of multiplier-less non-uniform filter bank transmultiplexer (Manoj and Alias, 2012), fuel management optimization (De Oliveira and Schirru, 2011), obtaining optimal control of automatic voltage regulator (AVR) systems (Gozde and Taplamacioglu, 2011) and capacitated vehicle routing problem (Szeto and Ho, 2011) and solving travelling salesman problem (TSP) with real coded version (Pathak and Tiwari, 2012). Additionaly, ABC algorithm is also employed as a data mining method for data clustering (Zhang et al., 2010; Karaboga and Ozturk, 2011).

Accelerated ABC (A-ABC) algorithm is also a modified version of the ABC algorithm proposed by Ozkis and Babalik (2013). In this study, ABC algorithm and A-ABC algorithm are used to cluster 4 different datasets (Vertebral -2 Coloumn, Iris, Wine, Dermatology) taken from UCI Machine Learning Repository database and compared their clustering performance with each other.

The paper is organized as main structure of the ABC and A-ABC algorithms and using strategy of both algorithms for clustering are explained in Section 2, experiments and obtained results presented in Section 3 and conclusion in Section 4.

2. Material and Methods

2.1. ABC Algorithm

The ABC algorithm simulates foraging behaviour of honey bee colony for optimization. This algorithm consist of three kind of bees: employeds, onlookers and scouts. Total number of employed and onlooker bees are equal to population and both type has same number. Main structure of the ABC algorithm is explained as follows (Karaboga, 2005; Karaboga and Akay, 2009):

1) *Generating initial food source position:* Scout bees determine randomly the first positions of food sources by using Eqs. (1).

$$X_{ij} = X^{min} + rand(0,1)(X^{max} - X^{min}) \qquad (1)$$

where i = 1. . .SN, j = 1. . .D. SN is the number of food sources and D is the number of optimization parameters. X is a SNxD dimensional matrix which describes the food sources generated by scout bees.

In this study, all attributes in datasets are normalized between 0 and 1. So, we set X^{min} as 0 and X^{max} as 1 in all dimensions.

2) *Exploitation mechanism of employed bees:* Scout bees turn to employeds and exploitation process begins. The exploitation mechanism implements local searching by using employed and onlooker bees near the food sources. Employeds and onlookers together try to develop the quality of the existing food sources by using Eqs. (2).

$$V_{ij} = X_{ij} + rand(-1,1)(X_{ij} - X_{kj}) \qquad (2)$$

Where X_{ij} describes the existing position of the food source and V_{ij} is a candidate source generated by modifiying one parameter randomly selected from X_{ij}. J is a randomly selected integer between [1,D] and k is another random integer between [1,SN], k and i must be different from each other.

3) *Selection probability of food sources by onlooker bees:* Onlookers determine the source which they will exploit, according to waggle dances of the employed bees. This dance points the quality of the food source. In the ABC, quality of the food sources is calculated by using Eqs. (3).

$$P_i = \frac{0.9 * Fitness_i}{Fitness_{best}} + 0.1 \tag{3}$$

Where $fitness_i$ indicates the nectar amount of the i_{th} food source, $fitness_{best}$ indicates nectar amount of the best food sources found until that cycle and p_i is chosen probability of i_{th} food source. As long as p_i value become bigger, chosen probability of the source also increases.

4) *Exceeding limit parameter and scouting:*In ABC algorithm, when all swarm complete their research for an iteration, all the sources are checked whether anyone is runs out or not. Counters record the number of failed attempts on modification process. If the number of a counter exceeds the limit parameter, the employed bee abandons own food source and chooses randomly new one as a scout by using Eqs. (4)

$$Limit = FoodNumber * D \tag{4}$$

Where FoodNumber is equal to number of employed bees and D is assign the dimension (for this study number of attributes) of problem.

2.2. A-ABC Algorithm

A-ABC algorithm which is a modified version of the ABC algorithm proposed by Ozkis and Babalik (2013). In A-ABC, two modifications are combined to develop the performance of the ABC algorithm. First modification called as MR which belong to Akay and Karaboga (2012) provides the opportunity to changing more than one parameter in a modification. Thus, MR contributes faster convergence than the standard ABC.

X_i	Dim_1	Dim_2	Dim_3	Dim_4	…….	Dim_m
$R_{ij}[0,1]$	0.337	0.647	0.108	0.814	…….	…….
MR	0.4	0.4	0.4	0.4	0.4	0.4
State ($MR>R_{ij}$)	√	⊗	√	⊗		

Fig. 1. MR modification which belong to Akay and Karaboga (2012).

In this study MR value is set as 0.4 such as adviced by Akay and Karaboga (2012). This process is shown with an example in Fig. 1.

Second modification related to greedy selection mechanism. Greedy selection mechanism is applied by using objective values instead of fitness value. This greedy selection modification which proposed by Banharnsakun et al. (2011) is shown in Fig. 2. Detailed information about the A-ABC algorithm is presented by Ozkis (2013) and Ozkis and Babalik (2013).

```
Greedy selection mechanism
If (Finding Minimum)
    If (f($x_i$))>f($v_i$))
        Replace the $v_i$ with $x_i$.
    Else
        Keep the $x_i$ solution.
Else If (Finding Maksimum)
    If (f($v_i$))>f($x_i$))
    Replace the $v_i$ with $x_i$.
    Else
Keep the $x_i$ solution.
```

Fig.2. Greedy selection mechanism proposed by Banharnsakun et al. (2011).

2.3. ABC and A-ABC for Clustering

One of the most popular clustering techniques is K-means algorithm which uses the Euclidian distance for clustering process (MacQueen, 1967). In K-means algorithm, the distances of each objects from the centers of clusters are calculated by using Euclidian distance. Each objects is labeled depend on the nearest center of cluster. The problem is described as follows by Karaboga and Ozturk (2011);

There is N objects, each object share to one of K clusters and the aim is minimizing the distance between each object and the center of the belonging cluster. This minimizing problem is shown in Eqs (5)

$$j(w,z) = \sum_{i=1}^{N} \sum_{j=1}^{K} w_{ij} \left(\left\| x_i - z_j \right\|\right)^2 \tag{5}$$

Where K assign the number of cluster, N the number of objects, x_i(i=1,…,N) position of the ith objects and z_j(j=1,…,K) is the center of the jth cluster calculated by Eqs.(6).

$$z_j = \frac{1}{N_j} \sum_{i=1}^{N} w_{ij}\, x_{ij} \tag{6}$$

Where N_j is the number of objects in the j_{th} cluster and w_{ij} is a binary value (0,1). If i_{th} object belongs to j_{th} cluster, w_{ij} is set as 1, otherwise 0.

In this study, it is aimed to determine optimum center of clusters vector by using ABC and A-ABC algorithms. Initially, candidate vectors are generated randomly by scout bees. Each vector includes candidate solution for each center of cluster. So, each vector consists of **number of class** *x* ***number of attributes*** dimensions.

Each vector indicates a food source which belong to an employed bee. Thus, total number of vectors are equal to number of employed bees. Classification error percentage (CEP) which means percentage of misclassified samples in testing process is used as fitness function by Karaboga and Ozturk (2011). In our study, we also used same way as fitness function and CEP formula is given in Eqs. (7).

$$CEP = \frac{misclassified\ examples}{size\ of\ test\ dataset} x100 \tag{7}$$

3. Experimental Study

In this work, 4 different datasets (Vertebral -2 Coloumn, Iris, Wine, Dermatology) taken from the UCI are used to compare the performance of ABC and A-ABC algorithms. From the datasets 75% of randomly selected data are used in training process and the remaning 25% part of the data are used in testing process.

3. 1. Datasets Description

Used datasets taken from UCI and datasets descriptions are given below (referans ver):

Vertebral Column dataset is related to vertebral column disorders of people. The dataset includes 310 objects having 6 real valued attributes in 2 classes or 3 classes. In this study we used the dataset in 2 classes named as normal and abnormal.

Iris dataset consist of 150 objects of flowers from the Iris species. There are 3 classes named as Setosa, Versicolor, Virginica and each class has 50 objects. Each objects have 4 attributes named as sepal length, sepal width, petal length, and petal width.

Wine dataset has totally 178 objects and 13 inputs for each object. There are 3 types of wine in the set.

Dermatology dataset has the biggest number of classes among datasets used in this study. There are 6 classes named as psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris and 366 objects with 34 attributes. Experiments are applied with 358 objects by removing 8 samples which has some missing values.

Table 1. UCI datasets used in performance analysis.

Datasets Name	**Data**	**Train**	**Test**	**Attribute**	**Class**
Vertebral Column – 2 Column	310	232	78	6	2
Iris	150	112	38	4	3
Wine	178	133	45	13	3
Dermatology	358	268	90	34	6

3. 2. Settings of Parameters

One of the most important features being ABC algorithm very simple is that having just 3 parameters: Number of swarm (NS), limit and maximum cycle number (MCN). In our study, for booth algorithm NS is selected as 20, limit value is determined by using Eqs. (4) and MCN is set as 500 iterations. So, total evaluation number is 10,000. Each experiments are repeated 30 times.

3. 3. Results and Discussion

From the datasets 75% of randomly selected data are used in training process and the remaning 25% part of the data are used in testing process. For each dataset best, worst, mean results of all runs are shown in Table2 and Table3 respectively. Here, *best* indicates the highest classification accuracy, w*orst* indicates the lowest classification accuracy, *mean* indicates the average value of classification accuracy for a dataset in 30 runs. Table4 also shows the comparision of ABC and A-ABC algorithms according to mean results of classification accuracy.

Table 2. Classification Accuracy of ABC algorithm (%).

Datasets Name	**Best**	**Worst**	**Mean**
Vertebral Column – 2 Column	78,20	67,94	74,31
Iris	100	94,73	96,40
Wine	95,55	82,22	91,33
Dermatology	88,88	51,11	72,77

Table 3. Classification Accuracy of A-ABC algorithm (%).

Datasets Name	Best	Worst	Mean
Vertebral Column – 2 Column	80,77	71,79	75,13
Iris	100	92,1	95,87
Wine	100	88,89	94,44
Dermatology	94,44	75,56	83,56

Table 4. Comparision of ABC and A-ABC Accuracy of classification on test datasets (%).

Datasets Name	ABC	A-ABC
Vertebral Column – 2 Column	74,31	**75,13**
Iris	**96,40**	95,87
Wine	91,33	**94,44**
Dermatology	72,77	**83,56**

As shown in Table2-Table4, A-ABC algorithm has better classification performance than the ABC algorithm on the Vertebral Column – 2 Column , Wine and Dermatology datasets. A-ABC algorithm has worst performance only on Iris data sets. Especially on Dermatology datasets A-ABC is quite better than ABC algorithm. Performance comparision of both algorithms are shown on Fig. 3 as graph.

4. Conclusion

In this study, we used ABC algorithm and Accelerated ABC (A-ABC) algorithm, which is a modified version of the ABC algorithm, for data clustering. Test datasets are taken from the UCI database to demonstrate the results of ABC and A-ABC algorithms. Obtained results show that our proposed A-ABC algorithm has better performance on three quarters comparision. ABC algorithm has better performance just on one (iris) comparision. This results show that our proposed algorithm is a successful method for these datasets and this method can be tested on other clustering problems and real world applications.

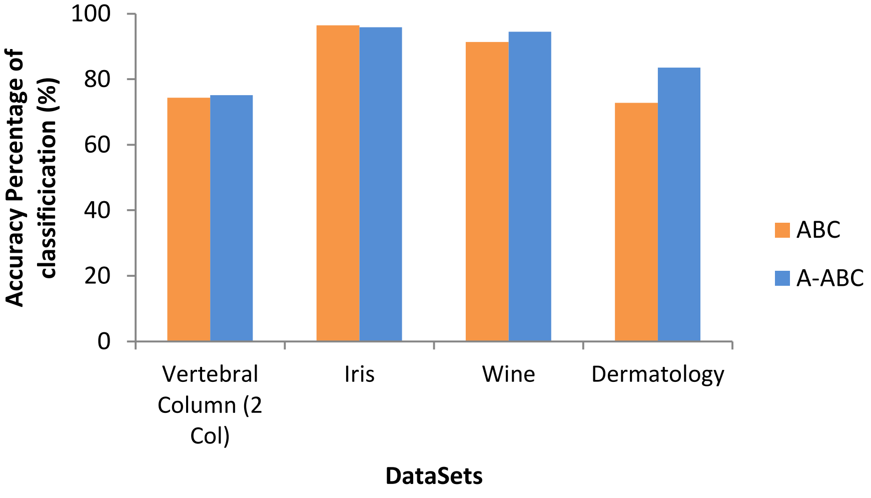

Fig. 3. Performance comparision of ABC and A-ABC algorithms.

References

Abraham, A., Das, S., Roy, S. (2008). Swarm Intelligence Algorithms for Data Clustering, Soft Computing for Knowledge Discovery and Data Mining, 279-313.

Akay B., Karaboga D. (2012). A modified Artificial Bee Colony algorithm for real-parameter optimization. Information Sciences 192,120–142.

Banharnsakun A., Achalakul T., Sirinaovakul B. (2011). The best-so-far selection in Artificial Bee Colony algorithm. Applied Soft Computing, 11(2), 2888–2901.

Baykasoglu A., Ozbakir L., Tapkan P. (2007). Swarm intelligence focus on ant and particle swarm optimization, Artificial Bee Colony Algorithm and Its Application to Generalized Assignment Problem, I-Tech Education and Publishing, Vienna, Austria, pp. 113–144.

De Oliveira I.M.S., Schirru R. (2011). Swarm intelligence of artificial bees applied to in-core fuel management optimization, Annals of Nuclear Energy 38 (2011) 1039–1045.

Gozde H., Taplamacioglu M.C. (2011). Comparative performance analysis of artificial bee colony algorithm for automatic voltage regulator (AVR) system, Journal of the Franklin Institute 348 (2011) 1927–1946.

JA Hartigan, (1975). ”Clustering Algorithms”, John Wiley EL Sons, New York.

Karaboga D. (2005). An idea based on honey bee swarm for numerical optimization, Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department.

Karaboga D., Akay B. (2007). An Artificial Bee Colony (ABC) Algorithm on Training Artificial neural networks.in: 15th IEEE Signal Processing and Communications Applications. SIU 2007 (Eskisehir.Turkiye).

Karaboga D., Akay B., Ozturk C. (2007). Modeling decisions for artificial intelligence. Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks. LNCS 4617, Springer-Verlag. pp 318–329.

Karaboga D., Akay, B. (2009). A comparative study of artificial bee colony algorithm. Applied Mathematics and Computation, 214(1), 108–132.

Karaboga D., Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. Journal of Global Optimization, 39(3), 459–471.

Karaboga D., Ozturk, C. (2011). A Novel Clustering Approach: Artificial Bee Colony (ABC) Algorithm, Applied Soft Computing, 11(1), 652–657.

Karaboga N. (2009). A new design method on ABC algorithm for digital IIR filters. Journal of the Franklin Institute, 346(4), 328–348.

Ma, M., Liang, J., Guo, M., Fan, Y., & Yin, Y. (2011). SAR image segmentation based on Artificial Bee Colony algorithm. Applied Soft Computing, 11(8), 5205–5214.

MacQueen J. (1967), Some methods for classification and analysis of multivariate observations, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297.

Manoj V.J., Elias E. (2012). Artificial bee colony algorithm for the design of multiplier- less nonuniform filter bank transmultiplexer, Information Sciences 192, 193–203.

Ozkis A. (2013). Sayısal Optimizasyon Problemlerinin Çözümü için Yapay Arı Kolonisi Algoritmasının İyileştirilmesi, MSc. Thesis, Selcuk University, Konya, Turkey.

Ozkis A., Babalik A. (2013). Accelerated ABC (A-ABC) Algorithm for Continuous Optimization Problems. Lecture Notes on Software Engineering vol. 1, no. 3, pp. 262-266.

Pathak, N., Tiwari, S. P. (2012). Travelling Salesman Problem Using Bee Colony With SPV, (3), 410–414.

Qingxian F., Haijun D. (2008). Bee colony algorithm for the function optimization, Science Paper Online. August 2008.

Szeto W.Y., Wu Y., Ho S.C. (2011). An artificial bee colony algorithm for the capacitated vehicle routing problem, European Journal of Operational Research 215, 126–135.

Van der Merwe, D. W., & Engelbrecht, A. P. (2003). Data clustering using particle swarm optimization. The 2003 Congress on Evolutionary Computation. CEC ’03., 215–220.

Zhang C., Ouyang D., Ning J. (2010). An artificial bee colony approach for clustering, Expert Systems with Applications 37, 4761–4767.

Web Sites:
Web-1: http://archive.ics.uci.edu/ml/datasets.html consulted 15 March 2014.