Data Clustering Using Bee Colony Optimization

Khadijeh Keshtkar mizooji
Department of Electrical & Computer Engineering,
Islamic Azad University, Qazvin Branch,
Qazvin, Iran
kh.keshtkar@gmail.com

Abolfazl Toroghi Haghighat
Department of Electrical & Computer Engineering,
Islamic Azad University, Qazvin Branch,
Qazvin, Iran
haghighat@qiau.ac.ir

Rana Forsati
Department of Electrical & Computer Engineering,
Islamic Azad University, Karaj Branch
Karaj, Iran
r_forsati@sbu.ac.ir

***Abstract*— The paper presents a comparative analysis of data clustering by Bee Colony Optimization (BCO) technique. Experiments over a standard benchmark demonstrate that applying Bee Colony Optimization in the context of clustering is a feasible approach and improves the clustering results. Superiority of the proposed algorithm is demonstrated by comparing it with some recently developed partitional clustering techniques.**

Keywords-Clustering; Swarm Intelligence; Function Optimization.

I. INTRODUCTION

Cluster analysis seeks to divide a set of objects into a small number of relatively homogeneous groups on the basis of their similarity over N variables [1]. Cluster analysis can be viewed either as a means of summarizing a data set or as a means of constructing a topology [12]. Patterns within a valid cluster are more similar to each other than to a pattern belonging to a different cluster. Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, data mining, document retrieval, image segmentation and pattern classification [33].

Our concern in this paper is based on partitioning clustering [7] methods which relocate instances by moving them from one cluster to another, starting from the initial partitioning.

Partitioning methods try to partition a collection of objects into a set of groups, so as to maximize a pre-defined objective value. The most popular partitional clustering algorithms are the prototype-based clustering algorithms where each cluster is represented by the center of the cluster and the used objective function is the sum of the distances from the patterns to the center [8].

The most popular class of partitional clustering methods is is K-means algorithm [11], where K denotes the number of clusters. The reasons for the algorithmic popularity is its ease of interpretation, simplicity of implementation, speed of convergence, adaptability to sparse data and works fast in most situations [1].

The disadvantages of this algorithm lie in the fact that the number of clusters, K, must be specified prior to application. Also, since the summary statistic is mean of the values for each cluster, so, the individual members of the cluster can have a high variance and mean may not be a good summary of the cluster members. In addition, as the number of clusters grow, for example to thousands of clusters, K-means clustering becomes untenable, approaching the O (n^2) comparisons where n is the number of documents. However, for relatively few clusters and a reduced set of pre-selected words, K-means can do well [12]. The other major drawback of K-means algorithm is sensitivity to initial states. Finally, K-means algorithm converges to the nearest local optimum from the starting position of the search and the final clusters may not be the optimal solution.

In order to overcome these problems that exist in traditional partition clustering methods new techniques have been proposed in this area by researchers from different fields. One of these techniques is optimization methods that tries to optimize a pre-defined function that can be very useful in data clustering.

Optimization techniques define a global function and try to optimize its value by traversing the search space.

Bee Colony Optimization (BCO) [26] is a nature-inspired metaheuristic optimization method, which is similar to the way bees in nature look for food, and the way optimization algorithms search for an optimum in combinatorial optimization problems. The performance of the BCO algorithm has been compared with those of other well-known modern heuristic algorithms such as genetic algorithm, differential evolutional algorithm, and particle swarm optimization algorithm for unconstrained optimization problems. The BCO belongs to the class of population-based and Swarm Intelligence techniques [26], which is considered to be applied to find solutions for difficult combinatorial optimization problems. The major idea behind the BCO is to create the multi agent system (colony of artificial bees) capable to efficiently solve hard combinatorial optimization problems. These features increase the flexibility of the BCO algorithm and produce better solutions. The artificial bee colony behaves to some

 ISBN: 978-1-61208-202-8

extent similar and to some extent in a different way from bee colonies in nature. They explore through the search space looking for the feasible solutions. In order to discover superior and superior solutions, artificial bees cooperate with each other and exchange information. Also, they focus on more promising areas and gradually discard solutions from the less promising areas via collective knowledge and giving out information among themselves.

In this paper, by modeling the partitioning problem as an optimization problem, a BCO-based clustering algorithm is proposed. The performance of the proposed algorithm by applying it to standard benchmark functions and also for clustering real-world data sets is evaluated. The reminder of this paper is organized as follows. In Section 2, some previous related works are summarized. In Section 3, the BCO-based clustering algorithm is described. Section 4 presents data sets used in our experiments, the performance evaluation of the proposed algorithm compared to K-means and GA-based and PSO-based clustering algorithms. Conclusion is discussed in Section 5.

II. DATA CLUSTERING METHODS: A BRIEF OVERVIEW

Data clustering can be hierarchical or partitional [2][3]. A hierarchical algorithm [4][5] creates a hierarchical decomposition of the given dataset forming a dendrogram—a tree which splits the dataset recursively into smaller subsets and represent the objects in a multi-level structure.

Hierarchical clustering algorithms can be agglomerative (bottom-up) or divisive (top-down) [6]. Agglomerative algorithms begin with each element as a separate cluster and merge them into larger clusters. Divisive algorithms begin with the whole set of data objects and proceed to divide it into successively smaller clusters [6].

Partitional clustering algorithms relocate instances by moving them from one cluster to another, starting from the initial partitioning. Such method requires the number of clusters to be preset by the user [1].

Although hierarchical methods are often said to have better quality in clustering, they usually do not provide the reallocation of objects, which may have been poorly classified in the early stages of the analysis [3] and the time complexity of them declared to be quadratic [9]. On the other hand, in recent years the partitioning clustering methods showed a lot of advantages in applications involving large datasets due to their relatively low computational requirements [٩][١٠]. The time complexity of the partitioning technique is almost linear, which makes it widely used.

In addition to the algorithms mentioned above, several heuristics algorithms, such as statistics [13], graph theory [14], expectation-maximization algorithms [15], evolutionary algorithms 18][30-32] and swarm intelligence algorithms [19-25][27] have been proposed for data clustering.

As the behavior of the K-means algorithm mostly is influenced by the number of clusters specified and the random choice of initial cluster centers, in this study, we present a novel algorithm based on the Bee Colony Optimization. BCO is applied in the clustering problem because of its robust, adaptive search method for performing global search. The proposed algorithm, called Bee Colony Clustering, which is good at finding promising areas of the search space but not as good as K-means at fine-tuning within those areas. To demonstrate the effectiveness and speed of proposed algorithm, we have applied these algorithms on various standard datasets and got very good results compared to the K-means and PSO-based clustering algorithm [22]. BCO and PSO algorithms fall into the same class of artificial intelligence optimization algorithms, population-based algorithms, and they are proposed by inspiration of swarm intelligence. Beside, comparing the BCO algorithm with PSO algorithm, the performance of BCO algorithm is also compared with a wide set of classification techniques. The evaluation of the experimental results shows considerable improvements and robustness of the proposed algorithm.

III. THE BASIC BEE COLONY BASED ALGORITHM TO DATA CLUSTERING

In order to cluster data using bee colony algorithm, we must first model the clustering problem as an optimization problem that locates the optimal centroids of the clusters rather than to find an optimal partition. This model offers us a chance to apply bee colony optimization algorithm on the optimal clustering of a collection of data. The following subsections describe the proposed algorithm.

A. Representation of Solutions

In order to apply BCO to solve clustering problem, we have used floating point arrays to encode cluster centers. The assignment matrix has the properties that each data must assigned exactly to one cluster. An assignment that represents K nonempty clusters is a legal assignment. In this model, each food source discovered by each bee is a candidate solution and corresponds to a set of K centroids. Let us denote by a finite set of pre-selected stages, where K is the number of stages. By B, we denote the number of bees to participate in the search process and by I the total number of iterations.

At each forward pass, bees are supposed to visit a single stage. All bees are located in the hive at the beginning of the search process. Each artificial bee allocates some of the data to the corresponding cluster with special probabilities in each stage, and in this way constructs a solution of the problem incrementally. Bees are adding solution components to the current partial solution until they visit all of the K stages. The search process is composed of iterations. The first iteration is finished when bees create feasible solutions. The best discovered solution during the first iteration is saved, and then the second iteration begins. In each iteration of proposed algorithm, for each cluster (stage), all the bees leave the hive to allocate some of the data to that cluster with special probabilities and come back to the hive to see the work of other bees until that time and decide whether to continue its way or select one of the other bees' solution and continue on that way.

B. Evaluation of solutions

A key characteristic of most clustering algorithms is that they use a global criterion function whose optimization drives the entire clustering process. For those clustering algorithms, the clustering problem can be stated as computing a clustering solution such that the value of a particular objective function is optimized. Our objective function is to minimize intra-cluster similarity while maximizing the inter cluster similarity.

Fitness value of each solution is measured by equation:

$$f = \sum_{j=1}^{K} \left(\sum_{i=1}^{n_j} D(d_{ij}, C_j) \right) \tag{1}$$

A food source represents a possible solution to the problem. The quantity of existing sources of pollen, nectar in the areas is explored by the bees corresponds to the quality of the solution represented by that food source. Bees search for food sources in a way that minimize the ration f where f is the proportional to the nectar amount of food sources discovered by bees. In this problem, the goal is to find the minimum of the objective function.

The each iteration of the proposed algorithm is detailed in the following steps:

Step 1. **Initialization**: If this is not the first iteration of the algorithm and the best discovered cluster centers during the previous iterations are available, the initial cluster centers for all the stages are set to the best answer of the previous iteration. Else if this is the first iteration, a set of initial cluster centers generated randomly from the dataset points will be set for each stage. There is a loop from 1 to K where in each loop the following two steps are performed:

Step 2. **Constructive moves in the forward pass**: In each forward pass, every artificial bee visits one stage, allocates the data to the corresponding cluster, and after that returns to the hive as detailed in step 3. For each cluster, the probability of a bee choosing the data i as a member of j^{th} cluster (c_j), p_{ij}, is expressed as follows:

$$p_{ij} = \frac{e^{-D(d_i - C_j)}}{\sum_{m=1}^{n} e^{-D(d_m - C_j)}}, \quad j = 1,2,\ldots,K \tag{2}$$

where $D(d_i - C_j)$ denotes the distance of i^{th} data to j^{th} cluster and n denotes the number of not previously chosen data. Within each forward pass a bee visited a certain number of nodes and created a partial solution. After solutions are evaluated (and normalized) the loyalty decision and recruiting process are performed as described in the following subsection.

Step 3. Backward pass (Bees' partial solutions comparison mechanism): After all of the bees completed step 2, they will be back to hive to compare their partial solutions with themselves. We assume that every bee can obtain the information about solutions' quality generated by all other bees. In this way, bees compare all generated partial solutions. Based on the quality of the partial solutions generated, every bee decides whether to abandon the created partial solution or dance and thus recruit the nest mates before returning to the created partial solution. Depending on the quality of the partial solutions generated, every bee possesses certain level of loyalty to the previously discovered partial solution. Our criterion to decide about the goodness of discovered solution in general is sum of the distance of each vector from its cluster center for all the vectors. We want this criterion to be as minimal as possible. So, as the bees are back at the hive, the probability that b-th bee (during stage u and iteration z) will be faithful to its previously generated partial solution (loyalty decision) is expressed as follows:

$$p_b(u+1, z) = e^{-\frac{O_b(u,z)}{u \times z}} \quad , \qquad b = 1,2,\ldots,B \tag{3}$$

where

$$O_b(u,z) = \frac{SumDistance_b(u,z) - SumDistance_{\min}(u,z)}{SumDistance_{\max}(u,z) - SumDistance_{\min}(u,z)} \tag{4}$$

where

$$SumDistance_b(u,z) = \sum_{i=1}^{u} \sum_{j=1}^{N} D^b{}_{ji}$$

$$D^b{}_{ji} = \begin{cases} \left(\sum_{k=1}^{m} (d^b{}_{jk} - c_{ik})^2\right)^{\frac{1}{2}} & , \textit{ if } d_j \textit{ has been selected by } b-th\ bee \\ 0 & , otherwise \end{cases}$$

$$SumDistance_{\min}(u,z) = \min_i \{SumDistance_i(u,z)\} \quad i = 1,2,\ldots,B$$

$$SumDistance_{\max}(u,z) = \max_i \{SumDistance_i(u,z)\} \quad i = 1,2,\ldots,B$$

We denote by O_b the normalized value of sum distance, with $sumDistance_b$ is sum of the distance of each vector from its cluster center for all the vectors that has been classified by bee number b and minSumDistance is minimum of this sum that exists among all bees.

$SumDistance_{\max}(u,z)$: the objective function value of the worst discovered partial solution from the beginning of the search process

$SumDistance_{\min}(u,z)$: the objective function value of the best discovered partial solution from the beginning of the search process

u : the ordinary number of the forward pass (e.g., u = 1 for first forward pass, u = 2 for second forward pass, etc.) that in each forward pass one of the clusters' members are decided and z denotes the iteration number.

Step 4. Recruiting process: In the case when at the beginning of a new stage a bee does not want to expand the previously generated partial solution, the bee will go to the dancing area and will follow another bee. Within the dance area the bee dancers (recruiters) 'advertise' different partial solutions. We have assumed in this paper that the probability that b's partial solution would be chosen by any uncommitted bee is equal to:

$$p_b = e^{-\theta O_b(u,z)}, \; b = 1,2,...,RC \tag{5}$$

where θ is a coefficient which is a double between 0 and 1 and RC denotes the number of recruiters and O_b denotes the normalized value for the objective function of partial solution created by the b^{th} bee advertised partial solution.

$$O_b = \frac{SumDis\tan ce_b(u+1,z) - SumDis\tan ce_{\min}(u+1,z)}{SumDis\tan ce_{\max}(u+1,z) - SumDis\tan ce_{\min}(u+1,z)} \tag{6}$$

where $maxSumDistance$ is the maximum sum of the distance of each vector from its cluster centers for all the data that has been classified until now that exists among all the bees.

This probability p_b is used in a roulette wheel selection or tournament selection algorithm and one of the bees is selected.

Using Eq. (5) and a random number generator, every uncommitted follower joins one bee dancer (recruiter). Recruiters fly together with their recruited nestmates in the next forward pass along the path discovered by the recruiter. So the bee that wants to continue another partial solution will set its partial solution exactly as the selected bee but will continue the algorithm independently. At the end of this path, all bees are free to independently search the solution space and generate the next iteration of constructive moves.

Step 5. **Set the cluster centers** (compute the Centroid of Clusters): At last, the cluster centers as the centroid of the vectors belong to each cluster for each bee are computed as follows: Each solution extracted by each bee corresponds to a clustering with assignment matrix A. Let $C = (c_1, c_2, ...c_i, ..., c_K)$ is set of K centroids for assignment matrix A. The centroid of the k^{th} cluster is $c_k = (c_{k1}, c_{k2}, ..., c_{Km})$ and is computed as follows:

$$c_{kj} = \frac{\sum_{i=1}^{n} (a_{ki}) d_{ij}}{\sum_{i=1}^{n} a_{ki}} \tag{7}$$

where m is the number of dimensions in all data.

Step 6. **Selecting the best answer**: In this phase, among all generated solutions, the best one is determined and is used to update the global best. The global best will be used for setting the cluster centers for all the stages in next iteration. At this point, all B solutions are deleted, and the new iteration starts. The BCO runs iteration by iteration until a stopping condition is met.

IV. EXPERIMENTAL RESULTS

In this section, we present the experimental evidences and results that were made on several standard datasets, and the comparisons that were made with other relevant works.

A. Dataset Description

In this work, five clustering problems from the UCI database [28], which is a well-known database repository are used to evaluate the performance of the proposed algorithm.

Data Set 1: Fisher's Iris plants database (n = 150, d = 4, K = 3): It is perhaps the best-known database to be found in the pattern recognition literature.

The data set contains four inputs, three classes, and150 data vectors.

Data Set 2: Glass (n = 214, d = 9, K = 6): The data were sampled from six different types of glass: 1) building windows float processed (70 objects); 2) building windows non float processed (76 objects); 3) vehicle windows float processed (17 objects); 4) containers (13 objects);5) tableware (9 objects); and 6) headlamps (29 objects).Each type has nine features: 1) refractive index; 2) sodium; 3) magnesium; 4) aluminum; 5) silicon;6) potassium; 7) calcium; 8) barium; and 9) iron.

Data Set 3: Wisconsin breast cancer data set (n = 683, d=9, K=2):The Wisconsin breast cancer database contains nine relevant features: 1) clump thickness; 2) cell size uniformity;3) cell shape uniformity; 4) marginal adhesion; 5) single epithelial cell size; 6) bare nuclei; 7) bland chromatin;8) normal nucleoli; and 9) mitoses. The data set has two classes. The objective is to classify each data vector into benign (239 objects) or malignant tumors (444 objects).

Data Set 4: (n = 178, d = 13, K = 3): This is a classification problem with "well-behaved" class structures. There are13 features, three classes, and 178 data vectors.

Data Set 5: Vowel data set (n = 871, d = 3, K = 6): This data set consists of 871 Indian Telugu vowel sounds. The data set has three features, namely F1, F2, and F3, corresponding to the first, second and, third vowel frequencies, and six overlapping classes {d (72 objects), a (89 objects), i (172 Objects), u (151 objects), e (207 objects), o (180 objects)}.

B. Experimental setup

In the next step, the K-means and the proposed algorithm are applied to the above mentioned data sets. The cosine correlation measure is used as the similarity metrics in each algorithm. The results shown in the rest of paper, for every dataset, are the average of over 20 independent runs of the algorithms (to make a fair comparison), each run with randomly generated initial solutions and different seeds of the random number generator. Also, for an easy comparison, the algorithms run 1,000 iterations in each run since the 1,000 generations are enough for convergence of algorithms.

C. Comparisons and discussions

In the previous subsection, the structure of datasets were explained. Now, in this section, we evaluate and compare the performances of the proposed algorithm according to its quality of generated clusters with K-mean [11], PSO [22] and a GA-based [29] clustering algorithm. For evaluation of the clustering results' quality, we use SICD metric which has been selected from internal measures. Whereas SICD examines how much the clustering satisfies the optimization constraints. The smaller the SICD value, the more compact the clustering solution is. Table 1 demonstrates the normalized SICD value of algorithms.

Looking at the results in Table 1, we can see that the results obtained by proposed algorithm are significantly comparable by results obtained by the other evolutionary based algorithms.

TABLE 1- SICD COMPARISONS AMONG PROPOSED ALGORITHM AND THE OTHER ALGORITHMS

		GA	TS	SA	ACO	K-means	PSO	Proposed Algorithm
Iris	Average	139.98	97.86	97.13	97.17	106.05	103.51	97.05
	Worst	193.78	98.57	97.26	97.81			97.33
	best	125.19	97.36	97.1	97.1	97.33	96.66	97.22
Wine	Average	16530.5	16785.46	16530.53	16530.5	18061	16311	16449.81
	Worst	16530.5	16837.54	16530.53	16530.5			16461.8
	best	16530.5	16666.22	16530.53	16530.5	16555.68	16294	16433.37
Glass	Average	_	_	_	_	260.4	291.33	225.19
	Worst	_	_	_	_			250.44
	best	_	_	_	_	215.68	271.29	214.85
Cancer	Average	_	_	_	_	2988.3	3334.6	2976.89
	Worst	_	_	_	_			2977.57
	best	_	_	_	_	2987	2976.3	2976.24
Vowel	Average	_	_	_	_	159242.9	168477	150881.16
	Worst	_	_	_	_			154469.62
	best	_	_	_	_	149422.3	163882	149466.61

V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a swarm-based data clustering technique. In the proposed algorithm, a group of bees created k centroids, as the cluster centers of each cluster and then data are assigned to the clusters. In other words in the proposed algorithm the solutions represented by the bees were considered as initial centroids for each center of the k-means clusters, which led to significant improvements. Also some relevant comparisons were made, to demonstrate the effectiveness of the algorithms. Our experimental results on different datasets showed that proposed algorithm produces better solutions with high quality in comparison with other algorithms and the difference is tremendous. Different improvements can be done to enhance the evaluation metrics. The bee colony based algorithm can be extended by K-means algorithm through different hybridization methods. For example by running k-means and bee colony colony alternatively 2 different procedure would be produced.

REFERENCES

[1] S. Hanuman, V. Babu, A. Govardhan, and S. C. Satapathy, "Data Clustering Using Almost Parameter Free Differential Evolution Technique", International Journal of Computer Applications, vol. 8, no. 13, pp. 1-7, 2010.

[2] J. Han, M. Kamber, and A. K. H. Tung, "Spatial Clustering Methods In Data Mining: A Survey", In Geographic Data Mining and Knowledge Discovery, New York, 2001.

[3] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review", ACM Computing Surveys (CSUR), vol. 31, no. 3, pp. 264-323, 1999.

[4] S. Guha, R. Rastogi, and K. Shim, "CURE: An Efficient Clustering Algorithm for Large Databases", In Proc. of ACM-SIGMOD Int. Conf. Management of Data (SIG-MOD98), pp. 73-84, 1998.

[5] G. Karypis, E. H. Han, and V. Kumar, "CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling", Computer, vol. 32, pp. 68-75, 1999.

[6] S. Xu and J. Zhang, "A Parallel Hybrid Web Document Clustering Algorithm and Its Performance Study" , The Journal of Supercomputing, vol. 30, pp. 117-131, 2004.

[7] P. S. Bradley, U. M. Fayyad, and C. A. Reina, "Scaling EM (Expectation Maximization) Clustering To Large Databases", Microsoft Research Technical Report, 1998.

[8] B. Mirkin, "Mathematical Classification and Clustering", Kluwer Academic Publishers, Dordrecht, the Netherlands, 1996.

[9] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques", KDD2000, Technical report of University of Minnesota,2000.

[10] J. Kennedy, R. C. Eberhart, and Y. Shi, "Swarm Intelligence", Morgan Kaufmann, New York, 2001.

[11] J. B. MacQueen, "Some Methods For Classification And Analysis Of Multivariate Observations", Proceedings of the Fifth Berkeley Symposium on Mathematical Statistic and Probability, University of California Press, Berkley, pp. 281-297, 1967.

[12] S. Vaithyanathan and B. Dom, "Model Selection in Unsupervised Learning with Applications to Document Clustering", In Proceedings International Conference on Machine Learning, 1999.

[13] E. W. Forgy, "Cluster Analysis of Multivariate Data: Efficiency Versus Interpret Ability of Classification", Biometrics, vol. 21, no. 3, pp. 768–769, 1965.

[14] C. T. Zahn, "Graph-Theoretical Methods For Detecting And Describing Gestalt Clusters", IEEE Trans. Comput., pp. 68–86, 1971.

[15] T. Mitchell, "Machine Learning", McGraw-Hill, New York, 1997.

[16] J. Mao and A. K. Jain, "Artificial Neural Networks For Feature Extraction And Multivariatedata Projection", IEEE Trans. Neural Netw, pp. 296–317, 1995.

[17] S. H. Liao and C. H. Wen, "Artificial Neural Networks Classification and Clustering of methodologies and Applications Literature Analysis From 1995 To 2005", ExpertSys. Appl, vol. 32, pp. 1–11, 2007.

[18] S. Paterlini and T. Minerva, "Evolutionary Approaches for Cluster Analysis", Soft Computing Applications, Springer–Verlag, pp. 167–178, 2003.

[19] C. H. Tsang and S. Kwong, "Ant Colony Clustering And Feature Extraction For Anomaly Intrusion Detection", Stud. Comput. Intell, vol. 34, pp. 101–123, 2006.

[20] R. Younsi and W. Wang, "A New Artificial Immune System Algorithm for Clustering", IDEAL 2004, LNCS 3177, Springer, Berlin, pp. 58–64, 2004.

[21] P. S. Shelokar, V. K. Jayaraman, and B. D. Kulkarni, "An Ant Colony Approach for Clustering", Anal. Chim. Acta 509 , pp. 187–195, 2004.

[22] S. Paterlini and T. Krink, "Differential Evolution and Particle Swarm Optimization In Partitional Clustering", Comput. Stat. Data Anal, pp. 1220–1247, 2006.

[23] Y. Kao and K. Cheng, "An ACO-Based Clustering Algorithm", in ANTS, LNCS 4150, Springer, Berlin, pp. 340–347, 2006.

[24] M. Omran, A. Engelbrecht, and A. Salman, "Particle Swarm Optimization Method for Image Clustering", Int. J. Pattern Recogn. Artif. Intell, vol. 19, no. 3, pp. 297–322, 2005.

[25] D. T. Pham, S. Otri, A. Afify, M. Mahmuddin, and H. Al-Jabbouli, "Data clustering using the bees algorithm," In: Proc. 40th CIRP International Manufacturing Systems Seminar, 2007, Liverpool.

[26] P. Lucic and D. Teodorovic, " Bee System: Modeling Combinatorial Optimization Transportation Engineering Problems by Swarm Intelligence, In preprints of the TRISTAN IV Triennial symposium on Transportation Analysis. Sao Miguel, Azores Island, pp. 441-445, 2001.

[27] K. Krishna and M. NarasimhaMurty, "Genetic K-Means Algorithm", IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics, vol. 29, no. 3, pp. 433-439, 1999.

[28] P. Murphy and D. Aha, "UCI Repository of Machine Learning Databases", 1995, URL http://www.sgi.com/Technology/mlc/db, [retrieved: 03, 2012].

[29] U. Mualik and S. Bandyopadhyay, "Genetic Algorithm Based Clustering Technique", Pattern Recognition, vol. 33, pp. 1455–1465, 2002.

[30] R. Forsati, A. Moayedikia, and B. Safarkhani, "Heuristic Approach To Solve Feature Selection Problem", DICTAP 2011, 2011, pp. 707-717.

[31] R. Forsati, M. Shamsfard, and P. Mojtahedpour, "An Efficient Meta Heuristic Algorithm For Pos-Tagging", Fifth International Multi-Conference on Computing in the Global Information Technology (ICCGI), pp. 93-98, 2010.

[32] R. Forsati, M. Mahdavi, M. Kangavari, and B. Safarkhani, "Web page clustering using Harmony Search optimization", Canadian Conference on Electrical and Computer Engineering, CCECE 2008, pp. 1601-1604, 2008.

[33] D. B. Kenneth, "Cluster Analysis. Sociological Methodology", vol. 6, pp. 59-128, American Sociological Association, 1975.