Application of the Bees Algorithm to Fuzzy Clustering

D.T. Pham, H. AL-Jabbouli, M. Mahmuddin, S. Otri and A. Haj Darwish

Intelligent Systems Laboratory, Manufacturing Engineering Centre, Cardiff University, Cardiff, CF24 3AA, UK

Abstract

This paper discusses the application of the Bees Algorithm to fuzzy clustering. The Bees Algorithm is used to optimise the performance of the fuzzy C-Means (FCM) algorithm and improve its clustering results. Computational experiments show that the Bees Algorithm gives a significant improvement over the FCM algorithm on its own and better results compared to the FCM algorithm combined with a Genetic Algorithm.

Keywords: Bees Algorithm, Genetic Algorithm, Fuzzy Data Clustering, Optimisation.

1. Introduction

Data clustering is known as the problem of grouping similar data points together, such that similar data points are in the same group. Traditional data clustering, also called crisp clustering, uses hard thresholds to group data points into separate groups while fuzzy clustering uses fuzzy logic to create overlapped groups of data.

In fuzzy clustering, each data object can belong to more than one cluster at the same time with a different possibility degree or membership function value. The membership function of a cluster may vary from 1 to 0. Conversely, membership function in crisp clustering is based on either belonging to the cluster, which normally is 1 and is otherwise 0. These membership functions are depicted in Figs. 1 and 2.

A Fuzzy clustering algorithm is known as one that produces more realistic results compare to crisp clustering techniques, and has been widely used especially in pattern recognition[1] and medical applications[2]. In this paper, the Bees Algorithms, which has been used to improve crisp clustering results by integrating it with the k-means algorithm [5], is now integrated with the (FCM) algorithm, one of the most well-known fuzzy clustering algorithms, to solve the problem of trapping in a local optimum.

The rest of the paper is organised as follows: section 2 reviews the (FCM) algorithm and the Bees Algorithm; section 3 describes the proposed fuzzy clustering algorithm that combines the Bees Algorithm with (FCM) algorithm; section 4 shows the results obtained. Finally, section 5 concludes the paper.

2. FCM and the Bees Algorithm

2.1. FCM

The fuzzy C-means (FCM) algorithm is the most popular algorithm in fuzzy clustering. It was introduced and developed by Dunn [3] then improved by Bezdek [1]. The fuzzy C-means algorithm is based on minimising of an objective function (Eq. 1) that indicates the sum of distances from each cluster centre to the data points in that cluster, so that the smaller the value of J, the better the clustering.

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \qquad \text{(Eq. 1)}$$

where:
m is the fuzziness degree of any real number greater than 1.
uij is the degree of membership of xi in the cluster j
cj is the centre of the cluster j
N is the number of data objects
xi is the ith d-dimensional measured data object.
||*|| is any norm expressing the similarity between any measured data and the centre

The total number of clusters is a required parameter. When performing fuzzy clustering, the FCM algorithm produces fuzzy clusters and the membership degree of each data object to each cluster by repeating the following steps:

1- Initialise $U = \lfloor u_{ij} \rfloor$

2- At k-step: calculate the centres vectors $C^{(k)} = [c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

3- Update $U^{(k)}$, $U^{(k+1)}$

$$U_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}}$$

4- If $\left\| U^{(k+1)} - U^{(k)} \right\| \langle \sigma$ then stop, otherwise go to step 2.

Fig. 3. FCM Algorithm

where:
K is the iteration step
σ is a termination criterion between 0 and 1

The main problem of this calculus-based optimisation approach is that it can become trapped in a local optimum. In this case, the Bees optimisation algorithm has been used to overcome the problem.

2.2. Honey Bees in nature

The Bees Algorithm is an optimisation algorithm proposed by Pham et al [4] based on the food foraging behaviour of honey_bees in nature. The Bees Algorithm has become one of the most successful optimisation algorithms due to its successful implementation in various applications for optimisation problems such as neural network training [5], identifying homogenous data clustering [6] and a preliminary design problem [7].

The foraging process of bees begins in a colony by scout bees being sent to search for promising flower patches. Flower patches with large amounts of nectar or pollen that can be collected with less effort tend to be visited by more bees, whereas patches with less nectar or pollen receive fewer bees.

During the harvesting season, a colony continues its exploration, keeping a percentage of the population as scout bees. When they return to the hive, those scout bees that have found a patch rated above a certain quality threshold deposit their nectar or pollen and go to the “dance floor” to perform a dance known as the “waggle dance”.

This mysterious dance is essential for colony communication, and contains three pieces of information regarding a flower patch: the direction in which it will be found, its distance from the hive and its quality rating (or fitness). This information helps the colony to send its bees to flower patches precisely, without using guides or maps. Each individual’s knowledge of the outside environment is gleaned solely from the waggle dance. This dance enables the colony to evaluate the relative merit of different patches according to both the quality of the food they provide and the amount of energy needed to harvest it. After waggle dancing on the dance floor, the dancer (i.e. the scout bee) goes back to the flower patch with follower bees that were waiting inside the hive. More follower bees are sent to more promising patches. This allows the colony to gather food quickly and efficiently.

While harvesting from a patch, the bees monitor its food level. This is necessary so that they can decide upon the next waggle dance when they return to the hive. If the patch is still good enough as a food source, then it will be advertised in the waggle dance and more bees will be recruited to that source.

2.3 The Bees Algorithm

The basic steps of the Bees Algorithm are explained in fig 4. The Algorithm requires a number of parameters to be set. These are; number of scout bees (n), number of sites selected for neighbourhood search (m), number of best “elite” sites out of m selected sites (e), number of bees recruited for the best e sites (nep), number of bees recruited for the other (m-e) selected sites (nsp) and a stopping criterion.

1. Initialise the solution population (each initial solution, or ‘bee’, being a set of randomly placed cluster centres).
2. Evaluate the fitness of the population.
3. While (stopping criterion is not met)
   // forming new population.
4. Select sites for neighbourhood search.
5. Recruit bees for selected sites (more bees for the best e sites) and evaluate fitnesses.
6. Select the fittest bee from each site.
7. Assign remaining bees to search randomly and evaluate their fitnesses.
8. End While.

Fig. 4. The Bees Algorithm

3. Proposed Fuzzy Clustering Method

The proposed clustering method exploits the search capability of the Bees Algorithm to overcome the local optimum problem of the FCM algorithm. Specifically, the task is to search for appropriate cluster centres such that the objective function (J) is minimised.

The basic steps of the proposed clustering operation are described in detail below:

1. Read data from the file.
2. Initialise the solution population.
3. Fill array *A* using formula u_{ij} and evaluate the fitness of the population using Eq. (1).
4. While (stopping criterion is not met)
5. Select sites for neighbourhood search.
6. Recruit bees for selected sites (more bees for the best e sites) and fill array *A* using formula u_{ij} and evaluate the fitness using Eq. (1).
7. Select the fittest bee from each site.
8. Assign remaining bees to search randomly then fill array *A* using formula u_{ij} and evaluate their fitnesses using Eq. (1).
9. End While.

Fig. 5. The Proposed Algorithm

The algorithm starts by reading data and providing a random population of clusters centres as in step 2. In step 3, a calculation is made using equation U in order to fill the array A with a membership function for all data objects. A repeat of generating a new population of solutions is done in steps 4-9, by selecting the best sites of former population as in step 5 then sending more bees to them. Step 6 fills the array A using equation U for each sent bee. In step 7, selection of the fittest bees is chosen from each site. Assigning the remaining bees to search randomly and fill the array A is step 8.

4. Experimental Results

This section presents the results for the Bees-Algorithm-based fuzzy clustering method compared with the FCM algorithm and GA -based fuzzy clustering method.

Table 1 shows the parameter values for each algorithm used in this test, while the value of m is set to 2 in all algorithms. The algorithms were applied to the following real data sets (Iris [8], Control Charts [9], Wood Defects [10] and Crude Oil [11]). The main characteristics of these data sets are summarized in Table 3.

The clustering criterion J Eq. (1) is used to evaluate the performance of the algorithms: the smaller the value of this metric, the better the fuzzy clustering results. The algorithms were executed 10 times and the average, minimum and maximum values of E were collected. The results obtained for each algorithm are listed in Table 2.

The results in Table 2 show that the Bees and GA enhanced algorithms outperform the traditional FCM algorithm in most cases. The Table also shows that the Bees Algorithm produces better results than the GA in all cases. It also shows that the FCM algorithm performs better on the Control Charts data set compared to the other algorithms.

Table 1
Parameters used in the clustering experiments

Algorithm	Parameters	Value
FCM	Maximum number of iterations	1000
GA	Crossover probability, μ_c	0.8
	Mutation probability, μ_m	0.001
	Population size, P	100
Bees Algorithm	Number of scout bees	21
	Number of sites selected for neighbourhood search	8
	Number of elite bees	2
	Number of bees recruited for the elite sites	5
	Number of bees recruited for the other selected sites	2
	Number of iterations	300

5. Conclusion

Combining the Bees Algorithm with the FCM algorithm improved the fuzzy clustering results compared to the traditional C-means algorithm in most cases. It also proved that the Bees Algorithms produces better results than those of the GA combined with FCM.

One of the main concerns about the current algorithm is that it requires a long computing time. Future work should therefore concentrate on improving the computational speed of the Bees Algorithm.

Table 2
Results (clustering criterion) for the tested clustering algorithms

Data set	Algorithm	Mean	Min	Max
Iris	FCM	61.72	60.58	65.81
	GA	65.54	63.34	69.23
	Bees	**60.58**	**60.58**	**60.58**
Ctrl Charts	FCM	**530.99**	**530.99**	**530.99**
	GA	802.08	791.23	819.28
	Bees	549.08	546.87	552.15
Wood Defects	FCM	64077395.11	59195094.29	66225010.49
	GA	168035.20	157508.00	174784.00
	Bees	**165132.52**	**153866.53**	**173523.00**
Crude Oil	FCM	1330.43	1242.30	1560.57
	GA	1252.61	1238.25	1286.04
	Bees	**1239.50**	**1237.24**	**1241.56**

Table 3
Data sets used in the experiments

Data Set	# of Object	# of Feature	# of Class
Iris	150	4	3
Control Charts	1500	60	6
Wood Defects	232	17	13
Crude oil	59	5	3

References

[1] Keller B.J.M., Krishnapuram R., and Pal N.R. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. 1. 1999: Springer.

[2] Wang S., Zhou M., and Geng G. Application of fuzzy cluster analysis for medical image data mining, in Mechatronics and Automation, 2005 IEEE International Conference. 2005. pp. 631- 636.

[3] Dunn J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 1973. 3: pp. 32-57.

[4] Pham D.T., Ghanbarzadeh A., Koc E., Otri S., and Zaidi M. The Bees Algorithm – A Novel Tool for Complex Optimisation Problems, in 2nd Virtual International Conference on Intelligent Production Machines and Systems (IPROMS 2006). 2006, Elsevier. Cardiff, UK. pp. 454-459.

[5] Pham D.T., Ghanbarzadeh A., Koc E., and Otri S. Application of the Bees Algorithm to the training of Radial Basis Function Networks for control chart pattern recognition, in 5th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering (ICME '06). 2006, CIRP: Ischia, Italy. pp. 711-716.

[6] Pham D.T., Otri S., Afify A., Mahmuddin M., and Al-Jabbouli H. Data clustering using the Bees Algorithm, in 40th CIRP International Seminar on Manufacturing Systems. 2007: Liverpool.

[7] Pham D.T., Castellani M., and Ghanbarzadeh A. Preliminary design using the Bees Algorithm, in 8th International Conference on Laser Metrology, CMM and Machine Tool Performance, LAMDAMAP. 2007, Euspen: Cardiff, UK. pp. 420-429.

[8] Asuncion A. and Newman D.J. UCI Machine Learning Repository. 2007, University of California, School of Information and Computer Science: Irvine, CA.

[9] Pham D.T. and Oztemel E. Control chart pattern recognition using neural networks. Journal of Systems Engineering, 1992. 2, 4: pp. 256-262.

[10] Pham D.T., Soroka A.J., Ghanbarzadeh A., Koc E., Otri S., and Packianather M. Optimising Neural Networks for Identification of Wood Defects Using the Bees Algorithm, in 4th International IEEE Conference on Industrial Informatics (INDIN'06). 2006, IEEE Computer Society: Singapore. pp. 1346-1351.

[11] Johnson R.A. and Wichern D.W. Applied Multivariate Statistical Analysis 1982: Prentice Hall, Englewood Cliffs, NJ

.