

# AN IMPROVED TEACHING-LEARNING BASED OPTIMIZATION APPROACH FOR FUZZY CLUSTERING

Parastou Shahsamandi E.<sup>1</sup> and Soheil Sadi-nezhad<sup>2</sup>

<sup>1,2</sup>Department of Industrial Engineering, Science & Research Branch,  
Islamic Azad University, Tehran, Iran

<sup>1</sup>p.shahsamandi@mau.ac.ir, <sup>2</sup>sadinejad@hotmail.com

## ABSTRACT

*Fuzzy clustering has been widely studied and applied in a variety of key areas of science and engineering. In this paper the Improved Teaching Learning Based Optimization (ITLBO) algorithm is used for data clustering, in which the objects in the same cluster are similar. This algorithm has been tested on several datasets and compared with some other popular algorithm in clustering. Results have been shown that the proposed method improves the output of clustering and can be efficiently used for fuzzy clustering.*

## KEYWORDS

*Meta-heuristic algorithm, Improved teaching-learning-based optimization, Fuzzy clustering*

## 1. INTRODUCTION

Data clustering is an important problem in data mining and knowledge discovery. The main objective of any clustering technique is grouping a set of objects in a number of clusters; in such a way that objects in one cluster are very similar and objects in different clusters are quite distinct.[1,2] One of the proper measures of similarity for data in  $K$  cluster is the distance between data and their cluster center (e.g. the Euclidean distance in fuzzy c-means algorithm proposed by [3]). In fact this unsupervised classification produces a  $K \times P$  optimum partition matrix  $U^*(x)$  of the given dataset  $X$ , consisting of  $m$  data samples,  $X = \{x_1, x_2, x_3, \dots, x_m\}$ , where each  $X_i$  in the universe  $X$  is an  $p$ -dimensional vector of  $m$  elements or  $m$  features; and  $i=1, 2, \dots, m$ . The partition matrix can be represented as  $U = \{u_{ki}\}$ ,  $K=1, 2, \dots, k$ , and  $u_{ki}$  is the membership of object  $X_i$  to  $k$ th cluster. For fuzzy clustering of data,  $0 < u_{ki} < 1$ , (i.e.  $u_{ki}$  denotes the degree of belonging of object  $X_i$  to the  $k$ th cluster.)

Finding the optimum matrix,  $U^*$ , is difficult for practical problems, so the application of advanced optimization techniques is required. Considering that, the clustering problem is NP-hard problem (when the number of data and number of cluster are exceeded), the application of metaheuristic method is necessary for partitioning data. [4]

The metaheuristic algorithms can be classified into different groups depending on the criteria being considered; the evolutionary algorithms (ES), (such as Genetic algorithm, Differential

Evolution) and Swarm intelligence algorithm (such as Particle Swarm Optimization, Ant Colony Optimization, and Artificial bee colony) are based on population criteria. Beside these algorithms,[5] there are some other algorithms which work on the principles of different natural phenomena, such as Harmony Search [6], Gravitational Search algorithm,[7] Teaching-Learning Based optimization. [8,9]

The metaheuristic algorithms can solve large problem faster and can obtain robust algorithms, more over these algorithms are simple to design and implement [4]. Many of these algorithms have been introduced to solve clustering problems. [10-12] have been used Genetic Algorithm(GA) to optimizing a specified objective function related to solve the clustering problem. [13,14] proposed a Tabu Search based heuristic for clustering. [15] considered the problem of clustering and Simulated Annealing approach has been proposed for solving problem. [16] presented an Ant Colony clustering algorithm, which simulate the way real ants find the shortest path from their nest to food source and back. They compared the performance of this algorithm with other algorithms. [17] hybridized Particle Swarm Optimization with K-means and Nelder-Mead simplex search method to improve the performance of algorithm for clustering problem.[18] presented an Artificial Bee Colony to optimally solve the clustering problem. The performance of this algorithm has been compared with other popular heuristics algorithm in clustering. [8,9] implemented Teaching-Learning-Based Optimization (TLBO) for automatic clustering and named it Auto-TLBO. This algorithm is evaluated on benchmark datasets and performance comparisons are made with some well-known clustering algorithms.

Different algorithm requires common controlling parameters such as population size, number of generations and its own algorithm specific control parameters (e.g. mutation rate and cross over rate in GA algorithm; inertia weight and social and cognitive parameters in PSO). Among of all algorithms, [8,9] showed that TLBO algorithm does not requires any algorithm-specific parameters. TLBO algorithm simulates the teaching-learning phenomenon of a classroom, where a group of learners are considered the population and different subjects offered to the learners are similar to different design variables of optimization problem. The best solution in the entire population is considered as teacher.

This algorithm has been improved by introducing more than one teacher for learners (i.e. increased the collective knowledge) and some other modifications. [19]

Thereby, in this paper the Improved TLBO is proposed for fuzzy clustering problem; the objective function of fuzzy c-means algorithm is used as fitness function and Euclidian distance metric as a distance metric. The minimum amount of this objective shows the better clustering. Clustering results are reported for a number of real-life and artificial datasets. The performance of algorithm compared with several other proposed clustering algorithms. This paper is organized as follows: the next section discussed the fuzzy c-mean algorithm. In section 3, the improved TLBO algorithm for data clustering is described. Section 4 presents the experimental results conducted on several data sets .Finally, section 5 concludes the article.

## 2. FUZZY CLUSTERING ALGORITHM

Let  $X_{m \times p}$  be the profile data matrix, with  $m$  rows (set of  $m$  objects) and  $p$  columns ( $p$ -dimensional), where each  $X_{ij}$  corresponds to the  $j$ th real value feature ( $j=1, 2, \dots, p$ ) of  $i$ th object ( $i=1, 2, \dots, m$ ). Given  $X_{m \times p}$  the goal of partitioned clustering algorithm is to find grouping or structures; such that objects which are assigned to the same cluster should be similar (i.e. Homogeneity), while objects which are assigned to different clusters should be different (i.e. Heterogeneity).

In most cases the data is in the form of real value vector. The Euclidean distance is a suitable measure of similarity for these datasets. The Euclidean id derived from Minkowski metric.(Eq.1)

$$d(x, y) = (\sum_{i=1}^p |x_i - y_i|^r)^{\frac{1}{r}} \xrightarrow{r=2} d(x, y) = (\sum_{i=1}^p |x_i - y_i|^2)^{\frac{1}{2}} \quad \text{Eq.1}$$

Fuzzy c-mean (FCM) is a widely used technique which allows a datum to belong to more than one cluster. (Eq.2), It is based on minimization of the following measure:

$$J_m = \sum_{i=1}^p \sum_{k=1}^K u_{ij}^{m'} d_{ij} \quad \text{Eq.2}$$

Where  $p$  is the number of data objects,  $K$  represent number of clusters,  $u$  is the fuzzy membership matrix;  $m'$  ( $m' > 1$ ) is the weighting exponent and controls the fuzziness of resulting clusters and  $d_{ij}$  is euclidian distance from data  $X_i$  to cluster center. This criterion is based on the compactness of data in clusters.[3]

### 3. ITLBO ALGORITHM BASED FUZZY CLUSTERING

TLBO algorithm simulates the teaching-learning process that every individual tries to learn something from other individual to improve themselves. The algorithm simulates two fundamental modes of learning: Teacher phase and Learner phase. A group of learner is considered the population of algorithm and the results of learner are the fitness value of optimization problem, which indicates its quality. [8,9]

In teacher phase, the learning of the learner through teacher is simulated. During this phase, a teacher conveys knowledge among the learners and makes an effort to increase the mean result of the class. At any iteration of algorithm, there are  $n$  number of learners (population size) and  $m$  number of subjects.  $M_{j,i}$  be the mean result of learners iv subject  $j$ th. The best overall result  $X_{total-kbest,i}$  is the result of the best learner. A teacher is the most experience person (the best learner) in the algorithm. The difference between the result of the teacher and the mean result of the learner in each subject in given by (Eq.3),

$$Difference\_Mean_{j,i} = r_i (X_{j,kbest,i} - T_F M_{j,i}) \quad \text{Eq.3}$$

Where  $X_{j,kbest,i}$  is the result of the teacher (best learner) in subject  $j$ .  $T_F$  is the teaching factor which decides the value of mean to be changed, and  $r_i$  is the random number in range  $[0,1]$ . Based on the  $Difference\_Mean_{j,i}$ , the existing solution is updated in the teacher phase according to the following expression:

$$X'_{j,k,i} = X_{j,k,i} + Difference\_Mean_{j,i}$$

Where  $X'_{j,k,i}$  is the update value; accept it if it gives a better function value. These accepted values become the input to the learner phase.

The learner phase of the algorithm simulates the learning of the learners; through interaction among themselves. The learners can also gain knowledge by interacting with other learners. A learner interacts randomly with other learners and learns new things if the other learner has more knowledge than him or her. Randomly select two learners  $P$  and  $Q$ , such that

$X'_{total-P,i} \neq X'_{total-Q,i}$ , (where these values are the updated value at the end of teacher phase). The following equations are for maximization problem.

$$X''_{j,P,i} = X'_{j,P,i} + r_i (X'_{j,P,i} - X'_{j,Q,i}) \quad , \\ \text{if } X'_{total-P,i} > X'_{total-Q,i}$$

$$X''_{j,P,i} = X'_{j,P,i} + r_i (X'_{j,Q,i} - X'_{j,P,i}) \quad , \\ \text{if } X'_{total-Q,i} > X'_{total-P,i}$$

Accept  $X''_{j,P,i}$ , if it gives a better function value.

The algorithm stops with the criteria such as the maximum iteration and the minimum change of objective function.

[19] improved the algorithm by introducing more than one teacher for learners and some modifications to adaptive teaching factor and self motivated learning; named it ITLBO. In this paper the ITBO proposed for clustering.

Based on the mentioned statements, the steps of ITLBO algorithm with detailed description of each are as follows:

**Step 1:** objective function; define the optimization problem as minimizing overall deviation of a partitioning or maximize compactness inside the clusters. This is simply computed as the overall summed distances between data items and their corresponding cluster center; (i.e. the objective function of fuzzy c- mean algorithm; Eq.2) the weighting exponent  $m'$  in Eq.2 is set to 2, which is a common choice for fuzzy clustering. Considering that, TLBO algorithm does not require any algorithm-specific parameter; so, setting the control parameter value is not necessary.

**Step 2:** Initialization; Initialize the population (N learner), in order to solve clustering, each candidate solution in the population consist of N number  $U_{k \times m}$  matrix, where each elements of this matrix represents the degree of belonging objects to  $k$ th clusters. The fuzzy matrix  $U$  is generated randomly according to population size, then the center of each cluster is compute to find the distance between each data and the centroids of clusters (as Eq.2). In the experiment, we set the population size or the number of learner as 100.

**Step 3:** Evaluation; evaluate the population, then select and assign the best solution as chief teacher to first rank (i.e.  $f(x)_{best}$ ).  $(X_{teacher})_1 = f(x)_1$ , where  $f(x)_1 = f(x)_{best}$ . Select the other teachers based on the chief teacher and rank them.  $f(x)_s = f(x)_1 - \text{rand} \times f(x)_1$ , where  $s$  is the number of teacher as selected; (if the equality is not met, select the  $f(x)_s$  closet to the value calculated above). We select four teachers in this algorithm.

**Step 4:** Assignment; assign the learners to the teachers according to their fitness value as:  $(X_{teacher})_s = f(x)_s$ , where  $s = 1, 2, \dots, T$  (in this paper  $T=4$ );

For  $k=1: (n-s)$

if  $f(x)_1 \leq f(x)_k < f(x)_2$ ,

assign the learner  $f(x)_k$  to teacher 1

else, if  $f(x)_2 \leq f(x)_k < f(x)_3$ ,  
 assign the learner  $f(x)_k$  to teacher 2  
 else, if  $f(x)_3 \leq f(x)_k < f(x)_4$ ,  
 assign the learner  $f(x)_k$  to teacher 3  
 else, assign the learner  $f(x)_k$  to teacher 4  
 end

**Step 5:** keep the elite solutions of each group; elitism is a mechanism to preserve the best individual from generation to generation. Therefore the system never loses the best individuals found during the optimization process. It can be done by placing one or more of the best individual directly in to the population for the next generation.

**Step 6:** Updating; calculate the mean result of each group of learners in each subject (i.e.  $(M_j)_s$ ) and evaluate the difference given by Eq. 3. For each teacher, the adaptive teaching factor is as:

$$(T_F)_i = \left( \frac{X_{total-k}}{X_{total-kbest}} \right), \quad k=1, \dots, n,$$

$$(T_F)_i = 1, \quad \begin{array}{l} \text{if } X_{total-kbest,i} \neq 0 \\ \text{if } X_{total-kbest,i} = 0 \end{array}$$

Where  $X_{total-k}$  is the result of any learner,  $X_{total-kbest}$  is the result of teacher at the same iteration, i. for each group, update the learners' knowledge with the help of teacher's knowledge, along with the knowledge acquired by the learners during the tutorial hours, according to: (where  $hh \neq s$ )

$$(X'_{j,k})_s = (X_{j,k} + \text{Difference}_{Mean_j})_s + \text{rand}(X_{hh} - X_k)_s, \quad \text{If } f(X)_{hh} < f(X)_s$$

$$(X'_{j,k})_s = (X_{j,k} + \text{Difference}_{Mean_j})_s + \text{rand}(X_k - X_{hh})_s, \quad \text{If } f(X)_s < f(X)_{hh}$$

For each group, update the learner's knowledge by utilizing the knowledge of some other learners, as well as by self learning, according to:

$$(X''_{j,k})_s = X'_{j,k,i} + \text{rand}(X'_{j,k} - X'_{j,p})_s + \text{rand}(X_{teacher} - E_f X'_{j,k})_s,$$

$$\text{if } f(X')_k < f(X')_p$$

$$(X''_{j,k})_s = X'_{j,k,i} + \text{rand}(X'_{j,p} - X'_{j,k})_s + \text{rand}(X_{teacher} - E_f X'_{j,k})_s,$$

$$\text{if } f(X')_p < f(X')_k$$

Where  $E_f$  = Exploration Factor = round (1+rand)

**Step 7:** Replace the worst solution of each group with an elite solution

**Step 8:** Eliminate the duplicate solutions; it is necessary to modify the duplicate solutions in order to avoid trapping in the local optima. These solutions are modified by randomly selected.

**Step 9:** combine all groups.

**Step 10:** check the termination criteria; if the termination criterion is not satisfied, repeat the step 3 to last step, otherwise stop the algorithm. In this experiment, the maximum iteration is 300 and the minimum improvement of objective function is  $10^{-6}$ .

#### 4. EXPERIMENTAL RESULTS

We tested the performance of the ITLBO algorithm for fuzzy clustering on three different real-life datasets (Iris, Thyroid and wine datasets) and two artificial datasets;<sup>1</sup> then the ability of algorithm has been compared with FCM [3], SA [15], TS [14], GA [10], ACO [16], PSO [17]. The well-known datasets are described below:

*The Iris dataset:* This dataset contains 3 clusters of 50 objects; where each cluster refers to a type of Iris plant as Setosa, Viginia and Versicolor. The data is four dimensional space(sepal length, sepal width, petal length and petal width). There is no missing attribute value.

*The wine dataset:* this dataset contains 178 data points along with 13 continuous features derived from chemical analysis (e.g. Alcohol, Malicacid, Ash) and it divided in to 3 clusters.

*The Thyroid dataset:* this data set contains 215 samples of patients suffering from three human thyroid disease; each individual was characterized by five features of laboratory tests.

*Artificial dataset:* this is a two dimensional data set consisting of 900 points. The data set has 9 classes. The data set is shown in fig. 1.

```
class1:[-3.3,-0.7]x[0.7,3.3]
class2:[-1.3,1.3]x[0.7,3.3]
class3:[0.7,3.3]x[0.7,3.3]
class4:[-3.3,-0.7]x[-1.3,1.3]
class5:[-1.3,1.3]x[-1.3,1.3]
class6:[0.7,3.3]x[-1.3,1.3]
class7:[-3.3,-0.7]x[-3.3,-0.7]
class8:[-1.3,1.3]x[-3.3,-0.7]
class9:[0.7,3.3]x[-3.3,-0.7]
```

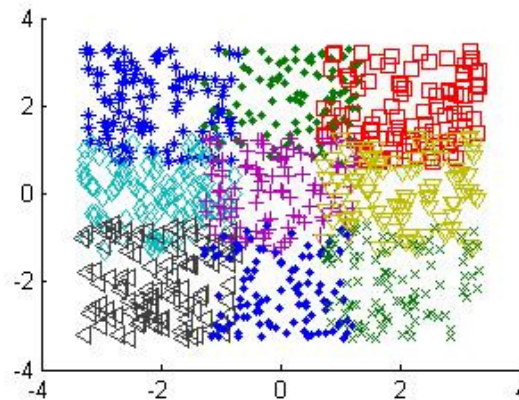


Figure 1. Artificial dataset

The algorithm is implemented in MATLAB Version 12 and the controlling parameters of other algorithms are set the same as their parameters on mentioned references. For maximize compactness inside the clusters, according to FCM algorithm, can calculate the overall summed distances between data and cluster centers, as defined in Eq. 1. Obviously the smaller objective function is, the higher quality of data clustering.

From table 1 we can see that the ITLBO algorithm has achieved the best performance in terms of average compactness criteria.

<sup>1</sup> These datasets are taken from: (<http://www.ics.uci.edu/~mllearn/MLRespository.html>) (<ftp://ftp.ics.uci.edu/pub/machine-learning-databases>)

Table1. Average of compactness index for some of popular algorithms

Algorithms name	Artificial dataset	Iris dataset	Wine dataset	Thyroid dataset
fuzzy c-mean (FCM)	253.32	605.58	17960.84	20642.59
simulated Annealing (SA)	101.82	97.13	16530.53	10114.04
Tabu Search (TS)	102.91	97.86	16785.46	10354.31
Genetic Algorithm (GA)	100.76	125.19	16530.53	10128.82
Ant Colony optimization (ACO)	98.90	97.17	16530.53	10112.13
Particle Swarm Optimization (PSO)	97.83	96.67	16293.00	10109.70
Artificial Bee Colony (ABC)	93.51	78.94	16260.52	10104.03
Improved Teaching Learning Based Optimization (ITLBO)	90.13	77.04	16070.63	10001.18

From the above results, we can obtain that ITLBO algorithm performed better than other mentioned algorithms in terms of intra-cluster distance.

## 5. CONCLUSIONS

This paper proposed an approach for data clustering based on Improved TLBO algorithm; this algorithm model the process of teaching-learning that every individual to learn something from other individuals to improve themselves. TLBO algorithm does not require any algorithm-specific parameter; because of this advantage the application of algorithm is easier than other meta-heuristics algorithm. To evaluate the performance of this algorithm, it is compared with genetic algorithm, simulated annealing, ant colony, tabu search, artificial bee colony and particle swarm optimization. This algorithm was tested on several datasets.

The experimental results over 4 dataset show that the proposed algorithm is efficient. The ITLBO algorithm performed better than other compared algorithms in terms of intera-cluster distance.

## REFERENCES

- [1] Gan, Gujun,Chaoqun Ma & Jianhong Wu, (2007) "Data Clustering: Theory, Algorithms, and Applications", ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA,
- [2] Everitt, B.S., (1993) "Cluster analysis", 3rd edition. New York, Toronto: Halsted Press,
- [3] J.C. Bezdek, (1981) "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York,.
- [4] Brucker, P. (1978) "On the complexity of clustering problems", In M. Beckmenn & H. P. Kunzi (Eds.), Optimisation and operations research. Lecture notes in economics and mathematical systems , Berlin: Springer, Vol. 157, pp. 45–54,.
- [5] El-ghazali Talbi.,(2009) "Metaheuristics : from design to implementation",John Wiley & Sons, Inc.
- [6] Geem, Z. W., Kim, J.H. & Loganathan G.V. (2001) "A new heuristic optimization algorithm: harmony search", Simulation, vol.76, pp 60-70.
- [7] Rashedi, E., Nezamabadi-pour, H. & Saryazdi, S. (2009) "GSA: A gravitational search algorithm", Information Sciences, vol. 179, pp 2232-2248
- [8] Rao, R.V., Savsani, V.J. & Vakharia, D.P. (2011) "Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems", Computer-Aided Design, Vol. 43, No.3, pp 303-315.
- [9] Rao, R.V., Savsani, V.J. and Vakharia, D.P. (2011) "Teaching-learning-based optimization: a novel optimization method for continuous non-linear large scale problems", Inform. Sci., Vol. 183, No.1, pp. 1–15
- [10] Murthy, C. A., & Chowdhury, N. (1996) "In search of optimal clusters using genetic algorithms", Pattern Recognition Letters, vol.17, pp 825–832.

- [11] Mualik, U., & Bandyopadhyay, S. (2002) "Genetic algorithm based clustering technique", Pattern Recognition, vol.33, pp 1455–1465.
- [12] Krishna, K., & Murty (1999) "Genetic K-means Algorithm", IEEE Transactions on Systems Man and Cybernetics B Cybernetics, Vol.29, pp 433–439.
- [13] Sung, C. S., & Jin, H. W. (2000) "A tabu-search-based heuristic for clustering", Pattern Recognition, Vol.33, pp 849–858.
- [14] Al-Sultan, K. S. (1995) "A tabu search approach to the clustering problem" Pattern Recognition, Vol.28, No.9, pp 1443–1451.
- [15] Selim, S. Z., & Al-Sultan, K. (1991) "A simulated annealing algorithm for the clustering problem", Pattern Recognition, Vol.24, No.10, pp 1003–1008.
- [16] Shelokar, P. S., Jayaraman, V. K., & Kulkarni, B. D. (2004) "An ant colony approach for clustering", Analytica Chimica Acta, Vol. 509, pp 187–195.
- [17] Kao, Y.-T., Zahara, E., & Kao, I.-W. (2008) "A hybridized approach to data clustering", Expert Systems with Applications, Vol. 34, No.3, pp 1754–1762.
- [18] Zhang C. , Ouyang D., & Ning J., (2010) "An artificial bee colony approach for clustering", Expert Systems with Applications, Vol.37, pp 4761-4767
- [19] Rao, R.V., Patel V., (2013) "An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems", Scientia Iranica, Vol.20, No.3, pp 710-720

## AUTHORS

Parastou Shahsamandi E. Received the B.Sc.degree in electrical engineering from Isfahan University of technology, in 1997 and the M.Sc. degree in industrial engineering from the Tarbiyat modares University, in 2007. She is currently pursuing her Ph.D. degree at Science and research branch of Islamic Azad University, Tehran, Iran, where her work focused on fuzzy clustering and optimization.



Soheil Sadi-Nezhad is Assistant Professor at Science and research branch of Islamic Azad University Received the M.sc and B.sc in Industrial engineering ( 1989 and 1987) from University of Science & Technology, Tehran, Iran. He got his PhD on Industrial Engineering, In 1999. His research Interests mostly are in Soft computing, Fuzzy Multiple Criteria Decision Making, System thinking, Fuzzy Ranking, FIS and Fuzzy clustering.

