*Journal of **AI** and **D**ata **M**ining*
Published online:

A multi-objective approach to fuzzy clustering using ITLBO algorithm

P. Shahsamandi Esfahani* and A. Saghaei

Department of Industrial engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Received 09 September 2015; Accepted 04 October 2016
**Corresponding author: parastoushahsamandi@yahoo.com (P. Shahsamandi).*

Abstract
Data clustering is one of the most important research areas in data mining and knowledge discovery. Recent research works in this area has shown that the best clustering results can be achieved using multi-objective methods. In other words, assuming more than one criterion as objective functions for clustering data can measurably increase the quality of clustering. In this work, a model with two contradictory objective functions based on maximum data compactness in clusters (the degree of proximity of data) and maximum cluster separation (the degree of remoteness of cluster centers) is proposed. In order to solve this model, the multi-objective improved teaching-learning–based optimization (MOITLBO) algorithm is used. This algorithm is tested on several datasets, and its clusters are compared with the results of some single-objective algorithms. Furthermore, with respect to noise, the comparison of the performance of the proposed model with another multi-objective model shows that it is robust to noisy datasets, and thus it can be efficiently used for multi-objective fuzzy clustering.

Keywords: *Fuzzy Clustering, Cluster Validity Measure, Multi-objective Optimization, Meta-heuristic Algorithms, Improved Teaching-learning–based Optimization.*

1. Introduction

Data clustering is an important topic in data mining and knowledge discovery. The main objective of any clustering technique is to group a set of objects into a number of clusters in such a way that the objects in one cluster are very similar and the objects in different clusters are quite different [1-3]. One measure of similarity for data in K clusters is the distance between the data and their cluster center (e.g. the Euclidean distance in the fuzzy c-means (FCM) algorithm proposed by [4]). In fact, this unsupervised classification produces a $K \times m$ optimum partition matrix $U^*(x)$ of the given dataset X that consists of m data samples $X = \{x_1, x_2, x_3, \ldots, x_m\}$,where each x_i in universe X is a p-dimensional vector of m elements or m features, where $i = 1, 2, \ldots, m$. The partition matrix can be represented as $\{u_{ki}\}$, $k = 1, 2, \ldots, K$. For fuzzy data clustering, $0 \le u_{ki} \le 1$ (where u_{ki} denotes the degree to which object x_i belongs to the kth cluster). Finding the optimum matrix U^* is difficult for practical problems. Hence, the application of advanced optimization techniques is required. As clustering is an NP-hard problem (since the number of data and clusters increases), the application of meta-heuristics is necessary for partitioning data [5]. Meta-heuristic algorithms can be classified into different groups depending on the criteria being considered. Evolutionary algorithms (e.g. genetic algorithms (GAs) and differential evolution) and swarm intelligence algorithms (e.g. particle swarm optimization (PSO), ant colony optimization, and artificial bee colonies) are based upon the population criteria. In addition to these algorithms [6], there are some other algorithms that work based on the principles of different natural phenomena such as harmony search [7], gravitational search [8], and teaching-learning–based optimization (TLBO) [28, 29].

Meta-heuristic algorithms can solve large problems quickly. Moreover, these algorithms can be simply designed and implemented [5, 6]. A large number of such algorithms have been introduced to solve single-objective clustering problems [26, 27, 35, 41], in most of which, the fitness function is based on maximizing the compactness of the data in a cluster. Recent research works have shown that more efficiency may be obtained by using more than one objective function for clustering. Therefore, it is necessary to optimize several cluster validity measures, simultaneously. There are some related studies that have applied multi-objective techniques to data clustering [13-23].

Different meta-heuristic algorithms require similar control parameters such as the population size and the number of generations as well as the algorithm-specific control parameters (e.g. mutation rate and cross-over rate for GA [26] or inertia weight, social, and cognitive parameters for PSO [27]). However, TLBO requires merely common controlling parameters. Thus TLBO can be said to be an algorithm-specific parameterless algorithm [39]. The TLBO algorithm has been designed based upon a teaching-learning process of several students and one teacher in a classroom. The learners are considered to be the population, and the best solution in the population is the teacher. Different subjects that have been suggested to the learners are comparable to different design variables of an optimization problem. TLBO is effective in terms of computational effort and consistency. This algorithm has been improved by introducing more than one teacher for the learners (i.e. increasing the collective knowledge) and some other modifications [24].

In this paper, we use the multi-objective improved TLBO (MOITLBO) [33] for the proposed multi-objective fuzzy clustering model. Two objective functions have been proposed in order to cluster data in a manner better than single objective algorithms. Measure of FCM algorithm (j_m) [4], partition coefficient and an exponential separation (PCAES) validity index [32] have been proposed to minimize the proximity of data in clusters and maximize the differentiation of clusters. The proposed objectives optimize the compactness and separation of the clusters independently.

Sometimes there can be noise in datasets, and as some validity indices are sensitive to noisy data, they cannot determine a good clustering. Therefore, we chose the PCAES validity index that was not sensitive to noise [25] so as to achieve more advantageous clustering results.

Clustering results have been reported for a number of real-life datasets as well as two artificial ones. The performance of this algorithm was compared with those of the single-objective improved TLBO (ITLBO) and FCM clustering algorithms. In order to demonstrate the robustness of the model to noise, it was compared with another multi-objective clustering model.

This paper is organized as what follows. The next section discusses the multi-objective optimization concept, and provides a brief literature review of multi-objective clustering. In Section 3, the proposed multi-objective clustering method and some validity measures are discussed. In Section 4, the MOITLBO algorithm for data clustering is described. Section 5 presents the experimental results of this method on several datasets. Finally, Section 6 concludes the study.

2. Multi-objective clustering optimization

In most practical situations, there are several objectives that must be optimized simultaneously to solve a problem. A multi-objective optimization problem deals with more than one objective function. It is typical that no unique solution exists in multi-objective optimization problems but a set of equally good mathematical solutions can be identified. These solutions are known as non-dominated or Pareto-optimal solutions. The best solution is often subjective, and depends on the needs of the decision-makers (DMs). The multi-objective problem can be categorized into three main methods. If the DMs state some considerations before starting to optimize the problem, the techniques are called priori; if the DMs make some decisions during the process of solving the multi-objective problem, they are called progressive or interactive; and if after solving the problem, some subsets of effective solutions are presented to the DMs to select the most satisfying solution, they are called posteriori.

However, it is not possible to use exact methods to solve real multi-objective problems that have large and complex dimensions. Therefore, approximate methods are often used to solve these problems. Regarding the approximate methods, considerable research works have focused on the multi-objective meta-heuristic algorithms [14]. Multi-objective optimization can be formally stated as follows:

$$\begin{cases} \max \mathrm{f}(\mathrm{x}) \\ s.t\, x \in X = \left\{ \mathrm{x} \in \mathrm{R}^{\mathrm{n}} | \mathrm{g}(\mathrm{x}) \le \mathrm{b},\ \mathrm{x} \ge 0 \right\} \end{cases} \tag{1}$$

where, $f(x)$ represents n conflicting objective functions, $g(x) \le b$ represents m constraints, and x is an n-vector of decision variables, $x \in R^n$. Solution x^* is said to be a Pareto-optimal solution if and only if there does not exist another $x \in X$, Such that $f_i(x) \geq f_i(x^*)$ for all i and $f_i(x) \neq f_i(x^*)$ for at least one *i*.

The use of multi-objective optimization has been gaining popularity since the last few years, and there are some instances in the related literature that have applied multi-objective techniques for data clustering. One of the earliest approaches to multi-objective clustering can be found in [13]. A bi-criterion clustering algorithm has been proposed, in which the objective functions representing homogeneity and separation of the clusters are optimized in a crisp clustering context using a deterministic method. The theoretical advantages of multi-objective clustering have been described in [15] but this paper is limited to an exclusive proof of the concept. A series of related studies on multi-objective clustering can be found in [14, 16, 18-21], in which the authors have developed the first multi-objective clustering algorithm [15]. The Voronoi-initialized evolutionary nearest-neighbor algorithm (VIENNA), which is based upon the Pareto envelope–based selection algorithm II (PESAII) [17], and employs a straightforward encoding of a clustering with a gene for each data item such that its allele value specifies the cluster to which the data item should belong. In [18], the authors have developed a method for selecting solutions from the Pareto front based on a null model, and also determining a better encoding that does not fix the number of clusters. These developments have been incorporated into a new algorithm called multi-objective clustering with automatic K-determination (MOCK). A brief summary of MOCK has been given in [19], where the authors have used a canonical problem to demonstrate that the best solution to some clustering problems is a trade-off between two objectives, and cannot be reached by methods that optimize these objectives individually. MOCK has been further extended in [20] to improve its scalability to large, high-dimensional datasets and data with a large number of clusters.

Most clustering algorithms may not be able to find the global optimal cluster that fits the dataset; these algorithms will stop if they find a locally-optimal partition of the dataset. The algorithms in the family of search-based clustering algorithms can explore the solution space beyond local optimality to find a globally-optimal clustering that fits the dataset [1]. In [22], a metaheuristic search procedure based on two well-known methodologies, Tabu search and Scatter search, has been proposed for multi-objective clustering problems.

A new multi-objective differential evolution-based fuzzy clustering technique has been developed in [23]. The authors have presented a new model that encodes the cluster centers in its vectors and optimizes multiple validity measures simultaneously. For this reason, the Xie-Beni (XB) index [12] and FCM [4] measures (j_m) are considered to be the two objective functions that must be minimized simultaneously. The tendency of the XB index is to increase monotonically when the number of clusters becomes very large and close to the number of patterns. In addition, this index is sensitive to noise (here, the term noise refers to the points that are separated from the other clusters but do not have enough potential to form a distinct cluster) [25]. The main characteristics of the aforementioned multi-objective clustering methods are summarized in table 1.

Table 1. Some main characteristics of related works in multi-objective clustering.

Researcher(s) (Year) – [Ref.]	Multi-objective clustering environment	Optimization methods	Objective functions
Delattre, M., Hansen, P., (1980)-[13]	Crisp	Exact method	Homogeneity and separation based on Single link clustering algorithm and graph-theoretic algorithm
Ferligoj, A., Batagelj, V., (1992)-[15]	Theoretical advantages of multi-criteria clustering		
Caballero, R., Laguna, M., Marti, R., Molina, J., (2006)-[22]	Fuzzy	Approximation method/Tabu search algorithm and Scatter search	A combination of the four functions: Partition diameter, Unadjusted within-cluster dissimilarity, Adjusted within-cluster dissimilarity, and Average within-cluster dissimilarity.
Handl, J., Knowles, J., (2004, 2005, 2006)-[18-21]	Fuzzy	Approximation method/ Multi-objective Clustering with automatic K-determination (MOCK) algorithm	Overall deviation(compactness) and connectivity
saha, I., Maulik, U., Plewczynski, D., (2011)-[23]	Fuzzy	Approximation method/ Differential Evolution algorithm	Compactness and separation- measure of FCM algorithm ($\boldsymbol{J_m}$) and validity index ($\boldsymbol{XB}$)

3. Proposed Multi-objective clustering model

The goal of a partitioned clustering algorithm is to find clusters, that the data that is assigned to the same cluster are similar (i.e. homogenous), while the data that is assigned to different clusters is different (i.e. heterogeneous) [1, 2].

The proposed multi-objective model is based upon two criteria, compactness and separation [21]. Compactness indicates the sameness of data, and separation indicates the dissimilarity among all data. Let $X_{m\times p}$ be the profile data matrix with *m* rows (for a set of *m* objects) and *p* columns (*p*-dimensional), in most cases, the data is in the form of real value vectors. The Euclidean distance is derived from the Minkowski metric, and is a suitable measure of similarity for these datasets [1]. Equation (2) is the Euclidian distance between the two points *x* and *y*.

$$d(x,y) = (\sum_{i=1}^{m} |x_i - y_i|^2)^{\frac{1}{2}} \tag{2}$$

Fuzzy c-means (FCM) is a widely-used technique that allows an object to belong to more than one cluster [4]. It is based on the minimization of the J_m measure, as shown in (3).

$$\text{Min } J_m = \sum_{i=1}^{m}\sum_{k=1}^{K} u_{ki}^{m'} d_{ki}^{2} \tag{3}$$

where m is the number of data objects, K represents the number of clusters, and u is the fuzzy membership matrix. Furthermore, m' ($m' > 1$) is the weighting exponent that controls the fuzziness of the resulting clusters, and d_{ki} is the Euclidian distance from data x_i to the center of the *k*th cluster. The first objective function of the proposed model is J_m, and this criterion is based on increasing the compactness of data in clusters by minimizing the degree of proximity of data [4].

The second objective is based on the partition coefficient and an exponential separation (PCAES) [31], and it seeks to calculate the global cluster variance (i.e. to maximize the separation between one cluster to the other $c-1$ clusters) and the intra-cluster compactness.

$$\text{Max } V_{PCAES} = \sum_{k=1}^{K} PCAES_k = \sum_{i=1}^{m}\sum_{k=1}^{K} u_{ki}^{2} \Big/ u_M - \sum_{k=1}^{K} \exp\left(-\min_{l\neq k}\left\{\frac{\|v_k - v_l\|^2}{\beta_T}\right\}\right) \tag{4}$$

where, u_M and β_T are defined as follow:

$$u_M = \min_{1\le k\le K} \sum_{i=1}^{m} u_{ik}^{2}, \beta_T = \frac{1}{K}\sum_{k=1}^{K} \|v_k - \bar{v}\|^2 \tag{4a}$$

$$\bar{v} = \frac{1}{K}\sum_{k=1}^{K} v_k \tag{4b}$$

A large V_{PCAES} value means that each one of these K clusters is compact, and separated from the other clusters. In addition, under the proposed PCAES objective, a noisy point will not have enough potential to be a cluster; hence, the algorithm will be robust in a noisy environment [32]. Under the proposed multi-objective model, the constraints are as follow:

$$\sum_{k=1}^{K} u_{ki} = 1 \qquad i = 1,2,\ldots,m \tag{5a}$$

$$0 \le u_{ik} \le 1 \qquad i = 1,\ldots,m \quad k = 1,\ldots,K \tag{5b}$$

$$\sum_{i=1}^{m} u_{ik} > 0 \qquad k = 1,2,\ldots,K \tag{5c}$$

As mentioned in [35], the maximum possible number of clusters that one should consider for a dataset is $\sqrt{m}$ $(2 \le k \le \sqrt{m})$. The performance of multi-objective clustering highly depends on the choice of objectives, which should be as contradictory as possible. A further important aspect to be considered when choosing two objective functions is their potential to balance each other’s tendency to increase or decrease the number of clusters. While the objective value associated with compactness is necessarily improved with an increasing number of clusters, the opposite is the case for separation among the centers of clusters. The interaction of the two is crucially important to keep the number of clusters dynamic and to explore interesting areas of the solution space.

3.1. Validity measures in clustering

In general, internal and external cluster validations determine the goodness of the partitions as well as the possibility of better partitioning. In addition, if the number of classes within the data is not known beforehand, a validation index may help to determine the optimal number of classes [11, 25]. Therefore, the role of a validity index is very important. In this research work, to evaluate the performance of the proposed multi-objective clustering algorithm, the partition coefficient (PC) [4, 9], Pakhira-Bandyopadhyay-Maulik (PBM) [34], and Davies-Bouldin (DB) [37] indices were used. Furthermore, the performance of the XB [12] index, as one of the objective functions in the MOITLBO algorithm, was compared with the PCAES index.

3.1.1. PC index

The PC index [4, 9] is based on minimizing the overall content of pairwise fuzzy inter-sections in partition matrix U. This index indicates the average relative amount of membership sharing done between pairs of fuzzy subsets in partition matrix U by combining into a single number the average contents of pairs of fuzzy algebraic products. The index is defined as:

$$V_{PC} = \frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}u_{ki}^{2} \tag{6}$$

A larger V_{PC} indicates a better clustering performance for dataset X.

3.1.2. DB index

This index is a function of the ratio of the sum of the within-cluster scatter to the between-cluster separation [37]. The scatter within the kth cluster may be computed as:

$$S_k = \frac{1}{|C_k|}\sum_{x\in C_k}\{x - v_k\} \tag{7}$$

The Euclidian distance between the centers of the kth and lth clusters is denoted by d_{kl}. This index is then defined as:

$$DB = \frac{1}{k}\sum_{k=1}^{K}R_{k,qt} \tag{7a}$$

where:

$$R_{k,qt} = \max_{l,\, l\neq k}\left\{\frac{S_{kq}+S_{lq}}{d_{kl}}\right\} \tag{7b}$$

Lower values for the DB index indicate better clustering.

3.1.3. PBM index

The PBM index [34] is a composition of three factors, namely $\frac{1}{k}$, $\frac{E}{J_m}$, and D_k.

$$V_{PBMF} = \left(\frac{1}{k}\times\frac{E}{J_m}\times D_k\right)^2 \tag{8}$$

In (8), the first factor indicates the divisibility of a k cluster system that decreases with increasing k. However, in this research work, its value is specified. Equation (8a) is factor E, the sum of the distances of each sample to the geometric center v_0, the centroid of the dataset, and is a measure of the compactness of a k cluster system.

$$E = \sum_{i=1}^{m}u_{ki}\,x_i - v_0 \tag{8a}$$

The third factor, as shown in (8b), is the maximum inter-cluster separation in a k cluster system, which is based on the maximum cluster separations.

$$D_k = \max_{1\le i,k\le K} v_k - v_i \tag{8b}$$

Hence, while the first factor decreases, the other two increase for increasing k.

Based on the above analysis, the maximum value for V_{PBMF} indicates the best clustering performance for dataset X.

3.1.4. XB index

Validity index V_{XB} focuses on compactness and separation [11].

$$V_{XB} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{m}u_{ik}^{m'}\, x_i - v_k{}^{2}}{m\,\min_{k,i} v_k - v_i{}^{2}} \tag{9}$$

As indicated in (9) for V_{XB}, the numerator indicates the compactness of the fuzzy partition, and the denominator indicates the strength of the separation between clusters. A small value for the compactness and a high value for the separation indicate a good partition

4. MOITLBO-based fuzzy clustering

The ITLBO algorithm proposed in [24] is a version of the basic TLBO algorithm with enhanced exploration and exploitation capacities. The TLBO algorithm simulates the teaching-learning process in which every individual tries to learn something from another individual to improve him/herself. The algorithm simulates two fundamental modes of learning, teacher phase and learner phase. A group of learners are considered to be the population of the algorithm, and the results of a learner are the fitness value of the optimization solution, which indicates its quality [28, 29]. In the teacher phase, learning of a learner through a teacher is simulated. The teacher is the most experienced person (the best learner) in the algorithm. During this phase, the teacher conveys knowledge to the learners and makes an effort to increase the mean results of the class. At any iteration of the algorithm, there are n number of learners (population size) and m number of subjects. Let $M_{j,i}$ be the mean result of learners in the jth subject. The difference between the result of the teacher and the mean result of the learners in each subject is given by:

$$Difference_{Mean\ j,i} = r_i\left(X_{j,lbest,i} - T_F\ M_{j,i}\right) \tag{10}$$

where, $X_{j,lbest,i}$ is the result of the teacher (best learner) in subject j, T_F is the teaching factor that decides the value of the mean to be changed, and r_i is a random number in the range $[0, 1]$. Based on the calculated difference, the existing solution is updated in the teacher phase and accepted if it gives a better function value. These accepted values become the input for the learner phase.

The learner phase of the algorithm simulates the learning of the learners through interaction among themselves. One learner interacts randomly with other learners who have more information than itself; hence, in this way, it can increase its knowledge. Randomly, two learners P and Q are selected such that $X'_{total-P,i} \neq X'_{total-Q,i}$ (where these values are the updated values at the end of the teacher phase), the following equations are for the maximization problem:

$$X''_{j,P,i} = X'_{j,P,i} + r_i\left(X'_{j,P,i} - X'_{j,Q,i}\right) \quad , \quad \text{if} \quad X'_{total-P,\,i} > X'_{total-Q,\,i}$$
$$X''_{j,P,i} = X'_{j,P,i} + r_i\left(X'_{j,Q,i} - X'_{j,P,i}\right), \quad \text{if} \quad X'_{total-Q,\,i} > X'_{total-P,\,i} \qquad (11)$$

According to (11), we accept $X''_{j,P,i}$ if it returns a better function value. The algorithm stops according to criteria such as the maximum number of iterations allowed or a minimum change in the objective function. In [24], the algorithm is improved by introducing more than one teacher for the learners to avoid premature convergence, and some other modifications such as the adaptive teaching factor that can automatically tune itself, and self-motivated learning. This algorithm is named ITLBO. In this work, MOITLBO [33] was used to optimize the multi-objective fuzzy clustering. At every iteration of the algorithm, the solutions are maintained in a fixed-size archive. If the solution is dominated by at least one member of the archive, it is not added to the archive; otherwise, the solution is added to the archive. The ε-dominance method is used to refine the solutions in the external archive [33]. In the ε-dominance method, the algorithm uses a grid. The size of each box in the grid is ε, and only one non-dominated solution is placed in each box [10].

Based on these statements, the steps of the MOITLBO algorithm for fuzzy clustering are described, in detail, as follows:

Step 1. Defining objective functions: Define the optimization problem as minimizing the overall deviation of partitioning and maximizing the separation among the centers of each cluster. The first objective is simply computed as the overall summed distances between the data items and their corresponding cluster center (i.e. the objective function of the FCM algorithm). The weighting exponent m' is set to two, which is a common choice for fuzzy clustering. The second objective calculates the global cluster variance and the intra-cluster compactness. However, the TLBO algorithm does not require any algorithm-specific parameter; therefore, setting the control parameter value is not necessary.

Step 2. Initialization: Initialize the external archive and population (N learners). To solve clustering, each candidate solution in the population consists of N $U_{k\times m}$ matrices, where each element of the matrix represents the degree of belonging of the *m*th object to the *k*th cluster. The fuzzy matrix U is generated randomly according to the population size, then the center of each cluster is computed to find the distance between each data and the centroids of the clusters. In the experiment, we set the population size or number of learners to 100.

Step 3. Evaluation: To evaluate the population, rank the evaluated solutions (in ascending order for the minimization problem, and descending order for the maximization problem), then select and assign the best solution as the chief teacher to the first rank.

$(X_{teacher})_1 = f(x^1)$, where $f\left(x^1\right) = f(x)_{best}$

Select the other teachers based on the chief teacher, and rank them:

$f\left(x^s\right) = f(x^1)) \pm \text{rand} \times f(x^1)$,

where s is the number of teachers selected. (If the equality cannot be met, select the $f(x^s)$ closet to the value calculated above.) We selected four teachers in this algorithm.

Step 4. Assignment: Assign the learners to the teachers according to their fitness values, as:

$(X_{teacher})_s = f(x^s)$ **,** where $s = 1, 2, \ldots, T$

For ***l*= 1: (n-s)**
if $f(x^1) \le f(x^l) < f(x^2)$
assign learner $f(x^l)$ ***to teacher 1***
else if $f(x^2) \le f(x^l) < f(x^3)$
assign learner $\boldsymbol{f(x^l)}$ ***to teacher 2***
else if $f(x^3) \le f(x^l) < f(x^4)$
assign learner $f(x^l)$ ***to teacher 3***
else
assign learner $f(x^l)$ ***to teacher 4***
end if
end for

In this work, the number of teachers T is 4.

Step 5. Updating: Calculate the mean result of each group of learners in each subject (i.e. $(M_j)_s$), where $f(x^l)$ is the result of any learner l associated with group s at iteration i, and $f\left(x^s\right)$ is the result of the teacher of the same group during the same iteration i. Evaluate the difference given by (10). For each teacher, the adaptive teaching factor is:

$$(T_F)_i = \left(\frac{f(x^l)}{f(x^s)}\right)_i \quad l=1,2,\ldots,n \quad ; \text{if } f(x^s)\neq 0$$
$$(T_F)_i = 1, \qquad ; \text{if } f(x^s)=0 \tag{12}$$

According to (13), for each group, update the learners' knowledge with the help of a teacher's knowledge or fellow classmates during tutorial hours.

$$\begin{aligned}
&(X'_{j,l})_s = (X_{j,l} + Difference_{Mean\,j})_s \\
&+ rand(X_h - X_l)_s \ , \quad if\ f(x^h) < f(x^l) \\
&(X'_{j,l})_s = (X_{j,l} + Difference_{Mean\,j})_s \\
&,\ \ if\ f(x^l) < f(x^h)\ \ + rand\ (X_l - X_h)_s
\end{aligned} \tag{13}$$

Here, $h \neq l$. According to (14), update the learner's knowledge for each group by utilizing the knowledge of some other learners as well as by self-learning.

$$\begin{aligned}
&(X''_{j,l})_s = X'_{j,l,i} rand(X'_{j,l} - X'_{j,p})_s \\
&+ rand\ (X_{teacher} - E_f\ \ X'_{j,l})_s \\
&if\ f(x'^l) <\ f(x'^p) \\
&(X''_{j,l})_s = X'_{j,l,i} + rand(X'_{j,p} - X'_{j,l}\ )_s \\
&+ rand\ (X_{teacher} - E_f\ \ X'_{j,l})_s\,, \\
&if\ \ \ f(x'^p) < f(x'^l)
\end{aligned} \tag{14}$$

where, E_f = Exploration Factor = round(1 + rand).

Step 6. Elimination: Eliminate duplicate solutions. It is necessary to modify the duplicate solutions to avoid becoming trapped in local optima. These solutions are modified by random selection.

Step 7. Combination: Combine all groups.

Step 8. External archive: Check the archive. If the archive is not full, add the new solution to it; otherwise, select a victim solution to be removed from the archive. The ε-dominance is used to maintain the archive; each dimension of the objective space is divided into segments of width ε. Initialize the grid on the archive. For each box in the grid, if any box dominates the other boxes, remove the dominated box and its related solutions. For the remaining boxes in the grid, if the box contains more than one solution, remove the dominated solution(s) from the box. If the box still contains more than one solution, keep the solution closest to the lower left corner of the box (for the minimization problem) and remove the others.

Step 9. Checking: Check the termination criteria. If neither termination criteria is satisfied, repeat steps 3–8; otherwise, stop the algorithm and output the external archive as a Pareto-optimal set. In this experiment, the maximum number of iterations was 300 and the minimum improvement of the objective function was 10^{-6}.

5. Results

The performance of multi-objective fuzzy clustering based on the MOITLBO algorithm was tested on four different real-life datasets (Iris, Thyroid, Wine, and Red Wine) and four artificial datasets [40]. The artificial datasets are shown in figure 1.

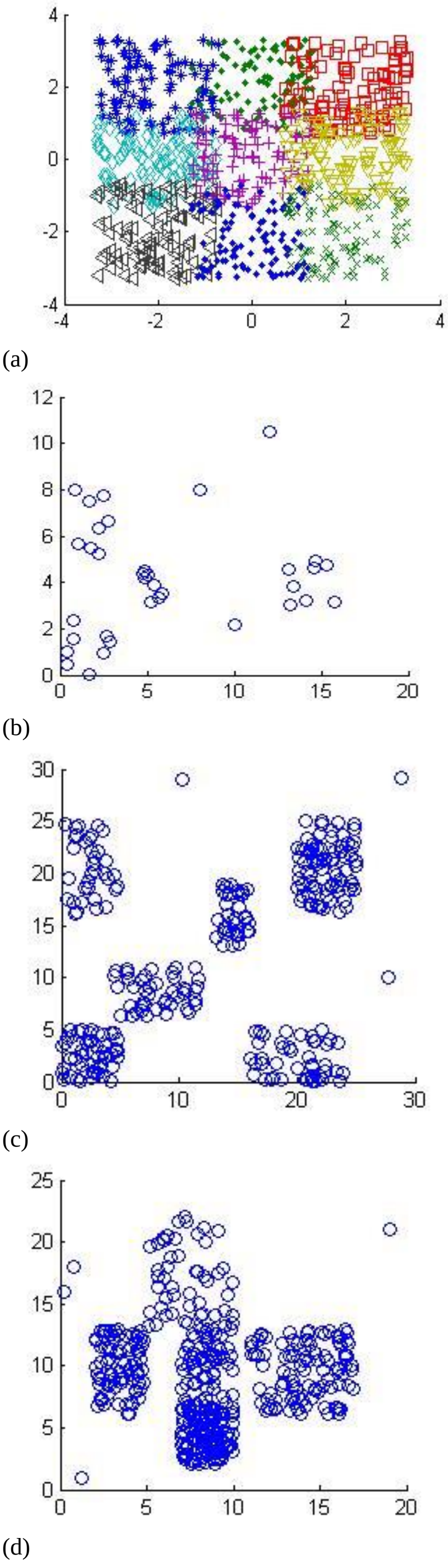

Figure 1. (a) Artificial dataset 1, (b) Artificial dataset 2, (c) Artificial dataset 3, (d) Artificial dataset 4.

The performance of the algorithm was also compared with FCM [4] and ITLBO [24]. The well-known datasets are described below.

Artificial dataset 1: This is a 2D data set consisting of 900 points. The dataset has nine classes.

Artificial dataset 2: There are 35 points in this dataset. It contains some noise and four classes.

Artificial dataset 3: The dataset contains 483 sample points and some random noise. There are five categories in the data.

Artificial dataset 4: This dataset contains 554 points with five classes and some random noise.

Iris dataset: This dataset contains 3 clusters of 150 objects, where each cluster refers to a type of Iris plant, Setosa, Virginica, or Versicolor. The data represents four dimensions (sepal length, sepal width, petal length, and petal width). There are no missing attribute values.

Wine dataset: This dataset contains 178 data points along with 13 continuous features derived from chemical analysis (e.g. Alcohol, Malic Acid, and Ash). It is divided into three clusters.

Thyroid dataset: This dataset contains 215 samples of patients suffering from three human thyroid diseases. Each individual was characterized by five features from laboratory tests.

Red Wine dataset: This dataset is related to red Vinho Verde wine samples from the north of Portugal. The number of instances is 4,898 and the number of attributes is 12. There are some outliers (noise) in this dataset. We refer the reader to [38] for more information about this dataset.

The algorithms were implemented in MATLAB, and the PC, DB, and PBM validity indices were calculated according to their definitions. Several runs of the algorithms were executed. The data in this work was crisp but their memberships were fuzzy. Table 2 shows the comparative results for all the 6 datasets. For the FCM algorithm, the fuzzy exponent m' was set to 2. The population size used for ITLBO algorithms was 100, and did not require any algorithm-specific parameters. Four teachers were used. The objective function in ITLBO I is to minimize j_m, the objective function in ITLBO II is to maximize V_{PCAES}, and the objective functions in MOITLBO I are to minimize j_m and XB. The objective functions in the proposed multi-objective model, namely, MOITLBO II are j_m and PCAES indices.

The low values of the PC and DB indices indicate that the multi-objective clustering performance for all datasets is better than that of single-objective clustering. Parameter m' in the PBM index was set to 2. Larger results of this value indicate a better clustering performance.

Table 2. Cluster index values for some algorithms on different datasets (averaged over 40 runs).

cluster validity indices	Algorithm name	Artificial dataset 1	Artificial dataset 2	Artificial dataset3	Artificial dataset4	Iris dataset	Wine dataset	Thyroid dataset	Red Wine dataset
PC	FCM	0.3210	0.2513	0.3343	0.5012	0.7833	0.5322	0.6510	0.2899
	ITLBOI	0.8864	0.3428	0.7061	0.6229	0.8770	0.7012	0.7943	0.3015
	ITLBO II	0.9032	0.3771	0.7187	0.6953	0.8992	0.7923	0.7734	0.3567
	MOITLBO I	0.9322	0.6105	0.8854	0.7402	0.9114	0.8714	0.8979	0.5188
	MOITLBO II	**0.9767**	**0.7127**	**0.9016**	**0.8916**	**0.9346**	**0.8809**	**0.9106**	**0.7931**
DB	FCM	0.4916	0.7892	0.6669	1.3944	0.9643	1.3944	2.0316	1.0231
	ITLBOI	0.3567	0.6690	0.4660	0.9915	0.8660	1.0975	1.9965	0.8041
	ITLBO II	0.3064	0.6721	0.4732	0.8920	0.8732	0.9962	1.4490	0.7569
	MOITLBO I	0.2031	0.5323	0.3661	0.4318	0.5165	0.7388	1.2338	0.5537
	MOITLBO II	**0.1908**	**0.3206**	**0.1980**	**0.2987**	**0.5089**	**0.7097**	**1.2531**	**0.2438**
PBM	FCM	14.3862	10.9359	54.0640	111.2321	32.4641	204.6350	78.9321	132.7210
	ITLBOI	23.4142	12.8471	40.0558	134.1657	38.3021	231.4027	86.3508	158.8721
	ITLBO II	28.8915	16.0113	42.3727	135.9878	40.9561	309.1602	87.0755	160.4215
	MOITLBO I	33.2092	20.5988	54.6849	167.6579	63.9981	318.0650	98.2698	162.0358
	MOITLBO II	**35.1470**	**26.1943**	**57.5101**	**172.1079**	**67.5482**	**322.8770**	**99.7463**	**181.1477**

A Wilcoxon's rank sum test [31] for independent samples was conducted at the 5% level. This method is a non-parametric statistical hypothesis test that is used when the data does not meet the requirements for a parametric test. It is appropriate for analyzing data from any distribution. Therefore, we used this test to assess whether the difference between the performances of the algorithms could have occurred merely by chance.

It is obvious from table 3 that the median values for MOITLBO II are better than those of the other algorithms. To show that these values are statistically significant, table 4 lists the P-values produced by Wilcoxon's rank sum test for MOITLBO II with respect to the FCM, ITLBOI, and ITLBO II algorithms. All the P-values reported in the table are less than the 5% significance level. As a null hypothesis, it is assumed that there are no significant differences between the median values of MOITLBO II and other algorithms. The alternative hypothesis states that there is a significant difference in the median values of the two groups. The P-values in table 3 indicate the rejection of the null hypothesis. For example, the rank sum test between algorithms MOITLBO II and ITLBO II for the Red Wine dataset provides a P-value of 0.0007, which is very small. This strongly indicates that the better median values of the performance metrics produced by MOITLBO II are statistically significant, and have not occurred by chance. Similar results were obtained for all the other indices and algorithms with respect to MOITLBO II.

Table 3. PBM index values of each algorithm for datasets (median over 40 runs).

Algorithm	Artificial dataset 1	Artificial dataset 2	Artificial dataset 3	Artificial dataset 4	Iris dataset	Wine dataset	Thyroid dataset	Red Wine dataset
FCM	13.8654	10.9774	53.2229	113.7129	32.3081	204.6352	77.8350	132.7231
ITLBOI	23.7221	13.0125	39.5157	134.9650	39.2355	228.1093	86.3491	151.9906
ITLBO II	25.0907	16.1056	42.3488	133.2571	43.6159	303.0045	86.9788	160.4251
MOITLBO I	33.1834	22.3780	56.1294	168.0878	64.0621	316.5639	94.0913	163.1189
MOITLBO II	34.8099	27.1834	58.7861	171.4372	67.4508	321.7338	100.3428	184.0602

Table 4. P-values of Wilcoxon's rank sum test for tested algorithms.

Algorithms	Artificial dataset 1	Artificial dataset 2	Artificial dataset 3	Artificial dataset 4	Iris dataset	Wine dataset	Thyroid dataset	Red Wine dataset
FCM	2.536×10^{-4}	2.789×10^{-4}	1.380×10^{-4}	2.055×10^{-4}	3.700×10^{-4}	1.675×10^{-4}	1.224×10^{-4}	1.532×10^{-4}
ITLBOI	0.0081	0.0038	0.0022	0.0075	0.0023	0.0055	0.0068	0.0022
ITLBO II	0.0023	0.0009	0.0024	0.0070	0.0044	0.0018	0.0038	0.0007
MOITLBO I	0.0015	0.0011	0.0010	0.0035	0.0030	0.0025	0.0014	0.0009

6. Conclusion

This paper proposed a multi-objective approach for fuzzy clustering based on two objective functions, namely the J_m and V_{PCAES} indices. An important aspect to be considered when choosing the two objective functions, is their potential to balance each other's tendency to increase or decrease the number of clusters. This interaction between the two objectives is crucially important to keep the number of clusters dynamic and explore interesting areas of the solution space. In order to optimize the model, the MOITLBO algorithm was applied. This algorithm modeled the process of teaching-learning, where every individual learned something from the other individuals in order to improve themselves. In clustering, the role of validity indices are very important, and these indices help determining the validity of the clustering. We used the PC, PBM, and DB indices to evaluate the performance of the clustering algorithms. To evaluate the clustering performance of the MOITLBO algorithm, a statistical test was performed to compare it with some single-objective algorithms, FCM and ITLBO. In addition, the performance of this model with respect to noise was compared using MOITLBO based on the two objectives, namely J_m and XB indices. The experimental results showed that the proposed MOITLBO algorithm based on the J_m and $PCAES$ indices achieved the best performance.

Although we introduced a multi-objective clustering model that can be used to generate research insights, there are some limitations that need to be addressed. These limitations clearly point to the potential future developments.

The proposed model in this study was tested on some real-life and artificial datasets; if the technique is applied on more big datasets and real life domains, the results may or may not be as valid.

Since the Euclidean measure was used as distance metric to measure similarity and dissimilarity between clusters, future research works could involve examination and evaluation using different distance measures (e.g. Mahalanobis distance measure) to determine the performance of the clustering model introduced in this paper.

Comparing and evaluating the proposed multi-objective clustering based on MOITLBO to other multi-objective meta-heuristic approaches will help further evaluation of the robustness of the model.

References

[1] Gan, G., Ma, C., & Wu, J. (2007). Data Clustering: Theory, Algorithms, and Applications. ASA-SIAM Series on Statistics and Applied Probability, Philadelphia, Alexandria: SIAM.

[2] Everitt, B. S. (1993) Cluster Analysis. (3rd ed.), New York, Toronto: Halsted Press.

[3] Mosavi, A. (2014). Data mining for decision making in engineering optimal design, Journal of AI and Data Mining, vol. 2, no. 1, pp.7-14.

[4] Bezdek, J. C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press.

[5] 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.

[6] EL-Ghazali, T. (2009). Metaheuristics: From Design to Implementation. (1st ed.), John Wiley and Sons.

[7] Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A New Heuristic Optimization Algorithm: Harmony Search. Simulation, vol. 76, pp. 60–70.

[8] Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). GSA: A Gravitational Search Algorithm. Information Sciences, vol. 179, pp.2232–2248.

[9] Jardin, N., & Sibson, R. (1971), Mathematical Taxonomy. (1st ed.), John Wiley and Sons.

[10] Deb, K., Mohan, M., & Mishra, S. (2005). Evaluating the Epsilon-Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of Pareto-optimal Solutions. Evolutionary Computations, vol. 13, no. 4, pp.501–525.

[11] Rezaee, M. R., Lelieveldt, B. P. F., & Reiber, J. H. C. (1998). A New Cluster Validity Index for the Fuzzy c-mean. Pattern Recognition Letters, vol. 19, pp. 237–246.

[12] Xie, X. L., & Beni, G. (1991). A Validity Measure for Fuzzy Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp.841–847.

[13] Delattre, M., & Hansen, P. (1980). Bicriterion Cluster Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 2, no. 4, pp. 277–291.

[14] Handl, J., & Knowles, J. (2007). An Evolutionary Approach to Multiobjective Clustering. IEEE Transactions on Evolutionary Computation, vol. 11, pp. 56–76.

[15] Ferligoj, A., & Batagelj, V. (1992). Direct Multicriterion Clustering. Journal of Classification, vol. 9, pp. 43–61.

[16] Handl, J., & Knowles, J. (2004). Evolutionary Multiobjective Clustering. In Proceedings 8th International Conference on Parallel Problem Solving from Nature, pp. 1081–1091.

[17] Corne, D. W., Jerram, N. R., Knowles, J. D., & Oates M. J. (2001). PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization. In Proceedings of the Genetic and Evolutionary Computation Conference, pp. 283–290.

[18] Handl, J., & Knowles, J. (2004). Multi-objective Clustering with Automatic Determination of the Number of Clusters. Technical Report TR-COMPSYSBIO, Institute of Science and Technology, University of Manchester.

[19] Handl, J., & Knowles, J. (2005). Exploiting the Tradeoff—The Benefits of Multiple Objectives in Data Clustering. In Proceedings 3rd International Conference on Evolutionary Multi-Criterion Optimization, pp. 547–560.

[20] Handl, J., & Knowles, J. (2005). Improvements to the Scalability of Multiobjective Clustering. In Proceedings 2005 IEEE Congress on Evolutionary Computation, vol. 3, pp. 2372–2379.

[21] Handl, J., & Knowles, J. (2006). Multiobjective Clustering and Cluster Validation. Computational Intelligence, vol. 16, pp.21–47.

[22] Caballero, R., Laguna, M., Marti, R., & Molina, J. (2006). Multiobjective clustering with metaheuristic optimization technology. Technical Report, Leeds School of Business at the University of Colorado at Boulder.

[23] Saha, I., Maulik, U., & Plewczynski, D. (2011). A New Multi-objective Technique for Differential Fuzzy Clustering. Applied Soft Computing, vol. 11, pp. 2765–2776.

[24] Rao, R. V., & Patel V. (2013). An Improved Teaching-Learning-Based Optimization Algorithm for Solving Unconstrained Optimization Problems. Scientia Iranica, vol. 20, no. 3, 710–720.

[25] Wang, W., & Zhang, Y. (2007). On Fuzzy Cluster Validity Indices. Fuzzy Sets and Systems, vol. 158, pp. 2095–2117.

[26] Mualik, U., & Bandyopadhyay, S. (2002). Genetic Algorithm Based Clustering Technique. Pattern Recognition, vol. 33, pp.1455–1465.

[27] 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.

[28] 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.

[29] Rao, R. V., Savsani, V. J., & Vakharia, D. P. (2012). Teaching-learning-based optimization: A Novel Optimization Method for Continuous Non-Linear Large Scale Problems. Information Sciences, vol. 183, no. 1, pp.1–15.

[30] Van Veldhuizen, D. A. (1999). Multi-objective Evolutionary Algorithms: Classifications. Analysis and New Innovations. Evolutionary Computation, vol. 8, no. 2, pp. 125–147.

[31] Hollander, M., & Wolfe, D. A. (2014), Nonparametric Statistical Methods (3rd ed.), John Wiley and Sons.

[32] Wu, K. L., & Yang, M. S. (2005). A Cluster Validity Index for Fuzzy Clustering. Pattern Recognition Letters, vol. 26, pp.1275–1291.

[33] Rao, R. V., & Patel, V. (2014). A Multi-Objective Improved Teaching Learning Based Optimization Algorithm for Unconstrained and Constrained Optimization Problems. International Journal of Industrial Engineering Computation, vol. 5, pp. 1–22.

[34] Pakhira, M. K., bandyopadhyay, S., & Maulik, U. (2004). Validity Index for Crisp and Fuzzy Clusters. Pattern Recognition, vol. 37, pp. 481–501.

[35] Murty, M. R., et al. (2014). Automatic Clustering Using Teaching Learning Based Optimization. Applied Mathematics, vol. 5, pp. 1202–1211.

[36] Pal, N. R., & Bezdek, J. C. (1995), On Cluster Validity for the Fuzzy c-means Model, IEEE Transactions on Fuzzy Systems, vol. 3, no. 3, pp. 370–379.

[37] Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, pp. 224–227.

[38] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling Wine Preferences by Data Mining from Physicochemical Properties, In Decision Support Systems, vol. 47, no. 4, pp. 547–553.

[39] Rao, R. V., & Patel, V. K. (2012). An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems, International Journal of Industrial Engineering Computations, vol. 3, pp. 535–560.

[40] These datasets, Available: (http://www.ics.uci.edu/~mlearn/MLResponsitory.html), (ftp://ftp.ics.edu/pub/machine-learning-databases).

[41] Gaffari, A., & Nobahar, S. (2015). FDMG, Fault detection method by using genetic algorithmin clustered wireless sensor networks, Journal of AI and Data Mining, vol. 3, no. 1, pp. 47–57.

نشریه هوش مصنوعی و داده کاوی

خوشه‌بندی چندهدفه فازی با استفاده از الگوریتم بهینه‌سازی یادگیری-یاددهی بهبود یافته

پرستو شاه‌سمندی اصفهانی* و عباس سقایی

گروه مهندسی صنایع، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران.

ارسال ۲۰۱۵/۰۹/۰۹؛ پذیرش ۲۰۱۶/۱۰/۰۴

چکیده:

خوشه‌بندی داده یکی از مهم‌ترین حوزه‌های تحقیقاتی در داده‌کاوی و کشف دانش می‌باشد. تحقیقات اخیر در این زمینه نشان می‌دهد در نظر گرفتن بیش از یک معیار به عنوان تابع هدف برای خوشه‌بندی داده منجر به نتایج بهتری خواهد شد و کیفیت خوشه‌بندی را افزایش می‌دهد. در این مطالعه مدلی با دو تابع هدف بر اساس بیشینه کردن فشردگی داده ها درون هر دسته و جدایی دسته ها از یکدیگر برای خوشه‌بندی فازی پیشنهاد گردیده است. مدل پیشنهادی با استفاده از الگوریتم بهینه‌سازی چندهدفه یادگیری-یاددهی بهبودیافته حل گردید و بر روی مجموعه داده‌های مختلف، آزمایش شد. همچنین مجموعه داده‌هایی با داده‌های پرت برای نشان دادن مقاوم بودن الگوریتم در نظر گرفته شد. با توجه به برخی از معیارهای اعتبار خوشه-بندی، خروجی حاصل از خوشه‌بندی فازی الگوریتم پیشنهادی با الگوریتم‌های دیگر مقایسه گردید. نتایج نشان دهنده عملکرد مؤثر این الگوریتم چندهدفه برای خوشه‌بندی فازی می‌باشد.

کلمات کلیدی: خوشه‌بندی فازی، معیارهای اعتبار خوشه‌بندی، بهینه‌سازی چندهدفه، الگوریتم‌های فرا ابتکاری، الگوریتم یادگیری-یاددهای بهبودیافته.