Available online at http://sanad.iau.ir/journal/ijim/
Int. J. Industrial Mathematics (ISSN 2008-5621)
Vol.16, No.1, 2024, Article ID IJIM-1629, 26 pages
Research Article

Science and Research Branch (IAU)

Fuzzy clustering ensemble by optimization approach based on clustering reliability

B. Minaei-Bidgoli* [1], A. Bagherinia[1], M. Hossinzadeh [2], H. Parvin[3]

[1] Department of Computer Engineering, Science and Research Branch, Islamic Azad university, Tehran, Iran,

[2] Iran University of Medical Sciences, Tehran, Iran,
Computer Science, University of Human Development, Sulaimaniyah, Iraq,

[3] Department of Engineering, Mamasani Branch, Islamic Azad university, Mamasani, Iran.

Submission Date: 2019/07/26, Revised Date: 2019/08/25, Date of Acceptance: 2019/10/17

Abstract

Although some ensemble clustering approaches have been developed in recent years to improve the quality of the clustering, but lack of a median fuzzy partition-based consensus function that considers more participate reliable fuzzy clustering, remains unsolved problem. In this paper, we convert the median fuzzy partition problem into an optimization problem based on the reliability-based co-association matrix that minimizes distances between co-association matrix of final clustering and co-association matrix of base-clustering in the ensemble. The optimization problem is a constrained nonlinear objective function and we solve it by sparse sequential quadratic programming (SSQP). Experimental results reveal the effectiveness of the proposed approach rather than the state-of-the-art methods in the quality-terms on various standard datasets.

Keywords: Fuzzy Clustering Ensemble, Fuzzy Clustering Reliability, Median Partition, Sequential Quadratic Programming.

* Corresponding author: Email: b_minaei@iust.ac.ir

1. Introduction

Clustering is a process of categorization the unlabeled data-objects into different groups, called clusters, such that data-objects within a cluster are more similar to each other than they are to those belonging to different clusters. Clustering has been applied for exploration, analysis and pattern discovery in an unsupervised manner in such domains as machine learning, image segmentation, document retrieval, marketing research, bioinformatics, data mining, etc.

Based on the relationship of each data-object to the clusters, the clustering algorithms can also be categorized into fuzzy and hard (crisp) clustering algorithms. In fuzzy clustering, the data-objects can belong to more than one cluster, and a membership degree associated with each of the data-objects which indicate the degree to which the data-objects belong to the different clusters. But some data-objects are inherently fuzzy, (i.e. the ones that will not be definitively assigned to a cluster), and are doubtful. For example, in satellite image segmentation application, a pixel corresponds to an area of the land space, may belongs to different types of land cover. In hard clustering a data-object certainly belongs to one cluster. Crisp clustering is a special case of fuzzy clustering, in which the membership degree of a data-object belonging to a cluster equals one and its membership degree belonging to the other clusters is zero. The foundation of the fuzzy clustering analysis is the basic FCM clustering algorithm which proposed by Don and completed by Bezdek [1] . In due time, the famous fuzzy clustering algorithms have been derived from FCM in order to adapt it to different datasets that contain clusters of various shapes, sizes or densities. We can mention Gustafson-Kessel algorithm (GK) [2], Gath-Geva algorithm (GG) [3], Kernel-based fuzzy clustering (KFCM) [4] algorithms as some examples.

In the data clustering context, various clustering algorithms have emerged, each uses a different similarity criterion. Therefore, they have different objective functions. Different algorithms or the same algorithm with different parameters or initialization, will produce different clustering results for the same dataset. In specific conditions, some of these algorithms might outperform others. For example, some algorithms have high computational complexity, some others have good accuracy rate, and the others fit the datasets with special characteristics (e.g., k-means fits the datasets with circular-shape clusters). That is, all these methods are heavily dependent on dataset. Hence, an alternative solution for managing all the objectives regarding the clustering that might be contradictory, is to combine some of these algorithms. This idea is named clustering aggregation and is called cluster ensemble in many scientific context [5], which recently has become popular in scientific community [6], [7], [8], [9], [10], [11], [11], [12], [13]. Clustering ensembles generally outperform the single clustering in several aspects, such as robustness, novelty, quality enhancement, knowledge reusability, multi view clustering, stability, parallel/distributed data processing [14] and mining heterogeneous data.

Clustering ensemble approach consists of two phases: In the first phase a set of base clusterings is generated (Phase 1 in Fig. 1), which are as diverse as possible, and then a suitable consensus function will be applied to combine different base clusterings into a single final clustering in the second phase (Phase 2 in Fig. 1; (This research does not address to phase 1, only focuses on phase 2). The goal of the second phase of the ensemble clustering is to reach the final clustering. This achieved through a consensus function. Since the clustering problem is unsupervised, producing the "final clustering" with maximum similarity to all base clusterings is very difficult and an NP problem [13]. For this purpose, various consensus functions are proposed, each with a specific approach and different information from the base clustering obtained from phase one, and sometimes by considering the initial characteristics of the data. The existing cluster ensemble approaches can be classified into following categories: a) intermediate space clustering ensemble methods [9], [15], b) co-association matrix based clustering ensemble methods [16]–[19], c) hyper-graph based clustering ensemble methods [5], [7], [19] d) expectation maximization clustering ensemble methods [13] and e) mathematical modeling clustering ensemble methods (median partition) [20], [21], f) voting- based approach [22]–[24], g) Quadratic Mutual Information approach [25]. Median partition based clustering ensembles, which selects a single candidate clustering solution from a set of candidate clustering solutions is so far considered as the best approach [26].

Despite the more generality of fuzzy clustering compared to crisp clustering, researches in the fuzzy clustering ensemble are still in their initial stages and there exist relatively few approaches for this field. The quality of the base clustering highly affects the consensus process in ensemble clustering. The consensus results may be badly affected by low-quality or even noised base clusterings. To deal with low-quality base clusterings, some researchers investigated the quality-evaluation and weighting of the base clusterings to improve the consensus functions quality [27]–[29]. However, these approaches are developed based on an implicit assumption that all of the clusterings in the ensemble have the same reliability. Due to the noise and inherent complexity of real-world datasets, the different clusterings in the ensemble may have different reliability. It is worth mentioning that among the available studies no researcher has considered the role of reliability of fuzzy clusterings in the ensemble. Therefore, without the need to access the data features or relying on specific assumptions made on data distribution, the key questions here are 1) how to measure the reliability of fuzzy clusterings and 2) how to weight the clusterings based on their measured reliabilities to enhance the accuracy and robustness of the consensus clustering. In summary, measuring clustering reliability in the fuzzy clustering ensemble poses a challenging task (**problem1**).

In the median partition-based consensus function approach, the final clustering is obtained by finding a clustering which has maximizes (minimize) the similarity (dissimilarity) from (with) all clusterings in the clustering ensemble. Although a great number of clustering ensemble methods have been proposed over the past years, there are relatively few researches in handling fuzzy clustering ensemble based on median partition (It is worth noting that partition and clustering are same) approach and none of them investigated ensemble method based on data-objects co-occurrence and median partition simultaneously (**problem2**).

This study devoted towards the development of a new fuzzy clustering ensemble framework based on ensemble-driven clustering reliability calculation and local weighting strategy to address the aforementioned challenges. The overall flowchart of this approach is illustrated in Fig. 1. The advantage of the fuzzy clustering quality is posed into a locally weighted scheme to enhance the consensus quality. Here, the reliability of each fuzzy clustering is computed according to information theory [5] based on the pair-wise fuzzy cluster similarity. In particular, for a given fuzzy clustering, its reliability is computed by defining a new metric. This measure named reliability driven fuzzy clustering indicator (*RDFCI*). Here, the point is that the crowd of diverse clusterings in the ensemble can provide an effective indication for evaluating every single clustering. By assessing and weighting the clusterings in the ensemble through the *RDFCI* measure, the concept of weighted co-association matrix whose weights are calculated based on reliability, which incorporates local similarity between fuzzy clusters into the conventional co-association (Co) matrix and is treated as a summary for the ensemble of diverse fuzzy clusterings (**dealing with problem 1**). Finally, finding the final consensus clustering was modeled as an optimization problem, and this problem is solved by sequential quadratic Programming (**dealing with problem 2**).

As a summary, in previous work [30], we proposed a framework for selecting fuzzy base-clusterings based on diversity and quality of fuzzy base-clusterings which applied hierarchical agglomerative clustering as consensus function over the co-association matrix to obtain final clustering. Then we decide to propose an approach that **1)** consider the reliability of each fuzzy clustering in constructing fuzzy co-association matrix, in addition **2)** a novel median based consensus function that is applied over the co-association matrix.

The contributions of this paper are as follows:

- A fuzzy reliability-driven clustering indicator is proposed to measure the reliability and to weight the fuzzy clusterings in the ensemble, which provides an indication of reliability at the clustering-level with a contribution to the local weighting plan.
- A consensus function based on the object's co-occurrence and median partition is proposed to obtain final fuzzy clustering from base clusterings concerning clustering reliability.
- A median fuzzy partition problem is defined formally, which seeks to estimate an optimum consensus fuzzy clustering.
- The proposed median fuzzy partition problem is solved by SSQP solver to obtain the final fuzzy clustering.
- experimental evaluation is performed on a set of datasets indicate that this proposed fuzzy clustering ensemble approach outperforms the state-of-the-art approaches in terms of clustering quality and robustness.

The rest of the paper is organized as follows: related work is presented in Sec. (2). The background knowledge about entropy and ensemble clustering is introduced in Sec. (3). The proposed fuzzy clustering ensemble approach is described in Sec. (4). The experimental results are reported in Sec. (5) and the conclusion and future work is presented in Sec. (6).

2. Related work

A lot of work has been performed in the field of crisp cluster ensemble. Here we can consider the studies related to fuzzy cluster ensemble, among which the following are briefed:

Berikov presented the probabilistic model for the fuzzy Clustering ensemble based on the weighted co-association matrix [31]. in this model, each of the base clustering is created by different clustering algorithms. Each algorithm performs a certain number of times on each data set (i.e., r times). The variance of the Hellinger distance [32] between each data-object pair in r times is considered as the weight of each base clustering in the calculation of the reverse co-association (the matrix is based on the distance of the data-object pair rather than the similarity of the data-object pair); in this algorithm, the variance of the distance between the data-object pair is considered as a consistency criterion. Then, the final clustering is obtained by applying the traditional hierarchy method al on the resulting matrix.

CSPA, MCLA proposed by Strel and Ghosh [5] and HBGF proposed by Fred and Brodly [7] approaches were later extended by Punera and Ghosh, to allow fuzzy base clusterings on the clustering ensemble, showing that the addition of information of fuzzy clusters on the ensemble is useful; the proposed models were the soft version of CSPA, of MCLA, and HBGF and named sCSPA, of sMCLA, and sHBGF, respectively. [7].

Another related research introduced by Saha et al. is SVMeFC [33]. In this method some fuzzy clustering methods like MoDEFC, GAFPSC, FCIDE, GAFC, FCM are applied to the data set. Some of the derived clustering are chosen to train SVM and the remaining objects of each clustering are re-labeled by SVM. At the end of this procedure, by applying CSPA to obtained clusters from SVM, final clusters are produced. Support vector machines are very powerful algorithms for data sorting and separating, especially when combined with other methods of machine learning. This procedure best fits the cases where excessive precision is required, as long as we choose mapping functions properly. But it is time-consuming because of high computational complexity and also consumes a lot of memory.

In [34] Alizadeh et al. converted the fuzzy ensemble clustering problem to a 0-1 bit string problem. Their proposed model consists of a constrained nonlinear objective function, named fuzzy string objective function (FSOF). FSOF simultaneously maximizes the agreement and minimizes the disagreement between the ensemble members. They solved this nonlinear model using genetic algorithm by applying two modified crossovers and a modified mutation operator. Based on these operators two consensus function that named FSCEOGA1 and FSCEOGA2 were proposed. It is worthy to be mentioned that in this method base clustering must be crisp.

A voting based method proposed by Sevillano et al. [23] to obtain consensus clustering from base fuzzy clustering in the ensemble. This method includes two procedures, 1) Disambiguation and 2) Voting. In disambiguation phase of clusters, the re-labeling problem is performed using the Hungarian algorithm [35] with $O(K^3)$ time complexity, k represents the number of clusters in each clustering. The final consensus clustering is obtained through the voting procedure. Two confidence-based voting methods named sum voting rule and product voting rule [36] and also two positional-based voting methods named Borda voting rule [37] and Copeland voting rule [38] are presented which time complexity of these four algorithms is $O(MK\beta \log(K))$ such that K represents the number of cluster in each clustering, M shows the number of data-objects and β indicates the number of base clusterings. Based on the combination of re-labeling and voting being direct or repetitive, there exist eight different consensus functions.

A heterogeneous clustering ensemble is proposed by Arlyd and Anna in [39] to increase the stability of fuzzy cluster analysis. First, they applied basic fuzzy clustering algorithms like FCM, GG, GK and KFCM and then applied the FCM algorithm to the co-association matrix and the final clustering is yield. In this ensemble the weights of participation of all clusters in the co-association matrix are equal.

Szabo and Nunes de Castro (2017) offer method for fuzzy clustering ensemble based on Particle Swarm optimization which can be applied to fuzzy and crisp clusters [40]. Initial clusters in this method are created using Particle Swarm Clustering (PSC) algorithm and through parameter change. Then β' among β initial clusterings ($\beta' < \beta$) are selected through the pruning process. In order to perform the pruning process, first the fitness of base clusterings are measured using one of internal cluster validity indices like Ball-Hall [41], Calinski-Harabasz [42], Dunn index [43], Silhouette index [44] or Xie-Beni [45] and then the elite clustering are chosen using one of the genetic selection mechanism like tournament or roulette wheel. Finally, the PSC algorithm is applied as consensus function in order to perform the final clustering. In this method, each particle represents a cluster, despite the other PSO-based methods which each particle represents a clustering.

In [17] Parvin et al. in for handling imbalanced clustering propose a weighted locally adaptive clustering algorithm (FWLAC). Because the performance of FWLAC algorithm is dependent to tune two parameters, they propose an elite clustering ensemble two mange to manage these parameters. Their proposed elite procedure first converts fuzzy clusters into crisp clustering and consider each cluster as a clustering, finally clustering NMI was used to assess each cluster.

3. PRELIMINARIES

Before explaining the proposed approach, the general formulation of the data, fuzzy clustering ensemble and entropy definitions should be introduced as follows:

Definition 1. A data-object is a multi-tuple $\left(d_i^1, d_i^2, \dots d_i^N\right)$ presented as $\overrightarrow{d_i}$, where d_i^j is the *j*-th feature from *i*-th data, $d_:^j$ is the *j*-th feature from whole data. N is the number of features, $N = |d_1^:|$ and M is the number of the data-objects‘ $M = \left|d_:^i\right|$.

Definition 2. Fuzzy clustering of data set x is a two-dimensional matrix with size $M * c$, where $M = |d_:^1|$ and c is the number of clusters, presented as $\pi(d)$ so that:

$$\forall \text{j} \in \{1, \dots, c\}, i: \pi\left(\overrightarrow{d_i}\right)^j \in [0,1] \qquad (1)$$

where $\forall$ i: $\sum_{j=1}^{c} \pi\left(\vec{d}_i\right)^j = 1$ (2)

where $\pi\left(\vec{d}_i\right)^j$ is the membership degree of i-th data-object belong to *j*-th cluster.

Definition 3. A clustering ensemble consists of β base-clusterings defined as:

$$\Pi = \left\{\pi^1, \dots, \pi^\beta\right\} \qquad (3)$$

where $\pi^m = \left\{C_1^m, \dots, C_{n^m}^m\right\}$ (4)

Where, π^m is the *m*-th base clustering in Π, C_i^m is the *i*-th cluster in clustering π^m and n^m is the number of clusters in π^m.

To sum up, the set of all clusters in the ensemble is presented as

$$C = \{C_1^1, \dots, C_{n^\beta}^\beta\} \qquad (5)$$

where C_i^j is the i-th cluster of clustering π^j , thus the number of all clusters in the base clusterings (c) is computed as:

$$c = n^1 + \dots + n^\beta \qquad (6)$$

Example 1. Three fuzzy clustering π^1, π^2 and π^3 ($\beta = 3$) on dataset x with 6 data-objects ($M = 6$) are shown in Table 1.

Table 1. Three fuzzy clustering π^1 , π^2 and π^3.

	c_1	c_2	c_3		c_1	c_2	c		c_1	c_2	c
x_1	0.1	0.7	0.2	x_1	0.6	0.3	0	x_1	0.7	0.2	0
x_2	0.0	0.8	0.2	x_2	0.8	0.2	0	x_2	0.9	0.1	0
x_3	0.1	0.4	0.5	x_3	0.5	0.5	0	x_3	0.9	0.0	0
x_4	0.0	0.2	0.8	x_4	0.2	0.7	0	x_4	0.2	0.6	0
x_5	0.8	0.1	0.1	x_5	0.0	0.5	0	x_5	0.1	0.9	0
x_6	0.7	0.1	0.2	x_6	0.1	0.2	0	x_6	0.0	0.2	0

4. Proposed approach

In this paper, a new fuzzy clustering ensemble approach based on ensemble-driven clustering reliability calculation and local weighting strategy is proposed. As can be seen in Fig. 1 the proposed approach lies in phase 2 of fuzzy clustering ensemble and decomposed it in to 4 steps. The steps of the proposed approach are shown in Fig. 2 (after phase 1). In the first step, the similarity of each fuzzy cluster in relation to other clusters is computed. In the second step, the reliability of each clustering is calculated in Sec. (4.2), In the third step the fuzzy reliability-based co-association matrix is computed in Sec. (4.3) and in step 4, the consensus clustering is obtained as described in Sec. (4.4).

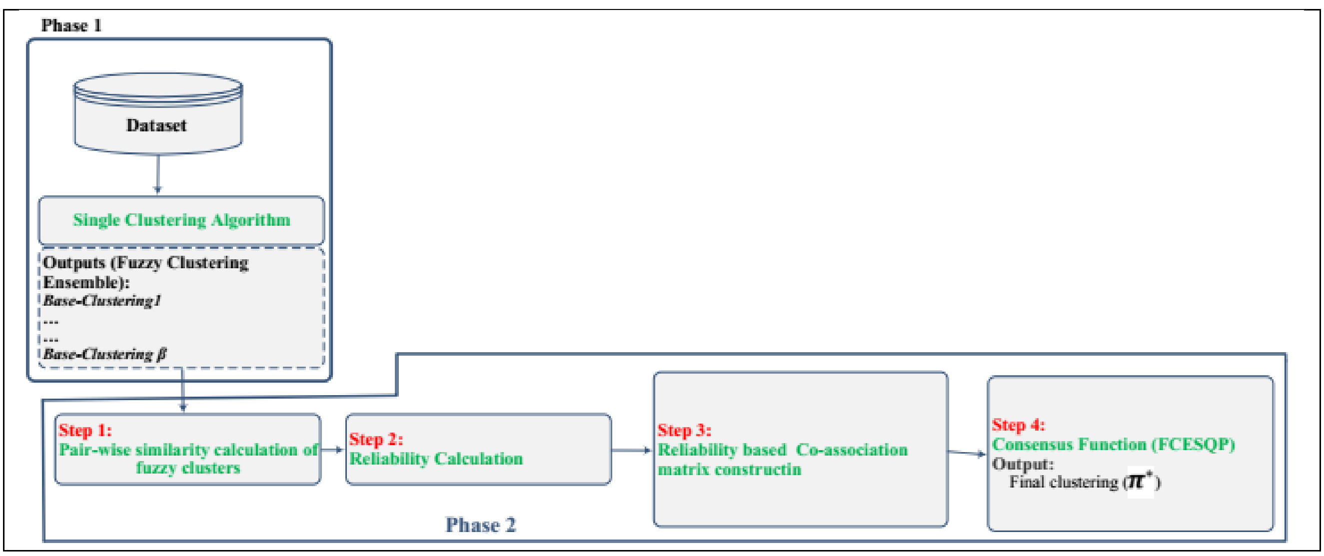

Fig. 1. Proposed approach steps

4.1 Computing the similarity between fuzzy cluster

The first step in Fig. 2 is to compute the similarity of each fuzzy cluster in relation to other fuzzy clusters within different clusterings; i.e., the probability of agreement between two clusters in different clusterings which is obtained according to Definition 7. We measure it by extending the 'similarity between fuzzy sets' that proposed by Gao Zheng [46]. Then this extended criterion measures similarity relationship between two fuzzy clusters.

Definition 5. The similarity of cluster C_i^{ki} (cluster $C_i \in \pi^{ki}$) in relation to cluster C_j^{kj} (cluster $C_j \in \pi^{kj}$), where $i \neq j$ is computed through Eq. (9):

$$sim\left(C_i^{ki}, C_j^{kj}\right) = \sum_{t=1}^{M} \frac{\min\left(1-\pi^{ki}(\overrightarrow{x_t})^i+\pi^{kj}(\overrightarrow{x_t})^j, 1+\pi^{ki}(\overrightarrow{x_t})^i-\pi^{kj}(\overrightarrow{x_t})^j\right)}{\max\left(1-\pi^{ki}(\overrightarrow{x_t})^i+\pi^{kj}(\overrightarrow{x_t})^j, 1+\pi^{ki}(\overrightarrow{x_t})^i-\pi^{kj}(\overrightarrow{x_t})^j\right)} * \min\left(\pi^{ki}(\overrightarrow{x_t})^i, \pi^{kj}(\overrightarrow{x_t})^j\right) \quad (7)$$

The value of the cluster acceptability (sim) computed from (7) is in the range [0, 1].

Example 2 (Continuation of example 1). The sim values of fuzzy clusterings related to Example 1 in Table 1 is shown in Table 2.

Table 2. The values of sim corresponding to fuzzy clusterings presented in Tables 1,2 and 3.

	c_1	**c_2**	**c_3**	**c_4**	**c_5**	**c_6**	**c_7**	**c_8**	**c_9**
c_1	-	0.24	0.42	0.10	0.45	1.07	0.05	0.80	0.77
c_2	0.24	-	1.35	1.92	0.70	0.17	1.79	0.26	0.30
c_3	0.42	1.35	-	0.77	1.68	0.21	0.47	0.89	0.22
c_4	0.10	1.92	0.77	-	1.72	0.28	1.56	0.27	0.29
c_5	0.45	0.70	1.68	1.72	-	1.32	0.49	1.15	0.23
c_6	1.07	0.17	0.21	0.28	1.32	-	0.15	0.40	0.75
c_7	0.05	1.79	0.47	1.56	0.49	0.15	-	0.35	0.47
c_8	0.80	0.26	0.89	0.27	1.15	0.40	0.35	-	0.44
c_9	0.77	0.30	0.22	0.29	0.23	0.75	0.47	0.44	-

4.2 clustering reliability calculation

The second step in Fig. 1 is devoted to computing reliability of each clustering in the ensemble. for a given clustering in the ensemble the clustering reliability is defined as the quantity certain knowledge of the ensemble about the clustering. Reliable clusterings are clusterings that share more information with other clusterings in the ensemble, in other words reliable clusterings are stable clusterings. Because ignoring reliability of base clusterings makes ensemble method vulnerable to low-quality clusterings, then we compute the reliability of each clustering and considered as clustering weight in construction co-association matrix. We assume that there is no knowledge about the original data-object features, the reliability of each fuzzy clustering is computed based on its shared information with other based clustering in the ensemble (sub section 4.2.2).

4.2.1 Estimating the cluster unreliability

The reliability of fuzzy clustering π^m in the ensemble Π means how much the clustering π^mshares information with other base clusterings in the ensemble Π. Then, The reliability of fuzzy clusterings is computed based on fuzzy normalized mutual information (FNMI) [30]. FNMI is obtained without the knowledge of the original data features or making any assumptions on data distribution. The fuzzy normalized mutual information between two fuzzy base clusterings π^a and π^b is calculated according to definition 6.

Definition 6. Fuzzy Normalized Mutual Information between two fuzzy clustering π^a, π^b is shown by $FNMI$ (π^a, π^b) and is computed *as:*

$$FNMI(\pi^a, \pi^b) = {FMI(\pi^a, \pi^b)} / {max(H({\pi^a}_{\pi^b}), ({\pi^b}_{\pi^a}))} \quad (8)$$

where $FMI(\pi^a, \pi^b)$ is the fuzzy mutual information between two clustering π^a and π^b, and is computed as,

$$FMI(\pi^a, \pi^b) = H({\pi^a}_{\pi^b}) + H({\pi^b}_{\pi^a}) - JH(\pi^a, \pi^b) \quad (9)$$

where $JH(\pi^a, \pi^b)$ is the joint entropy between two fuzzy clusterings π^a and π^b, and is computed by Eq. (12),

$$JH(\pi^a, \pi^b) = -\sum_{t=1}^{n^a} \sum_{l=1}^{n^b} \left(\frac{sim(C_t^a, C_l^b)}{MSsim(\pi^a, \pi^b)} * \log\left(\frac{sim(C_t^a, C_l^b)}{MSsim(\pi^a, \pi^b)}\right)\right) \quad (10)$$

and $H({\pi^a}_{\pi^b})$ is the entropy of fuzzy clustering π^a in relation to clustering π^b and is computed by Eq. (11) and $H({\pi^b}_{\pi^a})$ is the entropy of fuzzy clustering π^b in relation to clustering π^b

$$H({\pi^a}_{\pi^b}) = -\sum_{t=1}^{n^a} \frac{Ssim\left(C_t^a \ {}_{\pi^b}\right)}{MSsim(\pi^a, \pi^b)} \log \frac{Ssim\left(C_t^a \ {}_{\pi^b}\right)}{MSsim(\pi^a, \pi^b)} \quad (11)$$

where $sim(C_t^a, C_r^b)$ is the similarity between two fuzzy clusters $c_t \in \pi^a$, $c_r \in \pi^b$ and is computed according to Definition 5, $Ssim\left(C_t^a \ {}_{\pi^b}\right)$ is the sum of similarity between the fuzzy clusters $c_t \in \pi^a$ and all clusters $C_l^b \in \pi^b$ and is computed according to Eq. (12) and $MSsim(\pi^a, \pi^b)$ is the sum of similarity between each cluster of clustering π^a in relation to each cluster of clustering π^b (Mutual similarity) and is computed according to Eq. (13).

$$Ssim\left(C_t^a \ {}_{\pi^b}\right) = \sum_{l=1}^{n^b} sim(C_t^a, C_l^b) \quad (12)$$

$$MSsim(\pi^a, \pi^b) = \sum_{t=1}^{n^a} \sum_{l=1}^{n^b} sim(C_t^a, C_l^b) \quad (13)$$

Example 3 (Continuation of example 2). The FNMI between the fuzzy clustering pairs in Table 1 based on the similarity between fuzzy clusters in Table 2 has been calculated according to Definition 6 and the result is shown in Table 4.

Table 4 The values of *FNMI* corresponding to Table 1

	$\boldsymbol{\pi^1}$	$\boldsymbol{\pi^2}$	$\boldsymbol{\pi^3}$
$\boldsymbol{\pi^1}$	-	0.3131	0.3012
$\boldsymbol{\pi^2}$	0.3132	-	0.3012
$\boldsymbol{\pi^3}$	0.3012	0.3012	-

4.2.2 Computing the reliability of each fuzzy clustering in the ensemble

As mentioned above the reliability of a clustering in the ensemble is the certainty knowledge of the ensemble about the cluster. Because we supposed that we do not have any knowledge about the underlying dataset (we only access to an ensemble of clusterings) we must first compute the normalized mutual information between each pair of base clusterings in the ensemble. Then the reliability of each base clustering is the average of its normalized mutual information with respect to other base clusterings in the ensemble and is termed as reliability driven fuzzy clustering indicator (RDFCI) and computed as:

Definition 7. For an ensemble *Π* with β base clusterings, the weight of each fuzzy clustering in clustering ensemble (i.e., RDFCI: reliability driven fuzzy clustering indicator) for a clustering π^a is defined as

$$FDFCI(\pi^a, \Pi\ ) = \frac{\sum_{t=1}^{\beta} FNMI(\pi^a, \pi^t)}{\beta - 1} \qquad (14)$$

where $FNMI(\pi^a, \pi^t)$ is the normalized mutual information between fuzzy base clusterings π^a and π^t that computed according to Definition 6.

Example 4. The *RDFCI* values corresponding to Table 4 are computed according to equation (13) and the result is shown in Table 5

Table 5. The values of *RDFCI* corresponding to Table 4.

π^1	π^2	π^3
0.3072	0.3072	0.3012

4.3 Computing reliability based fuzzy co-association matrix

As can be seen in Fig. 1, the third step is computing the co-association matrix with regard to the reliability of each clustering in the ensemble. One of the most common methods used to combine the base clustering is the co-association matrix-based method. Evidence Accumulation Clustering (EAC), which was first proposed by Fred and Jain [18]. EAC maps the individual data-object clustering in a clustering ensemble into a new pairwise similarity measure.

Unlike in the general crisp evidence accumulation method, because a data-object doesn't belong to any base cluster absolutely we can't compute the co-association values by measuring how many times a pair of data-objects appears in the same cluster. In fuzzy clustering, each data-object belongs to each cluster with different membership degrees. Therefore, we should find a way to evaluate the strength of association between data-objects.

Definition 9. The fuzzy co-association clustering ensemble matrix is expressed as:

$$FCo_{i,j}^{\pi_{\cdot}(x)} = {}^{1}/_{\beta} \sum_{k=1}^{\beta} {}_{t=1}^{n^k}\sup\left(\inf\left(\pi^k(\vec{x_i})^t, \pi^k(\vec{x_j})^t\right)\right) \quad (15)$$

where $\vec{x_i}$ and $\vec{x_j}$ are the data-objects. If $\inf(x, y) = xy$ and $sup(x, y) = x + y$, then

$$FCo_{:}^{\pi^k} = \pi^k \times (\pi^k)^T \quad (16)$$

where $(\pi^k)^T$ is the transposition matrix of π^k.

As was mentioned in Definition 9, the co-association matrix reflects the strength of association between data-objects. In order to take the reliability of each clustering into account in the co-association matrix, *RDCI* would be considered as a multiplier term (weight) in co-association matrix computation, leading to computation of the weighted (reliability based) fuzzy co-association matrix according to Definition 10.

Definition 10. The reliability based fuzzy co-association clustering ensemble matrix ($RFCo$) is defined as:

$$RFCo_{i,j}^{\pi_{\cdot}(x)} = {}^{1}/_{\beta} \sum_{k=1}^{\beta} RDFCI(\pi^k, \Pi\ )_{t=1}^{n^k}\sup\left(\inf\left(\pi^k(\vec{x_i})^t, \pi^k(\vec{x_j})^t\right)\right) \quad (17)$$

Example 6 (Continuation of example 5). The $RFCo$ matrix of fuzzy clustering ensemble in example 1 (Table 1) with regard to calculated *RDFCI* in Table 5 With $\beta = 3$ is calculated as: $RDCI(C_1^1, \Pi)$=0.3218, $RDCI(C_2^1, \Pi)$=0.3666, $RDCI(C_3^1, \Pi)$=0.3407,$RDCI(C_1^2, \Pi)$=0.3230, $RDCI(C_2^2, \Pi)$=0.3324,$RDCI(C_3^2, \Pi)$=0.4116, $RDCI(C_1^3, \Pi) = 0.4427$, $RDCI(C_2^3, \Pi) = 0.4105$, $RDCI(C_3^3, \Pi) = 0.3819$,$\pi^1(\vec{x_1})^1 = 0.1$, $\pi^1(\vec{x_2})^1 = 0.0$,$\pi^1(\vec{x_1})^2 = 0.7$, $\pi^1(\vec{x_2})^2 = 0.8$, $\pi^1(\vec{x_1})^3 = 0.2$,$\pi^1(\vec{x_2})^3 = 0.2$, $\pi^2(\vec{x_1})^1 = 0.6$, $\pi^2(\vec{x_2})^1 = 0.8$,$\pi^2(\vec{x_1})^2 = 0.3$, $\pi^2(\vec{x_2})^2 = 0.2$, $\pi^2(\vec{x_1})^3 = 0.1$, $\pi^2(\vec{x_2})^3 = 0.0$, $\pi^3(\vec{x_1})^1 = 0.7$, $\pi^3(\vec{x_2})^1 = 0.9$,$\pi^2(\vec{x_1})^2 = 0.2$, $\pi^3(\vec{x_2})^2 = 0.1$, $\pi^3(\vec{x_1})^3 = 0.1$ and $\pi^3(\vec{x_2})^3 = 0.0$ according to Definition 10, $WFCo_{1,2} = {}^{1}/_{3}$ (sup $[RDCI(C_1^1, \Pi) * inf(\pi^1(\vec{x_1})^1, \pi^1(\vec{x_2})^1), RDCI(C_2^1, \Pi) *$ $\inf(\pi^1(\vec{x_1})^2), \pi^1(\vec{x_2})^2), RDCI(C_3^1, \Pi) * \inf(\pi^1(\vec{x_1})^3, \pi^1(\vec{x_2})^3)] +$ sup $[RDCI(C_1^2, \Pi) *$ $inf(\pi^2(\vec{x_1})^1, \pi^2(\vec{x_2})^1), RDCI(C_2^2, \Pi) * \inf(\pi^2(\vec{x_1})^2), \pi^2(\vec{x_2})^2), RDCI(C_3^2, \Pi) *$ $\inf(\pi^2(\vec{x_1})^3, \pi^2(\vec{x_2})^3)] +$ sup $[RDCI(C_1^3, \Pi) * inf(\pi^3(\vec{x_1})^1, \pi^3(\vec{x_2})^1), RDCI(C_2^3, \Pi) *$ $\inf(\pi^3(\vec{x_1})^2), \pi^3(\vec{x_2})^2), RDCI(C_3^3, \Pi) * \inf(\pi^3(\vec{x_1})^3, \pi^3(\vec{x_2})^3)]$)=${}^{1}/_{3}$ (sup [0.3218 $*$ inf(0.2,0.1),0.3666 $*$ inf(0.8,0.1) , 0.3407 $*$ inf (0.7,0.8)] + sup [0.3230 $*$ inf(0.4,0.2),0.3324 $*$ inf(0.6,0.1),0.4116 $*$ inf (0.0,0.7)] + sup [0.4227 $*$ inf(0.7,0.9),0.4105 $*$ inf(0.2,0.1) , 0.3819 $*$ inf (0.1,0.0)]] = 0.2488. In a similar way, the other entries of $RFCo$ matrix can be obtained (see Table 7).

Table 6. The *RFCo* matrix of fuzzy clustering ensemble presented in Table 1.

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0.0478	0.0555	0.0458	0.0286	0.0192	0.0153
x_2	0.0555	0.0676	0.0534	0.0268	0.0117	0.0082
x_3	0.0458	0.0534	0.0537	0.0353	0.0159	0.0137
x_4	0.0286	0.0268	0.0353	0.0517	0.0327	0.0214
x_5	0.0192	0.0117	0.0159	0.0327	0.0613	0.0382

x_6	0.0153	0.0082	0.0137	0.0214	0.0382	0.0545

4. 4 the consensus functions

The final step in Fig. 1 is computing the final clustering (consensus clustering). The process of extracting final clustering from the co-association matrix using the *RFCo* method (Definition 10) is named consensus function. To obtain consensus function at the first we define the objective function to drive final clustering in section 4. 4.1, then we explain its solution in section 4. 4.2.

4. 4.1 The objective function

A consensus function is used to derive the final fuzzy clustering π^* from *Π* by solving the Eq. (18). At the first, we formalize the problem of finding the final fuzzy clustering π^* from the fuzzy clustering ensemble *Π* as objective function. Objective function must take into account both fuzzy cluster diversity and fuzzy cluster reliability of ensemble *E* summarized in fuzzy co-association clustering ensemble matrix (*RFCo*). We try to find the final fuzzy clustering π^* that its co-association matrix approximately equals to *RFCo*. According to Eq. (16), the co-association of π^* is $\pi^* \times {\pi^*}^T$. Hence, we try to minimize the (dissimilarity) distance between *RFCo* and co-association matrix of π^* *(*minimize the sum of the absolute difference between co-association matrix of the final clustering and co-association matrix of ensemble *E)*, as formalized in Eq. (18-1). Also π^* is fuzzy clustering, each element must be satisfying the constraint (18-2). Additionally, because sum of membership of a data-object to all clusters in π^* must equal to 1 the constraint (18-3) was added to Eq. (18). This objective function is defined as Eq. (18).

$$minimize\ \sum_{k=1}^{M}\sum_{l=1}^{M}\left|RFCo_{kl}-\sum_{j=1}^{k}\left({\pi^*}_{kj}\times{\pi^*}_{jl}\right)\right| \qquad (18\text{-}1)$$

Subject to:

$$\forall\ j\in\{1,\dots,k\},\ k\in\{1,\dots,M\}\}|\ 0\le {\pi^*}_{kj}\le 1 \qquad (18\text{-}2)$$

$$\forall\ \ k\in\{1,\dots,M\}\}|\sum_{j=1}^{k}{\pi^*}_{kj}=1 \qquad (18\text{-}3)$$

Where *RFCo* is the fuzzy co-association matrix of base clusterings Π, π^* is the final clustering matrix (problem variable that must be found), *M* is the number of data-objects, *K* is the number of clusters in the final clustering. It is worth noting that π^* is a $M \times K$ membership matrix, where rows correspond to data-objects and columns to clusters and where each element of it represents the membership degree of a data-object belonging to a particular cluster and each row of it is a membership degree of each data-object to final clusters.

4. 4.2 Problem Solver

The proposed solver named the FCESQP (Fuzzy Clustering Ensemble by Sequential Quadratic Programming) is introduced in this section. AS mentioned in the previous section, the cluster ensemble is formulated as an optimization problem.

The goal of the optimization problem is to minimize the fuzzy clustering ensemble objective function that implicitly yields to a clustering in which its reliability among base clusterings is maximized. To solve the proposed model, any nonlinear optimization solver can be applied. Sequential quadratic programming (SQP) method has proved highly effective for solving constrained optimization problems with smooth nonlinear functions in the objective and constraints [47], [48]. The objective function is nonlinear with linear constraints. Because the coefficient matrix of constraints is spare, the sparse SQP (SSQP) is applied for solving our optimization problem.

The SSQP algorithm is fully described by Gill, Murray and Saunders [49]. It employs a sparse sequential quadratic programming (SQP) algorithm with limited-memory quasi-Newton approximations to the Hessian of the Lagrangian. It is especially effective for nonlinear problems with functions and gradients that are expensive to evaluate. The functions should be smooth but need not be convex. SSQP is suitable for large-scale for general nonlinear programs of the form

$$\min_x f_0(x) \qquad (19\text{-}1)$$

Subject to: $Ax = a$ (19-2)

$$x_l \le x \le x_u \qquad (19\text{-}3)$$

where *l* and *u* are lower and upper bounds (with constant values), $f_0(x)$ is a smooth scalar objective function, *A* is a sparse matrix that refers coefficient values of the constraints, and x_l, x_u are lower and upper bound of variable x. We map the objective function (18) to Eq. (19) form as follows:

At the first, the transformation of matrix π^* into the vector x (containing $M \times K$ scalar variables) according to (Eq. (20)) is necessary.

$$j = 1..k, i = 1M \ : x_t = \pi^*{}_{ij} \qquad (20)$$

where *t=i+(j-1)*k*

After this transformation $M \times 1$ vector x_l is set to zero, $M \times 1$ vector x_h is set to one, the *a* $M \times 1$ vector is set to one and the $(M) \times (M \times k)$ sparse matrix *A* is defined according to (Eq.(21))

$$i = 1...M, j = 1...k, t = 1 ... (M \times k) \qquad A_{it} = \begin{cases} 1 & if\ t == (i-1)*j + k \\ 0 & otherwise \end{cases} \qquad (21)$$

Example 7. suppose the structurer of final clustering ($a\ 6 \times 3\ matix\ \pi^*$) for fuzzy base clusterings in example 1 (suppose *K*=3) where its *RFCo* is shown in table 6, is shown in table 7. This matrix (π^*) is transformed to the 18× 1 vector *x* by Eq. (20) and it is as x=($\pi^*{}_{11}, \pi^*{}_{12}, \pi^*{}_{13}, \pi^*{}_{21}, \pi^*{}_{21}, \pi^*{}_{22}, \pi^*{}_{23}, \pi^*{}_{31}, \pi^*{}_{32}, \pi^*{}_{32}, \pi^*{}_{33}, \pi^*{}_{41}, \pi^*{}_{42}, \pi^*{}_{43},$ $\pi^*{}_{51}, \pi^*{}_{52}, \pi^*{}_{53}, \pi^*{}_{61}, \pi^*{}_{62}, \pi^*{}_{63}$). The corresponding 6×18 matrix A is computed by Eq. (21) and the its values is shown in table 8.

Table 7. The structurer of the final clustering corresponds to base clustering in Table 1	**Table 8.** Coefficient matrix A related to example 5
$\pi^* = \begin{matrix} {\pi^*}_{11} & {\pi^*}_{12} & {\pi^*}_{13} \\ {\pi^*}_{21} & {\pi^*}_{22} & {\pi^*}_{23} \\ {\pi^*}_{31} & {\pi^*}_{32} & {\pi^*}_{33} \\ {\pi^*}_{41} & {\pi^*}_{42} & {\pi^*}_{43} \\ {\pi^*}_{51} & {\pi^*}_{52} & {\pi^*}_{53} \\ {\pi^*}_{61} & {\pi^*}_{62} & {\pi^*}_{63} \end{matrix}$	$\begin{bmatrix} 1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1 \end{bmatrix}$

As the proposed algorithm is based on the SQP approach, we provide here a short *synopsis* of that method. Because the objective function (19) ($f_0(x)$) is nonlinear we approximate it to quadratic form with consider 3 first term of Taylor series as Eq. (22).

$$F(\bar{x}) = F(x) + g^T(x)(\bar{x} - x) + \frac{1}{2}(\bar{x} - x)^T H(x)(\bar{x} - x) \qquad (22)$$

where H and g are the Hessian matrix and gradient vector of the objective function, x is the current value of variable X. At each iteration, Eq. (22) has only one variable $\bar{x}$, therefore $F(x)$ is constant and it can be eliminated. The constraint of our objective function is linear, then they not need to approximation. Therefore, it is rewritten as Eq. (23).

$$F(\bar{x}) = g^T(x)(\bar{x} - x) + \frac{1}{2}(\bar{x} - x)^T H(x)(\bar{x} - x) \quad (23\text{-}1)$$

$$A(x) = a \quad (23\text{-}2)$$

$$B(x) \le b \quad (23\text{-}3)$$

Eq. (23) is in the form of quadratic and we can solve it by using a sequence of quadratic programming (QP) subproblems in each iteration. Since the objective function (13) consists of equality and inequality constraints, the active-set method of QP is used to solve it as follows:

Start from an arbitrary point x_0, then find the next iterate by setting $x_{k+1} = x_k + p_k d_k$, where p_k is step-length and d^k is search direction at iteration k.

At the current iterate x_k determine the index set of active the inequality constraints:

$$A^k = \{j \mid b_j^T x_k - b_j = 0,\ j = 1, \dots, m_2\} \qquad (24)$$

Then we find the direction (d) value by solving Eq. (25).

$$\min_d \left\{ g^T(x_k + d) + \frac{1}{2}(x_k + d)^T H(x_k + d) \right\} \qquad (25\text{-}1)$$

Subject to:

$$a_i^T(x_k + d) = a_i, i = 1, \dots, m_1 \qquad (25\text{-}2)$$

$$b_j^T(x_k + d) = b_j,\ j \in A^k \qquad (25\text{-}3)$$

where m_1 is the number of equation constraints.

If we expand the Eq. (25), and simplify these expressions and drop constants to Eq. (26) is obtained.

$$\min_{d}\left\{(Hx_k + g)^T d + \frac{1}{2} d^T H d\right\} \quad (26\text{-}1)$$

Subject to:

$$Ad = 0 \quad (26\text{-}2)$$

$$\tilde{B}d = 0 \quad (26\text{-}3)$$

$$\text{where } \tilde{B} = \begin{bmatrix} \vdots \\ {b_j}^T \\ \vdots \end{bmatrix},\ j \in A^k \quad (26\text{-}4)$$

To obtain the search direction d^k, solve the Eq. (27):

$$\min_{d}\{{}^1/_2\, d^T H d + [g^k]^T d\} \quad (27\text{-}1)$$

Subject to:

$$Ad = 0 \quad (27\text{-}2)$$

$$\tilde{B}d = 0 \quad (27\text{-}3)$$

where $g^k = Hx_k + g$.

The Karush–Kuhn–Tucker (KKT) optimality conditions [50] lead to the Eq. (28):

$$\begin{bmatrix} H & A^T & \tilde{B}^T \\ A & 0 & 0 \\ \tilde{B} & 0 & 0 \end{bmatrix}\begin{bmatrix} d \\ \lambda \\ \tilde{\mu} \end{bmatrix} = \begin{bmatrix} -g^k \\ 0 \\ 0 \end{bmatrix} \quad (28)$$

where μ, λ are Lagrange multipliers corresponding to active inequality and equality constraint, respectively.

If d_k is a solution of QP, then there are λ_k and $\tilde{\mu}_k$ such that

$$Hd_k + g^k + A^T \lambda_k + \tilde{B}^T \tilde{\mu}_k = 0 \quad (29\text{-}1)$$

$$Ad_k = 0 \quad (29\text{-}2)$$

$$\tilde{B}d_k = 0 \quad (29\text{-}3)$$

There are two cases: either d_k =0 or $d_k \neq 0$.

Case 1**:** d_k =0. the Eq. (29) above reduces to Eq. (30).

$$g^k + A^T \lambda_k + \tilde{B}^T \tilde{\mu}_k = 0 \quad (30)$$

Case 1-a: if $\tilde{\mu}_k \geq 0$, $x_{k+1} = x_k$ is a KKT point. Stop!

Case 1-b: If some components of $\tilde{\mu}_k$ are negative, then x_k is not an optimal solution. Let $\mu_{j_o} = \min\{\tilde{\mu}_j | \tilde{\mu}_j < 0,\ \ j \in A^k\}$. Remove the index j_o from A^k and solve the quadratic programming problem (31).

$$\min_{d}\{{}^1/_2\, d^T H d + [g^k]^T d\} \quad (31\text{-}1)$$

Subject to:

$$a_i^T d = 0 \quad (31\text{-}2)$$

$b_j^T d = 0, j \in A^k \backslash \{j_o\}$ (31-3)

Then the obtained direction is descent direction d_k for (QP).

Case 2: ($d_k \neq 0$): Determine a step-length p_k that guarantees $x_k + p_k d_k$ is feasible to QP.

A common p_k that guarantees the satisfaction of all constraints is

$$min\left\{1, \frac{b_j - b_j{}^T d_k}{b_j{}^T d_k} \middle| j \notin A^k \ and\ b_j{}^T d_k > 0\right\} \quad (32)$$

If p_k<1, then p=$\frac{b_j - b_j{}^T d_k}{b_j{}^T d_k}$ for some $j_o \notin A^k$. This implies that $b_{j_o}{}^T(x_i + p * d_k) = b_{j_o}$. This is, the inequality constraint in Eq. (23) corresponding to the index j_o becomes active and must be added to active set. Then $A^k = A^k \cup j_o\}$.

The SQP algorithm is summarized in the Algorithm1.

Algorithm 1. Consensus_clustering_SQP

// Input: Objective function, number of final clustering K

// output final clustering π^*

1. Transform final clustering membership π^* to vector form *x* by Eq. (20)

2. rewrite the objective function (18) as a nonlinear form (19) by Eq. (21)

3. convert nonlinear objective function (19-1) by Taylor series to quadratic form (29)

4: Give a start vector x_0 as initial solution for x;

5: Identify the active index set A^0;

6: Set k = 0;

7: while (no convergence) {

8: Compute $g^k = Hx_k + g$

9: Obtain d_k , λ_k and $\tilde{\mu}_k$ by solving the KKT-equations for

$\min_d\{{}^1/_2\, d^T H d + [g^k]^T d\}$ (27-1)

Subject to:

$a_i^T d = 0$ (27-2)

$b_j^T d = 0$ (27-3), $j \in A^k$

10: if (d_k = 0) {

11: if $\tilde{\mu}_k \geq 0$ {

12: STOP! x_k is a KKT point.

13: else

14: $\mu_{j_o} = \min\{\tilde{\mu}_j | \tilde{\mu}_j < 0,\ j \in A^k\}$

165: Update the index set $A^k \leftarrow A^k \backslash \{j_o\}$ and GOTO Step 6.

16: } // end if $\tilde{\mu}_k \geq 0$

17: } // end if (d_k = 0)

18. if ($d_k \neq 0$)

19. Compute the step-length as:

$$p_k = min\left\{1, \frac{b_j - b_j{}^T d_k}{b_j{}^T d_k} \middle| j \notin A^k \ and \ b_j{}^T d_k > 0\right\}$$

20. update $x_{k+1} = x_k + p_k * d_k$

21. Update active index-set: if p_k = 1, then $A^{k+1} = A^k$, else $A^{k+1} = A^k \cup \{j_o\}$, where $p_k = , \frac{b_{j_o} - b_{j_o}{}^T d_k}{b_{j_o}{}^T d_k} \ for \ b_{j_o}{}^T d_k > 0$

22. Update $k \leftarrow k + 1$.

23 } // end if ($d_k \neq 0$)

24. } //end while

25. transform vector *x* to $M \times K$ matrix π^* as final clustering

The approximate Hessian matrix Q is updated from iteration to iteration using one of the variable metric updating formulas [51]. Because the matrix of coefficient constraints in our objective function (Eq. (20)) is sparse, SQP algorithm must exploits sparsity in the constraint Jacobian and maintains a limited-memory quasi-Newton approximation H_k to the Hessian of the Lagrangian [52].

It is worth mentioning that the vector *X* is transformed to the $M \times K$ matrix π^* by Eq. (33)

$t = 1 \ldots M \times K, j = 1..K, i = 1 \ldots M \ \pi^*{}_{ij} = x_t$ where i = $\lfloor t/K \rfloor$ $and \ j = t - K \times \lfloor t/K \rfloor$ (33)

4. 4.3 the consensus algorithm

To obtain consensus clustering from base clustering *Π* according to the local reliability of each clustering in the ensemble, it is necessary to compute each clustering's reliability in the ensemble according to Eq. (14). For this purpose, the FNMI of each cluster in relation to other base clustering in *Π* needs to be computed according to Eq. (8). For this computation, computing the similarity between cluster pairs in the ensemble *Π* according to Eq. (7) is necessary. After computing the *RDFCI*, the weighted fuzzy co-association clustering ensemble matrix ($RFCo$) is obtained according to Eq. (16). Then based on $RFCo$ the objective function (Eq. (17)) is constructed. Finally, by applying the Consensus_clustering_SQP algorithm as solver over the Eq. (18) the final clustering (π^*) is obtained.

This algorithm is named FCESQP (Fuzzy Clustering Ensemble by SQP) is presented in Algorithm 2 with details. In this algorithm *Π* is the base clustering ensemble and k is the number of clusters in the final clustering.

Algorithm 2 FCESQP **(Fuzzy Clustering Ensemble by SQP Algorithm)** ***Input:*** *Π, k*.

// **Input**: *Π, k*

// *Π* is an ensemble of basic clusterings

// *k* is the number of final clusters

// output: π^* is final clustering

1: Compute the similarity of each cluster in relation to other clusters in each clustering π^m belong to the ensemble Π according to Definition 5.

2: for each two base clusterings in *Π* Compute the Fuzzy clustering pairwise *FNMI* according to Definition 6

3: Compute the *RDFCI* of the base clusterings in ensemble according to Definition 7.

4: Construct the *RFCo* matrix according to definition 10.

5: construct the objective function that consists of *RFCo* (Eq. (18)).

6. Obtain the final clustering with *k* clusters via Consensus_clustering_SQP (*Eq. (18)*,*k*)

//($\pi^* =$ Consensus_clustering_SQP $(Eq. (18), k)$

Output: the consensus clustering π^*.

5. EXPERIMENTS

5.1 Datasets and quality evaluation criterion

To evaluate the robustness and quality of the proposed fuzzy clustering ensemble approach, twelve data sets are selected from UCI Machine Learning Data Sets [53], the " Galaxy" dataset described in [54] and a well-known dataset HalfRing as the experimental datasets. The description of these datasets is shown in Table 9.

Two evaluation criteria *NMI* and *Dunn* are applied here to assess the quality of clustering.

NMI is normalized mutual information between two clusterings [5], and for two clusterings π^1 and π^2 is calculated as

$$NMI(\pi^1, \pi^2) = {I(\pi^1, \pi^2)}\Big/{(\sqrt{H(\pi^1)}\sqrt{H(\pi^2)})} \tag{34}$$

where $I(\pi^1, \pi^2)$ denotes the mutual information between two clusterings and $H(\pi^1)$ denotes the entropy of π^1. In this paper π^1 is final clustering and π^2 is ground truth of each dataset. A larger value of *NMI* indicate a better clustering result.

The Dunn Index [43] is defined as

$$Dunn(\pi^i) = \min_{\substack{j \in \{1,\dots,c^i\} \\ k \in \{1,\dots,c^i\} \\ j \neq k}} \left\{ \frac{\min\_dis(C_j^i, C_k^i)}{\max_{t \in \{1,\dots,c^i\}} (diam(C_t^i))} \right\} \tag{35}$$

where min $\_dis(C_j^i, C_k^i)$ is the distance between the two nearest data-objects in clusters C_j^i and C_k^i and $diam(C_t^i)$ is the diameter of the cluster C_t^i, similar to *NMI*, a higher value of the *Dunn* index indicates a better clustering result.

5.2 base clustering generation

To evaluate the consensus quality over various ensembles, base clustering are constructed through the FCM and K-means clustering algorithms. In order to construct diverse base clustering, the FCM and K-means are run with different numbers of clusters. The number of clusters for them is randomly chosen from the $[2, \sqrt{M}]$ interval, where M is the number of data-objects in the dataset under experiment.

The ensemble size for performance evaluation of the methods was assumed as $\beta = 10$. Base on empirical results, the best result is obtained when parameter $\emptyset = 0.84$ approximately. To rule out the occasional luck factor and provide a fair comparison, this proposed approach, the state-of-the-art fuzzy clustering ensemble methods were assessed by their quality criteria and AC robustness average over numerous runs (40 runs).

Table 9. Description of the data sets

Dataset	Number of data-objects (M)	Number of classes (k)	Number of features (N)
Breast	683	2	9
Galaxy	323	7	4
Glass	214	7	10
Haberman	306	2	3
Halfring	400	2	2
Ionesphere	351	2	34
Iris	178	3	13
Knowledge	258	4	5
Seeds	210	3	7
SAHeart	462	2	9
Wine	178	3	13
Vehicle	846	4	18

All experiments are run in Matlab R2014a 64-bit environment on a Windows Server 2008 64-bit, Intel Xeon CPU E5-2609(2.5 GHz 2.5 GHz) 2 processors and 16 GB of RAM workstation.

5.3 Comparison of the quality of the proposed approach against the other clustering methods

The proposed SQP approach were compared with eight clustering ensemble methods, i.e. GPMGLA [27], SVC [23], PVC [23], BVC [23], ISC [23], Berikov [31] and FSCEOGA1 [34]. The two quality-evaluating criteria, *NMI* and *Dunn* were applied to determine the quality of the final clustering resulted from the proposed methods and the baseline methods. The number of clusters in each dataset is the same as the number of pre-defined classes (ground truth) in each dataset.

For comparison purposes, each of the proposed methods and the baseline methods are executed 40 times. The average values of *NMI* and *Dunn* criteria of different methods over 40 runs are shown in Tables 10 and 11 respectively. The value in bold in the rows represents the best quality-term of each dataset yield by all the examined algorithms. The last row shows the average quality-term for each algorithm over all the datasets. Because the FSCEOGA1 is computationally expensive, these methods cannot handle large datasets because of large execution time. For this reason, the quality results of FSCEOGA1 method are missing on the Vehicle dataset. Therefore, the quality of FSCEOGA1 the Vehicle dataset denoted as a dash.

According to Table 10, FCESQP outperforms other algorithms on ten datasets, while GPMGLA and ISC outperform other algorithms only in one dataset. It is 2 times that the FCESQP algorithm obtains the third best results. We can see that FCESQP algorithm achieves the best average *NMI* with the value of *0.3753*.

To ensure the results do not happen by chance, and to assess quality of the proposed method running statistical analysis is a must. The Friedman test [55] is applied here to the results of Tables 10, subject to null hypothesis, where the mean ranks are equal for all the examined algorithms. The significant level is set to *0.05*. The experimental results, subject to Friedman test in Table 10 is shown in Fig.2. As observed in Fig. 2 and the null hypothesis that the mean rank of the *NMI* being equal in all algorithms is rejected, because p-value is *3.651E-6*, indicating that there exists a significant difference. As observed in the mean ranks, FCESQP has the highest *NMI* score followed by SVC and then ISC.

TABLE 10. The NMI resulted from different algorithms

Dataset	GPMGLA	SVC	PVC	BVC	ISC	Berikov	FSCEOGA1	FCESQP
Breast	0.0029	0.2789	0.1715	0.6985	0.3750	0.0457	0.6977	0.7685
Galaxy	0.2768	0.3083	0.0461	0.0445	0.3196	0.0986	0.2309	0.2746
Glass	0.3648	0.3129	0.0823	0.0944	0.3584	0.0901	0.2797	0.3410
Haberman	0.0002	0.0253	0.0021	0.0002	0.0231	0.0164	0.0003	0.1025
Halfring	0.3088	0.2838	0.0051	0.2238	0.2608	0.0264	0.2886	0.3130
Ionesphere	0.0165	0.1448	0.1485	0.1312	0.1403	0.0642	0.1227	0.1767
Iris	0.7869	0.5923	0.1272	0.5923	0.5458	0.0815	0.6813	0.7899
Knowledge	0.1115	0.2681	0.0815	0.1703	0.2502	0.0669	0.2455	0.2614
Seeds	0.6286	0.4161	0.2751	0.4698	0.7075	0.0741	0.5963	0.7156
Sheart	0.0000	0.0388	0.0303	0.0592	0.0379	0.0194	0.0741	0.0768
Wine	0.3927	0.2673	0.1256	0.3806	0.2315	0.0389	0.4033	0.4724
Vehicle	0.2027	0.1610	0.0720	0.0807	0.2061	0.0170	-	0.2111
Alg. Avg	0.2577	0.2581	0.0973	0.2455	0.2880	0.0533	0.3291	0.3753

Ranks

	Mean Rank
GPMGLA	4.32
SVC	5.32
PVC	2.64
BVC	3.91
ISC	5.27
Berikov	2.00
FSCEOGA1	5.09
FCESQP	7.45

Test Statistics[a]

N	11
Chi-Square	37.574
df	7
Asymp. Sig.	3.651E-6

a. Friedman Test

Fig. 2. Friedman test result of Table 10

According to Table 11, it is obvious FCESQP outperforms other algorithms on nine datasets, while FSCEOGA1 outperforms other algorithms on dataset Breast and PVC outperforms other algorithms on dataset Glass. GPMGLA and ISC achieve the best performance in terms of *Dunn* index on dataset Seeds. With respect to last row (average values on all datasets) it is obvious that FCESQP algorithm achieves the best average *Dunn* index with the value of *1.78*, GPMGLA has the second score and FSCEOGA1 has the third score.

The experimental results, subject to Friedman test in Table 11 is shown in Fig.3. As observed in Fig. 3 and the null hypothesis that the mean rank of the *Dunn* being equal in all algorithms is rejected, because p-value is *2.104E-7*, indicating that there exists a significant difference. As observed in the mean ranks, FCESQP has the highest *Dunn* score followed by FSCEOGA1 and then GPMGLA.

TABLE 11. The Dunn index resulted from different algorithms

Dataset	GPMGLA	SVC	PVC	BVC	ISC	Berikov	FSCEOGA1	FCESQP
Breast	0.18	0.26	0.57	0.85	0.39	0.08	1.63	1.46
Galaxy	1.39	1.10	0.10	0.43	1.27	0.05	0.69	1.45
Glass	0.33	0.51	1.77	0.10	0.75	0.01	0.37	1.74
Haberman	1.98	0.51	1.02	0.35	0.57	0.08	2.01	2.12
Halfring	2.46	1.07	0.85	0.62	1.46	0.01	2.47	2.68
Ionesphere	0.17	0.27	0.60	1.02	0.35	0.07	0.92	1.15
Iris	2.41	2.15	1.61	2.15	1.12	0.04	1.92	2.46
Knowledge	0.51	1.08	0.70	0.32	1.02	0.10	1.03	1.12
Seeds	2.35	1.30	0.28	0.54	2.35	0.04	1.88	2.33
SHeart	0.20	0.00	1.11	0.09	0.75	0.10	1.02	1.14

Wine	1.33	0.41	1.53	0.74	0.27	0.06	1.83	1.90
vehicle	1.73	0.50	0.44	0.56	1.56	0.03	-	1.80
Alg. Avg	1.25	0.76	0.88	0.65	0.99	0.06	1.43	1.78

Ranks

	Mean Rank
GPMGLA	4.77
FSCEOGA1	5.82
SVC	3.95
PVC	4.55
BVC	3.50
ISC	4.59
Berikov	1.18
FCESQP	7.64

Test Statistics[a]

N	11
Chi-Square	44.035
df	7
Asymp. Sig.	2.104E-7

a. Friedman Test

Fig. 3. Friedman test result of Table 11

6. Conclusion and future work

In this paper, a novel fuzzy cluster ensemble approach based on the computation of normalized mutual information of each fuzzy base clustering in the ensemble has been proposed. Then a new reliability driven fuzzy clustering indicator termed *RDFCI* was proposed. The *RDFCI* measure does not depend on the original data features and has no presumption on data distribution. A local weighting scheme to promote the conventional co-association matrix through the *RDFCI* weigh has been also introduced named *RFCo*. Instead of participating all fuzzy clusterings in the co-association matrix equally, in this approach each fuzzy clustering participates in the co-association matrix with respect to its reliability in the ensemble. In order to extraction final fuzzy clustering from matrix *RFCo* a constrained nonlinear optimization problem was formed. We solve this problem by sparse sequential quadratic programming (SSQP). The experimental results over twelve datasets confirm the quality improvement in comparison with other fuzzy clustering ensemble methods.

Propose a parallel solution that obtains the final clustering by solving the optimization problem can be considered as in a future work (mapping it to map-reduce and solving by cluster computing). Solving this nonlinear optimization problem by other methods can be discussed as a future work of this paper. Apply this approach in some real-world applications (especially engineering applications) will also be carried out.

References

[1] J. C. Bezdek, R. Ehrlich, and W. Full, “FCM: The fuzzy c-means clustering algorithm,” *Comput. Geosci.*, vol. 10, no. 2–3, pp. 191–203, 1984.

[2] D. E. Gustafson and W. C. Kessel, “Fuzzy clustering with a fuzzy covariance matrix,” in *Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on*, 1979, pp. 761–766.

[3] I. Gath and A. B. Geva, “Unsupervised optimal fuzzy clustering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 773–780, 1989.

[4] D. Graves and W. Pedrycz, “Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study,” *Fuzzy Sets Syst.*, vol. 161, no. 4, pp. 522–543, Feb. 2010.

[5] A. Strehl and J. Ghosh, “Cluster ensembles---a knowledge reuse framework for combining multiple partitions,” *J. Mach. Learn. Res.*, vol. 3, no. Dec, pp. 583–617, 2002.

[6] X. Z. Fern and C. E. Brodley, “Random projection for high dimensional data clustering: A cluster ensemble approach,” *Proc. Twent. Int. Conf. Mach. Learn.*, vol. 20, pp. 186–193, 2003.

[7] X. Z. Fern and C. E. Brodley, “Solving cluster ensemble problems by bipartite graph partitioning,” in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 36.

[8] D. Greene, A. Tsymbal, N. Bolshakova, and P. Cunningham, “Ensemble clustering in medical diagnostics,” in *Computer-Based Medical Systems, 2004. CBMS 2004. Proceedings. 17th IEEE Symposium on*, 2004, pp. 576–581.

[9] S. T. Hadjitodorov, L. I. Kuncheva, and L. P. Todorova, “Moderate diversity for better cluster ensembles,” *Inf. Fusion*, vol. 7, no. 3, pp. 264–275, 2006.

[10] L. I. Kuncheva, S. T. Hadjitodorov, and L. P. Todorova, “Experimental comparison of cluster ensemble methods,” in *Information Fusion, 2006 9th International Conference on*, 2006, pp. 1–7.

[11] a. Topchy, a. K. Jain, and W. Punch, “Combining multiple weak clusterings,” *Third IEEE Int. Conf. Data Min.*, pp. 0–7, 2003.

[12] A. P. Topchy, A. A. K. Jain, and W. F. Punch, “A Mixture Model for Clustering Ensembles.,” *Sdm*, pp. 379–390, 2004.

[13] A. Topchy, A. K. Jain, and W. Punch, “Clustering ensembles: models of consensus and weak partitions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1866–1881, Dec. 2005.

[14] S. VEGA-PONS and J. RUIZ-SHULCLOPER, “a Survey of Clustering Ensemble

Algorithms," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 25, no. 03, pp. 337–372, 2011.

[15] L. Franek and X. Jiang, "Ensemble clustering by means of clustering embedding in vector spaces," *Pattern Recognit.*, vol. 47, no. 2, pp. 833–842, 2014.

[16] A. L. N. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, 2005.

[17] H. P. B. Minaei-bidgoli, "A clustering ensemble framework based on selection of fuzzy weighted clusters in a locally adaptive clustering algorithm," pp. 87–112, 2015.

[18] a. L. N. Fred and a. K. Jain, "Data clustering using evidence accumulation," *Object Recognit. Support. by user Interact. Serv. Robot.*, vol. 4, pp. 276–280 vol.4, 2002.

[19] C. Zhong, X. Yue, Z. Zhang, and J. Lei, "A clustering ensemble: Two-level-refined co-association matrix with path-based transformation," *Pattern Recognit.*, vol. 48, no. 8, pp. 2699–2709, 2015.

[20] V. Singh, L. Mukherjee, J. M. Peng, and J. H. Xu, "Ensemble clustering using semidefinite programming with applications," *Mach. Learn.*, vol. 79, no. 1–2, pp. 177–200, 2010.

[21] S. Vega-Pons, J. Correa-Morris, and J. Ruiz-Shulcloper, "Weighted partition consensus via kernels," *Pattern Recognit.*, vol. 43, no. 8, pp. 2712–2724, 2010.

[22] H. G. Ayad and M. S. Kamel, "Cumulative voting consensus method for partitions with variable number of clusters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 160–173, 2008.

[23] X. Sevillano, F. Alías, and J. C. Socoró, "Positional and confidence voting-based consensus functions for fuzzy cluster ensembles," *Fuzzy Sets Syst.*, vol. 193, no. Supplement C, pp. 1–32, 2012.

[24] H. G. Ayad and M. S. Kamel, "On voting-based consensus of cluster ensembles," *Pattern Recognit.*, vol. 43, no. 5, pp. 1943–1953, 2010.

[25] B. Minaei-Bidgoli, A. Topchy, and W. F. Punch, "Ensembles of partitions via data resampling," *Int. Conf. Inf. Technol. Coding Comput. ITCC*, vol. 2, no. Cl, pp. 188–192, 2004.

[26] J.-P. Barthélemy and B. Leclerc, "The Median Procedure for Partitions.," *Partitioning data sets*, vol. 19, pp. 3–34, 1993.

[27] D. Huang, J.-H. Lai, and C.-D. Wang, "Combining multiple clusterings via crowd agreement estimation and multi-granularity link analysis," *Neurocomputing*, vol. 170, no. November 2016, pp. 240–250, 2015.

[28] T. Li and C. Ding, "Weighted consensus clustering," in *Proceedings of the 2008 SIAM International Conference on Data Mining*, 2008, pp. 798–809.

[29] Z. Yu *et al.*, "Hybrid clustering solution selection strategy," *Pattern Recognit.*, vol. 47, no. 10, pp. 3362–3375, 2014.

[30] A. Bagherinia, B. Minaei-Bidgoli, M. Hossinzadeh, and H. Parvin, "Elite fuzzy clustering ensemble based on clustering diversity and quality measures," *Appl. Intell.*, pp. 1–24, 2018.

[31] V. B. Berikov, "A Probabilistic Model of Fuzzy Clustering Ensemble," *Pattern Recognit. Image Anal.*, vol. 28, no. 1, pp. 1–10, 2018.

[32] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 1, pp. 52–60, 1967.

[33] I. Saha, U. Maulik, S. Bandyopadhyay, and D. Plewczynski, "SVMeFC : SVM Ensemble Fuzzy Clustering for Satellite Image Segmentation," vol. 9, no. 1, 2012.

[34] H. ALIZADEH, B. MINAEI-BIDGOLI, and H. PARVIN, "Optimizing Fuzzy Cluster Ensemble in String Representation," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 27, no. 02, p. 1350005, 2013.

[35] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logist.*, vol. 2, no. 1–2, pp. 83–97, 1955.

[36] M. Van Erp, L. Vuurpijl, and L. Schomaker, "An overview and comparison of voting methods for pattern recognition," in *Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on*, 2002, pp. 195–200.

[37] J. C. de Borda, "M{é}moire sur les {é}lections au scrutin," *Hist. l'Academie R. des Sci. pour 1781 (Paris, 1784)*, 1784.

[38] A. H. Copeland, "A reasonable social welfare function," in *Mimeographed notes from a Seminar on Applications of Mathematics to the Social Sciences, University of Michigan*, 1951.

[39] E. Bedalli, E. Mançellari, and O. Asilkan, "A Heterogeneous Cluster Ensemble Model for Improving the Stability of Fuzzy Cluster Analysis," *Procedia Comput. Sci.*, vol. 102, no. August, pp. 129–136, 2016.

[40] J. V. de Oliveira, A. Szabo, and L. N. de Castro, "Particle Swarm Clustering in clustering ensembles: Exploiting pruning and alignment free consensus," *Appl. Soft Comput.*, vol. 55, no. Supplement C, pp. 141–153, 2017.

[41] G. H. Ball and D. J. Hall, "ISODATA, a novel method of data analysis and pattern classification," 1965.

[42] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat.*

Methods, vol. 3, no. 1, pp. 1–27, 1974.

[43] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *J. Cybern.*, vol. 4, no. 1, pp. 95–104, 1974.

[44] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987.

[45] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, 1995.

[46] G. Zheng, "A Similarity Measure between Fuzzy Sets," in *Applied Mechanics and Materials*, 2012, vol. 229, pp. 2663–2666.

[47] R. T. Haftka and Z. Gurdal, "Elements of Structural Optimization, Third revised and expanded edition,(1992)." Kluwer Academic Publishers.

[48] T. F. Edgar, D. M. Himmelblau, and L. S. Lasdon, *Optimization of chemical processes*. McGraw-Hill, 2001.

[49] P. E. Gill, W. Murray, and M. A. Saunders, "SNOPT: An SQP algorithm for large-scale constrained optimization," *SIAM Rev.*, vol. 47, no. 1, pp. 99–131, 2005.

[50] H. W. Kuhn, "„& Tucker, AW (1951). Nonlinear programming," in *Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press*, pp. 481–492.

[51] S.-P. Han, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems," *Math. Program.*, vol. 11, no. 1, pp. 263–282, 1976.

[52] J. E. Dennis Jr and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, vol. 16. Siam, 1996.

[53] C. L. Blake and C. J. Merz, "UCI Repository of machine learning databases [http://www. ics. uci. edu/~ mlearn/MLRepository. html]. Irvine, CA: University of California," *Dep. Inf. Comput. Sci.*, vol. 55, 1998.

[54] S. C. Odewahn, E. B. Stockwell, R. L. Pennington, R. M. Humphreys, and W. A. Zumach, "Automated star/galaxy discrimination with neural networks," *Astron. J.*, vol. 103, pp. 318–331, 1992.

[55] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the fbietkan statistic," *Commun. Stat. Methods*, vol. 9, no. 6, pp. 571–595, 1980.