



A fuzzy clustering ensemble based on cluster clustering and iterative Fusion of base clusters

Musa Mojarad¹ · Samad Nejatian^{2,3} · Hamid Parvin^{4,5} · Majid Mohammadpoor^{3,4}

Published online: 25 January 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

For obtaining the more robust, novel, stable, and consistent clustering result, clustering ensemble has been emerged. There are two approaches in clustering ensemble frameworks: (a) the approaches that focus on creation or preparation of a suitable ensemble, called as ensemble creation approaches, and (b) the approaches that try to find a suitable final clustering (called also as consensus clustering) out of a given ensemble, called as ensemble aggregation approaches. The first approaches try to solve ensemble creation problem. The second approaches try to solve aggregation problem. This paper tries to propose an ensemble aggregator, or a consensus function, called as Robust Clustering Ensemble based on Sampling and Cluster Clustering (RCESCC). RCESCC algorithm first generates an ensemble of fuzzy clusterings generated by the fuzzy c-means algorithm on subsampled data. Then, it obtains a cluster-cluster similarity matrix out of the fuzzy clusters. After that, it partitions the fuzzy clusters by applying a hierarchical clustering algorithm on the cluster-cluster similarity matrix. In the next phase, the RCESCC algorithm assigns the data points to merged clusters. The experimental results comparing with the state of the art clustering algorithms indicate the effectiveness of the RCESCC algorithm in terms of performance, speed and robustness.

Keywords Clustering ensemble · Fuzzy c-means · Between-cluster similarity · Subsampling

1 Introduction

Clustering is one of the most essential tasks in machine learning [1, 2]. Its main aim is to find a data grouping in such a way that the similarities among data pairs of the same group are as much as possible and the similarities among data pairs of the different groups are as small as possible. It can be used in image processing [3] and image segmentation [4], feature extraction [5] and feature reduction [6], geospatial data clustering [7, 8].

However, applying different clustering algorithms or the same clustering algorithm with different initializations on a same given data-set can result in diverse base clusterings. One of the most challenging issues in clustering task can be to select the best method to cluster a given data-set or at least to find an appropriate clustering algorithm. Strehl and Ghosh [9] proposed a clustering ensemble which combines all clustering results rather than finds the best one. Although clustering ensemble has been applied to many scenarios, not all solutions make positive contribution to the final result. Many existing clustering ensemble algorithms combine all clustering solutions, however, some others find that only merging partial solutions produces better results than combining all solutions [10–17]. Wang et al. [14] regarded selecting solutions as the problem of unsupervised feature selection based on the significance of attribute in rough set theory. Naldi et al. [12] selected solutions based on relative validity indexes. Ni et al. [13] selected solutions based on fractal dimension and projection.

Clustering ensemble is characterized by scalability, high robustness, stability, and parallelism. It combines all results of different clustering algorithms and it can be extended with the increase of clustering algorithms. The experiments in [18] showed that clustering ensemble outperforms single clustering algorithm and it is suitable for more datasets than one single

✉ Samad Nejatian
nejatian@iauyasooj.ac.ir

¹ Department of Computer Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran

² Department of Electrical Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran

³ Young Researchers and Elite Club, Yasooj Branch, Islamic Azad University, Yasooj, Iran

⁴ Department of Computer Engineering, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

⁵ Young Researchers and Elite Club, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

clustering algorithm. Clustering ensemble is robust against noise and it works well on heterogeneous datasets [19, 20]. One of the process in clustering ensemble is combining multiple solutions and this makes clustering ensemble parallel. In addition, clustering ensemble has an advantage in privacy protection and knowledge reuse. It only needs to access clustering solutions rather than original data, so it provides privacy protection for original data [21]. Clustering ensemble produces final partition by using the results of different clustering algorithms. Therefore, clustering ensemble has the characteristics of knowledge reuse [22].

Traditional clustering ensemble contains two main steps: diversity generation and consensus function. The common methods of diversity generation are heterogeneous ensemble, homogeneous ensemble, resampling and projecting data into random subspace [21, 23–27]. The common consensus functions are voting approach, pairwise approach, graph-based approach, feature-based approach, and etc. ([28, 29]; Lam-On and Boongoen 2015; [30, 31]). In addition, the research on clustering ensemble is focused on the selection of solutions. Recently, Yousefnezhad et al. defined selective clustering ensemble in [16]. Selective clustering ensemble selects solutions from solutions library according to some benchmark. Then the consensus function only merges the selected solutions rather than combines all solutions.

Among different aspects to consensus, it is more introduced as a kind of machine learning system consisting of a group of individual and parallel work models. In order to obtain a unique solution to a problem, all inputs of the model are integrated through a decision integration strategy [32]. Based on the above definitions, ensembles methods were initially utilized in the supervised learning. Researchers sought to apply a unique unsupervised learning paradigm due to its successful clustering since the past decade. There are two reasons for application of these methods particularly in clustering problems. First, there is not any knowledge about expected types of specific features and structures or about an expected desired data solution [33, 34].

The clustering ensembles cover the extremely data-dependent performance of a large number of basic clustering algorithms. An appropriate data partition can be provided by a certain clustering algorithm, but it can have a weak performance for other types of data. Clustering algorithms may generally face two serious challenges; first, different types of algorithms may find similar dataset with unequal structures; For instance, k-means is probably the most suitable algorithm for globular clusters. Second, a unit clustering algorithm with a variety of parameters may indicate various structures in the same dataset. Accordingly, it is highly challenging to select the best type of clustering algorithm for a target dataset. A new method based on the integration of some basic partitions is another mechanism solving this challenge. This method is called the cluster ensemble.

A clustering ensemble, called the combined clustering, easily defines a similar clustering method that it in fact combines

several clustering partitions in a single consensus partition [9]. A clustering algorithm will be efficient if it provides higher reliability, accuracy and consistency than other individual algorithms of clustering.

Nevertheless, it is impossible to provide a clear conceptual definition for transforming the supervised to unsupervised learning since there are various problems in performing an ensemble clustering. Despite these issues, it is difficult to integrate clusters that are created by clustering partitions (each of these partitions are called a member).

It cannot be done in an ensemble method through a voting straightforward or average-based method because it is feasible in an ensemble classifier; hence, there is a need for complex consensus functions. There are not any efficient and consensus functions in the real world, but it will be possible over time.

The present paper sought to analyze and describe the role of a three-layer aggregator in framework of a clustering ensemble. First, we converted each cluster into a cluster representation, and then we measured the rate of similarity between initial clusters and recombined the highly similar clusters in order to create final k clusters. In the third step, we selected desired clusters for uncertain objects. We created the last partition through successive uncertain objects with the minimum impact on their quality. In summary, the present paper provided an ensemble clustering algorithm on the basis of integrating basic clustering partitions with a three-step process. We then examined this method on numerous datasets and compared it with several methods of the state of the art.

While it has been shown that soft clustering algorithms are capable of producing better clustering results [35], fuzzy clustering ensemble has a little share in clustering ensemble research. Fuzzy clustering ensemble, while being severely ignored, has turned out as a hot topic recently ([19, 23, 36–42]; Li and Chen, 2018). It has been shown that a clustering ensemble framework with fuzzy clusters can outperform the same clustering ensemble framework with hard clusters [39]. Therefore, we focus on extending the proposed method to be applicable to fuzzy clusters.

The paper can be considered as a descendant of DICLENS. However, it is completely different from DICLENS in terms of the following 3 items: (1) they use different base clustering algorithms, (2) they define clustering similarity different from each other, and finally (3) they use different mechanisms to assign data objects to meta clusters. In fact, this method can be also considered as an advanced version of the method in 2015 [43], and it presents a dual similar clustering ensemble method. DSCE algorithm was developed in terms of three aspects. First, sustainability of the DSCE in generation of the final clustering result was improved through the pre-defined k value even though members had a variety of clusters. Second, the impact of parameters namely, α_1 and α_2 , was decreased on the quality of results through a suitable adaptive strategy for their

threshold values. Third, the similarity of adjacent objects was calculated for uncertain objects, and thus no information was removed when an inappropriate cluster was omitted. The proposed method, while taking all advantages of the DSCE algorithm, removes its disadvantages: (1) the parameters namely, α_1 and α_2 , have been removed, and (2) fuzzy clusters can also be used instead of hard clusters. Therefore, the main motivation of the paper is to propose a fuzzy clustering ensemble, while using the cluster clustering idea of DSCE and DICLENS. The proposed method is also better than DSCE because it has no parameter to be tuned and is also applicable to fuzzy clusters. It is also better than DICLENS, because of its flexibility in cluster-cluster similarity definition and its applicability to fuzzy clusters.

Other sections of research were organized as follows. In the Section 2, we introduced the clustering ensemble method as well as a framework for the general clustering ensemble. In the Section 3, we explained the relevant studies; and described details of an appropriate clustering ensemble method and also its different stages in the Section 4. Section 5 presented empirical studies and section 6 provided results of real datasets. Section 7 presented the analyses of parameters and time complexity of our method, and ultimately, the Section 8 presented the conclusion.

2 Clustering ensemble

2.1 Problem definition

Dataset A dataset is denoted by D and is a set of data points. The i th data point in dataset D is denoted by D_i . The size of dataset D is denoted by $|D|$.

Partition A dataset partition or a dataset clustering over dataset D is denoted by $\phi(D, c)$ and is defined as a set of clusters, i.e. $\phi(D, c) = \{C(\phi(D, c), 1), C(\phi(D, c), 2), C(\phi(D, c), 3), \dots, C(\phi(D, c), c)\}$ where c is number of clusters, and $C(\phi(D, c), i)$ is the i th cluster in clustering $\phi(D, c)$.

Fuzzy cluster A cluster in a dataset partition or a dataset clustering $\phi(D, c)$ is denoted by $C(\phi(D, c), i)$ and is defined as a fuzzy set as follows

$$C(\phi(D, c), i) = \left\{ \left(D_1, \mu_{C(\phi(D, c), i)}(D_1) \right), \left(D_2, \mu_{C(\phi(D, c), i)}(D_2) \right), \dots, \left(D_{|D|}, \mu_{C(\phi(D, c), i)}(D_{|D|}) \right) \right\} \tag{1}$$

where $\mu_{C(\phi(D, c), i)}(D_j)$ is membership function indicating how much the j th data point belongs to the i th cluster in clustering $\phi(D, c)$. It worthy to be mentioned that for all $j \in \{1, 2, \dots, |D|\}$

we have $\sum_{i=1}^c \mu_{C(\phi(D, c), i)}(D_j) = 1$ and for all $j \in \{1, 2, \dots, |D|\}$ and $i \in \{1, 2, \dots, c\}$ we have $\mu_{C(\phi(D, c), i)}(D_j) \geq 0$.

Clustering ensemble An ensemble is a set of dataset clusterings over a dataset D and it is denoted by $\Phi(D, b) = \{\phi_1(D, c_1), \phi_2(D, c_2), \dots, \phi_b(D, c_b)\}$ where b is ensemble size, $\phi_j(D, c_j)$ is the j th clustering on the dataset D into c_j clusters.

It is worthy to be mentioned that the number of clusters in the members of ensemble may differ from each other and from the real number of clusters; i.e. $c_i \neq c_j \neq c$.

In the clustering ensemble field, there are two problems. The first one is to find a diverse ensemble; this problem is considered as ensemble generation problem. The second one is to find an aggregated clustering, called a consensus partition, which has the most agreement among the ensemble members; this problem is considered as consensus function problem. The consensus partition is denoted by $\phi_*(D, c)$. The mentioned two problems in clustering ensemble have been discussed in the next two subsections.

2.2 Ensemble creation problem

The first step for framework of clustering ensemble is the creation of ensemble members. This step aims to generate basic b fuzzy clustering results of $\{\phi_1(D, c_1), \phi_2(D, c_2), \dots, \phi_b(D, c_b)\}$. Furthermore, produced members should be different in order to uncover different data structures. These members can increase the performance of consensus clustering, and thus it is essential applying one or more creation methods to access to logical quality with this diversity.

Several researchers applied the problem-based creation method. Strehl and Ghosh [9] utilized some clustering algorithms with numerous data sub-spaces, applied several features, and generated members on them for large-dimension data. They also applied a sub-sampling for objects. Fern and Brodley [44] applied a same approach; for instance, data partitions were obtained for implementing data in a variety of sub-spaces.

Breiman [45] introduced a re-sampling method in machine learning, and it was then developed and improved by Minaei-Bidgoli et al. [46]. Parvin and Minaei-Bidgoli [41] applied an ensemble generation method by the fuzzy clustering. We can generate a different accurate ensemble through the Fuzzy Weighted Locally Adaptive Clustering (FWLAC) algorithm. According to bagging and boosting algorithms in the classification, Minaei-Bidgoli et al. [46] and Parvin et al. [47] also investigated the weighing and non-weighing-based sampling methods for the ensemble generation.

To make the use of a subset of initial members in the ensemble, Alizadeh et al. [23] introduced an ensemble clustering framework instead of applying all previous cases. Accordingly, the qualitative metric, the Normalized Mutual Information (NMI) was utilized to determine the target

Table 1 A representative assumptive ensemble with the size of five base clustering members

		D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}	D_{13}
$\phi_1(D, 2)$	$C(\phi_1(D, 2), 1)$	0.3	0.2	0.1	0.2	0.8	0.4	0.1	0.9	0.9	0.7	0.2	0.7	0
	$C(\phi_1(D, 2), 2)$	0.7	0.8	0.9	0.8	0.2	0.6	0.9	0.1	0.1	0.3	0.8	0.3	1
$\phi_2(D, 2)$	$C(\phi_2(D, 2), 1)$	0.7	0.1	0	0.7	0.3	0.1	0.9	1	0.9	0.2	0.8	0.9	0.4
	$C(\phi_2(D, 2), 2)$	0.3	0.9	1	0.3	0.7	0.9	0.1	0	0.1	0.8	0.2	0.1	0.6
$\phi_3(D, 3)$	$C(\phi_3(D, 3), 1)$	0	0	0	0.1	0	0	0	0	0	0.8	0.9	0.8	0
	$C(\phi_3(D, 3), 2)$	0.4	0.2	0.8	0.7	0.8	0.8	0.2	0.4	0.8	0.1	0.1	0.1	0.6
	$C(\phi_3(D, 3), 3)$	0.6	0.8	0.2	0.2	0.2	0.2	0.8	0.6	0.2	0.1	0	0.1	0.4
$\phi_4(D, 3)$	$C(\phi_4(D, 3), 1)$	0.7	0.8	0.2	0.2	0.3	0.1	0.9	0.6	0.2	0	0	0	0.5
	$C(\phi_4(D, 3), 2)$	0.3	0.2	0.7	0.8	0.7	0.9	0.1	0.3	0.8	0.1	0	0.1	0.5
	$C(\phi_4(D, 3), 3)$	0	0	0.1	0	0	0	0	0.1	0	0.9	1	0.9	0
$\phi_5(D, 3)$	$C(\phi_5(D, 3), 1)$	0.1	0.2	0.1	0	0.7	0.8	0.6	0.1	0	0.8	0.2	0.1	0.2
	$C(\phi_5(D, 3), 2)$	0.1	0.1	0	0.1	0.3	0.1	0.3	0.8	0.9	0	0.8	0.8	0.2
	$C(\phi_5(D, 3), 3)$	0.8	0.7	0.9	0.9	0	0.1	0.1	0.1	0.1	0.2	0	0.1	0.6

clustering. However, setting an appropriate NMI threshold depends on data and needs the domain knowledge.

The kmeans clustering algorithm is probably the highly multifaceted clustering algorithm for generating ensemble components due to its simplicity and fastness ([47]; and [46]). For instance, Alizadeh et al. [48] applied the kmeans clustering algorithm consisting of cluster centers with random start or a random c (cluster number) at a target clustering range. Strehl and Ghosh [9] made use of various distance functions for members to take the advantage of a graph clustering algorithm. Furthermore, Topchy et al. [49] applied a weak clustering algorithm with slightly accurate clustering outcomes than the random predictor.

Iam-on et al. [50] used a variety of methods consisting of several kmeans with a fixed number of clusters for all basic partitions as well as selecting a cluster number randomly from a set of $\{2, 3, \dots, \sqrt{|D|}\}$. Iam-on et al. [51] also made use of different ensemble creation methods.

Other researchers utilized a variety of basic clustering algorithms for generating ensemble elements. They also applied kmeans and hierarchical clustering algorithms as basic algorithms ([52]; and [53]).

Alizadeh et al. created ensembles based on the concept of the Wisdom of Crowds [48]. This phenomenon, which is established in social sciences, refers to the application of criteria in the group behavior. Group decisions are probably better than individual members due to the fulfilled criteria. Accordingly, the Wisdom of Crowds Cluster Ensemble (WOCCE) was introduced due to its ability to analyze necessary conditions for ensembles in order to indicate their collective wisdom. This process also covered diverse conditions in ensembles, decentralization of generated base clustering, and independence condition among base algorithms.

Akbari et al. [21] proposed the Hierarchical Cluster Ensemble Selection (HCES) method and the diversity measure for determining effects of diversity and quality on final results since there was an uncertain relationship between the diversity and quality. They utilized single, average, and full linkage agglomerative techniques for hierarchical members.

2.3 Consensus function

An consensus function is a function that gives the final consensus partition of $\phi_*(D, c)$ and gets the number of desired clusters, i.e. c , and ensemble $\Phi(D, b)$ (i.e. base clustering elements) as their inputs. Most researchers consider the consensus function step as the most important components in ensembles.

Alqurashi and Wang [54] analyzed a number of consensus functions. There are generally two groups of consensus functions in the clustering ensembles as follows: (a) heuristic category, and (b) the median clustering category.

The first group seeks to find an extracted consensus partition heuristically, for example from the Co-Association (CA) matrix. This category consists of direct approaches such as the conventional voting method; feature-based approaches such as Bayesian consensus clustering method; graph-based approaches such as the hyper-graph partitioning algorithm; and pairwise similarity-based approaches such as the CA-based method.

Some of the cluster ensemble methods have designed solutions for problem of integrating initial partitions as optimization processes; hence, the information is in bit strings of 0–1 among ensemble members. Similarly, Alizadeh et al. [23] introduced the objective and nonlinear *constraint* function called the fuzzy string objective function (FSOF) in order to find a median partition by maximizing the agreement between ensemble members and at the same time minimizing the disagreement.

Table 2 The similarity matrix computed between the fuzzy clusters of the ensemble presented in Table 1

	$C(\phi_1(D, 2), 1)$	$C(\phi_1(D, 2), 2)$	$C(\phi_2(D, 2), 1)$	$C(\phi_2(D, 2), 2)$	$C(\phi_3(D, 3), 1)$	$C(\phi_3(D, 3), 2)$	$C(\phi_3(D, 3), 3)$	$C(\phi_4(D, 3), 1)$	$C(\phi_4(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$
$\phi_1(D, 2)$	1	-1	0.29	-0.29	0.15	0.05	-0.26	-0.31	0.07	0.18	0.11	0.53	-0.6
$\phi_2(D, 2)$	$C(\phi_1(D, 2), 2)$	1	-0.29	0.29	-0.15	-0.05	0.26	0.31	-0.07	-0.18	-0.11	-0.53	0.6
$\phi_3(D, 3)$	$C(\phi_2(D, 2), 1)$	$C(\phi_2(D, 2), 2)$	1	-1	0.17	-0.3	0.1	0.13	-0.32	0.16	-0.44	0.76	-0.35
$\phi_4(D, 3)$	$C(\phi_3(D, 3), 1)$	$C(\phi_3(D, 3), 2)$	$C(\phi_3(D, 3), 3)$	1	-0.17	0.3	-0.1	-0.13	0.32	-0.16	0.44	-0.76	0.35
	$C(\phi_4(D, 3), 1)$	$C(\phi_4(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	1	-0.68	-0.2	-0.64	0.97	-0.68	0.09	0.31	-0.37
	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$	$C(\phi_4(D, 3), 1)$	$C(\phi_4(D, 3), 2)$	1	0.98	1	-0.21	-0.62	-0.05	-0.2	0.23
	$C(\phi_4(D, 3), 1)$	$C(\phi_4(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	1	-0.21	1	-0.64	-0.02	-0.2	0.25
	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$	$C(\phi_4(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_4(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	1	-0.64	1	0.09	-0.23	0.27
	$C(\phi_4(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_4(D, 3), 1)$	$C(\phi_5(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	1	0.09	1	-0.39	-0.47
	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	$C(\phi_4(D, 3), 3)$	$C(\phi_4(D, 3), 1)$	$C(\phi_4(D, 3), 2)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$	$C(\phi_4(D, 3), 1)$	1	-0.39	1	-0.63
	$C(\phi_5(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_4(D, 3), 1)$	$C(\phi_4(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_5(D, 3), 3)$	$C(\phi_4(D, 3), 2)$	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	1	-0.47	1
	$C(\phi_4(D, 3), 1)$	$C(\phi_4(D, 3), 2)$	$C(\phi_4(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$	$C(\phi_4(D, 3), 3)$	$C(\phi_5(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$	$C(\phi_4(D, 3), 1)$	$C(\phi_5(D, 3), 2)$	$C(\phi_5(D, 3), 3)$

3 Related work

Mapping ensembles into a CA matrix [9] is the most widely used clustering ensemble method. The consensus partition can be then obtained by application of a hierarchical clustering method.

The CA matrix-based clustering ensemble methods prevent the label matching problem by mapping consensus members inside a new representation with computed similarity matrix between a pair of objects as members with the same pair of objects. Fred and Jain [55] obtained the consensus clustering through utilizing average and single-linkage hierarchical clustering algorithms. When there is pairwise information in clustering ensembles, the CA is suitable for clustering as the CA merely considers the relationship of object pairs in consensus members.

Combining boosting and bagging strengths, Yang and Jiang [56] introduced a novel hybrid sampling for cluster ensembles.

Furthermore, Bai et al. [57] introduced a new information-entropy based community description model. They found that the information theory was an appropriate tool for analyzing clusters.

Strehl and Ghosh [9] provided a graph-based consensus function called the Cluster-based Similarity Partitioning Algorithm (CSPA) that considered the CA matrix as a hyper graph. Each row (data point) was considered as a node; and each column as a hyper edge in which the entry matrix value indicated the node membership value to the hyper edge. Afterwards, the hyper graph was partitioned into target numbers of clusters through a graph-partitioning algorithm like HMETIS [58]. They also introduced other hyper-graph partitioning algorithms such as the Meta Clustering Algorithm (MCLA) and the Hyper-Graph Partitioning Algorithm (HGPA) considering each data as a node and each cluster as a hyper-edge.

Fern and Brodley [59] introduced a membership binary matrix for a bipartite graph according to the CSPA, and then applied a graph-partitioning algorithm such as HMETIS for extracting consensus partitions. Their method was called the hybrid bipartite graph formulation (HBGF) algorithm.

Iam-on et al. [28] introduced another advanced CA on the basis of clustering ensemble in order to apply relationships of clusters and redefine the CA matrix. For generation of the final consensus partition, a common clustering algorithm applied this similarity matrix.

Mimaroglu and Aksehirli [67] introduced a new clustering ensemble approach on the basis of a new definition of the similarity. This approach applied a matrix like CA, but each entry determined the similarity of clusters.

This can be measured according to the co-association matrix between two objects in two various clusters indicating the similarity of clusters.

Alqurushi and Wang (2014) introduced a new approach to the clustering ensemble that applied the neighboring relationships of pairs of objects as well as real relationships of pairs in

order to measure the similarity matrix. Huang et al. [60] examined and developed the aforementioned approach.

Huang et al. [61] provided an approach to examination of a clustering ensemble problem as a linear binary programming problem for automatic determination of number of clusters.

4 Proposed consensus function

A consensus function is a very important element in clustering ensemble. Indeed, it has a great effect on the efficacy of clustering ensemble. We want to propose a consensus function which is efficient and fast. The proposed consensus function inspires from the co-association based clustering ensemble, but it also bypasses its high computational cost through hierarchical cluster clustering. It first defines a between-cluster similarity measure. Using the mentioned between-cluster similarity measure, a similarity matrix is generated where its columns and rows (in spite of co-association matrix) are base clusters produced in the ensemble generation phase. Then the clusters are merged into new defined clusters. A new cluster-sample similarity measure is also introduced to determine how much a data sample belongs to a newly generated clusters.

Between-Cluster Similarity measure is denoted by $Sim_{P,Q}^{CC}$ where P and Q are two fuzzy base clusters. It is defined based on eq. 2.

$$Sim_{P,Q}^{CC} = \frac{\sum_{j=1}^{|D|} \mu_P(D_j) \mu_Q(D_j) - |D| \times \frac{\sum_{j=1}^{|D|} \mu_P(D_j)}{|D|} \times \frac{\sum_{j=1}^{|D|} \mu_Q(D_j)}{|D|}}{(|D|-1) \times \sqrt{\frac{\sum_{j=1}^{|D|} \left(\mu_P(D_j) - \frac{\sum_{j=1}^{|D|} \mu_P(D_j)}{|D|} \right)^2}{(|D|-1)} \times \frac{\sum_{j=1}^{|D|} \left(\mu_Q(D_j) - \frac{\sum_{j=1}^{|D|} \mu_Q(D_j)}{|D|} \right)^2}{(|D|-1)}}}} = \frac{\sum_{j=1}^{|D|} \mu_P(D_j) \mu_Q(D_j) - \frac{\sum_{j=1}^{|D|} \mu_P(D_j) \times \sum_{j=1}^{|D|} \mu_Q(D_j)}{|D|}}{\sqrt{\frac{\sum_{j=1}^{|D|} \left(\mu_P(D_j) - \frac{\sum_{j=1}^{|D|} \mu_P(D_j)}{|D|} \right)^2}{(|D|-1)} \times \frac{\sum_{j=1}^{|D|} \left(\mu_Q(D_j) - \frac{\sum_{j=1}^{|D|} \mu_Q(D_j)}{|D|} \right)^2}{(|D|-1)}}}} = \frac{|D| \times \sum_{j=1}^{|D|} \mu_P(D_j) \mu_Q(D_j) - \sum_{j=1}^{|D|} \mu_P(D_j) \times \sum_{j=1}^{|D|} \mu_Q(D_j)}{|D| \times \sqrt{\frac{\sum_{j=1}^{|D|} \left(\mu_P(D_j) - \frac{\sum_{j=1}^{|D|} \mu_P(D_j)}{|D|} \right)^2}{(|D|-1)} \times \frac{\sum_{j=1}^{|D|} \left(\mu_Q(D_j) - \frac{\sum_{j=1}^{|D|} \mu_Q(D_j)}{|D|} \right)^2}{(|D|-1)}}}} \tag{2}$$

The mentioned similarity measure computes the correlation between two clusters [62]. The mentioned similarity measure returns a real number in the interval $[-1, 1]$. The mentioned similarity measure returns 0 for two completely different clusters, 1 for two similar clusters, -1 for two complementary clusters [63].

Assume the number of all clusters is denoted by a , therefore, it is defined based on eq. 3.

$$a = \sum_{i=1}^b c_i \tag{3}$$

A similarity matrix of size $a \times a$, where a is the number of clusters and it is computed based on equation 3, denoted by S_{ij} is defined based on equation 4.

$$S_{ij}^{CC} = Sim_{C(\phi_p(D,c_p),r),C(\phi_q(D,c_q),t)}^{CC} \tag{4}$$

where p and r are positive integers and computed based on equation 5.

$$i = \sum_{i=1}^{p-1} c_i + r \text{ s.t. } r \leq c_p \tag{5}$$

and q and t are positive integers and they are computed based on equation 6.

$$j = \sum_{i=1}^{q-1} c_i + t \text{ s.t. } t \leq c_q \tag{6}$$

An example has been presented here for more clarification.

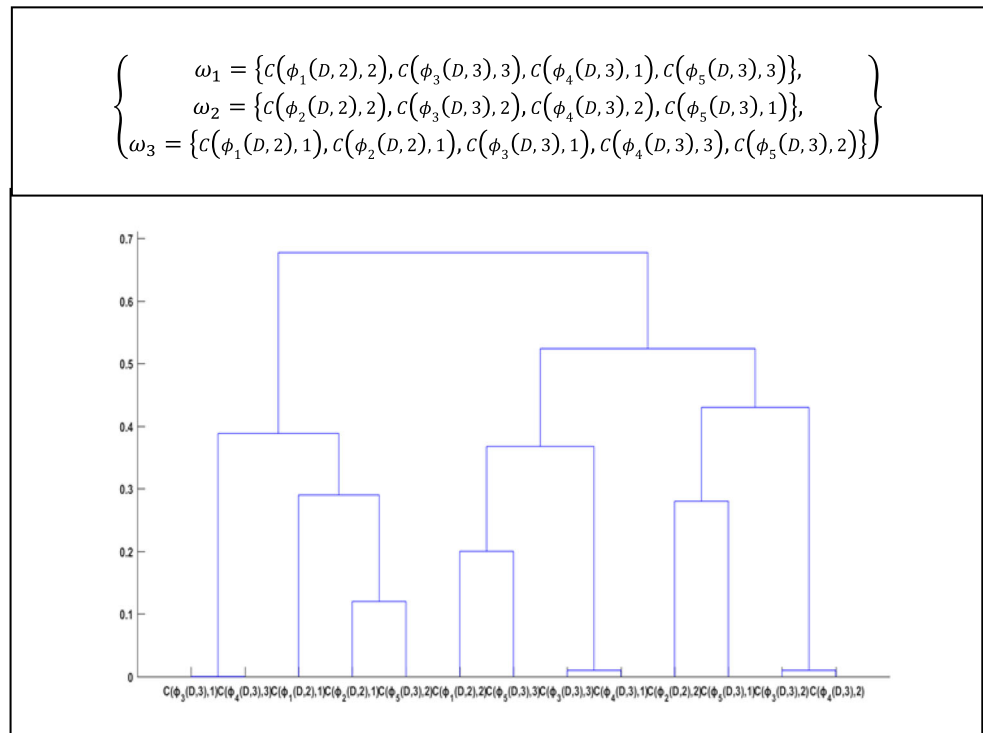
The similarity matrix S_{ij}^{CC} is transformed into a distance matrix in the next step. The distance matrix is denoted by \check{S}_{ij}^{CC} and is computed based on equation 7.

$$\check{S}_{ij}^{CC} = 1 - \frac{S_{ij}^{CC} + 1}{2} \tag{7}$$

Table 3 The distance (or dissimilarity) matrix of the base fuzzyclusters presented in Table 1

$\phi_1(D, 2)$	$C(\phi_1(D, 2), 1)$	0	1	0.35	0.65	0.42	0.47	0.63	0.65	0.46	0.41	0.45	0.23	0.8
	$C(\phi_1(D, 2), 2)$	1	0	0.65	0.35	0.58	0.53	0.37	0.35	0.54	0.59	0.55	0.77	0.2
$\phi_2(D, 2)$	$C(\phi_2(D, 2), 1)$	0.35	0.65	0	1	0.41	0.65	0.45	0.44	0.66	0.42	0.72	0.12	0.68
	$C(\phi_2(D, 2), 2)$	0.65	0.35	1	0	0.59	0.35	0.55	0.56	0.34	0.58	0.28	0.88	0.32
$\phi_3(D, 3)$	$C(\phi_3(D, 3), 1)$	0.42	0.58	0.41	0.59	0	0.84	0.8	0.82	0.81	0	0.45	0.34	0.69
	$C(\phi_3(D, 3), 2)$	0.47	0.53	0.65	0.35	0.84	0	0.6	0.56	0.01	0.84	0.52	0.6	0.39
	$C(\phi_3(D, 3), 3)$	0.63	0.37	0.45	0.55	0.8	0.6	0	0.01	0.62	0.79	0.54	0.6	0.38
$\phi_4(D, 3)$	$C(\phi_4(D, 3), 1)$	0.65	0.35	0.44	0.56	0.82	0.56	0.01	0	0.6	0.81	0.55	0.6	0.37
	$C(\phi_4(D, 3), 2)$	0.46	0.54	0.66	0.34	0.81	0.01	0.62	0.6	0	0.82	0.51	0.61	0.38
	$C(\phi_4(D, 3), 3)$	0.41	0.59	0.42	0.58	0	0.84	0.79	0.81	0.82	0	0.46	0.33	0.7
$\phi_5(D, 3)$	$C(\phi_5(D, 3), 1)$	0.45	0.55	0.72	0.28	0.45	0.52	0.54	0.55	0.51	0.46	0	0.69	0.74
	$C(\phi_5(D, 3), 2)$	0.23	0.77	0.12	0.88	0.34	0.6	0.6	0.6	0.61	0.33	0.69	0	0.81
	$C(\phi_5(D, 3), 3)$	0.8	0.2	0.68	0.32	0.69	0.39	0.38	0.37	0.38	0.7	0.74	0.81	0

Fig. 1 The output dendrogram and clusters of the average-linkage hierarchical clustering algorithm on the distance matrix depicted in Table 3



After obtaining the distance matrix, a partition is extracted on the base fuzzy clusters by applying a hierarchical average linkage clustering algorithm on the mentioned matrix; indeed, each of the base fuzzy clusters in the ensemble is considered as a data point. Each cluster of the base fuzzy clusters is denoted by ω_i . The number of clusters in this step is equal to the real number of clusters in the ground truth labels. After obtaining the clusters of the base fuzzy clusters, the step one is completed.

Example 1 Assume an assumptive dataset of size 13. Assume also we have run fuzzy kmeans on the mentioned assumptive dataset 5 times and the number of clusters requested from the clustering algorithm is 2, 2, 3, 3, and 3 consecutively. Assume the toy fuzzy ensemble on the mentioned assumptive dataset is the one which is depicted in Table 1. The similarity matrix computed between the fuzzy clusters of the ensemble presented in Table 1 is depicted in Table 2. The distance matrix has

Table 4 The similarity measure, the object-cluster membership matrix and consensus partition of the fuzzy ensemble presented in Table 1

		D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}	D_{13}
ω_1	$C(\phi_1(D, 2), 2)$	0.7	0.8	0.9	0.8	0.2	0.6	0.9	0.1	0.1	0.3	0.8	0.3	1
	$C(\phi_3(D, 3), 3)$	0.6	0.8	0.2	0.2	0.2	0.2	0.8	0.6	0.2	0.1	0	0.1	0.4
	$C(\phi_4(D, 3), 1)$	0.7	0.8	0.2	0.2	0.3	0.1	0.9	0.6	0.2	0	0	0	0.5
	$C(\phi_5(D, 3), 3)$	0.8	0.7	0.9	0.9	0	0.1	0.1	0.1	0.1	0.2	0	0.1	0.6
ω_2	$C(\phi_2(D, 2), 2)$	0.3	0.9	1	0.3	0.7	0.9	0.1	0	0.1	0.8	0.2	0.1	0.6
	$C(\phi_3(D, 3), 2)$	0.4	0.2	0.8	0.7	0.8	0.8	0.2	0.4	0.8	0.1	0.1	0.1	0.6
	$C(\phi_4(D, 3), 2)$	0.3	0.2	0.7	0.8	0.7	0.9	0.1	0.3	0.8	0.1	0	0.1	0.5
	$C(\phi_5(D, 3), 1)$	0.1	0.2	0.1	0	0.7	0.8	0.6	0.1	0	0.8	0.2	0.1	0.2
ω_3	$C(\phi_1(D, 2), 1)$	0.3	0.2	0.1	0.2	0.8	0.4	0.1	0.9	0.9	0.7	0.2	0.7	0
	$C(\phi_2(D, 2), 1)$	0.7	0.1	0	0.7	0.3	0.1	0.9	1	0.9	0.2	0.8	0.9	0.4
	$C(\phi_3(D, 3), 1)$	0	0	0	0.1	0	0	0	0	0	0.8	0.9	0.8	0
	$C(\phi_4(D, 3), 3)$	0	0	0.1	0	0	0	0	0.1	0	0.9	1	0.9	0
	$C(\phi_5(D, 3), 2)$	0.1	0.1	0	0.1	0.3	0.1	0.3	0.8	0.9	0	0.8	0.8	0.2
	$Sim_{OC}^{D_1}$	0.70	0.78	0.55	0.53	0.18	0.25	0.68	0.35	0.15	0.15	0.20	0.13	0.63
	$Sim_{OC}^{D_2}$	0.28	0.38	0.65	0.45	0.73	0.85	0.25	0.20	0.43	0.45	0.13	0.10	0.48
	$Sim_{OC}^{D_3}$	0.22	0.08	0.04	0.22	0.28	0.12	0.26	0.56	0.54	0.52	0.74	0.82	0.12
	S_{I1}	0.58	0.63	0.44	0.44	0.15	0.20	0.57	0.32	0.13	0.13	0.19	0.12	0.51
	S_{I2}	0.23	0.31	0.52	0.38	0.61	0.70	0.21	0.18	0.38	0.40	0.12	0.10	0.39
	S_{I3}	0.18	0.06	0.03	0.18	0.24	0.10	0.22	0.50	0.48	0.46	0.69	0.78	0.10
	Consensus Partition	1	1	2	1	2	2	1	3	3	3	3	3	1

Bold entries discriminate different partitions

Table 5 Detailed specifications of the employed datasets

ID	Name	# of Samples	# of Features	# of Clusters
BC	Breast Cancer	683	9	2
B	Bupa	345	6	2
G	Glass	214	9	6
W	Wine	178	13	3
Y	Yeast	1484	8	10
I	Iris	150	4	3
SAH	SAHeart	462	9	2
Ion	Ionosphere	351	34	2
Gx	Galaxy	323	4	7
HR	Half-Ring	400	2	2
A1	Artificial ₁	300	2	3
A2	Artificial ₂	300	2	3
A3	Artificial ₃	300	2	3
LS	Landsat-Satellite	6435	36	6
IS	ISOLET	7797	617	26
U	USPS	11,000	256	10
LR	Letter-Recognition	20,000	16	26

been computed based on similarity matrix and it is depicted in Table 3. Then the output dendrogram and clusters of the average-linkage hierarchical clustering algorithm on the distance matrix depicted in Table 3 have been represented in Fig.

1. According to the output clustering, three clusters have been emerged among the fuzzy base clusters.

The next step is to assign each data point to exactly one of the clusters defined on the base fuzzy clusters. We define an object-cluster membership matrix as follows.

$$S_{ij}^{OC} = \frac{Sim_{D_i, \omega_j}^{OC}}{\sum_{k=1}^c Sim_{D_i, \omega_k}^{OC}} \quad (8)$$

where Sim_{D_i, ω_j}^{OC} is a similarity measure between an object and a cluster and is defined based on equation 9.

$$Sim_{D_i, \omega_j}^{OC} = \frac{\sum_{P \in \omega_j} \mu_P(D_j)}{\sum_{P \in \omega_j} 1} \quad (9)$$

Example II Consider the previous example. The new clusters of ω_1, ω_2 and ω_3 of the base fuzzy clusters are depicted in Table 4. The similarity measure and the object-cluster membership matrix have been also depicted in the Table 4. In the last row of the Table 4, the consensus partition has been depicted.

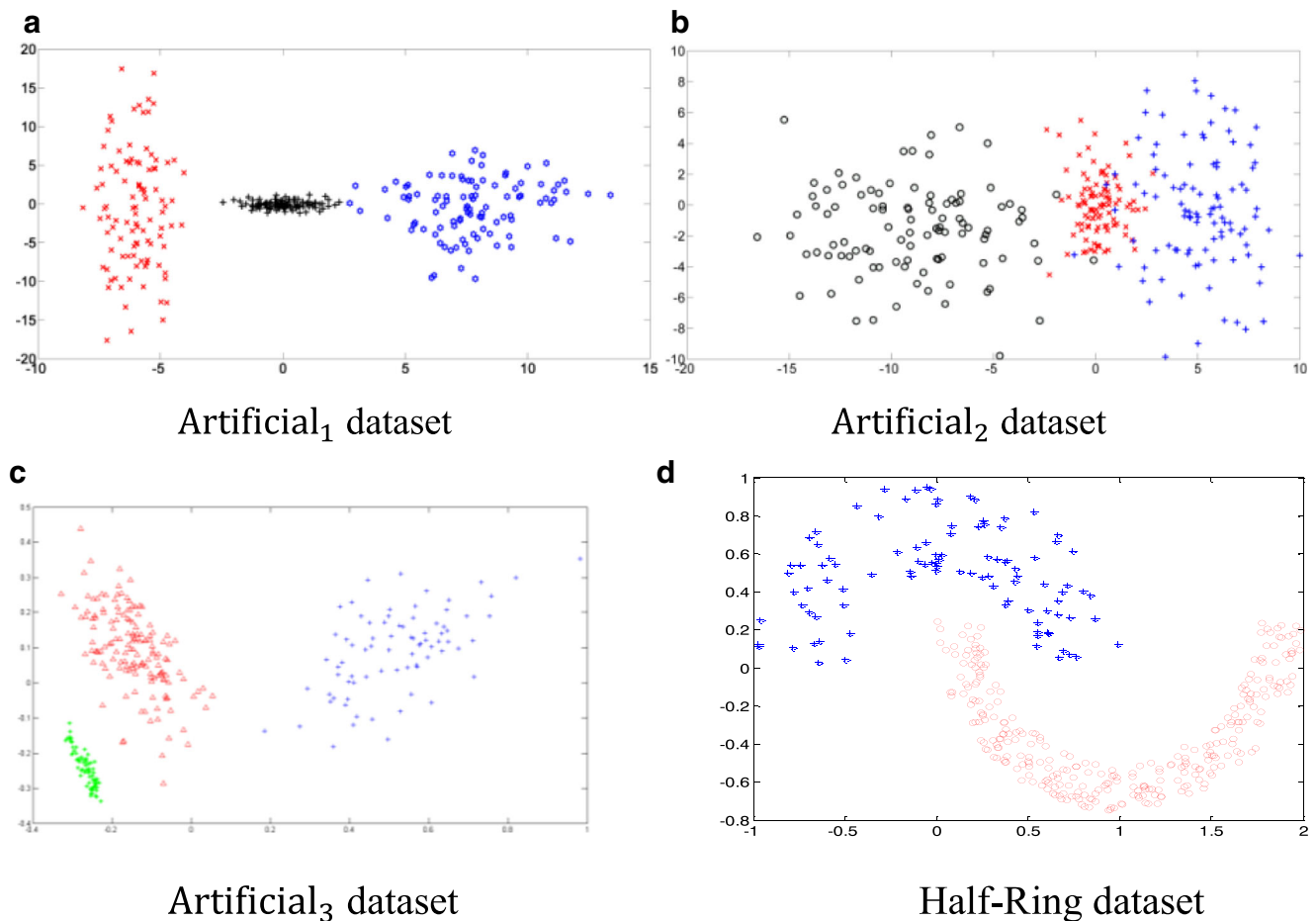
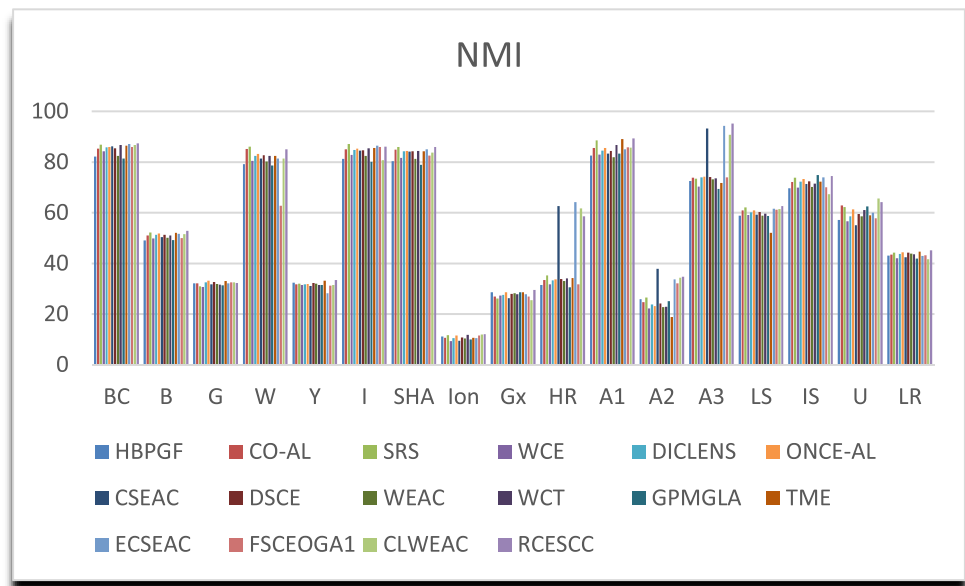
**Fig. 2** Artificial datasets

Fig. 3 The performances of different methods in terms of NMI



To sum up the proposed method, it first produces a number of fuzzy clusterings. Then, it transforms them into a number of fuzzy clusters. After that, it makes a cluster-cluster similarity matrix whose entries indicates how much two fuzzy clusters are similar. This similarity has been estimated by fuzzy correlation. In the next step, the fuzzy clusters have been clustered into a predefined number of meta clusters according to a hierarchical clustering algorithm. In the last step, each data is assigned to exactly one meta cluster according to mean aggregation.

5 Experimental study

5.1 Datasets

The proposed method, i.e. RCESCC, has been evaluated on a set of real and artificial datasets. The real datasets, except USPS dataset which is from [64], are from the UCI machine learning repository [65]. The detailed specifications of the employed datasets are presented in Table 5. All of the artificial datasets include 2 features and 3 classes. The artificial datasets have been plotted in Fig. 2. The artificial datasets Gaussian datasets. All datasets are standardized. It means that any feature in any datasets is transformed into a new feature which comes from $N(0, 1)$ distribution. Features with missing values are also removed from the datasets.

5.2 Parameter setting

Like the approach employed by Ren et al. [66], an ensemble of size 200, i.e. $b = 200$, is produced. A set of 100 subsets is

prepared, each of them is extracted by a bootstrapped subsampling out of the given dataset. Each of them contains 80% of all data points in the dataset. A fuzzy cmeans algorithm is run on each of bag, and then for the missing data points in the bag, according to distance of the data point to the representators of fuzzy clusters, the membership values of the data point to all

Table 6 Performances of different state of the art methods comparing with performance of the proposed method, i.e. RCESCC, in terms of NMI. A triple *W-D-L* indicates the numbers of that the proposed method wins/draws with/loses to other methods validated by paired t-test

Type	Name	W-D-L	Average
Hard	HBPGF	14-3-0	53.96
Hard	CO-AL	15-2-0	55.86
Hard	SRS	7-9-1	56.76
Hard	WCE	17-0-0	53.68
Hard	DICLENS	16-1-0	55.33
Hard	ONCE-AL	11-5-1	56.00
Fuzzy	CSEAC	14-1-2	58.23
Hard	DSCE	15-2-0	55.57
Hard	WEAC	16-1-0	54.17
Hard	WCT	13-2-2	55.60
Hard	GPMGLA	14-1-2	53.85
Hard	TME	10-7-0	55.17
Fuzzy	ECSEAC	9-7-1	59.18
Fuzzy	FSCEOGA1	14-3-0	54.42
Fuzzy	CLWEAC	11-4-2	58.45
Fuzzy	RCEIFBC	-	60.50

fuzzy clusters are computed. Therefore, it means that a fuzzy partition is obtained for each bag on all of data points. An additional set of 100 base fuzzy clusterings on the given dataset are created through applying fuzzy cmeans on 100 subspaces of the given dataset. Each subspace includes half of randomly selected features.

The recent works which have been used as baseline methods include *HBPGF* [59], *CO-AL* [55], *SRS* (Iam-On et al., 2008), *WCT* [28], *DICLENS* [67], *ONCE-AL*[54], *CSEAC* [68], *DSCE* [43], *WCE* [48], *WEAC* [69], *GPMGLA* [69], *TME* [29], *ECSEAC* [41], *DSCE* [43], *FSCEOGA1* [23] and *CLWEAC* [39]. All reported results in the paper are averaged over 100 independent runs. All methods use their default values mentioned in their corresponding papers for their parameters. Ensemble size is 200 throughout all the paper experiments.

5.3 Evaluation metrics

Four different metrics are used in the experimental results: Adjust Rand Index (ARI), Normalized Mutual Information (NMI), Accuracy (ACC) and F-Measure (FM). ARI is defined based on equation 10.

$ARI(P^*, L^t)$

$$ARI(P^*, L^t) = \frac{\sum_{i=1}^c \sum_{j=1}^c \binom{n_{ij}}{2} - \frac{\sum_{i=1}^k \binom{n_i^*}{2} \times \sum_{j=1}^k \binom{n_j^t}{2}}{\binom{n}{2}}}{\frac{\sum_{i=1}^c \binom{n_i^*}{2} + \sum_{j=1}^c \binom{n_j^t}{2}}{2} - \frac{\sum_{i=1}^c \binom{n_i^*}{2} \times \sum_{j=1}^c \binom{n_j^t}{2}}{\binom{n}{2}}} \quad (10)$$

where P^* is the consensus partition, L^t is the ground-truth labels of the given dataset, n is the total number of objects in given dataset X , n_{ij} is the number of objects in the intersection of the i th cluster of P^* and the j th cluster of L^t , n_i^* and n_j^t are the number of objects in the i th cluster in P^* and the number of objects in the j th cluster in L^t respectively. The expression $\binom{n}{2}$ stands for the binomial coefficient.

NMI is defined as follows:

$$NMI(P^*, L^t) = \frac{\sum_{i=1}^c \sum_{j=1}^c n_{ij} \log_2 \left(\frac{nn_{ij}}{n_i^* n_j^t} \right)}{\sqrt{\sum_{i=1}^c n_i^* \log_2 \left(\frac{n_i^*}{n} \right) \times \sum_{i=1}^c n_j^t \log_2 \left(\frac{n_j^t}{n} \right)}} \quad (11)$$

ACC is another metric based on that to evaluate quality of the different partitions. But to compute this metric, the partitions should be relabeled in order to match maximally with the ground-truth labels. This problem is handled by the Hungarian algorithm [70]. FM is the last criterion employed in this paper to assess quality of a partition and is defined based on equation 12.

$$FM(P^*, L^t) = \max_{\sigma} \sum_{j=1}^c \frac{2n_{\sigma(j)}^* \frac{n_{\sigma(j)j}}{n_j^t} \times \frac{n_{\sigma(j)j}}{n_{\sigma(j)}^*}}{n \left(\frac{n_{\sigma(j)j}}{n_j^t} + \frac{n_{\sigma(j)j}}{n_{\sigma(j)}^*} \right)} \quad (12)$$

where σ is a permutation of $\{1, 2, \dots, c\}$ and $\sigma(i)$ is the i th value in the permutation σ , i.e. $\forall i \in \{1, 2, \dots, c\} : \sigma(i) \in \{1, 2, \dots, c\}$ and $\forall i, j \in \{1, 2, \dots, c\} : (\sigma(i) = \sigma(j)) \rightarrow (i = j)$.

Fig. 4 The performances of different methods in terms of ARI

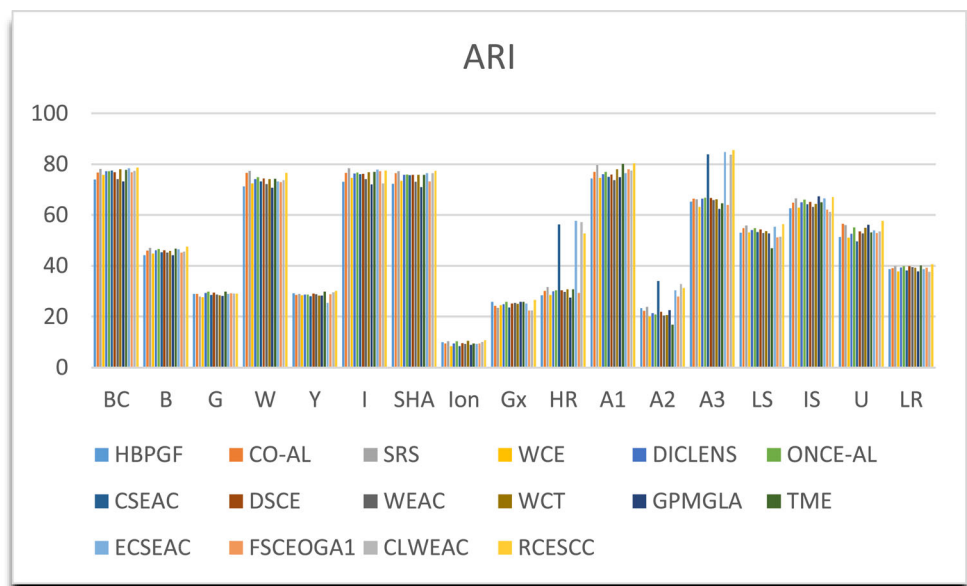
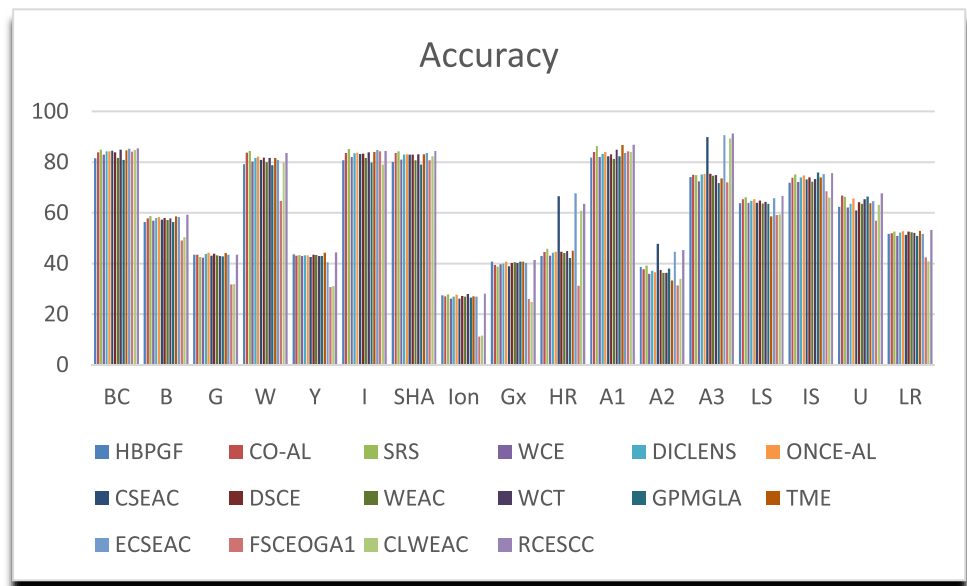


Fig. 5 The performances of different methods in terms of accuracy



Each of the mentioned criteria (ARI, NMI, FM, and ACC) is bounded in [0, 1]. While the values near to zero indicate the two input partitions are the independent clusterings, the values near to one indicate the two input partitions are the identical clusterings.

5.4 Experiments

Figure 3 depicts the performance of proposed method comparing with the state of the art methods in terms of NMI. The numbers of times that the proposed method wins/draws with/loses to other methods validated by paired t-test are summarized in Table 6. It is notable that we have separated the *clustering ensemble methods* which use *fuzzy clustering* from

those which use *hard clustering* in Table 6. Therefore, we have used both *fuzzy clustering ensemble* and *hard clustering ensemble* as baseline methods in the experimental results.

Figures 4, 5 and 6 represent the performances of different methods respectively in terms of ARI, ACC, and FM.

Although in 6 datasets (i.e. Glass, Wine, Iris, Half-Ring, Artificial₂, ISOLET) the proposed method has not shown the best results, it overshadows all other methods. Indeed, there is not a method overshadowing others, i.e. while a method may have the best result on a special dataset, it has not the best in others.

Considering the performances of the state-of-the-art methods in terms of accuracy, presented in Fig. 6, we notice several

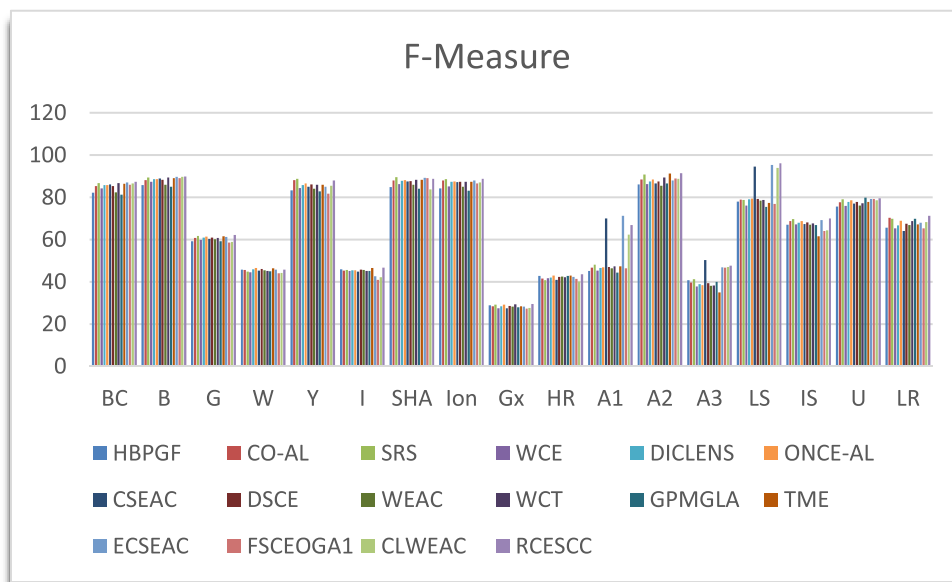
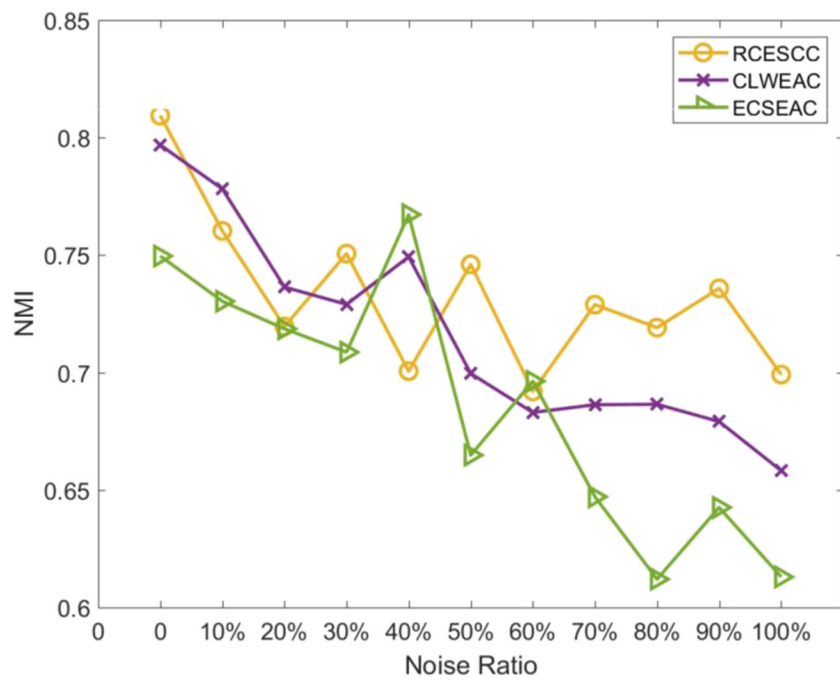
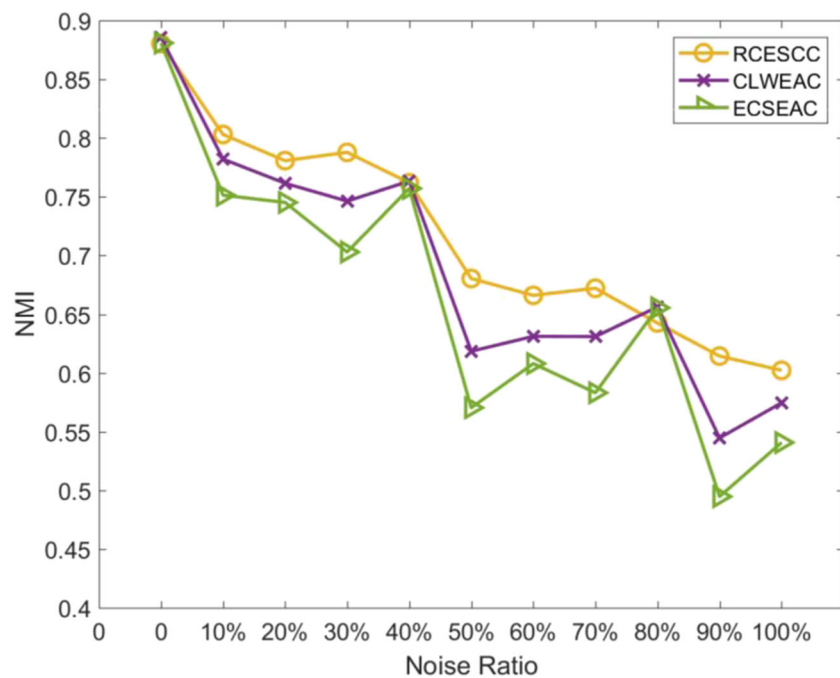


Fig. 6 The performances of different methods in terms of F-measure



(Iris)



(Wine)

Fig. 7 The NMI results for Iris and Wine datasets in the presence of different levels of noise

interesting outcomes. First, the performance of the proposed method is considerably better than all other methods in eight datasets, while its performance on other datasets is very close to the best. Second, the performance of the proposed method is considerably worse than some of the state-of-the-art methods only in three datasets, while its performance is still among the top three best methods. On the other datasets, the performance

of the proposed method is either very close to the best, or the best with slight margin.

5.5 Time complexity

The proposed method time complexity of the worst case is member of $O(c^2b^2 \log(cb))$ where the approximate number of

clusters by $c \times b$ (c is the real number of clusters and b is the ensemble size).

5.6 Parameter analysis

Noise effect has been examined in this subsection. For robustness analysis, we have selected the three best methods according to Table 6, i.e. RCESCC, ECSEAC and CLWEAC. We have added Gaussian noise with different energy levels and then applied these three methods. The results for Iris and Wine datasets have been depicted in Fig. 7.

6 Conclusions and future work

Clustering ensemble has been recently a hot topic in computer science communities. It makes it possible to overcome the instability of basic clustering algorithms and therefore it can boost the clustering efficacy. There are two sub-problems in clustering ensemble: ensemble generation problem, and ensemble aggregation problem. The paper works on the second problem by presenting an ensemble clustering framework to combine the ensemble members by a fast and robust two-phase clustering method. Indeed, it uses an ensemble of fuzzy clusterings generated by the fuzzy cmeans clustering algorithm. To produce the necessary diversity in the ensemble, the subsampling has been used. The experimental results on standard UCI datasets have shown that our algorithm which uses a fast and robust two-phase clustering method performs well on almost all datasets and outperforms all the state of the art clustering ensemble algorithms.

Acknowledgments This paper is extracted from a PhD thesis written by Musa Mojarad.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Wang B, Zhang J, Liu Y, Zou Y (2017) Density peaks clustering based integrate framework for multi-document summarization. *CAAI Transactions on Intelligence Technology* 2(1):26–30
- Ma J, Jiang X, Gong M (2018) Two-phase clustering algorithm with density exploring distance measure. *CAAI Transactions on Intelligence Technology* 3(1):59–64
- Deng Q, Wu S, Wen J, Xu Y (2018) Multi-level image representation for large-scale image-based instance retrieval. *CAAI Transactions on Intelligence Technology* 3(1):33–39
- Chakraborty D, Singh S, Dutta D (2017) Segmentation and classification of high spatial resolution images based on Hölder exponents and variance. *Geo-spatial Information Science* 20(1):39–45
- Yang H, Yu L (2017) Feature extraction of wood-hole defects using wavelet-based ultrasonic testing. *J For Res* 28(2):395–402
- Li C, Zhang Y, Tu W et al (2017) Soft measurement of wood defects based on LDA feature fusion and compressed sensor images. *J For Res* 28(6):1285–1292
- Alsaadeh B, Tateishi R, Phong DX, Hoan NT, Al-Hanbali A, Xiulian B (2017) New urban map of Eurasia using MODIS and multi-source geospatial data. *Geo-spatial Information Science* 20(1):29–38
- Song XP, Huang C, Townshend JR (2017) Improving global land cover characterization through data fusion. *Geo-spatial Information Science* 20(2):141–150
- Strehl A, Ghosh J (2003) Cluster ensembles—a knowledge reuse framework for multiple partitions. *The Journal of Machine Learning Research* 3:583–617
- Alizadeh H, Minaei-Bidgoli B, Parvin H (2014b) To improve the quality of cluster ensembles by selecting a subset of base clusters. *Journal of Experimental & Theoretical Artificial Intelligence* 26(1):127–150
- Mondal S, Banerjee A (2015) ESDF: ensemble selection using diversity and frequency. *Eprint Arxiv* 68(1):10–12
- Naldi MC, Carvalho AC, Campello RJ (2013) Cluster ensemble selection based on relative validity indexes. *Data Min Knowl Disc* 27(2):259–289
- Ni Z, Wu X, Ni L, Tang L, Xiao H (2015) The research on selective clustering ensemble algorithm based on fractal dimension and projection. *Journal of Computational Information Systems* 11(11):4025–4035
- X. Wang, D. Han, C. Han, Rough set based cluster ensemble selection, information FUSION (FUSION), 2013a
- Yang F, Li T, Zhou Q, Xiao H (2017) Cluster ensemble selection with constraints. *Neurocomputing* 235:59–70
- Yousefnezhad M, Reihanian A, Zhang D, Minaei-Bidgoli B (2016) A new selection strategy for selective cluster ensemble based on diversity and independency. *Eng Appl Artif Intell* 56(C):260–272
- Minaei-Bidgoli B, Parvin H, Alinejad-Rokny H, Alizadeh H, Punch WF (2013) Effects of resampling method and adaptation on clustering ensemble efficacy. *Artif Intell Rev* 41(1):27–48
- Yu Z, Chen H, You J, Wong HS (2014) Double selection based semi-supervised clustering ensemble for tumor clustering from gene expression profiles. *IEEE/ACM transactions on computational biology. Bioinformatics* 11(4):727–740
- Kao LJ, Huang YP (2013) Ejecting outliers to enhance robustness of fuzzy cluster ensemble. In: *IEEE international conference on systems, man, and cybernetics*, pp 3790–3795
- Mishra SP, Mishra D, Patnaik S (2015) An integrated robust semi-supervised framework for improving cluster reliability using ensemble method for heterogeneous datasets. *Karbala International Journal of Modern Science* 1(4):200–211
- Akbari E, Dahlan HM, Ibrahim R, Alizadeh H (2015) Hierarchical cluster ensemble selection. *Eng Appl Artif Intell* 39(39):146–156
- H. Wang, J. Qi, W. Zheng, M. Wang, Semi-supervised cluster ensemble based on binary similarity matrix, in: *The IEEE International Conference on Information Management and Engineering*, 2010, pp. 251–254
- Alizadeh H, Minaei B, Parvin H (2013) Optimizing fuzzy cluster Ensemble in String Representation. *International Journal of Pattern Recognition and Artificial Intelligence, IJPRAI, ISSN:0218–0014*
- Meng J, Hao H, Luan Y (2016) Classifier ensemble selection based on affinity propagation clustering. *J Biomed Inform* 60:234–242
- Soltanmohammadi E, Naraghi-Pour M, Schaar MVD (2016) Context-based unsupervised ensemble learning and feature ranking. *Mach Learn* 105(3):1–27
- Wang D, Li L, Yu Z, Wang X (2013b) AP2CE: double affinity propagation based cluster ensemble. In: *International conference on machine learning and cybernetics*, pp 16–23
- Yu Z, Luo P, You J, Wong HS, Leung H, Wu S, Zhang J, Han G (2016) Incremental semi-supervised clustering ensemble for high

- dimensional data clustering. *IEEE Transactions on Knowledge & Data Engineering* 28(3):701–714
28. Iam-On N, Boongoen T, Garrett S, Price C (2011) A link-based approach to the cluster ensemble problem. *IEEE transactions on Pattern Analysis & Machine Intelligence* 33(12):2396–2409
 29. Zhong C, Yue X, Zhang Z, Lei J (2015) A clustering ensemble: two-level-refined co-association matrix with path-based transformation. *Pattern Recogn* 48(8):2699–2709
 30. Wang LJ, Hao ZF, Cai RC, Wen W (2014) An improved local adaptive clustering ensemble based on link analysis. In: *International conference on machine learning and cybernetics*, pp 10–15
 31. Xiao W, Yang Y, Wang H, Li T, Xing H (2016) Semi-supervised hierarchical clustering ensemble and its application. *Neurocomputing* 173:1362–1376
 32. Wang W (2008) Some fundamental issues in ensemble methods. In: *proceedings of the IEEE international joint conference on neural networks, IEEE world congress on. Comput Intell*:2243–2250
 33. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM computing surveys (CSUR)* 31(3):264–323
 34. Tan PN, Steinbach M, Kumar V (2006) *Introduction to data mining*. In: Pearson Addison Wesley (Boston)
 35. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* 3:32–57
 36. Berikov VB (2018) A probabilistic model of fuzzy clustering ensemble. *Pattern Recognition and Image Analysis* 28(1):1–10. <https://doi.org/10.1134/S1054661818010029>
 37. Dimitriadou E, Weingessel A, Hornik K (2002) A combination scheme for fuzzy clustering. *Int J Pattern Recognit Artif Intell* 16(07):901–912
 38. Li T, Chen Y (2010) Fuzzy clustering ensemble with selection of number of clusters. *JCP* 5(7):1112–1119
 39. Nazari A, Dehghan A, Nejatian S, Rezaie V, Parvin H (2018) A comprehensive study of clustering ensemble weighting based on cluster quality and diversity. *Pattern Anal Applic*. <https://doi.org/10.1007/s10044-017-0676-x>
 40. Pan S, Changjing S, Qiang S (2015) A hierarchical fuzzy cluster ensemble approach and its application to big data clustering. *Journal of Intelligent & Fuzzy Systems* 28(6):2409–2421
 41. Parvin H, Minaei-Bidgoli B (2015) A clustering ensemble framework based on selection of fuzzy weighted clusters in a locally adaptive clustering algorithm. *Pattern Anal Appl* 18(1):87–112
 42. Sevillano X, JC S'õ, Alias F (2009) Fuzzy clusterers combination by positional voting for robust document clustering. *Procesamiento del lenguaje natural* 43:245–253
 43. Alqurashi T, Wang W (2015) A new consensus function based on dual-similarity measurements for clustering ensemble. In: *International conference on data science and advanced analytics (DSAA)*, IEEE. ACM, pp 149–155
 44. Fern XZ, Brodley CE (2003) Random projection for high dimensional data clustering: A cluster ensemble approach. In: *Proceedings of the 20th International Conference on Machine Learning*, pp 186–193, URL <http://www.aaai.org/Papers/ICML/2003/ICML03-027.pdf>
 45. Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140
 46. Minaei-Bidgoli B, Topchy A, Punch WF (2004) Ensembles of partitions via data resampling. In: *Proceedings of the international conference on information technology: coding and computing ITCC*, IEEE, vol 2, pp 188–192
 47. Parvin H, Minaei-Bidgoli B, Alinejad-Rokny H, Punch WF (2013) Data weighing mechanisms for clustering ensembles. *Computers & Electrical Engineering* 39(5):1433–1450
 48. Alizadeh H, Yousefnezhad M, Minaei-Bidgoli B (2015) Wisdom of crowds cluster ensemble. *Intell Data Anal* 19(3):485–503
 49. Topchy A, Jain AK, Punch W (2004) A mixture model of clustering ensembles. *Proceedings of the SIAM International Conference of Data Mining*. In
 50. Iam-on N, Boongoen T, Garrett S (2010) LCE: a link-based cluster ensemble method for improved gene expression data analysis. *Bioinformatics* 26(12):1513–1519
 51. Iam-On N, Boongoen T, Garrett S, Price C (2012) A link based cluster ensemble approach for categorical data clustering. *IEEE Trans Knowl Data Eng* 24(3):413–425
 52. Yi J, Yang T, Jin R, Jain AK, Mahdavi M (2012) Robust ensemble clustering by matrix completion. In: *proceedings of the IEEE 12th international conference on data mining (ICDM)*. IEEE:1176–1181
 53. Gionis A, Mannila H, Tsaparas P (2007) Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1(1):4–es
 54. Alqurashi T, Wang W (2014) Object-neighborhood clustering ensemble method. In: *Intelligent data engineering and automated learning (IDEAL)*. Springer, pp 142–149
 55. Fred AL, Jain AK (2005) Combining multiple clustering's using evidence accumulation. *IEEE Trans Pattern Anal Mach Intell* 27(6):835–850
 56. Yang Y, Jiang J (2016) Hybrid sampling-based clustering ensemble with global and local constitutions. *IEEE Transactions on Neural Networks and Learning Systems* 27(5):952–965
 57. Bai L, Cheng X, Liang J, Guo Y (2017) Fast graph clustering with a new description model for community detection. *Inf Sci* 388-389: 37–47
 58. Karypis G, Kumar V (1998) A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J Sci Comput* 20(1):359–392
 59. Fern XZ, Brodley CE (2004) Solving cluster ensemble problems by bipartite graph partitioning. In: *Proceedings of the 21st International Conference on Machine learning*, ACM, p 36
 60. Huang D, Lai J, Wang CD (2016b) Robust ensemble clustering using probability trajectories. *The IEEE Transactions on Knowledge and Data Engineering, Robust Ensemble Clustering Using Probability Trajectories*
 61. Huang D, Lai J, Wang CD (2016a) Ensemble clustering using factor graph. *Pattern Recogn* 50:131–142
 62. Houle ME (2008) The relevant-set correlation model for data clustering. *Statistical Analysis and Data Mining* 1(3):157–176
 63. Vinh NX, Houle ME (2010) A set correlation model for partitional clustering. *Advances in Knowledge Discovery and Data Mining*, Springer, In, pp 4–15
 64. D. Dueck, "Affinity propagation: Clustering data by passing messages," Ph.D. dissertation, University of Toronto, 2009
 65. Newman CBDJ, SS Hettich, C Merz (1998) UCI repository of Mach Learn databases, <http://www.ics.uci.edu/~mllearn/MLSummary.html>, (1998)
 66. Ren Y, Zhang G, Domeniconi C, Yu G (2013) Weighted object ensemble clustering. In: *Proceedings of the IEEE 13th International Conference on Data Mining (ICDM)*, IEEE, pp 627–636
 67. Mimaroglu S, Aksehirli E (2012) DICLENS: divisive clustering ensemble with automatic cluster number. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 9(2):408–420
 68. Alizadeh H, Minaei-Bidgoli B, Parvin H (2014a) Cluster ensemble selection based on a new cluster stability measure. *Intelligent Data Analysis* 18(3):389–408
 69. Huang D, Lai JH, Wang CD (2015) Combining multiple clusterings via crowd agreement estimation and multi-granularity link analysis. *Neurocomputing* 170:240–250
 70. Hubert L, Arabie P (1985) Comparing partitions. *J Classif* 2(1): 193–218

Musa Mojarad is from Iran. He has received M.Sc. degree in computer science from Islamic Azad University, Ahvaz, Iran. He is a PhD student in Department of Computer Science, Yasooj Branch, Islamic Azad University, Yasooj, Iran. He is currently under supervision of Hamid Parvin and Samad Nejatian.

Samad Nejatian received his B.Eng. degrees in Electronic Engineering from the University of Sistan and Baluchestan, Zahedan, Iran in 2003. M.Sc. degree in Telecommunication Technology, and PhD degree in Data Communication from the University of Mashhad, Iran, and Universiti Teknologi Malaysia in 2006 and 2014, respectively. Currently, he holds university assistant professor position at the Faculty of Electrical Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran. His Professional Affiliations are with Member of Institute Electrical and Electronic Engineers (IEEE), Member of Telecommunications Research

Group under Centre of Excellence in Telecommunication Technology, UTM. His research interest includes intelligent systems, Wireless communication, Software define Radio, Cognitive radio, and Statistical Pattern Recognition.

Hamid Parvin is from Nourabad Mamasani, Iran. He has received PhD degree in computer science under supervision of Behrouz Minaei at 2011. Hamid Parvin is a faculty member in Department of Computer Science, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran.

Majid Mohammadpoor is from Iran. He has received M.Sc. degree in computer science from Islamic Azad University, Yasooj, Iran. He is a PhD candidate in Islamic Azad University.