



A multi-level consensus function clustering ensemble

Kim-Hung Pho¹ · Hamidreza Akbarzadeh² · Hamid Parvin^{3,4} · Samad Nejatian^{5,6} · Hamid Alinejad-Rokny^{7,8,9}

Accepted: 8 March 2021 / Published online: 13 September 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

In order to improve the performance of a clustering on a data set, a number of primary partitions are generated and stored in an ensemble and their aggregated consensus partition is used as their clustering. It is widely accepted that the consensus partition outperforms the primary partitions. In this paper, an ensemble clustering method called multi-level consensus clustering (MLCC) is proposed. To construct the MLCC, a cluster–cluster similarity matrix which is achieved by an innovative similarity metric is first generated. The mentioned cluster–cluster similarity matrix is based on a multi-level similarity metric. In fact, it can be computed in a new defined multi-level space. Then, a point–point similarity matrix which is boosted using the mentioned cluster–cluster similarity matrix is generated. The new consensus function applies an average linkage hierarchical clusterer algorithm on the mentioned point–point similarity matrix to make consensus partition. MLCC is better than traditional clustering ensembles and simple versions of clustering ensembles on traditional cluster–cluster similarity matrix. Its computational cost is not very bad too. Accuracy and robustness of the proposed method are compared with those of the modern clustering algorithms through the experimental tests. Also, time analysis is presented in the experimental results.

Keywords Consensus partition · Multi-level similarity metric · Ensemble learning

1 Introduction

Numerous real-world issues are resolved by utilizing extremely beneficial straightforward machine learning models (Nejatian et al. 2019; Pirbonyeh et al. 2019; Hosseinpoor et al. 2019; Partabian et al. 2020; Shabaniyan

et al. 2019; Szetoa et al. 2020; Moradi et al. 2018; Jenghara et al. 2018a; Parvin et al. 2018; Alishvandi et al. 2016; Omidvar et al. 2018; Yasrebi et al. 2018). Nevertheless, since understandable machine learning models encounter difficulties in challenging issues (Nejatian et al. 2018; Jamalnia et al. 2018; Jenghara et al. 2018b), ensemble models have emerged as a new option in regard with the

✉ Hamid Parvin
parvin@iust.ac.ir

✉ Hamid Alinejad-Rokny
h.alinejad@unsw.edu.au

¹ Fractional Calculus, Optimization and Algebra Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

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

³ Department of Computer Engineering, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

⁴ Young Researchers and Elite Club, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

⁵ Department of Electrical Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran

⁶ Young Researchers and Elite Club, Yasooj Branch, Islamic Azad University, Yasooj, Iran

⁷ BioMedical Machine Learning Lab (BML), The Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, NSW 2052, Australia

⁸ UNSW Data Science Hub, The University of New South Wales (UNSW Sydney), Sydney, NSW 2052, Australia

⁹ Health Data Analytics Program, AI-enabled Processes (AIP) Research Centre, Macquarie University, Sydney 2109, Australia

machine learning classification tasks (Rashidi et al. Aug. 2019; Mojarad et al. 2019a,b; Nazari et al. 2019; Bagherinia et al. 2019).

To find a partition with the goal of assigning data points within identical groups is a prominent and complicated matter which has been employed in many real-world problems such as knowledge extraction and pattern recognition. According to no free lunch (NFL), there is no dominant algorithm among many algorithms for solving a problem (Huang et al. 2016b; Fred and Jain 2005). Therefore, we know that there is a dominant algorithm depending on the at-hand data set. However, lack of knowledge on distributions, structures and natures of the clusters in the at-hand data set results in some problems which prevent us from detection of the most suitable clustering algorithm to our specific application (Roth et al. 2002).

Cluster ensembles have attracted enormous interest in result of several successful applications of ensemble learning in supervised fields such as (Freund and Schapire 1995; Ho 1995; Friedman 2001; Soto et al. 2014; Yu et al. 2015,2016a,2017). Many classification applications have shown a tendency towards hybrid methods in the past few years (Faceli et al. 2006) similar to every sub-field of the supervised pattern detection. Generally, several fields including data mining (Hong et al. 2008; Zhang et al. 2012; Yu et al. 2012,2016b; Naldi et al. 2013; Franek and Jiang 2014; Jiang et al. 2015; Huang et al. 2017; Yousefnezhad et al. 2018), multimedia (Rafiee et al. 2013) and bio-informatics (Yu et al. 2011,2013; Hanczar and Nadif 2012) have used various cluster ensemble approaches which have intended to clarify innovative partitions because they are stronger than simple clustering algorithms (Ayad and Kamel 2008). Using ensembles as a group of primary partitions to devise a consensus partitions is the objective of the ensemble clustering (Domeniconi and Al-Razgan 2009; Ghosh and Acharya 2011; Ghaemi et al. 2011). Consensus ensemble uses every primary partition within the ensemble in construction of the consensus partition, and the consensus partition enhances a specific objective function.

Cluster ensemble consists of two stages. First stage is to generate a pool of primary partitions. These primary partitions usually are the moderate or low-quality ones (Topchy et al. 2003). Actually, they should be low-quality but not without-quality. It is widely acceptable to generate these primary partitions through a simple clusterer method (usually k -means) on a number of dissimilar parameters (Topchy et al. 2005), on a number of dissimilar initializations, on a number of dissimilar data subspaces (Ayad and Kamel 2008), or on a number of dissimilar data subsamples (Minaei-Bidgoli et al. 2004). The proposed method has used all of the mentioned approaches.

In second step, we aim at finding a consensus partition that has maximum similarity to the primary partitions in our ensemble generated in the first step. Evidence accumulation clustering (EAC) is proposed in Fred and Jain (2002) as an ideal consensus function in which the ensemble is initially transformed into a co-association matrix (CAM), and then, the consensus partition is found by employing a single linkage hierarchical clusterer algorithm.

Normalized mutual information (NMI) measure is proposed in Fred and Jain (2005) to assess a cluster which has numerous deficiencies described in Alizadeh et al. (2011a). MAX is an edition of the normalized mutual information measure which has also some flaws. Therefore, Alizadeh-Parvin-Moshki-Minaei (APMM) measure has been introduced in Alizadeh et al. (2011b) as a metric to enhance the MAX. All of mentioned metrics, i.e. NMI, MAX, and APMM, are weak in dealing with complement cluster effect.

In this paper, an innovative cluster-cluster similarity measure is introduced. Then, we extend it to EAC. MLCC is inspired by dual-similarity clustering ensemble (DSCE) method (Alqurashi and Wang 2014,2015). DSCE and MLCC have three differences. First, MLCC has less sensitivity to the usage of the real number of clusters in primary partitions. Second, the MLCC method has less parameters than DSCE. Third, MLCC is a CAM-based consensus clustering, but DSCE is a voting-based consensus clustering that is an important consensus function clustering (Strehl and Ghosh 2000; Fern and Brodley 2003; Breiman 1996; Alizadeh et al. 2015; Iam-On et al. 2010,2012; Gionis et al. 2007; Yi et al. 2012; Mirzaei and Rahmati 2010; Akbari et al. 2015).

This article has been continued as follows. In Sect. 2, the clustering ensemble problem of the paper is defined. In Sect. 3, the related works are presented. Section 4 has explained the proposed method. Section 5 presents the experimental results, and the paper conclusions will be presented in Sect. 6.

2 Related work

Encapsulating ensemble information into a CAM and application of a hierarchical clusterer algorithm on the obtained CAM are needed for extracting the consensus partition in an approach that is based on CAM (Fred and Jain 2002). These types of clustering ensembles take the advantage of bypassing the need for a relabeling phase. As an instance, consensus partition was achieved in Fred and Jain (2005) through applying an average linkage hierarchical clusterer which we have named CO-AL as a slow and CAM-based method. A new CAM which considers the

between cluster relations was introduced in Iam-On et al. (2008) where the implemented approach was named SRS which can be considered to be a slow method. It was based on CAM and used link similarity. After that, it was extended into WCT method (Iam-On et al. 2011).

Instead of an object–object similarity CAM, a cluster–cluster similarity matrix was introduced in Mimaroglu and Aksehirli (2012) where a method named DICLENS was offered as a CAM-based, slow, and robust approach. An object-neighbourhood-based similarity matrix was introduced in where a new cluster–cluster similarity matrix was obtained by neighbourhood and real relation between objects. This method that was named ONCE–AL together with its extension DSCE method introduced in Alqurashi and Wang (2015), can be considered to be a heuristic, fast, and parameter-sensitive approach. The mentioned method later extended (Huang et al. 2016b). An extended version of the DSCE was also proposed in Huang et al. (2015) which introduces WEAC and GPMGLA as a pair of two cluster weighting, CAM-based, slow and robust clustering ensemble methods.

CSEAC was proposed in Alizadeh et al. (2014a) as a CAM-based, very slow robust clustering ensemble method, and also Wisdom of Crowds Ensemble (WCE) was introduced in Alizadeh et al. (2015) as a CAM-based, slow, robust and flexible ensemble method. Another related work is ECSEAC which is a cluster selection-based ensemble clustering algorithm proposed in Parvin and Minaei-Bidgoli (2015) as a cluster weighting, CAM-based, slow and robust approach. TME is also an ensemble method which was proposed in Zhong et al. (2015) as a cluster weighting and CAM-based one; it is widely considered to be a very slow and robust method.

Clustering ensemble has been reformulated as a linear binary programming problem by Huang et al. (2016). This approach was also extended and used for categorical data set by Zhao et al. (2017).

A cluster ensemble was introduced through sampling by Yang and Jiang (2016) which was inspired by bagging and boosting theory. Also, information theory as a suitable tool for data clustering was used to introduce a cluster ensemble method by Bai et al. (2017).

Consensus partition was extracted by Strehl and Ghosh (2003) by partitioning a hyper-graph which is constructed on a CAM so that each row and each column in the CAM are, respectively, assumed to be a node and a hyper-edge. The hyper-graph partitioning algorithm used in the mentioned work could be HMETIS (Dimitriadou et al. 2002). This algorithm is named cluster-based similarity partitioning algorithm (CSPA). Hyper-graph partitioning algorithm (HGPA) and meta-clustering algorithm (MCLA) are some other examples of the methods in which each cluster is considered to be a hyper-edge and each data as a node.

CSPA has been extended by Fren and Brodley (2004) as a hybrid bi-partite graph formulation (HBGF) which outperforms HGPA, MCLA and CSPA.

A cluster ensemble method was proposed in Rashidi et al. (2019) which did not need the original features of data set and also did not consider any distribution in the data set. Moreover, it used assessment of the clusters' undependability and the weighting mechanism to suggest a clustering ensemble. In addition, diversity was defined in cluster level. During consensus function process, quality of the clusters and their diversity were used in order to improve the consensus partition. Any clusters' dependability was computed based on its undependability which itself was calculated according to the entropy in the labels of its data points throughout all the ensemble partitions. Also, an innovational cluster-wise weighing CAM was suggested according to measuring the clusters' reliability and then emphasizing those with higher reliability in the ensemble. Furthermore, a cluster-wise weighting bi-partite graph was suggested to offer the ensemble based on the related cluster-wise weighing co-association matrix. Finally, the consensus partition was extracted using two mechanisms including application of a simple hierarchical clustering algorithm for cluster-wise weighing of the co-association matrix as a similarity matrix and cluster-wise weighting bi-partite graph partitioning into a certain number of parts.

The idea proposed in Alizadeh et al. (2014a) is to find a subset of primary clusters which performs better in the ensemble, and through participating the clusters with more quality in consensus function process, the consensus partition performance was improved. In this regard, many clusters are first gained by specific generation of primary results. Then, a goodness function is used to evaluate and also to sort the achieved clusters. After that, with the goal of finding more efficient clusters for participating in the ensemble, a specific subset of primary clusters is offered for each class of datasets. The selected clusters are then combined to extract the final clusters which is followed by a post-processing.

A supreme consensus partition was produced in Alizadeh et al. (2013) by combination of spurious clustering results where the objective function was inspired by proposed immature formulation in which simultaneous maximization of the agreement between the ensemble members and minimization of the disagreement was implemented. They improve a nonlinear binary goal function which was introduced to propose a constrained nonlinear fuzzy objective function. The genetic algorithm was also used to solve the proposed model and while any other optimizer could be used to solve the model.

In the clustering ensemble proposed in Minaei-Bidgoli et al. (2014), a simple adaptive attitude was suggested for

generating primary partitions. It uses the clustering history. Since it is supposed in the clustering that the ground truth is not achievable in the class labels form, it is needed to use an alternative performance measure for an ensemble of partitions throughout process of the clustering. To determine clustering consistency of data points, a history of cluster assignments was evaluated in the generated sequence of the partitions for each data point which could adapt the data sampling during the partition generation to the current state of an ensemble that improves the confidence in cluster assignments.

By extending the previous clustering frameworks, ensemble and swarm concepts in the clustering field were suggested in Parvin et al. (2012) where an unprecedented clustering ensemble learning method was introduced according to the ant colony clustering algorithm. Since diversity of the ensemble is very important and swarm intelligence is inherently related to random processes, a new dataset space was introduced by diverse partitions of the acquired results from different runs of the ant colony clustering algorithm with different initializations on a data set and they were aggregated into a consensus partition by a simple clusterer algorithm. The contribution of swarm intelligence in reaching better results was also measured.

Benefits of the ensemble diversity together with quality of the fuzzy clustering-level were used to select a subset of fuzzy base-clusterings in Bagherinia et al. (2019). Integrating diversity of the base-clusterings and quality of the clustering level was resulted into enhancement of the final clustering quality. It first calculated a new fuzzy normalized mutual information (FNMI) and then calculated the diversity of each fuzzy base-clustering in relation to other fuzzy base-clusterings. After that, the calculated diversity was used to cluster all the base-clusterings which resulted in clusters of base-clusterings named base-clusterings-clusters. In subsequence, a base-clusterings-cluster is selected which satisfies the quality measure. Final clustering was achieved by using a novel consensus method derived by graph portioning algorithm named FCBGP algorithm or by forming extended fuzzy co-association matrix (EFCo) which is finally considered as similarity matrix.

Because of lower information or higher variances some features of under-consideration data set have more information than the others in that data set, weighted locally adaptive clustering (WLAC) algorithm was proposed in Parvin and Minaei-Bidgoli (2013) as a weighted clustering algorithm which could encounter the imbalanced clustering. WLAC algorithm is very sensitive to the two parameters which was resulted in proposing a simple clustering ensemble framework and an elite cluster ensemble framework to manage those two control parameters. It was later

extended into a fuzzy version, i.e. Fuzzy WLAC (FWLAC) algorithm (Parvin and Minaei-Bidgoli 2015).

A novel approach for clustering ensemble was also suggested in Abbasi and Nejatian (2019). It uses a subset of primary clusters. It proposed to evaluate the cluster quality by suggesting stab as a new validity measure so that the clusters that satisfy a stab threshold are qualified to be used in construction of the consensus partition. A set of methods for consensus function such as co-association based ones were used to combine the chosen clusters. Extended evidence accumulation clustering (EEAC) was used for co-association matrix construction from the selected clusters because it was not possible for evidence accumulation clustering (EAC) method to use a subset of clusters in construction of co-association matrix. Other class of consensus functions was also used based on hyper-graph partitioning algorithms. In another one, the chosen clusters were considered to be a new feature space and the consensus partition was extracted by using a simple clusterer algorithm.

Averaged Alizadeh–Parvin–Moshki–Minaei (AAPMM) criterion was introduced in Alizadeh et al. (2014b) for assessing a cluster quality. It was proposed to use AAPMM for obtaining quality of clusters. Then, in their clustering ensemble method, the clusters satisfying an AAPMM value threshold participate in constructing the co-association matrix. In addition, EEAC was proposed as a new method for matrix establishment. Finally, consensus partition extraction was performed by application of a hierarchical clusterer method over the obtained co-association matrix.

A primary partitions' generator was suggested in Parvin et al. (2013). It could be considered to be a boosting mechanism in clustering. As true labels are needed for each cluster to be available in weight updating of boosting, the clustering needs an alternative performance metric for an ensemble of partitions to update the probability sampling vector. This study has determined the clustering consistency of data points was determined to assess a cluster assignment history for every data point in the generated partitions of ensemble. In order to have a suitable data sampling, determination of the clustering consistency for produced partitions was performed during the partition generation. The adaptation concentrated the sampling distribution on problematic regions of the feature space to modify the confidence in cluster assignments. Accordingly, better approximation of the inter-cluster boundaries was performed by concentrating the notice on the data points which have the minimum consistent clustering assignments.

3 Clustering ensemble

3.1 Clustering ensemble definition

Let's assume a set of data points be denoted by \mathcal{X} . In the data set \mathcal{X} , i th feature of all data points is denoted by data set $\mathcal{X}_{:,i}$ and j th data point is denoted by $\mathcal{X}_{j,:}$. The number of data points in the data set \mathcal{X} is denoted by $|\mathcal{X}_{:,1}|$. The number of features in the data set \mathcal{X} is denoted by $|\mathcal{X}_{1,:}|$. A partition on the data set \mathcal{X} is denoted by $\pi_{\mathcal{X}}$ (also, $\pi_{\mathcal{X}}^k$ is k th partition on the data set \mathcal{X} where k is an integer value indicating the index of partition). The partition $\pi_{\mathcal{X}}$ contains a $|\pi_{\mathcal{X},:}|$ subsets of $\{1, 2, \dots, |\mathcal{X}_{:,1}|\}$. Let's denote p th subset of partition $\pi_{\mathcal{X}}$ by $\pi_{\mathcal{X},p}$. To name $\pi_{\mathcal{X}}$ a non-overlapping partition, we must have

$$\forall p_1, p_2 \in \{1, 2, \dots, |\pi_{\mathcal{X},:}|\} \bullet \pi_{\mathcal{X},p_1} \cap \pi_{\mathcal{X},p_2} = \emptyset \tag{1}$$

If the Eq. 1 holds for a partition, we name it a non-overlapping partition. From now on, without loss of generality, we assume all of the partitions, discussed in the paper, are non-overlapping. Any member of a partition, i.e. $\pi_{\mathcal{X},p}$, is considered to be a cluster. If the union of all clusters in a partition is $\{1, 2, \dots, |\mathcal{X}_{:,1}|\}$, then the partition is named a complete partition. It means the Eq. 2 must be correct for a partition $\pi_{\mathcal{X}}$ to name it a complete partition.

$$\bigcup_{p=1}^{\pi_{\mathcal{X}}} \pi_{\mathcal{X},p} = \{1, 2, \dots, |\mathcal{X}_{:,1}|\} \tag{2}$$

Let's assume a data set \mathcal{X} has a desired clustering similar to its ground-truth labels denoted by $\pi_{\mathcal{X}}^+$. Note that we have not assumed that any of $|\pi_{\mathcal{X},:}^k|$ should be equal to $|\pi_{\mathcal{X},:}^+|$. Each cluster, say $\pi_{\mathcal{X},p}$, has a central point denoted by $C^{\pi_{\mathcal{X},p}}$ defined according to Eq. 3.

$$C^{\pi_{\mathcal{X},p}} = \frac{1}{|\pi_{\mathcal{X},p}|} \sum_{j \in \pi_{\mathcal{X},p}} \mathcal{X}_{j,:} \tag{3}$$

We denote the complete version of partition $\pi_{\mathcal{X}}$ by $\tilde{\pi}_{\mathcal{X}}$ whose p th cluster is defined according to Eq. 4.

$$\tilde{\pi}_{\mathcal{X},p} = \{ \forall j \in \{1 \dots |\mathcal{X}_{:,1}|\} \bullet [\forall p_1 \in \{1 \dots |\pi_{\mathcal{X},:}|\} \bullet |C^{\pi_{\mathcal{X},p}} - \mathcal{X}_{j,:}| \leq |C^{\pi_{\mathcal{X},p_1}} - \mathcal{X}_{j,:}|] j \} \tag{4}$$

We consider all of the partitions $\pi_{\mathcal{X}}^k$ to be an ensemble (denoted by $\pi_{\mathcal{X}}$) with the ensemble size $|\pi_{\mathcal{X}}^k|$. Our aim is to combine them to extract a consensus partition denoted by $\pi_{\mathcal{X}}^*$ in such a way that it is as much robust and high-quality as possible. Formally, the ensemble clustering intends to find a better consensus partition by integrating the information of the primary partitions in the ensemble.

Clustering ensemble domain has examined the creation of an ensemble and also the consensus function construction which both have highly influenced the outcomes of the final clustering. Since variety of the results from the base learners is a crucial factor to make a successful ensemble, the diversity could be extended into the clustering ensemble field in regard to classifier ensemble.

3.2 Ensemble creation

First of all, the ensemble members should be created in clustering ensemble framework. To find various data substructures and also to enhance the potential performance of consensus partition, we must use at least one of the clustering generation methods that guarantee to produce an ensemble of partitions where any of them has a minimum quality and all of them have a minimum diversity.

Many problem-specific ensemble generation methods have been introduced. Subspace clustering has been done on different projections of data set by Strehl and Ghosh (2000). Along to feature subset selection, a subsampling (or instance subset selection) is also used for generating each partition. Almost the same approach was proposed by Fern and Brodley (2003). Breiman (1996) first presents subsampling in machine learning, and later, it was extended by Minaei-Bidgoli et al. (2004).

An alternative to the prior methods including a subset of primary members within the ensemble was used by Alizadeh et al. (2011a,2011b). They used a subsampling for generating each primary partition. Each partition is attained by running a randomly initialized k-means clustering algorithm. Then, a subset of primary partitions that has maximum diversity is selected as final ensemble. A similar study was also conducted by Nazari et al. (2019).

Since k -means clusterer algorithm is an easy and efficient clusterer algorithm (Minaei-Bidgoli et al. 2004), it is considered to be the most suitable clusterer algorithm in ensemble generation. Thus, for each base clustering, k -means clusterer method was used in Alizadeh et al. (2015) with two random elements: randomly initialized seed points and randomly selected number of clusters. These random sources cause to produce a diverse ensemble. It is widely recommended to select an integer value randomly in $[2; \sqrt{|\mathcal{X}_{:,1}|}]$ as the number of clusters (Iam-On et al. 2010,2012) during generating the ensemble. Different clusterer algorithms can be also a source for generating a diverse ensemble (Gionis et al. 2007; Yi et al. 2012; Mirzaei and Rahmati 2010; Akbari et al. 2015).

For ensemble creation, Wisdom of Crowds (WOC) was proposed in Alizadeh et al. (2015) as a concept in social sciences field that provides suitable criteria for group behaviour. Meeting these criteria could result in favourable

ensembles. Therefore, WOC cluster ensemble (WOCCE) has been proposed. It was later extended in Yousefnezhad et al. (2018).

3.3 Ensemble combiner

In the literature, researchers pay more attention to ensemble combiner. It can be viewed as a function with two inputs: (a) ensemble members or partitions and (b) the number of clusters, i.e. $|\pi_{\mathcal{X}}^+|$. It returns an output, widely referred to as consensus or aggregated partition, i.e. $\pi_{\mathcal{X}}^*$.

As there are some widely accepted approaches for ensemble combiner, we will present a brief summarization on them here. Any ensemble combiner is in one category of the following ones: (a) category of CAM based ensemble combiners, (b) category of median partition based ensemble combiners, (c) category of graph partitioning based ensemble combiners, (d) category of voting based ensemble combiners, and (e) category of intermediate space based ensemble combiners.

The methods in the first category transform information of ensemble partitions into a similarity matrix, named CAM, and then, they apply a second clustering, widely a hierarchical one, to find the consensus partition. The methods in the second category transform the finding of consensus partition into an optimization problem, and then, they solve it through an optimizer. The methods in the third category transform the finding of consensus partition into a graph clustering problem, and then, they solve it through a graph partitioning algorithm. The methods in the fourth category first make a relabeling between all of the partitions, and then, they find consensus partition by a voting or averaging mechanism. The methods in the fifth category assume the clusters of data points as new intermediate features, and then, they apply a secondary simple clusterer algorithm on the new intermediate data set to find consensus partition.

Figure 1 depicts framework of the proposed framework. In the proposed framework, the initial data set is subsampled several times and then by applying a k-means clusterer method on each of them to partition them into a random number of clusters, an ensemble of $|\pi_{\mathcal{X}}^+|$ partitions denoted by $\pi_{\mathcal{X}}^q$ will be generated, where $q = \{1, \dots, |\pi_{\mathcal{X}}^+|\}$. After that, each partition is extended to its complete version, i.e. $\tilde{\pi}_{\mathcal{X}}^q$. After that, the ensemble is converted into a cluster representation, and then, a similarity matrix is created to determine the similarities of the clusters. In the next step, a CAM boosted cluster similarity matrix is obtained. Finally, a merging mechanism (a hierarchical clusterer algorithm) and post-processing are used to determine the consensus partition.

4 Proposed consensus function

Ensemble combiner or consensus function is the main part in a clustering ensemble framework. It must be appropriately designed so as to maximally extract ensemble information. We first create a graph according to Eq. 5 for our clustering ensemble $\pi_{\mathcal{X}}$.

$$\mathcal{G}(\pi_{\mathcal{X}}) = (\mathcal{V}(\pi_{\mathcal{X}}), \mathcal{E}(\pi_{\mathcal{X}})) \tag{5}$$

In Eq. 5, $\mathcal{V}(\pi_{\mathcal{X}})$ is set of indices of all the data clusters that are in the ensemble, i.e. $\{1, 2, \dots, |\pi_{\mathcal{X}}^+|\}$ where $|\pi_{\mathcal{X}}^+|$ denotes the number of clusters in ensemble $\pi_{\mathcal{X}}$ and $\mathcal{E}(\pi_{\mathcal{X}})$ is set of edge weights for graph $\mathcal{G}(\pi_{\mathcal{X}})$. The term $|\pi_{\mathcal{X}}^+|$ is defined according to Eq. 6.

$$|\pi_{\mathcal{X}}^+| = \sum_{k=1}^{|\pi_{\mathcal{X}}^+|} |\pi_{\mathcal{X}}^k| \tag{6}$$

In Eq. 5, $\mathcal{E}(\pi_{\mathcal{X}})$ is defined according to Eq. 7.

$$\mathcal{E}(\pi_{\mathcal{X}}) = \{\forall j_1, j_2 \in \mathcal{V}(\pi_{\mathcal{X}}) \bullet w_{\pi_{\mathcal{X}}}(j_1, j_2) \neq 0 | (j_1, j_2, w_{\pi_{\mathcal{X}}}(j_1, j_2))\} \tag{7}$$

where $w_{\pi_{\mathcal{X}}}(j_1, j_2)$ is defined according to Eq. 8.

$$w_{\pi_{\mathcal{X}}}(j_1, j_2) = \begin{cases} 0 & j_1 = j_2 \\ SCCS_{\pi_{\mathcal{X}}}(\tilde{\pi}_{\mathcal{X}, p_{j_1, \pi_{\mathcal{X}}}}^{k_{j_1, \pi_{\mathcal{X}}}}, \tilde{\pi}_{\mathcal{X}, p_{j_2, \pi_{\mathcal{X}}}}^{k_{j_2, \pi_{\mathcal{X}}}}) & j_1 \neq j_2 \end{cases} \tag{8}$$

where $k_{j, \pi_{\mathcal{X}}}$ and $p_{j, \pi_{\mathcal{X}}}$ are, respectively, defined according to Eqs. 8 and 9.

$$k_{j, \pi_{\mathcal{X}}} = arg \max_K \sum_{k=1}^K |\pi_{\mathcal{X}}^k| \leq j \tag{9}$$

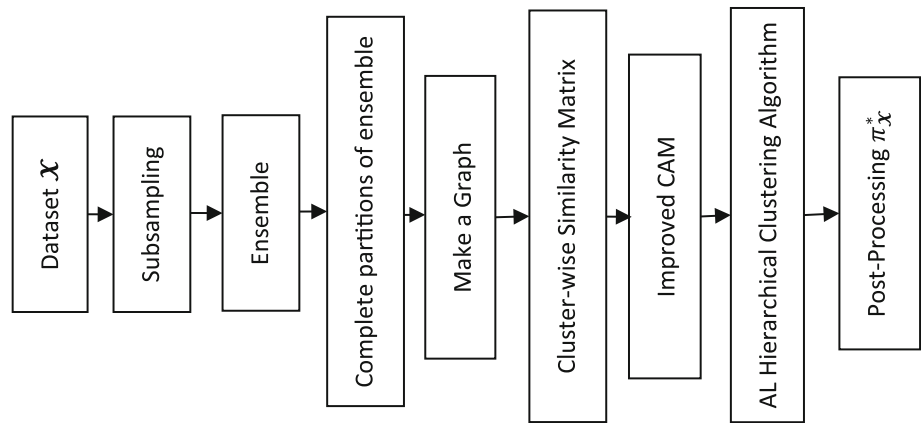
$$p_{j, \pi_{\mathcal{X}}} = j - k_{j, \pi_{\mathcal{X}}} \tag{10}$$

In Eq. 8, $SCCS_{\pi_{\mathcal{X}}}(A, B)$ stands for a simple cluster-cluster similarity function. It is the corrected version of the set correlation ratio. The set correlation first emerged in (Houle, 2008). It is defined according to Eq. 11.

$$SCCS_{\pi_{\mathcal{X}}}(\pi_{\mathcal{X}, a}^{k_1}, \pi_{\mathcal{X}, b}^{k_2}) = 0.5 + \frac{|\pi_{\mathcal{X}, a}^{k_1} \cap \pi_{\mathcal{X}, b}^{k_2}| \times |\mathcal{X}_{:,1}| - |\pi_{\mathcal{X}, a}^{k_1}| \times |\pi_{\mathcal{X}, b}^{k_2}|}{2 \times \sqrt{|\pi_{\mathcal{X}, a}^{k_1}| \times |\pi_{\mathcal{X}, b}^{k_2}| \times (|\mathcal{X}_{:,1}| - |\pi_{\mathcal{X}, a}^{k_1}|) \times (|\mathcal{X}_{:,1}| - |\pi_{\mathcal{X}, b}^{k_2}|)}} \tag{11}$$

It can be proven that $SCCS_{\pi_{\mathcal{X}}}(A, B)$ is less than or equal to one and greater than or equal to zero. If the cluster A and cluster B have maximum similarity, it will return one. If the cluster A and cluster B have minimum similarity, it will return zero (Vinh and Houle 2010). It can be proven that $w_{\pi_{\mathcal{X}}}$ is reflexive. Now, assuming we are at the j_1 th cluster,

Fig. 1 Framework of our ensemble clustering



i.e. $\tilde{\pi}_{\mathcal{X}, p_{j_1, \pi_{\mathcal{X}}}}^{k_{j_1, \pi_{\mathcal{X}}}}$, we intend to define a transition probability to cluster the j_2 th cluster, i.e. $\tilde{\pi}_{\mathcal{X}, p_{j_2, \pi_{\mathcal{X}}}}^{k_{j_2, \pi_{\mathcal{X}}}}$ after r steps denoted by $TP_{\pi_{\mathcal{X}}}^r(j_1, j_2)$. For $r = 1$, it is defined according to Eq. 12.

$$TP_{\pi_{\mathcal{X}}}^1(j_1, j_2) = \begin{cases} 0 & j_1 = j_2 \\ \frac{w_{\pi_{\mathcal{X}}}(j_1, j_2)}{\sum_{j=1}^{|\pi_{\mathcal{X}:}|} w_{\pi_{\mathcal{X}}}(j_1, j)} & j_1 \neq j_2 \end{cases} \quad (12)$$

Now, we have the matrix $TP_{\pi_{\mathcal{X}}}$ whose (i, j) th entry shows the probability that a traveller that is currently at the i th cluster will be at the j th cluster in the next step.

To compute the probability that a traveller that is currently at the i th cluster will be at the j th cluster in the next two steps, we should compute it by Eq. 13.

$$TP_{\pi_{\mathcal{X}}}^2(j_1, j_2) = \sum_{j=1}^{|\pi_{\mathcal{X}:}|} TP_{\pi_{\mathcal{X}}}^1(j_1, j) \times TP_{\pi_{\mathcal{X}}}^1(j, j_2) \quad (13)$$

Therefore, it is easy to prove that $TP_{\pi_{\mathcal{X}}}^2 = TP_{\pi_{\mathcal{X}}}^1 \times TP_{\pi_{\mathcal{X}}}^1$. Also, it is easy to show the transition probability matrix after r steps, i.e. $TP_{\pi_{\mathcal{X}}}^r$, is computed according to Eq. 14.

$$TP_{\pi_{\mathcal{X}}}^r = \prod_{k=1}^r TP_{\pi_{\mathcal{X}}}^1 \quad (14)$$

Now, we introduce a new feature space for clusters of the ensemble, denoted by $NF_{\pi_{\mathcal{X}}}^r$, as follows in Eq. 15.

$$NF_{\pi_{\mathcal{X}}}^r = [TP_{\pi_{\mathcal{X}}}^1 \ TP_{\pi_{\mathcal{X}}}^2 \ \dots \ TP_{\pi_{\mathcal{X}}}^r] \quad (15)$$

Note that each row corresponds to a cluster and each column corresponds to a feature in our new feature space. Therefore, we have $|\pi_{\mathcal{X}:}|$ rows and $|\pi_{\mathcal{X}:}| \times r$ defined features. According to a distance metric (here, we use Euclidian distance), the distances between pairs of the ensemble clusters are computed. The mentioned distances for all pairs of clusters in the ensemble are stored in a distance matrix denoted by $EuclidianDis_{\pi_{\mathcal{X}}}^r$ and defined based on Eq. 16.

$$EuclidianDis_{\pi_{\mathcal{X}}}^r(j_1, j_2) = \left(\frac{\sum_{j=1}^{|\pi_{\mathcal{X}:}|} (NF_{\pi_{\mathcal{X}}}^r(j_1, j) - NF_{\pi_{\mathcal{X}}}^r(j_2, j))^2}{|\pi_{\mathcal{X}:}|} \right)^{0.5} \quad (16)$$

Correspondingly, a distance matrix can be denoted by $CosineDis_{\pi_{\mathcal{X}}}^r$ and defined based on Eq. 17 if we use Cosine distance.

$$CosineDis_{\pi_{\mathcal{X}}}^r(j_1, j_2) = 1 - \left(\frac{\sum_{j=1}^{|\pi_{\mathcal{X}:}|} (NF_{\pi_{\mathcal{X}}}^r(j_1, j) \times NF_{\pi_{\mathcal{X}}}^r(j_2, j))}{\left(\sum_{j=1}^{|\pi_{\mathcal{X}:}|} (NF_{\pi_{\mathcal{X}}}^r(j_1, j))^2 \right)^{0.5} \times \left(\sum_{j=1}^{|\pi_{\mathcal{X}:}|} (NF_{\pi_{\mathcal{X}}}^r(j_2, j))^2 \right)^{0.5}} \right)^{0.5} \quad (17)$$

According to Eq. 18, the cluster–cluster similarity matrix, denoted by $CCS_{\pi_{\mathcal{X}}}^r$, is computed.

$$CCS_{\pi_{\mathcal{X}}}^r = \begin{cases} 1 - EuclidianDis_{\pi_{\mathcal{X}}}^r & \text{measure} = Euclidian \\ 1 - CosineDis_{\pi_{\mathcal{X}}}^r & \text{measure} = Cosine \end{cases} \quad (18)$$

Now, we define a point–point similarity matrix, denoted by $PPS_{\pi_{\mathcal{X}}}^r$, according to Eq. 19.

$$PPS_{\pi_{\mathcal{X}}}^r(j_1, j_2) = \frac{\sum_{b=1}^{|\pi_{\mathcal{X}}|} CCS_{\pi_{\mathcal{X}}}^r(CI_{\pi_{\mathcal{X}}}(\mathcal{X}_{j_1, :}, b), CI_{\pi_{\mathcal{X}}}(\mathcal{X}_{j_2, :}, b))}{|\pi_{\mathcal{X}}|} \quad (19)$$

where $CI_{\pi_{\mathcal{X}}}(\mathcal{X}_{j, :}, b)$ is the index of the cluster to which the j th data point belongs in the b th partition (note the index is computed among all clusters of ensemble). It is computed according to Eq. 20.

$$CI_{\pi_{\mathcal{X}}}(\mathcal{X}_{j, :}, b) = \sum_{i=1}^{b-1} \left| \pi_{\mathcal{X}, :}^i \right| + \arg \left[\mathcal{X}_{j, :} \in \tilde{\pi}_{\mathcal{X}, p}^b \right] \quad (20)$$

It is worthy to be mentioned that the matrix $PPS_{\pi_{\mathcal{X}}}^r$ is symmetric, i.e. $PPS_{\pi_{\mathcal{X}}}^r(j_1, j_2) = PPS_{\pi_{\mathcal{X}}}^r(j_2, j_1)$, just like $w_{\pi_{\mathcal{X}}}$.

Therefore, we propose to apply an agglomerative hierarchical complete linkage clustering algorithm on the attained point–point similarity matrix, i.e. $PPS_{\pi_{\mathcal{X}}}^r$, to cluster data points into a set of $|\pi_{\mathcal{X},:}^+|$ clusters. The mentioned obtained set of $|\pi_{\mathcal{X},:}^+|$ clusters is considered to be consensus partition.

Toy Example: We have assumed a dataset with 13 instances and three clusters. Let’s assume we want to generate an ensemble of size 5 using k-means clusterer algorithm. We have considered k-means clusterer algorithm running 5 times on our dataset to partition them into 2, 2, 3, 3 and 3 clusters, respectively. Let’s assume we have obtained an ensemble of size 5 on 13 instances. The assumptive ensemble is considered to be presented by Fig. 2. Therefore, we have ensemble $\pi_{\mathcal{X}} = \{\pi_{\mathcal{X},1}, \pi_{\mathcal{X},2}, \pi_{\mathcal{X},3}, \pi_{\mathcal{X},4}, \pi_{\mathcal{X},5}\}$ where first partition is as follows $\pi_{\mathcal{X},1} = \{\pi_{\mathcal{X},1}^1, \pi_{\mathcal{X},1}^2\}$, second partition is as follows $\pi_{\mathcal{X},2} = \{\pi_{\mathcal{X},2}^1, \pi_{\mathcal{X},2}^2\}$, third partition is as follows $\pi_{\mathcal{X},3} = \{\pi_{\mathcal{X},3}^1, \pi_{\mathcal{X},3}^2, \pi_{\mathcal{X},3}^3\}$, fourth partition is as follows $\pi_{\mathcal{X},4} = \{\pi_{\mathcal{X},4}^1, \pi_{\mathcal{X},4}^2, \pi_{\mathcal{X},4}^3\}$, and fifth partition is as follows $\pi_{\mathcal{X},5} = \{\pi_{\mathcal{X},5}^1, \pi_{\mathcal{X},5}^2, \pi_{\mathcal{X},5}^3\}$.

Also, clusters of first partition are as follows: $\pi_{\mathcal{X},1}^1$ is equal to $\{\mathcal{X}_{1,:}, \mathcal{X}_{2,:}, \mathcal{X}_{3,:}, \mathcal{X}_{4,:}, \mathcal{X}_{6,:}, \mathcal{X}_{7,:}, \mathcal{X}_{11,:}\}$, and $\pi_{\mathcal{X},1}^2$ is equal to $\{\mathcal{X}_{5,:}, \mathcal{X}_{8,:}, \mathcal{X}_{9,:}, \mathcal{X}_{10,:}, \mathcal{X}_{12,:}, \mathcal{X}_{13,:}\}$. Clusters of second partition are as follows: $\pi_{\mathcal{X},2}^1$ is equal to $\{\mathcal{X}_{1,:}, \mathcal{X}_{4,:}, \mathcal{X}_{7,:}, \mathcal{X}_{8,:}, \mathcal{X}_{9,:}, \mathcal{X}_{11,:}, \mathcal{X}_{12,:}\}$, $\pi_{\mathcal{X},2}^2$ is equal to $\{\mathcal{X}_{2,:}, \mathcal{X}_{3,:}, \mathcal{X}_{5,:}, \mathcal{X}_{6,:}, \mathcal{X}_{10,:}, \mathcal{X}_{13,:}\}$. Clusters of third partition are as follows: $\pi_{\mathcal{X},3}^1$ is equal to $\{\mathcal{X}_{1,:}, \mathcal{X}_{2,:}, \mathcal{X}_{3,:}, \mathcal{X}_{4,:}, \mathcal{X}_{11,:}, \mathcal{X}_{12,:}\}$, $\pi_{\mathcal{X},3}^2$ is equal to $\{\mathcal{X}_{8,:}, \mathcal{X}_{9,:}\}$, $\pi_{\mathcal{X},3}^3$ is equal to $\{\mathcal{X}_{5,:}, \mathcal{X}_{6,:}, \mathcal{X}_{7,:}, \mathcal{X}_{10,:}, \mathcal{X}_{13,:}\}$. Clusters of fourth partition are as follows: $\pi_{\mathcal{X},4}^1$ is equal to $\{\mathcal{X}_{3,:}, \mathcal{X}_{5,:}, \mathcal{X}_{6,:}, \mathcal{X}_{9,:}\}$, $\pi_{\mathcal{X},4}^2$ is equal to $\{\mathcal{X}_{10,:}, \mathcal{X}_{11,:}, \mathcal{X}_{12,:}, \mathcal{X}_{13,:}\}$, $\pi_{\mathcal{X},4}^3$ is equal to $\{\mathcal{X}_{1,:}, \mathcal{X}_{2,:}, \mathcal{X}_{4,:}, \mathcal{X}_{7,:}, \mathcal{X}_{8,:}\}$. Finally, clusters of last partition are as follows: $\pi_{\mathcal{X},5}^1$ is equal to $\{\mathcal{X}_{9,:}, \mathcal{X}_{10,:}, \mathcal{X}_{11,:}, \mathcal{X}_{12,:}, \mathcal{X}_{13,:}\}$, $\pi_{\mathcal{X},5}^2$ is equal to $\{\mathcal{X}_{3,:}, \mathcal{X}_{5,:}, \mathcal{X}_{6,:}\}$, $\pi_{\mathcal{X},5}^3$ is equal to $\{\mathcal{X}_{1,:}, \mathcal{X}_{2,:}, \mathcal{X}_{4,:}, \mathcal{X}_{7,:}, \mathcal{X}_{8,:}\}$.

$SCCS_{\pi_{\mathcal{X}}}$ for the ensemble presented in Fig. 2 is depicted in Fig. 3. Now, we have a graph $\mathcal{G}(\pi_{\mathcal{X}})$ where set of its vertices, i.e. $\mathcal{V}(\pi_{\mathcal{X}})$, is $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$ and set of its edges, i.e. $\mathcal{E}(\pi_{\mathcal{X}})$, is presented in Fig. 4. Figure 5 depicts cluster–cluster similarity matrix, i.e.

	$\pi_{\mathcal{X},1}$	$\pi_{\mathcal{X},2}$	$\pi_{\mathcal{X},3}$	$\pi_{\mathcal{X},4}$	$\pi_{\mathcal{X},5}$
$\mathcal{X}_{1,:}$					
$\mathcal{X}_{2,:}$					
$\mathcal{X}_{3,:}$					
$\mathcal{X}_{4,:}$					
$\mathcal{X}_{5,:}$					
$\mathcal{X}_{6,:}$					
$\mathcal{X}_{7,:}$					
$\mathcal{X}_{8,:}$					
$\mathcal{X}_{9,:}$					
$\mathcal{X}_{10,:}$					
$\mathcal{X}_{11,:}$					
$\mathcal{X}_{12,:}$					
$\mathcal{X}_{13,:}$					

Fig. 2 A clustering ensemble with four base partitions

$CCS_{\pi_{\mathcal{X}}}^2$, when (a) Euclidian distance, i.e. $EuclidianDis_{\pi_{\mathcal{X}}}^2$, or (b) Cosine distance, i.e. $CosineDis_{\pi_{\mathcal{X}}}^2$ is used for our toy ensemble. Finally, Fig. 6 depicts point–point similarity matrix, i.e. $PPS_{\pi_{\mathcal{X}}}^2$, when (a) Euclidian distance, i.e. $EuclidianDis_{\pi_{\mathcal{X}}}^2$, or (b) Cosine distance, i.e. $CosineDis_{\pi_{\mathcal{X}}}^2$ is used for our toy ensemble.

5 Experimental study

5.1 Datasets

Many real-world and synthetic data sets are here employed for assessment. The real-world data sets are as follows: Bupa, Breast-cancer, Glass, Galaxy, Ionosphere, Iris, ISOLET, Landsat-satellite, Letter-Recognition, SAHeart, Wine, and Yeast from UCI machine learning repository (Newman et al. 1998) together with an auxiliary real data set of USPS (Dueck 2009). Bupa, Breast-cancer, Glass, Galaxy, Ionosphere, Iris, ISOLET, Landsat-satellite, Letter-Recognition, SAHeart, USPS, Wine, and Yeast data sets have 345, 683, 214, 323, 351, 150, 7,797, 6,435, 20,000, 462, 11,000, 178, and 1,484 data samples, respectively; they also have 6, 9, 9, 4, 34, 4, 617, 36, 16, 9, 256, 13, and 8 features and 2, 2, 6, 7, 2, 3, 26, 6, 26, 2, 10, 3, and 10 clusters, respectively.

	$\pi_{x,1}^1$	$\pi_{x,1}^2$	$\pi_{x,2}^1$	$\pi_{x,2}^2$	$\pi_{x,3}^1$	$\pi_{x,3}^2$	$\pi_{x,3}^3$	$\pi_{x,4}^1$	$\pi_{x,4}^2$	$\pi_{x,4}^3$	$\pi_{x,5}^1$	$\pi_{x,5}^2$	$\pi_{x,5}^3$
$\pi_{x,1}^1$	1	0	0.54	0.46	0.77	0.27	0.39	0.47	0.31	0.71	0.23	0.57	0.71
$\pi_{x,1}^2$	0	1	0.46	0.54	0.23	0.73	0.61	0.53	0.69	0.29	0.77	0.43	0.29
$\pi_{x,2}^1$	0.54	0.46	1	0	0.62	0.7	0.23	0.31	0.47	0.71	0.55	0.2	0.71
$\pi_{x,2}^2$	0.46	0.54	0	1	0.38	0.3	0.77	0.69	0.53	0.29	0.45	0.8	0.29
$\pi_{x,3}^1$	0.77	0.23	0.62	0.38	1	0.3	0.13	0.36	0.53	0.61	0.45	0.43	0.61
$\pi_{x,3}^2$	0.27	0.73	0.7	0.3	0.3	1	0.33	0.59	0.36	0.55	0.55	0.38	0.55
$\pi_{x,3}^3$	0.39	0.61	0.23	0.77	0.13	0.33	1	0.58	0.58	0.35	0.51	0.66	0.35
$\pi_{x,4}^1$	0.47	0.53	0.31	0.69	0.36	0.59	0.58	1	0.28	0.24	0.41	0.91	0.24
$\pi_{x,4}^2$	0.31	0.69	0.47	0.53	0.53	0.36	0.58	0.28	1	0.24	0.92	0.32	0.24
$\pi_{x,4}^3$	0.71	0.29	0.71	0.29	0.61	0.55	0.35	0.24	0.24	1	0.19	0.28	1
$\pi_{x,5}^1$	0.23	0.77	0.55	0.45	0.45	0.55	0.51	0.41	0.92	0.19	1	0.28	0.19
$\pi_{x,5}^2$	0.57	0.43	0.2	0.8	0.43	0.38	0.66	0.91	0.32	0.28	0.28	1	0.28
$\pi_{x,5}^3$	0.71	0.29	0.71	0.29	0.61	0.55	0.35	0.24	0.24	1	0.19	0.28	1

Fig. 3 The $SCCS_{\pi_{x^i}}$ for the ensemble presented in Fig. 2

	$\pi_{x,1}^1$	$\pi_{x,1}^2$	$\pi_{x,2}^1$	$\pi_{x,2}^2$	$\pi_{x,3}^1$	$\pi_{x,3}^2$	$\pi_{x,3}^3$	$\pi_{x,4}^1$	$\pi_{x,4}^2$	$\pi_{x,4}^3$	$\pi_{x,5}^1$	$\pi_{x,5}^2$	$\pi_{x,5}^3$
$\pi_{x,1}^1$	0	0	0.1	0.09	0.14	0.05	0.07	0.09	0.06	0.13	0.04	0.11	0.13
$\pi_{x,1}^2$	0	0	0.08	0.1	0.04	0.13	0.11	0.09	0.12	0.05	0.14	0.08	0.05
$\pi_{x,2}^1$	0.1	0.08	0	0	0.11	0.13	0.04	0.06	0.09	0.13	0.1	0.04	0.13
$\pi_{x,2}^2$	0.08	0.1	0	0	0.07	0.05	0.14	0.13	0.1	0.05	0.08	0.14	0.05
$\pi_{x,3}^1$	0.14	0.04	0.11	0.07	0	0.06	0.02	0.07	0.1	0.11	0.08	0.08	0.11
$\pi_{x,3}^2$	0.05	0.13	0.12	0.05	0.05	0	0.06	0.1	0.06	0.1	0.1	0.07	0.1
$\pi_{x,3}^3$	0.07	0.11	0.04	0.14	0.02	0.06	0	0.11	0.11	0.06	0.09	0.12	0.06
$\pi_{x,4}^1$	0.08	0.09	0.05	0.12	0.06	0.11	0.1	0	0.05	0.04	0.07	0.16	0.04
$\pi_{x,4}^2$	0.06	0.13	0.09	0.1	0.1	0.07	0.11	0.05	0	0.04	0.17	0.06	0.04
$\pi_{x,4}^3$	0.13	0.05	0.13	0.05	0.11	0.1	0.06	0.04	0.04	0	0.03	0.05	0.18
$\pi_{x,5}^1$	0.04	0.14	0.1	0.08	0.08	0.1	0.09	0.07	0.17	0.03	0	0.05	0.03
$\pi_{x,5}^2$	0.1	0.08	0.04	0.14	0.08	0.07	0.12	0.16	0.06	0.05	0.05	0	0.05
$\pi_{x,5}^3$	0.13	0.05	0.13	0.05	0.11	0.1	0.06	0.04	0.04	0.18	0.03	0.05	0

Fig. 4 The edges of the transition graph, i.e. $\mathcal{E}(\pi_{x^i})$, for the ensemble presented in Fig. 2

Our benchmark includes two dimensions synthetic data sets including *Artificial*₁, *Artificial*₂, *Artificial*₃, and Halfring. Halfring data set contains two clusters and 400 instances, while other three data sets contain three clusters and 300 instances. The scatter plots of these synthetic data sets are depicted in Fig. 7. We have standardized all of our mentioned data sets so that in all datasets, each feature is transformed into a new range with distribution of $N(0, 1)$ and the features with missed values are removed from the data sets.

5.2 Parameters

The ensemble generation has been accomplished according to the technique introduced in Ren et al. (2013). The

k-means clustering algorithm was employed for 50 different subsets of a given data set to generate ensemble of size 50. The sampling ratio employed in this paper is 0.8.

The following recent clustering ensembles have been employed as a benchmark to this paper: HBGF (Fern and Brodley 2004), CO-AL (Fred and Jain 2005), SRS (Iam-On et al. 2008), WCT (Iam-On et al. 2011), DICLENS (Mimaroglu and Aksehirli 2012), ONCE-AL (Alqurashi and Wang 2014), CSEAC (Alizadeh et al. 2014a), DSCE (Alqurashi and Wang 2015), WEAC (Huang et al. 2015), GPMGLA (Huang et al. 2015), WCE (Alizadeh et al. 2015) ECSEAC (Parvin and Minaei-Bidgoli 2015) and TME (Zhong et al. 2015).

For a given clustering ensemble, its initialization is performed according to its corresponding paper. The

	$\pi_{\mathcal{X},1}^1$	$\pi_{\mathcal{X},1}^2$	$\pi_{\mathcal{X},2}^1$	$\pi_{\mathcal{X},2}^2$	$\pi_{\mathcal{X},3}^1$	$\pi_{\mathcal{X},3}^2$	$\pi_{\mathcal{X},3}^3$	$\pi_{\mathcal{X},4}^1$	$\pi_{\mathcal{X},4}^2$	$\pi_{\mathcal{X},4}^3$	$\pi_{\mathcal{X},5}^1$	$\pi_{\mathcal{X},5}^2$	$\pi_{\mathcal{X},5}^3$
$\pi_{\mathcal{X},1}^1$	1	0.64	0.61	0.56	0.61	0.65	0.57	0.57	0.58	0.64	0.57	0.57	0.64
$\pi_{\mathcal{X},1}^2$	0.64	1	0.57	0.62	0.59	0.58	0.64	0.65	0.63	0.55	0.64	0.63	0.55
$\pi_{\mathcal{X},2}^1$	0.61	0.57	1	0.64	0.63	0.59	0.58	0.57	0.58	0.65	0.57	0.57	0.65
$\pi_{\mathcal{X},2}^2$	0.56	0.62	0.64	1	0.56	0.63	0.63	0.66	0.63	0.54	0.64	0.64	0.54
$\pi_{\mathcal{X},3}^1$	0.61	0.59	0.63	0.56	1	0.7	0.69	0.59	0.57	0.67	0.57	0.57	0.67
$\pi_{\mathcal{X},3}^2$	0.65	0.58	0.59	0.63	0.7	1	0.69	0.58	0.7	0.61	0.64	0.62	0.61
$\pi_{\mathcal{X},3}^3$	0.57	0.64	0.58	0.63	0.69	0.69	1	0.69	0.64	0.54	0.65	0.65	0.54
$\pi_{\mathcal{X},4}^1$	0.57	0.65	0.57	0.66	0.59	0.58	0.69	1	0.7	0.59	0.65	0.6	0.59
$\pi_{\mathcal{X},4}^2$	0.58	0.63	0.58	0.63	0.57	0.7	0.64	0.7	1	0.59	0.6	0.65	0.59
$\pi_{\mathcal{X},4}^3$	0.64	0.55	0.65	0.54	0.67	0.61	0.54	0.59	0.59	1	0.59	0.58	0.64
$\pi_{\mathcal{X},5}^1$	0.57	0.64	0.57	0.64	0.57	0.64	0.65	0.65	0.6	0.59	1	0.67	0.59
$\pi_{\mathcal{X},5}^2$	0.57	0.63	0.57	0.64	0.57	0.62	0.65	0.6	0.65	0.58	0.67	1	0.58
$\pi_{\mathcal{X},5}^3$	0.64	0.55	0.65	0.54	0.67	0.61	0.54	0.59	0.59	0.64	0.59	0.58	1

(a)

	$\pi_{\mathcal{X},1}^1$	$\pi_{\mathcal{X},1}^2$	$\pi_{\mathcal{X},2}^1$	$\pi_{\mathcal{X},2}^2$	$\pi_{\mathcal{X},3}^1$	$\pi_{\mathcal{X},3}^2$	$\pi_{\mathcal{X},3}^3$	$\pi_{\mathcal{X},4}^1$	$\pi_{\mathcal{X},4}^2$	$\pi_{\mathcal{X},4}^3$	$\pi_{\mathcal{X},5}^1$	$\pi_{\mathcal{X},5}^2$	$\pi_{\mathcal{X},5}^3$
$\pi_{\mathcal{X},1}^1$	1	0.72	0.66	0.58	0.66	0.73	0.58	0.61	0.62	0.73	0.6	0.6	0.73
$\pi_{\mathcal{X},1}^2$	0.72	1	0.59	0.69	0.64	0.62	0.72	0.74	0.72	0.58	0.73	0.71	0.58
$\pi_{\mathcal{X},2}^1$	0.66	0.59	1	0.72	0.7	0.63	0.6	0.6	0.62	0.73	0.6	0.6	0.73
$\pi_{\mathcal{X},2}^2$	0.58	0.69	0.72	1	0.58	0.69	0.71	0.75	0.71	0.55	0.72	0.72	0.55
$\pi_{\mathcal{X},3}^1$	0.66	0.64	0.7	0.58	1	0.8	0.79	0.65	0.6	0.77	0.61	0.61	0.77
$\pi_{\mathcal{X},3}^2$	0.73	0.62	0.63	0.69	0.8	1	0.79	0.62	0.8	0.69	0.72	0.69	0.69
$\pi_{\mathcal{X},3}^3$	0.58	0.72	0.6	0.71	0.79	0.79	1	0.79	0.72	0.57	0.75	0.74	0.57
$\pi_{\mathcal{X},4}^1$	0.61	0.74	0.6	0.75	0.65	0.62	0.79	1	0.82	0.66	0.74	0.67	0.66
$\pi_{\mathcal{X},4}^2$	0.62	0.72	0.62	0.71	0.6	0.8	0.72	0.82	1	0.66	0.68	0.75	0.66
$\pi_{\mathcal{X},4}^3$	0.73	0.58	0.73	0.55	0.77	0.69	0.57	0.66	0.66	1	0.66	0.63	0.74
$\pi_{\mathcal{X},5}^1$	0.6	0.73	0.6	0.72	0.61	0.72	0.75	0.74	0.68	0.66	1	0.77	0.66
$\pi_{\mathcal{X},5}^2$	0.6	0.71	0.6	0.72	0.61	0.69	0.74	0.67	0.75	0.63	0.77	1	0.63
$\pi_{\mathcal{X},5}^3$	0.73	0.58	0.73	0.55	0.77	0.69	0.57	0.66	0.66	0.74	0.66	0.63	1

(b)

Fig. 5 Cluster–cluster similarity matrix, i.e. $CCS_{\pi_{\mathcal{X}}}^2$, when (a) Euclidian distance, i.e. $EuclidianDis_{\pi_{\mathcal{X}}}^2$, or (b) Cosine distance, i.e. $CosineDis_{\pi_{\mathcal{X}}}^2$ is used for the ensemble presented in Fig. 2

results of the paper are provided after averaging on 30 independent runs, while all methods are considered with the ensemble size of 100.

5.3 Evaluation metrics

Internal measures such as silhouette or sum of square errors (SSE) computed without caring about the ground-truth labels of the data could be used to assess a clustering result on a given data set. External measures computed considering the ground-truth labels of the data are other set of the

measures to assess a resultant partition on a given data set. Note that the ground-truth labels of the data are only used for assessment after obtaining the resultant partition, and they are not used during the clustering task. Although both the internal and external sets of the measures are popular, we have only used four external measures of adjust rand index (ARI), normalized mutual information (NMI) and accuracy (ACC) rate and F-measure (FM). Given the ground-truth labels $\pi_{\mathcal{X}}^+$, adjust rand index of a resultant partition $\pi_{\mathcal{X}}^*$ is achieved using the following equation

$$ARI(\pi_{\mathcal{X}}^*, \pi_{\mathcal{X}}^+) = \frac{\sum_{i=1}^{|\pi_{\mathcal{X}}^*|} \sum_{j=1}^{|\pi_{\mathcal{X}}^+|} \binom{|\pi_{\mathcal{X},i}^* \cap \pi_{\mathcal{X},j}^+|}{2} - \frac{\sum_{i=1}^{|\pi_{\mathcal{X}}^*|} \binom{|\pi_{\mathcal{X},i}^*|}{2} \times \sum_{i=1}^{|\pi_{\mathcal{X}}^+|} \binom{|\pi_{\mathcal{X},i}^+|}{2}}{\binom{|\mathcal{X}_{:,1}|}{2}}}{\frac{\sum_{i=1}^{|\pi_{\mathcal{X}}^*|} \binom{|\pi_{\mathcal{X},i}^*|}{2} + \sum_{i=1}^{|\pi_{\mathcal{X}}^+|} \binom{|\pi_{\mathcal{X},i}^+|}{2}}{\binom{|\mathcal{X}_{:,1}|}{2}} - \frac{\sum_{i=1}^{|\pi_{\mathcal{X}}^*|} \binom{|\pi_{\mathcal{X},i}^*|}{2} \times \sum_{i=1}^{|\pi_{\mathcal{X}}^+|} \binom{|\pi_{\mathcal{X},i}^+|}{2}}{\binom{|\mathcal{X}_{:,1}|}{2}}} \tag{21}$$

where $|\pi_{\mathcal{X},i}^* \cap \pi_{\mathcal{X},j}^+|$ is the number of data points shared between the i th cluster of consensus partition $\pi_{\mathcal{X}}^*$ and the j th cluster of target partition $\pi_{\mathcal{X}}^+$. The binomial function $\binom{x}{2}$ is obtained as

$$\binom{x}{2} = \frac{x!}{(x-2)!2!} = \frac{x \times (x-1)}{2} \tag{22}$$

Given the ground-truth partition $\pi_{\mathcal{X}}^+$, the normalized mutual information for a resultant partition $\pi_{\mathcal{X}}^*$ is achieved

	$\mathcal{X}_{1,:}$	$\mathcal{X}_{2,:}$	$\mathcal{X}_{3,:}$	$\mathcal{X}_{4,:}$	$\mathcal{X}_{5,:}$	$\mathcal{X}_{6,:}$	$\mathcal{X}_{7,:}$	$\mathcal{X}_{8,:}$	$\mathcal{X}_{9,:}$	$\mathcal{X}_{10,:}$	$\mathcal{X}_{11,:}$	$\mathcal{X}_{12,:}$	$\mathcal{X}_{13,:}$
$\mathcal{X}_{1,:}$	1	0.93	0.76	1	0.63	0.7	0.94	0.87	0.71	0.63	0.84	0.76	0.63
$\mathcal{X}_{2,:}$	0.93	1	0.83	0.93	0.7	0.77	0.87	0.8	0.63	0.7	0.76	0.69	0.7
$\mathcal{X}_{3,:}$	0.76	0.83	1	0.76	0.87	0.94	0.7	0.63	0.73	0.74	0.8	0.73	0.74
$\mathcal{X}_{4,:}$	1	0.93	0.76	1	0.63	0.7	0.94	0.87	0.71	0.63	0.84	0.76	0.63
$\mathcal{X}_{5,:}$	0.63	0.7	0.87	0.63	1	0.93	0.69	0.7	0.8	0.87	0.67	0.74	0.87
$\mathcal{X}_{6,:}$	0.7	0.77	0.94	0.7	0.93	1	0.76	0.63	0.73	0.8	0.74	0.67	0.8
$\mathcal{X}_{7,:}$	0.94	0.87	0.7	0.94	0.69	0.76	1	0.87	0.7	0.69	0.77	0.7	0.69
$\mathcal{X}_{8,:}$	0.87	0.8	0.63	0.87	0.7	0.63	0.87	1	0.84	0.7	0.7	0.78	0.7
$\mathcal{X}_{9,:}$	0.71	0.63	0.73	0.71	0.8	0.73	0.7	0.84	1	0.81	0.81	0.88	0.81
$\mathcal{X}_{10,:}$	0.63	0.7	0.74	0.63	0.87	0.8	0.69	0.7	0.81	1	0.79	0.87	1
$\mathcal{X}_{11,:}$	0.84	0.76	0.8	0.84	0.67	0.74	0.77	0.7	0.81	0.79	1	0.93	0.79
$\mathcal{X}_{12,:}$	0.76	0.69	0.73	0.76	0.74	0.67	0.7	0.78	0.88	0.87	0.93	1	0.87
$\mathcal{X}_{13,:}$	0.63	0.7	0.74	0.63	0.87	0.8	0.69	0.7	0.81	1	0.79	0.87	1

(a)

	$\mathcal{X}_{1,:}$	$\mathcal{X}_{2,:}$	$\mathcal{X}_{3,:}$	$\mathcal{X}_{4,:}$	$\mathcal{X}_{5,:}$	$\mathcal{X}_{6,:}$	$\mathcal{X}_{7,:}$	$\mathcal{X}_{8,:}$	$\mathcal{X}_{9,:}$	$\mathcal{X}_{10,:}$	$\mathcal{X}_{11,:}$	$\mathcal{X}_{12,:}$	$\mathcal{X}_{13,:}$
$\mathcal{X}_{1,:}$	1	0.94	0.8	1	0.7	0.76	0.96	0.9	0.77	0.71	0.86	0.81	0.71
$\mathcal{X}_{2,:}$	0.94	1	0.86	0.94	0.76	0.82	0.9	0.85	0.71	0.77	0.81	0.75	0.77
$\mathcal{X}_{3,:}$	0.8	0.86	1	0.8	0.9	0.96	0.76	0.71	0.8	0.82	0.86	0.8	0.82
$\mathcal{X}_{4,:}$	1	0.94	0.8	1	0.7	0.76	0.96	0.9	0.77	0.71	0.86	0.81	0.71
$\mathcal{X}_{5,:}$	0.7	0.76	0.9	0.7	1	0.94	0.75	0.76	0.85	0.92	0.76	0.82	0.92
$\mathcal{X}_{6,:}$	0.76	0.82	0.96	0.76	0.94	1	0.8	0.7	0.8	0.86	0.82	0.76	0.86
$\mathcal{X}_{7,:}$	0.96	0.9	0.76	0.96	0.75	0.8	1	0.9	0.77	0.75	0.82	0.77	0.75
$\mathcal{X}_{8,:}$	0.9	0.85	0.71	0.9	0.76	0.7	0.9	1	0.87	0.76	0.77	0.82	0.76
$\mathcal{X}_{9,:}$	0.77	0.71	0.8	0.77	0.85	0.8	0.77	0.87	1	0.86	0.87	0.92	0.86
$\mathcal{X}_{10,:}$	0.71	0.77	0.82	0.71	0.92	0.86	0.75	0.76	0.86	1	0.85	0.9	1
$\mathcal{X}_{11,:}$	0.86	0.81	0.86	0.86	0.76	0.82	0.82	0.77	0.87	0.85	1	0.94	0.85
$\mathcal{X}_{12,:}$	0.81	0.75	0.8	0.81	0.82	0.76	0.77	0.82	0.92	0.9	0.94	1	0.9
$\mathcal{X}_{13,:}$	0.71	0.77	0.82	0.71	0.92	0.86	0.75	0.76	0.86	1	0.85	0.9	1

(b)

Fig. 6 Point–point similarity matrix, i.e. $PPS_{\pi_{\mathcal{X}}}^2$, when (a) Euclidian distance, i.e. $EuclidianDis_{\pi_{\mathcal{X}}}^2$, or (b) Cosine distance, i.e. $CosineDis_{\pi_{\mathcal{X}}}^2$ is used for the ensemble presented in Fig. 2

$$NMI(\pi_{\mathcal{X}}^*, \pi_{\mathcal{X}}^+) = \frac{\sum_{i=1}^{|\pi_{\mathcal{X}}^*|} \sum_{j=1}^{|\pi_{\mathcal{X}}^+|} |\pi_{\mathcal{X},i}^* \cap \pi_{\mathcal{X},j}^+| \log_2 \left(\frac{|\mathcal{X}_{:,1}| \times |\pi_{\mathcal{X},i}^* \cap \pi_{\mathcal{X},j}^+|}{|\pi_{\mathcal{X},i}^*| |\pi_{\mathcal{X},j}^+|} \right)}{\sqrt{\sum_{i=1}^{c^*} \left(|\pi_{\mathcal{X},i}^*| \times \log_2 \left(\frac{|\pi_{\mathcal{X},i}^*|}{|\mathcal{X}_{:,1}|} \right) \right) \times \sum_{i=1}^c \left(|\pi_{\mathcal{X},j}^+| \times \log_2 \left(\frac{|\pi_{\mathcal{X},j}^+|}{|\mathcal{X}_{:,1}|} \right) \right)}} \tag{23}$$

by

Also, accuracy of π_* is computed by

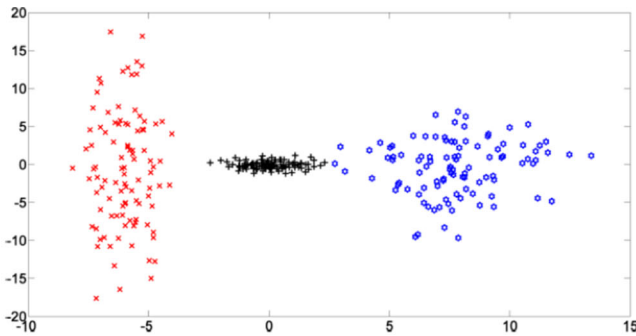
$$ACC(\pi_{\mathcal{X}}^*, \pi_{\mathcal{X}}^+) = \max_{\sigma} \sum_{j=1}^{|\pi_{\mathcal{X}}^+|} \frac{|\pi_{\mathcal{X},\sigma_j}^* \cap \pi_{\mathcal{X},j}^+|}{|\mathcal{X}_{:,1}|} \tag{24}$$

where the relabeling task σ is a permutation of $\{1, 2, \dots, |\pi_{\mathcal{X}}^*|\}$ so that $\sigma_j \in \{1, 2, \dots, |\pi_{\mathcal{X},:}^*|\}$, $\forall i \in \{1, 2, \dots, |\pi_{\mathcal{X},:}^*|\}$ and $\forall i, j \in \{1, 2, \dots, |\pi_{\mathcal{X},:}^*|\}$: $(i \neq j) \rightarrow (\sigma_i \neq \sigma_j)$.

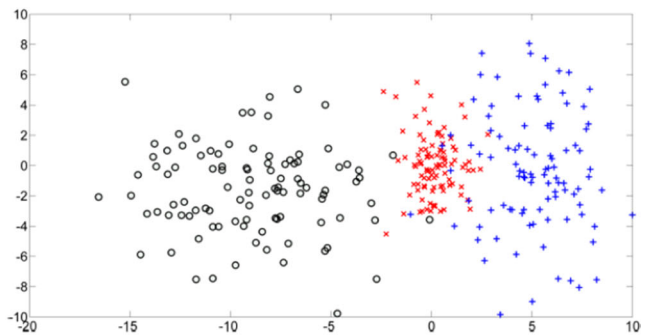
Finally, for a given target partition $\pi_{\mathcal{X},:}^+$, F-measure of a resultant partition $\pi_{\mathcal{X},:}^*$ is acquired by

$$FM(\pi_{\mathcal{X}}^*, \pi_{\mathcal{X}}^+) = \max_{\sigma} \sum_{j=1}^{|\pi_{\mathcal{X}}^+|} \frac{2 \times |\pi_{\mathcal{X},\sigma_j}^*| \times \frac{|\pi_{\mathcal{X},\sigma_j}^* \cap \pi_{\mathcal{X},j}^+|}{|\pi_{\mathcal{X},j}^+|} \times \frac{|\pi_{\mathcal{X},\sigma_j}^* \cap \pi_{\mathcal{X},j}^+|}{|\pi_{\mathcal{X},\sigma_j}^*|}}{|\mathcal{X}_{:,1}| \times \left(\frac{|\pi_{\mathcal{X},\sigma_j}^* \cap \pi_{\mathcal{X},j}^+|}{|\pi_{\mathcal{X},j}^+|} + \frac{|\pi_{\mathcal{X},\sigma_j}^* \cap \pi_{\mathcal{X},j}^+|}{|\pi_{\mathcal{X},\sigma_j}^*|} \right)} \tag{25}$$

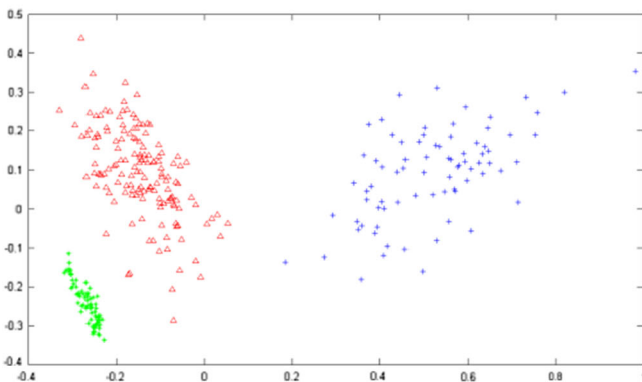
The mentioned measures vary in the range of [0, 1]. For any of them, the greater values indicate better quality of



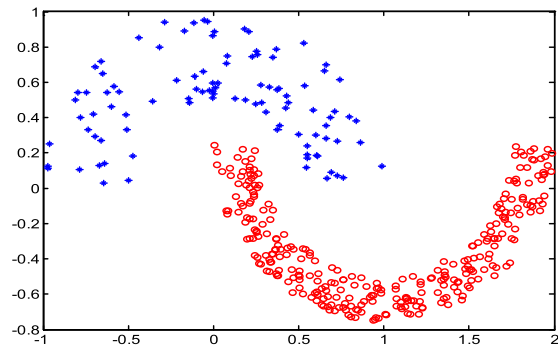
(a) *Artificial₁* dataset



(b) *Artificial₂* dataset



(c) *Artificial₃* dataset



(d) Halfring dataset

Fig. 7. 4 artificial benchmark datasets

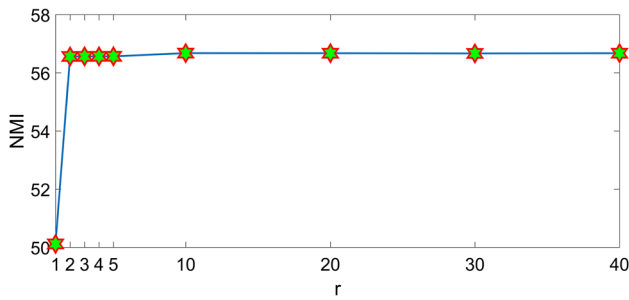


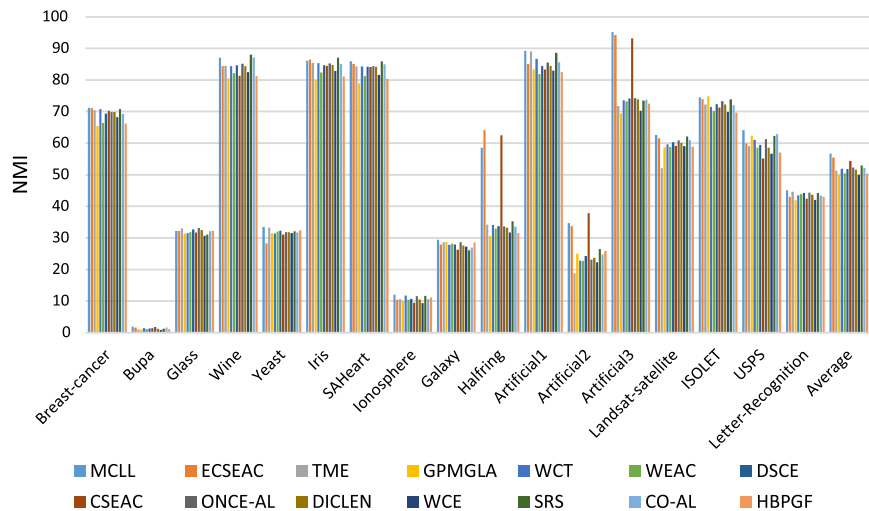
Fig. 8 Effect of parameter r on the performance of the MLCC method in terms of normalized mutual information

resultant partition π_{χ}^* . ACC and FM measures are asymmetric, but ARI and NMI are symmetric measures.

5.4 Numerical empirical analysis

The analysis of parameter r in our method, i.e. MLCC, is presented in Fig. 8. Figure 8 depicts performance of MLCC for different values of parameter r in terms of NMI.

Fig. 9 Comparison of the MLCC method with the state-of-the-art baseline methods on different data sets in terms of normalized mutual information



Friedman's ANOVA Table					
Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	1734.18	13	133.398	99.11	2.46882e-15
Error	2132.82	208	10.254		
Total	3867	237			

Table 1 Summary of the presented results in Fig. 9

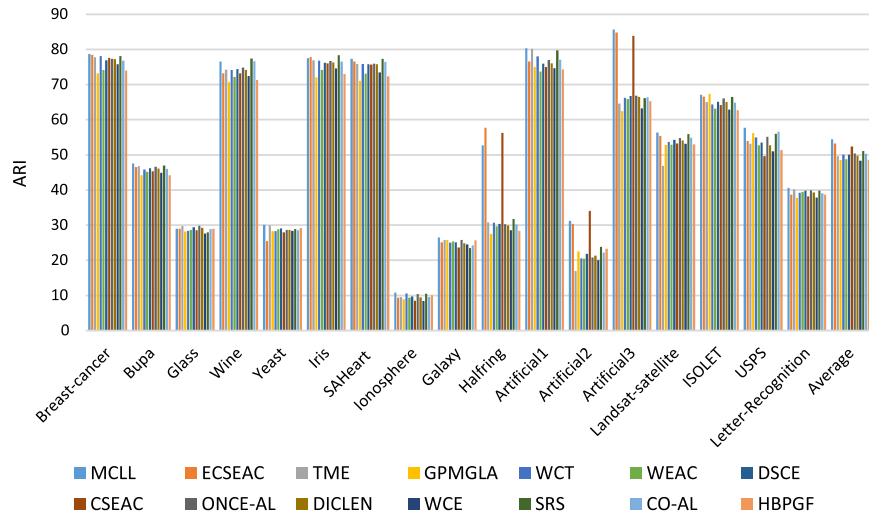
	HBPGF	CO-AL	SRS	WCE	DICLEN	ONCE-AL	CSEAC	DSCE	WEAC	WCT	GPMGLA	TME	ECSEAC	CFTLC
Mean	50.32	52.13	52.94	49.98	51.57	52.23	54.41	51.81	50.46	51.86	50.19	51.35	55.46	56.66
w-t-l	17-0-0	14-3-0	10-5-2	17-0-0	15-2-0	14-2-1	12-3-2	15-2-0	17-0-0	14-2-1	16-2-0	14-3-0	8-7-2	-

w , t , and l are three integer numbers in any triple $w-t-l$. The proposed method is the superior method and the loser method in w and l data sets, respectively. However, it is neither the loser nor the winner method in t data sets. All tests have been done by paired t -test

According to Figs. 10, 11 and 12, the following conclusions could be drawn. First, the MLCC algorithm outperforms the state-of-the-art baseline methods in most of the used benchmark data sets. Second, although the state-of-the-art baseline methods outperform the MLCC algorithm in some of the used benchmark data sets, it is always among the top three best methods. All of the results

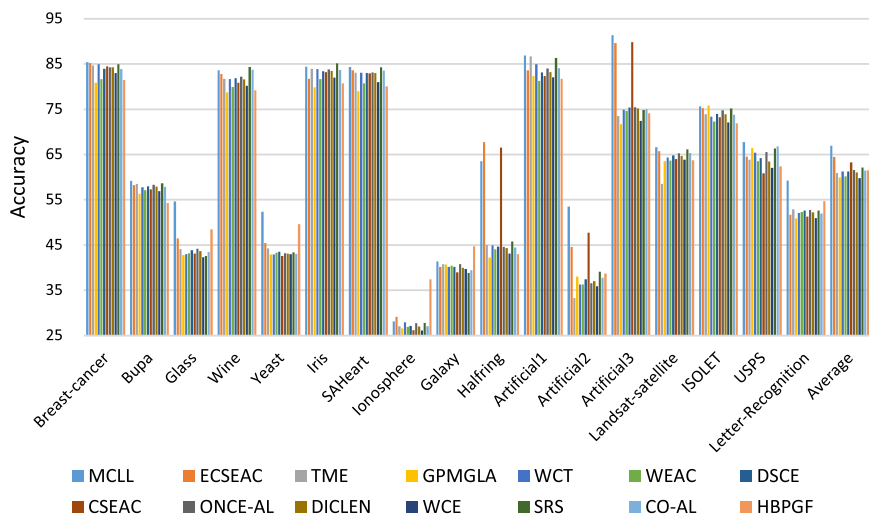
presented in Figs. 9, 10, 11 and 12 have been validated by Friedman ANOVA test. The Friedman ANOVA test shows that there is always significant difference. For further analysis, the detailed results of Figs. 10, 11, 12 and 13 are presented in Table 2. In Table 2, the results of different clustering ensembles in terms of our four metrics have been presented.

Fig. 10 Comparison of the MLCC method with the state-of-the-art baseline methods on different data sets in terms of adjacent rand index



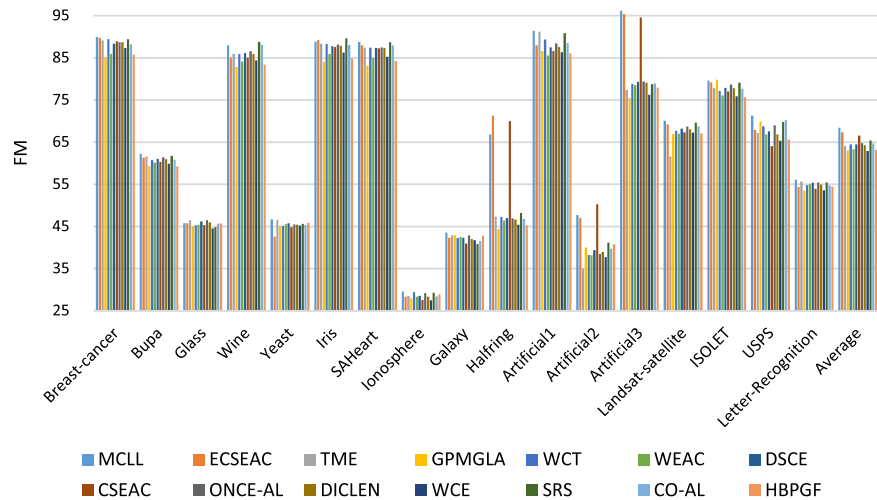
Friedman's ANOVA Table					
Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	1794.68	13	138.052	102.58	5.23934e-16
Error	2071.82	208	9.961		
Total	3866.5	237			

Fig. 11 Comparison of the MLCC method with the state-of-the-art baseline methods on different data sets in terms of accuracy

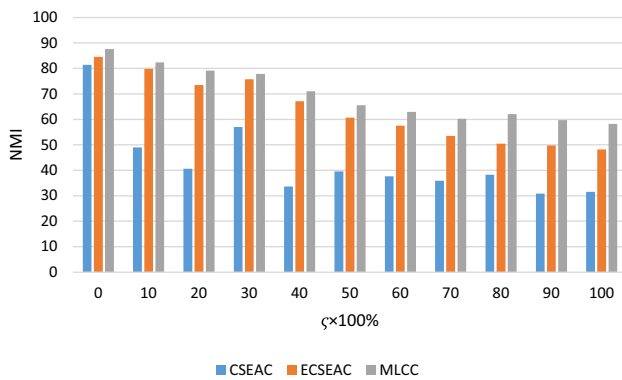


Friedman's ANOVA Table					
Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	1763.82	13	135.679	100.82	1.15256e-15
Error	2102.68	208	10.109		
Total	3866.5	237			

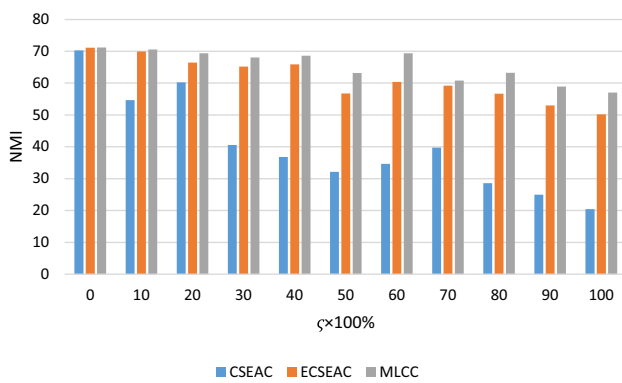
Fig. 12 Comparison of the MLCC method with the state-of-the-art baseline methods on different data sets in terms of f-measure



Friedman's ANOVA Table					
Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	1791.91	13	137.839	102.43	5.58999e-16
Error	2074.09	208	9.972		
Total	3866	237			



(a)



(b)

Fig. 13 Performances of different methods in the presence of different levels of noise in terms of normalized mutual information for (a) Iris data set and (b) Wine data set

We can also prove that the MLCC method has the time complexity of $O(r \times |\pi_{\mathcal{X}_i}|^3)$ or $O(r|\pi_{\mathcal{X}_i}|^3|\pi_{\mathcal{X}}|^3)$ in its worst case.

5.5 Noise-resistance analysis

We have examined the noise effect by adding Gaussian noise with different energy levels and analyzing robustness of the MLCC, ECSEAC and CSEAC as three best methods achieved by Table 1. The i th cluster of the data set is assumed to have the covariance of Σ_i , while \aleph_i is considered to be a temporary noise data which has the same size as the cluster of data set mentioned for a noise ratio of ς . Data samples of \aleph_i are with the covariance matrix of $\varsigma \times \Sigma_i$. The data samples of that temporary noise data are here added to the i th cluster's data samples. For all clusters in data set, the data samples of that temporary noise data with the covariance matrix $\varsigma \times \Sigma_i$ are added to the data samples of the i th cluster to obtain a new noisy data set with noise ratio $\varsigma \times 100\%$. We can easily observe that there is no noise when $\varsigma = 0$. Figure 13 includes the results of the Iris and Wine data sets where NMI value is shown on the vertical axis and amount of the noise ratio ς is shown on the horizontal axis. Apparently, by adding noise level, the gap between MLCC and other methods will be larger. It means MLCC is more robust against noise.

Table 2 The performance of the proposed method and the other methods in terms of NMIIARI/Accuracy/IFM

	MCLL	ECSEAC	TME	GPMGLA	WCT	WEAC	DSCE
Breast-cancer	71.1678,6785,4389,93	71.0978,4385,2389,71	70.4177,7784,6789,13	65.3573,2280,8385,08	70.7778,0984,9589,42	66.4174,1781,6385,93	69.4176,8783,9188,33
Bupa	1.8547,5759,1762,28	1.6546,4958,2561,32	1.0146,8158,5361,61	0.9244,2156,3359,3	1.3745,8757,7460,78	1.1645,1457,1260,13	1.346,1757,9961,04
Glass	32.1928,9754,6445,75	32.1728,9546,4545,74	33.0729,7644,1346,46	31.3128,1842,845,05	31.5528,442,9845,24	31.8128,6343,1845,45	32.7129,4443,8646,17
Wine	87.0476,5483,6388,03	84.3573,2282,8385,08	84.574,2581,786	80.5870,7278,7282,86	84.4274,1881,6485,94	82.1572,1479,9184,12	84.6874,4181,8486,14
Yeast	33.3930,0552,3846,71	28.2725,4445,4942,62	33.2129,8844,2346,56	31.4328,2942,8945,14	31.4328,2942,8945,14	31.9928,7943,3145,59	32.2629,0343,5245,81
Iris	86.0877,4784,4288,86	86.4977,8481,7389,19	85.4876,9383,9688,38	80.0772,0679,8584,06	85.3576,8283,8788,28	82.4174,1781,6385,93	84.776,2383,3787,76
SAHeart	85.9677,3684,3388,77	85.0376,5383,6288,02	84.2975,5683,6087,43	78.9371,0478,9983,14	84.3175,8883,0887,45	81.2573,1380,7585	84.1975,7782,9887,35
Ionosphere	12110,828,1229,6	10.419,3729,1228,33	10.619,5927,0628,49	9.978,9726,5827,98	11.7510,5827,9329,4	10.49,3626,928,32	10.739,6627,1528,58
Galaxy	29.4626,5141,3943,57	27.8725,0840,1842,3	28.6425,7840,7742,91	28.6125,7540,7442,89	27.8325,0540,1542,26	28.1825,3640,4242,54	27.8925,1440,242,31
Halfing	58.5852,7263,5266,86	64.1457,7367,7571,31	34.1830,7644,9847,34	30.5527,542,2244,44	34.1230,7144,9347,3	32.9929,6944,0746,39	33.7730,3944,6747,02
Artificial	89.2880,3586,8591,42	85.0276,5283,6288,02	89.0780,1686,6991,26	83.2974,9682,386,63	86.7278,0584,9189,38	81.9173,7281,2585,53	84.3975,9583,1487,51
Artificial2	34.6731,2153,4947,74	33.7130,3444,6246,97	18.7816,933,2735,02	25.0122,5138,0140,01	22.7920,5136,3238,23	22.7420,4736,2838,19	24.2321,8137,4139,38
Artificial3	95.1785,6591,3396,14	94.2684,8389,6495,41	71.7464,5773,5277,39	69.3662,4271,71175,49	73.5466,1974,8978,83	73.1765,85174,6178,54	74.1366,72175,3479,3
Landsat-satellite	62.6356,3766,670,1	61.5355,3865,7669,22	52.0346,8358,5461,62	58.6352,7763,5666,9	59.6553,6964,3367,72	58.7652,8863,6667,01	60.2854,2564,8168,22
ISOLET	74.5167,7467,7671,33	73.9866,5875,2279,18	72.2865,0573,9377,82	74.8367,3575,8779,86	71.4964,3473,3377,19	70.1363,12172,376,1	72.3465,1173,9877,87
USPS	64.1657,7467,7671,33	59.9553,9664,5667,96	59.0153,1163,8567,21	62.4256,1866,4469,94	61.0254,9265,3868,82	58.6152,7463,5466,88	59.4753,5264,267,58
Letter-Recognition	45.0940,5859,2756,07	42.9838,6851,6654,38	44.6140,1512,9855,69	41.9237,7350,8653,54	43.5239,1752,0854,82	43.8139,4352,3155,05	44.2139,7952,6655,37
Average	56.6654,4566,9468,4	55.4653,2664,4567,34	51.3549,6560,9364,14	50.1948,4659,9263,08	51.8650,0461,2664,48	50.4648,7560,1763,34	51.8150,0161,2364,46

	CSEAC	ONCE-AL	DICLEN	WCE	SRS	CO-AL	HBPFG
Breast-cancer	70.2877,5784,588,95	69.8777,2884,2688,7	69.8677,2784,2588,69	68.2175,798387,37	70.7878,184,9589,42	69.3176,7883,8488,25	66.2173,9981,4885,77
Bupa	1.3945,35157,360,31	1.7546,5858,3361,4	1.2246,1157,9360,98	0.8444,8656,8859,87	1.1946,97158,6661,75	1.5845,97157,8260,86	1.0944,1854,3159,27
Glass	31.7128,5343,0945,36	33.1429,8344,1946,51	32.4829,2343,6845,98	30.6827,6142,3244,54	31.0127,9142,5744,81	32.1328,9243,4245,7	32.1628,9448,4445,73
Wine	81.3573,2280,8385,08	85.1774,8582,2186,54	84.3874,1481,6185,9	82.4972,4480,1784,39	88.0177,4184,3788,81	87.1476,6383,7188,11	81.2371,3179,2183,38
Yeast	31.0727,9642,6144,86	31.8528,6743,2145,48	31.7728,5943,1545,42	31.5228,3742,9645,22	32.0528,8543,3645,64	31.6828,5143,0845,34	32.3529,1249,5945,88
Iris	84.4776,0283,287,58	85.2376,7183,7788,18	84.7976,3183,4487,83	82.8874,5981,9986,3	87.0778,3685,1789,66	85.0876,5783,6688,06	81.1873,0680,7184,94
SAHeart	84.1175,782,9287,29	84.3475,9183,1187,47	84.2275,883,0187,38	81.673,4481,0285,28	85.8877,2984,2788,7	84.9476,4583,5587,95	80.3572,3280,0784,28
Ionosphere	9.438,4926,1727,54	11.5110,3512,7429,2	10.489,4326,9628,38	9.338,426,0927,46	11.6110,4412,8229,28	10.589,5212,0428,46	11.1310,02137,46128,9
Galaxy	26.2523,6338,9541	28.6625,7940,7842,93	27.5924,2339,9742,07	27.2324,5139,6941,78	26.0723,4638,8140,86	26.9224,2339,4641,54	28.5825,7244,7242,86
Halfing	62.5356,2866,5270,02	33.6530,2944,5746,92	33.2829,9544,2946,62	31.7428,5743,1245,39	35.2231,7145,7748,18	33.4930,1444,4546,79	31.5528,4442,9845,24
Artificial	83.337582,3386,66	85.567784,0388,45	84.4676,0183,1987,57	82.9674,6682,0586,37	88.5979,7386,3390,87	85.5977,0384,0588,47	82.5874,3281,7686,06
Artificial2	37.8234,0447,7450,26	23.0720,7636,5338,46	23.7221,3537,0338,98	22.2119,9935,8837,77	26.4923,8439,1341,19	24.6922,2237,7639,75	25.9223,3338,7440,74
Artificial3	93.2283,989,8594,58	74.2666,8375,4479,41	73.8866,4975,1579,1	70.2663,2372,476,21	73.4666,1174,8378,77	73.7566,3875,0579	72.4965,2474,0977,99
Landsat-satellite	59.1753,2563,9767,34	60.9154,8265,2968,73	60.0954,0864,6768,07	59.0653,1563,8967,25	62.0755,8666,1769,66	60.9654,8665,3368,77	58.8652,9763,7367,09
ISOLET	71.3264,1973,277,06	73.3466,0174,7478,67	72.2665,0373,9277,81	69.8962,972,1275,91	73.8966,5175,1679,11	72.0864,8773,7877,66	69.6262,6671,9175,7
USPS	55.1149,660,8864,09	61.2655,1365,5669,01	58.5352,6863,4866,82	56.6650,9962,0665,33	62.2656,0366,3269,81	62.8856,5966,7970,3	57.0851,3762,3865,66
Letter-Recognition	42.4538,2151,2653,96	44.3439,9152,7455,47	43.6739,3152,1954,94	42.0237,8250,9453,62	44.2739,8452,6555,42	43.3939,0551,9854,71	43.0338,7354,7154,42
Average	54.4152,4163,2566,58	52.2350,4461,5664,8	51.5749,861,0564,27	49.9848,3159,862,94	52.9451,0862,1465,41	52.1350,2861,4664,69	50.3248,5761,5463,17

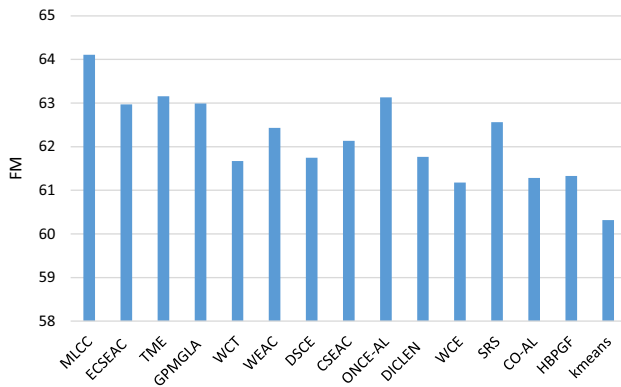


Fig. 14 Performances of different methods on Hamshahri text data set

5.6 News clustering application

A widely used labelled Persian text benchmark was introduced as Hamshahri data set in AleAhmad et al. (2009) which includes 9 classes and 36 subclasses.

We have sampled a subset of Hamshahri data set containing sport, economic, rural, adventure, and foreign classes each of which has 200 texts (1000 texts totally). Efficacy of the k-means clusterer algorithm after obtaining the data set like Shahriari et al. (2015) without using a thesaurus is 60.32% in terms of the f-measure, while efficacy of the MLCC, CSEAC and ECSEAC algorithms are, respectively, 64.11%, 62.13% and 62.97%. Since CSEAC, ECSEAC, and MLCC are the best methods according to Table 2, this experiment is carried out using these methods and the results have been summarized in Fig. 14.

6 Conclusions and future work

This work proposes a new consensus function in clustering ensemble named multi-level consensus clustering (MLCC). Since we use a multi-level clustering instead of direct object clustering, MLCC is a really efficient clusterer algorithm. An innovative similarity metric for measuring similarity between clusters is proposed. After obtaining an ensemble, a cluster–cluster similarity matrix using the mentioned metric is generated. The mentioned cluster–cluster similarity matrix is used to make an object–object or point–point similarity matrix. Then, we apply an average hierarchical clusterer algorithm on the point–point similarity matrix to make consensus partition. While MLCC is better than traditional clustering ensembles and simple version of clustering ensembles on traditional cluster–cluster similarity matrix, but its computational cost is not very bad. Accuracy and robustness of the proposed method are compared with the art clustering algorithms through the experimental tests. Also, time analysis is presented in the

experimental results. Comparison of the proposed method with the state-of-the-art clustering ensemble methods in terms of 4 different metrics on numerous real-world and synthetic data sets reveals that the proposed method is the superior method which did not need relabeling mechanism.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abbasi S, Nejatian S et al (2019) Clustering ensemble selection considering quality and diversity. *Artif Intell Rev* 52:1311–1340
- Akbari E, Mohamed Dahlan H, Ibrahim R, Alizadeh H (2015) Hierarchical cluster ensemble selection. *Eng Appl AI* 39:146–156
- AleAhmad A, Amiri H, Darrudi E, Rahgozar M, Oroumchian F (2009) Hamshahri: a standard Persian text collection. *J Knowl-Based Syst* 22(5):382–387
- Alishvandi H, Gouraki GH, Parvin H (2016) An enhanced dynamic detection of possible invariants based on best permutation of test cases. *Comput Syst Sci Eng* 31(1):53–61
- Alizadeh H, Minaei-Bidgoli B, Parvin H (2011a) A new criterion for clusters validation. In: *Artificial intelligence applications and innovations (AIAI 2011)*, IFIP, Springer, Heidelberg, Part I, pp 240–246
- Alizadeh H, Minaei-Bidgoli B, Parvin H, Moshki M (2011b) An asymmetric criterion for cluster validation, developing concepts in applied intelligence. *Stud Comput Intell* 363:1–14
- Alizadeh H, Minaei-Bidgoli B, Parvin H (2013) Optimizing fuzzy cluster ensemble in string representation. *Int J Pattern Recognit Artif Intell* 27(2):1350005
- Alizadeh H, Minaei-Bidgoli B, Parvin H (2014a) To improve the quality of cluster ensembles by selecting a subset of base clusters. *J Exp Theor Artif Intell* 26(1):127–150
- Alizadeh H, Minaei-Bidgoli B, Parvin H (2014b) Cluster ensemble selection based on a new cluster stability measure. *Intell Data Anal* 18(3):389–408
- Alizadeh H, Yousefnezhad M, Minaei-Bidgoli B (2015) Wisdom of crowds cluster ensemble. *Intell Data Anal* 19(3):485–503
- Alqurashi T, Wang W (2014) Object-neighborhood clustering ensemble method. In *Intelligent data engineering and automated learning (IDEAL)*, Springer, pp 142–149
- 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
- Ayad HG, Kamel MS (2008) Cumulative Voting Consensus Method for Partitions with a Variable Number of Clusters. *IEEE Trans Pattern Anal Mach Intell* 30(1):160–173
- Bagherinia A, Minaei-Bidgoli B, Hossinzadeh M, Parvin H (2019) Elite fuzzy clustering ensemble based on clustering diversity and quality measures. *Appl Intell* 49(5):1724–1747

- 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
- Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140
- Dimitriadou E, Weingessel A, Hornik K (2002) A combination scheme for fuzzy clustering. *Int J Pattern Recognit Artif Intell* 16(07):901–912
- Domeniconi C, Al-Razgan M (2009) Weighted cluster ensembles: methods and analysis. *ACM Trans Knowl Disc Data (TKDD)* 2(4):1–42
- Dueck D (2009) Affinity propagation: clustering data by passing messages. Ph.D. dissertation, University of Toronto
- Faceli K, Marcilio CP, Souto D (2006) Multi-objective clustering ensemble. In: *Proceedings of the sixth international conference on hybrid intelligent systems*
- X. Z. Fern and C. E. Brodley, “Random projection for high dimensional data clustering: A cluster ensemble approach”, In: *Proceedings of the 20th International Conference on Machine Learning*, (2003), pp. 186–193.
- 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
- Franek L, Jiang X (2014) Ensemble clustering by means of clustering embedding in vector spaces. *Pattern Recogn* 47(2):833–842
- Fred A, Jain AK (2002) Data clustering using evidence accumulation. In: *Intl. conf. on pattern recognition, ICPR02, Quebec City*, pp 276–280
- Fred A, Jain AK (2005) Combining multiple clustering’s using evidence accumulation. *IEEE Trans Pattern Anal Mach Intell* 27(6):835–850
- Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. *Comput Learn Theory* 55:119–139
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232
- Ghaemi R, ben Sulaiman N, Ibrahim H, Mustapha N (2011) A review: accuracy optimization in clustering ensembles using genetic algorithms. *Artif Intell Rev* 35(4):287–318
- Ghosh J, Acharya A (2011) Cluster ensembles. *Data Min Knowl Disc* 1(4):305–315
- Gionis A, Mannila H, Tsaparas P (2007) Clustering aggregation. *ACM Trans Knowl Discov Data* 1(1):4
- Hanczar B, Nadif M (2012) Ensemble methods for biclustering tasks. *Pattern Recognit* 45(11):3938–3949
- Ho TK (1995) Random decision forests. In: *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1, pp 278–282. <https://doi.org/10.1109/ICDAR.1995.598994>
- Hong Y, Kwong S, Chang Y, Ren Q (2008) Unsupervised feature selection using clustering ensembles and population based incremental learning algorithm. *Pattern Recogn* 41(9):2742–2756
- Hosseinpoor MJ, Parvin H, Nejatian S, Rezaie V (2019) Gene regulatory elements extraction in breast cancer by Hi-C data using a meta-heuristic method. *Russ J Genet* 55(9):1152–1164
- Huang D, Lai JH, Wang CD (2015) Combining multiple clusterings via crowd agreement estimation and multi-granularity link analysis. *Neurocomputing* 170:240–250
- Huang D, Lai J, Wang CD (2016) Ensemble clustering using factor graph. *Pattern Recogn* 50:131–142
- Huang D, Lai J, Wang CD (2016b) Robust ensemble clustering using probability trajectories. *The IEEE Trans Knowl Data Eng*
- Huang D, Wang CD, Lai JH (2017) Locally weighted ensemble clustering. *IEEE Trans Cybern* 99:1–14. <https://doi.org/10.1109/TCYB.2017.2702343>
- N. Iam-On, T. Boongoen and S. M. Garrett, “Refining Pairwise Similarity Matrix for Cluster Ensemble Problem with Cluster Relations”, *Discovery Science*, (2008), 222–233.
- 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
- Iam-On N, Boongoen T, Garrett S, Price C (2011) A link based approach to the cluster ensemble problem. *IEEE Trans Pattern Anal Mach Intell* 33(12):2396–2409
- 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
- Jamalinia H, Khalouei S, Rezaie V, Nejatian S, Bagheri-Fard K, Parvin H (2018) Diverse classifier ensemble creation based on heuristic dataset modification. *J Appl Stat* 45(7):1209–1226
- Jenghara MM, Ebrahimpour-Komleh H, Parvin H (2018a) Dynamic protein–protein interaction networks construction using firefly algorithm. *Pattern Anal Appl* 21(4):1067–1081
- Jenghara MM, Ebrahimpour-Komleh H, Rezaie V, Nejatian S, Parvin H, Syed-Yusof SK (2018b) Imputing missing value through ensemble concept based on statistical measures. *Knowl Inf Syst* 56(1):123–139
- Jiang Y, Chung FL, Wang S, Deng Z, Wang J, Qian P (2015) Collaborative fuzzy clustering from multiple weighted views. *IEEE Trans Cybern* 45(4):688–701
- Mimaroglu S, Aksehirli E (2012) DICLENS: Divisive clustering ensemble with automatic cluster number. *IEEE/ACM Trans Comput Biol Bioinf* 9(2):408–420
- Minaei-Bidgoli B, Topchy A, Punch WF (2004) Ensembles of partitions via data resampling. In: *Intl. conf. on information technology, ITCC 04, Las Vegas*, pp 188–192
- Minaei-Bidgoli B, Parvin H, Alinejad-Rokny H, Alizadeh H, Punch WE (2014) Effects of resampling method and adaptation on clustering ensemble efficacy. *Artif Intell Rev* 41(1):27–48
- Mirzaei A, Rahmati M (2010) A Novel hierarchical-clustering-combination scheme based on fuzzy-similarity relations. *IEEE Trans Fuzzy Syst* 18(1):27–39
- Mojarad M, Parvin H, Nejatian S, Rezaie V (2019a) Consensus function based on clusters clustering and iterative fusion of base clusters. *Int J Uncertain Fuzziness Knowl-Based Syst* 27(1):97–120
- Mojarad M, Nejatian S, Parvin H, Mohammadpoor M (2019b) A fuzzy clustering ensemble based on cluster clustering and iterative Fusion of base clusters. *Appl Intell* 49(7):2567–2581
- Moradi M, Nejatian S, Parvin H, Rezaie V (2018) CMCABC: Clustering and memory-based chaotic artificial bee colony dynamic optimization algorithm. *Int J Inf Technol Decis Mak* 17(04):1007–1046
- Naldi MC, De Carvalho ACM, Campello RJ (2013) Cluster ensemble selection based on relative validity indexes. *Data Min Knowl Disc* 27(2):259–289
- Nazari A, Dehghan A, Nejatian S, Rezaie V, Parvin H (2019) A comprehensive study of clustering ensemble weighting based on cluster quality and diversity. *Pattern Anal Appl* 22(1):133–145
- Nejatian S, Parvin H, Faraji E (2018) Using sub-sampling and ensemble clustering techniques to improve performance of imbalanced classification. *Neurocomputing* 276:55–66
- Nejatian S, Rezaie V, Parvin H, Pirbonyeh M, Bagherifard K, Yusof SKS (2019) An innovative linear unsupervised space adjustment by keeping low-level spatial data structure. *Knowl Inf Syst* 59(2):437–464
- Newman CBDJ, Hettich SS, Merz C (1998) UCI repository of machine learning databases. <http://www.ics.uci.edu/mlearn/MLSummary.html>.

- Omidvar MN, Nejatian S, Parvin H, Rezaie V (2018) A new natural-inspired continuous optimization approach, *Journal of Intelligent & Fuzzy Systems*, 1–17,
- Partabian J, Rafe V, Parvin H, Nejatian S (2020) An approach based on knowledge exploration for state space management in checking reachability of complex software systems. *Soft Comput* 24(10):7181–7196
- H. Parvin, B. Minaei-Bidgoli, A clustering ensemble framework based on elite selection of weighted clusters, *Advances in Data Analysis and Classification* (2013) 1–28.
- 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
- Parvin H, Beigi A, Mozayani N (2012) A clustering ensemble learning method based on the ant colony clustering algorithm. *Int J Appl Comput Math* 11(2):286–302
- Parvin H, Minaei-Bidgoli B, Alinejad-Rokny H, Punch WF (2013) Data weighing mechanisms for clustering ensembles. *Comput Electr Eng* 39(5):1433–1450
- Parvin H, Nejatian S, Mohamadpour M (2018) Explicit memory based ABC with a clustering strategy for updating and retrieval of memory in dynamic environments. *Appl Intell* 48(11):4317–4337
- Pirbonyeh A, Rezaie V, Parvin H, Nejatian S, Mehrabi M (2019) A linear unsupervised transfer learning by preservation of cluster-and-neighborhood data organization. *Pattern Anal Appl* 22(3):1149–1160
- Rafiee G, Dlay SS, Woo WL (2013) Region-of-interest extraction in low depth of field images using ensemble clustering and difference of Gaussian approaches. *Pattern Recognit* 46(10):2685–2699
- Rashidi F, Nejatian S, Parvin H, Rezaie V (2019) Diversity based cluster weighting in cluster ensemble: an information theory approach. *Artif Intell Rev* 52(2):1341–1368
- 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
- Roth V, Lange T, Braun M, Buhmann J (2002) A resampling approach to cluster validation. *Intl. conf. on computational statistics, COMPSTAT*
- Shabaniyan T, Parsaei H, Aminsharifi A, Movahedi MM, Jahromi AT, Pouyesh S, Parvin H (2019) An artificial intelligence-based clinical decision support system for large kidney stone treatment. *Australas Phys Eng Sci Med* 42(3):771–779
- Shahriari A, Parvin H, Monajati A (2015) Exploring weights of hierarchical and equivalency relationship in general Persian texts. *EANN Workshops* 7(1):7
- Soto V, Garcia-Moratilla S, Martinez-Munoz G, Hernandez-Lobato D, Suarez A (2014) A double pruning scheme for boosting ensembles. *IEEE Trans Cybern* 44(12):2682–2695
- Strehl A, Ghosh J (2000) Value-based customer grouping from large retail data sets. In *AeroSense, International Society for Optics and Photonics*, pp 33–42
- Strehl A, Ghosh J (2003) Cluster ensembles—a knowledge reuse framework for multiple partitions. *J Mach Learn Res* 3:583–617
- Szetoa PM, Parvin H, Mahmoudi MR, Tuan BA, Pho KH (2020) Deep neural network as deep feature learner. *J Intell Fuzzy Syst.* <https://doi.org/10.3233/JIFS-191292>
- Topchy AP, Jain AK, Punch WF (2003) Combining multiple weak clusterings. In: *IEEE international conference on data mining*, pp 331–338
- Topchy A, Jain AK, Punch W (2005) A mixture model of clustering ensembles. *Proc SIAM Int Conf Data Min, Citeseer* 27(12):1866–1881
- N. X. Vinh and M. E. Houle, “A set correlation model for partitional clustering”, In: *Advances in Knowledge Discovery and Data Mining*, Springer, (2010) pp. 4–15.
- Yang Y, Jiang J (2016) Hybrid sampling-based clustering ensemble with global and local constitutions. *IEEE Trans Neural Netw Learn Syst* 27(5):952–965
- Yasrebi M, Eskandar-Baghban A, Parvin H, Mohammadpour M (2018) Optimisation inspiring from behaviour of raining in nature: droplet optimisation algorithm. *Int J Bio-Inspired Comput* 12(3):152–163
- 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, pp 1176–1181
- Yousefnezhad M, Huang SJ, Zhang D (2018) WoCE: a framework for clustering ensemble by exploiting the wisdom of crowds theory. *IEEE Trans Cybernetics* 48(2):486–499
- Yu Z, Wong HS, You J, Yang Q, Liao H (2011) Knowledge based cluster ensemble for cancer discovery from biomolecular data. *IEEE Trans Nanobiosci* 10(2):76–85
- Yu Z, You J, Wong HS, Han G (2012) From cluster ensemble to structure ensemble. *Inf Sci* 198:81–99
- Yu Z, Chen H, You J, Han G, Li L (2013) Hybrid fuzzy cluster ensemble framework for tumor clustering from biomolecular data. *IEEE/ACM Trans Comput Biol Bioinf* 10(3):657–670
- Yu Z, Li L, Liu J, Han G (2015) Hybrid Adaptive Classifier Ensemble. *IEEE Transactions on Cybernetics* 45(2):177–190
- Yu Z, Zhu X, Wong HS, You J, Zhang J, Han G (2016a) Distribution-based cluster structure selection. *IEEE Trans Cybern* 99:1–14. <https://doi.org/10.1109/TCYB.2016.2569529>
- Yu Z, Chen H, Liu J, You J, Leung H, Han G (2016b) Hybrid k-nearest neighbor classifier. *IEEE Trans Cybern* 46(6):1263–1275
- Yu Z, Lu Y, Zhang J, You J, Wong HS, Wang Y, Han G (2017) Progressive semisupervised learning of multiple classifiers. *IEEE Trans Cybern* 99:1–14
- Zhang S, Wong HS, Shen Y (2012) Generalized adjusted rand indices for cluster ensembles. *Pattern Recognit* 45(6):2214–2226
- Zhao X, Liang J, Dang C (2017) Clustering ensemble selection for categorical data based on internal validity indices. *Pattern Recognit* 69:150–168
- 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

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