Artificial Intelligence Review
https://doi.org/10.1007/s10462-020-09862-1

Consensus function based on cluster-wise two level clustering

Mohammad Reza Mahmoudi[1,2] **· Hamidreza Akbarzadeh**[3] **· Hamid Parvin**[4,5] **· Samad Nejatian**[6,7] **· Vahideh Rezaie**[7,8] **· Hamid Alinejad-Rokny**[9,10,11]

Abstract
The ensemble clustering tries to aggregate a number of basic clusterings with the aim of producing a more consistent, robust and well-performing consensus clustering result. The current paper wants to introduce an ensemble clustering method. The proposed method, called consensus function based on two level clustering (CFTLC), introduces a new consensus clustering where it makes a cluster clustering task through applying an average hierarchical clustering on a cluster–cluster similarity matrix obtained by an innovative similarity metric. By applying the average hierarchical clustering algorithm, a set of meta clusters has been attained. Considering each meta cluster as a consensus cluster in the consensus clustering output, it then assigns each data point to a meta cluster through defining an object-cluster similarity. Before doing anything, CFTLC converts the primary partitions into a binary cluster representation where the primary ensemble has been broken into a number of basic binary clusters (BC). CFTLC first combines the basic BCs with the maximum cluster–cluster similarity. This step is iterated as long as a predefined number of meta clusters are ready. At the subsequent step, it assigns each data point to exactly one meta cluster. The proposed method has been experimentally compared with the state of the art clustering algorithms in terms of accuracy and robustness.

Keywords Consensus clustering · K-means · Similarity criterion · Machine learning · Data mining

1 Introduction

In numerous applications, machine learning functions are extremely beneficial (Pattanasri 2012; Yang and Yu 2017; Li et al. 2017; Deng et al. 2018; Chakraborty et al. 2017). In order to resolve numerous real world issues, straightforward machine learning models are utilized. Moreover, understandable machine learning models are frail in facing

✉ Hamid Parvin
parvin@iust.ac.ir

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

Extended author information available on the last page of the article

challenging issues (Song et al. 2017; Alsaaideh et al. 2017). Thus, regarding machine learning classification tasks, ensemble models have emerged as a new option (Derakhshani 2011; Wu 2011; Wagner 2011).

Partitioning is a prominent and complicated issue with numerous real world applications (Wang et al. 2017; Ma et al. 2018). The goal of partitioning is to locate data within identical groups (Duda et al. 2001). Partitioning has numerous applications such as knowledge extraction (Duda et al. 2001) and pattern recognition (Fred and Jain 2005). If previous information does not exist on clusters i.e. distributions, structures and natures, it will be challenging to define the most suitable clustering algorithm (Roth et al. 2002).

Several ensemble learning applications in classifier fields (Freund and Schapire 1995; Ho 1995; Friedman 2011; Soto et al. 2014; Yu et al. 2015, 2016a, b, 2017) guides researchers in turning their focus to cluster ensembles. Similar to all sub-fields within pattern recognition and classification has a tendency towards hybrid methods in the past few years (Faceli et al. 2006). Various cluster ensemble approaches are applied in general within the data mining (Hong et al. 2008; Zhang et al. 2012; Yu et al. 2012; Naldi et al. 2013; Franek and Jiang 2014; Jiang et al. 2015; Yu et al. 2016a, b; Huang et al. 2017; Yousefnezhad et al. 2018), multimedia (Rafiee et al. 2013) and bio-informatics fields (Yu et al. 2011, 2013; Hanczar and Nadif 2012). Cluster ensembles aim to clarify innovative partitions that are stronger in comparison to simple clustering algorithms that implement simple clustering algorithms (Ayad and Kamel 2008). The objective of clustering ensembles (Domeniconi and Al-Razgan 2009; Ghosh and Acharya 2011; Ghaemi et al. 2011) is to devise a consensus partitions by using a group of primary clusterings named ensembles. This consensus ensemble represents every primary clustering within the ensemble. The consensus partition enhances a specific objective function.

Two steps are carried out by the cluster ensemble. First, it creates a quantity of base clusterings. The clusterings are typically weak but not too weak. They are either moderate or weak clusterings (Topchy et al. 2003). In order to acquire a quantity of weak base clusterings, a basic clustering algorithm such as kmeans may be utilized with various parameters (Topchy et al. 2003), initializations, various data perspectives (Ayad and Kamel 2008), or various data samples (Minaei-Bidgoli et al. 2004). The mentioned methods have been implemented in the suggested method.

For the second cluster ensemble step, a partition named consensus partition is taken out of the produced ensemble from the first step in order to maintain maximum similarity. The mapping function that has its input as the ensemble and provides a consensus partition as its output is called the consensus function. Fred and Jain initially introduced EAC, as an ideal consensus function which primarily turns the ensemble into a CAM and then implements a single linkage hierarchical clustering algorithm for the purpose of finding the consensus partition (Fred and Jain 2002).

In order to assess a cluster, the Normalized Mutual Information Measure (Fred and Jain 2005) has numerous deficiencies which are explained in (Alizadeh et al. 2011a). An edition of this assessment tool named MAX is also presented which is not without its flaws. In order to enhance the MAX, another metric, named APMM, has been suggested (Alizadeh et al. 2011b). The named criteria have experienced some sort of complement cluster effect in one way or another. Thus, this paper suggests an innovative criterion for cluster assessment. This criterion predicts the average similarity between the cluster and other clusters by eliminating the impact of its complimentary cluster. All of the mentioned mechanisms will be dealt with along with their weaknesses. The advantages of the proposed method will be discussed in Sect. 4. This section will also include the previously stated improper complement cluster effect along with its impact on the consensus partition.

This paper presents an innovative cluster ensemble method. This method utilizes a fraction of base clusters. A modern strength criterion called Stab is used to assess the cluster quality. Every cluster that meets a Stab threshold criterion is considered to take part in determining consensus partition. In order to combine the selected clusters, a group of consensus function mechanisms are applied. One of the classes of the implemented consensus function is the co-association based consensus functions. Due to the inability of the EAC in deriving the co-association matrix within a subset of clusters, Extended Evidence Accumulation Clustering (E-EAC) is applied to create the co-association matrix from the selected clusters subset. Hyper graph partitioning algorithms are the basis of the consensus functions' second class. Furthermore, the remaining used consensus functions' class regards the selected clusters as a new feature space and implements a straightforward clustering algorithm in order to remove the consensus partition.

In spite of instance-based ensembles, cluster-based ensembles (Liang et al. 2018) are a set of approaches where they aggregate the ensemble at cluster-level. Our method, i.e. CFTLC, as a cluster-based ensemble, can be actually considered as an expanded version of dual-similarity clustering ensemble (*DSCE*) method (Alqurashi and Wang 2015). The CFTLC method improves the DSCE method in two shapes: (a) the CFTLC method is less sensitive to usage of the real number of clusters in primary clusterings, and (b) the CFTLC method is parameter less. Indeed, proposing a new cluster–cluster similarity is not our contribution. Our contribution is to measure similarity of clusters in a new feature space where its features are similarities of those two clusters to all other clusters along with similarities of those two clusters to all other clusters with an intermediate cluster.

The paper organization is as follows. Next section defines the problem. The paper presents the related work in section three. After that, it explains the proposed method. Experimental results will be presented in section five. Finally, the paper is concluded in section seven.

2 Clustering ensemble

2.1 Clustering ensemble definition

For a set of *d* data objects represented by $\mathrm{D} = \left\{\mathrm{D}_1, \mathrm{D}_2, \mathrm{D}_3, \dots, \mathrm{D}_d\right\}$, a π_p determined by $\pi_p = \left\{\pi_p^1, \pi_p^2, \pi_p^3, \dots, \pi_p^{c_p}\right\}$ represents a clustering result with c_p clusters acquired by a clustering algorithm as *p*th clustering. It is assumed that the basic clustering algorithms are unshared and complete, thus $\pi_p^i \bigcap \pi_p^j = \varphi$ and $\bigcup_{j=1}^{c_p} \pi_p^j = \mathrm{D}$.

Recognizing that members may not unquestionably possess the quantity of clusters within their partitions; i.e. $c_p \neq c_r \neq c$ where c is the favorable number of clusters. A clustering ensemble is responsible for extracting a consensus partition π_* out of the dataset D by the combination of ensemble partitions $\left\{\pi_1, \pi_2, \pi_3, \dots, \pi_B\right\}$ where *B* is the ensemble size, and π_* may potentially be more favorable than ensemble partitions $\left\{\pi_1, \pi_2, \pi_3, \dots, \pi_B\right\}$ in regard to quality and robustness.

To express the purpose of ensemble clustering in formal manner, it can be said that "it is to integrate the information of the primary partitions in the ensemble in order to find a better consensus clustering result".

Two tasks have previously been examined by researchers in the clustering ensemble domain: (a) creation of an ensemble and (b) consensus function construction. Both highly influence the final clustering outcomes.

In regard to classifier ensemble, the variety of results from the base learners are a crucial factor, thus the diversity may be extended into the clustering ensemble field. By utilizing diversity within a clustering ensemble, clustering ensembles of more favorable performance can be developed. Both tasks will be discussed further.

2.2 Ensemble creation

The first step is to create ensemble members of clustering ensemble framework. The aim of this step is to create B fundamental clustering results $\left\{\pi_1, \pi_2, \pi_3, \dots, \pi_B\right\}$. Additionally, the produced members should vary since greater diversity results in the uncovering of various data substructures. This also enhances the potential performance of consensus clustering, thus it is vital to utilize at least one creation method to access logical quality.

Totally these methods are categorized into three approaches: (a) homogeneous approach, where a single clustering algorithm is run multiple times with different initializations, (b) subspace clustering approach where a single clustering algorithm is run multiple times on different dataset projections, and (c) multiple clustering algorithms are run on a same dataset (Zhao et al. 2018).

Researchers have previously applied problem dependent creation methods. Strehl and Ghosh (2000) applied numerous clustering algorithms with several data subspaces. For larger dimensions of data, they implemented a few features subsets and created each member on each one. An object sub-sampling was also applied. Fern and Brodley (2003) also implemented a similar approach and obtained data partitions by projecting data into numerous subspaces.

Breiman (1996) introduced the resampling method in machine learning which was then further developed by Minaei-Bidgoli et al. (2004). Parvin et al. (2013b) implemented a method to generate ensemble via crisp clustering. By using Weighted Locally Adaptive Clustering (WLAC) algorithm, a diverse and precise ensemble can be produced. Later on, they extended their work to fuzzy clusters (Parvin and Minaei-Bidgoli 2015). Similar studies by Parvin et al. (2013a, 2013b) also inspired by boosting and bagging algorithms in classification include the examination of non-weighing and weighing-based sampling methods for ensemble generation (Parvin et al. 2013b).

Alizadeh et al. (2014a, b) presented an ensemble clustering framework that utilizes a subset of primary members within the ensemble as an alternative to prior methods. Thus, the Normalized Mutual Information (NMI) quality metric is used to determine these target clusterings. Essentially, a suitable NMI threshold depends on data and necessitates domain knowledge. Similar studies have been conducted by Nazari et al. (2017).

Kmeans is potentially the most versatile clustering algorithm for the production of ensemble elements because it is straightforward and fast (Minaei-Bidgoli et al. 2004). For example, Alizadeh et al. (2015) utilized kmeans clustering algorithm with random cluster center start or random c (number of clusters) selection within a predetermined scope for each base clustering. Strehl and Ghosh (2003) took advantage of a graph clustering algorithm by using various distance functions for every member. Topchy et al. (2005) utilized a weak clustering algorithm where its clustering outcomes were marginally superior to random predictors in accuracy terms.

Iam-On et al. (2010) experimented with various mechanisms that include implementing a set of kmeans with a constant quantity of clusters for every base partition and random cluster number selection from a set $\left\{2, 3, \ldots, \sqrt{\mathrm{d}}\right\}$. Additionally, Iam-on et al. (2012) utilized numerous creation methods to create ensembles.

Other researchers implemented several basic clusterings algorithms to create ensemble elements. They applied kmeans and hierarchical clustering algorithms as their main clustering algorithms (Gionis et al. 2007; Yi et al. 2012).

Alizadeh et al. (2015) carried out ensemble creation on the basis of the Wisdom of Crowds concept. It is concept in the social sciences field that provides criteria suited to group behavior. Thus, when these criteria are met, group decisions can be more favorable than individual members. Hence, the Wisdom of Crowds Cluster Ensemble (WOCCE) is presented with the competency to assess circumstances required for ensembles to exhibit collective wisdom. Decentralization conditions for base clusterings generation (Zhao et al. 2018), diversity conditions within ensembles and independence conditions among base algorithms are also included. Yousefnezhad et al. (2018) later developed this concept.

Hierarchical Cluster Ensemble is first introduced by Mirzaei et al. (Mirzaei and Rahmati 2010; Mirzaei et al. 2008). Due to the uncertain relationship between quality and diversity, Akbari et al. (2015) suggested the Hierarchical Cluster Ensemble Selection (HCES) method along with the diversity measure to define the impact of quality and diversity on the final outcomes. Specifically, they utilize complete linkage, average linkage and single linkage agglomerative methods to hierarchically choose members.

2.3 Aggregator

An aggregator uses ensembles $\left\{\pi_1, \pi_2, \pi_3, \ldots, \pi_B\right\}$ i.e. base clustering elements and the quantity of required clusters i.e. c as inputs and generates the final consensus clustering result π_*. Most researchers consider the aggregator stage as the most vital element of an ensemble.

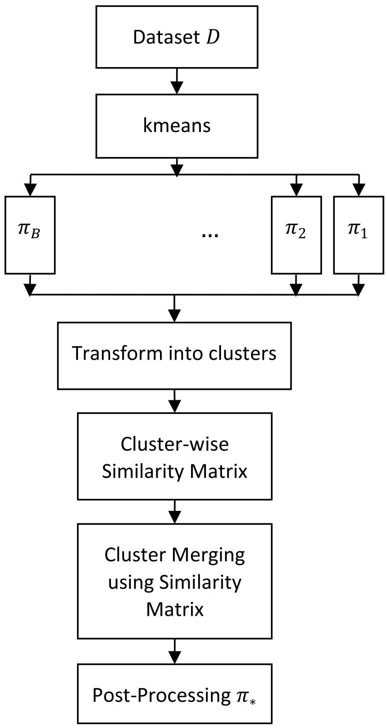
Several aggregators have been assessed by Alqurashi and Wang (2014). Two classifications of clustering ensembles include: (a) traditional category (like co-association (CA)) and (b) median clustering category. Totally the traditional consensus functions can be categorized into five approaches: (a) graph-based approach, (b) feature-based approach, (c) similarity-based approach (d) cluster-wise approach, and finally (e) relabeling and voting-based approach (Zhao et al. 2018).

For the first category, the methods aim to select a consensus partition which is extracted from a co-association matrix. Straightforward approaches such as the conventional voting method exhibit a feature based approach such as the Bayesian consensus clustering method, graph based approach like hyper-graph partitioning algorithm, and pair-wise similarity based approach like CA based method.

A set of new cluster ensemble approaches consider the issue of primary partition combination as an optimization procedure. Thus, the data among ensemble members is formulated as 0–1 bit strings. Based on this terminology, Alizadeh et al. (2013) presented a constrained nonlinear objective function named fuzzy string objective function (FSOF) to select a median partition. This is done by minimizing the disagreement among ensemble members while maximizing their agreement.

The framework used in the paper is depicted in Fig. 1. Based on the suggested clustering ensemble framework, the initial dataset is provided to the k-means clustering algorithm

Fig. 1 The proposed clustering ensemble framework



B times and an ensemble of *B* size is acquired. The ensemble is then turned into a cluster representation. A similarity matrix that determines the similarities among clusters is then created. Ultimately, the consensus partition is determined using a merging mechanism and post-processing.

3 Related work

CA-based clustering ensemble approaches encapsulate ensemble information into a CA matrix (Fred and Jain 2002). After that, they apply a hierarchical clustering algorithm on the CA matrix and extract the consensus clustering result. One of the advantages of the CA-based clustering ensemble approaches is the fact that they do not need the relabeling phase. For example, Fred and Jain (2005) attain the consensus clustering result through applying an average linkage hierarchical clustering algorithm. We name their method as *CO-AL* and consider it as a slow and CA-based one. Iam-On et al. (2008) introduce a new CA matrix considering the between-cluster relations. Their method *SRS* can be considered as a CA-based, slow, link based similarity one. They (Iam-On et al. 2011) extended it later in *WCT* method.

A similar work has been done by Mimaroglu and Aksehirli (2012) where they introduce a cluster–cluster similarity matrix instead of object–object similarity matrix (or CA matrix). Their method, named *DICLENS*, can be considered as a CA-based, slow, and robust one. Alqurashi and Wang (2014) have introduced an object-neighborhood-based similarity matrix. They (Alqurashi and Wang 2014) have introduced a new cluster–cluster similarity matrix where it has been obtained using neighborhood and real relation between objects. Their method is named $ONCE - AL$. Their extended method (Alqurashi and Wang 2015) is named $DSCE$. Both $ONCE - AL$ and $DSCE$ methods can be considered as heuristic, fast, and parameter sensitive ones. Huang et al. (2016b) have proposed an extended version of the $DSCE$ method. Huang et al. (2015) have introduced two cluster weighting, CA-based, slow, robust clustering ensemble methods named *WEAC* and *GPMGLA*.

While a large number of cluster–cluster similarity or dissimilarity measures have been proposed (such as single link or complete link in the agglomerative clustering algorithms, distance between centers of clusters, Jaccard measure), a few of them are robust. uses the number of the common objects included by the two clusters to reflect their similarity. Later, a weighted similarity measure (Liang et al. 2018) has been proposed to solve some of the previous measures. Here, we propose a new ensemble framework that uses a new similarity measure (while, without loss of generality, it can still use the previous measures), in order to improve the captured similarity measures.

Alizadeh et al. (2014b) have proposed a CA-based, very slow, and robust clustering ensemble, named $CSEAC$. Alizadeh et al. (2015) introduced a Wisdom of Crowds Ensemble (*WCE*) which is a CA-based, slow, robust, and flexible ensemble method. A cluster selection-based ensemble clustering algorithm, named *ECSEAC*, proposed by Parvin and Minaei-Bidgoli (2015), as a cluster weighting, CA-based, slow, and robust method is another related work. *TME* (Zhong et al. 2015) is another ensemble method which is a cluster weighting, CA-based, very slow, and robust method

Huang et al. (2016a) have transformed clustering ensemble as a linear binary programming problem. Zhao et al. (2017) deal clustering ensemble approaches for categorical dataset.

Inspired by bagging and boosting theory, Yang and Jiang (2016) have introduced a cluster ensemble through sampling. Bai et al. (2017) have introduced a cluster ensemble through information theory. The information theory has been shown to be a suitable tool for data clustering.

Although graph-based clustering ensemble approaches, firstly introduced by Strehl and Ghosh (2003), use a CA matrix like CA-based clustering ensemble approaches, they extract consensus clustering result through partitioning a hyper-graph. Indeed, they look at CA matrix as a hyper-graph, i.e. they consider each row and each column as a node and a hyper-edge respectively. The hyper-graph partitioning algorithm used in their work could have been HMETIS (Dimitriadou et al. 2002). This method is named Cluster based Similarity Partitioning Algorithm (CSPA). Hyper-Graph Partitioning Algorithm (HGPA) and Meta CLustering Algorithm (MCLA) can be considered as some other algorithms from these approaches. These methods consider each cluster as a hyper-edge and each data as a node. Hybrid Bipartite Graph Formulation (HBGF) algorithm (Fern and Brodley 2004) is an extension to CSPA. *HBGF* is better than HGPA, MCLA, and CSPA (Fern and Brodley 2004).

4 Proposed consensus function

An ensemble aggregator is a significant element in a clustering ensemble framework. It should be designed in such a way that benefits from as much ensemble information as possible. The aim of the proposed consensus function is to find a better consensus clustering result.

We first transform the ensemble $\{\pi_1, \pi_2, \ldots, \pi_B\}$ into a matrix with d rows and r columns. Each row stands for a data point and each column stands for a cluster; i.e. r is equal to $\sum_{i=1}^{B} c_i$. The mentioned matrix is denoted by $\tau^{\{\pi_1, \pi_2, \ldots, \pi_B\}}$.

Now, we present the proposed method. In the first step, the proposed algorithm transforms the ensemble into a binary cluster representation. Indeed, each cluster is a column and each data point is a row in the new binary cluster representation of the ensemble. The transformed binary cluster representation of the ensemble $\{\pi_1, \pi_2, \ldots, \pi_B\}$ is denoted by $\tau^{\{\pi_1, \pi_2, \ldots, \pi_B\}}$ where it is a matrix of size $d \times r$ where r is the number of clusters and computed based on Eq. 1.

$$r = \sum_{i=1}^{B} c_i \tag{1}$$

The matrix $\tau_{i_1 i_2}^{\{\pi_1, \pi_2, \ldots, \pi_B\}}$ stands for the (i_1, i_2)-th entry of $\tau^{\{\pi_1, \pi_2, \ldots, \pi_B\}}$ and it is defined based on Eq. 2.

$$\tau_{i_1 i_2}^{\{\pi_1, \pi_2, \ldots, \pi_B\}} = \begin{cases} 1 & D_{i_1} \in \pi_{\rho_2}^{\rho_1} \\ 0 & D_{i_1} \notin \pi_{\rho_2}^{\rho_1} \end{cases} \tag{2}$$

where ρ_1 and ρ_2 are two positive integer numbers which should hold Eq. 3.

$$\sum_{i=1}^{\rho_2 - 1} c_i + \rho_1 = i_2 \tag{3}$$

Toy Example 1 Let's consider an assumptive dataset with 12 instances with three clusters. Also, assume kmeans clustering algorithm has been run on the dataset 5 times and partitions the dataset into respectively 2, 2, 3, 3 and 3 clusters; therefore, we have an ensemble of size 5 on 12 instances. Let's assume that the assumptive ensemble is as presented in Fig. 2. Therefore, according to Eq. 2, the mentioned ensemble has been transformed into the binary representation as given in Fig. 3.□

Fig. 2 A clustering ensemble which contains four base partitions

	π_1	π_2	π_3	π_4	π_5
$\boldsymbol{D_1}$	a	α	2	3	γ
$\boldsymbol{D_2}$	a	β	2	3	γ
$\boldsymbol{D_3}$	a	β	2	1	ϑ
$\boldsymbol{D_4}$	a	α	2	3	γ
$\boldsymbol{D_5}$	b	β	1	1	ϑ
$\boldsymbol{D_6}$	a	β	1	1	ϑ
$\boldsymbol{D_7}$	a	α	1	3	γ
$\boldsymbol{D_8}$	b	α	3	3	γ
$\boldsymbol{D_9}$	b	α	3	1	μ
$\boldsymbol{D_{10}}$	b	β	1	2	μ
$\boldsymbol{D_{11}}$	a	α	3	2	μ
$\boldsymbol{D_{12}}$	b	α	3	2	μ

	π_1^1	π_1^2	π_2^1	π_2^2	π_3^1	π_3^2	π_3^3	π_4^1	π_4^2	π_4^3	π_5^1	π_5^2	π_5^3
$\boldsymbol{D_1}$	0	1	0	1	0	0	1	0	0	1	0	0	1
$\boldsymbol{D_2}$	0	1	1	0	0	0	1	0	0	1	0	0	1
$\boldsymbol{D_3}$	0	1	1	0	0	0	1	1	0	0	0	1	0
$\boldsymbol{D_4}$	0	1	0	1	0	0	1	0	0	1	0	0	1
$\boldsymbol{D_5}$	1	0	1	0	1	0	0	1	0	0	0	1	0
$\boldsymbol{D_6}$	0	1	1	0	1	0	0	1	0	0	0	1	0
$\boldsymbol{D_7}$	0	1	0	1	1	0	0	0	0	1	0	0	1
$\boldsymbol{D_8}$	1	0	0	1	0	1	0	0	0	1	0	0	1
$\boldsymbol{D_9}$	1	0	0	1	0	1	0	1	0	0	1	0	0
$\boldsymbol{D_{10}}$	1	0	1	0	1	0	0	0	1	0	1	0	0
$\boldsymbol{D_{11}}$	0	1	0	1	0	1	0	0	1	0	1	0	0
$\boldsymbol{D_{12}}$	1	0	0	1	0	1	0	0	1	0	1	0	0

Fig. 3 The binary cluster representation of the ensemble presented in Fig. 2

After producing the binary representation of an ensemble, the proposed method contains two steps. In the first step, it merges the binary clusters into c meta clusters. In the second step, considering each meta cluster as a consensus cluster, the proposed method assigns each data point to exactly one meta clusters. These two steps are explained in detail as follows.

Step I:

The proposed method first makes a cluster–cluster similarity matrix as follows. It constructs a transition graph $\mathbb{G} = (\mathbb{N}, \mathbb{E})$ whose set of nodes, i.e. $\mathbb{N}$, contains the binary clusters and is defined according to Eq. 4.

$$\mathbb{N} = \left\{\mathbb{m}_1, \mathbb{m}_2, \ldots, \mathbb{m}_r\right\} = \left\{\tau_{:1}^{\{\pi_1,\pi_2,\ldots,\pi_B\}}, \tau_{:2}^{\{\pi_1,\pi_2,\ldots,\pi_B\}}, \ldots, \tau_{:r}^{\{\pi_1,\pi_2,\ldots,\pi_B\}}\right\} \tag{4}$$

where r is computed according to Eq. 1. Also, the edges of the mentioned transition graph, i.e. $\mathbb{E}$, are relative similarities of those binary clusters as defined in Eq. 5.

$$\mathbb{e}_{\mathbb{m}_i\mathbb{m}_j} = \begin{cases} \dfrac{CCS(\mathbb{m}_i,\mathbb{m}_j)}{\sum_{k\in\{1,2,\ldots,r\}\setminus\{i\}} CCS(\mathbb{m}_i,\mathbb{m}_k)} & i \neq j \\ 0 & i = j \end{cases} \tag{5}$$

where $CCS(\mathbb{m}_i, \mathbb{m}_j)$ is Cluster Correlation Similarity measure indicating the similarity between clusters $\tau_{:i}^{\{\pi_1,\pi_2,\ldots,\pi_B\}}$ and $\tau_{:j}^{\{\pi_1,\pi_2,\ldots,\pi_B\}}$ which is defined based on Eq. 6.

$$CCS(\mathbb{m}_i, \mathbb{m}_j) = \frac{1 + \dfrac{\tau_{:i}^{\{\pi_1,\ldots,\pi_B\}^t}\tau_{:j}^{\{\pi_1,\ldots,\pi_B\}} - \frac{\tau_{:i}^{\{\pi_1,\ldots,\pi_B\}^t}\tau_{:i}^{\{\pi_1,\ldots,\pi_B\}}\tau_{:j}^{\{\pi_1,\ldots,\pi_B\}^t}\tau_{:j}^{\{\pi_1,\ldots,\pi_B\}}}{d}}{\sqrt{\tau_{:i}^{\{\pi_1,\ldots,\pi_B\}^t}\tau_{:i}^{\{\pi_1,\ldots,\pi_B\}}\tau_{:j}^{\{\pi_1,\ldots,\pi_B\}^t}\tau_{:j}^{\{\pi_1,\ldots,\pi_B\}}\left(1-\frac{\tau_{:i}^{\{\pi_1,\ldots,\pi_B\}^t}\tau_{:i}^{\{\pi_1,\ldots,\pi_B\}}}{d}\right)\left(1-\frac{\tau_{:j}^{\{\pi_1,\ldots,\pi_B\}^t}\tau_{:j}^{\{\pi_1,\ldots,\pi_B\}}}{d}\right)}}}{2} \tag{6}$$

where t indicates transpose sign. The Cluster Correlation Similarity measure is a modified version of a criterion named the set correlation and introduced by Houle (2008). It is a real number in interval [0, 1], where 1 represents that two clusters are similar and the value 0 represents that two clusters are complementary (Vinh and Houle 2010). It is worthy to be mentioned that it is not our contribution. Indeed, proposing a new cluster–cluster similarity is not our contribution at all. Our contribution is to measure similarity of clusters in a new feature space where its features are similarities of those two clusters to all other clusters along with similarities of those two clusters to all other clusters with an intermediate cluster. By this way, we acclaim that every cluster–cluster similarity measure can be improved. To do so, considering $\mathbb{E}$ as a square $r \times r$ matrix, we define a $r \times (r\iota)$ matrix $\mathbb{F} = \left[\mathbb{E}'\mathbb{E}^2\mathbb{E}^3 \dots \mathbb{E}^\iota\right]$ where $\mathbb{E}^q$ stands for matrix $\mathbb{E}$ to the q power and $\mathbb{E}'$ is defined as Eq. 7.

$$E'_{ij} = \begin{cases} E_{ij} & i \neq j \\ 1 & i = j \end{cases} \tag{7}$$

A similarity matrix is now defined according to Eq. 8.

$$S_{ij} = \sqrt{\sum_{s=1}^{r\iota} \left|\mathbb{F}_{is} - \mathbb{F}_{js}\right|^2} \tag{8}$$

In this step, the similarity matrix obtained by Eq. 8 is finally given to a hierarchical average link clustering algorithm to partition the binary clusters into a given number of meta clusters.

Step II:

The proposed method assigns the data points to meta clusters. After that, considering each meta cluster as a consensus cluster, the proposed method produces a consensus partition. Therefore, the proposed method creates the consensus partition in the second step.

Toy Example 2: In ensemble depicted by Fig. 2 and its binary cluster representation depicted by Fig. 3, the edges of the transition graph $\mathbb{E}$ is depicted by Fig. 4. The distance (or dissimilarity) matrix S of the binary clusters presented in Fig. 3 is presented in Fig. 5 where $t = 2$. Then, the output meta clusters of the average linkage hierarchical clustering algorithm are as follows.

$$\left\{\left\{\pi_2^1, \pi_3^1, \pi_4^1, \pi_5^2\right\}, \left\{\pi_1^2, \pi_3^3, \pi_4^3, \pi_5^3\right\}, \left\{\pi_1^1, \pi_2^2, \pi_3^2, \pi_4^2, \pi_5^1\right\}\right\}$$

As you can observe, three meta clusters have been obtained as the number of meta clusters is predefined as a priori. These meta clusters are denoted by π_*^1, π_*^2 and π_*^3.□

The second step is to assign all data points to exactly one of the meta clusters. We first define an *O*bject *C*luster membership matrix as follows.

	π_1^1	π_1^2	π_2^1	π_2^2	π_3^1	π_3^2	π_3^3	π_4^1	π_4^2	π_4^3	π_5^1	π_5^2	π_5^3
π_1^1	0	0.01	0.09	0.09	0.1	0.13	0.04	0.10	0.11	0.06	0.13	0.08	0.06
π_1^2	0.01	0	0.09	0.09	0.08	0.05	0.14	0.08	0.07	0.12	0.05	0.10	0.12
π_2^1	0.09	0.09	0	0.01	0.13	0.04	0.10	0.13	0.08	0.06	0.07	0.14	0.06
π_2^2	0.09	0.09	0.01	0	0.05	0.14	0.08	0.05	0.10	0.12	0.11	0.03	0.12
π_3^1	0.10	0.08	0.13	0.05	0	0.05	0.05	0.11	0.09	0.07	0.08	0.12	0.07
π_3^2	0.13	0.05	0.04	0.14	0.05	0	0.05	0.08	0.12	0.07	0.14	0.05	0.07
π_3^3	0.04	0.14	0.10	0.08	0.05	0.05	0	0.08	0.06	0.13	0.05	0.09	0.13
π_4^1	0.10	0.08	0.13	0.05	0.11	0.08	0.08	0	0.05	0.04	0.08	0.16	0.04
π_4^2	0.11	0.07	0.08	0.10	0.09	0.12	0.05	0.05	0	0.05	0.16	0.06	0.05
π_4^3	0.06	0.12	0.06	0.12	0.07	0.07	0.13	0.04	0.05	0	0.04	0.05	0.17
π_5^1	0.13	0.05	0.07	0.11	0.08	0.14	0.05	0.08	0.16	0.04	0	0.05	0.04
π_5^2	0.08	0.10	0.15	0.04	0.12	0.05	0.09	0.16	0.06	0.05	0.05	0	0.05
π_5^3	0.06	0.12	0.06	0.12	0.07	0.07	0.13	0.04	0.05	0.17	0.04	0.05	0

Fig. 4 The edges of the transition graph $\mathbb{E}$ for the ensemble given by Fig. 2

	π_1^1	π_1^2	π_2^1	π_2^2	π_3^1	π_3^2	π_3^3	π_4^1	π_4^2	π_4^3	π_5^1	π_5^2	π_5^3
π_1^1	0	1.41	1.30	1.29	1.28	1.24	1.37	1.28	1.26	1.35	1.24	1.31	1.35
π_1^2	1.41	0	1.29	1.30	1.31	1.36	1.22	1.31	1.34	1.25	1.36	1.29	1.25
π_2^1	1.30	1.29	0	1.41	1.24	1.37	1.29	1.24	1.32	1.35	1.33	1.21	1.35
π_2^2	1.29	1.3	1.41	0	1.36	1.22	1.31	1.36	1.28	1.25	1.27	1.39	1.25
π_3^1	1.28	1.31	1.24	1.36	0	1.36	1.35	1.26	1.30	1.33	1.31	1.24	1.33
π_3^2	1.24	1.36	1.37	1.22	1.36	0	1.36	1.32	1.24	1.33	1.22	1.36	1.33
π_3^3	1.37	1.22	1.29	1.31	1.35	1.36	0	1.32	1.35	1.23	1.37	1.30	1.23
π_4^1	1.28	1.31	1.24	1.36	1.26	1.32	1.32	0	1.35	1.37	1.32	1.19	1.37
π_4^2	1.26	1.34	1.32	1.28	1.30	1.24	1.35	1.35	0	1.36	1.19	1.34	1.36
π_4^3	1.35	1.25	1.35	1.25	1.33	1.33	1.23	1.37	1.36	0	1.38	1.36	1.17
π_5^1	1.24	1.36	1.33	1.27	1.31	1.22	1.37	1.32	1.19	1.38	0	1.35	1.38
π_5^2	1.31	1.29	1.21	1.39	1.24	1.36	1.30	1.19	1.34	1.36	1.35	0	1.36
π_5^3	1.35	1.25	1.35	1.25	1.33	1.33	1.23	1.37	1.36	1.17	1.38	1.36	0

Fig. 5 The distance (or dissimilarity) matrix S of the binary clusters presented in Fig. 2 where $t = 2$

$$OC_{ij} = OCS\left(x_i, \pi_*^{j}\right) \tag{9}$$

where $OCS\left(x_i, \pi_*^{j}\right)$ is defined as an object cluster similarity metric based on Eq. 10.

$$OCS\left(x_i, \pi_*^{j}\right) = \frac{1}{\max\limits_{k \in \{1,2,\ldots,d\}} \left(M_{kj}\right)} M_{ij} \tag{10}$$

where M is a matrix with integer entries whose M_{ij} indicates the number of the primary clusters that (a) contain the data point x_i and (b) are member of the meta cluster π_*^{j}. It is defined based on Eq. (11).

$$M_{ij} = \sum_{z \in \pi_*^{j}} \theta\left(x_i \in z\right) \tag{11}$$

where $\theta(Term)$ is one if *Term* is true, otherwise it is zero.

Now, the data points are partitioned into two categories: (a) certain ones and (b) uncertain ones. A certain data point is the one whose margin between the maximum object cluster membership and the other object cluster memberships is equal to or greater than a threshold α (by default it is 0.5). For example, in Fig. 6, the data point D_1 is a certain one (because $1 - 0.20 \geq 0.5$) and the data point D_3 is a uncertain one (because $0.75 - 0.50 \geq 0.5$). A certain data point is assigned to the cluster with the maximum object cluster membership value.

Cluster certainty ratio is defined as difference between the average of values below 0.5 and the average of values above 0.5 in object cluster similarity matrix. For example, certainty ratio for cluster π_*^1 is 0.71.

An uncertain data point x_i is assigned to the cluster with the maximum $R\left(\pi_*^j\right) \times OCS_{ij}$ where $R\left(\pi_*^j\right)$ is certainty ratio of the jth meta cluster.

Toy Example 3: Consider previous Example again. The object-cluster membership matrix is computed in Fig. 6. As you may notice, in each row of the left or right matrixes of Fig. 6, at most only one “1” can occur (this can be proved and it is assigned

	π_*^1	π_*^2	π_*^3
D_1	$\frac{4}{4}$	$\frac{0}{4}$	$\frac{1}{5}$
D_2	$\frac{4}{4}$	$\frac{1}{4}$	$\frac{0}{5}$
D_3	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{0}{5}$
D_4	$\frac{4}{4}$	$\frac{0}{4}$	$\frac{1}{5}$
D_5	$\frac{0}{4}$	$\frac{4}{4}$	$\frac{1}{5}$
D_6	$\frac{1}{4}$	$\frac{4}{4}$	$\frac{0}{5}$
D_7	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{5}$
D_8	$\frac{2}{4}$	$\frac{0}{4}$	$\frac{3}{5}$
D_9	$\frac{0}{4}$	$\frac{1}{4}$	$\frac{4}{5}$
D_{10}	$\frac{0}{4}$	$\frac{2}{4}$	$\frac{3}{5}$
D_{11}	$\frac{1}{4}$	$\frac{0}{4}$	$\frac{4}{5}$
D_{12}	$\frac{0}{4}$	$\frac{0}{4}$	$\frac{5}{5}$

$\Rightarrow$

	π_*^1	π_*^2	π_*^3
D_1	1	0	.20
D_2	1	.25	0
D_3	.50	.75	0
D_4	1	0	.20
D_5	0	1	.20
D_6	.25	1	0
D_7	.75	.25	.20
D_8	.50	0	.60
D_9	0	.25	.80
D_{10}	0	.50	.60
D_{11}	.25	0	.80
D_{12}	0	0	1
Certainty Ratio	0.71	0.72	0.66

Fig. 6 The object cluster membership matrix for the ensemble presented in Fig. 2

	π_*^1	π_*^2	π_*^3
D_1	1	0	.20
D_2	1	.25	0
D_3	.50	.75	0
D_4	1	0	.20
D_5	0	1	.20
D_6	.25	1	0
D_7	.75	.25	.20
D_8	.50	0	.60
D_9	0	.25	.80
D_{10}	0	.50	.60
D_{11}	.25	0	.80
D_{12}	0	0	1

$\Rightarrow$

	π_*^1	π_*^2	π_*^3
D_1	1	0	0
D_2	1	0	0
D_3	.50	.75	0
D_4	1	0	0
D_5	0	1	0
D_6	0	1	0
D_7	1	0	0
D_8	.50	0	.60
D_9	0	0	1
D_{10}	0	.50	.60
D_{11}	0	0	1
D_{12}	0	0	1

$\Rightarrow$

	π_*^1	π_*^2	π_*^3
D_1	1	0	0
D_2	1	0	0
D_3	0	1	0
D_4	1	0	0
D_5	0	1	0
D_6	0	1	0
D_7	1	0	0
D_8	0	0	1
D_9	0	0	1
D_{10}	0	0	1
D_{11}	0	0	1
D_{12}	0	0	1

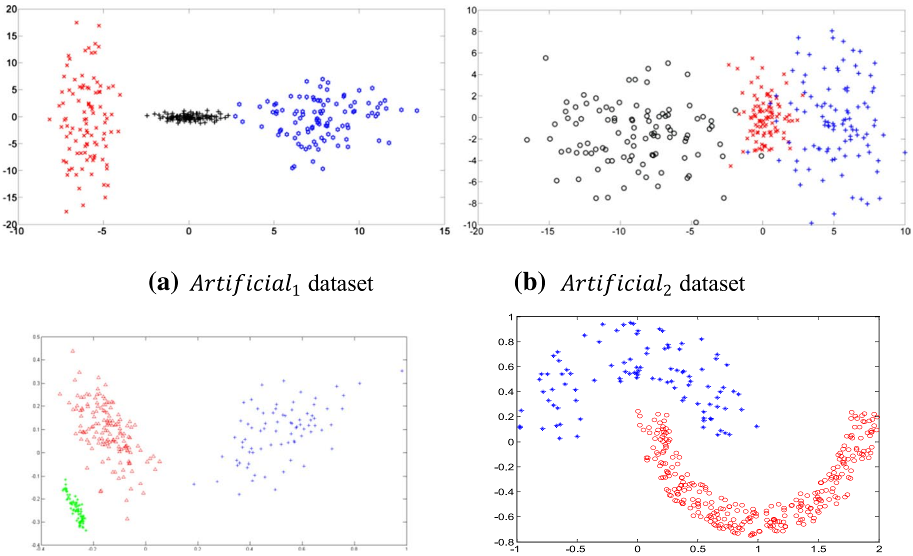
Fig. 7 The consensus clustering result for the given ensemble in Fig. 2

to the curious readers). After accomplishing the assignment task, the consensus partition will be as depicted in Fig. 7.

5 Experimental study

5.1 Datasets

To evaluate the different clustering ensemble approaches, we have used a number of real and artificial datasets. A set of 12 different real datasets from UCI machine learning repository (Newman et al. 1998) has been a subset of all benchmark datasets for assessment



(a) $Artificial_1$ dataset **(b)** $Artificial_2$ dataset

(c) $Artificial_3$ dataset **(d)** Halfring dataset

Fig. 8 The 4 artificial benchmark datasets

of the CFTLC algorithm. Another real dataset named USPS (Dueck 2009) is among the benchmark datasets for assessment of the CFTLC algorithm. Four artificial datasets, which are depicted in Fig. 8, have been used for assessment of the CFTLC algorithm. All of 17 datasets have been standardized. It means that any feature in all datasets have

Table 1 Description of the used datasets

ID	*Abbreviation*	*Datasize* : *Dimension* : *Clusters*
A_i	ith artificial dataset, $1 \le i \le 3$	300 : 2 : 3
B	*Breast*	683 : 9 : 2
G_2	*Galaxy*	323 : 4 : 7
G_1	*Glass*	214 : 9 : 6
H_1	*HalfRing*	400 : 2 : 2
H_2	*Heart*	462 : 9 : 2
I_1	*Ionosphere*	351 : 34 : 2
I_2	*Iris*	150 : 4 : 3
I_3	*ISOLET*	7797 : 617 : 26
L_1	*Landsat*	6435 : 36 : 6
L_2	*Letter*	20000 : 16 : 26
L_3	*Liver*	345 : 6 : 2
U	*USPS*	11000 : 256 : 10
W	*Wine*	178 : 13 : 2
Y	*Yeast*	1484 : 8 : 10

been transformed into a new ranges whose distribution is $N(0, 1)$. The features with missing value are managed by their removal. The description of datasets has been presented in Table 1.

5.2 Parameters

To produce the ensemble members, we use the method presented by Ren et al. (2013). To generate ensemble of size 50, the kmeans clustering algorithm has been applied on 50 different subsets of the dataset. Sampling ratio is 80%. In clustering of each bootstrap, the out-of-bag instances are assigned to the nearest clusters. To add a further 50 base clustering members to our ensemble on the given dataset, the kmeans clustering algorithm is applied on 50 different subspaces of the given dataset. Half of the dataset features is randomly chosen for each subspace. Parameter t is set to 2 by default.

The state of the art algorithms being used as benchmark in the paper, are as follows: *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. 2014b), *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 any baseline method, its initialization is as said in its corresponding paper. The ensemble size of all methods is hundred. Any reported result in the paper is averaged on 30 independent runs.

5.3 Evaluation metrics

To assess a clustering result on a given dataset, we can use the internal measures (like silhouette or SSE (sum of square errors)) which are computed regardless of the ground-truth labels of the data. Another set of the measures which can be used to assess a clustering result on a given dataset are external measures. The external measures are computed considering the ground-truth labels of the data. It is important to be mentioned that the ground-truth labels of the data are not used during the clustering task; it is only used after obtaining the output clustering results and for assessment. Both sets of the measures are popular. In this paper, we only use four external measures. Adjust Rand Index (ARI), Normalized Mutual Information (NMI), Accuracy (ACC) Rate and F-Measure (FM) are the used measures. Adjust rand index of a clustering result π_* given the ground-truth labels π_{gt} can be obtained by Eq. 12.

$$ARI\left(\pi_*, \pi_{gt}\right) = \frac{\sum_{i=1}^{c_*} \sum_{j=1}^{c} \binom{\left|\pi_*^i \cap \pi_{gt}^j\right|}{2} - \frac{\sum_{i=1}^{c_*} \binom{\left|\pi_*^i\right|}{2} \times \sum_{j=1}^{c} \binom{\left|\pi_{gt}^j\right|}{2}}{\binom{d}{2}}}{\frac{\sum_{i=1}^{c_*} \binom{\left|\pi_*^i\right|}{2} + \sum_{j=1}^{c} \binom{\left|\pi_{gt}^j\right|}{2}}{2} - \frac{\sum_{i=1}^{c_*} \binom{\left|\pi_*^i\right|}{2} \times \sum_{j=1}^{c} \binom{\left|\pi_{gt}^j\right|}{2}}{\binom{d}{2}}} \tag{12}$$

where c_* is the number clusters in the consensus clustering result π_*. Term $\binom{d}{2}$ means the binomial coefficient obtained as follows.

$$\binom{d}{2} = \frac{d!}{(d-2)!2!} = \frac{d \times (d-1)}{2} \tag{13}$$

Normalized mutual information for a clustering result π_* given the ground-truth labels π_{gt} can be obtained by Eq. 14.

$$NMI\left(\pi_*, \pi_{gt}\right) = \frac{\sum_{i=1}^{c_*} \sum_{j=1}^{c} \left|\pi_*^i \cap \pi_{gt}^j\right| \log_2\left(\frac{d \times \left|\pi_*^i \cap \pi_{gt}^j\right|}{\left|\pi_*^i\right| \times \left|\pi_{gt}^j\right|}\right)}{\sqrt{\sum_{i=1}^{c_*}\left(\left|\pi_*^i\right| \times \log_2\left(\frac{\left|\pi_*^i\right|}{d}\right)\right) \times \sum_{i=1}^{c}\left(\left|\pi_{gt}^i\right| \times \log_2\left(\frac{\left|\pi_{gt}^i\right|}{d}\right)\right)}} \tag{14}$$

Accuracy of a clustering result π_* given the ground-truth labels π_{gt} is denoted by $ACC\left(\pi_*, \pi_{gt}\right)$. It is computed based on Eq. 15.

$$ACC\left(\pi_*, \pi_{gt}\right) = \max_{\sigma} \sum_{i=1}^{c} \frac{\left|\pi_*^{\sigma_i} \cap \pi_{gt}^i\right|}{d} \tag{15}$$

where σ is a permutation of $\{1, 2, \ldots, c\}$ and σ_i is ith value in the permutation σ, i.e. $\forall i \in \{1, 2, \ldots, c\} : \sigma(i) \in \{1, 2, \ldots, c\}$ and $\forall i, j \in \{1, 2, \ldots, c\} : (i \neq j) \rightarrow (\sigma(i) \neq \sigma(j))$. The term σ is the relabeling task.

Finally, F-Measure of a clustering result π_* given the ground-truth labels π_{gt} is denoted by $FM\left(\pi_*, \pi_{gt}\right)$ and computed based on Eq. 16.

$$FM\left(\pi^*, \pi^t\right) = \max_{\sigma} \sum_{i=1}^{c} \frac{2 \times \left|\pi_*^{\sigma_i}\right| \times \frac{\left|\pi_*^{\sigma_i} \cap \pi_{gt}^i\right|}{\left|\pi_{gt}^i\right|} \times \frac{\left|\pi_*^{\sigma_i} \cap \pi_{gt}^i\right|}{\left|\pi_*^{\sigma_i}\right|}}{d \times \left(\frac{\left|\pi_*^{\sigma_i} \cap \pi_{gt}^i\right|}{\left|\pi_{gt}^i\right|} + \frac{\left|\pi_*^{\sigma_i} \cap \pi_{gt}^i\right|}{\left|\pi_*^{\sigma_i}\right|}\right)} \tag{16}$$

The range of all of the 4 mentioned measures is [0, 1]. The greater the values of these measures, the better. While the ARI and NMI are symmetric measures, the ACC and FM are asymmetric measures.

Table 2 The state of the art baseline methods comparing with CFTLC method on different datasets in terms of normalized mutual information

	CFTLC	*ECSEAC*	*TME*	*GPMGLA*	*WCT*	*WEAC*	*DSCE*	*CSEAC*	*ONCE-AL*	*DICLEN*	*WCE*	*SRS*	*CO-AL*	*HBPGF*
A_1	89.28	85.02	89.07	83.29	86.72	81.91	84.39	83.33	85.56	84.46	82.96	88.59	85.59	82.58
A_2	34.67	33.71	18.78	25.01	22.79	22.74	24.23	37.82	23.07	23.72	22.21	26.49	24.69	25.92
A_3	95.17	94.26	71.74	69.36	73.54	73.17	74.13	93.22	74.26	73.88	70.26	73.46	73.75	72.49
B	71.16	71.09	70.41	65.35	70.77	66.41	69.41	70.28	69.87	69.86	68.21	70.78	69.31	66.21
G_1	32.19	32.17	33.07	31.31	31.55	31.81	32.71	31.70	33.14	32.48	30.68	31.01	32.13	32.16
G_2	29.46	27.87	28.64	28.61	27.83	28.18	27.89	26.25	28.66	27.59	27.23	26.07	26.92	28.58
H_1	58.58	64.14	34.18	30.55	34.12	32.99	33.77	62.53	33.65	33.28	31.74	35.22	33.49	31.55
H_2	85.96	85.03	84.29	78.93	84.31	81.25	84.19	84.11	84.34	84.22	81.60	85.88	84.94	80.35
I_1	12.00	10.41	10.61	9.97	11.75	10.40	10.73	9.43	11.50	10.48	9.33	11.60	10.58	11.13
I_2	86.08	86.49	85.48	80.07	85.35	82.41	84.70	84.47	85.23	84.79	82.88	87.07	85.08	81.18
I_3	74.51	73.98	72.28	74.83	71.49	70.13	72.34	71.32	73.34	72.26	69.89	73.89	72.08	69.62
L_1	62.63	61.53	52.03	58.63	59.65	58.76	60.28	59.17	60.91	60.09	59.06	62.07	60.96	58.86
L_2	45.09	42.98	44.61	41.92	43.52	43.81	44.21	42.45	44.34	43.67	42.02	44.27	43.39	43.03
L_3	1.85	1.65	1.01	0.92	1.37	1.16	1.30	1.39	1.75	1.22	0.84	1.19	1.58	1.09
U	64.16	59.95	59.01	62.42	61.02	58.60	59.47	55.11	61.26	58.53	56.66	62.26	62.88	57.08
W	87.04	84.35	84.50	80.58	84.42	82.15	84.68	81.35	85.17	84.38	82.49	88.01	87.14	81.23
Y	33.39	28.27	33.20	31.43	31.43	31.99	32.26	31.07	31.85	31.77	31.52	32.05	31.68	32.35

Table 3 Summary results presented in Table 2

	HBPGF	*CO-AL*	*SRS*	*WCE*	*DICLENS*	*ONCE-AL*	*CSEAC*	*DSCE*	*WEAC*	*WCT*	*GPMGLA*	*TME*	*ECSEAC*	*CFTLC*
Mean	53.97	55.86	56.76	53.68	55.33	55.99	58.23	55.57	54.17	55.60	53.85	55.17	59.17	**60.59**
w-t-l	14-3-0	14-3-0	8-7-2	15-2-0	16-1-0	13-3-1	13-2-2	14-3-0	16-1-0	14-1-2	14-1-2	10-7-0	7-9-1	–

Any triple w-t-l contains three integer numbers, i.e. w, t, and l. The number w indicates the number of datasets on which the proposed method is the superior method. The number l indicates the number of datasets on which the proposed method is the loser method. The number t indicates the number of datasets on which the proposed method is neither the loser nor the winner method. All tests have been done by paired t-test

Table 4 The state of the art baseline methods comparing with CFTLC method on different datasets in terms of adjacent rand index

	CFTLC	*ECSEAC*	*TME*	*GPMGLA*	*WCT*	*WEAC*	*DSCE*	*CSEAC*	*ONCE-AL*	*DICLEN*	*WCE*	*SRS*	*CO-AL*	*HBPGF*
A_1	80.35	76.52	80.16	74.96	78.05	73.72	75.95	75.00	77.00	76.01	74.66	79.73	77.03	74.32
A_2	31.20	30.34	16.90	22.51	20.51	20.47	21.81	34.04	20.76	21.35	19.99	23.84	22.22	23.33
A_3	85.65	84.83	64.57	62.42	66.19	65.85	66.72	83.90	66.83	66.49	63.23	66.11	66.38	65.24
B	78.67	78.43	77.77	73.22	78.09	74.17	76.87	77.57	77.28	77.27	75.79	78.10	76.78	73.99
G_1	28.97	28.95	29.76	28.18	28.40	28.63	29.44	28.53	29.83	29.23	27.61	27.91	28.92	28.94
G_2	26.51	25.08	25.78	25.75	25.05	25.36	25.10	23.63	25.79	24.83	24.51	23.46	24.23	25.72
H_1	52.72	57.73	30.76	27.50	30.71	29.69	30.39	56.28	30.29	29.95	28.57	31.70	30.14	28.40
H_2	77.36	76.53	75.86	71.04	75.88	73.13	75.77	75.70	75.91	75.80	73.44	77.29	76.45	72.32
I_1	10.80	9.37	9.55	8.97	10.58	9.36	9.66	8.49	10.35	9.43	8.40	10.44	9.52	10.02
I_2	77.47	77.84	76.93	72.06	76.82	74.17	76.23	76.02	76.71	76.31	74.59	78.36	76.57	73.06
I_3	67.06	66.58	65.05	67.35	64.34	63.12	65.11	64.19	66.01	65.03	62.90	66.50	64.87	62.66
L_1	56.37	55.38	46.83	52.77	53.69	52.88	54.25	53.25	54.82	54.08	53.15	55.86	54.86	52.97
L_2	40.58	38.68	40.15	37.73	39.17	39.43	39.79	38.21	39.91	39.30	37.82	39.84	39.05	38.73
L_3	47.57	46.49	46.81	44.21	45.87	45.14	46.17	45.35	46.58	46.10	44.86	46.97	45.97	44.18
U	57.74	53.96	53.11	56.18	54.92	52.74	53.52	49.60	55.13	52.68	50.99	56.03	56.59	51.37
W	76.54	73.22	74.25	70.72	74.18	72.14	74.41	73.22	74.85	74.14	72.44	77.41	76.63	71.31
Y	30.05	25.44	29.88	28.29	28.29	28.79	29.03	27.96	28.67	28.59	28.37	28.85	28.51	29.12

Table 5 The state of the art baseline methods comparing with *CFTLC* method on different datasets in terms of accuracy

	CFTLC	*ECSEAC*	*TME*	*GPMGLA*	*WCT*	*WEAC*	*DSCE*	*CSEAC*	*ONCE-AL*	*DICLEN*	*WCE*	*SRS*	*CO-AL*	*HBPGF*
A_1	86.85	83.62	86.69	82.30	84.91	81.25	83.14	82.33	84.03	83.19	82.05	86.33	84.05	81.76
A_2	53.49	44.62	33.27	38.01	36.32	36.28	37.41	47.74	36.53	37.03	35.88	39.13	37.76	38.70
A_3	91.33	89.64	73.52	71.71	74.89	74.61	75.34	89.85	75.44	75.15	72.40	74.83	75.05	74.09
B	85.43	85.23	84.67	80.83	84.95	81.63	83.91	84.50	84.26	84.25	83.00	84.95	83.84	81.48
G_1	54.64	46.45	44.13	42.80	42.98	43.18	43.86	43.09	44.19	43.68	42.32	42.57	43.42	48.44
G_2	41.39	40.18	40.77	40.74	40.15	40.42	40.20	38.95	40.78	39.97	39.69	38.81	39.46	44.72
H_1	63.52	67.75	44.98	42.22	44.93	44.07	44.67	66.52	44.57	44.29	43.12	45.77	44.45	42.98
H_2	84.33	83.62	83.06	78.99	83.08	80.75	82.98	82.92	83.10	83.01	81.02	84.27	83.55	80.07
I_1	28.12	29.12	27.06	26.58	27.93	26.90	27.15	26.17	27.74	26.96	26.09	27.82	27.04	37.46
I_2	84.42	81.73	83.96	79.85	83.87	81.63	83.37	83.20	83.77	83.44	81.99	85.17	83.66	80.71
I_3	75.63	75.22	73.93	75.87	73.33	72.30	73.98	73.20	74.74	73.92	72.12	75.16	73.78	71.91
L_1	66.60	65.76	58.54	63.56	64.33	63.66	64.81	63.97	65.29	64.67	63.89	66.17	65.33	63.73
L_2	59.27	51.66	52.90	50.86	52.08	52.30	52.60	51.26	52.74	52.19	50.94	52.65	51.98	54.70
L_3	59.17	58.25	58.53	56.33	57.74	57.12	57.99	57.30	58.33	57.93	56.88	58.66	57.82	54.31
U	67.76	64.56	63.85	66.44	65.38	63.54	64.20	60.88	65.56	63.48	62.06	66.32	66.79	62.38
W	83.63	82.83	81.70	78.72	81.64	79.91	81.84	80.83	82.21	81.61	80.17	84.37	83.71	79.21
Y	52.38	45.49	44.23	42.89	42.89	43.31	43.52	42.61	43.21	43.15	42.96	43.36	43.08	49.59

Table 6 The state of the art baseline methods comparing with CFTLC method on different datasets in terms of f-measure

	CFTLC	*ECSEAC*	*TME*	*GPMGLA*	*WCT*	*WEAC*	*DSCE*	*CSEAC*	*ONCE-AL*	*DICLEN*	*WCE*	*SRS*	*CO-AL*	*HBPGF*
A_1	91.42	88.02	91.26	86.63	89.38	85.53	87.51	86.66	88.45	87.57	86.37	90.87	88.47	86.06
A_2	47.74	46.97	35.02	40.01	38.23	38.19	39.38	50.26	38.46	38.98	37.77	41.19	39.75	40.74
A_3	96.14	95.41	77.39	75.49	78.83	78.54	79.30	94.58	79.41	79.10	76.21	78.77	79.00	77.99
B	89.93	89.71	89.13	85.08	89.42	85.93	88.33	88.95	88.70	88.69	87.37	89.42	88.25	85.77
G_1	45.75	45.74	46.46	45.05	45.24	45.45	46.17	45.36	46.51	45.98	44.54	44.81	45.70	45.73
G_2	43.57	42.30	42.91	42.89	42.26	42.54	42.31	41.00	42.93	42.07	41.78	40.86	41.54	42.86
H_1	66.86	71.31	47.34	44.44	47.30	46.39	47.02	70.02	46.92	46.62	45.39	48.18	46.79	45.24
H_2	88.77	88.02	87.43	83.14	87.45	85.00	87.35	87.29	87.47	87.38	85.28	88.70	87.95	84.28
I_1	29.60	28.33	28.49	27.98	29.40	28.32	28.58	27.54	29.20	28.38	27.46	29.28	28.46	28.90
I_2	88.86	89.19	88.38	84.06	88.28	85.93	87.76	87.58	88.18	87.83	86.30	89.66	88.06	84.94
I_3	79.61	79.18	77.82	79.86	77.19	76.10	77.87	77.06	78.67	77.81	75.91	79.11	77.66	75.70
L_1	70.10	69.22	61.62	66.90	67.72	67.01	68.22	67.34	68.73	68.07	67.25	69.66	68.77	67.09
L_2	56.07	54.38	55.69	53.54	54.82	55.05	55.37	53.96	55.47	54.94	53.62	55.42	54.71	54.42
L_3	62.28	61.32	61.61	59.30	60.78	60.13	61.04	60.31	61.40	60.98	59.87	61.75	60.86	59.27
U	71.33	67.96	67.21	69.94	68.82	66.88	67.58	64.09	69.01	66.82	65.33	69.81	70.30	65.66
W	88.03	85.08	86.00	82.86	85.94	84.12	86.14	85.08	86.54	85.90	84.39	88.81	88.11	83.38
Y	46.71	42.62	46.56	45.14	45.14	45.59	45.81	44.86	45.48	45.42	45.22	45.64	45.34	45.88

5.4 Experiments

The state of the art baseline methods, mentioned in Sect. 5.2, have been compared by the results of Table 2 with *CFTLC* algorithm in terms of NMI on a number of benchmark datasets. The results of Table 3 summarize the results of Table 2. In the summary results presented in Table 2, any triple w-t-l contains three integer numbers, i.e. w, t, and l. The number w indicates the number of datasets on which the proposed method is the superior method. The number l indicates the number of datasets on which the proposed method is the loser method. The number t indicates the number of datasets on which the proposed method is neither the loser nor the winner method. All tests have been done by paired t test. According to Table 3, the proposed method is superior to all state of the art baseline methods.

The state of the art baseline methods comparing with *CFTLC* method on all 17 datasets in terms of adjacent rand index, accuracy and f-measure are respectively presented in Tables 4, 5 and 6.

The results presented in Table 4 (and also Tables 5, 6) indicates: (a) the *CFTLC* algorithm outperforms the state of the art baseline methods in most of the used benchmark datasets, (b) the *CFTLC* algorithm is outperformed by the state of the art baseline methods in three of the used benchmark datasets (while it is outperformed by the state of the art baseline methods in three datasets, it is still among the top three best methods), and (c) the *CFTLC* algorithm is always (for all of the used benchmark datasets) among the top three best methods.

It can be proved that the time complexity of the CFTLC method is $O\left(t \times r^3\right)$ or $O\left(tc^3B^3\right)$ in its worst case. In Table 7, a comparison of the consumed time (in terms of second) by different algorithms is presented on a dataset with 10,000 randomly generated instances and two clusters (each cluster has 5000 instance).

5.5 Parameter analysis

Noise effect has been examined in this subsection. For robustness analysis, we have selected the three best methods according to Table 3, i.e. RCEIFBC, ECSEAC and CSEAC. We have added Gaussian noise with different energy levels and then applied these three methods. To do this, assume that Σ_i be covariance of the ith cluster in the dataset. For a noise ratio ς, we define a temporary noise data with the same size of the data samples in the mentioned cluster of dataset and name it as $\aleph_i$. The data samples of $\aleph_i$ are produced in such a way that its covariance matrix is $\varsigma \times \Sigma_i$. Now the data samples of that temporary noise data are added to the data samples of the ith cluster. We repeat this procedure for all clusters in dataset (adding a temporary noise data with the covariance matrix $\varsigma \times \Sigma_i$) and obtain a new dataset that we considered as noisy dataset with noise ratio $\varsigma \times 100\%$. It is obvious when $\varsigma = 0$ there is no noise. The results for Breast-cancer and Wine datasets have been presented in Table 8. In Table 8, the vertical axis shows the NMI value and horizontal axis shows the amount of noise ratio, or ς.

5.6 News clustering application

Hamshahri dataset introduced by AleAhmad et al. (2009) is a widely used labelled Persian text benchmark. Hamshahri dataset has 9 classes (36 subclasses).

Table 7 Consumed time when 10,000 instances is used. "-" means more than 200 s

	HBPGF	*CO-AL*	*SRS*	*WCE*	*DICLENS*	*ONCE-AL*	*CSEAC*	*DSCE*	*WEAC*	*WCT*	*GPMGLA*	*TME*	*ECSEAC*	*CFTLC*
time	16		–	15	14	17	–	14	15	56	29	197	169	14

Table 8 The NMI results for Iris and Wine datasets in the presence of different levels of noise

	I_2			W		
	CSEAC	*ECSEAC*	*CFTLC*	*CSEAC*	*ECSEAC*	*CFTLC*
0	81.44	84.54	87.59	70.28	71.09	71.16
10	49.00	81.82	82.40	54.67	70.95	72.59
20	40.64	75.49	79.17	60.24	67.47	69.37
30	57.04	80.72	77.80	40.57	67.16	68.07
40	33.65	69.12	71.06	36.80	68.89	68.61
50	39.59	63.68	65.52	12.14	61.78	63.18
60	37.65	61.48	62.90	14.68	69.38	69.38
70	20.94	58.47	60.24	49.73	59.21	60.77
80	38.24	60.45	62.09	28.56	62.72	63.27
90	20.85	56.77	59.67	54.93	56.97	58.89
100	41.55	58.19	58.24	10.44	55.23	57.01

Table 9 The *CSEAC*, *ECSEAC*, and *CFTLC* algorithms' efficacy on Hamshahri text dataset

k-means	*CSEAC*	*ECSEAC*	*CFTLC*
60.32	62.13	62.97	**64.11**

A subset of Hamshahri dataset has been sampled. Our sampling contains 5 classes (sport, economic, rural, adventure, and foreign) and 1000 texts, i.e. 200 texts in each class. After obtaining the dataset like Shahriari et al. (2015) without using a thesaurus, the efficacy of the k-means clustering algorithm is 60.32% in terms of the f-measure. The efficacies of the *CFTLC*, *CSEAC* and *ECSEAC* algorithms are respectively 64.11%, 62.13% and 62.97% in terms of the f-measure. The *CSEAC*, *ECSEAC*, and *CFTLC* methods have been chosen for this experiment because they are the best methods according to Table 3. The results have been summarized in Table 9.

6 Conclusions and future work

A new consensus function in clustering ensemble has been introduced in this paper. It is named Consensus Function based on Two Level Clustering (CFTLC). It is scalable and fast as it uses cluster clustering instead of direct object clustering. After cluster clustering task, considering each cluster of clusters as a meta cluster, it defines an object clustering based on similarities of data points to the meta clusters. To do so, the paper defines two metrics: (a) cluster-wise similarity measure, and (b) cluster object similarity measure.

The proposed method has been compared with the state of the art clustering ensemble methods in terms of 4 different metrics on numerous real and artificial datasets. The experiments indicate the proposed method is superior to the state of the art clustering ensemble methods. The proposed method does not need relabeling mechanism. Besides, the proposed method is based on cluster clustering, not object clustering. Therefore, it is more scalable to the state of the art clustering ensemble methods.

Acknowledgements This study is supported by grants to HAR and HP. HAR is supported by UNSW Scientia Program Fellowship and is a member of the UNSW Graduate School of Biomedical Engineering.

Author contributions HP and HAR designed the study. HP and HAR developed hypothesis and experiments. MRM and HP wrote the manuscript; MRM, HP, and HA edited the manuscript with help from HAR; MRM, HP, HA, SN, and VR carried out the analyses, the implementation of the codes, and the statistical analyses; MRM, HP, and HA generated all figures and tables. All authors have read and approved the final version of the paper.

Compliance with ethical standards

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

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

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. IJPRAI 27(2):1350005

Alizadeh H, Minaei-Bidgoli B, Parvin H (2014a) Cluster ensemble selection based on a new cluster stability measure. Intell Data Anal 18(3):389–408

Alizadeh H, Minaei-Bidgoli B, Parvin H (2014b) 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, 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

Alsaaideh 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 Spat Inf Sci 20(1):29–38

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

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

Chakraborty D, Singh S, Dutta D (2017) Segmentation and classification of high spatial resolution images based on Hölder exponents and variance. Geo Spat Inf Sci 20(1):39–45

Deng Q, Wu S, Wen J, Xu Y (2018) Multi-level image representation for large-scale image-based instance retrieval. CAAI Trans Intell Technol 3(1):33–39

Derakhshani RR (2011) An ensemble method for classifying startle eyeblink modulation from high-speed video records. IEEE Trans Affect Comput 2(1):50–63

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 TKDD 2(4):1–42

Duda RO, Hart PE, Stork DG (2001) Pattern classification, 2nd edn. Wiley, Hoboken

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

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. http://www.aaai.org/Papers/ICML/2003/ICML03-027.pdf

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: International conference 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. Computational learning theory, pp 119–139

Friedman JH (2011) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

Ghaemi R, bin 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 TKDD 1(1):4

Hanczar B, Nadif M (2012) Ensemble methods for biclustering tasks. Pattern Recogn 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

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

Houle ME (2008) The relevant-set correlation model for data clustering. Stat Anal Data Min 1(3):157–176

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 (2016a) Ensemble clustering using factor graph. Pattern Recogn 50:131–142

Huang D, Lai J, Wang CD (2016b) Robust ensemble clustering using probability trajectories. IEEE Trans Knowl Data Eng 28(5):1312–1326

Huang D, Wang CD, Lai JH (2017) Locally weighted ensemble clustering. IEEE Trans Cybern 99(1):1–14. https://doi.org/10.1109/TCYB.2017.2702343

Iam-On N, Boongoen T, Garrett SM (2008) Refining pairwise similarity matrix for cluster ensemble problem with cluster relations. Discovery Science, pp 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, Boongeon 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

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

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

Liang JY, Shi QY, Zhao XW (2018) Multi-view data ensemble clustering: a cluster-level perspective. Int J Mach Intell Sens Sig Process 2(2):97–120

Ma J, Jiang X, Gong M (2018) Two-phase clustering algorithm with density exploring distance measure. CAAI Trans Intell Technol 3(1):59–64

Mimaroglu S, Aksehirli E (2012) DICLENS: divisive clustering ensemble with automatic cluster number. IEEE ACM TCBB 9(2):408–420

Minaei-Bidgoli B, Topchy A, Punch WF (2004) Ensembles of partitions via data resampling. In: International conference on information technology, ITCC 04, Las Vegas, pp 188–192

Mirzaei A, Rahmati M (2010) A novel hierarchical-clustering-combination scheme based on fuzzy-similarity relations. IEEE Trans Fuzzy Syst 18(1):27–39

Mirzaei A, Rahmati M, Ahmadi M (2008) A new method for hierarchical clustering combination. Intell Data Anal 12(6):549–571

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 (2017) A comprehensive study of clustering ensemble weighting based on cluster quality and diversity. Pattern Anal Appl. https://doi.org/10.1007/s10044-017-0676-x

Newman CBDJ, Hettich SS, Merz C (1998) UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLSummary.html

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, Alinejad-Rokny H, Minaei-Bidgoli B, Parvin S (2013a) A new classifier ensemble methodology based on subspace learning. J Exp Theor Artif Intell 25(2):227–250

Parvin H, Minaei-Bidgoli B, Alinejad-Rokny H, Punch WF (2013b) Data weighing mechanisms for clustering ensembles. Comput Electr Eng 39(5):1433–1450

Pattanasri N (2012) Learning to estimate slide comprehension in classrooms with support vector machines. IEEE Trans Learn Technol 5(1):52–61

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 Recogn 46(10):2685–2699

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. In: International conference on computational statistics, COMPSTAT

Shahriari A, Parvin H, Monajati A (2015) Exploring weights of hierarchical and equivalency relationship in general persian texts. EANN Workshops 7(1):7

Song XP, Huang C, Townshend JR (2017) Improving global land cover characterization through data fusion. Geo Spat Inf Sci 20(2):141–150

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

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. In: Proceedings of the SIAM international conference of data mining, Citeseer, vol 27, no 12, pp 1866–1881

Vinh NX, Houle ME (2010) A set correlation model for partitional clustering. In: Advances in knowledge discovery and data mining. Springer, pp 4–15

Wagner J (2011) Exploring fusion methods for multimodal emotion recognition with missing data. IEEE Trans Affect Comput 2(4):206–218

Wang B, Zhang J, Liu Y, Zou Y (2017) Density peaks clustering based integrate framework for multi-document summarization. CAAI Trans Intell Technol 2(1):26–30

Wu CH (2011) Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels. IEEE Trans Affect Comput 2(1):10–21

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

Yang H, Yu L (2017) Feature extraction of wood-hole defects using wavelet-based ultrasonic testing. J For Res 28(2):395–402

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 Cybern 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 Bioinform 10(3):657–670

Yu Z, Li L, Liu J, Han G (2015) Hybrid adaptive classifier ensemble. IEEE Trans Cybern 45(2):177–190

Yu Z, Chen H, Liu J, You J, Leung H, Han G (2016a) Hybrid k-nearest neighbor classifier. IEEE Trans Cybern 46(6):1263–1275

Yu Z, Zhu X, Wong HS, You J, Zhang J, Han G (2016b) Distribution- based cluster structure selection. IEEE Trans Cybern 99(1):1–14. https://doi.org/10.1109/TCYB.2016.2569529

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):1–14

Zhang S, Wong HS, Shen Y (2012) Generalized adjusted rand indices for cluster ensembles. Pattern Recogn 45(6):2214–2226

Zhao X, Liang J, Dang C (2017) Clustering ensemble selection for categorical data based on internal validity indices. Pattern Recogn 69:150–168

Zhao XW, Cao FY, Liang JY (2018) A sequential ensemble clusterings generation algorithm for mixed data. Appl Math Comput 335:264–277

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

Affiliations

Mohammad Reza Mahmoudi[1,2] · Hamidreza Akbarzadeh[3] · Hamid Parvin[4,5] · Samad Nejatian[6,7] · Vahideh Rezaie[7,8] · Hamid Alinejad-Rokny[9,10,11]

[1] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

[2] Department of Statistics, Faculty of Science, Fasa University, Fasa, Fars, Iran

[3] Department of Computer Engineering, 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

[6] Department of Electrical Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran

[7] Young Researchers and Elite Club, Yasooj Branch, Islamic Azad University, Yasooj, Iran

[8] Department of Mathematics, Yasooj Branch, Islamic Azad University, Yasooj, Iran

[9] Systems Biology and Health Data Analytics Lab, The Graduate School of Biomedical Engineering, UNSW Australia, Sydney 2052, Australia

[10] School of Computer Science and Engineering, The University of New South Wales (UNSW Sydney), Sydney 2052, Australia

[11] Health Data Analytics Program Leader, AI-enabled Processes (AIP) Research Centre, Macquarie University, Sydney 2109, Australia