THEORETICAL ADVANCES

A comprehensive study of clustering ensemble weighting based on cluster quality and diversity

Ahmad Nazari[1] · Ayob Dehghan[1] · Samad Nejatian[2,3] · Vahideh Rezaie[3,4] · Hamid Parvin[1,5]

Abstract

Clustering as a major task in data mining is responsible for discovering hidden patterns in unlabeled datasets. Finding the best clustering is also considered as one of the most challenging problems in data mining. Due to the problem complexity and the weaknesses of primary clustering algorithm, a large part of research has been directed toward ensemble clustering methods. Ensemble clustering aggregates a pool of base clusterings and produces an output clustering that is also named consensus clustering. The consensus clustering is usually better clustering than the output clusterings of the basic clustering algorithms. However, lack of quality in base clusterings makes their consensus clustering weak. In spite of some researches in selection of a subset of high quality base clusterings based on a clustering assessment metric, cluster-level selection has been always ignored. In this paper, a new clustering ensemble framework has been proposed based on cluster-level weighting. The certainty amount that the given ensemble has about a cluster is considered as the reliability of that cluster. The certainty amount that the given ensemble has about a cluster is computed by the accretion amount of that cluster by the ensemble. Then by selecting the best clusters and assigning a weight to each selected cluster based on its reliability, the final ensemble is created. After that, the paper proposes cluster-level weighting co-association matrix instead of traditional co-association matrix. Then, two consensus functions have been introduced and used for production of the consensus partition. The proposed framework completely overshadows the state-of-the-art clustering ensemble methods experimentally.

Keywords Data clustering · Clustering ensemble · Consensus function · Weighting

1 Introduction

Main idea of data clustering is the separation of data samples from each other and placing them into the similar groups; it means that data samples similar to each other should be placed into a same group and should have maximum dissimilarity with the samples of other group(s) [16, 29, 35].

✉ Ayob Dehghan
  ayubdehqan@gmail.com

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

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

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

4 Department of Mathematics, Yasooj Branch, Islamic Azad University, Yasooj, Iran

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

Each group in the above definition is usually referred to as a cluster. All clusters together are named a clustering. It is a very traditional task in any field for us to intend to group a host of data points. It can be due to lack of any insight into dataset. So, a first grouping in data that there is no primary insight can provide us with a good initial understanding and insight. Because of its utility, clustering has emerged as one of the leading methods of multivariate analysis. It has been shown that cluster analysis (CA) is among the most versatile concepts for analyzing datasets with many (more than three) features [68].

Indeed, data clustering is a necessary tool for finding groups (or clusters) in unlabeled datasets [45, 48, 50, 52, 53, 58]. One of the most important objective functions in data clustering is maximization of intra-cluster similarity simultaneous with minimization of inter-cluster similarity. In other words, data clustering objective functions usually intend to maximize inter-cluster variance and minimize intra-cluster variance. It is worthy to be mentioned that there is not any supervision during data clustering to guide the

clusterer. It is considered as a subfield in statistical pattern recognition (SPR), machine learning (ML), and data mining (DM) fields. Two main tasks in learning [46, 49, 72] are considered as data classification [6, 64] and data clustering [2, 17, 19, 56]. Data clustering is an interesting topic, and it has diverse applications [47, 50, 54, 55, 59]. There are many reasons why data clustering has many applications. Generally data clustering is used for two aims: (a) receiving a primary understanding of a huge amount of raw data and (b) reducing the size of a huge amount of raw data.

Based on Kleinberg's theorem, there is no base clustering method that is able to learn any dataset [37]. It means that each basic clustering algorithm is suitable for some datasets having a shared special feature. So we need a new approach that combines basic clustering algorithms in such a way that surpasses the basic clustering algorithms. Due to the lack of supervision in data clustering, choosing a special basic clustering algorithm for an unknown dataset is an impossible and incorrect topic. The lack of a unique clustering assessment metric is also an additive reason to make us be unable to select an absolutely superior clustering algorithm even out of a set of two basic clustering algorithms. Because choosing any objective function for data clustering, it is sometimes desired to find a data clustering algorithm that produces a robust output or even a stable output [7, 10, 27, 30, 61]. Reaching a robust clustering algorithm or a stable clustering algorithm is very hard with simple data clustering algorithms and is easier by ensemble-based data clustering algorithms.

Lack of a guideline for selecting the best data clustering algorithm for a given dataset can be considered as one of the most important delicacies in this field. Indeed, there is no automatic organized method for selection of the most suitable data clustering algorithm for an unknown given dataset.

One of the most important steps in pattern recognition is feature selection. In a more general manner, feature selection can be viewed as feature weighting. Both of them are frequently are considered during data classification, but they are usually hard to be considered during data clustering. It can be very challenging to consider any weighting mechanism in data clustering algorithms. In this paper, a weighting mechanism has been used in data clustering framework, but with the help of ensemble concept.

Data clustering with the help of an ensemble of clusterers is named clustering ensemble. The main aim of clustering ensemble is to search a better and more stable result with the aggregation of the information extracted from a set of base data clusterings. The better and more stable result that is extracted from the set of base data clusterings (also called clusterings pool) is named consensus partition or consensus clustering [22, 62, 63]. There are many researches that have been accomplished in classifier ensemble [1, 36, 57, 54, 55]. Inspired by

classifier ensemble success during past researches, clustering ensemble have emerged as a new concept in pattern recognition communities recently. It has been shown that clustering ensemble can be better in terms of both clustering quality and clustering stability. It is worthy to be mentioned that their parallelization ability is suitable for distributed data mining. Clustering ensemble can produce consensus partitions that are more robust, more novel, more stable, and more flexible than the clusterings produced by the base data clustering algorithms [8, 22, 25, 65].

In summary, clustering ensemble composes of the following two steps [22, 62]:

Ensemble generation step: in this step a set of base data clusterings should be produced. These base data clusterings should have two fundamental features to be considered as a suitable ensemble: quality and diversity. It is widely known as an effective approach to produce a large set of base data clusterings and then select a diverse subset of those base data clusterings with the highest qualities [3, 4, 58].

Aggregation of base data clusterings step: in this step all base data clusterings produced in the previous step should be aggregated. The aggregation mechanism is named ensemble consensus function (ECF).

Assigning weights to voter based on their qualities can be an effective mechanism to improve an ensemble. The process of weighting to basic learners in an ensemble of learners can be optimally set if accuracy of each base learner is known [38]. But if the learners are of type clusterers, the accuracy for learner is meaningless [29]. So, quality of each quality clusterer can be used as its approximation for its accuracy. Fred and Lourenco [25] have proposed a clustering ensemble based on selection of a subset of the base data clusterings with the highest qualities. They first assessed base clusterings then selected those with the most qualities. They finally employed an ECF named evidence accumulation clustering (EAC). They ignored the diversity of the base data clusterings (indeed they consider it implicitly).

Alizadeh et al. [4] have proposed a similar work in the cluster level. They have first produced an ensemble, and then they have transformed the ensemble into a cluster representation. After that, they have attached a quality value to each cluster based on its stability in a reference set. Finally, they have selected the clusters with the most qualities. They have improved their previous work by introducing a new metric [3]. Like the previous work of Fred and Lourenco [25], their proposed frameworks also ignored the diversity of the base data clusterings.

In general, applying weighting in clustering has been discussed recently and rarely. As the first and the most complete methods that used weighting mechanism, we can refer to a paper by Gullo et al. [29], and another paper by Domeniconi and Al-Razgan [14]. These methods are Weighted Similarity

Partitioning Algorithm (WSPA) [29] and Weighted Bipartite Partitioning Algorithm (WBPA) [14] where the WBPA is the better one.

In all previous researches, both diversity and clustering weighting have not considered simultaneously. They have dealt with either weighting primary data clusterings based on their qualities or selecting of a subset of the primary data clusterings based on their diversity. But diversity and quality have been rarely considered simultaneously. They are considered simultaneously in some works by Parvin and Minaei-Bidgoli [52] and Parvin and Minaei-Bidgoli [53], but they are considered only implicitly. Also in some works, Alizadeh et al. [3] and Alizadeh et al. [4] have considered both of them explicitly and simultaneously, but they have done it in the level of clustering. Also they only used binary weighting. It means that they first have selected a subset of clusterings with the best quality values and have named it the high quality clusterings. Then they have selected a subset of diverse clusterings among the high quality clusterings. Seeking to tackle the abovementioned issue, in this paper, an original data clustering ensemble method has been proposed. So, in the present paper we assign a weight to each cluster based on their quality value. Then, the most diverse subset of clusters is selected into the final ensemble. To reach the final consensus results, a new function termed cluster-level weighting evidence accumulation clustering (CLWEAC). Using the model of CLWEAC, consensus clustering is finally produced by an agglomerative hierarchical clustering algorithm. Figure 1 illustrates the diagram of the proposed method. We make use of the diversity concept at the cluster level and incorporate the cluster quality into a weighting structure for boosting the consensus clustering.
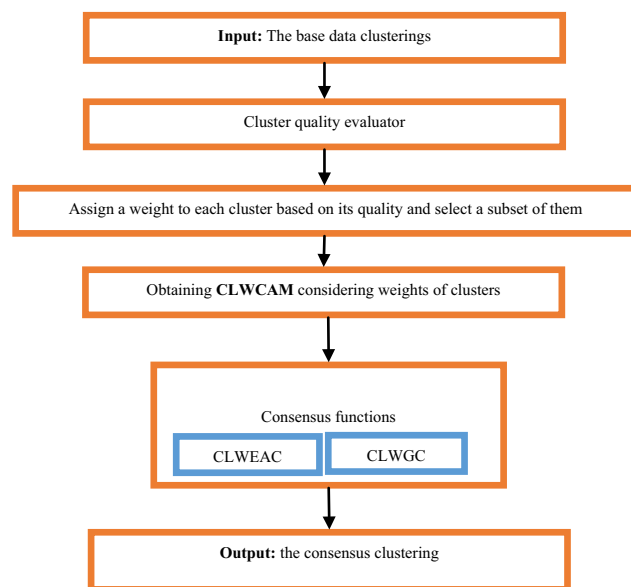

Fig. 1 Flow diagram of the proposed approach

To emphasize the main contributions of the paper, they are briefly listed here:

- A new metric has been proposed to assess quality of a cluster based on coherence in its data points. It is worthy to be mentioned that the metric don not need the original features of the given dataset.
- A weighting framework has been introduced to weight up the better clusters (clusters with higher qualities).
- A new novel consensus function that considers the cluster quality and diversity has been introduced.
- A large number of experimentations have been performed on many real standard benchmarks showing the proposed clustering ensemble performs better than the state-of-the-art clustering ensembles.

This manuscript is arranged as follows. Section 2 contains the related work. Section 3 explains the proposed clustering ensemble approach. Section 4 presents the experimental study. Finally, conclusion will be presented in Sect. 5.

2 Related work

Clustering is the problem of putting a set of data samples into similar categories (widely named as clusters) where similarity can be calculated based on Euclidean distance. In clustering, the algorithm aims at two objectives: (a) Groups have maximal intra-cluster similarities (or data samples in a same group have minimal Euclidean distances with each other or with cluster center) and (b) different groups have minimal inter-cluster similarities (it can be interpreted either as "each pair of data samples belonging to two different clusters has maximal Euclidean distances" or as "cluster centers of different clusters have maximal Euclidean distances with each other") [38]. So, clustering means the separating data into a number of groups in such a way that all pairs of the samples in a same group have the least possible dissimilarities from (or the most possible similarities with) each other, and all pairs of the samples assigned to different groups have the least possible similarities with each other.

In some application, we need a robust clustering (RC); for example if the clustering mechanism is employed in a preliminary stage of a classification algorithm (to find and omit outliers). Another reason to desire a RC is hidden in the strong relationship between the CA and Robust Statistics (RS) [60]. RC methods are discussed in detail by García-Escudero et al. [27]. Coretto and Hennig [10] have shown experimentally that the RIMLE method, their previously proposed method, can be recommended as optimal in some situations and acceptable in all situations.

Finding the best clustering for a given dataset is almost impossible [37]; the lack of a totally acceptable criterion to

assess the quality of a clustering is the main reason. It means that the measurement is subjective and so the best is also sensitive to that measurement. One of the most challenging tasks in CA is discovery of a mechanism how the stability of a clustering is obtained. Hennig [30] acclaimed that this problem is mainly due to the fact that robustness and stability in CA are not only data dependent, but even cluster dependent. Therefore, some researchers consider stability as the measurement to find which clustering algorithm is the best. It means that if clustering algorithm A is producing more stable results than clustering algorithm B, then we can assume that clustering algorithm B is better than clustering algorithm A. To find stable clustering algorithms, the clustering ensemble methods have emerged. They attempt to find a better and more stable clustering solution by fusing information from several primary data clusterings [3, 4].

The transfer distance (TD) between clusterings [9, 12] is defined as the minimum number of data points that should be moved from one cluster to another (possibly empty) to turn one clustering into another. This distance has been studied under the name of "partition-distance". Based on the concept of the TD between clusterings, a consensus clustering has been proposed by Guénoche [28]. Given a host of primary clusterings over a dataset, Guénoche [28] has proposed a consensus clustering containing a maximum number of joined or separated pairs in the given dataset that are also joined or separated in the primary clusterings. To do so, he has defined a utility function mapping a clustering to score value. He first transforms the problem to a graph clustering. Then, using some mathematical tricks converts it to an integer linear programming optimization. Finally, he seeks for a partition that maximizes the score. The score can be maximized by an integer linear programming only in certain cases. He has shown that in those certain cases that the method converges to a solution the method produces a consensus partition that is very close to the optimum. Although the method is a very simple and clear one, it fails in many real-world cases as the author mentions it.

Although some researchers have tried to find the best consensus clustering out of an ensemble of base data clusterings, it is widely believed that all ensembles do not perform well. It has been shown that an ensemble needs two elements to be considered as good ensemble. Diversity among the members of an ensemble is one of the most fundamental factors in the ensemble. It means that an ensemble will fail if the diversity among its members is not high. While the diversity is important, the qualities of primary clusterings are also important, so both of them are recently considered by the researchers. But rarely both of them are dealt with in an integrated approach. A new algorithm for choosing a subset of many primary clusters is proposed in this paper where it simultaneously considers both of them. A good selection is very decisive. The selection of some base clusterings (or some base clusters) can be done by a heuristic algorithm. To do this, a criterion to evaluate a cluster is necessary. To evaluate a cluster, normalized mutual information (NMI), a metric based on information theory, can be employed [21]. Alternatively to evaluate a cluster, edited normalized mutual information, ENMI, can be employed [3].

To preserve the diversity, some researchers [23] have suggested a clustering ensemble approach which selects a subset of base data clusterings. They form a smaller but better-performing clustering ensemble than using all primary clusterings. The ensemble selection method is designed based on quality and diversity, the two factors that have been shown to influence cluster ensemble performance. Their method attempts to select a subset of primary clusterings which simultaneously has both the highest quality and diversity. The Sum of Normalized Mutual Information, SNMI [22, 24, 70], is used to measure the quality of an individual partition with respect to other partitions. Also, the NMI is employed to measure the diversity among clusterings. Although the ensemble size in this method is relatively small, this method achieves significant performance improvement over full ensembles.

Inspiring by bagging and boosting algorithms in classification, Parvin et al. [58] have proposed boosting clustering ensemble. They have proposed a subsampling with and without replacement bagging and boosting methods. They have examined the effect of ensemble size, sampling size, and cluster number over quality of consensus clustering. Then, they have proposed a weighted locally adaptive clustering ensemble [53].

Law et al. [40] have proposed a multi-objective data clustering method based on the selection of individual clusters produced by several clustering algorithms through an optimization procedure. This technique chooses the best set of objective functions for different parts of the feature space from the results of base clustering algorithms.

Fred and Jain [24] have offered a new clustering ensemble technique which learns the pair-wise similarity between points in order to facilitate a proper clustering of the data without the a priori information of the number of clusters and of the nature of the clusters. This method which is based on cluster stability assesses the principal clustering results instead of final clustering.

The current article proposes a new clustering ensemble attitude in that it is tried to choose a subset of clusters instead of a subset of clusterings. Also here NMI is utilized to assess a cluster. A scheme introduced by Alizadeh et al. [3] is employed to match two clusters based on which two clusterings are created that they are used in cluster-based NMI calculation.

To make the consensus clustering, a technique based on co-association matrix (CAM) is employed. The results are reported on a variety of experimentations. Extensive

experiments on a variety of real-world datasets have shown that our approach exhibits significant advantages in clustering accuracy and efficiency over the state-of-the-art approaches. Experimental results are reported on some real standard dataset from UCI repository [51], and they prove the efficacy of the proposed technique.

The pair-wise co-occurrence-based approaches typically construct a CAM by considering how many times two objects occur in the same cluster among the multiple base clusterings [22]. By exploiting the CAM as the similarity matrix, the conventional clustering techniques, such as the agglomerative clustering methods [44], can be exploited to build the final clustering result. Fred and Jain [22] for the first time presented the concept of CAM and proposed the evidence accumulation clustering (EAC) method. Wang et al. [68] extended the EAC method by taking the sizes of clusters into consideration and proposed the probability accumulation method. Iam-On et al. [34] refined the CAM by considering the shared neighbors between clusters to improve the consensus results. Wang [67] introduced a dendrogram-like hierarchical data structure termed CA-tree to facilitate the co-association-based ensemble clustering process.

The graph partitioning-based approaches [18, 63] address the ensemble clustering problem by constructing a graph model to reflect the ensemble information. The consensus clustering is then obtained by partitioning the graph into a certain number of segments. Strehl and Ghosh [63] proposed three graph partitioning-based ensemble clustering algorithms, i.e., cluster-based similarity partitioning algorithm (CSPA), hypergraph partitioning algorithm (HGPA), and meta-clustering algorithm (MCLA). Fern and Brodley [18] constructed a bipartite graph for the clustering ensemble by treating both clusters and objects as graph nodes and obtain the consensus clustering by partitioning the bipartite graph.

The median partition-based approaches [11, 20] formulate the ensemble clustering problem into an optimization problem, which aims to find a median partition (or clustering) by maximizing the similarity between this clustering and the multiple base clusterings. The median partition problem is NP hard [66]. Finding the globally optimal solution in the huge space of all possible clusterings is computationally infeasible for large datasets. Cristofor and Simovici [11] proposed to obtain an approximate solution using the genetic algorithm, where clusterings are treated as chromosomes. Topchy et al. [66] cast the median partition problem into a maximum likelihood problem and approximately solve it by the EM algorithm. Franek and Jiang [20] cast the median partition problem into an Euclidean median problem by clustering embedding in vector spaces. The median vector is found by the Weiszfeld algorithm [69] and then transformed into a clustering again, which is treated as the consensus clustering.

Although all approaches have tried to find the solution of the ensemble clustering problem in different perspectives [13, 31, 39], they have a shared limitation that they generally look at all local clusters in the ensemble equally and possibly encounter low-quality clusters. In some of previous published works, this challenge has been discussed [26, 32, 42, 66, 68, 71]. In these papers, a weight is assigned to each base data clustering, but weighting is considered at the clustering level not in cluster level.

3 Proposed clustering ensemble

3.1 Notations

Consider a dataset with N data instances. Let us assume we have an ensemble of the basic data clusterings denoted by φ. The ensemble size is denoted by β, i.e., $\varphi = \left\{ \varphi^1, \ldots, \varphi^\beta \right\}$. The β is set to 30 in this paper throughout all experimentations. Each base data clustering is defined as $\varphi^i = \left\{ \varphi^i_1, \ldots, \varphi^i_{n_i} \right\}$ where $\bigcup_{j=1}^{n_i} \varphi^i_j = \{1, 2, \ldots, N\}$ and n_i is the number of clusters in the ith base data clustering and is greater than or equal to 2 and less than or equal to $4k$ where k is the real number of clusters in the given dataset. The jth cluster the ith base data clustering is denoted by φ^i_j where $\varphi^i_j \subset \{1, 2, \ldots, N\}$, $\varphi^i_j \neq \emptyset$, and for all $p \neq q$ the term $\varphi^i_p \cap \varphi^i_q = \emptyset$ is hold. So $N = \sum_{j=1}^{n_i} \left| \varphi^i_j \right|$ where $|Q|$ is the cardinality of a set Q. We denote the number of all base clusters, i.e., $\sum_{i=1}^{\beta} n_i$, by B. While it may be assumed that the consensus function needs the original data features of dataset as inputs along with the base clusterings [73], the proposed consensus function does not need the original features of dataset as it is widely an ordinary approach [3, 4, 7, 10, 14, 27, 29, 30, 58, 61, 63].

3.2 Backgrounds

In this study, it is illustrated that the diverse clusters in the ensemble can play a role as an indicator for evaluating the reliability of each individual cluster. Next, the conventional CAM is revised into a weighted co-association matrix (WCAM). First, we introduce CAM as a function where its input is an ensemble of data clusterings φ and its output is a matrix of size $N \times N$. The (i, j)th element of CAM is denoted by $\text{CAM}(\varphi)_{ij}$ and defined by Eq. (1).

$$\text{CAM}(\varphi)_{ij} = \frac{1}{\beta} \sum_{q=1}^{\beta} \sum_{p=1}^{n_q} \pi \left(\{i, j\} \subseteq \varphi^q_p \right) \tag{1}$$

where $\pi(a)$ is defined by Eq. (2).

$$\pi(a) = \begin{cases} 1 & \text{if } a == \text{true} \\ 0 & \text{if } a == \text{false} \end{cases} \tag{2}$$

Let us assume Q is a subset of data points (or a cluster), i.e., $Q \subseteq \{1, 2, \dots, N\}$. Now, we define $_Q\varphi$ as projection of a clustering ensemble on the given cluster Q. The projection of the clustering φ^i on cluster Q is defined according to Eq. (3).

$$_Q\varphi^i = \varphi^i \cap Q \tag{3}$$

So it means $_Q\varphi^i_j = \varphi^i_j \cap Q$. It can be inferred that the term $\bigcup_{j=1}^{n_i} {_Q\varphi^i_j} = Q$ holds for any $i \in \{1, 2, \dots, \beta\}$. The average

CAM of a clustering ensemble is denoted by $\text{ACAM}(\varphi)$, and it is defined according to Eq. (4).

$$\text{ACAM}(\varphi) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{CAM}(\varphi)_{ij} \tag{4}$$

The average CAM of the projection of a clustering ensemble on the given cluster Q, i.e., $_Q\varphi$, is denoted by $\text{ACAM}\left(_Q\varphi\right)$, and it is defined as an extension of $\text{ACAM}(\varphi)$. It is defined based on Eq. (5).

$$\text{ACAM}\left(_Q\varphi\right) = \frac{1}{|Q|^2} \sum_{i \in Q} \sum_{j \in Q} \text{CAM}\left(_Q\varphi\right)_{ij} \tag{5}$$

If $\text{ACAM}\left(_Q\varphi\right)$ hits its maximum at one, the cluster Q can be considered as the most reliable cluster for the given ensemble φ. If $\text{ACAM}\left(_Q\varphi\right)$ hits its minimum at $\frac{1}{|Q|}$, the cluster Q can be considered as the least reliable cluster for the given ensemble φ. So the reliability of a cluster in an ensemble φ is defined as follows in Eq. (6).

$$\text{Reliability}(Q, \varphi) = \frac{|Q| \times \text{ACAM}\left(_Q\varphi\right) - 1}{|Q| - 1} \tag{6}$$

The traditional CAM proposed first by Fred and Jain [22] is a $N \times N$ matrix whose (i,j)th element reflects how many times the ith data object is put into the same cluster with the jth data object in the base data clusterings of the ensemble.

3.3 Proposed CLWCAM

Although it is a widely versatile approach to use the CAM in solving the ensemble clustering problem [3, 4, 23, 22, 58], it considers the entire clusters and also the entire primary clusterings in the ensemble in the same way. So it is unable to weight the ensemble members based on their dependability. Huang et al. [32] have utilized the NCAI index to weigh the base clusterings and thereby build a weighted co-association matrix (WCAM). On the other hand, it only considers the dependability of primary clusterings and still pays no attention to the cluster-level quality. Now we introduce a revised version of CAM, called CLWEAC and denoted by cluster-level weighted co-association matrix (CLWCAM). The (i,j)th element of the mentioned matrix denoted by $\text{CLWCAM}(\varphi)_{ij}$ is defined by Eq. (7).

$$\text{CLWCAM}(\varphi)_{ij} = \frac{\sum_{p=1}^{\beta} \sum_{q=1}^{n_p} \pi\left(\text{Reliability}\left(\varphi^p_q, \varphi\right) > \text{MR}\right) \times \text{Reliability}\left(\varphi^p_q, \varphi\right) \times \text{CAM}\left(\varphi^p_q\right)_{ij}}{\sum_{p=1}^{\beta} \sum_{q=1}^{n_p} \pi\left(\text{Reliability}\left(\varphi^p_q, \varphi\right) > \text{MR}\right) \times \text{CAM}\left(\varphi^p_q\right)_{ij}} \tag{7}$$

where MR is the median of the base clusters' reliabilities (or even a less value as later will be explained) and $\text{CAM}\left(\varphi^p_q\right)$ is co-association matrix of the base cluster φ^p_q. The (i,j)th element of $\text{CAM}\left(\varphi^p_q\right)$ is denoted by $\text{CAM}\left(\varphi^p_q\right)_{ij}$ and defined by Eq. (8).

$$\text{CAM}\left(\varphi^p_q\right)_{ij} = \pi\left(\{i,j\} \subseteq \varphi^q_p\right) \tag{8}$$

Using the following example, the proposed method has been illustrated. Assume we have a dataset with 14 samples. Also assume we have produced an ensemble φ with eight base data clusterings on those data. Each clustering contains three clusters except base data clustering φ^5, φ^7 and φ^8 each of which contains 2 clusters.

$$\varphi = \begin{cases} \{\varphi^1_1 = \{1,2,3,8,9,10,12\}, \varphi^1_2 = \{4,5,6,7\}, \varphi^1_3 = \{11,13,14\}\}, \\ \{\varphi^2_1 = \{1,2,3,4,7\}, \varphi^2_2 = \{5,6,8,9,10\}, \varphi^2_3 = \{11,12,13,14\}\}, \\ \{\varphi^3_1 = \{1,2,3,4,5,6,7\}, \varphi^3_2 = \{8,9,10,11,12\}, \varphi^3_3 = \{13,14\}\}, \\ \{\varphi^4_1 = \{1,2,3,5,6,8,9\}, \varphi^4_2 = \{4,7,10,12\}, \varphi^4_3 = \{11,13,14\}\}, \\ \{\varphi^5_1 = \{1,8,9,10,11,12,13,14\}, \varphi^5_2 = \{2,3,4,5,6,7\}\}, \\ \{\varphi^6_1 = \{10,11,12,13\}, \varphi^6_2 = \{1,2,3,4,5,14\}, \varphi^6_3 = \{6,7,8,9\}\}, \\ \{\varphi^7_1 = \{1,4,5,8,9,12,13\}, \varphi^7_2 = \{2,3,6,7,10,11,14\}\}, \\ \{\varphi^8_1 = \{3,4,7,8,11,12\}, \varphi^8_2 = \{1,2,5,6,9,10,13,14\}\} \end{cases} \tag{9}$$

The reliabilities of clusters have been computed based on the proposed method, and then, they are shown in Table 1.

After computing the reliabilities of clusters, the proposed method has selected the half most reliable base clusters as shown in Table 1. Indeed, MR is first set to median of the sorted vector of the base clusters' reliabilities, i.e., 0.51, so the selected base clusters are as mentioned in Table 1. Then, CLWCAM(φ) of our example is computed in Table 2. About this matrix, any value in the main diagonal should at least be a positive value indicating that any data point should be appeared in at least one base selected cluster. If there exists at least one data point whose corresponding value in the main diagonal is zero, we will select the subsequent most reliable base cluster; i.e., MR will be equal to the subsequent smaller value in the sorted vector of the base clusters' reliabilities. After that CLWCAM(φ) is computed, and again if there exists at least one data point whose corresponding value in the main diagonal is zero, we will select the subsequent most reliable base cluster i.e., MR will be equal to the subsequent smaller value in the sorted vector of the base clusters' reliabilities. This procedure is continued until there is no zero in the main diagonal.

After determining MR, we compute CLWCAM(φ). As you can expect the most value of ith row (or ith column) must be in ith column (in ith row) in CLWCAM(φ). But it does not hold in Table 2 (it is simple to show why it happens). So the value of major diagonal in each row is first replaced with the most value possible, i.e., 1. Next, we transform the matrix into a distance matrix (DisM). To do so, we compute the distance matrix based on Eq. (10).

$$DisM(\varphi) = 1 - CLWCAM(\varphi) \tag{10}$$

It is worthy to be mentioned that the major diagonal of DisM is zero.

3.4 Consensus functions

The cluster reliability has been first introduced in this study. Then, based on cluster reliability, we select a subset of the most reliable clusters. Finally by employing a cluster-level weighting approach, 2 novel consensus functions have been introduced: (a) cluster-level weighting evidence accumulation clustering (CLWEAC) and (b) cluster-level weighting graph clustering (CLWGC).

Cluster-level weighting evidence accumulation clustering We first transform an ensemble into a CLWCAM(φ). Then DisM(φ) is computed and using a *hierarchical agglomerative clustering*, the final consensus clustering is produced. The metric of cluster similarity is "*average linkage*" method. This method is abbreviated as CLWEAC.

Table 1 Computed reliabilities of clusters of Eq. (9) and selected clusters for the final ensemble

Clusters	Reliability scores	Order of selection
φ_1^1	0.48	Discard
φ_2^1	0.58	6
φ_3^1	0.67	3
φ_1^2	0.60	4
φ_2^2	0.51	Discard
φ_3^2	0.56	8
φ_1^3	0.57	7
φ_2^3	0.55	9
φ_3^3	0.75	1
φ_1^4	0.54	11
φ_2^4	0.40	Discard
φ_3^4	0.67	2
φ_1^5	0.46	Discard
φ_2^5	0.59	5
φ_1^6	0.54	10
φ_2^6	0.51	Discard
φ_3^6	0.46	Discard
φ_1^7	0.42	Discard
φ_2^7	0.39	Discard
φ_1^8	0.40	Discard
φ_2^8	0.43	Discard

Table 2 Computed CLWCAM(φ) of the ensemble presented in Eq. (9)

0.56	0.58	0.58	0.61	0.58	0.57	0.64	0.52	0.52	0.48	0.46	0.48	0.46	0.50
0.58	0.59	0.59	0.61	0.59	0.58	0.64	0.54	0.54	0.49	0.00	0.49	0.00	0.53
0.58	0.59	0.59	0.61	0.59	0.58	0.64	0.54	0.54	0.49	0.00	0.49	0.00	0.53
0.61	0.61	0.61	0.59	0.61	0.63	0.60	0.00	0.00	0.42	0.00	0.42	0.00	0.53
0.58	0.59	0.59	0.61	0.61	0.61	0.64	0.60	0.60	0.00	0.00	0.00	0.00	0.53
0.57	0.58	0.58	0.63	0.61	0.58	0.59	0.56	0.56	0.49	0.00	0.49	0.00	0.00
0.64	0.64	0.64	0.60	0.64	0.59	0.59	0.53	0.53	0.42	0.00	0.42	0.00	0.00
0.52	0.54	0.54	0.00	0.60	0.56	0.53	0.55	0.55	0.52	0.53	0.52	0.46	0.46
0.52	0.54	0.54	0.00	0.60	0.56	0.53	0.55	0.55	0.52	0.53	0.52	0.46	0.46
0.48	0.49	0.49	0.42	0.00	0.49	0.42	0.52	0.52	0.55	0.60	0.55	0.60	0.55
0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.53	0.60	0.66	0.60	0.67	0.66
0.48	0.49	0.49	0.42	0.00	0.49	0.42	0.52	0.52	0.55	0.60	0.55	0.60	0.55
0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.46	0.60	0.67	0.60	0.70	0.70
0.50	0.53	0.53	0.53	0.53	0.00	0.00	0.46	0.46	0.55	0.66	0.55	0.70	0.67

Cluster-level weighting graph clustering We first transform an ensemble into a bipartite graph. Each base cluster and each data point are considered as a node. The links are defined as follows. (1) There is no link between any pair of data points. (2) There is no link between any pair of base clusters. (3) There is a link between any pair of nodes when one node is a data point and the other is a base cluster; the weight of the link is the reliability of that cluster. So the graph $G(\varphi) = (V(\varphi), E(\varphi))$ is defined as follows. $V(\varphi)$ is defined as a vector whose first N members are dedicated to the N data points of dataset and the last members are the selected base clusters. $E(\varphi)$ is defined based on Eq. (11).

$$
E(\varphi)_{ij} = \begin{cases} \text{Reliability}\left(V_i, \varphi\right) & V_i \in \left\{ \varphi_p^q \middle| \text{Reliability}\left(\varphi_p^q\right) \right\rangle \text{MR} \right\} \wedge V_j = \{1, \ldots, N\} \wedge V_j \in V_i \\ \text{Reliability}\left(V_j, \varphi\right) & V_j \in \left\{ \varphi_p^q \middle| \text{Reliability}\left(\varphi_p^q\right) \right\rangle \text{MR} \right\} \wedge V_i = \{1, \ldots, N\} \wedge V_i \in V_j \\ 0 & \text{else} \end{cases} \tag{11}
$$

This method is abbreviated as CLWGC. After building the CLWGC based on Eq. (11), the graph is partitioned through the TCUT method where it is a fast method for bipartite graphs [43]. The data points located in each partition are considered as a consensus cluster.

In our example, the consensus partition or clustering obtained by CLWEAC and CLWGC is $\{\{11, 12, 13, 14\}, \{8, 9, 10\}, \{1, 2, 3, 4, 5, 6, 7\}\}$. But the consensus clustering obtained by EAC is $\{\{11, 13, 14\}, \{8, 9, 10, 12\}, \{1, 2, 3, 4, 5, 6, 7\}\}$. It is clear that the consensus partition obtained by CLWEAC and CLWGC is better than the one obtained by EAC.

4 Experimentations

The proposed CLWEAC and CLWGC have been assessed in the current section. Also the state-of-the-art ensemble clustering methods have been evaluated comparatively. The experimentations have been done a number of standard datasets.

4.1 Datasets and evaluation metric

In our experiments, 10 real benchmark datasets have been utilized: Semeion (with 1593 data points, 256 features, and 10 classes), multiple features (MF) (with 2000 data points, 649 features, and 10 classes), image segmentation (IS) (with 2310 data points, 19 features, and 7 classes), Forest-Cover-Type (FCT) (with 3780 data points, 54 features, and 7 classes), MNIST(with 5000 data points, 784 features, and 10 classes), optical digit recognition (ODR) (with 5620 data points, 64 features, and 10 classes), Landsat Satellite (LS) (with 6435 data points, 36 features, and 6 classes), ISOLET (with 7797 data points, 617 features, and 26 classes), USPS

(with 11,000 data points, 256 features, and 10 classes), and letter recognition (LR) (with 20,000 data points, 16 features, and 26 classes). Except the MNIST benchmark dataset [41] and the USPS benchmark dataset [15], the other eight dataset benchmarks have been from the UCI machine learning repository [51].

We apply the normalized mutual information (NMI) [22, 62, 63] to assess the quality of any clustering. The NMI measure indicates the information shared between a pair of clusterings. It has been one of the most widespread evaluation metrics to analyze a data clustering. The bigger the NMI, the higher the clustering quality.

The base clusterings are produced by *k*-means and fuzzy *c*-means clustering algorithms. It is worthy to be mentioned that clusters produced by the fuzzy *c*-means clustering algorithm are first transformed into crisp clusters. Then in following, the all process for both base clustering algorithms is the same.

The number of clusters in each base clustering is randomly chosen from range $[2, 4k]$. Each result is average of 30 independent runs. For each run of the algorithm, an ensemble of size 30 is produced. After computing the consensus clustering over the ensemble of size 30 with a clustering method, the NMI between the consensus clustering and the ground-truth labels is stored. After repeating the process for 30 independent runs, 30 NMI values are stored. The average of these 30 NMI values is reported as the performance of the clustering method over that dataset.

4.2 Simple clustering algorithms

It is widely acceptable in SPR communities that the clustering ensembles outperform the simple base clusterings in terms of quality and robustness. The performance of the proposed CLWGC with the *k*-means and fuzzy *c*-means clustering algorithms as its simple basic clusterer is compared with the ones of the *k*-means and fuzzy *c*-means clustering algorithms in terms of NMI in Table 3. Each result is an averaged NMI over 30 independent runs. The proposed CLWGC clearly outperforms the simple *k*-means and fuzzy *c*-means clustering algorithms on all benchmark datasets.

4.3 State-of-the-art clustering algorithms

The proposed CLWEAC and CLWGC methods with the *k*-means and fuzzy *c*-means clustering algorithms as its simple basic clusterer are evaluated, and their results have

Table 3 Performance comparison of the base clustering algorithms with CLWGC when *k*-means and fuzzy *c*-means are used as its base clustering algorithm

Dataset	*k*-means	Fuzzy *c*-means	CLWGC with *k*-means	CLWGC with fuzzy *c*-means
Semeion	48.34	49.97	66.81	66.43
MF	50.60	52.41	68.89	69.13
IS	49.49	49.34	67.05	67.26
FCT	19.30	20.13	23.31	23.38
MNIST	44.41	44.67	65.26	64.16
ODR	57.71	57.91	83.12	82.57
LS	38.17	38.73	63.28	61.42
ISOLET	64.96	66.05	76.38	75.51
USPS	41.65	43.57	65.68	65.56
LR	38.16	40.09	41.47	41.69

been compared to 12 state-of-the-art ensemble clustering methods, that is, CSPA [63], HGPA [63], MCLA [63], hybrid bipartite graph formulation (HBGF) [18], SimRank similarity-based method (SRS) [33], weighted connected triple-based method (WCT) [34], cluster selection evidence accumulation clustering (CSEAC) [4], weighted evidence accumulation clustering (WEAC) [32], Wisdom of Crowds Ensemble (WCE) [5], graph partitioning with multi-granularity link analysis (GPMGLA) [32], and two-level co-association matrix ensemble (TME) [73], elite cluster selection evidence accumulation clustering (ECSEAC) [53].

Each method is executed 30 independent trials for a dataset. So, each of the reported results are the averaged value on 30 independent trials. Table 4 reports the results of the proposed CLWEAC and CLWGC methods with the *k*-means and fuzzy *c*-means clustering algorithms as its simple basic

Table 4 Performances of the proposed CLWEAC and CLWGC methods with the *k*-means and fuzzy c-means clustering algorithms as its simple basic clusterer compared with 12 state-of-the-art ensemble clustering methods

Dataset	CSPA	HGPA	MCLA	HBGF	CSEAC	SRS	WCT	WEAC	GPMGLA	TME	ECSEAC	WCE	CLWEAC	CLWGC
Semeion														
Fcm	0.531	0.485	0.577	0.619	0.619	0.630	0.635	0.624	0.622	0.585	0.619	0.616	0.645	0.640
K-means	0.539	0.454	0.574	0.623	0.618	0.634	0.630	0.630	0.625	0.577	0.624	0.617	0.657	0.644
MF														
Fcm	0.607	0.520	0.635	0.640	0.614	0.632	0.632	0.620	0.642	0.683	0.652	0.650	0.681	0.677
K-means	0.619	0.481	0.629	0.638	0.599	0.621	0.616	0.611	0.640	0.689	0.638	0.649	0.661	0.684
IS														
Fcm	0.593	0.483	0.614	0.609	0.602	0.606	0.612	0.601	0.618	0.556	0.609	0.618	0.629	0.646
K-means	0.612	0.448	0.623	0.611	0.601	0.606	0.605	0.602	0.621	0.478	0.602	0.606	0.623	0.631
FCT														
Fcm	0.202	0.154	0.214	0.215	0.211	0.212	0.219	0.214	0.219	0.210	0.218	0.217	0.229	0.226
K-means	0.209	0.220	0.220	0.190	0.205	0.211	0.213	0.208	0.201	0.201	0.220	0.205	0.232	0.202
MNIST														
Fcm	0.509	0.408	0.556	0.600	0.593	0.602	0.616	0.605	0.610	0.566	0.552	0.597	0.537	0.628
K-means	0.523	0.298	0.556	0.609	0.594	0.605	0.615	0.617	0.618	0.555	0.604	0.598	0.648	0.637
ODR														
Fcm	0.723	0.603	0.774	0.801	0.789	0.798	0.804	0.802	0.807	0.796	0.795	0.791	0.820	0.813
K-means	0.738	0.409	0.775	0.810	0.781	0.799	0.798	0.801	0.813	0.794	0.790	0.784	0.829	0.816
LS														
Fcm	0.502	0.366	0.527	0.610	0.577	0.582	0.602	0.593	0.609	0.501	0.579	0.587	0.612	0.628
K-means	0.485	0.312	0.518	0.618	0.559	0.569	0.602	0.601	0.619	0.510	0.578	0.592	0.616	0.644
ISOLET														
Fcm	0.650	0.617	0.699	0.729	0.729	0.733	0.735	0.729	0.727	0.692	0.720	0.717	0.736	0.729
K-means	0.629	0.486	0.665	0.742	0.730	0.726	0.719	0.734	0.740	0.691	0.716	0.715	0.745	0.743
USPS														
Fcm	0.493	0.353	0.533	0.577	0.562	0.576	0.584	0.582	0.589	0.581	0.592	0.587	0.641	0.630
K-means	0.501	0.120	0.531	0.581	0.555	0.568	0.579	0.583	0.597	0.573	0.576	0.566	0.633	0.614
LR														
Fcm	0.327	0.339	0.384	0.420	0.411	0.410	0.414	0.415	0.420	0.407	0.410	0.411	0.426	0.428
K-means	0.282	0.176	0.353	0.387	0.367	0.373	0.386	0.386	0.394	0.355	0.382	0.379	0.418	0.413
Average														
Fcm	0.514	0.433	0.551	0.582	0.571	0.578	0.585	0.579	0.586	0.558	0.575	0.579	0.596	0.605
K-means	0.514	0.34	0.544	0.581	0.561	0.571	0.576	0.577	0.587	0.542	0.573	0.571	0.606	0.603

clusterer along with those of the state-of-the-art ensemble clustering methods.

According to Table 4, the proposed CLWEAC and CLWGC methods outperform all the state-of-the-art ensemble clustering methods irrespective to the simple basic clusterer (it should be either the *k*-means clustering algorithm or the fuzzy *c*-means clustering algorithm). Table 5 summarizes Table 4. According to Table 5, both of the proposed CLWEAC and CLWGC methods are always the best two methods among all of the state-of-the-art ensemble clustering methods.

4.4 Robustness analysis

The performances of the proposed CLWEAC and CLWGC methods and the state-of-the-art ensemble clustering methods have been evaluated on all dataset with different ensemble sizes, i.e., β. For each dataset, for each method, and for each ensemble size, the experiment has been done 30 times and the averaged performance has been considered to make our results reliable. Then the average NMIs over all datasets are depicted in Fig. 2 when the *k*-means clustering algorithm is considered as its simple basic clusterer. The same results are repeated in Fig. 3 when the fuzzy *c*-means clustering algorithm is employed as its simple basic clusterer.

According to Figs. 2 and 3, in terms of the average NMI scores over all 10 benchmark datasets with different β, the

proposed CLWEAC and CLWGC methods have the best consensus performances. They are also robust to ensemble size.

4.5 Complexity analysis

In this section, the run time of multiple ensemble clustering methods varying data sizes is compared. The experiments are yielded on various subsets of the LR dataset made up of totally 20,000 data objects. Different methods are executed on randomly selected subsets of N' objects from the LR dataset to evaluate their run time. As depicted in Fig. 4, the MCLA method and the proposed CLWGC method are, respectively, the fastest methods to process the entire LR dataset. So, CLWGC method is superior to CLWEAC method, both in terms of output clustering quality and time complexity.

5 Conclusion and future works

This study has proposed a new clustering ensemble framework based on cluster-level weighting. The paper has considered the certainty amount the given ensemble has about a cluster as the reliability of that cluster. The certainty amount the given ensemble has about a cluster has been computed by the accretion amount of that cluster by the ensemble. Then

Table 5 Summary of the results presented in Table 4

	CSPA	HGPA	MCLA	HBGF	CSEAC	SRS	WCT	WEAC	GPMGLA	TME	ECSEAC	WCE	CLWEAC	CLWGC
#Times being ranked in top 2														
Fcm	0	0	0	0	0	0	2	0	0	1	0	0	9	8
K-means	0	0	0	0	0	0	0	0	1	1	1	0	8	9
Overall average NMI														
Fcm	0.514	0.433	0.551	0.582	0.571	0.578	0.585	0.579	0.586	0.558	0.575	0.579	0.596	0.605
K-means	0.514	0.34	0.544	0.581	0.561	0.571	0.576	0.577	0.587	0.542	0.573	0.571	0.606	0.603

Fig. 2 Effect of different ensemble sizes on different clustering methods when the *k*-means clustering algorithm is considered as their simple basic clusterer. NMI values of the different clustering methods with varying ensemble sizes
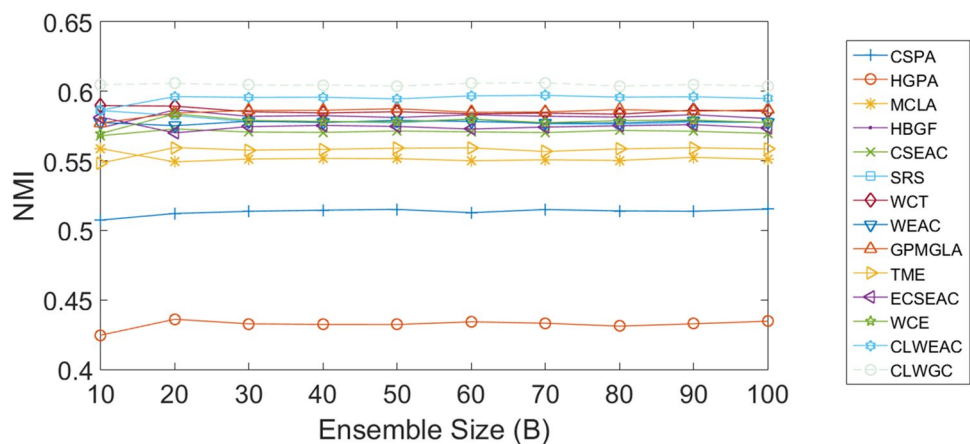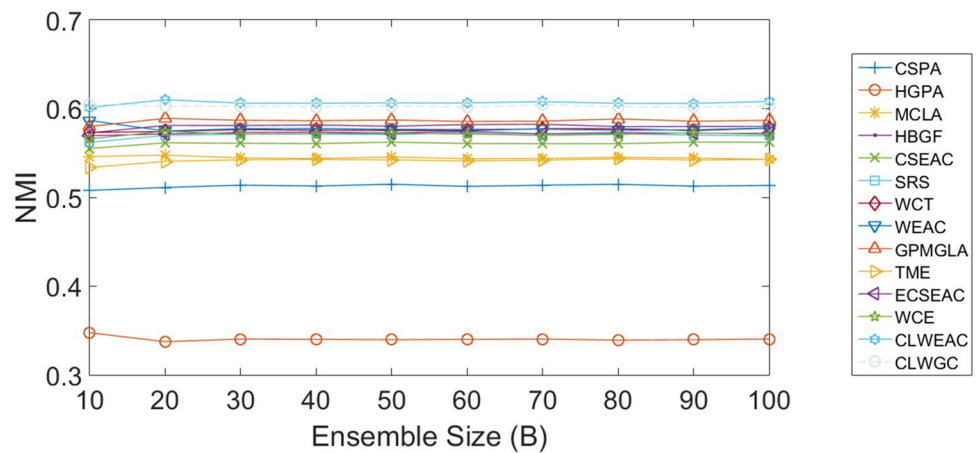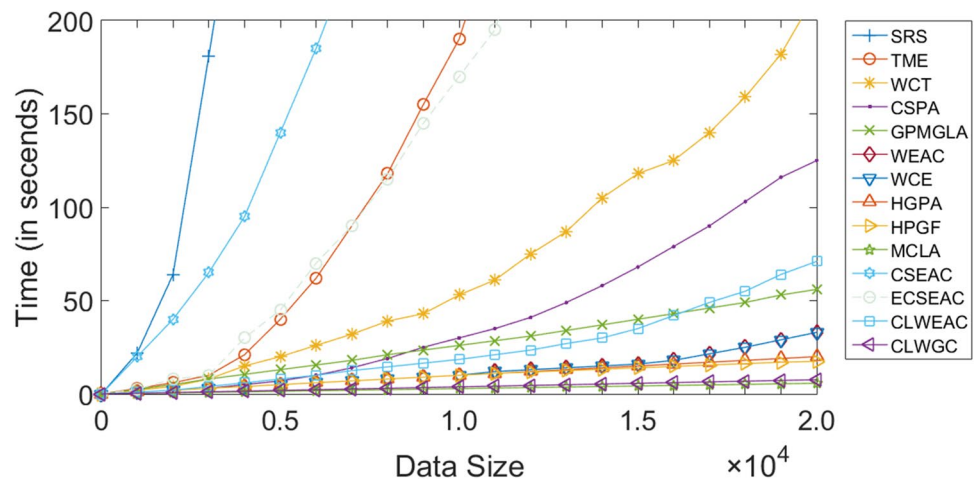
Fig. 3 Effect of different ensemble sizes on different clustering methods when the fuzzy *c*-means clustering algorithm is considered as their simple basic clusterer. NMI values of the different clustering methods with varying ensemble sizes

Fig. 4 Run time of different methods with varying data sizes

by selecting the best clusters and assigning a weight to each selected cluster based on its reliability, the final ensemble has been created. At the next step, the paper has proposed CLWCAM instead of traditional CAM. Then, two consensus functions have been introduced and used for production of the consensus partition. The proposed framework completely overshadows the state-of-the-art clustering ensemble methods experimentally.

For the future work, one can explore more diversity-oriented reliability metrics. Using optimization methods to select an optimized subset of clusters can be also explored. For example, the most diverse subset can be sought while considering cluster-level weighting.

References

1. Alizadeh H, Minaei-Bidgoli B, Amirgholipour SK (2009) A new method for improving the performance of K nearest neighbor using clustering technique. Int J Converg Inf Technol JCIT. ISSN: 1975-9320

2. Alizadeh H, Minaei-Bidgoli B, Parvin H (2013) Optimizing fuzzy cluster ensemble in string representation. IJPRAI 27(2). https://doi.org/10.1142/S0218001413500055

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

4. Alizadeh H, Minaei-Bidgoli B, Parvin H (2014) To improve the quality of cluster ensembles by selecting a subset of base clusters. J Exp Theor Artif Intell 26(1):127–150

5. Alizadeh H, Yousefnezhad M, Minaei-Bidgoli B (2015) Wisdom of crowds cluster ensemble. Intell Data Anal 19(3):485–503

6. Aminsharifi A, Irani D, Pooyesh S, Parvin H, Dehghani S, Yousofi K, Fazel E, Zibaie F (2017) Artificial neural network system to predict the postoperative outcome of percutaneous nephrolithotomy. J Endourol 31(5):461–467

7. Ana LNF, Jain AK (2003) Robust data clustering. In: Proceedings. 2003 IEEE computer society conference on computer vision and pattern recognition, 2003, vol 2. IEEE, pp II-128

8. 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

9. Charon I, Denoeud L, Guénoche A, Hudry O (2006) Maximum transfer distance between partitions. J Classif 23(1):103–121

10. Coretto P, Hennig C (2010) A simulation study to compare robust clustering methods based on mixtures. Adv Data Anal Classif 4:111–135

11. Cristofor D, Simovici D (2002) Finding median partitions using information-theoretical-based genetic algorithms. J Univ Comput Sci 8(2):153–172

12. Denoeud L (2008) Transfer distance between partitions. Adv Data Anal Classif 2:279–294

13. Di Gesù V (1994) Integrated fuzzy clustering. Fuzzy Sets Syst 68(3):293–308

14. Domeniconi C, Al-Razgan M (2009) Weighted cluster ensembles: methods and analysis. ACM Trans Knowl Discov Data (TKDD) 2(4):17

15. Dueck D (2009) Affinity propagation: clustering data by passing messages. Ph.D. dissertation, University of Toronto

16. Faceli K, Marcilio CP, Souto D (2006) Multi-objective clustering ensemble. In: Proceedings of the sixth international conference on hybrid intelligent systems (HIS'06)

17. Fern XZ, Brodley CE (2003) Random projection for high dimensional data clustering: a cluster ensemble approach. In: ICML, vol 3, pp 186–193

18. Fern XZ, Brodley CE (2004) Solving cluster ensemble problems by bipartite graph partitioning. In Proceedings of international conference on machine learning (ICML)

19. Fern XZ, Lin W (2008) Cluster ensemble selection. In: SIAM international conference on data mining

20. Franek L, Jiang X (2014) Ensemble clustering by means of clustering embedding in vector spaces. Pattern Recognit 47(2):833–842

21. Fred A (2001) Finding consistent clusters in data partitions. In: Multiple classifier systems. Springer, Berlin, Heidelberg, pp 309–318

22. Fred A, Jain AK (2002) Data clustering using evidence accumulation. In: Proceedings of the 16th international conference on pattern recognition, ICPR02, Quebec City, pp 276–280

23. Fred ALN, Jain AK (2005) Combining multiple clusterings using evidence accumulation. IEEE Trans Pattern Anal Mach Intell 27(6):835–850

24. Fred A, Jain AK (2006) Learning pairwise similarity for data clustering. In: International conference on pattern recognition

25. Fred A, Lourenco A (2008) Cluster ensemble methods: from single clusterings to combined solutions. Stud Comput Intell (SCI) 126:3–30

26. Friedman JH, Meulman JJ (2002) Clustering objects on subsets of attributes. Technical report, Stanford University

27. García-Escudero LA, Gordaliza A, Matrán C, Mayo-Iscar A (2010) A review of robust clustering methods. Adv Data Anal Classif 4:89–109

28. Guénoche A (2011) Consensus of partitions: a constructive approach. Adv Data Anal Classif 5:215–229

29. Gullo F, Tagarelli A, Greco S (2009) Diversity-based weighting schemes for clustering ensembles. SIAM, pp 437–448

30. Hennig C (2008) Dissolution point and isolation robustness: robustness criteria for general cluster analysis methods. J Multivar Anal 99:1154–1176

31. Hu X, Yoo I (2004) Cluster ensemble and its applications in gene expression analysis. In: Proceedings of the second conference on Asia-Pacific bioinformatics-Volume 29. Australian Computer Society, Inc, pp 297–302

32. Huang D, Lai JH, Wang CD (2015) Combining multiple clusterings via crowd agreement estimation and multi-granularity link analysis. Neurocomputing 170:240–250

33. Iam-On N, Boongoen T, Garrett S (2008) Refining pairwise similarity matrix for cluster ensemble problem with cluster relations. In: Proceedings of international conference on discovery science (ICDS), pp 222–233

34. 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

35. Jain AK (2010) Data clustering: 50 years beyond k-means. Pattern Recognit Lett 31(8):651–666

36. Jamalinia H, Khalouei S, Rezaie V, Nejatian S, Bagheri-Fard K, Parvin H (2017) Diverse classifier ensemble creation based on heuristic dataset modification. J Appl Stat. https://doi.org/10.1080/02664763.2017.1363163 **(in press)**

37. Kleinberg J (2002) An impossibility theorem for clustering. In: Proceedings of Neural Information Processing Systems'02 (NIPS 2002). pp 446–453

38. Kuncheva LI (2004) Combining pattern classifiers, methods and algorithms. Wiley, New York

39. Kuncheva LI, Hadjitodorov ST (2004) Using diversity in cluster ensembles. In: 2004 IEEE international conference on systems, man and cybernetics, vol 2. IEEE, pp 1214–1219

40. Law MHC, Topchy AP, Jain AK (2004) Multiobjective data clustering. In: IEEE conference on computer vision and pattern recognition, vol 2, pp 424–430

41. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

42. Li T, Ding C (2008) Weighted consensus clustering. In: Proceedings of SIAM international conference on data mining (SDM)

43. Li Z, Wu XM, Chang SF (2012) Segmentation using superpixels: a bipartite graph partitioning approach. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR)

44. Lior R, Maimon O (2005) Data mining and knowledge discovery handbook. Springer, Berlin

45. Minaei-Bidgoli B, Parvin H, Alinejad-Rokny H, Alizadeh H, Punch WF (2014) Effects of resampling method and adaptation on clustering ensemble efficacy. Artif Intell Rev 41(1):27–48

46. Mohammadi Jenghara M, Ebrahimpour-Komleh H, Parvin H (2017) Dynamic protein–protein interaction networks construction using firefly algorithm. Pattern Anal Appl. https://doi.org/10.1007/s10044-017-0626-7 **(in press)**

47. Mohammadi Jenghara M, Ebrahimpour-komleh H, Rezaie V, Nejatian S, Parvin H, Syed-Yusof SK (2017) Imputing missing value through ensemble concept based on statistical measures. Knowl Inf Syst. https://doi.org/10.1007/s10115-017-1118-1 **(in press)**

48. Monti S, Tamayo P, Mesirov J, Golub T (2003) Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach Learn 52(1–2):91–118

49. Nejatian S, Omidvar R, Mohamadi H, Eskandar-Baghbani A, Rezaie V, Parvin H (2017) An optimization algorithm based on behavior of see-see partridge chicks. J Intell Fuzzy Syst. https://doi.org/10.3233/JIFS-161718 **(in press)**

50. Nejatian S, Parvin H, Faraji E (2017) Using sub-sampling and ensemble clustering techniques to improve performance of imbalanced classification. Neurocomputing. https://doi.org/10.1016/j.neucom.2017.06.082 **(in press)**

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

52. Parvin H, Minaei-Bidgoli B (2013) A clustering ensemble framework based on elite selection of weighted clusters. Adv Data Anal Classif 7(2):181–208

53. 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

54. Parvin H, Alizadeh H, Minaei-Bidgoli B (2009a) A new method for constructing classifier ensembles. Int J Digit Content Technol Appl JDCTA. ISSN: 1975-9339 (in press)

55. Parvin H, Alizadeh H, Minaei-Bidgoli B (2009b) Using clustering for generating diversity in classifier ensemble. Int J Digit Content Technol Appl JDCTA 3(1):51–57. ISSN: 1975-9339

56. 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
57. Parvin H, Alinejad-Rokny H, Minaei-Bidgoli B, Parvin S (2013) A new classifier ensemble methodology based on subspace learning. J Exp Theor Artif Intell 25(2):227–250
58. Parvin H, Minaei-Bidgoli B, Alinejad-Rokny H, Punch WF (2013) Data weighing mechanisms for clustering ensembles. Comput Electr Eng 39(5):1433–1450
59. Parvin H, Mirnabibaboli M, Alinejad-Rokny H (2015) Proposing a classifier ensemble framework based on classifier selection and decision tree. Eng Appl AI 37:34–42
60. Schynsa M, Haesbroeck G, Critchley F (2010) RelaxMCD: smooth optimisation for the minimum covariance determinant estimator. Comput Stat Data Anal 54:843–857
61. Sevillano X, Cobo G, Alías F, Socoró JC (2006) Feature diversity in cluster ensembles for robust document clustering. In: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval. ACM, pp 697–698
62. Strehl A, Ghosh J (2002) Cluster ensembles-a knowledge reuse framework for combining partitionings. In: AAAI/IAAI, pp 93–99
63. Strehl A, Ghosh J (2003) Cluster ensembles: a knowledge reuse framework for combining multiple partitions. J Mach Learn Res 3:583–617
64. Tavana M, Parvin H, Rezazadeh F (2017) Parkinson detection: an image processing approach. J Med Imaging Health Inf 7:464–472
65. Topchy A, Jain AK, Punch WF (2003) Combining multiple weak clusterings. In: Proceedings of the 3rd IEEE international conference on data mining, pp 331–338
66. Topchy A, Jain AK, Punch W (2005) Clustering ensembles: models of consensus and weak partitions. IEEE Trans Pattern Anal Mach Intell 27(12):1866–1881
67. Wang T (2011) CA-Tree: a hierarchical structure for efficient and scalable co-association-based cluster ensembles. IEEE Trans Syst Man Cybern Part B Cybern 41(3):686–698
68. Wang X, Yang C, Zhou J (2009) Clustering aggregation by probability accumulation. Pattern Recognit 42(5):668–675
69. Weiszfeld E, Plastria F (2009) On the point for which the sum of the distances to n given points is minimum. Ann Oper Res 167(1):7–41
70. Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13(4):841–846
71. Yu Z, Li L, Gao Y, You J, Liu J, Wong HS, Han G (2014) Hybrid clustering solution selection strategy. Pattern Recognit 47(10):3362–3375
72. Zare N, Shameli H, Parvin H (2017) An innovative natural-derived meta-heuristic optimization method. Appl Intell. https://doi.org/10.1007/s10489-016-0805-z (in press)
73. Zhong C, Yue X, Zhang Z, Lei J (2015) A clustering ensemble: two-level-refined co-association matrix with path-based transformation. Pattern Recognit 48(8):2699–2709