

Cluster Ensemble Selection Based on a New Cluster Stability Measure*

Hosein Alizadeh[†], Behrouz Minaei-Bidgoli, Hamid Parvin
School of Computer Engineering, Iran University of Science and Technology (IUST), Tehran, Iran

Abstract. Many stability measures, such as Normalized Mutual Information (NMI), have been proposed to validate a set of partitionings. It is highly possible that a set of partitionings may contain one (or more) high quality cluster(s) but is still adjudged a bad cluster by a stability measure, and as a result, is completely neglected. Inspired by evaluation approaches measuring the efficacy of a set of partitionings, researchers have tried to define new measures for evaluating a cluster. Thus far, the measures defined for assessing a cluster are entirely based on the well-known NMI measure. The drawback of this commonly used approach is discussed in this paper, after which a new asymmetric criterion, called the Alizadeh–Parvin–Moshki–Minaei criterion (APMM), is proposed to assess the association between a cluster and a set of partitionings. The APMM criterion overcomes the deficiency in the conventional NMI measure. We also propose a clustering ensemble framework that incorporates the APMM’s capabilities in order to find the best performing clusters. The framework uses Average APMM (AAPMM) as a fitness measure to select a number of clusters instead of using all of the results. Any cluster that satisfies a predefined threshold of the mentioned measure is selected to participate in an elite ensemble. To combine the chosen clusters, a co-association matrix-based consensus function (by which the set of resultant partitionings are obtained) is used. Because Evidence Accumulation Clustering (EAC) can not derive the co-association matrix from a subset of clusters, a new EAC-based method, called Extended EAC (EEAC), is employed to construct the co-association matrix from the chosen subset of clusters. Empirical studies show that our proposed approach outperforms other cluster ensemble approaches.

Keywords: Clustering Ensemble, APMM Stability Measure, Extended Evidence Accumulation Clustering, Cluster Evaluation.

1 Introduction

Data clustering is an important and very challenging problem. The objective of clustering is to partition a set of unlabeled objects into homogeneous groups or clusters [9]. There are many applications that use clustering techniques to discover structures in data. These include data mining [9], information retrieval, image segmentation, and machine learning [11]. In real-world problems, the clusters can have different shapes, sizes, and degrees of separation and sparseness. Clustering techniques require the definition of a similarity measure between patterns that have extreme effects on the set of resultant clusters. Due to a lack of any prior knowledge about cluster shapes, choosing a specialized clustering method is not easy [23]. Studies in the past few years have tended to combinational methods. Cluster ensemble methods attempt to find a better, more robust clustering solution by fusing as much information as possible among several distinct executions of a partitioning algorithm over primary data [1].

Clustering ensemble [33,34] is the problem of deriving a consensus **partitioning** given a number of resultant outputs of a clustering algorithm. The resultant outputs of the clustering algorithm used is called an *ensemble*. The consensus **partitioning** should be the most representative one for all members in the ensemble, and it optimizes a certain objective function. Generally, there are two main steps in the clustering ensemble: (a) the creation of some weak partitionings (the output of a clustering algorithm is called a partitioning), and (b) the aggregation of the primary partitionings obtained.

The first step is the creation of some weak partitionings. Because every primary partitioning reveals a hidden aspect of some data, their ensemble can compensate for their individual downside. Therefore, the primary resultant partitionings need to be as diverse as possible in order to give more information about the underlying patterns in the data. Many methods to create the necessary diversity for the primary results have been proposed. However, the use of different clustering algorithms is the most straightforward method of obtaining diverse

*This work is based on our earlier work: Alizadeh, H., Minaei-Bidgoli, B., Parvin, H., and Moshki M (2011), An Asymmetric Criterion for Cluster Validation. Developing Concepts in Applied Intelligence, Studies in Computational Intelligence, vol. 363, pp. 1-14.

[†] Corresponding author. E-mail: halizadeh@iust.ac.ir

primary resultant partitionings. Other methods include the selection of different initializations, different algorithm parameters, subset of features, mapping the data to other feature spaces [1], and resampling of the data [19]. In this paper, techniques based on the use of different base algorithms, different initializations, and different parameters, and the resampling method are considered to provide the necessary diversity for the primary results.

The second step in the clustering ensemble is to combine the primary partitionings obtained in the first step. The co-association matrix-based aggregator is one of the most common methods used to combine the primary partitionings; it is also employed in this study. EAC, which was first proposed by Fred and Jain [12], maps the individual data partitionings in a clustering ensemble into a new between-patterns similarity measure, summarizing inter-pattern structures perceived from these clusterings. The final data partitioning is obtained by applying the single-link method to this new similarity matrix.

In this paper, a new clustering ensemble method that uses a subset of primary clusters is proposed. A new validity measure, called AAPMM, to evaluate the cluster goodness is also proposed. Each cluster that satisfies a threshold of the measure can be considered to participate in the construction of the co-association matrix. In addition, a new method, called EEAC, is proposed for establishing the matrix. Finally, a hierarchical method is applied over the obtained matrix to extract the consensus partitioning. In summary, the main contributions of this paper are as follows:

- 1- We propose a general framework for cluster ensemble selection.
- 2- After examining the strengths and weaknesses of the well-known NMI measure, we propose the APMN criterion, which evaluates individual clusters.
- 3- We offer a simple but efficient EEAC method to accumulate the selected clusters from primary partitionings in the co-association matrix.
- 4- We provide a set of experimental results that gives credence to our idea and demonstrate the efficacy of our proposed method.

The rest of this paper is organized as follows. The next section reviews background as well as recent related works on the cluster ensemble problem. In Section 3, we introduce our framework and discuss the NMI, APMN, and EEAC methods in detail. In Section 4, we present a number of empirical results obtained using diverse datasets and compare them directly with other well-known methods. Finally, we conclude this paper in Section 5 with some directions for future research.

2 Background

Similar to its closely related problem in classification, combining classifiers [36-39], the cluster ensemble has evoked an increasing interest among researchers. We next review some of the newest and most important studies of cluster ensemble methods.

Although significant research has been conducted on the issue of clustering thus far, it has remained a challenging problem. In this respect, researchers in the field of clustering have turned to *projective clustering* and *clustering ensembles*. In a work done by Gullo et al., a novel clustering problem, called projective clustering ensembles, was introduced [30], and subsequently enhanced [31].

Minaei et al. [32] proposed non-adaptive and adaptive resampling schemes for the integration of multiple independent and dependent clusterings. They also examined the efficacy of the bagging method, with and without replacement. In their work, primary partitionings in the ensemble are sequentially generated by clustering specially selected subsamples of the given dataset.

The main challenge faced by a clustering ensemble that is based on a subset of selected primary clusters or partitionings is the manner in which clusters or partitionings are evaluated. As data clustering is an unsupervised problem, its validation process is a most troublesome task. Baumgartner et al. [2] presented a resampling-based technique to validate the results of exploratory fuzzy clustering analysis. Since the concept of cluster stability was introduced as a means to assess the validity of data partitionings, it has been increasingly used in the literature [13]. This idea, which is based on the resampling method, was initially described in [5] and later generalized differently in [15]. Roth et al. [23] proposed a resampling based technique to validate a cluster. The basic element in their method, which is a complementary version of the past methods, is cluster stability. The stability measures the association between obtained partitionings from two individual clustering algorithms. The large value of the stability measure lies in the fact that applying the clustering algorithm several times on a dataset probably yields fixed results [21]. Roth and Lange [22] presented a new algorithm for data clustering based on feature selection. In their method, the resampling based stability measure is used to set the algorithm parameters. There are several cluster validation methods that are based on the stability concept [18]. Ben-Hur et al. [3] defined a stability measure for clustering solutions. It is based on the perturbation of a dataset. In their approach, stability is characterized by the distribution of the pairwise similarities between clusterings obtained

from sub-samples of the data. First, the co-association matrix is acquired using the resampling method. Then, the Jaccard coefficient is extracted from this matrix as the stability measure. Estivill-Castro and Yang [8] proposed a method in which Support Vector Machines (SVMs) are used to evaluate the separation of the clustering results. By filtering noise and outliers, this method can identify robust and potentially meaningful clustering results.

Moller and Radke [20] introduced an approach that validates clustering results based on the partitioning of stability measure. The approach uses perturbation produced by adding some noise to the data. Their empirical study indicates that the perturbation usually outperforms bootstrapping and subsampling. While the empirical choice of the subsampling size is often difficult [7], the choice of perturbation strength is not very crucial. The method uses a Nearest Neighbor Resampling (NNR) approach that offers a solution to problems of both information loss and empirical control over the degree of modification of the original data. The NNR technique was first used for time series analysis [4]. Inokuchi et al. [16] proposed a kernelized validity measure, where kernel refers to the kernel function used by SVMs. Two objectives are considered in this measure. The first is the sum of the traces of the fuzzy covariances within clusters, while the second is a kernelized Xie–Beni measure [25]. The validity measure is applied to determine the number of clusters and to evaluate the robustness of different partitionings. Das and Sil [6] proposed a method that determines the number of clusters and validates the clusters using a splitting and merging technique in order to obtain the optimal set of clusters.

Fern and Lin [10] proposed a clustering ensemble approach that selects a subset of solutions to form a smaller but better performing cluster ensemble than all primary solutions as a cluster ensemble. The ensemble selection method is designed based on quality and diversity, the two factors that have been shown to improve the cluster ensemble performance [10,36]. Their method attempts to select a subset of primary partitionings that simultaneously has both the highest quality and the highest diversity. The Sum of Normalized Mutual Information (SNMI) is used to measure the quality of any individual partitioning with respect to the other ones [11,24]. In addition, the NMI is employed to measure the diversity between partitionings. Although the ensemble size in their method is relatively small, the method can achieve a significant performance improvement over a full ensemble. Law et al. [17] proposed a multi-objective data clustering method based on the selection of individual clusters produced by several clustering algorithms, through an optimization procedure. The method chooses the best set of objective functions for different parts of the feature space from the results of base clustering algorithms. Fred and Jain [13] presented a new clustering ensemble method that learns the pairwise similarity between points in order to facilitate a proper partitioning of the data without a priori knowledge of the number of clusters or the shapes of the clusters. The method is based on cluster stability and it evaluates the primary clustering results instead of those of the final clustering.

Similar to Fern and Lin [10], Alizadeh et al. [28] recently proposed a selection procedure that makes an ensemble of better performing clusters. They also introduced a new metric, called MAX, to evaluate clusters and to compare them. They empirically showed that choosing better clusters (with respect to the MAX metric) yields better consensus partitioning. This means that there is analogy between the MAX value of the chosen clusters and the performance of final clustering. However, Azimi and Fern [35] showed that choosing better primary results based on NMI will not always yield better final results.

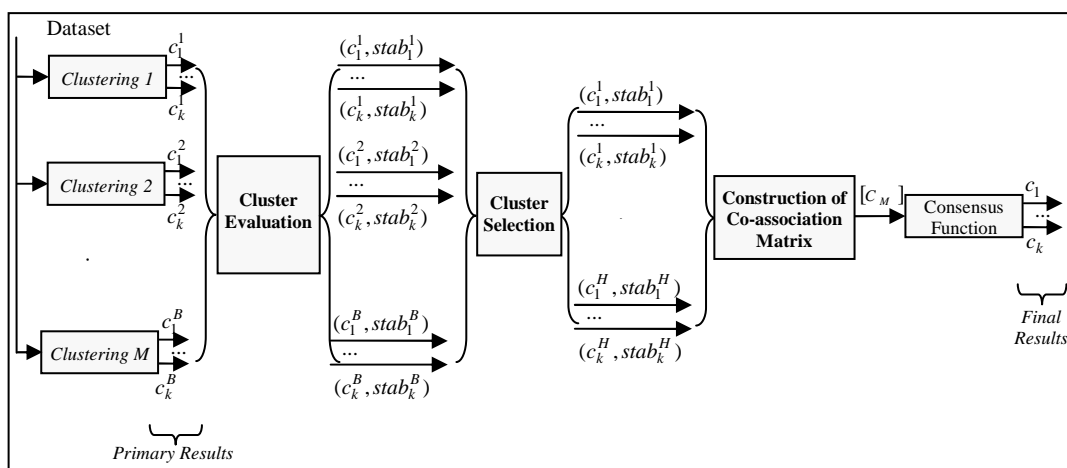


Fig. 1. Proposed clustering ensemble framework.

3 Proposed Clustering Ensemble Method

In this section, the proposed clustering ensemble method is first briefly outlined, and then its phases are described in further detail in the subsequent subsections.

The main idea underlying the proposed clustering ensemble method is the utilization of a subset of the best performing primary clusters in the ensemble instead of all of them. It seems that not every cluster is a good-quality cluster. Hence, in this method only clusters with stability greater than a specified threshold are chosen to participate in the combination. The cluster selection is done based on cluster stability that is defined according to NMI.

Fig. 1 depicts our proposed clustering ensemble framework. First, a set of B primary partitionings is provided using k -means (and hierarchical linkage clustering methods) to create the necessary diversity for an ensemble. Next, the AAPMM is computed for all clusters of each resultant clustering output. A detailed description of the manner in which the AAPMM is computed is given in the next sections. After that, a subset of the most stable clusters is selected to participate in the final decision committee. This is done simply by applying a threshold over the AAPMM value of each cluster. In the next step, the selected clusters construct the co-association matrix. Several methods of combining the primary results have been proposed [12], [14], [24]. Here, the difference is the absence of some clusters in primary partitionings. Since the original EAC method [12] cannot truly identify the pairwise similarity while there is only a subset of clusters, in this paper, a new method to construct the co-association matrix called EEAC is presented. Finally, the average-link method is employed to extract the final clusters from this matrix.

3.1 Cluster Evaluation Using APMM Criteria

A stable cluster is one that has a high likelihood of recurrence across consecutive applications of a clustering algorithm. Stable clusters are usually preferable, since they are robust with respect to minor changes in the dataset [17].

Now, let us assume that we want to compute the stability of cluster C_i . In our proposed method, first a set of partitionings over a sampled dataset (called the reference set) is provided. In this notation, D is resampled data and $P(D)$ is a partitioning over D . Now, the problem is “How many times is the cluster C_i repeated in the reference partitionings?” The problem is formally denoted by $NMI(C_i, P(D))$, which signifies the normalized mutual information between cluster C_i and a reference partitioning $P(D)$. Most previous works only compared a partitioning with another partitioning [24]. However, the stability measure used in [17] evaluates the similarity between a cluster and a partitioning by transforming cluster C_i to partitioning form and then employing common partitioning-to-partitioning measures. To illustrate this method, let $P_1 = P^a = \{C_i, D/C_i\}$ be a partitioning with two clusters, where D/C_i denotes the set of data points in D that are not in C_i .

We can then compute a second partitioning $P_2 = P^b = \{C^*, D/C^*\}$, where C^* denotes the union of all “positive” clusters in $P(D)$ and others are in D/C^* . A cluster C_j in $P(D)$ is positive if more than half of its data points are in C_i . Now, define $NMI(C_i, P(D))$ by $NMI(P^a, P^b)$, which is calculated as follows [11]:

$$NMI(P^a, P^b) = \frac{-2 \sum_{i=1}^{k_a} \sum_{j=1}^{k_b} n_{ij}^{ab} \log \left(\frac{n_{ij}^{ab} \cdot n}{n_i^a \cdot n_j^b} \right)}{\sum_{i=1}^{k_a} n_i^a \log \left(\frac{n_i^a}{n} \right) + \sum_{j=1}^{k_b} n_j^b \log \left(\frac{n_j^b}{n} \right)} \quad (1)$$

where n is the total number of samples; n_{ij}^{ab} denotes the number of shared patterns between clusters $C_i^a \in P^a$ and $C_j^b \in P^b$; n_i^a is the number of patterns in cluster i of partitioning a ; and n_j^b is the number of patterns in cluster j of partitioning b .

This computation is done between cluster C_i and all partitionings available in the reference set. Fig. 2 depicts this method.

NMI_i in Fig. 2 shows the stability of cluster C_i with respect to the i -th partitioning in the reference set. The total stability of cluster C_i is defined as follows:

$$Stability(C_i) = \frac{1}{M} \sum_{i=1}^M NMI_i \quad (2)$$

where M is the number of partitionings available in the reference set. This procedure is applied to each cluster of every primary partitioning.

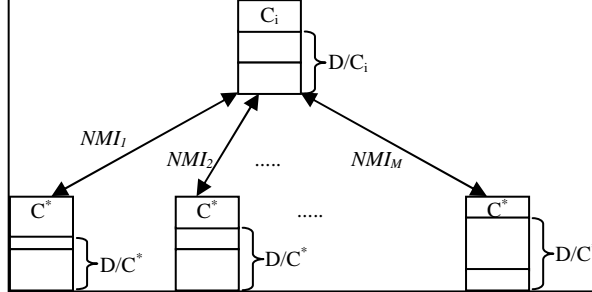


Fig. 2. Computing the stability of cluster C_i .

This approach as an evaluation of cluster stability has some drawbacks that can appear under certain circumstances. In this section, a drawback of this procedure is revealed, and a novel asymmetric measure, called APMM, is proposed to handle this drawback.

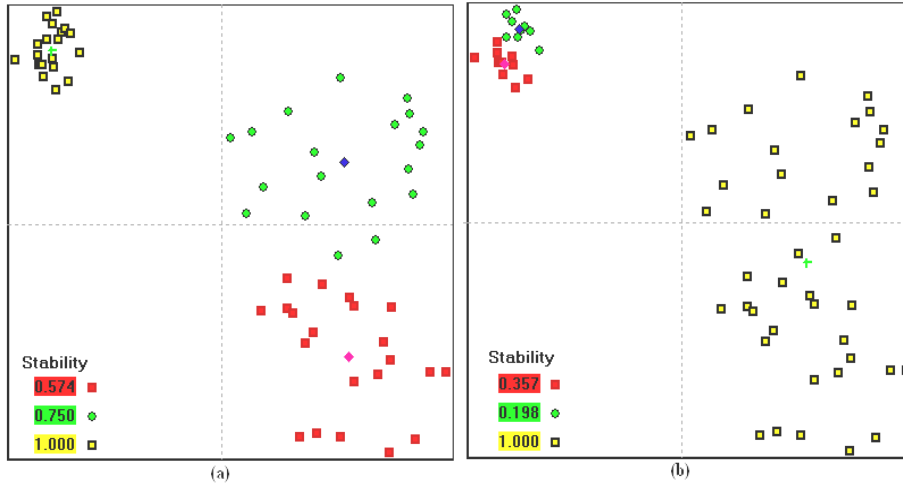


Fig. 3. Two primary partitionings with $k=3$: (a) True clustering (b) Spurious clustering.

Fig. 3 shows two primary partitionings along with the evaluated stability of each cluster and its computation process. In this example, k -means is applied as the base clustering algorithm and the true number of clusters, $k=3$, is fed to the k -means algorithm. For this example, the number of partitionings in the reference set was set to 40. In 36 of the partitionings, the result was similar to Fig. 3(a), while in the other 4, the top left cluster was divided into two clusters, as in Fig. 3(b). Fig. 3(a) shows a true clustering. Since the well-separated cluster in the top left corner is repeated several times (90% of cases) in partitionings of the reference set, it has to acquire a large stability value (but not equal to 1), however it acquires the stability value of 1. The two clusters on the right hand side of Fig. 3(a) are relatively jointed and sometimes they are not recognized in the reference set as well; they acquire rather low stability values. Fig. 3(b) shows a spurious clustering in which the two clusters on the right are incorrectly merged. Since a fixed number of clusters are forced to the base algorithm, the top left cluster is consequently divided into two clusters. Here, the weakness of the stability measure significantly emerges. Although it is obvious that this partitioning and the corresponding big cluster on the right rarely appear in the reference set (10% of cases), the stability of the big cluster on the right is evaluated as being equal to 1. Since the NMI is a symmetric criterion, the stability of the top left cluster in Fig. 3(a) is exactly equal to that of the big cluster on the right in Fig. 3(b); however, they are repeated 90% and 10%, respectively. In other word, when two clusters are complements of each other, their stabilities are always equal. This drawback occurs when the number of positive clusters in the partitioning of the reference set considered is greater than 1. It also means that when cluster C^* is obtained by merging two or more clusters, the undesirable stability value results are obtained.

Here, we propose a new criterion that can solve this problem. Assume that the problem is evaluating the APMM criterion for a cluster C_i in Fig. 4(a) with respect to clustering obtained in Fig. 4(b). The main idea in this method is to eliminate the symmetrical property that inherently exists in the NMI equation. In this approach, except for cluster C_i , all the clusters in P^a are taken out. Furthermore, all the clusters in P^b that are not included in the samples of cluster C_i are eliminated. In the next step, the other samples that are not in C_i of P^a are removed from the clusters in P^b (from the clusters that contain some of these samples). This process is depicted in Fig. 5.

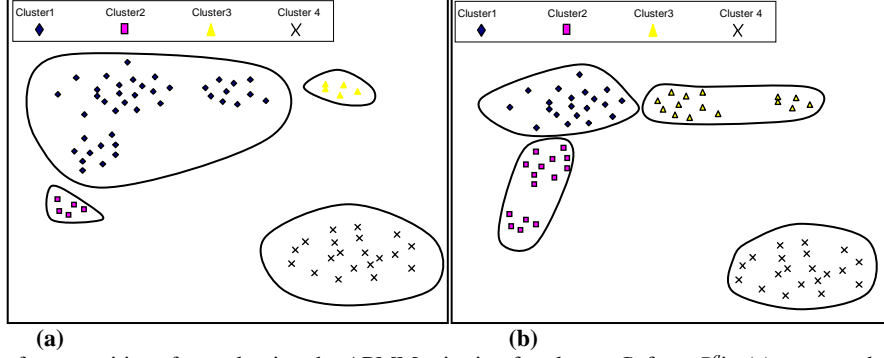


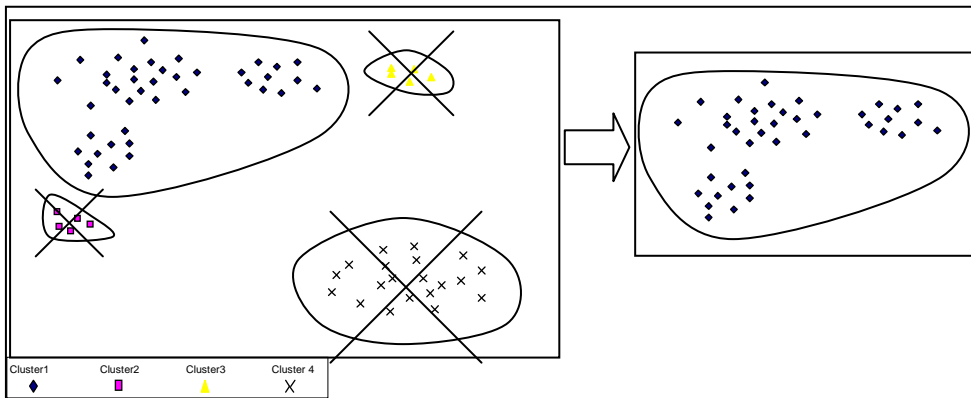
Fig.4.Examples of two partitions for evaluating the APMM criterion for cluster C_i from P^a in (a) compared to P^b in (b).

Now, the entropy between the clusters remaining in partitionings P^a and P^b is computed as depicted in Fig. 6. Considering that the other samples involved are eliminated, this criterion is not symmetric.

All the previous works are based on the NMI definition in equation 1. Even to evaluate the occurrence of a cluster in a partitioning, the problem is modified in some way to become a problem that involves comparison between two partitionings so that the NMI equation can be used. However, in this study, the problem is not changed to correspond to the definition of NMI; instead, the NMI equation is modified so that the occurrence of a cluster in a partitioning is computed. This is done by evaluating the entropy between the considered cluster and other pseudo clusters in the corresponding partitioning. The APMM is defined between a cluster C_i from P^a and the partitioning P^{b*} from P^b , as in the equation below:

$$APMM(C_i^a, P^{b*}) = \frac{-2n_i^a \log\left(\frac{n}{n_i^a}\right)}{n_i^a \log\left(\frac{n_i^a}{n}\right) + \sum_{j=1}^{k_{b^*}} n_j^{b^*} \log\left(\frac{n_j^{b^*}}{n}\right)} \quad (3)$$

where n is the number of the samples available in the dataset, and k_{b^*} is the number of clusters in P^{b*} .



(a)

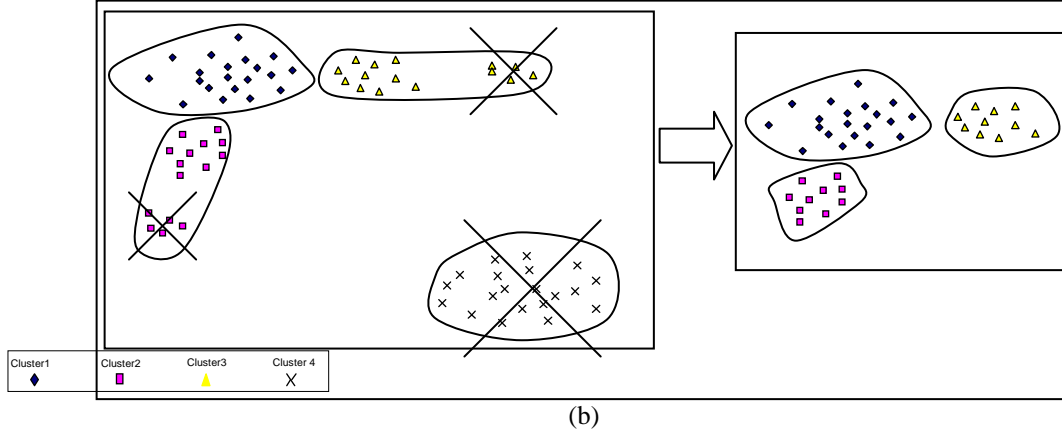


Fig. 5. Provision of data to evaluate the APMM criterion: (a) Deletion of all clusters except C_1 from P^a . (b) Derivation of P^{b*} from P^b , the corresponding samples of C_1 in P^b .

We propose AAPMM as a measure of stability for a primary cluster C_i compared to the partitionings available in the reference set defined in the following equation:

$$AAPMM(C_i) = \frac{1}{M} \sum_{j=1}^M APMM(C_i, P_j^{b*}) \quad (4)$$

where P_j^{b*} is from the j -th partitioning in the reference set.

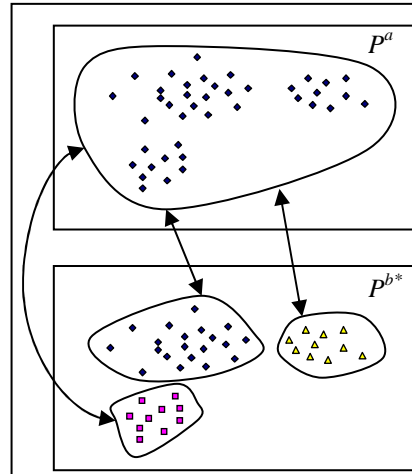


Fig. 6. Computing the entropy between cluster C_1 from P^a and P^{b*} from P^b .

3.2 Consensus Function

The final activity in our clustering ensemble framework is aggregation of the selected clusters. One way of doing this is to consider the selected clusters as inputs of the Hyper-Graph Partitioning Algorithm (HGPA), Meta-Clustering Algorithm (MCLA), and Cluster-based Similarity Partitioning Algorithm (CSPA) [24]. The outputs of these algorithms are the final partitioning, which is also called consensus partitioning. For example, in Fig. 7, four partitionings π_1 – π_4 are extracted from a simple dataset with 12 data points and 2 real clusters by the k -means clustering algorithm. The k parameters of k -means clustering are set to 3, 4, 2, and 2, respectively. These partitionings are broken into the 11 clusters depicted in Fig. 8. The clusters are then served as input to the HGPA, MCLA, and CSPA algorithms.

The second way to extract the final partitioning from the selected clusters is to consider clusters as a new data space, and employ a clustering algorithm, such as fuzzy k -means, to partition the mapped data. As is already known about the fuzzy k -means clustering algorithm, it assigns each data point to all clusters with

different membership values. To extract the final partitioning from the output of a fuzzy k -means algorithm as consensus function, each data point was assigned to the cluster with the highest membership value. For more information about fuzzy k -means algorithm, readers should refer to [40]. Let us again consider the example in Fig. 7. The partitionings are broken into the 11 clusters depicted in Fig. 8, according to the previous method. The clusters in Fig. 7 are considered the mapped data in a new feature space and the fuzzy k -means algorithm extracts consensus partitioning out of them.

	π_1	π_2	π_3	π_4
\mathbf{x}_1	2	1	2	1
\mathbf{x}_2	2	1	2	1
\mathbf{x}_3	2	1	2	2
\mathbf{x}_4	2	1	2	2
\mathbf{x}_5	3	2	1	2
\mathbf{x}_6	3	2	2	2
\mathbf{x}_7	1	3	1	1
\mathbf{x}_8	1	3	1	1
\mathbf{x}_9	1	4	1	2
\mathbf{x}_{10}	1	3	1	1
\mathbf{x}_{11}	1	4	1	1
\mathbf{x}_{12}	1	4	1	1

Fig.7. Four partitionings, π_1 – π_4 , extracted from a simple dataset with 12 data points and 2 real clusters using the k -means clustering algorithm. The k parameter in k -means was set to 3, 4, 2, and 2, respectively.

An alternative way to reach the consensus partitioning is to use co-association based methods. In these methods, the selected clusters are first used to construct a co-association matrix. In the EAC method, the primary results from re-sampled data are accumulated in an $n \times n$ co-association matrix. Each entry in the matrix is computed from equation 5.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}
\mathbf{x}_1	0	1	0	1	0	0	0	0	1	1	0
\mathbf{x}_2	0	1	0	1	0	0	0	0	1	1	0
\mathbf{x}_3	0	1	0	1	0	0	0	0	1	0	1
\mathbf{x}_4	0	1	0	1	0	0	0	0	1	0	1
\mathbf{x}_5	0	0	1	0	1	0	0	1	0	0	1
\mathbf{x}_6	0	0	1	0	1	0	0	0	1	0	1
\mathbf{x}_7	1	0	0	0	0	1	0	1	0	1	0
\mathbf{x}_8	1	0	0	0	0	1	0	1	0	1	0
\mathbf{x}_9	1	0	0	0	0	0	1	1	0	0	1
\mathbf{x}_{10}	1	0	0	0	0	1	0	1	0	1	0
\mathbf{x}_{11}	1	0	0	0	0	0	1	1	0	1	0
\mathbf{x}_{12}	1	0	0	0	0	0	1	1	0	1	0

Fig.8. The clusters extracted from the partitionings in Fig. 7.

$$C(i, j) = \frac{n_{i,j}}{m_{i,j}} \quad (5)$$

where n_{ij} counts the number of clusters shared by objects with indices i and j in the B ensemble partitionings. In addition, m_{ij} is the number of partitionings where the pair of objects is simultaneously present. There is only a fraction of all the primary clusters available, after thresholding. Therefore, the common EAC method cannot truly recognize the pairwise similarity for computing a proper co-association matrix. In our novel method (Extended Evidence Accumulation Clustering, or EEAC), each entry of the co-association matrix is computed using equation 6:

$$C(i, j) = \frac{n_{i,j}}{\max(n_i, n_j)} \quad (6)$$

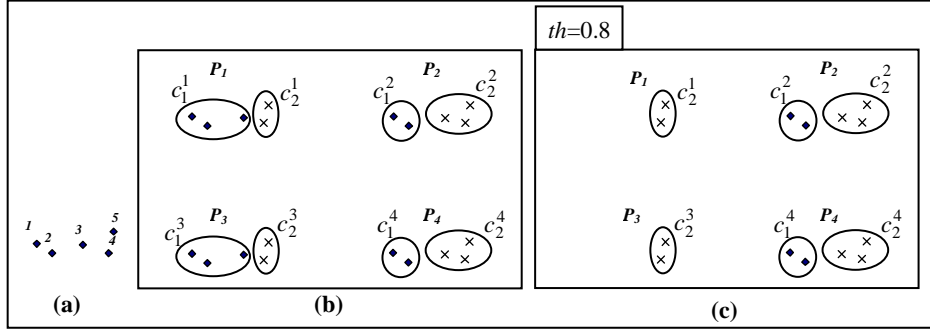


Fig.9. Computing the co-association matrix using the EEAC method: (a) Data samples. (b) Four primary clusterings. (c) Remaining clusters after applying threshold, $th=0.8$.

where n_i and n_j are, respectively, the number of occurrences of the i -th and j -th data points in the selected or remaining (after stability thresholding) clusters. In addition, n_{ij} represents the number of remaining clusters that contain both data points indexed by i and j . To further explain, consider this example. Let us assume that we have five samples (Fig. 9(a)), and that four primary clusterings are applied (Fig. 9(b)).

In addition, suppose that the stability of the clusters in Fig. 9(b) is as follows:

$$Stability(c_2^1) = Stability(c_2^3) = 1$$

$$Stability(c_1^2) = Stability(c_1^4) = 1$$

$$Stability(c_2^2) = Stability(c_2^4) = 0.82$$

$$Stability(c_1^1) = Stability(c_1^3) = 0.55$$

By choosing $th=0.8$ the first clusters from P1 and P3 are deleted (Fig. 9(c)). According to equation 4, each entry in the co-association matrix is as follows:

$$C(1,2) = \frac{2}{\max(2,2)} = \frac{2}{2} = 1$$

$$C(1,3) = C(2,3) = \frac{0}{\max(2,2)} = \frac{0}{2} = 0$$

$$C(3,4) = C(3,5) = \frac{2}{\max(2,4)} = \frac{2}{4} = 0.5$$

$$C(4,5) = \frac{4}{\max(4,4)} = \frac{4}{4} = 1$$

In Figs. 9(a)–9(c), the data points may be “tracked” by their geometrical arrangement. Example: in computing $C(3,4)$, note that points 3 and 4 are both in cluster 2 of partitionings P2 and P4, so that numerator $n_{34}=2$; in addition, note that $n_3=2$, since point 3 is only in cluster 2 of P2 and P4, but $n_4=4$ since point 4 is not only in these clusters, but also in cluster 2 of P1 and P3. Before and after threshold, the co-association matrix is given by equations 7 and 8, respectively:

$$C_{before} = \begin{bmatrix} 1 & 1 & 0.5 & 0 & 0 \\ 1 & 1 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 1 & 1 \\ 0 & 0 & 0.5 & 1 & 1 \end{bmatrix} \quad (7)$$

In this matrix, the third object can be considered to be both clusters with an equal probability of 50%. The stability measure adds some information to this matrix by applying the threshold.

$$C_{after} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 1 & 1 \\ 0 & 0 & 0.5 & 1 & 1 \end{bmatrix} \quad (8)$$

Comparing these two matrices and also considering the stability values, we can see that the omission of unstable clusters improves the co-association matrix. By eliminating the unstable cluster, samples {1, 2, 3}, which are spuriously created by primary clusterings, are neglected.

After computing the co-association matrix using the EEAC method, a consensus function is employed to extract the final clusters from the matrix. Here, the average-link method is generally used for this task. In the above example, the final two clusters are {{1, 2}, and {3, 4, 5}}.

Because different values of k are chosen in the base partitionings, there is a high probability that even the most stable cluster is divided into two or more unstable ones. Consequently, the base partitionings are diverse enough to present a good coverage of clusters with different stabilities to be chosen. This means that our proposed algorithm is fed a large number of options including different clusters with different properties and stability values. As a result, the probability of missing any sample in the retained clusters is very low.

Although samples are rarely overlooked, the method does not guarantee that all the observations will be included in at least one of the retained clusters. Therefore, post-processing in order to label the missing samples is inevitable. The post-processing carried out includes the assignment of each missing sample to the nearest cluster center in the feature space using the Euclidian distance.

4 Experimental Results

This section discusses experimental results that validate how our proposed framework, which uses the APMM measure, can compete with its rivals and its position in relation to the other methods. Evaluation metrics are discussed in the first subsection, while the details of the datasets used are given in the subsequent section. The settings used in the experiments are then given and, finally, experimental results are presented.

4.1 Evaluation Metrics

After producing the consensus partitioning, the most important challenge may be the assessment phase. The NMI measure by itself is generally considered an evaluation metric for a partitioning. Therefore, a partitioning is evaluated here by measuring the NMI between the consensus partitioning and the real labels of the dataset.

An alternative that can be used to evaluate a partitioning is accuracy metric, provided that the number of clusters and their true assignments are known. To compute the final performance of a clustering algorithm in terms of accuracy, one can first re-label the clusters obtained in such a way that they have maximal matching with the ground true labels, and then count the percentage of truly classified samples. Therefore, the error rate can be determined after solving the correspondence problem between the labels of the derived and the known clusters. The Hungarian algorithm [26], which has been shown to efficiently solve this label correspondence problem, can be employed to solve the minimal weight bipartite matching problem.

F-measure is another measure for evaluating a partitioning. It considers both precision and recall to compute the score. Precision p is the number of correct results divided by the number of all results returned, while recall r is the number of correct results divided by the number of results that should have been returned. The F-measure can be interpreted as a weighted average of the precision and the recall, where an F-measure reaches its best value at 1 and its worst score at 0 [29]. The F-measure used in this paper is computed based on equation 7.

$$FM(P^a, P^b) = \max_{\pi} \sum_{i=1}^{k_a} \frac{2 \times n_i^a \times \left(\frac{n_{i\pi(i)}^{ab}}{n_i^a} \times \frac{n_{i\pi(i)}^{ab}}{n_{\pi(i)}^b} \right)}{n \times \left(\frac{n_{i\pi(i)}^{ab}}{n_i^a} + \frac{n_{i\pi(i)}^{ab}}{n_{\pi(i)}^b} \right)} \quad (9)$$

where n_{ij}^{ab} again denotes the number of shared patterns between clusters $C_i^a \in P^a$ and $C_j^b \in P^b$; n_i^a is the number of patterns in cluster i of partitioning a ; n_j^b is the number of patterns in cluster j of partitioning b ; and π is a permutation of $\{1, \dots, k_b\}$ for the relabeling of P^b .

It is worth mentioning that in all the experiments, the NMI, F-measure, and accuracy were scaled in the range [0–100] for consistency, thus they are always reported in the range [0–100].

Table 1. Brief summary of the datasets used

	<i>Dataset Name</i>	<i># of Classes</i>	<i># of Features</i>	<i># of Samples</i>
1	Breast-Cancer*	2	9	683

2	Iris*	3	4	150
3	Bupa*	2	6	345
4	SAHeart*	2	9	462
5	Ionosphere	2	34	351
6	Glass*	6	9	214
7	HalfRing	2	2	400
8	Galaxy*	7	4	323
9	Yeast*	10	8	1484
10	Wine*	3	13	178

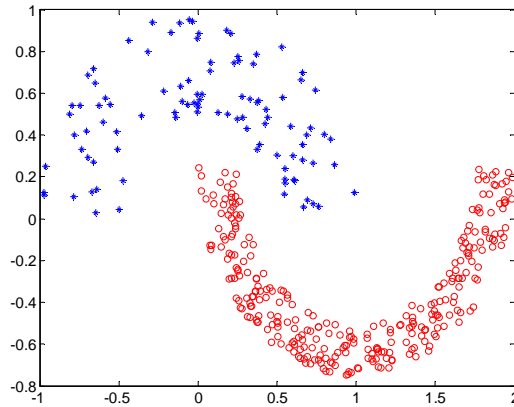


Fig.10. HalfRing dataset

4.2 Datasets

Our proposed method was examined over nine different standard datasets and one artificial dataset. Since a large variety in the datasets used can more effectively validate the obtained results, the datasets were chosen in such a way that they had adequate diversity in the number of true classes, features, and samples. The datasets used are briefly summarized in Table 1. More information is available in [27].

Note that the datasets that are marked with an asterisk (*) in Table 1 were normalized. The experiments were conducted over the normalized features in the starred dataset. This means that each feature was normalized with a mean of 0 and variance of 1, $N(0, 1)$. The artificial HalfRing dataset is depicted in Fig. 10. It is considered one of the most challenging datasets for clustering algorithms [32].

4.3 Experimental Settings

To be more general and fairer, all experiments were averaged over 10 independent runs. In all the experiments, there were 120 independent partitionings obtained by 120 independent runs of the k -means clustering algorithm with different initialized seed points and different values for the k parameter, ranging from k to $2k$.

We should mention here that there are no rigid rules for determining the exact ways in which diversity in the primary results can be achieved. Therefore, we set our diversity generation policies empirically. In other words, one can change these parameters and achieve different results. However, we believe that any differences in the results will not be so significant as to surpass the results attained in this paper.

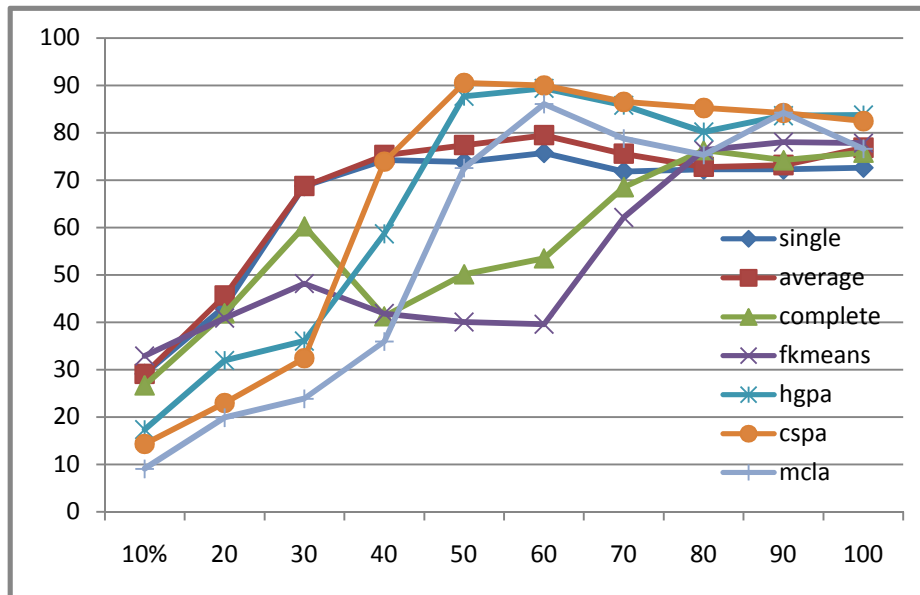
After selecting a subset of clusters, to extract the final partitioning out of them, the real number of clusters, i.e., the third column in Table 1, is served by the consensus functions.

4.4 Experimental Results

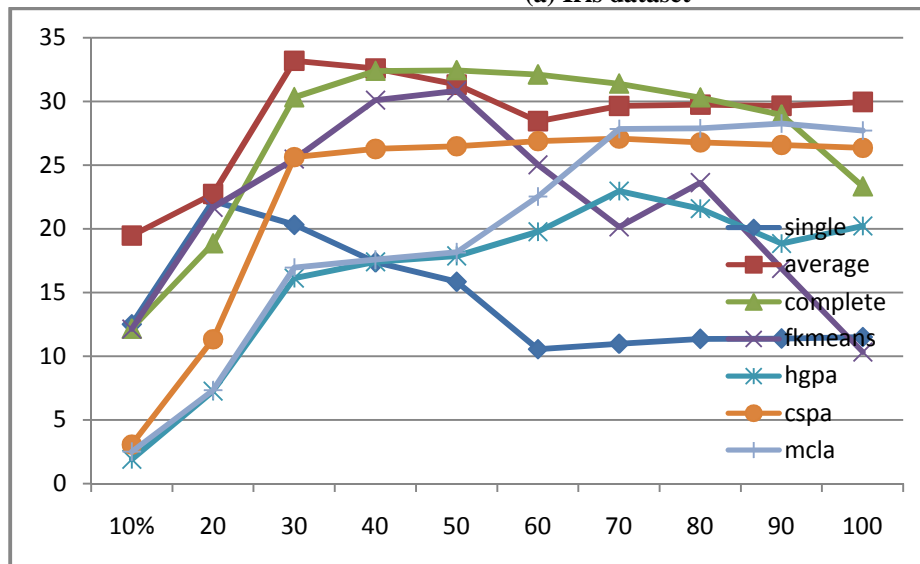
The NMI values in terms of employing different ratios from the most stable clusters in the final ensemble over different datasets are depicted in Fig. 11. As is obvious for the Iris dataset (Fig. 11(a)), the NMI value increased as the ratio of the most stable clusters participating in the final ensemble increased up to 50%.

However, after 50%, it did not increase, and even decreased in some cases. The same figure for the Yeast dataset is depicted in Fig. 11(b). In this case, the NMI value rose for the first 30%, after which it abated in almost all cases.

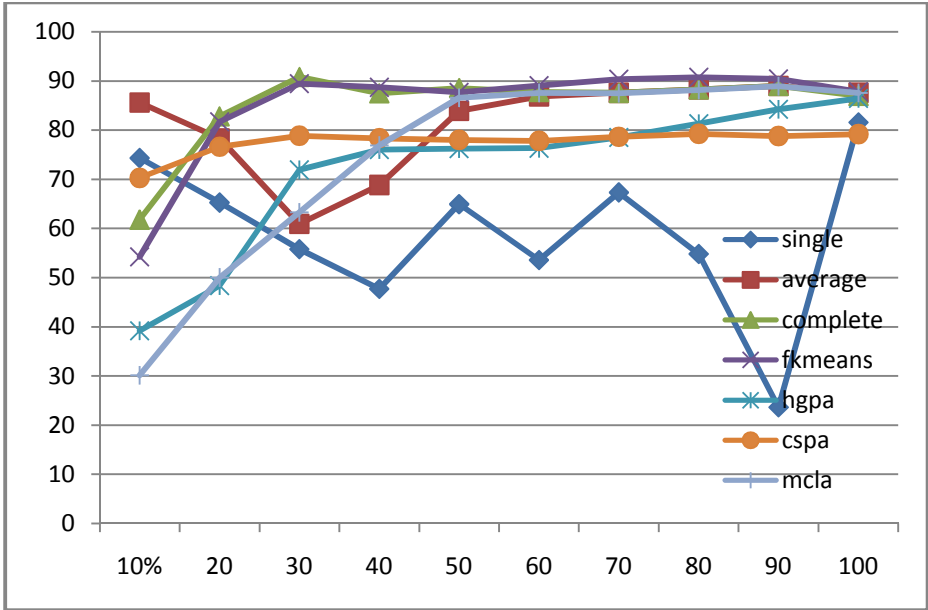
Like the results for the Iris dataset, for both the Wine dataset in Fig 11(c) and the Galaxy dataset in Fig 11(d), the NMI value increased up to 50% almost in all cases. However, after 50%, it did not increase, and even decreased in some cases. For the Wine dataset, the single-linkage aggregator was an exception. In the single-linkage aggregator, the best ensemble occurred when all clusters participated in the ensemble, but its peak was not comparable with others. However, fuzzy k -means stood out from the other consensus functions in the Galaxy test case. In the fuzzy k -means aggregator, the best ensemble occurred when 10% of the most stable clusters contributed to the ensemble. However, its peak did not give a result that was comparable with that of the others.



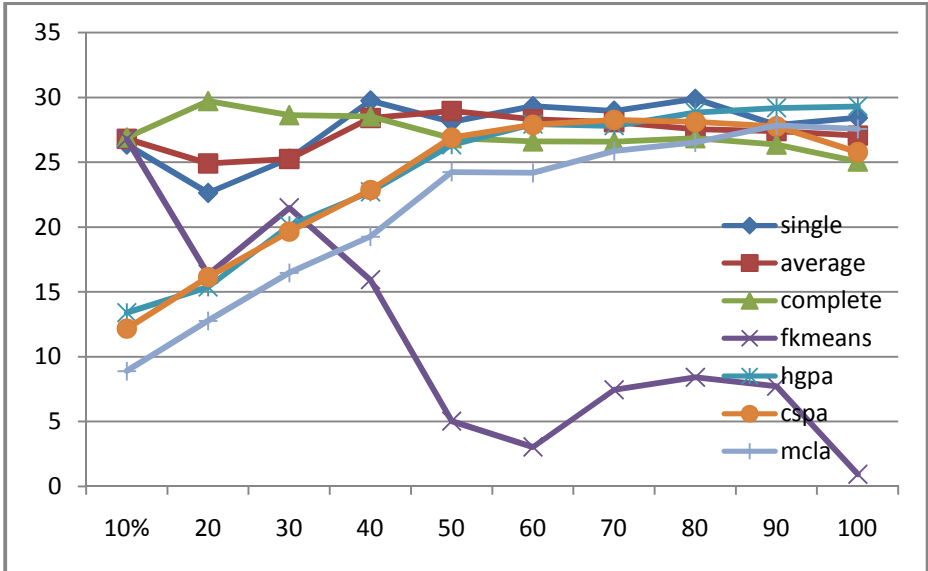
(a) Iris dataset



(b) Yeast dataset



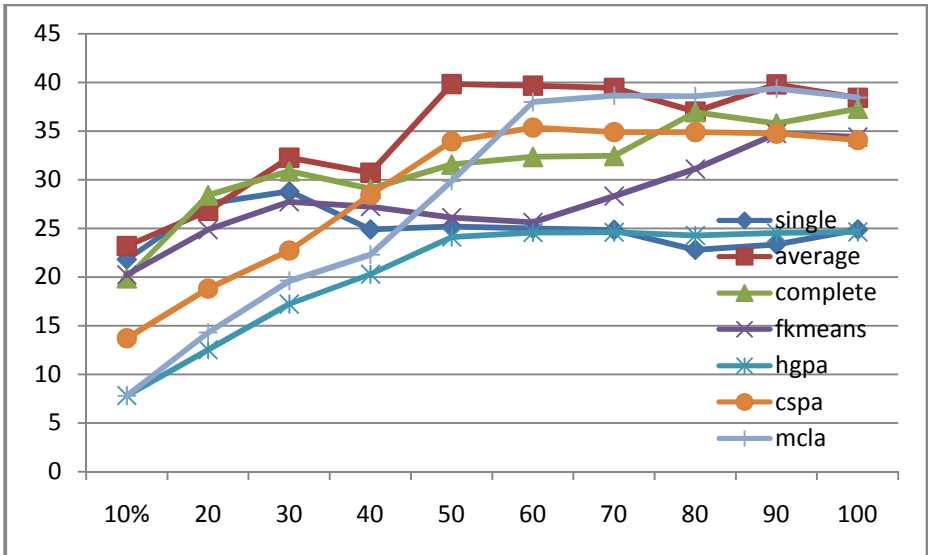
(c) Wine dataset



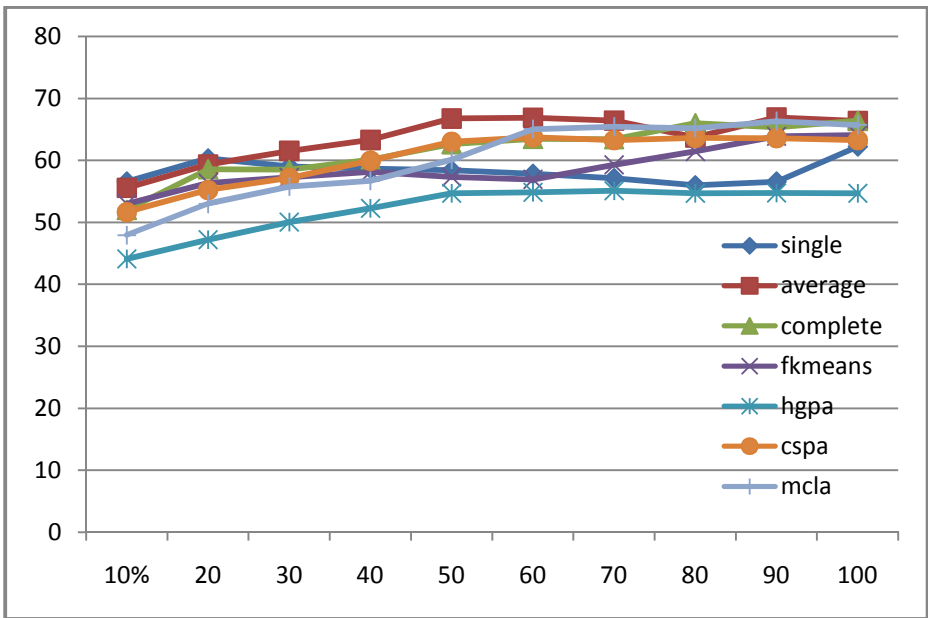
(d) Galaxy dataset

Fig.11. Effect of selection size on the performance in terms of NMI for (a) Iris, (b) Yeast, (c) Wine, and (d) Galaxy datasets. The horizontal axis represents the rate of stable clusters that were selected. The vertical axis represents the NMI value for the datasets.

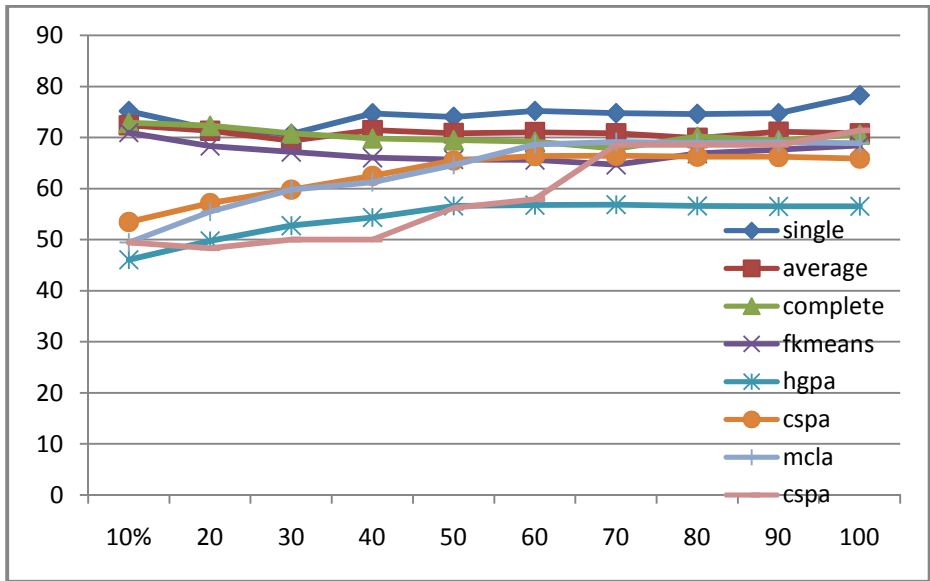
The averaged NMI value in terms of choosing different ratios from the most stable clusters in the final ensemble over all ten datasets is shown in Fig. 12. In Fig. 12(a), it is easy to see that in most cases the NMI value increased as the rate of participants grew to 50%. However, after 50%, it did not increase significantly; and even decreased in some cases. However, our proposed method peaked when 50% of the most stable clusters were taken into the final ensemble and by employing the average-linkage algorithm as the consensus function along with the EEAC method. The accuracy result of Fig. 12(b) confirms the NMI result (of Fig. 12(a)). In addition, the F-measure outcome of Fig. 12(c) also confirms both previous results, except in the case where the single-linkage aggregator is superior to the average-linkage aggregator. However, in this case, the average-linkage aggregator is the second best method.



(a) NMI



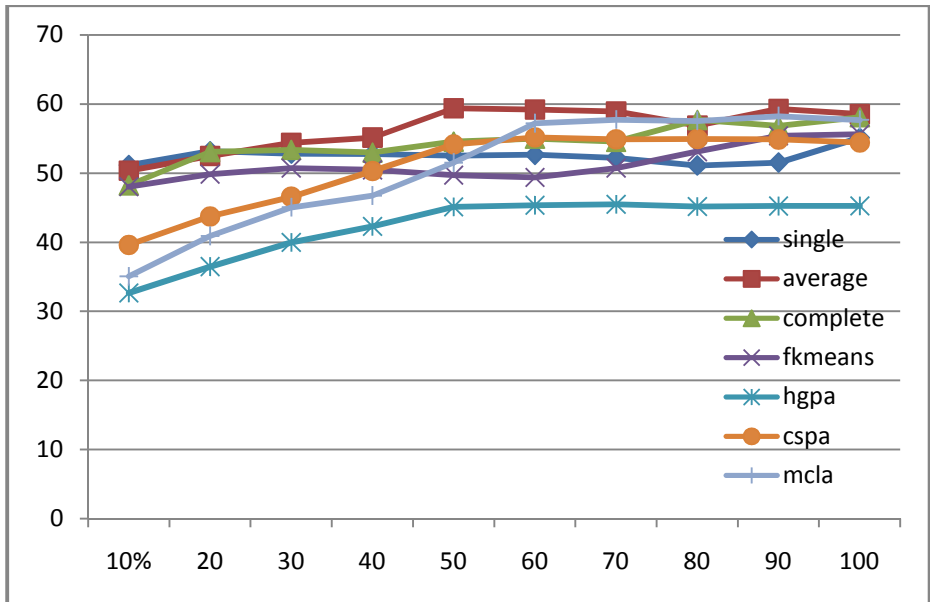
(b) Accuracy



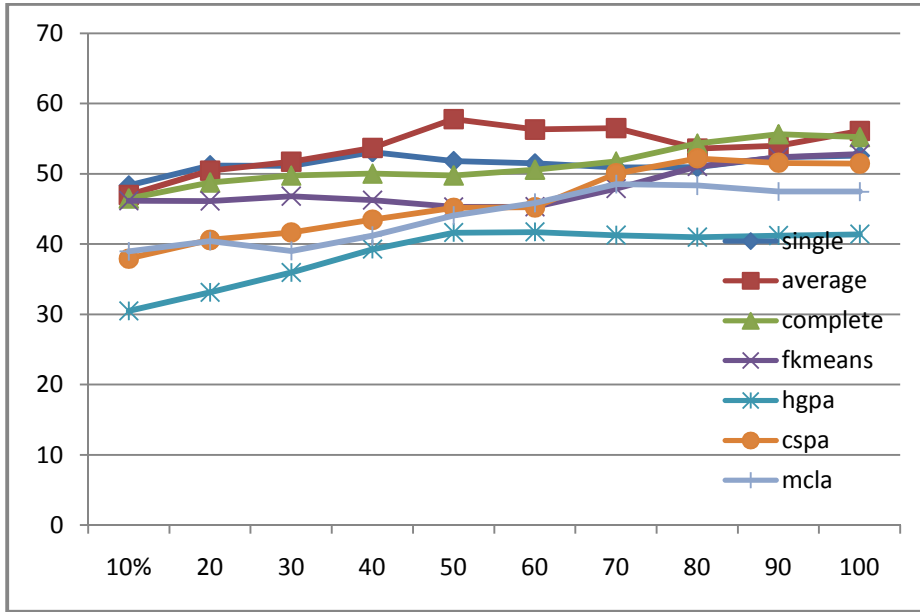
(c) F-measure

Fig.12. Effect of selection size on performance in terms of (a) NMI, (b) Accuracy, and (c) F-measure overall datasets. The horizontal axis represents the rate of stable clusters that were selected. The vertical axis represents the averaged F-measure over all ten datasets in Table 1.

Fig. 13 illustrates the effect of selection size on performance in terms of averaged NMI, accuracy, and F-measure over all ten datasets. Supporting the results of Fig. 12, this figure confirms that the average-linkage aggregator is superior to all aggregators. Therefore, we can conclude that using the average-linkage algorithm as aggregator along with the EEAC method is the best option for consensus function. Furthermore, it verifies that choosing the selection ratio of the most stable clusters to construct the final ensemble at 50% hits the peak. By comparing the average of the vertical axes of Figs.13(a) and 13(b), we can infer that employing APMM as the measure to evaluate a cluster generally yields better consensus partitioning.



(a) APMM



(b) MAX

Fig.13. Effect of selection size on performance in terms of averaging all three validation measures: NMI, accuracy, and F-measure overall datasets. In (a) the stability criterion is APMM, while in (b) it is MAX [28]. The horizontal axis represents the rate of stable clusters selected. The vertical axis represents the averaged vertical axes of Figs.12(a), 12(b), and 12(c).

A more general understanding over the three cluster evaluation criteria used—NMI, MAX, and APMM—can be captured from Table 2. The average-linkage algorithm is used as aggregator along with the EEAC method in Table 2. As can be seen, except for the eighth dataset, APMM and MAX completely outperformed NMI. In addition, the APMM-based cluster ensemble is slightly superior to the MAX-based cluster ensemble. As a result, in this paper, we conclude that using APMM for cluster evaluation is a better option than using its rivals.

Table 2. Accuracy of consensus partitioning produced by cluster selection based on NMI and MAX measures [28] and APMM.

Evaluation Measure	Dataset Number										Average
	1*	2	3*	4*	5	6*	7	8*	9*	10	
NMI	95.73	76.13	54.33	63.36	70.60	50.76	74.48	31.27	42.93	69.38	62.90
MAX [28]	96.49	84.87	57.42	63.87	67.75	49.35	74.55	29.85	51.27	70.00	64.54
APMM	95.56	90.00	55.07	63.87	70.66	50.79	74.00	30.65	53.10	70.23	65.39

The first four columns in Table 3 show the results of some base clustering algorithms. The results show that although each of the base algorithms can obtain a superior result over a specific dataset, it does not perform well over other datasets. For example, as shown in Table3, the *k*-means algorithm gives a more superior clustering result over the Wine dataset than that of the linkage methods. However, it has a lower performance than that of the linkage methods in the case of the Bupa dataset. Moreover, the complete linkage gives an efficient performance for the Breast-Cancer dataset compared with that of the others; however, this is not so for all datasets.

The last five columns show the performance of some ensemble methods compared with that of our proposed method. A glance at the last five columns and comparison with the first four columns reveals that the ensemble methods do better than simple base algorithms in the case of performance and stability against different datasets. For all cluster ensemble algorithms, the 90% sampling from dataset is used to create diversity in the primary results. Sub-sampling (without replacement) is used as the sampling method. In addition, the random initialization of the seed points of the *k*-means algorithm helps them to be more diverse. Finally, the average-linkage algorithm is employed as consensus function for deriving the final clusters from the co-association matrix.

The first column of the ensemble methods gives the results of an ensemble of 100 k -means that were combined by the EAC method. The second column from ensemble methods is the full ensemble, which uses several clustering algorithms to generate the primary results. Here, 70 k -means with the above-mentioned parameters in addition to 30 linkage methods provide the primary results. Since different runs of a specific linkage method always yield the same result, there is a limitation for using them as the base clustering algorithms. Here, forcing a set of different number of clusters, $k \pm 2$, is used to create diversity, in which k is the true number of clusters.

Table 3. Experimental results

	Simple Methods (%)				Ensemble Methods (%)				
Dataset	Single Linkage	Average Linkage	Complete Linkage	k -means	k -means Ensemble	Full Ensemble	Cluster Selection by NMI Measure	Cluster Selection by MAX Measure [28]	Cluster Selection by APMM Measure
Wine*	37.64	38.76	83.71	96.63	96.63	97.08	97.75	98.27	98.31
Breast-C*	65.15	70.13	94.73	95.37	95.46	95.10	95.75	96.49	95.56
Yeast*	34.38	35.11	38.91	40.20	45.46	47.17	42.93	51.27	53.10
Glass*	36.45	37.85	40.65	45.28	47.01	47.83	50.76	49.35	50.79
Bupa*	57.68	57.10	55.94	54.64	54.49	55.83	54.33	57.42	55.07

The last three columns show the results of ensemble methods that use only a subset of primary clusters. The first one shows the results of an ensemble that uses the traditional NMI-based stability for cluster validation [17]. The second one is our proposed clustering ensemble method, which uses the MAX measure [28] for cluster validation. The last column is our proposed clustering ensemble method, which uses the APMM measure for cluster validation. The selected clusters in the co-association matrix in all three methods are accumulated using the proposed EEAC method. Finally, the average-linkage algorithm is applied over the co-association matrix to extract final clusters. The primary clustering results are provided similar to the full ensemble. We used the idea of cluster ranking and thresholding to select the best subset of base clusters. More precisely, the threshold that is used for cluster selection is determined adaptively. First, it is adjusted to 95%. If less than 10% of the samples are absent in the selected clusters, the threshold is then reduced to 90%. This procedure (reducing 5% from the threshold value) is continued until the selected clusters include more than 90% of the samples. The results in the last three columns verify that although these approaches use a subset of the primary clusters, they usually outperform the full ensemble. In addition, comparing the last three columns shows the power of APMM-based stability in comparison with the NMI-and MAX-based stability. Examinations by 10 independent runs over different datasets robustly show the quality of the APMM measure with respect to the NMI and MAX measures.

5 Conclusions

In this paper, a new clustering ensemble method that is based on a subset of total primary spurious clusters is proposed. Since the qualities of primary clusters are not equal and the presence of some of them can even result in poor performance, we present a method to select a subset of more effective clusters. A common cluster validity criterion that is required to derive this subset is based on normalized mutual information. In this paper, some drawbacks of this criterion is discussed and an alternative criterion—the Alizadeh–Parvin–Moshki–Minaei (APMM) criterion. The results of experiments conducted indicate that the APMM criterion generally performs slightly better than the NMI (and also the MAX) criterion; moreover, it significantly outperforms the NMI criterion in the case of synthetic datasets. Because of the symmetry concealed in the NMI criterion and also in NMI-based stability, decreased performance results are obtained whenever symmetry also appears in the dataset.

After selecting a subset of clusters, several consensus functions are employed to extract consensus partitioning. This paper empirically showed that the most effective consensus function results from the average-linkage hierarchical clustering algorithm. Another novel contribution of this paper is a method for constructing the co-association matrix where some of the clusters and some of the corresponding samples do not exist in partitionings. This new method is called Extended Evidence Accumulation Clustering (EEAC). The empirical studies conducted over several datasets robustly show that the quality of our proposed method is usually better than that of other methods.

This paper finally concludes that using APMM for cluster evaluation is the best option. We can also conclude that using the average-linkage algorithm as an aggregator along with the EEAC method is the best option for consensus function. The performance of consensus partitioning peaks when the ratio of the most stable clusters to participate in the final ensemble is 50%.

As future work, the effect of data sampling on the performance of the final ensemble should be investigated. The subsampling and bootstrap methods in addition to the adaptive cluster ensemble [32] approach can be incorporated into our framework. We believe that enforcing diversity in primary results can expand the range of possible choices for selection, as it appears that selection from a wider range of possible clusters can improve the final results obtained from our proposed method. Furthermore, in this paper, we investigate and conclude that using the most stable clusters improves the final partitionings. However, to maintain the stability of the final ensemble, one can investigate other aspects of clusters such as diversity. Moreover, similar to many other research in the field of cluster ensemble [9-15,32-35,38,39], we assumed that our data set is standard and free of noise and missing values. In the future, the effect of noise and missing values of the data on the APMM evaluation measure should be investigated. All of these examinations will be carried out in our future studies.

Acknowledgements

This work was supported in part by the Research Institute for ICT – ITRC grant program. We would like to appreciate the ITRC for this support.

References

1. Ayad, H.G. and Kamel, M.S.: Cumulative Voting Consensus Method for Partitions with a Variable Number of Clusters. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, VOL. 30, NO. 1, 160-173 (2008)
2. Baumgartner, R., Somorjai, R., Summers, R., Richter, W., Ryner, L., and Jarmasz, M.: Resampling as a Cluster Validation Technique in fMRI, *JOURNAL OF MAGNETIC RESONANCE IMAGING* 11: pp. 228–231, (2000)
3. Ben-Hur, A., Elisseeff, A., and Guyon, I.: A stability based method for discovering structure in clustered data. *Pacific Symposium on Biocomputing*, vol. 7, pp. 6-17 (2002)
4. Brandsma, T. and Buishand, T.A.: Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. *Hydrology and Earth System Sciences* 2, pp. 195-209 (1998)
5. Breckenridge, J.: Replicating cluster analysis: Method, consistency and validity. *Multivariate Behavioral research* (1989)
6. Das, A.K. and Sil, J.: Cluster Validation using Splitting and Merging Technique. *Int. Conf. on Computational Intelligence and Multimedia Applications, ICCIMA* (2007)
7. Davison, A.C., Hinkley, D.V., and Young, G.A.: Recent developments in bootstrap methodology. *Statistical Science* 18, pp. 141-157 (2003)
8. Estivill-Castro, V. and Yang, J.: Cluster Validity Using Support Vector Machines. *DaWaK 2003, LNCS 2737*, pp. 244–256 (2003)
9. Faceli, K., Marcilio, C.P., and Souto, D.: Multi-objective Clustering Ensemble. *Proceedings of the Sixth International Conference on Hybrid Intelligent Systems* (2006)
10. Fern, X.Z. and Lin, W.: Cluster Ensemble Selection. *SIAM International Conference on Data Mining* (2008)
11. Fred, A. and Jain, A.K.: Combining Multiple Clusterings Using Evidence Accumulation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(6):835–850 (2005)
12. Fred, A. and Jain, A.K.: Data Clustering Using Evidence Accumulation. *Intl. Conf. on Pattern Recognition, ICPR02, Quebec City*, pp. 276 – 280 (2002)
13. Fred, A. and Jain, A.K.: Learning Pairwise Similarity for Data Clustering. *Int. Conf. on Pattern Recognition*, (2006)
14. Fred, A. and Lourenco, A.: Cluster Ensemble Methods: from Single Clusterings to Combined Solutions. *Studies in Computational Intelligence (SCI)*, 126, 3–30 (2008)
15. Fridlyand, J. and Dudoit, S.: Applications of resampling methods to estimate the number of clusters and to improve the accuracy of a clustering method. *Stat. Berkeley Tech Report. No. 600* (2001)
16. Inokuchi, R., Nakamura, T., and Miyamoto, S.: Kernelized Cluster Validity Measures and Application to Evaluation of Different Clustering Algorithms. *IEEE Int. Conf. on Fuzzy Systems, Canada, July 16-21* (2006)
17. Law, M.H.C., Topchy, A.P., and Jain, A.K.: Multiobjective data clustering. *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pp. 424–430 (2004)
18. Lange, T., Roth, V., Braun, M.L., and Buhmann, J.M.: Stability-based validation of clustering solutions. *Neural Computation*, 16(6):1299–1323 (2004)

19. Minaei-Bidgoli, B., Topchy, A., and Punch, W.F.: Ensembles of Partitions via Data Resampling. Intl. Conf. on Information Technology, ITCC 04, Las Vegas, (2004)
20. Möller, U. and Radke, D.: Performance of data resampling methods based on clustering. *Intelligent Data Analysis* 10(2) (2006)
21. Rakhlin, A. and Caponnetto, A.: Stability of k-means clustering. In *Advances in Neural Information Processing Systems* 19, MIT Press, Cambridge, MA (2007)
22. Roth, V. and Lange, T.: Feature Selection in Clustering Problems. In *Advances in Neural Information Processing Systems* (2004)
23. Roth, V., Lange, T., Braun, M., and Buhmann, J.: A Resampling Approach to Cluster Validation, Intl. Conf. on Computational Statistics, COMPSTAT (2002)
24. Strehl, A. and Ghosh, J.: Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(Dec):583–617 (2002)
25. Xie, X.L. and Beni, G.: A Validity measure for Fuzzy Clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.13, No.4, pp. 841–846 (1991)
26. Munkres, J.: Algorithms for the Assignment and Transportation Problems, *Journal of the Society for Industrial and Applied Mathematics*, 5(1): 32-38 (1957)
27. Newman C.B.D.J., Hettich S. and Merz C., UCI repository of machine learning databases, (1998) <http://www.ics.uci.edu/~mllearn/MLSummary.html>
28. Alizadeh H., Minaei-Bidgoli B., Parvin H.: A New Criterion for Clusters Validation. *Artificial Intelligence Applications and Innovations (AIAI 2011)*, IFIP, Springer, Heidelberg, Part I, pp. 240-246 (2011)
29. Marxer, R., Holonowicz, P., Purwins, H., and Hazan, A.: Dynamical hierarchical self-organization of harmonic, motivic, and pitch categories. In *Music, Brain and Cognition. Part 2: Models of Sound and Cognition*, held at NIPS (2007)
30. Gullo F., Domeniconi C., Tagarelli A. Enhancing single-objective projective clustering ensembles. In: *proceedings of the IEEE international conference on data mining (ICDM)*, pp 833–838(2010)
31. Gullo F., Domeniconi C., Tagarelli A.: Projective clustering ensembles, *Data Mining and Knowledge Discovery*, online, (2012)
32. Minaei-Bidgoli B., Parvin H., Alinejad-Rokny H., Alizadeh H., and Punch W.F.: Effects of resampling method and adaptation on clustering ensemble efficacy, *Artificial Intelligence Review*, in press (2011)
33. Domeniconi C., Al-Razgan M. Weighted cluster ensembles: methods and analysis. In: *ACM Trans Knowl Disc Data (TKDD)*, 2(4) (2009)
34. Ghosh J., Acharya A. Cluster ensembles. *Wiley interdisciplinary reviews. Data Min Knowl Disc* 1(4):305–315 (2011)
35. Azimi J., Fern X., “Adaptive Cluster Ensemble Selection”, *IJCAI* (2009)
36. Kuncheva L. I., *Combining Pattern Classifiers, Methods and Algorithms*. New York: Wiley, (2004)
37. Sun S., Local within-class accuracies for weighting individual outputs in multiple classifier systems. *Pattern Recognition Letters* 31: 119-124, (2010)
38. Parvin H., Alizadeh H., Classifier Ensemble Based Class Weighting, *American Journal of Scientific Research, AJSR* (19)84-90, (2011)
39. Parvin H., Minaei B., Alizadeh H., A Heuristic Classifier Ensemble for Huge Datasets, *AMT*: 29-38, (2011)
40. Bezdek J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, (1981)