Journal of King Saud University – Computer and Information Sciences 34 (2022) 3828–3842

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement

Teng Li [a,*], Amin Rezaeipanah [b,*], ElSayed M. Tag El Din [c]

[a] Artificial Intelligence and Big Data College, Chongqing College of Electronic Engineering, Chongqing 401331, China
[b] Department of Computer Engineering, Persian Gulf University, Bushehr, Iran
[c] Electrical Engineering Department, Faculty of Engineering and Technology, Future University in Egypt, New Cairo 11845, Egypt

ARTICLE INFO

Article history:
Received 10 February 2022
Revised 2 April 2022
Accepted 14 April 2022
Available online 22 April 2022

Keywords:
Hierarchical clustering
Meta-clusters
Ensemble clustering
Model selection
Similarity measurement
Clusters clustering

ABSTRACT

The advent of architectures such as the Internet of Things (IoT) has led to the dramatic growth of data and the production of big data. Managing this often-unlabeled data is a big challenge for the real world. Hierarchical Clustering (HC) is recognized as an efficient unsupervised approach to unlabeled data analysis. In data mining, HC is a mechanism for grouping data at different scales by creating a dendrogram. One of the most common HC methods is Agglomerative Hierarchical Clustering (AHC) in which clusters are created bottom-up. In addition, ensemble clustering approaches are used today in complex problems due to the weakness of individual clustering methods. Accordingly, we propose a clustering framework using AHC methods based on ensemble approaches, which includes the clusters clustering technique and a novel similarity measurement. The proposed algorithm is a Meta-Clustering Ensemble scheme based on Model Selection (MCEMS). MCEMS uses the bi-weighting policy to solve the model selection associated problem to improve ensemble clustering. Specifically, multiple AHC individual methods cluster the data from different aspects to form the primary clusters. According to the results of different methods, the similarity between the instances is calculated using a novel similarity measurement. The MCEMS scheme involves the creation of meta-clusters by re-clustering of primary clusters. After clusters clustering, the number of optimal clusters is determined by merging similar clusters and considering a threshold. Finally, the similarity of the instances to the meta-clusters is calculated and each instance is assigned to the meta-cluster with the highest similarity to form the final clusters. Simulations have been performed on some datasets from the UCI repository to evaluate MCEMS scheme compared to state-of-the-art algorithms. Extensive experiments clearly prove the superiority of MCEMS over HMM, DSPA and WHAC algorithms based on Wilcoxon test and Cophenetic correlation coefficient.

1. Introduction

Clustering is one of the most common methods of unsupervised learning in which data is segmented based on the similarity of instances (Hamidi et al., 2019). Clustering involves assigning a set of instances to subgroups called clusters, so that the members of each cluster are very similar to each other and are very different from the members of other clusters. Clustering has many applications for solving real-world problems such as community identification, anomaly detection, pattern recognition, and image processing that can be used in the variety of situations (Aslanpour et al., 2018; Berahmand et al., 2021). In addition, clustering is very important in data analysis because it performs segmentation among unlabeled data (Shakarami et al., 2021).

In general, the clustering process by different algorithms has the same basis, but there are differences in the method of measuring similarity (or distance) as well as determining the cluster of instances (Berahmand et al., 2022). These differences have led to the development of different schemes and frameworks of clustering algorithms (Beauchemin, 2015; Shakarami et al., 2021). In general, there are four main classes of clustering algorithms including Partitional Clustering (PC), Hierarchical Clus-

* Corresponding authors.
E-mail addresses: liteng@cqcet.edu.cn (T. Li), rezaeipanah@pgu.ac.ir (A. Rezaeipanah), elsayed.tageldin@fue.edu.eg (ElSayed M. Tag El Din).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2022.04.010
1319-1578/

tering (HC), Model-based Clustering (MC), Grid-based Clustering (GC) and Density-based Clustering (DC), as shown in Fig. 1.

In PC, also known as flat clustering, the number of clusters is already known and each instance belongs to only one cluster (Xu and Wunsch, 2005; Rezaeipanah et al., 2021). The K-means algorithm is one of the most well-known and popular PC algorithms. PC algorithms are usually applied based on the optimization of an objective function (Berkhin, 2006; Boongoen and Iam-On, 2018). For example, K-means is applied to minimize the sum of the squared distance between the instances and the cluster's centroid. HC shows the result of clustering at any level of distance by a dendrogram or cluster tree in which the levels are hierarchical (Yang and Jiang, 2018). These algorithms first consider each instance as a cluster and then merge two clusters with the highest similarity (or shortest distance) to create a new cluster. This step is repeated until there is only one cluster with all the instances. HC uses linkage methods to measure the distance between two clusters (Rai and Singh, 2010). One of the most popular algorithms for HC is DIANA (DIvisie ANAlysis) (Struyf et al., 1997). In DC, instances are identified with high density and placed in a cluster. These algorithms are also able to detect clusters of convex form, however, they are incapable of clustering concave data (Xu and Wunsch, 2005; Shahidinejad et al., 2021). One of the most popular algorithms in this category is DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Ester et al., 1996). In GC, a multi-resolution grid data structure is used for the clustering process (Nagpal and Mann, 2011). The GC divides the data into several regions called cells and performs clustering within them, where these speeds up processing time independent of the size of the data. In MC, first the main model of clustering is retrieved from the data and then the allocation of instances to the clusters is defined by the model (Boongoen and Iam-On, 2018). The purpose of these algorithms is to estimate the statistical distribution parameters along with the hidden variable, which is introduced as the cluster label in the model. A common metric for estimating model parameters is maximum likelihood.

Basically, each of the mentioned algorithms has different limitations for data clustering (Mojarad et al., 2021). Most PC-based methods suffer from problems of high sensitivity to phase initialization, outliers and noise, overlapping of instances, and poor cluster descriptors. Disadvantages of most HC-based methods include inability to change after merging/splitting and not interpreting cluster descriptors. On the other hand, most DC-based methods are sensitive to the value of input parameters. Although GC-based methods have the advantage of fastest processing time, however they suffer from the problem of determining the density threshold or grid size. MC-based methods provide robust clustering approaches but are not resistant to noise and outliers. Based on these problems, it can be concluded that choosing a clustering method for a particular data is very difficult. Hence, ensemble clustering methods are introduced which are a combination of several individual clustering methods (Sasirekha and Baby, 2013; Mojarad et al., 2021). These methods combine several basic clustering methods with the aim of achieving higher accuracy and quality. However, the selection of ensemble components is very important because it can lead to the formation of higher quality clusters with less time complexity (Khedairia and Khadir, 2022). In addition, ensemble clustering schemes are often based on PC methods, and HC-based ensemble clustering has been less studied (Rai and Singh, 2010). For this reason, we address the problem of ensemble clustering with a focus on model selection.

Zhou et al. (2002) showed that model selection in combined learning can have better results than combining all models. The problem of model selection in ensemble clustering involves the selection of a subset of available individual models that can provide better results than using all sets for clustering data. Although there are many approaches in the literature to ensemble clustering, but AHC-based ensemble clustering algorithms still suffer from model selection problems, for example, similarity matrix calculation, initialization sensitivity, consensus function, and detecting the automatic number of clusters (Sasirekha and Baby, 2013).

This paper proposes a Meta-Clustering Ensemble scheme based on Model Selection (MCEMS). The MCEMS scheme has selected multiple AHC individual methods based on bi-weighting policy to improve the quality and diversity factors in ensemble clustering. MCEMS forms primary clusters based on selected models from AHC. We use a novel similarity measurement to calculate the similarity matrix between instances, where similarity is measured by observing all primary clusters. The ensemble clustering approach in MCEMS is done through the clusters clustering technique and

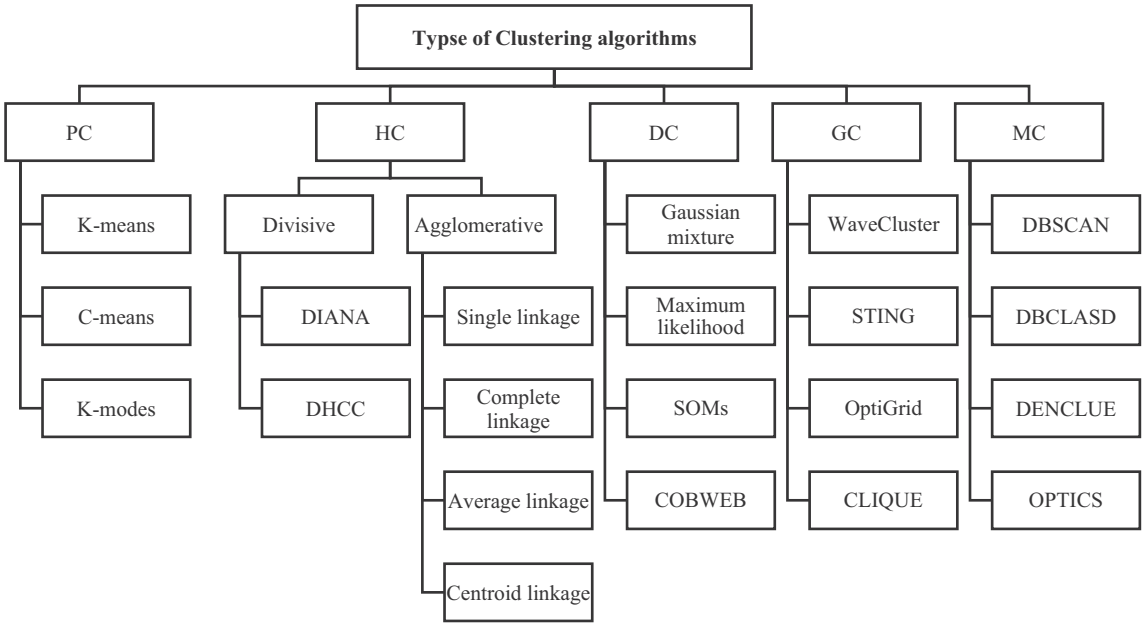

Fig. 1. Types of clustering algorithms.

the creation of meta-clusters as a consensus function. Each meta-cluster consists of one or more clusters and is created by re-clustering the primary clusters. The final clusters are formed by assigning each instance to a meta-cluster with the highest similarity, in which the meta-clusters contain clustering knowledge of several AHC methods. Specifically, MCEMS automatically determines the number of clusters by merging similar clusters and considering a threshold.

The main contributions in this paper can be highlighted as follows.

- Solving the problem of model selection through the proposed bi-weighting policy to improve the quality and diversity factors in ensemble clustering
- Development of a meta-clustering ensemble scheme using the clusters clustering technique
- Proposing a novel similarity measurement focusing on linkage-based AHC clustering methods

The rest of this paper is organized as follows. Section 2 describes agglomerative hierarchical clustering. Section 3 describes several ensemble clustering approaches through literature review. The proposed scheme to improve ensemble clustering is highlighted in Section 4. Experiments and discussion results are performed to evaluate the MCEMS scheme in Section 5. Finally, the conclusions are drawn with the future direction of the research in Section 6.

2. Agglomerative hierarchical clustering

In statistics and data mining, HC analysis is a strategy of cluster analysis that seeks to create a hierarchy of clusters (Murtagh and Contreras, 2012). In general, HC involves two methods: agglomerative (bottom-up) and divisive (top-down) (Jiang et al., 2021). In Agglomerative Hierarchical Clustering (AHC) methods, each instance is considered as a cluster and then the clusters are merged to create larger clusters (Fernández and Gómez, 2008). This continues until all the clusters are merged into one large cluster that contains all the instances. In Divisive Hierarchical Clustering (DHC) methods, all instances first belong to one cluster and then the large cluster is hierarchically and recursively divided into smaller clusters until each cluster has a single instance (Guénoche et al., 1991). Basically, DHC is more complex than AHC, because in DHC a flat method is required as a "subroutine" to divide each cluster as long as each instance has its own singleton cluster (Nagpal and Mann, 2011). Although DHC is also represented by a dendrogram, but this method is less commonly used (Nagpal and Mann, 2011). This is because there are limitations to dividing one cluster into other sub-clusters. In addition, common AHC methods are easy to implement because they start with only one instance in each cluster and gradually join the nearest clusters (Davidson and Ravi, 2005). Hence, this paper proposes an AHC-based method for clustering.

In statistics and data mining, HC analysis is a strategy of cluster analysis that seeks to create a hierarchy of clusters (Murtagh and Contreras, 2012). Basically, there are two types of HC analysis strategies including agglomerative and divisive. DHC is more complex than AHC, because in DHC a flat method is required as a "subroutine" to divide each cluster as long as each instance has its own singleton cluster (Nagpal and Mann, 2011). Although DHC is also represented by a dendrogram, but this method is less commonly used. This is because there are limitations to dividing one cluster into other sub-clusters. In addition, common AHC methods are easy to implement because they start with only one instance in each cluster and gradually join the nearest clusters (Davidson and Ravi, 2005). Hence, this paper proposes an AHC-based method for clustering.

Algorithm 1. Pseudocode of the agglomerative hierarchical clustering (AHC)

Input: Dataset S with n instances as $S = \{x_1, x_2, \cdots, x_n\}$ and and cluster distance function $D(c_i, c_j)$.

Output: Black partition $Dendrogram_k$, for each $1 \le k \le n$.

$c_i = \{x_i\}, \quad \forall i = 1, 2, \cdots, n$
for $k = n$ down to 1 **do**
  $Dendrogram_k = \{c_1, c_2, \cdots, c_k\}$
  $d(i,j) = D(c_i, c_j), \quad \forall i,j = 1, 2, \cdots, k$
  Let $(a,b) = argmin_{(i,j)}\{D(c_i, c_j) : 1 \le i < j \le k+1\}$
  $c_a = Join(c_a, c_b)$
  $Remove(c_b)$
end

AHC is known as a bottom-up approach that does not need to determine the number of clusters (Nagpal and Mann, 2011). The bottom-up methods first consider each instance as a single cluster and then sequentially combine the cluster pairs until all the clusters are merged into a single cluster that contains all the instances. This means that there are fewer clusters at higher levels (Sasirekha and Baby, 2013). The pseudo-code of standard AHC is shown in Algorithm 1.

In this regard, ensemble clustering includes several single clustering methods that can create more accurate final clusters by combining the results. Let $S = \{x_1, x_2, \cdots, x_n\}$ be a dataset with n instances, where x_i is an instance of dataset S. According to ensemble clustering, $p_q = \{c_1^q, c_2^q, \cdots, c_k^q\}$ is the result of the q-th individual clustering method with the number of k clusters, where c_j^q refers to the instances assigned to j-th cluster of q-th method. Ensemble clustering can be done by m different methods such as $\Gamma = \{p_1, p_2, \cdots, p_m\}$, where the consensus of the results of different methods for creating final clusters can be done based on F similarity measurement. This paper intends to use an ensemble approach according to AHC specifications for clustering, where F is introduced as a consensual similarity measure to calculate similarities between instances.

Given that AHC is used as the basic clustering method in the proposed algorithm, we will have an overview of these AHC-related methods in the following. So far, many methods have been introduced based on AHC where the difference between them is in how the distance is defined. Here, four known methods including single linkage, average linkage, centroid linkage and complete linkage are examined (Nagpal and Mann, 2011).

- Single linkage: This method defines the distance between two clusters based on the minimum distance of each pair of members from the two clusters. According to this definition, two clusters in each step are combined with the smallest linkage distance.
- Average linkage: This method defines the distance between two clusters based on the average distance between all members in these clusters. According to this definition, two clusters in each step are combined with the smallest average linkage distance.
- Centroid linkage: This method defines the distance between two clusters based on the average vector distance of these clusters. According to this definition, two clusters in each step are combined with the smallest center distance.

- Complete linkage: This method defines the distance between two clusters based on the maximum distance between all members in these clusters. According to this definition, two clusters in each step are combined with the smallest complete linkage distance.

As it turns out, the difference between these methods is in how the distance is defined. Suppose $c_a = \langle x_1, x_2, \cdots, x_r \rangle$ and $c_b = \langle y_1, y_2, \cdots, y_l \rangle$ are the details of the instances assigned to cluster 1 and cluster 2, respectively. Also, let $d(x_i, y_j)$ be the distance between the instance x_i and y_j. Accordingly, Table 1 shows the mathematical form of distance functions for different linkage methods based on c_a and c_b.

For better clarity, the following is an example of different linkage methods. In these methods, clustering begins with calculating the distance between each instance pair. This process creates a distance symmetric matrix with a diagonal of zero. An example of different linkage methods is given in Fig. 2 for single linkage, average linkage, centroid linkage and complete linkage. Here, in addition to the distance matrix, the final dendrogram for each method is given.

Basically, none of the individual clustering methods work best for all datasets (Rai and Singh, 2010). Ensemble clustering approaches can produce better results by combining several individual clustering methods and using the strengths of each method (Berikov and Pestunov, 2017). Generally, ensemble clustering architecture consists of two main phases, including the creation of primary clusters and combining the results with a consensus function, where maintaining diversity and improving quality is an important challenge. Fig. 3 shows the architecture of the ensemble clustering methods. Ensemble clustering has often been investigated for PC-based methods, and the use of HC methods has received less attention (Jiang et al., 2021; Shi et al., 2021).

3. Literature review

In recent years, ensemble clustering methods as well as AHC have been widely used in various studies to enhance the quality of clustering results (Yu et al., 2018; Jiang et al., 2021; Shi et al., 2021). In the following, we will briefly review some of these researches. Fred and Jain (2005) introduced an algorithm for combining different base clusters called Evidence Accumulation Clustering (EAC). This algorithm uses a new similarity technique to combine multiple clusters. The EAC combines the results obtained with the correlation matrix through HC methods (for example, single linkage, average linkage and complete linkage). Li et al. (2007) proposed the HC ensemble approach based on normalized edges, which aims to improve the clustering quality of the co-association matrix. The authors use normalized edges to calculate the similarity between clusters. Hence, in the HC process, two clusters are combined with the maximum value of normalized edges. According to the analysis, the worst complexity of this approach is $O(n^2 \log n)$.

Table 1
Mathematical form of distances for different linkage methods.

Linkage method	Distance function	Description
Single linkage	$D(c_a, c_b) = \min_{i,j} d(x_i, y_j)$	This method combines two clusters with the shortest distance between members.
Average linkage	$D(c_a, c_b) = \frac{1}{k \times l} \sum_{i=1}^{r} \sum_{j=1}^{l} d(x_i, y_j)$	This method combines two clusters with the least average distance of all members.
Centroid linkage	$D(c_a, c_b) = d\left(\frac{\sum_{x \in c_a} x)}{r}, \frac{\sum_{y \in c_b} y)}{l}\right)$	This method combines two clusters with the smallest distance between the average vectors of the clusters center.
Complete linkage	$D(c_a, c_b) = \max_{i,j} d(x_i, y_j)$	This method combines two clusters with the longest distance members.

Mirzaei et al. (2008) proposed a new method for combination of HC. The authors converted the primary dendrograms into matrices and then aggregated them into a final matrix to obtain final clusters. In this method, the matrix summation operator is used to aggregate the primary matrices. Wang et al. (2009) proposed the Probability Accumulation (PA) algorithm for AHC. PA algorithm uses original clusters to determine the optimal cluster size. Here, the PA matrix is calculated for each cluster and then the correlation matrix is measured based on the average of all PA matrices. Finally, the final clustering is determined by applying the minimum spanning tree to the correlation matrix. Yi et al. (2012) proposed an ensemble clustering algorithm based on the matrix completion strategy. The authors first create a similarity matrix based on the observed data and then apply the matrix completion strategy to complete the similarity matrix. Finally, clustering is created by a spectral clustering method through the similarity matrix. Beauchemin (2015) proposed an approach to clustering using density-based similarity matrix construction that is implemented based on K-means. K-means based on sub-bagging method has been used to improve the accuracy of density estimation. This method has shown the same efficiency in experiments with the EAC.

Berikov and Pestunov (2017) introduced an ensemble clustering method based on weighted co-association matrices. This method uses a fitness function to measure the weight between instances and generate co-association matrices. Moreover, the authors use an AHC to create the final clustering based on the co-association matrices. Huang et al. (2017) proposed the Locally Weighted Meta-Clustering (LWMC) algorithm, which is a cluster-based algorithm. The LWMC uses the Jaccard coefficient to measure the weight of links between clusters. In addition, LWMC uses the normalized cut algorithm to generate meta-clusters, where each meta-cluster contains a set of clusters. In this method, clustering is achieved through weighted voting based on silhouette criteria. Huang et al. (2018) proposed an approach called Locally Weighted Graph Partitioning (LWGP) for clustering data at different scales. LWGP is applied based on bipartite graph partitioning, where the link weight between the instance and the clusters is measured according to the silhouette criterion.

Hamidi et al. (2019) introduced an ensemble clustering approach with graph similarity partitioning in which the number of clusters is automatically determined by graph cutting. The purpose of this method is to prune the similarity graph and automatically remove outlier clusters. In this method, first the main graph is partitioned and then the sub-graphs are combined to create meta-clusters. In order to calculate the weight between the two clusters, the authors use the average of the links between the two clusters based on the Jaccard coefficient, and finally generate the clustering based on the weight majority vote. Khedairia and Khadir (2022) introduced the Iterative Combining Clustering Method (ICCM). The ICCM first generates the base clusters and then forms the sub-clusters by voting between instances. In each iteration, the instances with the highest votes are assigned to the respective sub-clusters and the other instances are clustered in the next iterations. Finally, an AHC method clusters the subclusters to create the final clusters. Jafarzadegan et al. (2019) used Principal Component Analysis (PCA) to develop ensemble clustering. They combined the results of linkage-based methods of AHC (i.e., single linkage, average linkage, and centroid linkage) using PCA as an aggregation operator to create an ensemble clustering approach. Here, a meaningful relation is considered between all the elements

Distances matrix

	1	2	3	4	5	6
1	0	0	0	0	0	0
2	11	0	0	0	0	0
3	7	8	0	0	0	0
4	5	4	6	0	0	0
5	6	2	9	10	0	0
6	7	9	2	11	3	0

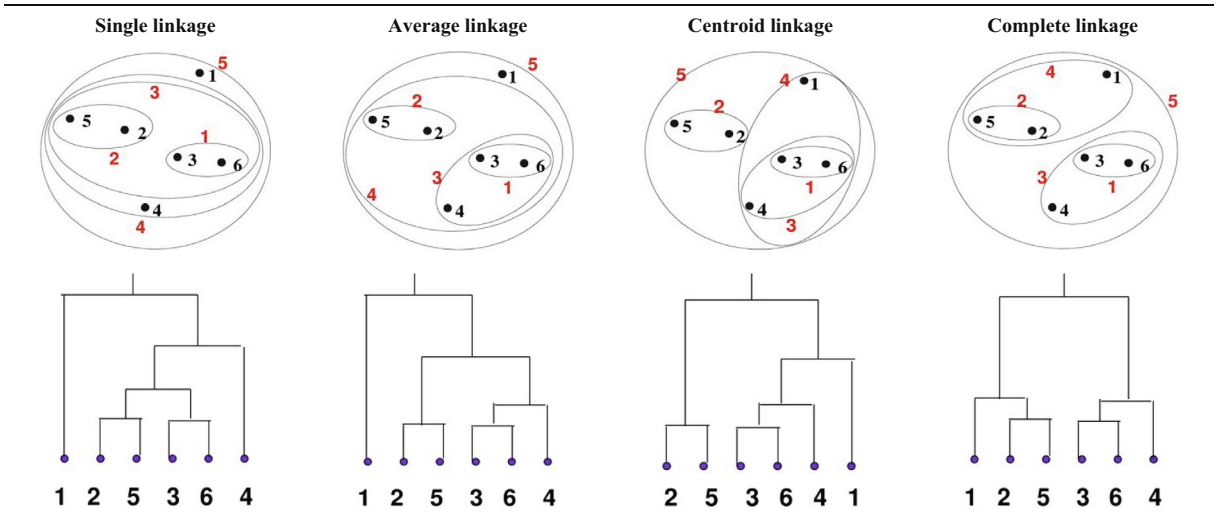

Fig. 2. An example of different linkage methods in AHC.

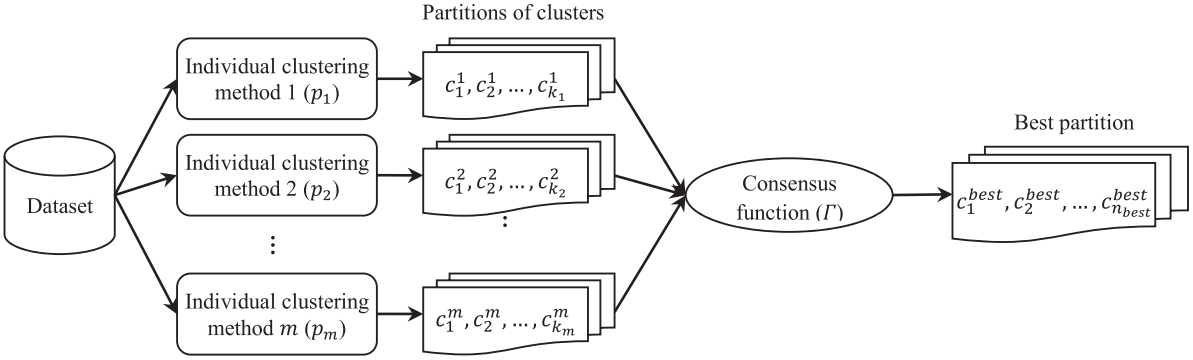

Fig. 3. Architecture of the ensemble clustering methods.

of the descriptor matrix, where the descriptor matrix is created from the base clustering methods. In this method, final clusters are created based on the extraction of a dendrogram from the descriptor matrix.

Niu et al. (2020) introduced an ensemble clustering method based on K-medoids that provides Locally Reliable Clustering. In this method, first a set of different clusters is created by K-medoids and then a weighted graph is made based on the clusters created. This graph can show the relationships between different clusters. Each node in this graph acts as a cluster and the similarity between the clusters is considered based on the weight of the link between them. Mojarad et al. (2021) proposed an ensemble clustering approach to model the behavior of inherited diseases. In this method, first several flat clustering algorithms are applied to create the primary clusters and then the similarity between each instance pair is calculated by an innovative similarity criterion. The final clusters are then formed based on the maximum similarities between the primary clusters and the merging of the clusters with the highest similarity. In addition, the authors use an uncertainty improvement algorithm to increase the quality of the final clusters. Pasupathi et al. (2021) used AHC methods for trend analysis in time series big data. The authors used AHC-based clustering and association rules extraction techniques in the application of road accidents. Here, the diagnosis of Paradigmatic Time Sequence (PTS) data in each cluster is performed using dynamic time warping. PTS can examine the details of different geographical areas and provide a cluster for data reporting.

Yang and Jiang (2018) proposed a meta-clustering ensemble approach considering bi-weighting that uses the Hidden Markov Model (HMM) to clustering temporal data. The purpose of HMM is to solve problems related to primary clusters and model selection. The bi-weighting approach can improve the fusion of consensus functions through adaptive partitioning. In addition, HMM has a dendrogram-based optimal consensus partition modification process that can determine the number of clusters adaptively and automatically. Yang and Jiang (2018) introduced the

Dendrogram-based Similarity Partitioning Algorithm (DSPA) for temporal data clustering. DSPA is a bi-weighted ensemble algorithm that clusters data based on HMM. DSPA solves primary clustering and model selection issues through HMM. In addition, DSPA automatically determines the intrinsic number of clusters by applying HMM during the consensus process. Banerjee et al. (2021) proposed the Weighted Hierarchical Agglomerative Clustering Ensemble (WHAC). WHAC is a new similarity measurement at the cluster level that can determine the merit of an individual clustering method. In addition to the concept of agreement, WHAC also considers the concept of disagreement between clusters. WHAC uses the merit criterion based on a polynomial heuristic for coupled ensemble selection.

4. Proposed scheme

In this paper, the MCEMS scheme is proposed to improve ensemble clustering. MCEMS addresses the challenges of model selection and consensus function in ensemble clustering. The main idea of MCEMS in model selection is to use a bi-weighting policy that selects a subset of individual AHC methods with simultaneous consideration of diversity and quality. MCEMS formulates the consensus function in ensemble clustering through the clusters clustering technique and the creation of meta-clusters. In addition, MCEMS uses a novel similarity measurement to calculate the similarity between instances, which simultaneously assumes knowledge of several clustering algorithms. MCEMS creates final clusters by assigning each instance to a meta-cluster with the highest similarity, where the number of optimal clusters is automatically determined by merging similar clusters and considering a threshold. The proposed ensemble clustering flowchart is presented in Fig. 4.

According to the proposed scheme, the model selection process is performed based on the available m AHC method, which leads to the selection of the m' AHC method to create the ensemble. The data are clustered based on selected methods. Then, the proposed similarity measurement is applied based on all the primary clusters created to calculate the similarity between each instance pair. The proposed similarity measurement emphasizes the size of the clusters and the presence of the same instances in the clusters. After creating primary clusters, meta-clusters are created using the clusters clustering technique. This technique develops the concept of clustering by considering a set of primary clusters from the instance level to the cluster level. MCEMS automatically determines the number of optimal clusters by merging clusters based on a threshold. Finally, the meta-clusters act as a consensus function, and each instance is assigned to a meta-cluster to obtain the final clusters. Here, the instances are assigned to the meta-cluster with the highest similarity based on the proposed novel similarity measurement.

4.1. Model selection

In general, diversity in the results of individual clustering methods can lead to the development of better-quality ensemble clustering approaches. Creating primary clusters with high diversity means analyzing data from different aspects. Although diversity can increase the quality of ensemble clustering results, the important consideration is the selection of individual primary clustering methods to create the ensemble. In this paper, the set P with m of the individual AHC method is available, where the purpose is to select the subset P^* with the m' method of the set P. Hence, ensemble clustering can be created with less complexity by P^* set. We examine the problem of model selection through a bi-weighting scheme. This scheme assigns appropriate weights to the m input method based on diversity as well as to each cluster created in each method based on quality. Here, the weighted average in the bi-weighting scheme is used to ensure the effectiveness of individual AHC methods as well as their clusters in ensemble formation. Finally, ρ percent of the P set methods are used as a P^* subset to create ensemble clustering. According to the simulations and evaluation results, the most appropriate value for ρ is 0.35.

Let w_k^1 be the weight of $p_k \in P$, where p_k refers to the k-th of the individual AHC method. Here, Shannon's entropy (Maruyama et al., 2005) is used to calculate w_k^1. Also, let $A(p_k)$ be Shannon's entropy for p_k, where a larger value means more difference in clustering results. In general, more differences in the results of the two clustering methods can lead to higher diversity in model selection. Accordingly, w_k^1 diversity represents the p_k method for model selection policy and ensemble clustering. w_k^1 is calculated according to Eq. (1).

$$w_k^1 = \frac{1}{m}\sum_{i=1}^{m} \|A(p_k) - A(p_i)\| \tag{1}$$

where, $\|*\|$ is defined as the Kullback–Leibler distance (Do and Vetterli, 2002) and $A(p_k)$ is Shannon's entropy for the clustering results of the p_k method. In fact, $A(p_k)$ indicates uncertainty, randomness and disorder, as defined in Eq. (2).

$$A(p_k) = -\sum_{i=1}^{n} P(r_k^i).\log P(r_k^i) \tag{2}$$

where, r_k^i refers to the label predicted for i-th instance by p_k method, and n is the total number of instances in the dataset. Also, $P(r_k^i)$ is the probability distribution in entropy.

The second weight in the bi-weighting scheme depends on the set of clusters in each method, which is defined as w_k^2. w_k^2 refers to the dissimilarity of the clusters (as inter-cluster distance) and also to the dissimilarity of the instances inside the clusters (as intra-cluster distance) in the p_k method. In general, the purpose is to form clusters that have more similarity of instances in the same clusters and more differences between different clusters. Thus, w_k^2 indicates the better quality of individual AHC methods. In this regard, w_k^2 can be calculated with Eq. (3).

$$w_k^2 = \frac{\sum_{(c_i,c_j)\in p_k} d(c_i,c_j)}{\sum_{c\in p_k}\sum_{(x_a,x_b)\in c} d(x_a,x_b)} \tag{3}$$

where, c_i and x_a refer to the i-th cluster and the a-th instance, respectively.

4.2. Primary clusters

The MCEMS scheme uses the combination of results of several individual AHC methods to improve clustering results. This process is performed by m' individual method to obtain the results $P^* = [p_1, p_2, \cdots, p_j, \cdots p_{m'}]$. Here, p_j refers to the results of the j-th clustering method, where it can be methods such as single linkage, average linkage, centroid linkage and complete linkage. In this paper, all methods in P^* set is used to calculate the similarity between each pair of instances. The proposed novel similarity measurement emphasizes the presence of more similar instances in the clusters as well as the clusters size. Each p_j presents the results hierarchically, where each level can be a form of clustering. Here, we consider each level of p_j as an independent cluster. For better understand, we report the clusters for each HC method from the example in Fig. 2 in Table 2.

The novel similarity measurement as F is a consensus technique for combining the results of m individual AHC methods, which are calculated according to Eq. (4).

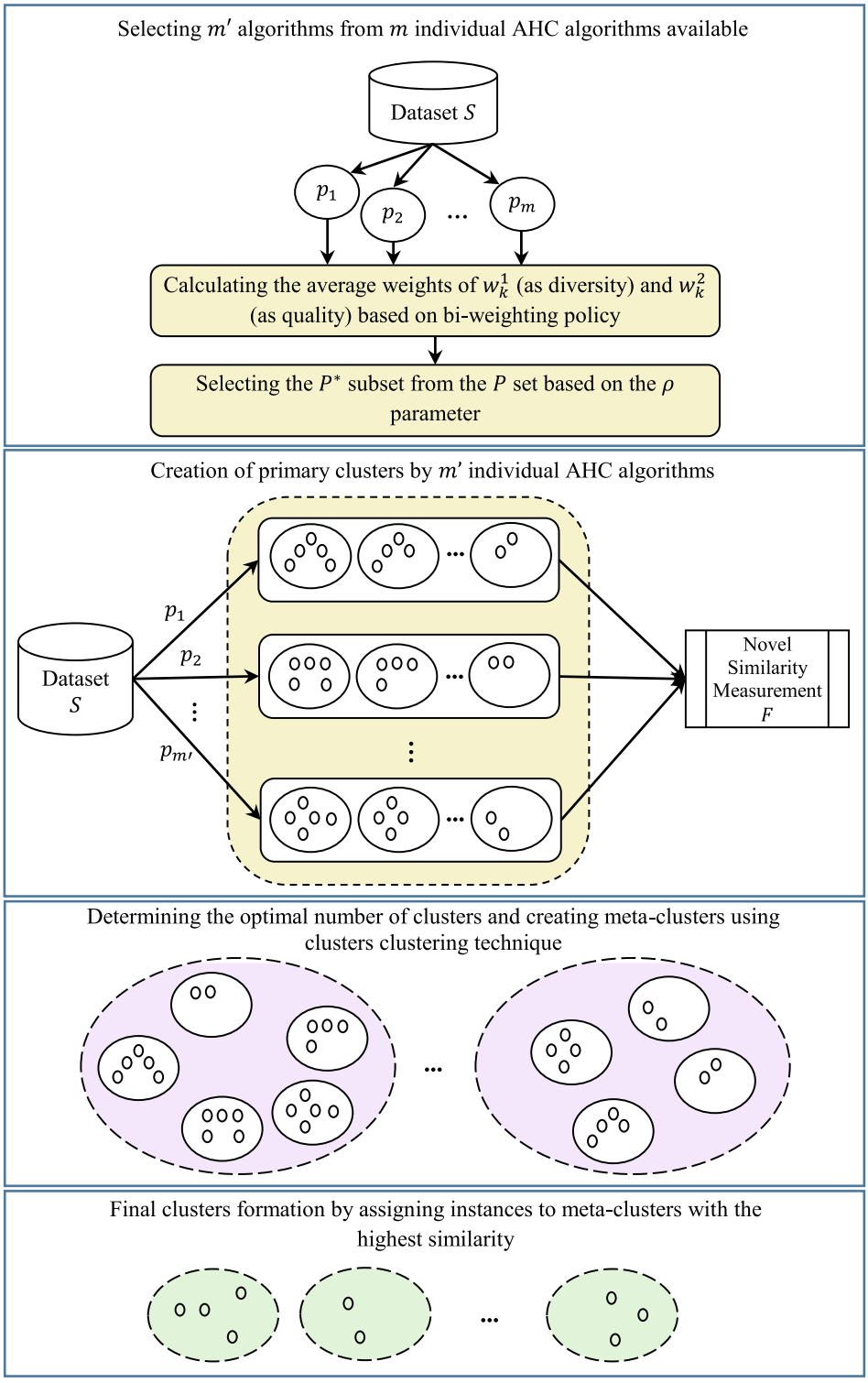

Fig. 4. Flowchart of MCEMS scheme.

Table 2
The considered clusters of each method in the example presented in Fig. 2.

Linkage method	Instances available in each cluster				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Single	3,6	2,5	2,3,5,6	2,3,4,5,6	1,2,3,4,5,6
Average	3,6	2,5	3,4,6	2,3,4,5,6	1,2,3,4,5,6
Centroid	3,6	2,5	3,4,6	1,3,4,6	1,2,3,4,5,6
Complete	3,6	2,5	3,4,6	1,2,5	1,2,3,4,5,6

$$f(x_i,x_j)=\sum_{q=1}^{m'}\sum_{z=1}^{k_q}\begin{cases}\beta^{|c_z^q|} & if\ (x_i,x_j)\in c_z^q\\ 0 & otherwise\end{cases},\quad \forall(x_i,x_j)\in S \tag{4}$$

where, $f(x_i,x_j)\in$ F provides the similarity of the instances $x_i\in S$ and $x_j\in S$. $|c_z^q|$ is the number of instances in the z-th cluster of the q-th method and k_q is the number of clusters obtained by q-th method. Also, β is the attenuation coefficient in the range [0, 1], which assigns less similarity to large clusters, because the similarity between the instances in smaller clusters can be more important.

In order to improve the proposed novel similarity measurement in creating the final clusters, we normalize F in the range [0, 1]. Let PC be the set of all primary clusters of all levels for the m' AHC methods. Considering all clusters of all individual clustering methods, some clusters from the PC set may be exactly the same. To reduce complexity, we remove clusters with the same content from the PC set. The pseudo-code for creating primary clusters and the similarity measurement process is shown in Algorithm 2.

Algorithm 2. Creating primary clusters and similarity measurement

Input: Dataset S, β and $m\prime$ individual AHC methods .

Output: Primary clusters and similarity matrix F.

```
1:  for q = 1 to m' do
2:      Creating primary clusters on the dataset S by q-th
        method from individual AHC methods.
3:  end
4:  for q = 1 to m' do
5:      For z = 1 to k_q do
6:          if (x_i, x_j) ∈ c_z^q then
7:              f(x_i, x_j) = f(x_i, x_j) + β^|c_z^q|.
8:      else
9:              The value of f(x_i, x_j) is unchanged.
10:         End
11:     end
12: end
13: Normalization of similarity measurement F in the range
    [0, 1].
14: All primary clusters of all levels are placed in the PC set.
15: Removing same clusters from the PC set.
```

In this algorithm, lines 1 to 3 create the primary clusters by m' individual AHC method. Line 4 is repeated to check the similarity calculation on each AHC method. Line 5 relates to examining all clusters in an individual AHC method. Lines 6 to 10 are presented to calculate the similarity between the two instances. Line 13 provides a normalized similarity matrix. Finally, line 14 creates the PC set that contains all the primary clusters, and the same clusters are removed based on line 15.

4.3. Clusters clustering technique

Using ensemble clustering can extend the clustering process from instance level to cluster level. The PC set includes all primary clusters of all individual AHC methods. Considering each cluster of PC as an instance, it can be re-clustered. Clusters clustering technique leads to the formation of meta-clusters, where each meta-cluster contains one or more independent clusters. Meta-clusters have clustering knowledge of several different clustering methods that can help improve the formation of final clusters. Let ϕ be the method used to create meta-clusters. Clusters clustering technique can be done with a AHC method such as single linkage, average linkage, centroid linkage and complete linkage. In this paper, the average linkage method is used as ϕ, because it provides better results than other linkage methods (Yang and Jiang, 2018). However, we compare different AHC methods to evaluate effectiveness. The similarity of the two instances of the dataset is available by $f(x_i,x_j)$. However, in meta-clusters the concept of similarity from instance level to cluster level is generalized. Moreover, the similarity of all pairs of instances in the clusters must be calculated to measure the similarity of the clusters, as shown in Eq. (5).

$$\Psi(mc_\alpha,mc_\beta)=\frac{1}{|mc_\alpha|.|mc_\beta|}\sum_{v=1}^{|mc_\alpha|}\sum_{w=1}^{|mc_\alpha|}\left[\frac{\sum_{i=1}^{|c_v|}\sum_{j=1}^{|c_w|}\Gamma(x_i,x_j)}{|c_v|.|c_w|}\right],\quad \forall x_i\in c_v,\ x_j\in c_w \tag{5}$$

where, $\Psi(mc_\alpha,mc_\beta)$ shows the average similarity between the meta-cluster mc_α and mc_β. $|mc_\alpha|$ and $|mc_\beta|$ the number of clusters assigned to meta-clusters mc_α and mc_β, respectively. Also, $|c_v|$ and $|c_w|$ shows the number of instances in clusters c_v and c_w, respectively.

4.4. Final clusters

Final clusters can be formed through meta-clusters. This is done through allocating each instance to a meta-cluster with the most similarity. However, AHC-based methods present the results as dendrograms in which the number of optimal clusters is unknown. The optimal number of clusters can be achieved through different levels of dendrograms, where each level can provide a form of data clustering. Here, the number of optimal clusters is determined through merging the primary clusters and considering a threshold value. Let K be the number of optimal clusters. By determining the value of K, the level of the dendrogram that has the closest cluster number to K is selected as the final clustering. Accordingly, MCEMS can automatically find the optimal number of clusters. Here, K is estimated by merging the primary clusters (i.e., all clusters of the PC set) until the threshold θ is reached, as shown in Eq. (6). Clusters merging is repeated until the similarity of any pair of clusters is greater than θ.

$$if\ \sigma(c_a,c_b)\geqslant\theta\begin{cases}hence\ merged & True\\ not\ merged & False\end{cases},\quad \forall a,b\in PC \tag{6}$$

where, c_a and c_b are two clusters of the PC. Also, $\sigma(c_a,c_b)$ indicates to the average similarity between c_a and c_b.

After determining K, the desired level is selected from the dendrogram and the meta-clusters are identified. In this regard, final clusters are generated based on the allocation of instances to the meta-cluster with the highest similarity. Therefore, the number of meta-clusters refers to the number of final clusters. We calculate the similarity of each instance (e.g., x_i) to a meta-cluster (e.g., mc_α) according to the average similarity of x_i to total instances in mc_α, as shown in Eq. (7).

$$\psi(x_i,mc_\alpha)=\frac{1}{|mc_\alpha|}\cdot\sum_{k=1}^{|mc_\alpha|}\frac{\sum_{z=1}^{|c_k^\alpha|}\Gamma(x_i,x_j)}{|c_k^\alpha|},\quad \forall x_j\in c_k^\alpha \tag{7}$$

where, $\psi(x_i,mc_\alpha)$ shows the average similarity between x_i and mc_α. $|mc_\alpha|$ the number of clusters in mc_α. Also, $|c_k^\alpha|$ the number of instances belongs to the k-th cluster of mc_α.

For better clarity, we have given the pseudo-code related to the final clustering process in Algorithm 3.

Algorithm 3. Creating final clusters

Input: Primary clusters in the *PC* set and meta-clusters.

Output: Final clusters, where c_k refers to k-th cluster.

```
Calculating K based on the PC set and Eq. (6).
Selecting the appropriate level of dendrogram based on the value of K.
Creating K empty clusters.
for i = 1 to n do
    for α = 1 to K do
        ψ(x_i, mc_α) is calculated based on Eq (7).
    End
    Instance x_i is assigned to the cluster c_k, If ψ is the maximum for mc_α.
end
```

5. Experimental results

In this section, extensive experiments are performed on 20 datasets with different dimensions to evaluate the effectiveness of the proposed MCEMS scheme from different perspectives. All datasets used for comparison are collected from the UCI machine learning repository, where they are approved by Bagherinia et al. (2019), Yang and Jiang (2018) and Banerjee et al. (2021) for clustering work. The performance of MCEMS is evaluated in comparison with linkage-based methods (i.e., single linkage, average linkage, centroid linkage and complete linkage) and several state-of-the-art clustering ensemble algorithms. Here, the state-of-the-art algorithms used for the comparison work are HMM (Yang and Jiang, 2018), DSPA (Yang and Jiang, 2018) and WHAC (Banerjee et al., 2021). Meanwhile, due to the possibility of slight variation of results in different runs and providing assurance of results, we report the results of all experiments based on 25 distinct runs. All simulations are performed by MATLAB R2019a on Lenovo laptop IdeaPad 320 with Intel Core i7-7500U processor at 3.5 GHz and 8 GB RAM.

This section includes the fifth subsection. The subsection 5.1 is dedicated to experimental setup. The subsection 5.2 reviews the evaluation criteria. The description of the datasets is given in subsection 5.3. In section 5.4 the parameters of MCEMS are analyzed and subsection 5.5 provides comparisons and discussions.

5.1. Experimental setup

The performance of the MCEMS scheme is evaluated based on several datasets. However, the presence of noise was not considered in this study. Therefore, there is no concern about the presence of noise and its effect on determining the number of clusters. The MCEMS scheme has four input parameters whose proper adjustment can increase the efficiency of clustering results. These parameters include ρ (to the model selection), β (to calculate similarity in Eq. (4)), θ (to merge primary clusters in Eq. (6)) and ϕ (to create final clusters). In this paper, different scenarios for these parameters are analyzed and the results are used for comparison based on the best values. This analysis is performed in subsection 5.4 and the best results are obtained based on $\rho = 0.35$, $\beta = 0.05$, $\theta = 0.25$ and $\phi = averagelinkage$.

5.2. Evaluation metrics

In this paper, various evaluation criteria are used to compare AHC-based clustering methods. These criteria include the Wilcoxon hypothesis test (Banerjee et al., 2021) and the CoPhenetic Correlation (CPC) coefficient (Bagherinia et al., 2019), which are discussed below.

The Wilcoxon test is a nonparametric statistical test used to evaluate the matching of paired groups that depend on sum rank scale. In this test, the variables can have various distances because the Wilcoxon test considers the size of the various between the rankings. In Wilcoxon test, first hypotheses zero and one are defined and then the significant level of α is determined. According to the value of α, the statistical value is measured and finally the hypothesis is rejected or accepted. In general, statistical tests consist of two hypotheses, namely zero and one. The one hypothesis is something that needs to be proven, and zero hypothesis is the opposite. Basically, the zero hypothesis is correct, and the probabilistic matching value of the instances is calculated through this hypothesis. This is known as meaningfulness significance (or p-value). If the p-value is less than the probability of one hypothesis (i.e., α), zero hypothesis is rejected. Based on this, one hypothesis can be accepted at the significance level of α.

In addition to the Wilcoxon test, the CPC coefficient can analyze the dendrogram compatibility created by an AHC method. This process is performed based on the comparison of the calculated distance matrix from the input data and clustering. Here, distances for the input data are measured based on the Euclidean metric of the instances pair and is known as the adjacency matrix. The distance matrix in AHC methods is measured through the descriptor matrix. Let Y be the adjacency matrix and W be the descriptor matrix. Accordingly, the CPC coefficient for Y and W is calculated based on Eq. (11).

$$CPCC = \frac{\sum_{i<j}\left(Y_{ij} - \bar{Y}\right)\left(W_{ij} - \bar{W}\right)}{\sqrt{\sum_{i<j}\left(Y_{ij} - \bar{Y}\right)^2}\sqrt{\sum_{i<j}\left(W_{ij} - \bar{W}\right)^2}} \tag{11}$$

where, Y_{ij} and W_{ij} are the distances between instances i and j, in Y and W, respectively, and $\bar{Y}$ and $\bar{W}$ are the mean values for Y and W, respectively. The CPC coefficient output is in the range of $[-1, 1]$ so values close to 1 indicate a suitable clustering method.

Each cluster in an AHC method is known as a dendrogram, where the dendrograms show the structure of the instances hierarchically. This hierarchical structure is a descriptor matrix with dimensions $n \times n$ (Jafarzadegan et al., 2019). This matrix expresses the relative distance between the instances in the dendrogram. There are several approaches to creating a descriptor matrix, such as Partition Membership Divergence (PMD), Sub-dendrogram Membership Divergence (SMD), Cluster Membership Dendrogram (CMD), Cophenetic Difference (CD), and Maximum Number of Edge Distance (MNED). PMD indicates the number of clusters in the dendrogram where two specific instances are not in the same cluster. SMD shows the number of sub-dendrograms that are not present in two specific instances. CMD shows the number of instances in the nearest cluster containing two specific instances. CD shows the height of the nearest cluster that connects two specific instances. MNED expresses the level of the nearest cluster that connects two specific instances.

5.3. Datasets

The UCI machine learning repository is one of the oldest sources of data on the Web. Most of the datasets in this repository are pre-

pared with the participation of users and have different levels of cleanliness and purification. It is possible to download the dataset directly from the UCI machine learning repository. In this paper, 25 different datasets are used to evaluate the proposed MCEMS scheme compared to other similar algorithms, all of which are accessible through the UCI machine learning repository. This dataset by Bagherinia et al. (2019), Banerjee et al. (2021) and Yang and Jiang (2018) have also been used. Due to the possible impact of dataset dimensions (number of instances, number of features and number of classes) in the clustering process, datasets with different dimensions have been selected for evaluation. Some datasets have a small number of instances, while some have many instances. This also applies to the number of features and the number of classes. The description of the dataset used is shown in Table 3.

5.4. Parameters analysis

This section is dedicated to the analysis of MCEMS scheme parameters including ρ, β, θ and φ. Our goal is to find the best values for the parameters that lead to optimal clustering. Due to the wide range of datasets used, all results in this section are reported based on the average of all datasets. We analyze each parameter independently with different values to find the most optimal values. Optimal values are based on the average of Wilcoxon test and CPC coefficient. When analyzing each parameter, the other parameters of the MCEMS scheme are set to default values. The default values are $\rho = 0.35$, $\beta = 0.05$, $\theta = 0.25$ and $\phi = average\ linkage$.

The ρ parameter in MCEMS indicates how many percent of the available individual AHC methods are used to participate in ensemble clustering. In this paper, linkage-based methods including single linkage, average linkage, centroid linkage and complete linkage based on Euclidean, Cosine and Manhattan distance functions have been used as available AHC methods. In other words, each linkage-based method is considered with three different types of distances. Therefore, $m = 12$ and m' must be determined based on ρ, where $m' = m * \rho$. Table 4 shows the results of the ρ parameter analysis with different values. Each column of results is related to an evaluation metric, and the last column refers to the average results. Highlighted values mean the best results for each evaluation metric. As illustrated, MCEMS performs best at both the Wilcoxon test and CPC coefficient metrics with $\rho = 0.35$. Accordingly, the value of m' in the simulations is approximately equal to 4. Examination of the model selection process shows that in most cases, single linkage, average linkage, centroid linkage and complete linkage methods are selected based on Euclidean distance to produce ensemble clustering.

Table 4
Analysis of ρ parameter in MCEMS scheme.

Value	Evaluation metric		
	Wilcoxon test	CPC coefficient	Average
0.2	0.7036 ± 0.06	0.8076 ± 0.04	0.7556 ± 0.05
0.25	0.7145 ± 0.06	0.8076 ± 0.06	0.7610 ± 0.06
0.3	0.7150 ± 0.07	0.8087 ± 0.05	0.7618 ± 0.06
0.35	**0.7161 ± 0.07**	**0.8102 ± 0.04**	**0.7631 ± 0.06**
0.4	0.7149 ± 0.05	0.8093 ± 0.06	0.7621 ± 0.06

Table 5
Analysis of β parameter in MCEMS scheme.

Value	Evaluation metric		
	Wilcoxon test	CPC coefficient	Average
0.01	0.7089 ± 0.05	0.8088 ± 0.07	0.7588 ± 0.05
0.05	0.7138 ± 0.07	**0.8102 ± 0.04**	**0.7620 ± 0.06**
0.1	**0.7161 ± 0.07**	0.8037 ± 0.06	0.7599 ± 0.07
0.15	0.7100 ± 0.06	0.8053 ± 0.06	0.7576 ± 0.06
0.2	0.7158 ± 0.06	0.8011 ± 0.05	0.7584 ± 0.05

The results of β-parameter analysis is presented in Table 5. This parameter is very important for calculating the proposed novel similarity measurement. The comparison of the β parameter with different values is based on the Wilcoxon test and the CPC coefficient, where the highlighted results refer to the superior values. The average results for these two metrics show that $\beta = 0.05$ provides better performance for MCEMS than other values. However, $\beta = 0.1$ is also promising and gives better results for the Wilcoxon test.

Table 3
Description of the dataset used.

Index	Dataset	#Instances	#Features	#Classes
DS1	Wine	178	13	3
DS2	Vehicle	846	18	4
DS3	Seeds	210	7	3
DS4	Pima-diabetes	768	8	2
DS5	Landsat-satellite	6435	36	6
DS6	Image-segmentation	2310	19	7
DS7	Mammographic	961	5	2
DS8	Glass	214	10	7
DS9	Bupa	323	4	7
DS10	Breast	683	9	2
DS11	Avila	20,867	10	12
DS12	Yeast	1484	8	10
DS13	Ecoli	336	8	8
DS14	Digits	5620	63	10
DS15	Banana	2000	2	2
DS16	Ring3	1500	2	3
DS17	Imbalance	2250	2	2
DS18	Bupa	345	6	2
DS19	Aggregation	788	2	7
DS20	SAHeart	462	9	2
DS21	Ionosphere	351	34	2
DS22	Galaxy	323	4	7
DS23	Half-Ring	400	2	2
DS24	Hand-made1	300	2	3
DS25	Letter-recognition	20,000	16	26

Table 6
Analysis of θ parameter in MCEMS scheme.

Value	Evaluation metric		
	Wilcoxon test	CPC coefficient	Average
0.2	0.7158 ± 0.09	0.8070 ± 0.05	0.7614 ± 0.07
0.25	**0.7161 ± 0.07**	**0.8102 ± 0.04**	**0.7631 ± 0.06**
0.3	0.7147 ± 0.07	0.8088 ± 0.05	0.7617 ± 0.06
0.35	0.7135 ± 0.08	0.8074 ± 0.04	0.7604 ± 0.06
0.4	0.7140 ± 0.10	0.8036 ± 0.03	0.7588 ± 0.07

Table 7
Analysis of ϕ parameter in MCEMS scheme.

Linkage method	Evaluation metric		
	Wilcoxon test	CPC coefficient	Average
Single	0.7086 ± 0.11	0.8037 ± 0.09	0.7561 ± 0.10
Average	**0.7161 ± 0.07**	**0.8102 ± 0.04**	**0.7631 ± 0.06**
Centroid	0.7144 ± 0.05	0.8084 ± 0.06	0.7614 ± 0.06
Complete	0.7159 ± 0.06	0.8096 ± 0.04	0.7627 ± 0.05

Another parameter of the MCEMS scheme that plays an important role in determining the number of optimal clusters automatically is the parameter θ. The results of the parameter θ analysis is presented in Table 6. Highlight values represent the best results for each metric. The average results show that $\theta = 0.25$ has a better performance for MCEMS than other values.

The MCEMS scheme uses a linkage-based AHC method to create meta-clusters (e.g., single linkage, average linkage, centroid linkage, and complete linkage). This parameter is defined as ϕ. The analysis of this parameter to find the best method for creating meta-clusters is reported in Table 7. Highlight values represent the best results for each metric. As illustrated, the most suitable method for clusters clustering and creating meta-clusters is the average linkage method. In addition, the results of the complete linkage method are also competitive.

5.5. Results and comparisons

In this section, various experiments have been performed based on different criteria to evaluate and compare the proposed algorithm in comparison with other similar algorithms. The comparison is based on Wilcoxon test and CPC coefficient. Here, linkage-based methods include single linkage, average linkage, centroid linkage and complete linkage, as well as state-of-the-art ensemble clustering algorithms including HMM (Yang and Jiang, 2018), DSPA (Yang and Jiang, 2018) and WHAC (Banerjee et al., 2021) are used for comparison. In all these algorithms, dendrogram reconstruction from consensus function is performed by average linkage and CD method is considered to convert primary clusters to descriptor matrix.

The Wilcoxon rank sum test is performed to validate the performance and dissimilarity of MCEMS (as the proposed algorithm) with other comparable algorithms. For this experiment, all linkage-based methods as well as HMM, DSPA and WHAC algorithms for p-value analysis are compared. We consider a significant threshold similar to previous studies to be 0.05 (Jafarzadegan et al., 2019; Banerjee et al., 2021). Therefore, the Wilcoxon rank sum test is analyzed using the zero hypothesis over 0.05% significance level (i.e., 95% confidence level). We calculate the p-value of all algorithms based on the Silhouette index (Mojarad et al., 2021), which is a measure of the clustering quality coefficient based on pairwise difference between- and within-cluster distances. The results of this test at $\alpha = 0.05$ for MCEMS scheme compared to other algorithms are reported in Table 8.

Each row of this table refers to the p-value and Significant Value (SG) results for a dataset, which are calculated based on an average of 25 runs, where the last row is dedicated to the average results for all datasets. In this regard, the p-value refers to the probability of the Wilcoxon test. If $p \leq 0.05$, zero hypothesis is rejected, otherwise zero hypothesis is accepted. In the SG column, the symbols '-' and '-' indicate zero hypothesis rejection, where '-' indicates better p-value results for MCEMS, and '-' indicates the superiority of other

Table 8
Comparison of MCEMS with other algorithms in Wilcoxon rank sum test at $\alpha = 0.05$.

Dataset	Single		Average		Centroid		Complete		HMM		DSPA		WHAC	
	p-value	SG	p-value	SG	p-value	SG	p-value	SG	p-value	SG	p-value	SG	p-value	SG
DS1	0.0020	–	0.0021	–	0.0023	–	0.0156	–	0.0227	–	0.0163	–	0.0181	–
DS2	0.0010	–	0.0013	–	0.0006	–	0.0008	–	0.0055	–	0.0013	–	0.0020	–
DS3	0.0001	–	0.0001	–	0.0001	–	0.0000	–	0.0015	–	0.0013	–	0.0001	–
DS4	0.0008	–	0.0006	–	0.0004	–	0.0008	–	0.0016	–	0.0028	–	0.0008	–
DS5	0.0009	–	0.0005	–	0.0017	–	0.0019	–	0.0858	+	0.0413	–	0.0021	–
DS6	0.0002	–	0.0001	–	0.0001	–	0.0002	–	0.0018	–	0.0014	–	0.0002	–
DS7	0.0005	–	0.0005	–	0.0005	–	0.0005	–	0.0010	–	0.0035	–	0.0008	–
DS8	0.0001	–	0.0001	–	0.0001	–	0.0001	–	0.0015	–	0.0013	–	0.0001	–
DS9	0.0005	–	0.0005	–	0.0007	–	0.0008	–	0.0070	–	0.0013	–	0.0027	–
DS10	0.0006	–	0.0006	–	0.0006	–	0.0008	–	0.0011	–	0.0188	–	0.0085	–
DS11	0.0003	–	0.0009	–	0.0005	–	0.0004	–	0.0010	–	0.0016	–	0.0011	–
DS12	0.0001	–	0.0004	–	0.0004	–	0.0003	–	0.0010	–	0.0120	–	0.0051	–
DS13	0.0002	–	0.0000	–	0.0002	–	0.0001	–	0.0010	–	0.0013	–	0.0002	–
DS14	0.0016	–	0.0011	–	0.0014	–	0.0011	–	0.0126	–	0.0017	–	0.0527	+
DS15	0.0010	–	0.0008	–	0.0011	–	0.0005	–	0.0010	–	0.0059	–	0.0020	–
DS16	0.0002	–	0.0002	–	0.0002	–	0.0002	–	0.0011	–	0.0021	–	0.0002	–
DS17	0.0011	–	0.0015	–	0.0016	–	0.0013	–	0.0033	–	0.0368	–	0.0186	–
DS18	0.0008	–	0.0009	–	0.0004	–	0.0007	–	0.0405	–	0.0013	–	0.0245	–
DS19	0.0001	–	0.0004	–	0.0002	–	0.0002	–	0.0018	–	0.0020	–	0.0005	–
DS20	0.0049	–	0.0042	–	0.0032	–	0.0067	–	0.0294	–	0.0073	–	0.0169	–
DS21	0.0014	–	0.0021	–	0.0024	–	0.0017	–	0.0370	–	0.0025	–	0.0183	–
DS22	0.0459	–	0.0552	–	0.0484	–	0.0488	–	0.3307	+	0.3270	+	0.2174	+
DS23	0.0004	–	0.0000	–	0.0002	–	0.0003	–	0.0010	–	0.0201	–	0.0091	–
DS24	0.0023	–	0.0022	–	0.0023	–	0.0024	–	0.0111	–	0.0075	–	0.0028	–
DS25	0.0023	–	0.0020	–	0.0031	–	0.0029	–	0.0256	–	0.0034	–	0.0118	–
Average	0.0020	–	0.0019	–	0.0021	–	0.0024	–	0.0251	–	0.0209	–	0.0166	–

algorithms. Rejection of zero hypothesis means that the average of the two comparable algorithms is significant. Further, the symbol '+' indicates the acceptance of zero hypothesis by MCEMS compared to other algorithms. Therefore, $p > 0.05$ means that the two compared algorithms are significantly different and MCEMS failed based on the dissimilarity test.

According to Table 8, we conclude that MCEMS is superior to all linkage-based methods and that all comparisons are significant, except compared to the average linkage method for the DS22 dataset. MCEMS is superior to all datasets compared to HMM, however the results are not significant in the DS5 and DS22 datasets. Moreover, MCEMS has surpassed DSPA in all datasets except DS11. In this comparison, the results are not significant for the DS22 dataset. Also, MDBO reported better results than WHAC in 23 datasets. In this comparison, the *p*-value results are not significant for the DS14 and DS22 datasets. However, the results in the DS14 dataset are competitive. Based on the results of the average rank sum in the last row, MCEMS has 17.6%, 15.2%, 17.8% and 18.1% superiority over single linkage, average linkage, centroid linkage and complete linkage, respectively. This superiority compared to HMM, DSPA and WHAC algorithms was reported to be 3.2%, 2.7% and 2.1%, respectively. In general, the results of the Wilcoxon rank sum test show the efficiency of MCEMS because it has reached a *p*-value of less than 0.05 compared to all algorithms and for most datasets. These results can indicate the rejection of zero hypothesis and the prove of the effectiveness of MCEMS for clustering real data.

The CPC coefficient is used to evaluate the degree of conformity between the dendrogram and the distance matrix. The larger the value of this coefficient, the greater the conformity between the dendrogram in different algorithms and the calculated distance matrix. Therefore, the nature of the data changes less in the CPC coefficient. A value greater than 0.9 indicates high conformity between the dendrogram and the distance matrix, whereas CPC values less than 0.74 indicate no conformity (Yang et al., 2007). Table 9 shows the CPC-related comparison results for different algorithms on the 25 datasets used. Each row is related to a dataset and the last row is dedicated to the average results. Highlight values represent the best results for each dataset.

According to the reported results, MCEMS is superior to all linkage-based methods (i.e., single linkage, average linkage, centroid linkage and complete linkage). The average of this superiority is 15.2%, 4.0%, 5.8% and 5.3%, respectively. The proposed algorithm performs better on all datasets except D14 compared to the HMM algorithm. The superiority of MCEMS over DSPA and WHAC algorithms is also clearly described. The proposed MCEMS scheme surpassed DSPA algorithm in all datasets except D13 and D24 and also outperformed WHAC algorithm in all datasets except D1 and D10. On average, the superiority of MCEMS compared to HMM, DSPA and WHAC has been reported at 3.2%, 2.4% and 0.7%, respectively.

The MCEMS scheme can automatically determine the most appropriate number of clusters by merging primary clusters based on the θ threshold. Here, K is considered as the number of optimal clusters. Because the number of clusters in AHC-based methods is determined through the dendrogram and the appropriate level selection, MCEMS selects the level of the dendrogram that has the closest number of clusters to K as the number of final clusters. Accordingly, MCEMS performs an automated process to detect the number of clusters in AHC-based methods. In the following experiments, the efficiency of MCEMS in determining the number of optimal clusters is evaluated. Fig. 5 shows the comparison results of MCEMS and ensemble clustering algorithms (i.e., HMM, DSPA and WHAC) for the 25 datasets used. Two values are reported for each dataset: the number of actual clusters and the number of clusters detected by the clustering algorithms. The results clearly show that the MCEMS curve is less different from the curve related to the number of actual clusters, and then the WHAC algorithm has obtained better results. On average, the difference between the number of clusters and the number of actual clusters in MCEMS is 12.3. These differences are reported for HMM, DSPA and WHAC algorithms 18.9, 17.6 and 15.0, respectively.

In addition to the number of clusters, the clustering accuracy of MCEMS in compared to other ensemble clustering algorithms is compared. The results of this comparison are reported in Fig. 6 for different algorithms. The clustering accuracy depends on the number of instances of the dataset that are correctly assigned to the respective class. Therefore, clustering accuracy is the ratio of

Table 9
Comparison of MCEMS with other algorithms in CPC coefficient.

Dataset	Single	Average	Centroid	Complete	HMM	DSPA	WHAC	MCEMS
DS1	0.4864 ± 0.02	0.7950 ± 0.08	0.7669 ± 0.06	0.7767 ± 0.10	0.7591 ± 0.05	0.8185 ± 0.09	**0.8350 ± 0.05**	0.8318 ± 0.03
DS2	0.5500 ± 0.07	0.7962 ± 0.05	0.7785 ± 0.07	0.7903 ± 0.07	0.8131 ± 0.10	0.8065 ± 0.06	0.8230 ± 0.13	**0.8235 ± 0.07**
DS3	0.5502 ± 0.12	0.7441 ± 0.04	0.7420 ± 0.05	0.7341 ± 0.15	0.7701 ± 0.06	0.7655 ± 0.07	0.7718 ± 0.09	**0.7772 ± 0.03**
DS4	0.8631 ± 0.02	0.9345 ± 0.03	0.9272 ± 0.04	0.9210 ± 0.03	0.9411 ± 0.05	0.9411 ± 0.05	0.9409 ± 0.06	**0.9429 ± 0.01**
DS5	0.7592 ± 0.03	0.7704 ± 0.03	0.7824 ± 0.03	0.7837 ± 0.05	0.8150 ± 0.13	0.8315 ± 0.04	0.8358 ± 0.04	**0.8562 ± 0.00**
DS6	0.7566 ± 0.04	0.9049 ± 0.04	0.9033 ± 0.04	0.9033 ± 0.06	0.9141 ± 0.04	0.9265 ± 0.05	0.9313 ± 0.05	**0.9342 ± 0.01**
DS7	0.5828 ± 0.04	0.6459 ± 0.06	0.6089 ± 0.05	0.6267 ± 0.05	0.6653 ± 0.05	0.6814 ± 0.06	0.7091 ± 0.06	**0.7119 ± 0.03**
DS8	0.8034 ± 0.09	0.8222 ± 0.10	0.8186 ± 0.08	0.8162 ± 0.11	0.8479 ± 0.09	0.8441 ± 0.10	0.8489 ± 0.11	**0.8503 ± 0.08**
DS9	0.7687 ± 0.16	0.8255 ± 0.02	0.7904 ± 0.06	0.8080 ± 0.04	0.8427 ± 0.10	0.8363 ± 0.09	0.8431 ± 0.15	**0.8686 ± 0.07**
DS10	0.7045 ± 0.05	0.7260 ± 0.05	0.7222 ± 0.04	0.7222 ± 0.06	0.7601 ± 0.04	0.7434 ± 0.05	**0.7723 ± 0.04**	0.7626 ± 0.01
DS11	0.4505 ± 0.03	0.6196 ± 0.03	0.6208 ± 0.04	0.6213 ± 0.10	0.5924 ± 0.05	0.6487 ± 0.06	0.6512 ± 0.07	**0.6579 ± 0.02**
DS12	0.5997 ± 0.09	0.6789 ± 0.06	0.6763 ± 0.05	0.6781 ± 0.05	0.6722 ± 0.05	0.6898 ± 0.06	0.6846 ± 0.06	**0.6931 ± 0.03**
DS13	0.9064 ± 0.01	0.9098 ± 0.04	0.9056 ± 0.05	0.9043 ± 0.08	0.7970 ± 0.07	**0.9127 ± 0.07**	0.9106 ± 0.10	0.9115 ± 0.04
DS14	0.6275 ± 0.04	0.7037 ± 0.03	0.7121 ± 0.04	0.6867 ± 0.05	**0.7211 ± 0.04**	0.6965 ± 0.05	0.7197 ± 0.06	0.7187 ± 0.01
DS15	0.8567 ± 0.03	0.8995 ± 0.07	0.8981 ± 0.03	0.9007 ± 0.05	0.9164 ± 0.05	0.9195 ± 0.06	0.9251 ± 0.05	**0.9353 ± 0.02**
DS16	0.7379 ± 0.03	0.7663 ± 0.03	0.7640 ± 0.05	0.7567 ± 0.08	0.7886 ± 0.07	0.7853 ± 0.07	0.7959 ± 0.10	**0.7988 ± 0.04**
DS17	0.6917 ± 0.03	0.6958 ± 0.03	0.6635 ± 0.05	0.6530 ± 0.04	0.7036 ± 0.04	0.7074 ± 0.05	0.7145 ± 0.05	**0.7265 ± 0.00**
DS18	0.7752 ± 0.10	0.8657 ± 0.04	0.8558 ± 0.04	0.8592 ± 0.04	0.8733 ± 0.04	0.7627 ± 0.05	0.8825 ± 0.05	**0.8905 ± 0.01**
DS19	0.5837 ± 0.01	0.6864 ± 0.03	0.6627 ± 0.04	0.6783 ± 0.05	0.6910 ± 0.04	0.6931 ± 0.05	0.7037 ± 0.06	**0.7108 ± 0.01**
DS20	0.8797 ± 0.02	0.9067 ± 0.08	0.9030 ± 0.12	0.9036 ± 0.05	0.9175 ± 0.05	0.9234 ± 0.06	0.9274 ± 0.05	**0.9345 ± 0.03**
DS21	0.8610 ± 0.06	0.8709 ± 0.15	0.8653 ± 0.06	0.8671 ± 0.13	0.8871 ± 0.08	0.8926 ± 0.13	0.9064 ± 0.15	**0.9119 ± 0.14**
DS22	0.8647 ± 0.05	0.8991 ± 0.10	0.8955 ± 0.08	0.8966 ± 0.08	0.9183 ± 0.13	0.9189 ± 0.09	0.9225 ± 0.09	**0.9290 ± 0.07**
DS23	0.8964 ± 0.03	0.8998 ± 0.04	0.8956 ± 0.04	0.8922 ± 0.06	0.9114 ± 0.05	0.9107 ± 0.06	0.9265 ± 0.07	**0.9406 ± 0.02**
DS24	0.4410 ± 0.03	0.5019 ± 0.06	0.4241 ± 0.05	0.4600 ± 0.05	0.5017 ± 0.11	**0.5234 ± 0.06**	0.5165 ± 0.06	0.5178 ± 0.03
DS25	0.5941 ± 0.02	0.6051 ± 0.11	0.5606 ± 0.09	0.5980 ± 0.10	0.6021 ± 0.10	0.6097 ± 0.10	0.6190 ± 0.11	**0.6193 ± 0.09**
Average	0.7036 ± 0.06	0.7789 ± 0.05	0.7657 ± 0.05	0.7693 ± 0.06	0.7848 ± 0.06	0.7915 ± 0.07	0.8046 ± 0.08	**0.8102 ± 0.04**

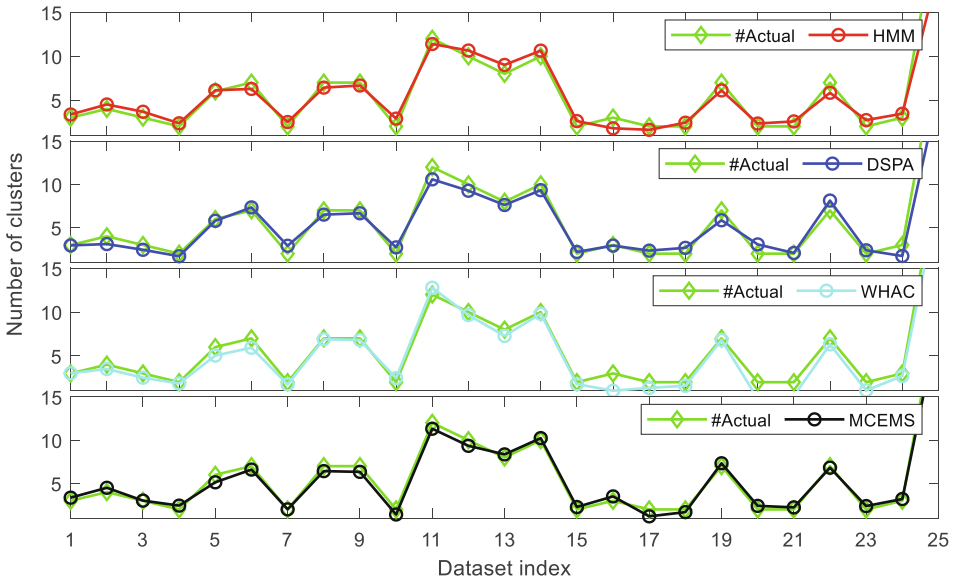

Fig. 5. Comparison of MCEMS with HMM, DSPA and WHAC algorithms in determining the number of optimal clusters.

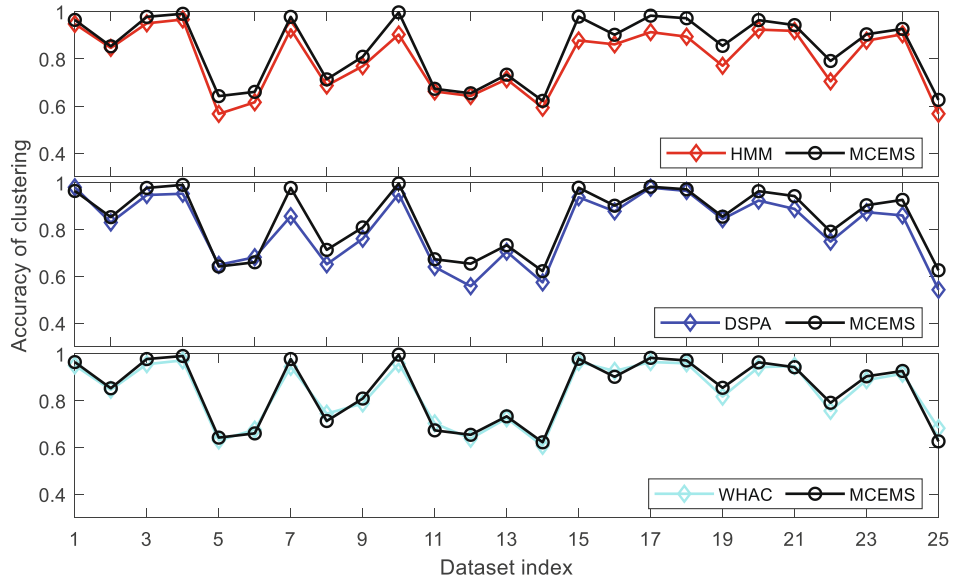

Fig. 6. Comparison of MCEMS with HMM, DSPA and WHAC algorithms in clustering accuracy.

Table 10
Comparison of MCEMS with HMM, DSPA and WHAC algorithms in different descriptors.

Descriptor method	Single	Average	Centroid	Complete	HMM	DSPA	WHAC	MCEMS
PMD	0.7020 ± 0.14	0.7709 ± 0.11	0.7611 ± 0.08	0.7654 ± 0.11	0.7779 ± 0.06	0.7863 ± 0.10	0.8012 ± 0.12	**0.8064 ± 0.05**
SMD	0.7004 ± 0.08	0.7664 ± 0.09	0.7623 ± 0.10	0.7639 ± 0.07	0.7807 ± 0.07	0.7884 ± 0.08	0.8022 ± 0.11	**0.8070 ± 0.04**
CMD	0.6984 ± 0.11	0.7754 ± 0.06	0.7641 ± 0.06	0.7614 ± 0.12	0.7811 ± 0.05	0.7890 ± 0.11	0.7994 ± 0.07	**0.8065 ± 0.07**
CD	0.7036 ± 0.06	0.7789 ± 0.05	0.7657 ± 0.05	0.7693 ± 0.06	0.7848 ± 0.06	0.7915 ± 0.07	0.8046 ± 0.08	**0.8102 ± 0.04**
MNED	0.7011 ± 0.08	0.7780 ± 0.07	0.7649 ± 0.05	0.7686 ± 0.08	0.7803 ± 0.05	0.7886 ± 0.11	0.8018 ± 0.08	**0.8085 ± 0.05**
Average	0.7001 ± 0.09	0.7739 ± 0.08	0.7636 ± 0.07	0.7657 ± 0.09	0.7810 ± 0.06	0.7888 ± 0.09	0.8019 ± 0.09	**0.8077 ± 0.05**

the number of correctly clustered instances to the total number of instances. As illustrated, MCEMS offers higher accuracy for most datasets than other algorithms. On average, MCEMS has 4.6%, 3.5% and 1.1% superiority for data clustering compared to HMM, DSPA and WHAC, respectively.

Evaluations of some structural descriptors (i.e., PMD, SMD, CMD, CD and MNED) for use in model selection and final dendrogram generation are presented in Table 10. This comparison presents the average CPC coefficient results for all datasets. Each row of the table is assigned a descriptor method, and the last

row shows the average results for all structural descriptors. Highlight values represent the best results for each dataset. The results show that the CD descriptor has the best performance for all algorithms. However, MCEMS performs better on all structural descriptors than all algorithms. However, it has been observed that in some datasets WHAC with CD descriptors reports better results. These results show that MCEMS is independent of descriptor methods and has promising performance with any descriptor method. In general, the MCEMS scheme is 15.4%, 4.4%, 5.8% and 5.5% superior to single linkage, average linkage, centroid linkage and complete linkage, respectively. This superiority over HMM, DSPA and WHAC algorithms is reported to be 3.4%, 2.4% and 0.72%, respectively.

The results of comparing the proposed MCEMS scheme with other algorithms based on Wilcoxon rank sum test and CPC coefficient confirmed the superiority of MCEMS as an ensemble clustering framework. In general, MCEMS provides similarity values in the clustering process with greater correlation than all individual AHC-based methods. This represents appropriate descriptions of the relationships between the data by novel similarity measurement proposed. On the other hand, the results clearly show that the dimensions of the dataset (number of instances, number of features and number of classes) have a relatively small effect on the effectiveness of MCEMS for different data clustering. In addition, MCEMS works well with various structural descriptors, which means that the proposed scheme is independent of descriptors.

6. Conclusion and future work

Clustering is one of the most important challenges in data analysis. Although many clustering methods have been studied in the subject literature, but there is no specific method that works better for all problems. In recent years, many efforts have been made to reduce the sensitivity of clustering to problem type, which can be referred to ensemble clustering. On the other hand, HC-based methods are more complex than PC-based methods, yet they provide higher quality clusters. Therefore, HC-based ensemble clustering approaches can provide good effectiveness but suffer from computational complexity problems. In this paper, reducing the complexity of HC-based ensemble clustering approaches through model selection methods is proposed. We proposed a meta-clustering ensemble scheme based on model selection called MCEMS. MCEMS uses multiple individual AHC methods to construct the ensemble clustering approach, where the bi-weighting policy is used to select the model with considering the quality and diversity. In addition, MCEMS has a novel similarity measurement that emphasizes observations on all primary clusters created by AHC methods. MCEMS solves the consensus function challenge through meta-clusters. Here, meta-clusters are created by re-clustering the primary clusters and then the final clusters are formed by assigning each instance to a meta-cluster with the highest similarity. The use of clusters clustering technique has caused MCEMS to consider the knowledge of different clustering methods in determining the cluster of each instance. In addition, MCEMS has overcome the computational complexity of AHC-based ensemble approaches by selecting a subset of AHC methods. Another advantage of MCEMS is the automatic determination of the number of clusters by considering a threshold and merging similar clusters, which can determine the optimal level on the dendrogram. The simulation results based on evaluation criteria and various statistical tests proved the superiority of the proposed algorithm over state-of-the-art clustering ensemble algorithms. Experiments showed that MCEMS is not dependent on the dimensions of the dataset and works well on various descriptive methods. Addressing the influence of reliability in model selection for ensemble clustering configuration can be considered as future work. Moreover, analysis of instance selection and feature selection on MCEMS performance can also be explored in future work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by Chongqing Municipal Education Commission grant number KJ QN201903110.

References

Aslanpour, M.S., Dashti, S.E., Ghobaei-Arani, M., Rahmanian, A.A., 2018. Resource provisioning for cloud applications: a 3-D, provident and flexible approach. J. Supercomput. 74 (12), 6470–6501.
Bagherinia, A., Minaei-Bidgoli, B., Hossinzadeh, M., Parvin, H., 2019. Elite fuzzy clustering ensemble based on clustering diversity and quality measures. Appl. Intel. 49 (5), 1724–1747.
Banerjee, A., Pujari, A.K., Rani Panigrahi, C., Pati, B., Chandan Nayak, S., Weng, T.H., 2021. A new method for weighted ensemble clustering and coupled ensemble selection. Connect. Sci. 33 (3), 623–644.
Beauchemin, M., 2015. A density-based similarity matrix construction for spectral clustering. Neurocomputing 151, 835–844.
Berahmand, K., Mohammadi, M., Faroughi, A., Mohammadiani, R.P., 2022. A novel method of spectral clustering in attributed networks by constructing parameter-free affinity matrix. Cluster Comput. 25, 869–888.
Berahmand, K., Nasiri, E., Pir mohammadiani, R., Li, Y., 2021. Spectral clustering on protein-protein interaction networks via constructing affinity matrix using attributed graph embedding. Comput. Biol. Med. 138, 104933.
Berikov, V., Pestunov, I., 2017. Ensemble clustering based on weighted co-association matrices: error bound and convergence properties. Pattern Recogn. 63, 427–436.
Berkhin, P., 2006. A survey of clustering data mining techniques. In: Kogan, J., Nicholas, C., Teboulle, M. (Eds.), Grouping Multidimensional Data. Springer-Verlag, Berlin/Heidelberg, pp. 25–71.
Boongoen, T., Iam-On, N., 2018. Cluster ensembles: a survey of approaches with recent extensions and applications. Comput. Sci. Rev. 28, 1–25.
Davidson, I., Ravi, S. S. (2005). Agglomerative hierarchical clustering with constraints: Theoretical and empirical results. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 59-70). Springer, Berlin, Heidelberg.
Do, M.N., Vetterli, M., 2002. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. IEEE Trans. Image Process. 11 (2), 146–158.
Ester, M., Kriegel, H. P., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).
Fernández, A., Gómez, S., 2008. Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. J. Classif. 25 (1), 43–65.
Fred, A.L., Jain, A.K., 2005. Combining multiple clusterings using evidence accumulation. IEEE Trans. Pattern Anal. Mach. Intell. 27 (6), 835–850.
Guénoche, A., Hansen, P., Jaumard, B., 1991. Efficient algorithms for divisive hierarchical clustering with the diameter criterion. J. Classif. 8 (1), 5–30.
Hamidi, S.S., Akbari, E., Motameni, H., 2019. Consensus clustering algorithm based on the automatic partitioning similarity graph. Data Knowl. Eng. 124, 101754.
Huang, D., Wang, C.D., Lai, J.H., 2017. LWMC: A locally weighted meta-clustering algorithm for ensemble clustering. In: International Conference on Neural Information Processing. Springer, Cham, pp. 167–176.
Huang, D., Wang, C.D., Lai, J.H., 2018. Locally weighted ensemble clustering. IEEE Trans. Cybern. 48 (5), 1460–1473.
Jafarzadegan, M., Safi-Esfahani, F., Beheshti, Z., 2019. Combining hierarchical clustering approaches using the PCA method. Expert Syst. Appl. 137, 1–10.
Jiang, D., Lin, W., Raghavan, N., 2021. A Gaussian mixture model clustering ensemble regressor for semiconductor manufacturing final test yield prediction. IEEE Access 9, 22253–22263.
Khedairia, S., Khadir, M.T., 2022. A multiple clustering combination approach based on iterative voting process. J. King Saud Univ.-Comput. Inf. Sci. 34 (1), 1370–1380.
Li, Y., Yu, J., Hao, P., Li, Z. (2007). Clustering ensembles based on normalized edges. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 664-671). Springer, Berlin, Heidelberg.
Maruyama, T., Kawachi, T., Singh, V.P., 2005. Entropy-based assessment and clustering of potential water resources availability. J. Hydrol. 309 (1–4), 104–113.
Mirzaei, A., Rahmati, M., Ahmadi, M., 2008. A new method for hierarchical clustering combination. Intell. Data Anal. 12 (6), 549–571.

Mojarad, M., Sarhangnia, F., Rezaeipanah, A., Parvin, H., Nejatian, S., 2021. Modeling hereditary disease behavior using an innovative similarity criterion and ensemble clustering. Curr. Bioinform. 16 (5), 749–764.
Murtagh, F., Contreras, P., 2012. Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Rev.: Data Min. Knowledge Discov. 2 (1), 86–97.
Nagpal, P.B., Mann, P.A., 2011. Comparative study of density based clustering algorithms. Int. J. Comput. Appl. 27 (11), 421–435.
Niu, H., Khozouie, N., Parvin, H., Alinejad-Rokny, H., Beheshti, A., Mahmoudi, M.R., 2020. An ensemble of locally reliable cluster solutions. Appl. Sci. 10 (5), 1891.
Pasupathi, S., Shanmuganathan, V., Madasamy, K., Yesudhas, H.R., Kim, M., 2021. Trend analysis using agglomerative hierarchical clustering approach for time series big data. J. Supercomput. 77 (7), 6505–6524.
Rai, P., Singh, S., 2010. A survey of clustering techniques. Int. J. Comput. Appl. 7 (12), 1–5.
Rezaeipanah, A., Amiri, P., Nazari, H., Mojarad, M., Parvin, H., 2021. An energy-aware hybrid approach for wireless sensor networks using re-clustering-based multi-hop routing. Wireless Pers. Commun. 120 (4), 3293–3314.
Sasirekha, K., Baby, P., 2013. Agglomerative hierarchical clustering algorithm-a. Int. J. Sci. Res. Publ. 83, 83.
Shahidinejad, A., Ghobaei-Arani, M., Masdari, M., 2021. Resource provisioning using workload clustering in cloud computing environment: a hybrid approach. Cluster Comput. 24 (1), 319–342.
Shakarami, A., Shahidinejad, A., Ghobaei-Arani, M., 2021. An autonomous computation offloading strategy in Mobile Edge Computing: a deep learning-based hybrid approach. J. Network Comput. Appl. 178, 102974.
Shi, P., Zhao, Z., Zhong, H., Shen, H., Ding, L., 2021. An improved agglomerative hierarchical clustering anomaly detection method for scientific data. Concurr. Comput.: Pract. Exp. 33 (6), e6077.
Struyf, A., Hubert, M., Rousseeuw, P., 1997. Clustering in an object-oriented environment. J. Stat. Softw. 1 (4), 1–30.
Wang, X., Yang, C., Zhou, J., 2009. Clustering aggregation by probability accumulation. Pattern Recogn. 42 (5), 668–675.
Xu, R., Wunsch, D., 2005. Survey of clustering algorithms. IEEE Trans. Neural Networks 16 (3), 645–678.
Yang, Y., Jiang, J., 2018a. Adaptive bi-weighting toward automatic initialization and model selection for HMM-based hybrid meta-clustering ensembles. IEEE Trans. Cybern. 49 (5), 1657–1668.
Yang, S.Z., Feng, Y.Y., Yeh, F.Y., 2007. Application of ordinal clustering to the taxonomy of the genus Entada (Fabaceae) in Taiwan. Bangladesh J. Plant Taxon. 14 (2), 93–100.
Yang, Y., Jiang, J., 2018b. Bi-weighted ensemble via HMM-based approaches for temporal data clustering. Pattern Recogn. 76, 391–403.
Yi, J., Yang, T., Jin, R., Jain, A.K., Mahdavi, M., 2012. Robust ensemble clustering by matrix completion. In: 2012 IEEE 12th International Conference on Data Mining. IEEE, pp. 1176–1181.
Yu, Z., Luo, P., Liu, J., Wong, H.S., You, J., Han, G., Zhang, J., 2018. Semi-supervised ensemble clustering based on selected constraint projection. IEEE Trans. Knowl. Data Eng. 30 (12), 2394–2407.
Zhou, Z.H., Wu, J., Tang, W., 2002. Ensembling neural networks: many could be better than all. Artif. Intell. 137 (1–2), 239–263.