An Ensemble Clusterer of Multiple Fuzzy k-Means Clusterings to Recognize Arbitrarily Shaped Clusters

Liang Bai, Jiye Liang, and Yike Guo

***Abstract*—Fuzzy cluster ensemble is an important research component of ensemble learning, which is used to aggregate several fuzzy base clusterings to generate a single output clustering with improved robustness and quality. However, since clustering is unsupervised, where "accuracy" does not have a clear meaning, it is difficult for existing ensemble methods to integrate multiple fuzzy k-means clusterings to find arbitrarily shaped clusters. To overcome the deficiency, we propose a new ensemble clusterer (algorithm) of multiple fuzzy k-means clusterings based on a local hypothesis. In the new algorithm, we study the extraction of local-credible memberships from a base clustering, the production of multiple base clusterings with different local-credible spaces, and the construction of cluster relation based on indirect overlap of local-credible spaces. The proposed ensemble clusterer not only inherits the scalability of fuzzy k-means but also overcomes the inability to find arbitrarily shaped clusters. We compare the proposed algorithm with other cluster ensemble algorithms on several synthetical and real datasets. The experimental results illustrate the effectiveness and efficiency of the proposed algorithm.**

***Index Terms*—Arbitrarily shaped clusters, fuzzy cluster ensemble, fuzzy k-means, local hypothesis.**

I. Introduction

CLUSTERING is an important problem in statistical multivariate analysis, data mining, and machine learning. The goal of clustering is to group a set of objects into clusters so that the objects in the same clusters are highly similar but remarkably dissimilar with objects in other clusters. To tackle this problem, various types of clustering algorithms have been developed in the literature (e.g., [1] and references therein), including partitional, hierarchical, density-based, grid-based clustering, etc. Among them, fuzzy k-means [2], [3] is one of the most computationally efficient clustering techniques, which is widely used to effectively solve many problems in real applications, such as image processing, automatic control, information retrieval, and bioinformatics. Its advantage is that it has linear time complexity and can deal with large-scale datasets. However, its disadvantage is that it is sensitive to the selection of initial points and can only find out spherical and uniform-sized clusters [4]. Currently, several complex clustering algorithms, such as spectral clustering [5], [6], density-based clustering [7], [8], and kernel clustering [9], have been developed to recognize arbitrarily shaped clusters. However, they need expensive time costs, i.e., the pairwise-objects distance calculations, which is not suitable for large-scale datasets. Therefore, it has been an urgent issue how to rapidly recognize different-shaped clusters.

In this paper, we wish to integrate multiple fuzzy k-means clusterings (MFKM) to quickly cluster datasets with different distributions, instead of complex algorithms. Cluster ensemble [10], [11] is a very popular technique to integrate several base clusterings into a final clustering with improved robustness and quality. Currently, there are various types of cluster ensemble methods, such as pairwise similarity, graph-based, relabeling-based, and feature-based methods [12]. Among them, some ensemble algorithms have been developed to integrate fuzzy clusterings. Su *et al.* [13] proposed link-based consensus methods for the ensemble of fuzzy k-means. Yu *et al.* [14] proposed a random double clustering based fuzzy cluster ensemble framework to perform tumor clustering based on gene expression data. Rathore *et al.* [15] proposed a fuzzy cluster ensemble framework based on random projection, which uses a cumulative agreement (voting) method to merge fuzzy base clusterings.

However, different from classifier ensemble, where the "accuracy" has a clear meaning, cluster ensemble is thought as an unsupervised ensemble learning [12]. It is very difficult for cluster ensemble to recognize the major strength and weakness of a base clustering on an unlabeled dataset [16]. Therefore, the ensemble objective of most existing cluster ensemble methods is to obtain the most consensus clustering with all the base clusterings. Their ensemble results strongly depend on the qualities of base clusterings. Thus, they cannot integrate multiple clusterings with low qualities into a good final clustering to realize "multiple weak clusterings equal to a strong clustering." To solve the problem, we propose a novel ensemble clusterer of MFKMs to simulate a complex clustering. We assume that a cluster center of a base clustering can well represent the objects in its neighborhood. Based on the assumption, we propose an evaluation function of membership credibility and an MFKM algorithm to produce multiple clusterings with different local-credible spaces. Furthermore, we construct a relation graph for all the clusters from base clusterings based on the indirect

Manuscript received November 29, 2017; revised March 21, 2018; accepted May 8, 2018. Date of publication May 11, 2018; date of current version November 29, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61773247, Grant 61432011, Grant 61573229, and Grant U1435212; in part by the Technology Research Development Projects of Shanxi (2015021100 and 201701D221097); and in part by the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi under Grant 2015107. *(Corresponding author: Jiye Liang.)*

L. Bai and J. Liang are with the School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China (e-mail: sxbailiang@hotmail.com; ljy@sxu.edu.cn).

Y. Guo is with the Department of Computing, Imperial College London, London SW7 2AZ, U.K. (e-mail: y.guo@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2018.2835774

overlap of their local-credible spaces. Finally, we determine the final clustering based on the membership credibility function and relation graph.

The outline of the rest of this paper is as follows. Section II reviews the related work of the cluster ensemble problem. Section III presents an ensemble clusterer of MFKMs. Section IV demonstrates the performance of the proposed ensemble clusterer. Section V concludes the paper with some remarks.

II. Related Work

Cluster ensemble, also called consensus clustering, is a kind of unsupervised ensemble learning. Generally speaking, cluster ensemble includes two major research tasks: 1) constructing a generator to produce a base clustering set, and 2) devising an ensemble strategy to produce the final partition. Their results affect the performance of a cluster ensemble method. In the following, we introduce the related work of the two tasks, respectively.

In ensemble learning, it is observed that the diversity among classification results of base classifiers or clusterers, to some extent, can enhance the performance of the ensemble learner. Currently, several heuristics have been proposed to produce different clusterings on a dataset, which can be classified into three categories.

1) Repeatedly run a single clustering algorithm with different parameters to produce base clusterings [17]–[19]. Fred and Jain [17] applied k-means with the different numbers of clusters to produce a clustering set. Kuncheva and Vetrov [18] used k-means with randomly selected different cluster centers. Liu *et al.* [19] aggregated multiple spectral clusterings with different kernel parameters.
2) Run different types of clustering algorithms to produce base clusterings [11], [20]. Gionis *et al.* [11] used several hierarchical clustering and k-means to produce a clustering set. Law *et al.* [20] applied multiple clustering algorithms with different objective functions as base clusterings and transformed a clustering ensemble problem as a multiobjective optimization.
3) Run one or more clustering algorithms on different subspaces or subsamples from a dataset [15], [21]–[23], [23]–[27]. Fischer and Buhmann [21] applied the bootstrap method to obtain several data subsets. Fern *et al.* [25] used the random projection method to obtain several feature subspaces. Yang and Jiang [27] proposed a novel hybrid sampling method for cluster ensemble by combining the strengths of boosting and bagging.

For ensemble strategy, there are several representative methods that can be classified into the following four categories.

1) The pairwise similarity approach that makes use of co-occurrence relationships between all pairs of data objects to aggregate multiple clusterings [13], [14], [28]–[31]. Fred and Jain [28] proposed an ensemble algorithm based on evidence accumulation and constructed a co-association (CO) matrix. Yang *et al.* [29] made use of clustering validity functions as weights to construct a weighted similarity matrix. Iam-On *et al.* [30], [31] defined a link-based similarity matrix that sufficiently considers the similarity between clusters. Su *et al.* [13] extended the link-based similarity matrix to deal with fuzzy clusterings. In the fuzzy cluster ensemble framework, Yu *et al.* [14] measure the label consistency between two objects on different subspace clusterings to construct the pairwise similarity matrix.
2) The graph-based approach that expresses the base clustering information as an undirected graph and, then, derives the ensemble clustering via graph partitioning [10], [32]–[34]. Strehl and Ghosh [10] proposed three hypergraph ensemble algorithms: cluster-based similarity partitioning algorithm (CSPA), hyper graph partitioning algorithm (HGPA), and metaclustering algorithm (MCLA). CSPA creates a similarity graph, where the vertices represent objects and the weights of edges represent similarity. HGPA constructs a hypergraph, where the vertices represent objects and the same weighted hyperedges represent clusters. MCLA generates a graph where the vertices represent clusters and the weights of edges reflect the similarity between clusters. Fern and Brodley [32] proposed the hybrid bipartite graph formulation algorithm where vertices represent both objects and clusters.
3) The relabeling-based approach that expresses the base clustering information as label vectors and, then, aggregates via label alignment [22], [23], [35], [37], [38]. Its representative methods can be classified into two types: crisp label correspondence and soft label correspondence. The crisp methods [22], [23], [37] transfer the relabeling problem into a minimum cost one-to-one assignment problem. Long *et al.* [38] used an alternating optimization strategy to solve the soft label alignment problem. Rathore *et al.* [15] proposed an efficient fuzzy ensemble framework that uses a cumulative agreement scheme to aggregate fuzzy clusters.
4) The feature-based approach that treats the problem of cluster ensemble as the clustering of categorical data [39]–[44]. Cristofor and Simovici [39] integrated the information theory and genetic algorithms to search for the most consensus clustering. Topchy *et al.* [40] proposed a probabilistic framework and used the EM algorithm for finding the consensus clustering. Nguyen *et al.* [43] made use of the k-modes [44] as the consensus function for cluster ensemble.

It is worth noting that the research objective of this paper is different from those of existing cluster ensemble algorithms. Most existing algorithms mainly focus on how to obtain the most consensus clustering from base clusterings, which can improve the clustering quality and robustness. However, these algorithms do not consider the credibility of memberships, which imposes difficulties at realizing “multiple weak clusterings equal to a strong clustering.” Therefore, this paper mainly study how to integrate multiple fuzzy k-means clusterings to rapidly recognize a complex clustering.

III. New Cluster Ensemble Algorithm

A. Cluster Ensemble Problem

Let $X = \{\mathbf{x}_i\}_{i=1}^N$ be a set of N objects, $\Pi = \{\pi_h\}_{h=1}^T$ be a set of T base clusterings, $\pi_h = \{C_{hl}\}_{l=1}^{k_h}$ be the hth base clustering, where k_h is the number of clusters and C_{hl} is the lth cluster in π_h, $\mathbb{W} = \{W_h\}_{h=1}^T$ be a set of membership matrices and $W_h = [w_{hli}]_{1\leq l\leq k_h, 1\leq i\leq N}$ be the membership matrix of the hth clustering, where w_{hli} is the membership of object $\mathbf{x}_i$ to cluster C_{hl}. $K = \{k_h\}_{h=1}^T$ be a set of the number of clusters in each base clustering. The cluster ensemble problem aims to find out a final clustering π^* of dataset X based on the clustering set Π.

In this paper, the fuzzy k-means algorithm is used as a base clusterer. Its objective function F is described as

$$F(W_h, V_h) = \sum_{l=1}^{k_h}\sum_{i=1}^{N} (w_{hli})^m \|\mathbf{x}_i - \mathbf{v}_{hl}\|^2$$

where $V_h = \{\mathbf{v}_{hl}\}_{l=1}^{k_h}$ and $\mathbf{v}_{hl}$ is the lth cluster center and $\sqrt{\|\mathbf{x}_i - \mathbf{v}_{hl}\|^2}$ is Euclidean distance between the object $\mathbf{x}_i$ and the center $\mathbf{v}_{hl}$ of the lth cluster. Fuzzy k-means makes use of alternatively updating W_h and V_h to solve the problem of minimizing F in finding cluster solutions. Its clustering results are often different, while it runs with different initial cluster centers. Therefore, we attempt to produce multiple base clusterings by fuzzy k-means and integrate them to rapidly generate a good clustering result on datasets. However, there are three important factors that often affect the effectiveness of cluster ensemble as follows:

1) *The membership credibility:* In a base clustering, there are some objects whose memberships may be correct. If these objects have consistently incorrect memberships in the base clusterings, these memberships are combined into the final clustering, which leads to reducing the effectiveness of ensemble. It is a key task for enhancing the ensemble effectiveness to provide an evaluation criterion for membership credibility.
2) *The difference among base clusterings:* In cluster ensemble, people wish base clusterings are different to some extent from each other. The ensemble learning uses the difference to find out a robust clustering result. If most base clusterings in Π are very similar, it is not worth integrating them. Thus, we wish to obtain multiple complementary clusterings of fuzzy k-means to adequately describe the entire data.
3) *The relation of clusters:* Unlike classification, each base clustering may have a different representation of labels. Thus, we need to judge which cluster labels represent the same clusters. Obtaining a good relation of clusters is the prerequisite to cluster ensemble. It is noted that the relation of clusters is different from that of most existing relabeling methods. Since the clusters from the same clustering also may represent the same cluster, we should reflect the relation of all the clusters from the same and different base clusterings.

In the following, we will propose an ensemble clusterer of MFKMs, which can fully consider these factors.

B. Membership Credibility Function

In fuzzy k-means, a cluster center is used to represent a cluster. However, if a cluster is nonlinearly separable with other clusters, the objects represented by a cluster center may come from different clusters. Take a clustering of fuzzy k-means shown in Fig. 1 for example. We can see that Cluster 1 consists of objects from different "true" clusters. Thus, the cluster center obtained by fuzzy k-means is not suitable to represent a nonlinear cluster. According to Fig. 1, we can also find that as the size of a local space represented by the cluster center is gradually reduced, the "true" cluster labels of objects in the local space are more consistent.

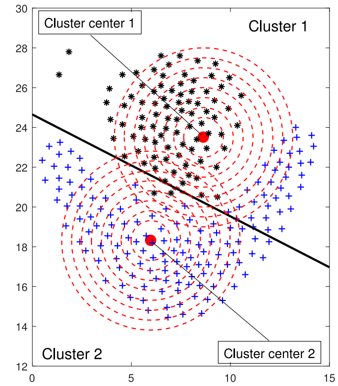


Fig. 1. Clustering of fuzzy k-means.

Therefore, we evaluate the credibility of a cluster membership based on a local hypothesis. We think that a cluster center can only represent the objects in its neighborhood space, and the membership credibility of an object to a cluster should be inversely proportional to the distance between the object and the cluster center.

Thus, we use the following equation to evaluate the membership credibility:

$$\frac{\exp\left(-||\mathbf{x}_i - \mathbf{v}_{hl}||^2\right)}{\sum_{j=1}^{N}\exp\left(-||\mathbf{x}_j - \mathbf{v}_{hl}||^2\right)} \tag{1}$$

which is the probability that object $\mathbf{x}_i$ is as the neighbor of $\mathbf{v}_{hl}$. The closer object $\mathbf{x}_i$ is to $\mathbf{v}_{hl}$, the higher the probability is. We use the probability to reflect the membership credibility of $\mathbf{x}_i$ to C_{hl}. The higher the probability is, the more possibly $\mathbf{v}_{hl}$ is representative of $\mathbf{x}_i$. Therefore, based on (1), the membership credibility function is defined as

$$\lambda_{hli} = \begin{cases} \frac{\exp\left(-||\mathbf{x}_i - \mathbf{v}_{hl}||^2\right)}{\sum_{j=1}^{N}\exp\left(-||\mathbf{x}_j - \mathbf{v}_{hl}||^2\right)}, & \text{if} \quad \mathbf{x}_i \in B(\mathbf{v}_{hl}) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $B(\mathbf{v}_{hl}) = \{\mathbf{x}_j \in X | \sqrt{||\mathbf{x}_j - \mathbf{v}_{hl}||^2} \leq \epsilon\}$ is the ϵ-neighborhood of the cluster center $\mathbf{v}_{hl}$, which is also called the local-credible space of the cluster C_{hl}, for $1 \leq i \leq N$ and $1 \leq h \leq T$. The definition shows that we only retain the membership information of the objects in the ϵ-neighborhood of a cluster center.

C. Production of Multiple Base Clusterings

To obtain MFKMs with different local-credible spaces, we extend the objective function of fuzzy k-means to define an optimization problem of producing base clusterings as follows:

$$\min_{\mathbb{W}} \left[Z(\mathbb{W}) = \sum_{h=1}^{T} \sum_{i=1}^{N} \theta_{hi} \sum_{l=1}^{k_h} \lambda_{hli} (w_{hli})^m \|\mathbf{x}_i - \mathbf{v}_{hl}\|^2 \right]. \tag{3}$$

In this function, except for λ, we introduce a new parameter $\theta_{hi} \in [0, 1]$, which is used to reflect the importance of $\mathbf{x}_i$ playing a role in producing the hth base clustering. The more the θ_{hi} value is, the more important the role of object $\mathbf{x}_i$ is. We hope that different clusterings are produced based on different distribution of θ.

We propose an incremental learning method to solve the optimization problem. The method gradually produces multiple base clusterings by trying to optimize an incremental problem at each stage. The incremental problem is described as follows. Given $\mathbb{W}'$ including the membership matrices of the first hth obtained base clusterings ($0 \le h < T$)

$$\min_{W_{h+1}} Z(\mathbb{W}' \cup \{W_{h+1}\}) \tag{4}$$

subject to

$$\theta_{h+1i} = \frac{\theta_{hi} \exp\left(-\max_{l=1}^{k_h} \lambda_{hli}\right)}{\sum_{j=1}^{N} \theta_{hj} \exp\left(-\max_{l=1}^{k_h} \lambda_{hlj}\right)} \tag{5}$$

for $1 \le i \le N$. According to (5), we see that the higher $\max_{l=1}^{k_h} \lambda_{hli}$ is, the lower θ_{h+1i} is. This means that its importance in producing next clustering is reduced, if object $\mathbf{x}_i$ has a high credible membership in the hth clustering. Such setting θ can help us to produce clusterings with different credible memberships.

The incremental learning method, called the MFKM algorithm, is described in Algorithm 1. In the method, we initially set $h = 1, \theta_{hi} = \frac{1}{N}$ for $1 \le i \le N$ and $S = X$, which is used to store objects whose credible memberships are equal to zero at the obtained clusterings. At the hth stage, we randomly select k_h objects as initial cluster centers from S and apply fuzzy k-means with a new updated formula of V_h to cluster the dataset. In this case, the cluster centers are updated by only considering the objects in their ϵ-neighborhoods, which makes the final obtained cluster centers better represent the objects in their local-credible spaces. After fuzzy k-means runs, we update $S = S - S'$, where S' is a set of the objects whose maximum of local-credible memberships is more than 0 in the hth base clustering. Furthermore, we update θ_{h+1i} for $1 \le i \le N$. The above-mentioned procedure is repeated until the number of the objects in S is less than k or the number of base clusterings is equal to T, which is the desired number of base clusterings. If $|S| < k_h$, we cannot select k_h initial cluster centers. In this case, the number of obtained clusterings may be less than T. The incremental procedure makes the cluster centers obtained at each time represent different local-credible spaces.

The time complexity of the MFKM algorithm is $O(N \sum_{h=1}^{T} t_h k_h)$, where t_h is the number of iterations of fuzzy k-means in the process of producing the hth base clustering and T is the number of the produced base clusterings. The outputs of the algorithm are membership matrices $\mathbb{W} = \{W_h, 1 \le h \le T\}$ and cluster center sets $\mathbb{V} = \{V_h, 1 \le h \le T\}$ of base clusterings.

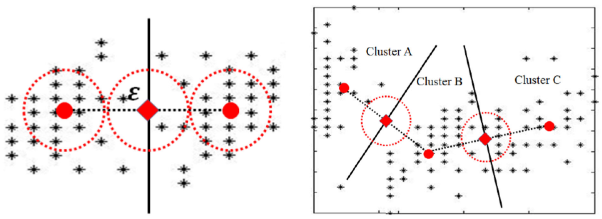
Algorithm 1 The MFKM Algorithm.

Input: X, K, ϵ, T
Output: $\mathbb{W}, \mathbb{V}$
Initialize $\Pi = \emptyset$, $V = \emptyset$, $h = 1$, $S = X$ and $\theta_{1i} = \frac{1}{N}$ for $1 \le i \le N$;
while $h \le T$ **do**
  if $|S| < k_h$ **then**
    Break;
  V_h is made up of randomly selected k_h objects on S;
  Compute λ_{hli} for $1 \le l \le k_h$ and $1 \le i \le N$;
  while $F < F'$ **do**
    $F' = F$;
    for $i = 1 : N$ **do**
      for $l = 1 : k_h$ **do**
        $w_{hli} = \frac{1}{\sum_{f=1}^{k_h} \left[\frac{\|\mathbf{x}_i - \mathbf{v}_{hl}\|^2}{\|\mathbf{x}_i - \mathbf{v}_{hf}\|^2} \right]^{1/(m-1)}}$;
    for $l = 1 : k_h$ **do**
      $\mathbf{v}_{hl} = \frac{\sum_{i=1}^{N} \theta_{hi} \lambda_{hli} (w_{hli})^m \mathbf{x}_i}{\sum_{i=1}^{N} \theta_{hi} \lambda_{hli} (w_{hli})^m}$;
    Update λ_{hli} for $1 \le l \le k_h$ and $1 \le i \le N$;
    $F = \sum_{i=1}^{N} \theta_{hi} \sum_{l=1}^{k_h} \lambda_{hli} (w_{hli})^m \|\mathbf{x}_i - \mathbf{v}_{hl}\|^2$;
  Update $S = S - \{\mathbf{x}_i \in S | \max_{l=1}^{k_h} \lambda_{hli} > 0\}$, θ_{h+1i} for $1 \le i \le N$, $\mathbb{W} = \mathbb{W} \bigcup \{W_h\}$, $\mathbb{V} = \mathbb{V} \bigcup \{V_h\}$, and $h = h + 1$;

D. Construction of Cluster Relation

Unlike classification where the labels represent specific classes, the cluster labels only express grouping characteristics of the data and cannot be directly comparable across different clusterings in cluster analysis. Therefore, in cluster ensemble, the labels of different clusterings should be aligned. Besides, since the fuzzy k-means algorithm only can recognize linearly separable clusters, two clusters from a base clustering may represent the same cluster. Therefore, we need to analyze the relation of all the clusters from base clusterings.

Currently, there are several similarity or dissimilarity measures between clusters proposed in existing cluster ensemble algorithms [12]. Among these measures, the degree of overlap between two clusters, i.e., the number of their common objects, is widely used to reflect their similarity, which can be seen in the graph-based algorithms proposed by Strehl *et al.* [10] and the relabeling-based algorithms proposed by Zhou *et al.* [23]. However, this measure cannot be used to evaluate the similarity between clusters from the same clusterings, since they have no common objects. To solve the problem, Iam-On *et al.* [30] proposed a link-based similarity measure between clusters, which compares the overlap of them with other clusters. Although these existing measures already have good practical contributions, they do not consider the credibility of cluster memberships that may affect the performance of these measures. Therefore, we need to design a new similarity measure to overcome the shortcoming.



(a) (b)

Fig. 2. (a) Latent cluster between clusters. (b) Similarity between clusters.

According to the MFKM algorithm, we know that the produced base clusterings Π are with different local-credible spaces. Thus, we hope to measure the overlap between the local-credible spaces of two clusters to reflect their similarity. Let C_{hl} and C_{pq} be two clusters, $\mathbf{v}_{hl}$ and $\mathbf{v}_{pq}$ be their cluster centers. If $\sqrt{||\mathbf{v}_{hl} - \mathbf{v}_{pq}||^2}$ is no more than 2ϵ, their local-credible spaces are overlapped. However, for any two clusters, the overlap of their local-credible spaces is generally small or null, due to the producing mechanism of the base clusterings by the MFKM algorithm. Therefore, we introduce a latent cluster to evaluate their "indirect" overlap. Let $\frac{\mathbf{v}_{hl}+\mathbf{v}_{pq}}{2}$ be the midpoint of the two centers $\mathbf{v}_{hl}$ and $\mathbf{v}_{pq}$. We assume there is a latent cluster $C_{(hl,pq)}$ whose cluster center is $\frac{\mathbf{v}_{hl}+\mathbf{v}_{pq}}{2}$. If $\sqrt{||\mathbf{v}_{hl} - \mathbf{v}_{pq}||^2}$ is no more than 4ϵ, the local-credible spaces of both the clusters C_{hl} and C_{pq} are overlapped with that of the latent cluster $C_{(hl,pq)}$, which can be seen in Fig. 2(a). In this case, we define that the local-credible spaces of C_{hl} and C_{pq} are indirectly overlapped with respect to the latent cluster.

We consider the following two factors to measure the "indirect" overlap between the local-credible spaces of clusters C_{hl} and C_{pq} as follows:

1) The similarity between their cluster centers.
2) The density in the local-credible space of their latent cluster.

The similarity between two cluster centers is defined as

$$\sigma_{(hl,pq)} = \begin{cases} \exp\left(-||\mathbf{v}_{hl} - \mathbf{v}_{pq}||^2\right), & \text{if } \sqrt{||\mathbf{v}_{hl} - \mathbf{v}_{pq}||^2} \leq 4\epsilon \\ 0, & \text{otherwise} \end{cases}. \tag{6}$$

We know that the smaller $||\mathbf{v}_{hl} - \mathbf{v}_{pq}||^2$ is, the more overlapped the local-credible spaces between them and $C_{(hl,pq)}$ are. Therefore, we think their "indirect" overlap should be proportional to $\sigma_{(hl,pj)}$. Besides, since the fuzzy k-means algorithm is a linear clusterer, the spaces of any two clusters are separated by the midline between their cluster centers. If the surrounding area of their midpoint includes few objects, they can be clearly distinguished. Let us consider an example in Fig. 2(b). We see that the center distance between clusters A and B is equal to that between clusters B and C. However, we find out that the boundary between clusters A and B is clearer than that between clusters B and C. Thus, if the clarity of the boundary between clusters is considered, clusters A and B are better separated than clusters B and C. Therefore, we think that the "indirect" overlap of two clusters should be proportional to the density in the local-credible space of their latent cluster. In this case, we use the sum of membership credibility of each object to the latent cluster to reflect its density. The density in the local-credible space of $C_{(hl,pq)}$ is defined as

$$\rho_{(hl,pq)} = \begin{cases} \sum_{i=1}^{N} \lambda_{(hl,pq)i}, & \text{if } \sqrt{||\mathbf{v}_{hl} - \mathbf{v}_{pq}||^2} \leq 4\epsilon \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $\lambda_{(hl,pq)i}$ is the membership credibility of object $\mathbf{x}_i$ to cluster $C_{(hl,pq)}$, which can be computed by (2). Therefore, we integrate σ and ρ to define the similarity measure for two clusters as follows:

$$\delta_{(hl,pq)} = \left(\frac{\sigma_{(hl,pq)} - \min \sigma}{\max \sigma - \min \sigma}\right)\left(\frac{\rho_{(hl,pq)} - \min \rho}{\max \rho - \min \rho}\right). \tag{8}$$

According to the definition, we see that the similarity measure is the product of the normalizations of σ and ρ. Based on the similarity measure, we construct an undirected and weighted graph $G \leq A, \Delta >$ to reflect the relation of these clusters. In the graph G, $A = \{hl\}_{1\leq h\leq T, 1\leq l\leq k_h}$ is a set of vertices each representing a cluster label from Π. Thus, A is also seen as a set of all the cluster labels in Π. $\Delta = \{\delta_{(hl,pq)}\}_{1\leq h,g\leq T, 1\leq l,j\leq k_h}$ is a weight set of edges between clusters. For any two clusters, we use their similarity as the weight of the edge between them. The higher similarity they have, the more possibly they represent the same cluster.

After the weighted graph is obtained, the problem of constructing a cluster relation can be transferred to a normalized graph cut problem, which is described as follows [5]:

$$\min_{\Omega} \left[Q(\Omega) = \frac{1}{k} \sum_{j=1}^{k} \frac{\sum_{hl\in A_j, pq\in A-A_j} \delta_{(hl,pq)}}{\sum_{hl\in A_j, pq\in A} \delta_{(hl,pq)}} \right] \tag{9}$$

where $\Omega = \{A_j\}_{j=1}^{k}$ is a partition of vertices in the graph G and A_j is the jth subset of A. We wish to obtain such a partition by minimizing the objective function Q that the vertices in the same subsets have very high similarity but are very dissimilar with vertices in other subsets. In order to solve the optimization problem, we apply the normalized spectral clustering (NSC) algorithm [6] to obtain a final partition of A. The vertices in the same subsets are used to represent a cluster. Thus, let $L(C_{hl})$ be the label of the subset that C_{hl} belongs to, we have

$$L(C_{hl}) = j, \text{ if } hl \in A_j \tag{10}$$

for $1 \leq j \leq k$. The time complexity of constructing cluster relation is $O(N(\sum_{h=1}^{T} k_h)^2)$. Let us consider the example of the dataset Flame to show a procedure of constructing cluster relation. The MFKM algorithm produces 12 clusters. Fig. 3(a) shows their relation graph. We employ the NSC algorithm to obtain a min-cut of this graph, which is shown in Fig. 3(b). All the clusters in each subgraph are used to represent the same cluster.

After relabeling the clusters from base clusterings, the membership matrix W^* of the final clustering π^* is obtained as follows:

$$w^*_{ji} = \frac{\sum_{hl\in A_j} \lambda_{hli} w_{hli}}{\sum_{pq\in A} \lambda_{pqi} w_{pqi}} \tag{11}$$

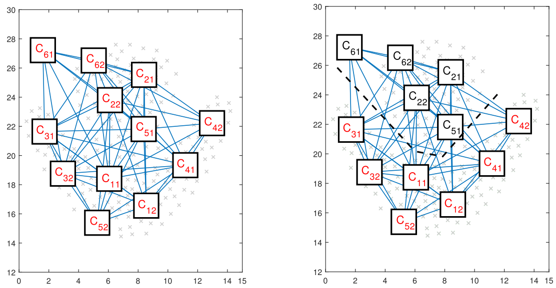


Fig. 3. Example about a procedure of constructing cluster relation. (a) Graph of cluster relation. (b) Min-cut of graph.

Algorithm 2 The FKMCE Algorithm.

Input: X, k, K, ϵ, T
Output: π^*
$\mathbb{W} = \arg\min Z(\mathbb{W})$ by Algorithm 1;
Compute λ_{hli} by (2), for $1 \leq h \leq T$, $1 \leq l \leq k_h$, $1 \leq i \leq N$;
A = a set including all the cluster labels in Π;
for $hl, pq \in A$ **do**
  Compute $\delta_{(hl,pq)}$ by (8);
Obtain a graph of cluster relation $G \leq A, \Delta >$;
$\Omega = \arg\min Q(\Omega)$ by the NSC algorithm;
Obtain the final membership matrix W^* by (11);
Obtain the final clustering π^* by (12);

for $1 \leq i \leq N$ and $1 \leq j \leq k$. Given W^*, we can obtain the final clustering as follows:

$$\pi^*(\mathbf{x}_i) = \arg\max_{j=1}^{k} w^*_{ji} \tag{12}$$

for $1 \leq i \leq N$. The time complexity of generating the final clustering is $O(NT)$.

E. Overall Implementation

We integrate the above steps to form a new multiple fuzzy k-means clustering ensemble (FKMCE) algorithm. This algorithm is described in Algorithm 2. The overall time complexity of the FKMCE algorithm is $O(N\sum_{h=1}^{T} t_h k_h + N\sum_{h=1}^{T} k_h + N(\sum_{h=1}^{T} k_h)^2 + NT)$. We see that the time complexity is linear with the number of objects. Generally, $\left(\sum_{h=1}^{T} k_h\right)^2 \ll N$. In this case, the time complexity is less than $O(N^2)$. We know that the time complexities of most existing complex clustering algorithms are no less than $O(N^2)$. This indicates that the FKMCE algorithm is suitable to deal with large-scale datasets, compared to existing complex clustering algorithms.

IV. Experimental Analysis

In this section, we carry out the FKMCE algorithm on eight synthetic and five real datasets to illustrate its effectiveness and efficiency.

TABLE I
Description of Datasets: Number of Data Objects (N), Number of Dimensions (D), Number of Clusters (k)

	Data set	N	D	k
Synthetic data	Ring [48]	1,500	2	3
	Jain [47][48]	373	2	2
	Flame [47][48]	240	2	2
	Agg [47][48]	788	2	7
	T4.8k [47][48]	7,235	2	6
	T7.1k [48]	3,031	2	9
	Chain [48]	1,000	3	2
	Atom [48]	800	3	2
Real data	Iris [47][49]	150	4	3
	Wine [47][49]	178	13	3
	Breast [47][49]	569	30	2
	Digits [49]	5,620	63	10
	Statlog [49]	6,435	36	7

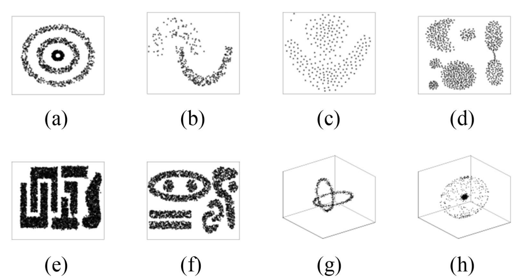


Fig. 4. Data distribution of synthetic data. (a) Ring. (b) Jain. (c) Flame. (d) Agg. (e) T4.8k. (f) T7.1k. (g) Chain. (h) Atom.

TABLE II
Notation for the Contingency Table for Comparing Two Partitions

$C\backslash P$	p_1	p_2	$\cdots$	$p_{k'}$	$Sums$
c_1	n_{11}	n_{12}	$\cdots$	$n_{1k'}$	b_1
c_2	n_{21}	n_{22}	$\cdots$	$n_{2k'}$	b_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
c_k	n_{k1}	n_{k2}	$\cdots$	$n_{kk'}$	b_k
$Sums$	d_1	d_2	$\cdots$	$d_{k'}$	

A. Datasets

Table I shows the details of these tested datasets. The data distributions of the synthetic datasets are shown in Fig. 4. These sets can be downloaded from [47]–[49].

B. Evaluation Criteria

We employ the two widely-used external criteria adjusted rand index (ARI) [50] and normalized mutual information (NMI) [51] to measure the similarity between the clustering result and the true partition on a dataset. Given a dataset X with N objects and two partitions of these objects, namely $C = \{c_1, c_2, \ldots, c_k\}$ (the clustering result) and $P = \{p_1, p_2, \ldots, p_{k'}\}$ (the true partition), the overlappings between C and P can be summarized in a contingency table (see Table II), where n_{ij} denotes the number of common nodes of groups c_i and p_j: $n_{ij} = |c_i \cap p_j|$. The ARI [50] is defined as

$$\mathrm{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{b_i}{2}\sum_j \binom{d_j}{2}]/\binom{N}{2}}{\frac{1}{2}[\sum_i \binom{b_i}{2} + \sum_j \binom{d_j}{2}] - [\sum_i \binom{b_i}{2}\sum_j \binom{d_j}{2}]/\binom{N}{2}}$$

TABLE III
ARI MEASURES OF DIFFERENT METHODS

Methods	Synthetic data sets								Real data sets				
	Ring	Jain	Flame	Agg	T4.8k	T7.1k	Chain	Atom	Iris	Wine	Breast	Digits	Statog
CO-AL	0.1305	0.5853	0.4880	0.6245	0.5098	0.3726	0.0927	0.1456	0.7302	0.8471	0.7302	0.6050	0.5700
WCT-AL	0.1382	0.5853	0.4880	0.7342	0.4952	0.3635	0.0927	0.1456	0.7302	0.8471	0.7302	0.6046	0.5699
WTQ-AL	0.1389	0.5853	0.4880	0.7081	0.3326	0.3705	0.0927	0.1456	0.7302	0.8471	0.7302	0.6049	0.5699
CSM-AL	0.1448	0.5853	0.4880	0.7192	0.4956	0.4199	0.0927	0.1456	0.7302	0.8471	0.7302	0.6146	0.5699
FLink-AL	0.0009	0.5853	0.4880	0.8205	0.4707	0.3422	0.0903	0.2321	0.7149	0.8498	0.7305	0.2888	0.5086
FCTS-AL	0.0016	0.5853	0.4880	0.6131	0.4796	0.3280	0.0903	0.1672	0.7149	0.8498	0.7305	0.2609	0.3891
CSPA	0.3163	0.2774	0.4312	0.5365	0.5010	0.3418	0.0927	0.0021	0.6521	0.7808	0.3414	0.7573	0.4329
HGPA	0.0004	0.0021	0.0038	0.3621	0.4012	0.1966	0.0010	0.0013	0.1026	0.1286	0.0007	0.3750	0.2619
MCLA	0.0004	0.5853	0.4880	0.5778	0.5018	0.3736	0.0927	0.1554	0.7302	0.8471	0.7302	0.6935	0.5127
SV	0.0847	0.5853	0.4763	0.3343	0.2443	0.1406	0.1002	0.1736	0.0067	0.8685	0.7302	0.3244	0.4533
SWV	0.1809	0.5853	0.4763	0.4612	0.2621	0.1966	0.1002	0.1736	0.0002	0.8685	0.7302	0.4641	0.4546
CAFCM	0.0440	0.5853	0.4880	0.7241	0.4272	0.4204	0.0915	0.2827	0.7149	0.8498	0.7305	0.2121	0.5248
EM	0.0302	0.5151	0.4164	0.5682	0.4775	0.3240	0.0896	0.2617	0.6008	0.7855	0.6328	0.6205	0.5074
IVC	0.3231	0.1288	0.3708	0.5783	0.4894	0.4097	0.0927	0.1178	0.5970	0.6875	0.0487	0.6006	0.4188
NSC	**1.0000**	**1.0000**	0.8382	0.9045	0.9260	0.9848	**1.0000**	**1.0000**	0.7455	**0.9310**	0.7493	0.7536	0.5308
DBSCAN	**1.0000**	0.2824	0.2270	0.6294	0.7780	0.8513	0.4947	0.3786	0.5162	0.3587	0.0478	0.5052	0.4319
CFSFDP	0.3227	0.6438	0.9337	0.9898	0.6098	0.8043	0.6853	0.4154	0.7028	0.7414	0.7305	0.7584	0.4963
FKMCE	**1.0000**	**1.0000**	**0.9539**	**0.9909**	**0.9786**	**0.9891**	**1.0000**	**1.0000**	**0.8296**	0.8834	**0.7700**	**0.8430**	**0.6544**

where n_{ij}, b_i, and d_j are values from the contingency table (see Table II). The NMI [51] is defined as

$$\mathrm{NMI} = \frac{2\sum_{\mathrm{i}}\sum_{\mathrm{j}} \mathrm{n_{ij}} \log \frac{\mathrm{n_{ij}N}}{\mathrm{b_i d_j}}}{-\sum_{\mathrm{i}} \mathrm{b_i} \log \frac{\mathrm{b_i}}{\mathrm{N}} - \sum_{\mathrm{j}} \mathrm{d_j} \log \frac{\mathrm{d_j}}{\mathrm{N}}}.$$

If a clustering result is close to the true partition, then its ARI and NMI values are high.

C. Compared Methods

In order to properly examine the performance of the proposed algorithm, we compare it with the following cluster ensemble algorithms. The codes of these compared algorithms are open and accessible, which can be found from the personal homepage of these authors.

1) Pairwise similarity algorithms include the CO similarity matrix proposed by Fred and Jain [17] and the three link-based similarity matrices WCT, WTQ, and CSM proposed by Iam-On *et al.*[30], and the two fuzzy similarity matrices FLink and FCTS proposed by Su *et al.* [13]. The average-link (AL) algorithm is used to derive the final solution.
2) Graph-based algorithms include the CSPA, HGPA, and MCLA proposed by Strehl and Ghosh [10].
3) Relabeling-based algorithms include the selectively voting (SV) algorithm and the selectively weighted voting (SWV) algorithm proposed by Zhou and Tang [23], the cumulative agreement-based fuzzy k-means (CAFCM) algorithm proposed by Rathore *et al.* [15].
4) Feature-based algorithms include the expectation maximization (EM) algorithm for cluster ensemble proposed by Topchy *et al.* [40] and the iterative voting consensus (IVC) algorithm proposed by Nguyen *et al.*[43].

Besides, we compare FKMCE with three complex clustering algorithms including the NSC algorithm [6], the density-based spatial clustering of applications with noise (DBSCAN) [7] and the clustering by fast search and find of density peaks (CFSFDP) [8]. The aim of the comparison is to show the simulation of FKMCE for complex clustering.

D. Experimental Settings

To compare these different algorithms, we need to introduce the settings of their related parameters, which are listed as follows:

1) For the compared cluster ensemble algorithms, we run fuzzy k-means T times, each with a random and different initialization of cluster centers, to produce base clusterings on a dataset. The number of clusters k_h in each base clustering is equal to the true number of classes on each of the given datasets. We set the number of base clusterings $T = 40$ and the fuzzy index $m = 2$. For other parameters of these algorithms, we set them according to the suggestions of the authors.
2) The DBSCAN, CFSFDP, and FKMCE algorithms are required to input the parameter ϵ. We estimate the ϵ value by using $\bar{d} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{||\mathbf{x}_i - \bar{\mathbf{x}}||^2}$ where $\bar{\mathbf{x}} = \sum_{j=1}^{n}\frac{\mathbf{x}_j}{n}$. However, each of these algorithms may need different ϵ values on a dataset. Thus, we select the parameter in the interval $[\bar{d}/10, \bar{d}]$ with the step size as $\bar{d}/10$. We test each of these algorithms with the ten different values and select the highest ARI and NMI values on each dataset for comparison. However, different from DBSCAN and CFSFDP, the FKMCE algorithm has a certain randomness. Therefore, we need to run the FKMCE algorithm 50 times on each dataset and compute the average ARI and NMI values for comparison. For the parameters k_h, T, and m of FKMCE, we set the same values as those of other cluster ensemble algorithms.
3) For the NSC algorithm, we use Gaussian kernel to obtain a pairwise-objects similarity matrix and set the kernel parameter δ^2 in the interval [0.1,2] with the step size as 0.1. We select the highest ARI and NMI values for comparison.

E. Experimental Results

We first test these algorithms on the given datasets to compare their clustering accuracies. Tables III and IV show the ARI and

TABLE IV
NMI MEASURES OF DIFFERENT METHODS

Methods	Synthetic data sets								Real data sets				
	Ring	Jain	Flame	Agg	T4.8k	T7.1k	Chain	Atom	Iris	Wine	Breast	Digits	Statog
CO-AL	0.2112	0.5533	0.4420	0.7522	0.6601	0.6343	0.0686	0.2631	0.7582	0.8347	0.6231	0.7307	0.6322
WCT-AL	0.2162	0.5533	0.4420	0.8291	0.6546	0.6302	0.0686	0.2631	0.7582	0.8347	0.6231	0.7305	0.6321
WTQ-AL	0.2174	0.5533	0.4420	0.8003	0.5027	0.6370	0.0686	0.2631	0.7582	0.8347	0.6231	0.7306	0.6321
CSM-AL	0.2211	0.5533	0.4420	0.7993	0.6563	0.6630	0.0686	0.2631	0.7582	0.8347	0.6231	0.7309	0.6321
FLink-AL	0.0020	0.5533	0.4420	0.8874	0.6301	0.6099	0.0669	0.3133	0.7304	0.8336	0.6152	0.4307	0.5840
FCTS-AL	0.0032	0.5533	0.4420	0.8033	0.6234	0.5900	0.0669	0.1646	0.7304	0.8336	0.6152	0.4031	0.5175
CSPA	0.3785	0.3631	0.4049	0.7200	0.6233	0.6071	0.0686	0.0024	0.6803	0.7771	0.2981	0.7857	0.5425
HGPA	0.0008	0.0000	0.0000	0.4088	0.5170	0.3656	0.0000	0.0000	0.1609	0.1705	0.0007	0.4932	0.326
MCLA	0.0013	0.5533	0.4420	0.7515	0.6418	0.6334	0.0686	0.2713	0.7582	0.8347	0.6231	0.7627	0.5903
SV	0.1758	0.5533	0.4343	0.3690	0.3672	0.2049	0.0743	0.2863	0.0183	0.8529	0.6231	0.3782	0.4481
SWV	0.2487	0.5533	0.4343	0.6481	0.3971	0.4339	0.0743	0.2863	0.0110	0.8529	0.6231	0.6085	0.5248
CAFCM	0.1548	0.5533	0.4420	0.8163	0.5786	0.6368	0.0678	0.2243	0.7304	0.8336	0.6152	0.3676	0.5248
EM	0.1495	0.4869	0.3780	0.7295	0.6197	0.5730	0.0663	0.3404	0.6727	0.7980	0.5400	0.7271	0.5837
IVC	0.3813	0.1217	0.3360	0.7303	0.6342	0.6467	0.0686	0.1942	0.6801	0.7281	0.0415	0.7208	0.5256
NSC	**1.0000**	**1.0000**	0.7770	0.9271	0.9538	0.9853	**1.0000**	**1.0000**	0.7980	**0.9016**	0.6328	0.8119	0.6243
DBSCAN	**1.0000**	0.2561	0.2070	0.6835	0.7926	0.8719	0.4828	0.2773	0.5904	0.4451	0.0303	0.7163	0.5021
CFSFDP	0.3792	0.5960	0.8883	0.9851	0.7131	0.8451	0.6544	0.4592	0.7277	0.7528	0.6152	0.8645	0.5644
FMKCE	**1.0000**	**1.0000**	**0.9028**	**0.9869**	**0.9840**	**0.9946**	**1.0000**	**1.0000**	**0.8381**	0.8667	**0.6667**	**0.8919**	**0.6774**

TABLE V
STANDARD DEVIATION OF THE FKMCE ALGORITHM FOR THE ARI AND NMI MEASURES

Indices	Synthetic data sets								Real data sets				
	Ring	Jain	Flame	Agg	T4.8k	T7.1k	Chain	Atom	Iris	Wine	Breast	Digits	Statog
ARI(std)	0.0000	0.0000	0.0486	0.0032	0.0614	0.0121	0.0000	0.0000	0.0906	0.0431	0.0650	0.0517	0.0293
NMI(std)	0.0000	0.0000	0.0542	0.0038	0.0367	0.0127	0.0000	0.0000	0.0508	0.0384	0.0631	0.0249	0.0182

NMI values of existing cluster ensemble algorithms on synthetic and real datasets, respectively. The bold values in these tables represent the highest value of the compared algorithms on each data set. According to these tables, we see that the clustering accuracies of the FKMCE algorithm are obviously superior to other cluster ensemble algorithms on these tested datasets. The main reason is that the base clusterings produced by fuzzy k-means include lots of incredible memberships, while we are clustering these datasets with different-shaped clusters. The existing ensemble algorithms cannot integrate them to recognize these clusters, due to the lack of evaluation about the membership credibilities. But our proposed algorithm can recognize the credible memberships to effectively discover different-shaped clusters and improve the performance of the fuzzy k-means algorithm. Besides, Tables III and IV also show the comparison results of the FKMCE algorithm with the NSC, DBSCAN, and CFSFDP clustering algorithms on the given datasets. We can see that the clustering validity of the FKMCE algorithm is superior or close to the best results of these algorithms. The experiments tell us that the proposed algorithm can well simulate complex clustering results. Furthermore, we compare the efficiency of the FKMCE algorithm with the NSC, DBSCAN, and CFSFDP algorithms on the KDD-CUP'99 dataset. In the experiment, we fix $k = 2$ and $\epsilon = 0.14$. Fig. 5 shows the running time of these algorithms with different numbers of objects. We can see that the proposed algorithm is very efficient, compared to other algorithms. This indicates that the FKMCE algorithm is a good choice for clustering large-scale datasets.

Due to the fact that the FKMCE algorithm has a certain randomness, we test it 50 times on each datasets. Table V shows the standard deviation (std) of the ARI and NMI values for its

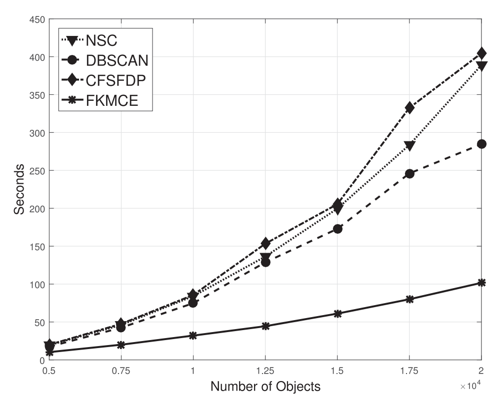
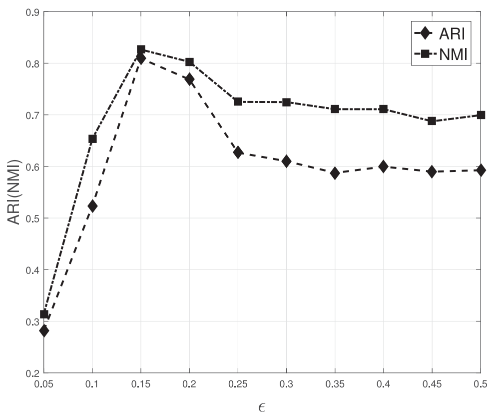


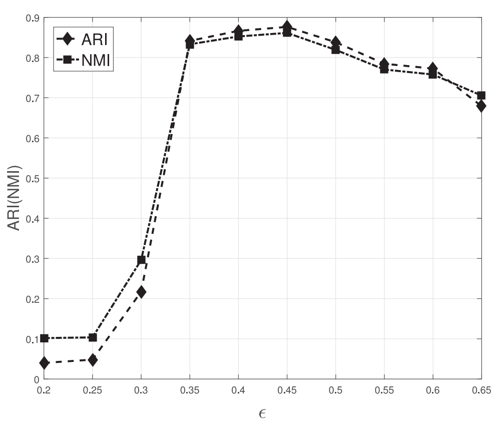
Fig. 5. Time comparison of different algorithms.

50 clustering results. We can see that the std value is less than 0.1 on each dataset. This indicates that the randomness has a limited impact on the performance of the FKMCE algorithm.

Furthermore, we test the effect of the parameter ϵ on the performance of the FKMCE algorithm by the experiments. We take the iris and wine data for example. According to Fig. 6, we see that the clustering accuracy of the FKMCE algorithm is very poor while the ϵ value is very low. As the ϵ value is further growing, the performance of the algorithm is becoming better. However, while the ϵ value is increasing to a certain extent, the clustering accuracy is decreasing. This experimental result tells us that the ϵ value is too large or small to obtain a good ensemble result. Since the performance of the FKMCE algorithm depends on the parameter setting, we should select a suitable value of ϵ on each dataset. However, there are few theoretical guidelines for setting the parameter. In this paper, we



(a)



(b)

Fig. 6. (a) Effect of the parameter ϵ on the iris data. (b) Effect of the parameter ϵ on the wine data.

provide a rule of thumb that the parameter value is selected from the interval $[\bar{d}/10, \bar{d}]$, where $\bar{d}$ is the average distance between each object and the center of a dataset. We tested the DBSCAN, CFSFDP, and FKMCE with different parameter values on the given datasets. We found that these algorithms can obtain better clustering results if the parameter is selected from the interval.

V. Conclusion

Fuzzy k-means is a widely used clustering algorithm for its low computational cost. However, it is a linear clusterer and its performance tends to be affected by data distributions. In this paper, we have proposed a new cluster ensemble algorithm for integrating MFKMs, which is called FKMCE. The new algorithm includes three main steps: producing MFKMs, evaluating the local credibility of memberships, and building the relation between clusters. It improves the robustness and quality of fuzzy k-means and can rapidly recognize different-shaped clusters. In the experimental analysis, we have compared the FKMCE algorithm with existing cluster ensemble algorithms and three complex clustering algorithms on synthetic and real datasets. The comparison results have illustrated that the performance of the proposed algorithm is very effective. Furthermore, we have analyzed the efficiency of the FKMCE algorithm, which is suitable to deal with large-scale datasets.

Acknowledgment

The authors are very grateful to the editors and reviewers for their valuable comments and suggestions.

References

[1] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[2] E. R. Ruspini, "A new approach to clustering," *Inf. Control*, vol. 15, no. 1, pp. 22–32, 1969.

[3] J. C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.

[4] J. Liang, L. Bai, C. Dang, and F. Cao, "The k-means type algorithms versus imbalanced data distributions," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 728–745, Aug. 2012.

[5] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[6] A. Y. Ng, M. I. Jordan, and Y. Weiss ,"On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*, vol. 14, T. G. Dietterich, S. Becker, Z. Ghahramani, Eds. Cambridge, MA, USA: MIT Press, 2002.

[7] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, 1996, pp. 226–231.

[8] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[9] Y. Wang, X. Liu, Y. Dou, and R. Li, "Multiple kernel clustering framework with improved kernels," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2999–3005.

[10] A. Strehl and J. Ghosh, "Cluster ensembles: A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2002.

[11] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, pp. 1–30, 2007.

[12] Z. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[13] P. Su, C. Shang, and Q. Shen, "Link-based pairwise similarity matrix approach for fuzzy c-means clustering ensemble," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2014, pp. 1538–1544.

[14] Z. Yu *et al.*,"Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 12, no. 4, pp. 887–901, Jul./Aug. 2015.

[15] P. Rathore, J. C. Bezdek, S. M. Erfani, S. Rajasegarar and M. Palaniswami, "Ensemble fuzzy clustering using cumulative aggregation on random projections," *IEEE Trans. Fuzzy Syst.*, to be published.

[16] E. Gonzlez and J. Turmo, "Unsupervised ensemble minority clustering," *Mach. Learn.*, vol. 98, pp. 217–268, 2015.

[17] A. Fred and A. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, Jun. 2005.

[18] L. Kuncheva and D. Vetrov, "Evaluation of stability of k-means cluster ensembles with respect to random initialization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1798–1808, Nov. 2006.

[19] H. Liu, J. Wu, T. Liu, D. Tao, and Y. Fu, "Spectral ensemble clustering via weighted k-means: Theoretical and practical evidence," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 5, pp. 1129–1143, May 2017.

[20] M. Law, A. Topchy, and A. Jain, "Multiobjective data clustering," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. II-424–II-430.

[21] B. Fischer and J. Buhmann, "Bagging for path-based clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1411–1415, Nov. 2003.

[22] A. Topchy, B. Minaei-Bidgoli , and A. Jain, "Adaptive clustering ensembles," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, pp. 272–275.

[23] Z. Zhou and W. Tang, "Clusterer ensemble," *Knowl.-Based Syst.*, vol. 19, no. 1, pp. 77–83, 2006.

[24] Y. Hong, S. Kwong, H. Wang, and Q. Ren, "Resampling-based selective clustering ensembles," *Pattern Recognit. Lett.*, vol. 41, no. 9, pp. 2742–2756, 2009.

[25] X. Fern and C. Brodley, "Random projection for high dimensional data clustering: A cluster ensemble approach," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 186–193.

[26] F. Gullo and C. Domeniconi, "Metacluster-based projective clustering ensembles," *Mach. Learn.*, vol. 98, no. 1/2, pp. 1–36, 2013.
[27] Y. Yang and J. Jiang, "Hybrid sampling-based clustering ensemble with global and local constitutions," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 5, pp. 952–965, May 2016.
[28] A. Fred and A. K. Jain, "Data clustering using evidence accumulation," in *Proc. 16th Int. Conf. Pattern Recognit.*, 2002, pp. 276–280.
[29] Y. Yang and K. Chen, "Temporal data clustering via weighted clustering ensemble with different representations," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 2, pp. 307–320, Feb. 2011.
[30] N. Iam-On , T. Boongoen, S. Garrett, and C. Price, "A link-based approach to the cluster ensemble problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2396–2409, Dec. 2011.
[31] N. Iam-On , T. Boongoen, S. Garrett, and C. Price, "A link-based cluster ensemble approach for categorical data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 413–425, Mar. 2012.
[32] X. Fern and C. Brodley, "Solving cluster ensemble problems by bipartite graph partitioning," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, pp. 36–43.
[33] D. Huang, J. Lai, and C. D. Wang, "Robust ensemble clustering using probability trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 5, pp. 1312–1326, May 2016.
[34] M. Selim and E. Ertunc, "Combining multiple clusterings using similarity graph," *Pattern Recognit.*, vol. 44, no. 3, pp. 694–703, 2011.
[35] C. Boulis and M. Ostendorf, "Combining multiple clustering systems," in *Proc. Eur. Conf. Principles Pract. Knowl. Discovery Databases*, 2004, pp. 63–74.
[36] A. Topchy, B. Minaei-Bidgoli , and A. Jain, "Adaptive clustering ensembles," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, pp. 272–275.
[37] P. Hore, L. O. Hall, and B. Goldgo, "A scalable framework for cluster ensembles," *Pattern Recognit.*, vol. 42, no. 5, pp. 676–688, 2009.
[38] B. Long, Z. Zhang, and P. S. Yu, "Combining multiple clusterings by soft correspondence," in *Proc. 4th IEEE Int. Conf. Data Mining*, 2005, pp. 282–289.
[39] D. Cristofor and D. Simovici, "Finding median partitions using information theoretical based genetic algorithms," *J. Universal Comput. Sci.*, vol. 8, no. 2, pp. 153–172, 2002.
[40] A. Topchy, A. Jain, and W. Punch, "Clustering ensembles: Models of consensus and weak partitions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1866–1881, Dec. 2005.
[41] H. Wang, H. Shan, and A. Banerjee, "Bayesian cluster ensembles," *Statist. Anal. Data Mining*, vol. 4, no. 1, pp. 54–70, 2011.
[42] Z. He, X. Xu, and S. Deng, "A cluster ensemble method for clustering categorical data," *Inf. Fusion*, vol. 6, no. 2, pp. 143–151, 2005.
[43] N. Nguyen and R. Caruana, "Consensus clusterings," in *Proc. 7th IEEE Int. Conf. Data Mining*, 2007, pp. 607–612.
[44] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining Knowl. Discovery*, vol. 2, no. 3, pp. 283–304, 1998.
[45] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, no. 3, pp. 301–315, Jun. 1998.
[46] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, Aug. 1995.
[47] P. Fränti *et al.*, "Clustering datasets," 2015. [Online]. Available: http://cs.uef.fi/sipu/datasets/
[48] "Clustering benchmarks," 2017. [Online]. Available: https://github.com/deric/clustering-benchmark
[49] UCI Machine Learning Repository, 2017. [Online]. Available: http://www.ics.uci.edu/mlearn/MLRepository.html
[50] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Statist. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
[51] S. A. Teukolsky, W. T. Vetterling, W. H. Press, and B. P. Flannery, "Conditional entropy and mutual information," in *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. New York, NY, USA: Cambridge Univ. Press, 2007.

Liang Bai received the Ph.D. degree in computer science from Shanxi University, Taiyuan, China, in 2012.

He is currently an Associate Professor with the School of Computer and Information Technology, Shanxi University and a Postdoctoral Worker with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He has authored or coauthored several journal papers in his research fields, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON FUZZY SYSTEMS, and *Data Mining and Knowledge Discovery*. His research focuses on cluster analysis.

Jiye Liang received the Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2001.

He is currently a Professor with the School of Computer and Information Technology, Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China. He has authored or coauthored more than 80 articles in international journals. His research interests include artificial intelligence, granular computing, data mining, and machine learning.

Yike Guo received the Ph.D. degree in logic and declarative programming from Imperial College, University of London, London, U.K., in 1993.

He is currently a Professor in computing science with the Department of Computing, Imperial College, University of London. His research interests include large-scale scientific data analysis, data mining algorithms and applications, parallel algorithms, and cloud computing.