ELSEVIER

Fuzzy Sets and Systems 128 (2002) 305–322

FUZZY sets and systems

www.elsevier.com/locate/fss

# Unsupervised fuzzy clustering with multi-center clusters

C.W. Tao

*Department of Electrical Engineering, National I-Lan Institute of Technology, I-Lan 260, Taipei, Taiwan*

Received 12 January 2000; received in revised form 3 May 2001; accepted 31 August 2001

**Abstract**

A new unsupervised fuzzy clustering algorithm is provided in this paper to cluster the data patterns without a priori information about the number of clusters. The initial guesses of the locations of the cluster centers or the initial guesses of the membership values are not necessary. With the minimization of a new objective function, cluster centers are generated one by one. Related centers are defined to belong to the same cluster. Multi-centers are adopted to represent the non-spherical shape of clusters. Thus, the clustering algorithm with multi-center clusters can handle non-traditional curved clusters. The proposed algorithm is tested on different data sets with a variety of cluster shapes, cluster densities, and number of points in each cluster. Also, the results are compared with some other clustering algorithms to show the effectiveness of the algorithm. Moreover, the designed unsupervised fuzzy clustering algorithm is applied to cluster the pixels in a color image to show the efficiency of the algorithm. 

## 1. Introduction

Clustering [7,21,36] plays an important role in pattern recognition [36], image processing, and computer vision [5,40]. With a clustering technique, a collection of objects or feature vectors is partitioned into clusters. The objects in the same cluster have similar characteristics, while a characteristic distinction exists between different clusters. Unsupervised clustering consists of partitioning the unlabeled objects into an unknown number of clusters. To cluster the unlabeled objects based on a model of mixture normal density [6,28,34,39], the parameter vectors of the multivariate normal densities could be either estimated by maximum likelihood approach or the unsupervised Bayesian learning approach [3,34]. Then the objects can be allocated to their correct population according to these normal densities. Unsupervised clustering can also be achieved by the mode identifying procedures [15,25,37]. A hierarchical clustering [13,33] sequentially merges (divides) clusters and records the merging (divisive) sequence. Two typical hierarchical clustering methods are the single-link [23] and complete-link [19] methods. Gordon [16] clusters the objects by minimizing the within-cluster sum of square distances. The $K$-means algorithm [11,29] assigns the objects to the closest cluster center. Friedman and Rubin [12] define a hill-climbing method to change the cluster label when the objective function is improved.

*E-mail address:* cwtao@mail.ilantech.edu.tw (C.W. Tao).

0165-0114/02/$ - see front matter 
PII: S0165-0114(01)00191-9

Unlike the hard clustering techniques above (each object is assigned to one and only one cluster), fuzzy clustering [1,8,30] allows an object to belong to a cluster with a grade of membership. Moreover, when there is not enough information about the structure of the data, fuzzy set theory [43] can handle this uncertainty better, and has been widely applied to the data clustering area [2,9,18]. One of the most popular fuzzy clustering algorithms is the fuzzy $C$-means algorithm [2]. Like many hard clustering techniques (e.g. $K$-means algorithm [35]), the number of clusters is known in advance for the fuzzy $C$-means algorithm [2]. However, we have to cluster objects without knowing the number of clusters in a lot of cases. To solve this difficulty, the validity of the resulting clustering is evaluated for a range of possible numbers of clusters. And the final number of clusters is selected to be the number with which the clustering process has the optimal validity [2,14,19,32]. The progressive approach to obtain the number of clusters consists of finding one good cluster at a time, and the clustering is completed when no more good clusters can be found [24,44]. Unlike these approaches to selecting the number of clusters, in this paper the number of clusters is determined by minimizing an objective function. There are many clustering criteria that can be used to derive the objective function [2]. The distance criterion is the most commomly used criterion; and is also adopted in our design.

It is known that the improper initial guesses of the cluster centers may degrade the performance of $K$-means (fuzzy $C$-means) algorithm. To eliminate the difficulty of initial guesses of the cluster centers, the mountain clustering method [42] and the subtractive clustering approach [4] can be used to find the initial cluster centers when the number of clusters is known. As mentioned in the book of Jang et al. [22], the computation load will be heavy when a large number of data points are clustered with the subtractive clustering approach. And the same difficulty (heavy computation load) occurs when high-dimensional data points are clustered with the mountain clustering method. In our study, the hierarchical subtractive clustering approach (HSC) is proposed to reduce the computational load.

Since, the data clustering will encounter the difficulty of the occurrence of variations in cluster shapes, researchers have devoted substantial efforts to cluster hyperellipsoidal data [14], multimodal data [15], and shell-type data [26]. Without derivation of the "exponential" distance measure [14] and the characteristic function of multimodal data set [15], the nonspherical cluster is represented with multi-centers in this paper. Lately, a fuzzy min–max clustering neural network was constructed in Simpson's work [31] to create and refine the clusters, and one cluster (called class in [31]) can consist of more than one hyperbox. However, the approach which is used to assign hyperboxes into one cluster is not detailed in [31]. Also, the clustering algorithm *DBSCAN* [10,27] considers that a cluster contains the points which has more than a specified threshold number of points in a specified neighborhood. The cluster is constructed by finding a maximum set of the connected points which satisfy the density (points in the neighborhood) condition. That is, *DBSCAN* expands the cluster in a pointwise manner. Although *DBSCAN* can find the clusters with arbitrary shape, a priori information (specified neighborhood and density threshold) is necessary. A similar algorithm (which creates the cluster pointwise) *DBCLASD* [41] is designed to cluster data points into arbitrarily shaped clusters without any prior information. *DBCLASD* is efficient for clustering large databases. However, the densities in the clusters are assumed to be uniform and a minimum number (30) of data point is required to apply the $\chi^2$-test used in the algorithm. Thus, *DBCLASD* might not be appropriate to find clusters with a small cardinality. To achieve a successful clustering without these difficulties, a new fuzzy clustering algorithm is introduced in this paper. Instead of generating clusters in a pointwise manner, candidate cluster centers are found to represent clusters. In order to partition data points into clusters with arbitrary shape which is non-traditional curved, it is reasonable to represent clusters using multi-centers. The connection test between centers is defined to associate the centers within one cluster. If the centers are connected, then they belong to the same cluster. With the benefits of the fuzzy technique, the proposed fuzzy clustering with multi-center clusters would be expected to handle non-traditional curved clusters well. The simulation results included empirically justify the effectiveness of the proposed unsupervised fuzzy clustering with multi-center clusters.

The remainder of this paper is organized as follows. The subtractive clustering algorithm is reviewed in Section 2. Section 3 describes the hierarchical subtractive clustering approach (HSC) in detail. The objective

function is designed and the cluster centers are selected to minimize the objective functon in Section 4. Section 5 presents the algorithms to assign multi-centers to one cluster to represent the nonspherical clusters. A complete algorithm for the proposed fuzzy clustering technique is presented in Section 6. Data with unequal population clusters, two-elliptic clusters, and linearly unseparable clusters are clustered in Section 7. The clustering for the iris data and a color image is also included. Moreover, the proposed clustering algorithm is compared with some other clustering algorithms. Finally, Section 8 provides the conclusion.

## 2. Subtractive clustering algorithm

In this section, the subtractive clustering algorithm [4] is reviewed for the comparison with the proposed hierarchical subtractive clustering algorithm described in Section 3. Let the collection of data points be $C_x = \{x_1, x_2, \ldots, x_n\}$, and each data point $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ be a $p$-dimensional vector. The subtractive clustering algorithm is presented in Algorithm 1:

**Algorithm 1** (Subtractive clustering)

1. Let $m = 0$, $x_{c0} = 0$, and $D_{c0} = 0$. ($m$ is the number of iterations, $x_{cm}$ is the cluster centers and $D_{cm}$ is the density of $x_{cm}$.)
2. Calculate the density function for each data point $x_i$,

$$D_i = \sum_{j=1}^{n} \exp\left(\frac{-\|x_i - x_j\|^2}{(r_a/2)^2}\right) - \sum_{k=0}^{m} D_{ck} \exp\left(\frac{-\|x_i - x_{ck}\|^2}{(r_b/2)^2}\right), r_a, r_b: \text{ constants} \tag{1}$$

3. Let $m = m + 1$.
4. Select the cluster center $x_{cm}$ to be the point with largest density value

$$D_{cm} = \max_i D_i.$$

5. If a sufficient number of cluster centers have been generated, then stop.
   else go back to step 2. □

From the density function in Eq. (1), it is easy to see that the computation load will be heavy when the number of data points $n$ is large. To reduce the computation load, a two-level hierarchical subtractive clustering algorithm is proposed.

## 3. Hierarchical subtractive clustering algorithm

The hierarchical subtractive clustering approach partitions the collection of data into several subcollections, and calculates the density functions for the data points in each subcollection. The density function is defined to be the number of points inside the neighborhood of each data point. For each subcollection, instead of taking the data point with highest value of density function as the cluster center, the centroid of the data points in the neighborhood of the point with highest value of density function is calculated as the cluster center. To find the second cluster center, the data points used for derivation of the first cluster center has to be taken off from the collection of data. With the same approach, the cluster centers for each subcollection can be found. Then, all the cluster centers from all the subcollections are considered as a new collection of data to be further clustered. Thus, the calculation load can be reduced with this hierarchical subtractive clustering approach.

Let the collection of data points be $C_x = \{x_1, x_2, \ldots, x_n\}$ with each data point $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$. A two-level hierarchical subtractive clustering algorithm is proposed to reduce the computation load. In the first level of HSC, the collection of data is divided into several subcollections with an approximately equal number of data points. Each subcollection has to include a sufficient number of points. That is, if the number of data point $n$ is small, the collection division is not necessary and the first level of HSC can be omitted. Assume that $n = mb + d$ ($b, d, m$ are integers and $m > 1, d < m$), $C_x$ can be divided into $m$ subcollections $C_x^\zeta$; $\zeta = 1, 2, \ldots, m$ with

$$C_x^\zeta = \{x_1^\zeta, x_2^\zeta, \ldots, x_b^\zeta\}, \quad \zeta = 1, 2, \ldots, m-1$$

and

$$C_x^m = \{x_1^m, x_2^m, \ldots, x_b^m, x_{b+1}^m, \ldots, x_{b+d}^m\}.$$

To define a neighborhood of a data point, the scatter of data points on each dimension is considered to be an important factor. For example, if the scatter of data points is large and the neighborhood of points is small, then there might be no points included in the neighborhood of any one of the points. On the other hand, if the scatter is small and the neighborhood is large, then the entire data set might be in the neighborhoods of all the data points. The problem can be reasonably solved by selecting different neighborhoods for different collections of data points with different scatter. It is known that a standard deviation can be used to approximately represent the scatter of data points. Thus, the neighborhood of a multidimensional point $x_i$ can be considered as a function of standard deviations on each dimension. In our work, the neighborhood of a data point $x_i$ is defined to be a hypersphere (centered at $x_i$) with the radius selected to be the minimum standard deviation among all the dimensions. It is assumed that the data points have been properly normalized. That is, the point $x_j$ is said to be in the neighborhood of data point $x_i$, if the distance $\|x_j - x_i\|$ is less than the value $r_{\min}$ which is defined to be

$$r_{\min} = \min_p(\text{standard deviation of } \{x_{1p}, x_{2p}, \ldots, x_{np}\}).$$

For each point $x_i^\zeta$ in the subcollection $C_x^\zeta$, the density function $D_i^\zeta$ of $x_i^\zeta$ is defined as

$$D_i^\zeta(x_i^\zeta) = \text{the number of points in the neighborhood of } x_i^\zeta,$$
$$= \sum_{j=1}^{b} u(r_{\min} - \|x_j^\zeta - x_i^\zeta\|) \tag{2}$$

is calculated. Note that the unit step function $u$ in Eq. (2) is defined as

$$u(\alpha) = \begin{cases} 1 & \text{if } \alpha \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

If the data point $x_{c1}^\zeta \in C_x^\zeta$ has the largest density value, i.e.,

$$D_{c1}^\zeta = \max_i D_i^\zeta,$$

the first center $cc_1^\zeta$ of the subcollection $C_x^\zeta$ is derived as

$$cc_1^\zeta = \text{mean}(\text{data points in the neighborhood of } x_{c1}^\zeta). \tag{3}$$

Instead of calculating Eq. (1) every time to find the cluster center in the subtractive clustering algorithm, the density function is only evaluated once in the first level of the hierarchical subtractive clustering algorithm.

Let the collection $C_{c1}^{\zeta}$ be the data points which are in the neighborhood of $x_{c1}^{\zeta}$. The centroid of the data points in the neighborhood of a data point $x_{c2}^{\zeta}$ ($\in C_x^{\zeta} - C_{c1}^{\zeta}$) is considered as the second cluster center if

$$D_{c2}^{\zeta} = \max_i D_i^{\zeta} \quad \text{for } x_i \in C_x^{\zeta} - C_{c1}^{\zeta},$$

where $D_{c2}^{\zeta}$ is the density of $x_{c2}^{\zeta}$. Likewise, we can continue to find the cluster centers of $C_x^{\zeta}$ until

$$C_x^{\zeta} - \sum_{k=1} C_{ck}^{\zeta} = \emptyset \text{ (empty set)}.$$

Thus, the first level of HSC can be summarized in the following algorithm:

**Algorithm 2** (First level of HSC)

1. Divide the collection $C_x$ into $m$ subcollections $C_x^{\zeta}$, $\zeta = 1, 2, \ldots, m$.
2. Let $\zeta = 1$.
3. Find the centers of subcollection $C_x^{\zeta}$ as follows,
   (a) Calculate the density function for each data point $x_i^{\zeta}$,

$$D_i^{\zeta} = \sum_{j=1}^{b} u(r_{\min} - \|x_j^{\zeta} - x_i^{\zeta}\|). \tag{4}$$

   (b) Let $k = 1$
   (c) Find

$$x_{ck}^{\zeta} \in C_x^{\zeta} - \sum_{l=1}^{k} C_{c(l-1)}^{\zeta};$$

   such that the density of $x_{ck}^{\zeta}$, $D_{ck}^{\zeta}$, satisfies

$$D_{ck}^{\zeta} = \max_i D_i^{\zeta} \quad \text{for } x_i^{\zeta} \in C_x^{\zeta} - \sum_{l=1}^{k} C_{c(l-1)}^{\zeta},$$

   where $C_{ck}^{\zeta}$ is the collection of data points in the neighborhood of $x_{ck}^{\zeta}$, and ($C_{c0}^{\zeta} = \emptyset$).
   (d) Take the centroid of the data points in the neighborhood of $x_{ck}^{\zeta}$ to be the center $cc_k^{\zeta}$.
   (e) If $C_x^{\zeta} - \sum_{l=1}^{k} C_{c(l-1)}^{\zeta} \neq \emptyset$ go back to step 3($c$) with $k = k + 1$. else go to step 4.
4. If $\zeta = m$, then stop.
   else go back to step 3 with $\zeta = \zeta + 1$. □

With the collection of cluster centers obtained from the first level of HSC as a new collection of data points, the second level of HSC mainly repeats the key steps in Algorithm 2 (the first level of HSC). And the algorithm of the second level of HSC is presented as:

**Algorithm 3** (Second level of HSC)

1. Reset $C_x$ to be the collection of all the centers obtained from Algorithm 2 with the elements $x_i$; $i = 1, 2, \ldots, n_c$, where $n_c$ is the number of centers.
2. Calculate the density function for each data point $x_i$, $i = 1, 2, \ldots, n_c$

$$D_i = \sum_{j=1}^{n_c} u(r_{\min} - \|x_j - x_i\|). \tag{5}$$

3. Let $k=1$.
4. Take the centroid of the data points in the neighborhood of $x_{ck}$ to be the cluster center $cc_k$ if

$$x_{ck} \in C_x - \sum_{l=1}^{k} C_{c(l-1)} \quad (C_{c0} = \emptyset)$$

and

$$D_{ck} = \max_i D_i \quad \text{for } x_i \in C_x - \sum_{l=1}^{k} C_{c(l-1)}.$$

5. If $C_x - \sum_{l=1}^{k} C_{c(l-1)} \neq \emptyset$ go back to step 4 with $k = k+1$.
   else stop with cluster centers, $cc_j$, $j=1,2,\ldots,n_p$, and $n_p$ being the number of cluster centers. □

From Eqs. (1) and (4), it is easy to see that for each data point $x_i$, the distance function $\|x_j - x_i\|$ is calculated $n$ times in Eq. (1) of the subtractive clustering algorithm, while it is calculated only $n/m$ (Eq. (4)) and $n_c$ (Eq. (5)) times for each data point in the first and second levels of HSC, respectively. Therefore, when $n$ is large, the computation load is significantly reduced with the hierarchical subtractive clustering algorithm. Note that $n_c$ is usually much smaller than $n$. To automatically determine the number of clusters, the centers obtained with the HSC algorithm (Algorithms 2 and 3) are used as the primary cluster centers for the further processing which is presented in Section 4.

## 4. Selection of candidate cluster centers

The candidate cluster centers are selected from the set of primary cluster centers,

$$C_c^{\mathrm{p}} = \{cc_1, cc_2, \ldots, cc_{n_p}\},$$

by minimizing the objcetive function defined in Eq. (6). There are many clustering criteria that can be used to derive the objective functions [2]. The distance criterion is the most commonly used criterion. Therefore, our objective function is designed based on the distance criterion. Consider the objective function of the fuzzy-$C$ means algorithm, i.e.,

$$J_{\mathrm{f\,cm}} = \sum_{i=1}^{n} \sum_{j=1}^{n_p} \mu_{ji}^{m} \|x_i - cc_j\|^2, \tag{6}$$

where $\mu_{ji}$ is the membership function and $m>1$ is the fuzzifier. It is shown in [2] that the optimal memberships $\mu_{ji}$ are given by

$$\mu_{ji} = \frac{1}{\sum_{k=1}^{n_p} (\|(x_i - cc_j)\|^2 / \|(x_i - cc_k)\|^2)^{2/(m-1)}}.$$

It can be seen that the objective function $J_{\mathrm{f\,cm}}$ in Eq. (6) is a measure of the differnces between the data points and the cluster centers. When $J_{\mathrm{f\,cm}}$ is minimized, the sum of the differences between the data points and the cluster centers are minimized. Thus, $J_{\mathrm{f\,cm}}$ can be minimized to obtain good clustering results when the number of clusters is fixed. However, it is known that if the number of cluster is not known, to minimize $J_{\mathrm{f\,cm}}$ would results $n$ clusters, where $n$ is the number of data poins. That is, the effective number of clusters cannot

be determined simply by minimizing the objective function $J_{\mathrm{fcm}}$ [19]. Instead of using the validity measures [2,38,40], the effective number of cluster is determined by minimizing the modified objective function $J^*_{\mathrm{fcm}}$,

$$J^*_{\mathrm{fcm}} = \sum_{i=1}^{n}\sum_{j=1}^{\beta} \mu_{ji}^m \|x_i - cc_j^{\mathrm{c}}\|^2 + \omega J_{cc}$$

and

$$J_{cc} = \sum_{i=1}^{\beta}\sum_{j=i+1}^{\beta} \|cc_i^{\mathrm{c}} - cc_j^{\mathrm{c}}\|^2, \tag{7}$$

where $\beta$ is the number of candidate cluster centers picked from $C_c^{\mathrm{p}}$; and $cc_i^{\mathrm{c}}$, $i = 1,2,\ldots,\beta$, are candidate cluster centers. Also, $J_{cc} = 0$ when $\beta = 1$. With the objective function $J^*_{\mathrm{fcm}}$, the entire energy is defined to not only consider the energy in each clusters, but also consider the energy between clusters. In this case, with one more cluster center selected, it decreases the energy in clusters ($J_{\mathrm{fcm}}$ is decreased) but it increases the energy between clusters ($J_{cc}$ is increased). Therefore, the system will only benefit from selecting the centers which will reduce more energy in clusters than the energy increased between clusters (if they are selected). That is, if the value of the objective function $J^*_{\mathrm{fcm}}$ is decreased with an additional cluster center $cc_i$ selected from $C_c^{\mathrm{p}}$, then the value of $\beta$ is increased by one and $cc_i$ is decided to be one candidate cluster center, i.e., $cc_\beta^{\mathrm{c}} = cc_i$. Otherwise, $cc_i$ is decided not to be a candidate cluster center and we will try the next primary cluster center of $C_c^{\mathrm{p}}$. Note that if there is a priori information which indicates that the clusters are highly close to each other, we can weight $J_{cc}$ by a factor $\omega > 1$ to further emphasize the closeness of cluster centers. The selection of the candidate cluster centers is implemented with Algorithm 4.

**Algorithm 4** (Selection of candidate cluster centers)

1. Let $\beta = 1$, $J^{**}_{\mathrm{fcm}} = \infty$, and $\delta = 1$.
2. Set $cc_\beta^{\mathrm{c}} = cc_\delta$. ($cc_\delta \in C_c^{\mathrm{p}} = \{cc_1, cc_2, \ldots, cc_{n_p}\}$.
3. Calculate the objective function $J^*_{\mathrm{fcm}}$ in Eq. (7) with the membership function

$$\mu_{ji} = \frac{1}{\sum_{k=1}^{\beta} (\|(x_i - cc_j^{\mathrm{c}})\|^2 / \|(x_i - cc_k^{\mathrm{c}})\|^2)^{2/(m-1)}}.$$

4. If $J^*_{\mathrm{fcm}} < J^{**}_{\mathrm{fcm}}$, then
   (a) $J^{**}_{\mathrm{fcm}} = J^*_{\mathrm{fcm}}$.
   (b) $\beta = \beta + 1$.
5. If $\delta > n_p - 1$, stop with $\beta = \beta - 1$.
   else go back to step 2 with $\delta = \delta + 1$. □

With Algorithm 4, the number of candidate cluster centers can be automatically determined. Since, the non-spherical clusters are represented with multi-centers, the number of cluster centers is not equal to the number of clusters. In the next section, non-spherical clusters are detected and the number of clusters is decided.

## 5. Association of related centers to the same cluster

Multi-centers are assigned to one cluster to represent the non-spherical cluster in this section. For two centers $cc_i^{\mathrm{c}}$ and $cc_j^{\mathrm{c}}$, let $v_1, v_2, \ldots, v_h$ be the equal distance points on the line connecting $cc_i^{\mathrm{c}}$ and $cc_j^{\mathrm{c}}$ such that the points $v_1, v_2, \ldots, v_h$ lie between $cc_i^{\mathrm{c}}$ and $cc_j^{\mathrm{c}}$. That is,

$$\|v_1 - cc_i^{\mathrm{c}}\| = \|v_2 - v_1\| = \cdots = \|v_h - v_{h-1}\| = \|cc_j^{\mathrm{c}} - v_h\| = \frac{\|cc_j^{\mathrm{c}} - cc_i^{\mathrm{c}}\|}{h+1}.$$

As the density function in Eq. (2), the densities of $cc_i^c$ and $cc_j^c$ are

$$D_{cci} = \sum_{l=1}^{n} u(r_{\min} - \|x_l - cc_i^c\|),$$

$$D_{ccj} = \sum_{l=1}^{n} u(r_{\min} - \|x_l - cc_j^c\|). \tag{8}$$

Also, the density $D_{vk}$; $k = 1, 2, \ldots, h$ is calculated as

$$D_{vk} = \sum_{l=1}^{n} u(r_{\min} - \|x_l - v_k\|), \quad k = 1, 2, \ldots, h.$$

In order to find centers which are connected, clusters are first assumed to be well separated. Then the densities of the points $v_k$, $k = 1, 2, \ldots, h$, between two centers ($cc_i^c$ and $cc_j^c$) are examined, if the condition in Eq. (9) is satisfied for all $k$, $cc_i^c$ and $cc_j^c$ are defined to be connected. After the connection relationship between any two candidate cluster centers are decided, the transitive law is applied to find the largest collections of connected centers, and assign each collection of connected centers to a cluster. Thus, multi-center clusters are created. For example, if the centers $cc_i^c$ and $cc_j^c$ are connected, and the centers $cc_j^c$ and $cc_k^c$ also connected, then $cc_i^c$, $cc_j^c$, and $cc_k^c$ are connected. The centers $cc_i$, $cc_j$, and $cc_k$ are assigned to the same cluster. To detail the connection criterion in Eq. (9), let $cc_i^c$ and $cc_j^c$ be two centers from two different clusters, and let there be no other centers on the line segment connecting $cc_i^c$ and $cc_j^c$. It is reasonable to assume that the distance between $cc_i^c$ and $cc_j^c$ is larger than $5r_{\min}$. Recall that $r_{\min}$ is defined as the radius of the neighborhood of a point and $r_{\min}$ is the minimum standard deviation among all the dimensions. That is, two different clusters are expected to be well separated or just slightly overlapping. Since there is no prior information about the structure of data points, the distributions of the two centers, $cc_i^c$ and $cc_j^c$, are assumed to be approximated by a normal distribution. Thus, for each center, the neighborhood is supposed to contain around 68 percent of the data points which belong to this corresponding center. Let the two centers be further assumed to have almost equivalent number of points in their neighborhoods. In this case, we can find a point $v_i$ (between $cc_i^c$ and $cc_j^c$, and closer to $cc_i^c$ than $cc_j^c$) which contains less than 17 percent of points of any one center (16 percent from the neighborhood of the center $cc_i^c$, and 1 percent of points from the neighborhood of the center $cc_j^c$). Likewise, the density of $v_i$ is less than $\frac{0.17}{0.68} = \frac{1}{4}$ of the density of each center. On the other hand, if such a point $v_i$ cannot be found on the line segment between $cc_i^c$ and $cc_j^c$, then the centers $cc_i^c$ and $cc_j^c$ are said to be connected. Therefore, the criterion for the connection of two centers can be defined as follows.

If the density $D_{vk}$; $\forall k$ satisfies

$$D_{vk} > \frac{\max(D_{cci}, D_{ccj})}{4} \quad \forall k, \tag{9}$$

then the centers $cc_i^c$ and $cc_j^c$ are considered to be connected and assigned to the same cluster. Otherwise, $cc_i^c$ and $cc_j^c$ belong to two different clusters. Note that the connection criterion could be adjusted if there is prior information about the clusters and the structure of data points. The algorithm for the multi-center cluster detection is presented as follows.

**Algorithm 5** (Detection of multi-center cluster). (Associate related centers to the same cluster)
1. Assign the center $cc_l^c$ to the cluster $l$, $l = 1, 2, \ldots, \beta$. That is, $Cluster_l = \{cc_l^c\}$, $l = 1, 2, \ldots, \beta$.
2. Create two index sets $indx1$, $indx2$ with $indx1(i) = 0$ and $indx2(j) = 0$, for $i, j = 1, 2, \ldots, \beta$.
3. Let $i = 1$, and $j = i + 1$.

4. Find the equal distance points $v_1, v_2, \ldots, v_h$ which are on the line segment connecting centers $cc_i^c$ and $cc_j^c$ and lying between $cc_i^c$ and $cc_j^c$.
5. Calculate the densities $D_{cci}, D_{ccj}, D_{vk}$, $k = 1, 2, \ldots, h$.
6. If the density $D_{vk}$; $\forall k$ satisfies the condition in Eq. (9), then

   if $indx1(i) = 0$ and $indx2(j) = 0$,
   $Cluster_j = Cluster_j \cup Cluster_i$, (assume $cc_i^c$ and $cc_j^c$ to belong to the same cluster),
   $indx1(i) = j$, and $indx2(j) = i$.
   elseif $indx1(i) \neq 0$ and $indx2(j) = 0$,
   $Cluster_{\max(j, indx1(i))} = Cluster_{\max(j, indx1(i))} \cup Cluster_{\min(j, indx1(i))}$,
   $indx1(i) = \max(j, indx1(i))$, and $indx2(j) = i$.
   elseif $indx1(i) = 0$ and $indx2(j) \neq 0$,
   $Cluster_{indx1(indx2(j))} = Cluster_{indx1(indx2(j))} \cup Cluster_i$, and
   $indx1(i) = indx1(indx2(j))$.
   elseif $indx1(i) \neq 0$ and $indx2(j) \neq 0$, let $temp = indx1(indx2(j))$,
   $Cluster_{\max(indx1(i), temp)} = Cluster_{\max(indx1(i), temp)} \cup Cluster_{\min(indx1(i), temp)}$,
   and $indx1(i) = \max(indx1(i), temp)$.
   else $Cluster_i = \{cc_i^c\}$, $Cluster_j = \{cc_j^c\}$ (the centers $cc_i^c$ and $cc_j^c$ represent two different clusters).
7. If $i \leqslant \beta$ and $j \leqslant \beta$, go back to step 4 with $j = j + 1$.
   elseif $i \leqslant \beta$, go back to step 4 with $i = i + 1$ and $j = i + 1$.
   else (Find the number of clusters and the set of final clusters)
   (a) Let $i = 1$ again, and $FC = \emptyset$ (set of final clusters).
   (b) If $indx1(i) \neq 0$, then $\beta = \beta - 1$, and delete $Cluster_i$.
   else $FC = FC \cup i$.
   (c) $i = i + 1$.
   (d) If $i \leqslant \beta$, go back to step 7(b).
   else stop with the cluster number $\beta$, and the set of final clusters $FC$. □

To complete the fuzzy clustering process of a given data, the proposed fuzzy clustering algorithm is presented in the next section.

## 6. A complete algorithm for the proposed fuzzy clustering

With the three parts in the previous Sections 3–5, a complete algorithm is presented in this section for the proposed fuzzy clustering of a given data set.

**Algorithm 6** (A complete fuzzy clustering algorithm)

1. Let the collection of data points be $C_x = \{x_1, x_2, \ldots, x_n\}$. And each data point $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is a $p$-dimensional vector.
2. If the number of data points is large, Run the first level of HSC (Algorithm 2) for all the possible cluster centers,
   else go to step 4 with each data point being a possible cluster center.
3. Run the second level of HSC (Algorithm 3) to reduce the number of cluster centers, and to generate the primary cluster centers.
4. Run Algorithm 4 to select the candidate cluster centers from the set of primary cluster centers to minimize the objective function.
5. Run Algorithm 5 to associate the related centers to the same cluster. Thus, the final number of clusters and the final clusters are obtained.

6. (This step may be by taking the centers of the final clusters as the final cluster centers and going to step 7 directly.) Run the standard fuzzy-$C$ means algorithm with the centers of the final clusters as initial cluster centers and then stop.
7. (if step 6 is skipped) The membership values $\mu_{ji}$ of data point $x_i$; $\forall i$ associated with cluster center $cc_j^c$; $\forall j$ are calculated with

$$\mu_{ji} = \frac{1}{\sum_{k=1}^{n}(\|(x_i - cc_j^c)\|^2/\|(x_i - cc_k^c)\|^2)^{2/(m-1)}} \quad \forall i, j,$$

and then stop. □

Note that the membership value of the data point $x_i$ in a multi-center cluster is the sum of the membership values of the data point $x_i$ associated with cluster centers which are assigned to represent the multi-center cluster. To indicate the effectiveness of the proposed unsupervised fuzzy clustering algorithm, data with unequal population clusters, two-elliptic clusters, linearly unseparable clusters, the iris data, and a color image are clustered in Section 7.

## 7. Experimental results

Three sets of two dimensional data with unequal population clusters, two elliptic clusters, and linearly unseparable clusters are generated for experiments. The iris data is also clustered in this section. Moreover, the classification of the color pepper image which contains $512 \times 512$ three-dimensional data points is implemented to indicate the time savings of the proposed fuzzy clustering algorithm compared to the subtractive clustering algorithm.

*Data with unequal population clusters*: Two unequal population clusters with 50 and 10 points are given in Fig. 1. The centers obtained from the HSC algorithm are shown with data points in Fig. 2. The candidate cluster centers are selected from the centers in Fig. 2 to minimize the objective function. And the candidate cluster centers are presented in Fig. 3. With the multi-center cluster detection algorithm (Algorithm 5), the

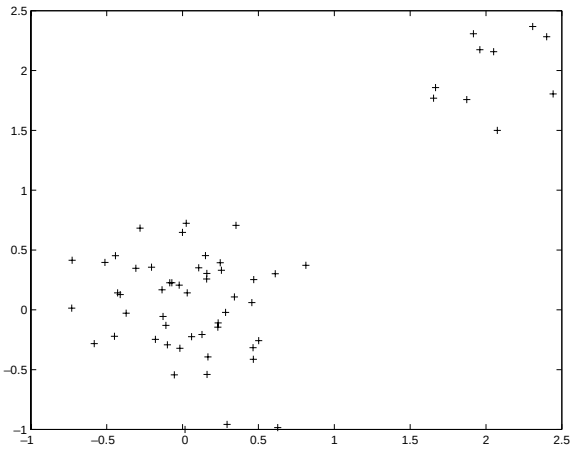


Fig. 1. Data with unequal population clusters.

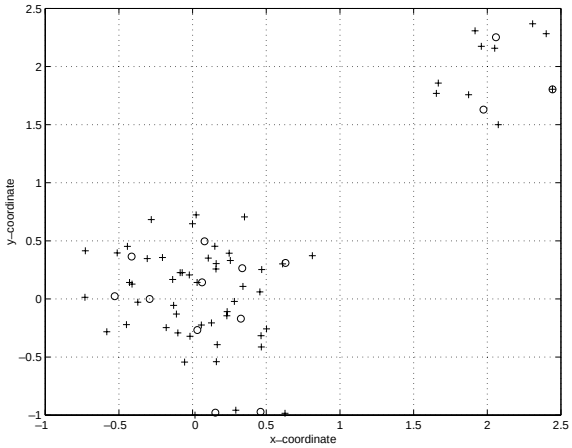


Fig. 2. Cluster centers ('o') and data with unequal population clusters ('+').

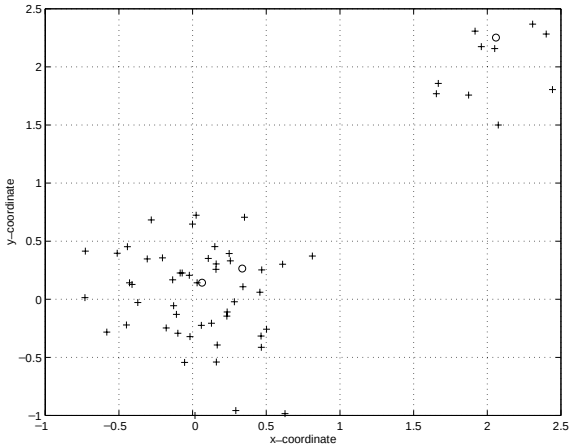


Fig. 3. Primary cluster centers ('o') and data with unequal population clusters ('+').

number of cluster is determined to be two. The coordinates of the centers are

$$\begin{bmatrix} 0.0630 & 0.1419 \\ 0.3378 & 0.2637 \end{bmatrix} \quad \text{and} \quad [2.0588 \quad 2.2521],$$

respectively. Note that there is a cluster with two centers. With the candidate cluster centers, the clustering is implemented by using the fuzzy $C$-means algorithm. For this set of data, the proposed clustering algorithm can cluster the data correctly without any misclustered data points.

*Data with two elliptic clusters*: Two elliptic clusters with 140 points are shown in Figs. 4 and 5 indicates the centers obtained from the HSC algorithm with data points. The candidate cluster centers are selected with Algorithm 4 to minimize the objective function. Fig. 6 shows the candidate cluster centers. The number of clusters is determined to be two. Each of the clusters has two centers. Two elliptic clusters are clustered correctly without any misclustered data points.

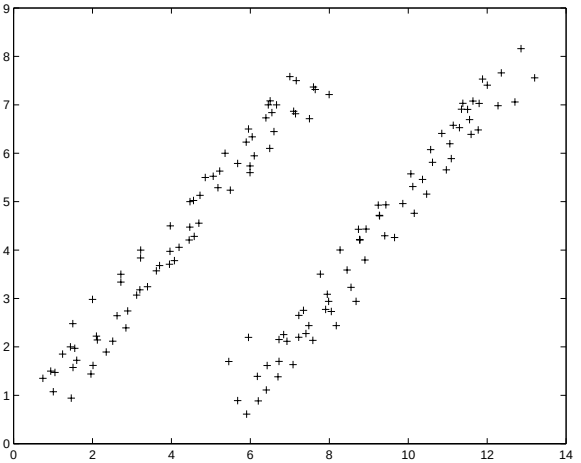

Fig. 4. Data with two elliptic clusters.

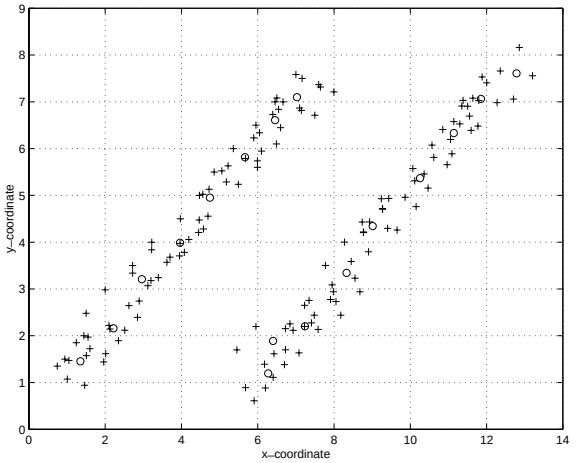


Fig. 5. Cluster centers ('o') and data with two elliptic clusters ('+').

*Data with linearly unseparable clusters*: In Fig. 7, data with linearly unseparable clusters is created. The HSC algorithm produces the cluster centers in Fig. 8. Five candidate cluster centers are denoted with data in Fig. 9. Four centers are assigned to the same cluster; and the number of clusters is two. No data points are misclustered.

*The iris data*. The iris data is obtained from three different species, and there are 50 data points for each specie. The 150 four-dimensional iris data are clustered with the proposed clustering algorithm. The candidate cluster centers are

$$\begin{bmatrix} 5.1 & 3.43 & 1.44 & 0.21 \\ 5.65 & 2.75 & 4.15 & 1.3 \\ 6.7 & 3.1 & 4.7 & 1.5 \\ 6.7 & 3.3 & 5.7 & 2.1 \end{bmatrix},$$

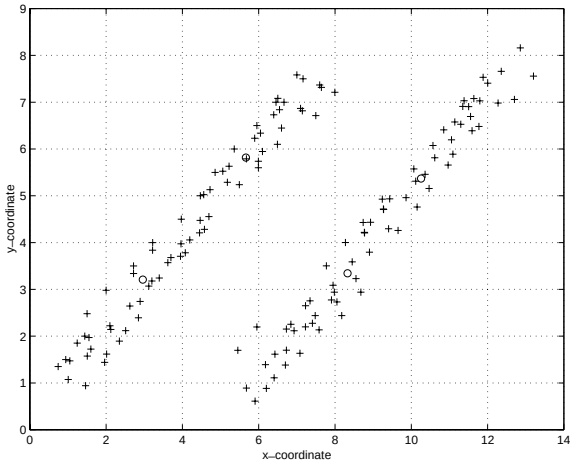


Fig. 6. Primary cluster centers ('o') and data with two elliptic clusters ('+').

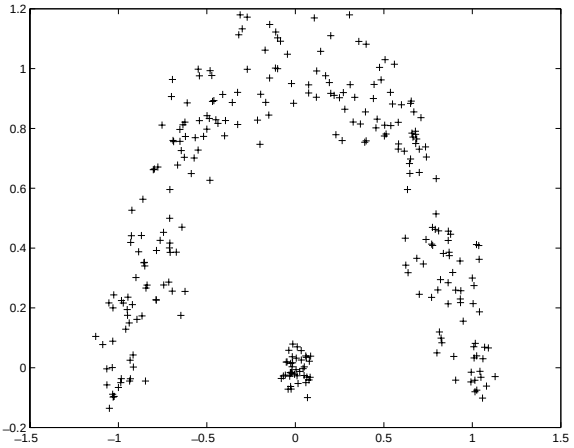


Fig. 7. Data with linearly unseparable clusters.

and the second and third centers (the second and third columns) are assigned to the same cluster. Thus, the number of clusters is three. With these candidate cluster centers, the first and second species can be correctly clustered without any misclustered data points. For the third specie, 13 data points out of 50 are misclassified, i.e., the percentage of error classification is 26 percent. Note that the percentage of classification error is because the clustering algorithm is not optimized for classification accuracy but rather for clustering soundness.

*The color image of peppers*: To emphasize the low computation load (compared to the subtractive clustering) and to show the application of the proposed clustering algorithm to image segmentation, the color peppers image is used as the data set for clustering. The image data set contains $512 \times 512$ three-dimensional (RGB) data points. For the first level HSC, we separate the image data set into 512 subsets, and each subsets includes 512 points. With the clustering algorithm (coded in Matlab), it takes 20 min to run through the first level of HSC, and 1432 cluster centers are obtained. The second level of HSC is then applied with this

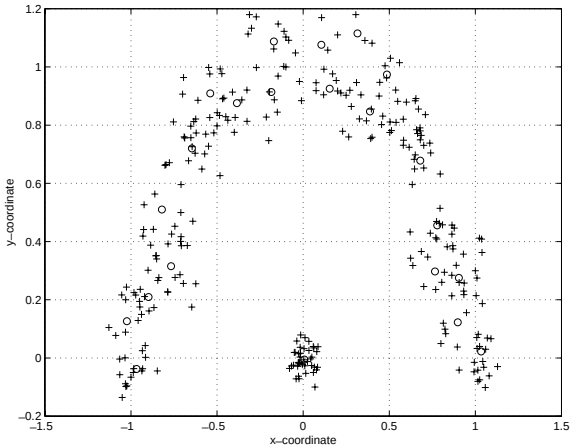


Fig. 8. Cluster centers ('o') and data with linearly unseparable clusters ('+').

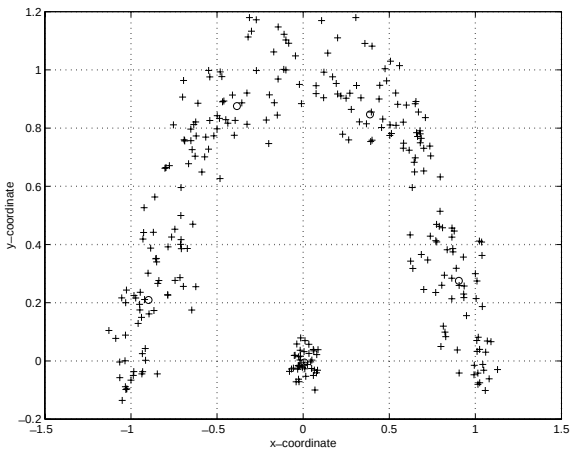


Fig. 9. Primary cluster centers ('o') and data with linearly unseparable clusters ('+').

collection of cluster centers, and the processing time of the second level of HSC is about 25 min. Note that because the number of data points in the second level of HSC (1432) is about 3 times larger than the number of points in one subset in the first level of HSC (512), the processing time of the second level of HSC is even larger than the processing time of the first level of HSC. There are 412 primary cluster centers generated by the second level of HSC. It takes another 15 min to select the 6 candidate cluster centers from the primary cluster centers. The candidate cluster centers are listed as

$$\begin{bmatrix} 123.6 & 172.7 & 98.8 \\ 120.7 & 118.6 & 52.4 \\ 31.9 & 21.1 & 0 \\ 188.6 & 212.1 & 165.4 \\ 186.4 & 58.4 & 26.5 \\ 135.9 & 165.1 & 90.2 \end{bmatrix},$$


Fig. 10. The image "peppers" used in the experiment.

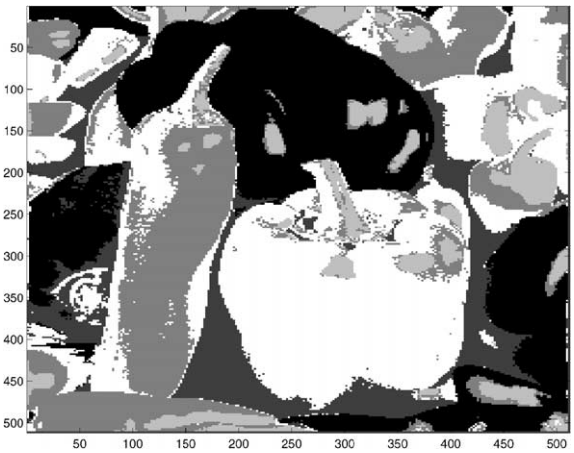


Fig. 11. Segmentation result of the color pepper image with the proposed algorithm.

and the (second, and sixth), candidate cluster centers are found to be associated with the same cluster. Thus, the final number of clusters is 5. The fuzzy-$C$ means algorithm is used to cluster the original peppers image data sets with the set of candidate cluster centers as the initial cluster centers. For simplicity, the original color peppers image is shown in Fig. 10. The clusters are shown with different gray levels $(0, 62, 125, 187, 255)$ shown in Fig. 11. From Fig. 11, it can be seen that the red area is clearly indicated to be black points (gray level 0). The green color area is clustered and shown with white points (255). The yellow area is represented by gray level 125. The area with high reflection is presented with gray level (187). The shadow area in the image is indicated as gray level (62). Note that the total amount of time taken for the proposed unsupervised fuzzy clustering algorithm to get the final results is around 1.5 h. However, if the original subtractive clustering algorithm is used, it cannot even generate one cluster during 1.5 h. From our experience in clustering the image data, the extension of HSC algorithm to have more than two levels is expected to be able to reduce the computation load for huge data sets (e.g. data mining clustering).

Table 1
Comparison of clustering algorithms (the numbers in the cell indicate the error percentage)

| Data type | Clustering algorithms | | | |
|---|---|---|---|---|
| | FCM (%) | GK (%) | FCS (%) | Proposed algorithm |
| Unequal population clusters | 0 | 0 | 18.3 | 0 |
| Two elliptic clusters | 27.1 | 0 | 24.3 | 0 |
| Linearly unseparable clusters | 42.7 | 39.9 | 13.9 | 0 |

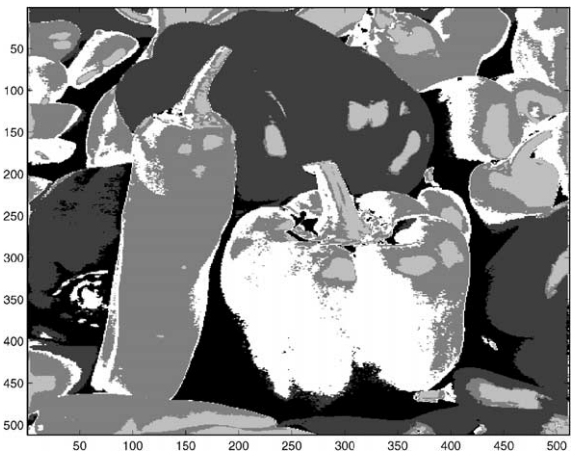


Fig. 12. Segmentation result of the color pepper image using plain FCM algorithm.

*Comparison with other clustering algorithms*: To show that the proposed unsupervised fuzzy clustering algorithm is more effective for clusters of various shape than some other "traditional" clustering algorithms, the data with unequal population clusters, two elliptic clusters, and linearly unseparable clusters are also clustered with other clustering algorithms (fuzzy-$C$ means (FCM), Gustafson–Kessel (GK) [17], fuzzy-$C$ shell (FCS) algorithms [26]). The correct cluster number and the fair initial conditions are given for clustering data with Algorithms FCM, GK, and FCS. The clustering results are listed in the Table 1. Moreover, with the cluster number assumed to be known as $c = 5$, and fairly selected initial membership value matrix, the color pepper image is segmented with the plain fuzzy-$C$ means algorithm for comparison. From Fig. 12, it can be seen that segmentation result with plain fuzzy-$C$ means algorithm is similar to the result with the proposed algorithm (see Fig. 11). However, if the initial membership value matrix is improperly selected, the degraded segmentation result will be obtained with the plain fuzzy-$C$ means algorithm in Fig. 13. Note that with the advantage of the matrix operation in Matlab, the segmentation of color pepper image with the plain fuzzy-$C$ means algorithm takes about 7 min.

## 8. Conclusions

In this paper, a new unsupervised fuzzy clustering algorithm is presented. A new objective function is defined. The number of clusters can be reasonably determined automatically by minimizing the objective function. And the final clusters with multi-centers are generated for the data. Also, the centers of the final

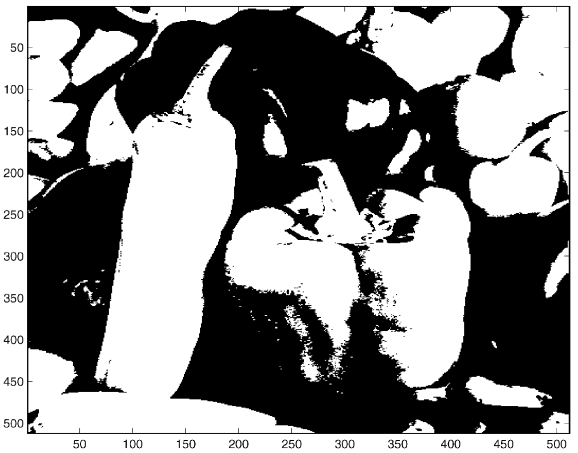


Fig. 13. Segmentaion result of the color pepper image using plain FCM algorithm (with improper initial membership value matrix).

clusters can form good initial cluster centers for the fuzzy-$C$ means algorithm. Since multi-centers are used to represent clusters, the nonspherical clusters can be correctly detected. Moreover, the proposed clustering algorithm can handle non-traditional, curved clusters. Note that the merging criteria used in this work prefers a cluster with a "thick" shape than a thin curve. Also, our clustering algorithm might have difficulty in clustering highly overlapped clusters. The clustering algorithm in this paper is efficient when the number of data points is large. However, for huge data sets (e.g. data mining sets), the clustering algorithm with HSC may still suffer from a heavy computation load. Thus, the extension of HSC to be more than two levels can be considered. The experimental results indicate the effectiveness of the proposed clustering algorithm.

## References

[1] J.C. Bezdek, Cluster validity with fuzzy sets, J. Cybernet. 3 (1974) 58–73.
[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function, Plenum, New York, 1981.
[3] D.A. Binder, Bayesian cluster analysis, Biometrika 65 (1978) 31–38.
[4] S.L. Chiu, Fuzzy model identification based on cluster estimation, J. Intell. Fuzzy Systems 2(3) (1994) 267–278.
[5] R.N. Dave, R. Krishnapuram, Robust clustering methods: a united view, IEEE Trans. Fuzzy Systems 5(2) (1997) 270–293.
[6] N.E. Day, Estimating the components of a mixture of two normal distribution, Biometrika 56 (1969) 463–474.
[7] R.O. Duda, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[8] J.C. Dunn, Well-separated clusters and optimal fuzzy partitions, J. Cybernet. 4 (1974) 95–104.
[9] Y. EL-Sonbaty, M.A. Ismail, Fuzzy clustering for symbolic data, IEEE Trans. Fuzzy Systems 6(2) (1998) 195–204.
[10] M. Ester, H.P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, Proc. 2nd Internat. Conf. on Knowledge Discovery and Data Mining, 1996, pp. 226–231.
[11] E. Forgy, Cluster analysis of multivariate data: efficiency versus interpretability of classification, Biometrics 21 (1965) 768.
[12] H.P. Friedman, J. Rubin, On some invariant criteria for grouping data, J. Amer. Statist. Assoc. 62 (1981) 1159–1178.
[13] J.H. Friedman, L.C. Rafsky, Graphics for the multivariate two-sample problem, J. Amer. Statist. Assoc. 76 (1981) 277–287.
[14] I. Gath, A.B. Geva, Unsupervised optimal fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 11(7) (1989) 773–781.
[15] I. Gitman, M.D. Levine, An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique, IEEE Trans. Comput. 19(7) (1970) 583–593.
[16] A.D. Gordon, J.T. Henderson, Algorithm for Euclidean sum of squares classification, Biometrics 33 (1977) 355–362.
[17] E.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, Proc. IEEE CDC Conf., 1979, pp. 761–766.
[18] F. Hoeppner, Fuzzy shell clustering algorithms in image procesing: fuzzy $c$-rectangular and 2-rectangular shells, IEEE Trans. Fuzzy Systems 5(4) (1997) 599–613.
[19] F. Hoppner, F. Klawonn, R. Kruse, T. Runkler, Fuzzy Cluster Analysis, Wiley, Englewood Cliffs, NJ, 1999.
[20] L.J. Hubert, Some applications of graph theory to clustering, Psychometrika 39 (1974) 698–704.

[21] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[22] J.-S.R. Jang, C.-T. Sun, E. Mizutani, Neural-Fuzzy and Soft Computing, Prentice-Hall, Englewood Cliffs, NJ, 1997.
[23] N. Jardine, R. Sibson, Mathematical Taxonomy, Wiley, New York, 1971.
[24] J.M. Jolion, P. Meer, S. Bataouche, Robust clustering with applications in computer vision, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 791–802.
[25] J. Kittler, A local sensitive method for cluster analysis, Pattern Recognition 8 (1976) 22–33.
[26] R. Krishnapuram, O. Nasraoui, H. Frigui, The fuzzy $C$ spherical shells algorithms: a new approach, IEEE Trans. Neural Networks 3 (1992) 663–671.
[27] A. Lazarevic, X. Xu, T. Fiez, Z. Obradovic, Clustering-regression-ordering steps for knowledge discovery in spatial database, Proc. Internat. Joint Conf. on Neural Network, 1999, pp. 2530–2534.
[28] G.J. McLachlan, The classification and mixture maximum likelihood approaches to cluster analysis, in: Handbook of Statistics, Vol. 2, 1982, pp. 199–208.
[29] J.B. McQueen, Some methods of classification and analysis of multivariate observations, Proc. of 5th Berkley Symp. on Mathematical Statistics and Probability, 1994, pp. 281–297.
[30] E.H. Ruspini, A new approach to clustering, Inf. Control 15 (1969) 22–32.
[31] P.K. Simpson, Fuzzy min-max neural network — part 2: clustering, IEEE Trans. Fuzzy Systems 1(1) (1993) 32–45.
[32] P. Smyth, Clustering using Monte Carlo cross-validation, Proc. KDD-96, 1996, pp. 126–133.
[33] P.H. Sneath, R.R. Sokal, Numerical Taxonomy, W.H. Freeman Company, San Francisco, CA, 1973.
[34] M.J. Symons, Approximate clustering criteria and multivariate normal mixtures, Biometrics 37 (1981) 35–43.
[35] C.W. Therrien89, Decision, Estimation, and Classification, Wiley, Englewood Cliffs, NJ, 1989.
[36] S. Thodoridis, K. Koutroumbas, Pattern Recognition, Academic Press, San Diego, 1999.
[37] A.I. Torn, Cluster analysis using seed points and density-determined hyperspheres as an aid to global optimization, IEEE Trans. Systems Man Cybernet. 24 (1994) 1279–1284.
[38] M.P. Windham, Cluster validity for the Fuzzy $c$-means clustering algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1982) 357–363.
[39] J.H. Wolfe, Pattern clustering by multivariate mixture analysis, Multivar. Beha. Res. 5 (1970) 329–350.
[40] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13(8) (1991) 841–847.
[41] X. Xu, M. Ester, H.P. Kriegel, J. Sander, A distribution-based clustering algorithm for mining in large spatial databases, Proc. 14th Internat. Conf. on Data Engineering, 1996, pp. 324–331.
[42] R.R. Yager, D.P. Filev, Approximate clustering via the mountain method, IEEE Trans. Systems, Man Cybernet. 24 (1994) 1279–1284.
[43] L.A. Zadeh, Fuzzy sets, Inform. and Control 8(3) (1965) 338–353.
[44] X. Zhuang, Y. Huang, K. Palaniappan, Y. Zhao, Gaussian mixture density modeling, decomposition and applications, IEEE Trans. Image Process. 5 (1996) 1293–1302.