Alternative c-means clustering algorithms

Kuo-Lung Wu, Miin-Shen Yang *

Department of Mathematics, Chung Yuan Christian University, Chung-Li 32023, Taiwan

Received 23 October 2000; received in revised form 2 October 2001

Abstract

In this paper we propose a new metric to replace the Euclidean norm in c-means clustering procedures. On the basis of the robust statistic and the influence function, we claim that the proposed new metric is more robust than the Euclidean norm. We then create two new clustering methods called the alternative hard c-means (AHCM) and alternative fuzzy c-means (AFCM) clustering algorithms. These alternative types of c-means clustering have more robustness than c-means clustering. Numerical results show that AHCM has better performance than HCM and AFCM is better than FCM. We recommend AFCM for use in cluster analysis. Recently, this AFCM algorithm has successfully been used in segmenting the magnetic resonance image of Ophthalmology to differentiate the abnormal tissues from the normal tissues. © 2002 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Hard c-means; Fuzzy c-means; Alternative c-means; Fixed-point iterations; Robustness; Noise

1. Introduction

Cluster analysis is a method for clustering a data set into groups of similar individuals. It is a branch in multivariate analysis and an unsupervised learning in pattern recognition. The clustering applications in various areas such as taxonomy, medicine, geology, business, engineering systems and image processing, etc., are well documented (see for example Refs. [1–3]). In these clustering methods the hard c-means (or called k-means) are the most well-known conventional (hard) clustering methods which restrict each point of the data set to exactly one cluster.

Since Zadeh [4] proposed fuzzy set theory which produced the idea of uncertainty of belonging described by a membership function, fuzzy clustering has been widely studied and applied in a variety of substantive areas (see Refs. [5,6]). In the fuzzy clustering literature, the fuzzy c-means clustering algorithm, proposed by Dunn [7] and extended by Bezdek [6] is the most used and discussed (see Refs. [8–10]). Fuzzy c-means (FCM) are extensions of hard c-means (HCM). FCM has been shown to have better performance than HCM. FCM has become the most well-known and powerful method in cluster analysis. However, these FCM algorithms have considerable trouble in a noisy environment and inaccuracy with a large number of different sample sized clusters. A good clustering algorithm should be robust and able to tolerate these situations that often happen in real application systems.

In this paper we propose a new metric. The robustic property of this new metric is discussed based on the statistical point of view with the influence function. This proposed metric is more robust than the common-used Euclidean norm. We then replace the Euclidean norm with the new metric in c-means clustering. Thus, we created two new clustering methods called the alternative hard c-means (AHCM) and alternative fuzzy c-means (AFCM) clustering algorithms. These proposed algorithms actually improve the weaknesses in HCM and FCM. In Section 2 the new metric is presented and its properties are discussed. We then claim that the proposed new metric is more robust than the Euclidean norm on the basis of the robust statistic and the influence function. In Section 3, based on the new metric, we propose the AHCM clustering algorithm and create the AFCM in Section 4. Numerical examples and comparisons between these

* Corresponding author. Tel.: +886-3-456-3171x3119; fax: +886-3-456-3160.

E-mail address: msyang@math.cycu.edu.tw (M.-S. Yang).

algorithms are made. Finally, we make conclusions in Section 5.

2. A new metric

The concept of a metric space is basic and important in Mathematics. Different metric functions construct different metric spaces. Based on these metric spaces, mathematical theorems will be studied and applied in various areas of applications. The performance of these application is always affected by different chosen metrics. The Euclidean norm is well known and commonly used as a metric. However, the parameter estimate resulting from an objective function based on this Euclidean metric may not be robust in a noisy environment. Suppose that $X = \{x_1, \ldots, x_n\}$ is a data set where x_j is a feature vector in m-dimensional Euclidean space R^m, a good estimate of the center z can be acquired using a minimum mean square error procedure. Thus, minimizing

$$\sum_{j=1}^{n} ||x_j - z||^2 \tag{1}$$

with respect to z, we have a sample mean minimizer

$$z = \frac{\sum_{j=1}^{n} x_j}{n}. \tag{2}$$

Here, the distance measure between x_j and center z is the Euclidean norm. This method of finding estimators in statistic is called the least-squares (LS) method. An artificial data set $\{3, 4, 4.4, 4.7, 4.9, 5, 5.1, 5.3, 5.6, 6, 7\}$ is tested using this procedure. This data set is tight in the data point 5 and the estimate of z is 5 by solving Eq. (2). However, when we add a noisy point in the data set with the value 30, this procedure obtains its estimate as $z = 7.08$, which is outside the original data range. This is obviously not robust because the result is heavily affected by the noisy point. These results are shown in Figs. 1(a) and (c). The curve in Fig. 1 is made with the values of Eq. (1) with respective to the estimate z. The location of the minimizer is calculated with Eq. (2). Obviously, the minimizer in Fig. 1(c) is affected by the noise and is quite different from the original estimate 5 with its new estimate 7.08.
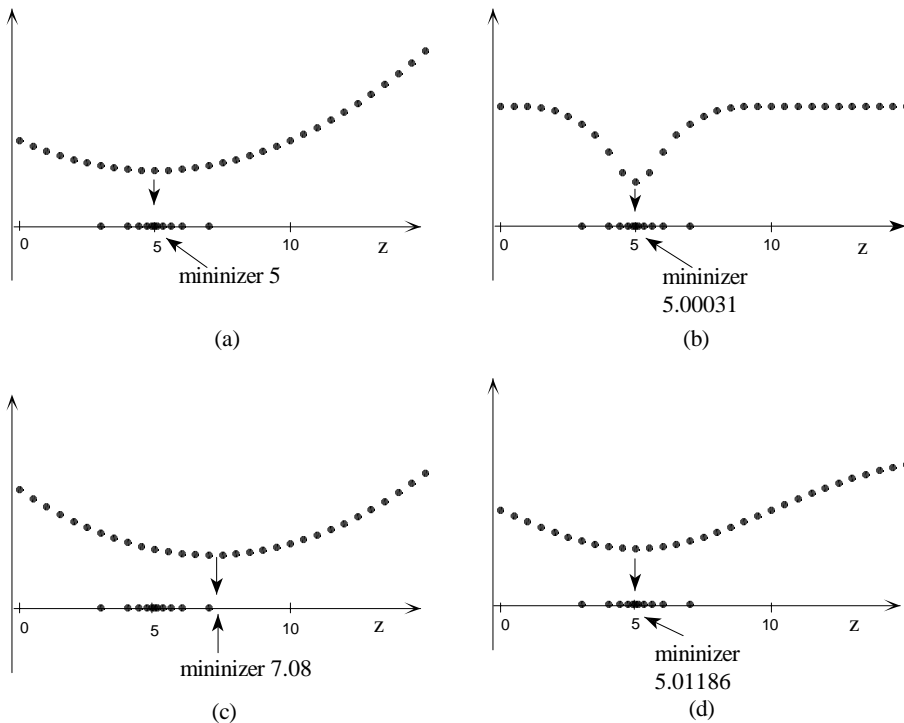

Fig. 1. The acquired center z using a minimum mean square error procedure with different metrics. (a) and (b) are results for a noise-free data set (the solid circle points) under the Euclidean norm and the proposed new metric, respectively. (c) and (d) are results using the Euclidean norm and the proposed new metric in a noisy environment with the added noise value 30. The curves in (a) and (c) are made with the values of Eq. (1) with respective to the estimate z. The curves in (b) and (d) are made with the values of Eq. (4) with respective to the estimate z.

Note that the sample mean resulting from Eq. (2) can be written as

$$z = \sum_{j=1}^{n} \frac{w_j}{\sum_{j=1}^{n} w_j} x_j \tag{3}$$

with equal weights $w_j = 1$ even though a data point may be far away from other data points. Thus, if we can give a small weight to those noisy points or outliers and a large weight to those compact points in the data set, then the weighted sum of x_j shall be more robust. In order to accomplish this goal, we propose a new distance function as

$$d(x, y) = 1 - \exp(-\beta||x - y||^2), \tag{4}$$

where β is a positive constant. Note that the distance function $d(x, y)$ is a monotone increasing function of $||x - y||$. We know that any distance function $d(x, y)$ is a metric if it satisfies the following three conditions (see Rudin [11]):

(i) $d(x, y) > 0 \quad \forall x \neq y, \quad d(x, x) = 0,$
(ii) $d(x, y) = d(y, x),$
(iii) $d(x, y) \leqslant d(x, z) + d(z, y) \quad \forall z.$

We will claim that the new distance function (4) is a metric.

Property 1. The distance $d(x, y) = 1 - \exp(-\beta||x - y||^2)$ is a metric.

Proof. Conditions (i) and (ii) are satisfied obviously. It only needs to prove the triangle inequality (iii).

$$d(x, z) + d(z, y) - d(x, y)$$
$$= 1 - \exp(-\beta||x - z||^2) - \exp(-\beta||z - y||^2)$$
$$\quad + \exp(-\beta||x - y||^2)$$
$$\geqslant 1 - \exp(-\beta||x - z||^2) - \exp(-\beta||z - y||^2)$$
$$\quad + \exp(-\beta(||x - z||^2 + ||z - y||^2))$$
$$= \{1 - \exp(-\beta||x - z||^2)\}\{1 - \exp(-\beta||z - y||^2)\}$$
$$\geqslant 0.$$

Thus, $d(x, y) \leqslant d(x, z) + d(z, y) \ \forall z$. The proof is completed. $\square$

Under the metric (4), we have an estimate z using the same procedure with minimizing

$$\sum_{j=1}^{n} \{1 - \exp(-\beta||x_j - z||^2)\} \tag{5}$$

with respect to z. It gives the necessary conditions with the equation

$$z = \sum_{j=1}^{n} \frac{\exp(-\beta||x_j - z||^2)}{\sum_{j=1}^{n} \exp(-\beta||x_j - z||^2)} x_j$$

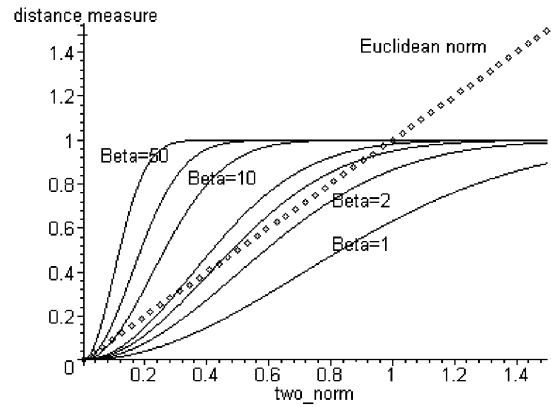$$= \sum_{j=1}^{n} \frac{w_j}{\sum_{j=1}^{n} w_j} x_j. \tag{6}$$

Fig. 2. The distance measure plot for the proposed new metric with different Beta. The Euclidean norm presents a straight line.

As we expect, it gives larger weights w_j to those x_j which are closer to z and smaller weights to those data points far away from z. Note that z in Eq. (6) cannot be solved directly. However, we can use the fixed-point iterative method to approximate it.

Fixed-point iteration. *Let the right-hand side of Eq. (6) be $f(z)$. The first step is to specify the initial value $z^{(0)}$ and then compute $f(z^{(0)})$ and set it to be $z^{(1)}$. Repeat the step until the $(l + 1)$th solution $z^{(l+1)}$ is very close to the lth solution.*

The same artificial data set is tested using the new procedure and the results are shown in Figs. 1(b) and (d). Based on the new metric (4), we obtain a more robust estimate of z with z very close to 5 in both noise-free and noise cases.

In Fig. 2, we plot the distance measure with the Euclidean norm and the proposed metric (4) with different β. We see that this new metric is bounded and monotone increasing with distance measure zero as $||x_j - z|| = 0$ and distance measure one as $||x_j - z||$ tends to infinite. If $||x_j - z||$ is larger than a level (i.e. x_j keeps away from z) then the distance measure will be close to its maximum value and give a small weight to x_j. However, the distance measure with the Euclidean norm presents a straight line. As we see in Fig. 2, if the normalized parameter β tends to infinity, each data point will have no neighborhood. That is, the distance between each point will attach its maximum distance measure. Each point become an individual because the distances between them are all the farthest. If β tends to zero, the space will degenerate to a point. The distances between them are the smallest distance measure zero.

A good clustering algorithm should be robust and be able to tolerate noise and outliers that often happen in real application systems. Therefore, many robust statistics were developed to accomplish this objective, such as the M-, L- and R-estimators [12]. We will show our estimator resulting from Eq. (5) is an M-estimator and robust according to the influence function analysis.

Let $\{x_1, \ldots, x_n\}$ be an observed data set and θ is an unknown parameter to be estimated. An M-estimator is generated by minimizing the from

$$\sum_{j=1}^{n} \rho(x_j; \theta), \tag{7}$$

where ρ is an arbitrary function that can measure the loss of x_j and θ. Suppose that $X_1, \ldots, X_n$ denote a random sample from the probability density function (p.d.f.) $f_x(x; \theta)$ and we take $\rho(x_j; \theta) = -\log f_x(x; \theta)$. The M-estimator is then equivalent to the maximum likelihood estimator (MLE). Here, we are interested in a location estimate that minimizes

$$\sum_{j=1}^{n} \rho(x_j - \theta) \tag{8}$$

and the M-estimator is generated by solving the equation

$$\sum_{j=1}^{n} \varphi(x_j - \theta) = 0, \tag{9}$$

where $\varphi(x_j - \theta) = (\partial/\partial\theta)\rho(x_j - \theta)$. If we take $\rho(x_j - \theta) = (x_j - \theta)^2$, the M-estimator is the sample mean which is equivalent to the classical least-squares (LS) estimator and if $\rho(x_j - \theta) = |x_j - \theta|$, the M-estimator is the median. Eq. (9) can be solved by rewritting as

$$\sum_{j=1}^{n} w_j(x_j - \theta) = 0, \tag{10}$$

where

$$w_j = \varphi(x_j - \theta)/(x_j - \theta) \tag{11}$$

is called the weight function. This gives the M-estimator $\hat{\theta}$ as a weighted mean

$$\hat{\theta} = \sum_{j=1}^{n} \frac{w_j}{\sum_{j=1}^{n} w_j} x_j. \tag{12}$$

Note that the results by solving Eq. (12) may not be a close form for $\hat{\theta}$. We can apply the fixed-point iteration or Newton's method to obtain a solution to Eq. (12) iteratively.

The influence function or influence curve (IC) can help us to assess the relative influence of individual observations toward the value of an estimate. The M-estimator has been shown that its influence function is proportional to its φ function [12]. In the location problem, we have the influence function of an M-estimator with

$$\text{IC}(x; F, \theta) = \frac{\varphi(x - \theta)}{\int \varphi'(x - \theta) \, dF_X(x)}, \tag{13}$$

where $F_X(x)$ denotes the distribution function of X. If the influence function of an estimator is unbounded, an outlier might cause trouble. Many important measures of robustness can be observed from the influence function. One of these important measures is the gross error sensitivity γ^*, defined by

$$\gamma^* = \sup_x |\text{IC}(x; F, \theta)|. \tag{14}$$

This quantity can interpret the worst approximate influence that the addition of an infinitesimal point mass can have on the value of the associated estimator. Our estimate of z (replaced by θ here) resulting from Eq. (5) is an M-estimator with

$$\rho(x - \theta) = 1 - \exp(-\beta(x - \theta)^2) \tag{15}$$

and

$$\varphi(x - \theta) = \frac{-2\beta(x - \theta)}{\exp(\beta(x - \theta)^2)}. \tag{16}$$

By applying the L'Hospital's rule, we have

$$\lim_{x \to \infty} \varphi(x - \theta) = \lim_{x \to -\infty} \varphi(x - \theta) = 0 \tag{17}$$

we can also obtain the maximum and minimum values of $\varphi(x - \theta)$ by solving

$$(\partial/\partial x)\varphi(x - \theta) = 0. \tag{18}$$

According to above, the function $\varphi(x - \theta)$ with Eq. (16) is bounded and continuous. Thus, our new estimator has a bounded and continuous influence function and also has a finite gross error sensitivity. Hence it is robust.

The use of Euclidean norm, corresponding to the LS method, has $\rho(x - \theta) = (x - \theta)^2$ and $\varphi(x - \theta) = 2(\theta - x)$. The influence function of LS estimator is unbounded and the gross error sensitivity $\gamma^* = \infty$ and hence it is not robust to the outliers and noisy points. Huber [13] discusses a robust statistic, that has a finite gross error sensitivity, with

$$\rho_k(x - \theta) = \begin{cases} \frac{1}{2}(x - \theta)^2 & \text{if } |x - \theta| \leqslant k, \\ k|x - \theta| - \frac{1}{2}(x - \theta)^2 & \text{if } |x - \theta| > k, \end{cases} \tag{19}$$

which corresponds to a $\varphi_k(x - \theta)$ with

$$\varphi_k(x - \theta) = \begin{cases} (x - \theta) & \text{if } |x - \theta| \leqslant k, \\ k \, \text{sign}(x - \theta) & \text{if } |x - \theta| > k, \end{cases} \tag{20}$$

where $\text{sign}(t) = -1, 0, 1$ as $t <, =, > 0$. We now illustrate the difference of influence functions among LS estimator, Huber's statistic and our new estimator. The influence of an extremely large or small x will be very great in LS estimator. In Huber's statistic, the influence of $x \in [0 - k, \theta + k]$ is equivalent to the LS estimator but the influence outside the range will be restricted to a constant. The influence of an extremely large or small x of Huber's statistic is a finite constant that is equivalent to the influence of median when $k = 1$. However, an extremely large or small x will make no influence on our new estimator. This property tells us why the new estimator is robust to the outliers. The data point x which is far away from θ in both median and Huber's statistic has a finite constant influence. In our new

estimator, however, it will be classified to be a new observation which is from an unknown new population and hence no influence on our new estimator. The above discussion gives a theoretical foundation to asseverate the new clustering algorithms (see Sections 3 and 4) based on the proposed new metric are more robust than the Euclidean norm from the robust statistics point of view.

According to this new metric (4), we can accomplish the robustic property and then new alternative. hard c-means (AHCM) and alternative fuzzy c-means (AFCM) clustering algorithms will be established in Sections 3 and 4.

3. Alternative hard c-means clustering

A good clustering method will cluster data set $X = \{x_1, \ldots, x_n\}$ into c well partitions with $2 \leqslant c \leqslant n - 1$. Since we have no priori information about unlabelled data set X, a reasonable criteria or objective function is important for a clustering method. Intuitively, each cluster shall be as compact as possible. Thus, a well-known hard c-means (HCM) clustering objective function J_{HCM} is created with the Euclidean norm as

$$J_{\text{HCM}} = \sum_{i=1}^{c} \sum_{j \in I_i} ||x_j - z_i||^2, \tag{21}$$

where z_i is the ith cluster center. The necessary condition of minimizing J_{HCM} is

$$z_i = \frac{\sum_{j \in I_i} x_j}{|I_i|}, \tag{22}$$

where j is in the index set I_i if $||x_j - z_i||^2$ is a minimum of $\{||x_j - z_1||^2, \ldots, ||x_j - z_c||^2\}$. Thus, an HCM clustering algorithm is a Picard's fixed point iteration with Eq. (22) and the update index set I_i.

According to the analysis in Section 2, the Euclidean norm is sensitive to noise or outliers. Hence HCM clustering algorithm should be affected by noise and outliers. We will show it in Example 1. Now, the new proposed metric (4) is used to replace the Euclidean norm in the HCM objective function (21). Thus, an alternative hard c-means (AHCM) clustering objective function is proposed as

$$J_{\text{AHCM}} = \sum_{i=1}^{c} \sum_{j \in I_i} \{1 - \exp(-\beta ||x_j - z_i||^2)\}, \tag{23}$$

where β is a constant which can be defined by

$$\beta = \left(\frac{\sum_{j=1}^{n} ||x_j - \bar{x}||^2}{n} \right)^{-1} \quad \text{with } \bar{x} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{24}$$

and $j \in I_i$ if

$$1 - \exp(-\beta ||x_j - z_i||^2) = \min_k \{1 - \exp(-\beta ||x_j - z_k||^2)\},$$

$$k = 1, \ldots, c. \tag{25}$$

The effect of parameter β has been discussed in Section 2. In order to fix the effect of parameter β, we should define it as a constant using data points. We set it to be the inverse of sample covariance such as (24). Thus, the necessary condition of minimizing AHCM objective function is as follows:

$$z_i = \frac{\sum_{j \in I_i} (\exp(-\beta ||x_j - z_i||^2)) x_j}{\sum_{j \in I_i} \exp(-\beta ||x_j - z_i||^2)}. \tag{26}$$

According to this necessary condition, we have the exact AHCM clustering algorithm as follows:

Exact AHCM Algorithm. *Choose some initial values* $z_1, \ldots, z_c$.
Step 1. *Classify n data points by assigning them to the class of the smallest distance* (4).
Step 2. *Get new centers z_i with Eq.* (26) *using the fixed-point iteration.*
IF no new z_i is found, THEN stop;
ELSE go to step 1.

Note that, since z_i in the necessary condition Eq. (26) cannot be solved directly, we must use an iterative method to achieve step 2. It is an inefficient and time-consuming application. We use a one-step method to approximate it in the following second alternative hard c-means algorithm.

AHCM Algorithm. *Let $f(z_i)$ be the right term of Eq.* (26) *and set the iteration counter $l = 0$ and choose the initial values $z_i^{(0)}$, $i = 1, \ldots, c$. Give $\varepsilon > 0$.*
Step 1. *Classify the n data points by assigning them to the class of the smallest distance measure* (4).
Step 2. *Find $z_i^{(l+1)}$ by $z_i^{(l+1)} = f(z_i^{(l)})$.*
IF $\max_i ||z_i^{(l+1)} - z_i^{(l)}|| < \varepsilon$, THEN stop;
ELSE $l = l + 1$ and go to step 1.

Obviously, if both algorithms converge, then their solutions shall satisfy the necessary condition (26). In our simulations, the AHCM algorithm is faster than the Exact AHCM algorithm, but the results of both algorithms are almost the same. Thus, we recommend AHCM algorithm and also use it in our examples.

In AHCM, if $\beta \neq 0$ then Eq. (25) is equivalent to

$$||x_j - z_i||^2 = \min_k ||x_j - z_k||^2, \quad k = 1, \ldots, c. \tag{27}$$

As β tends to zero, $\exp(-\beta ||x_j - z_i||^2)$ tends to one. The necessary condition (26) will then tend to necessary condition (22). Thus, we have the connection between AHCM and HCM in which the AHCM will tend to HCM as β tends to zero. The following examples will show the robustness of the AHCM algorithm and a comparison between AHCM and HCM is also made.

Example 1. We add an outlier with its coordinate (100, 0) in the data set. The centers and partition results of HCM and AHCM are shown in Figs. 3(a) and (c), respectively. The solid black dots are cluster centers. The HCM cluster centers
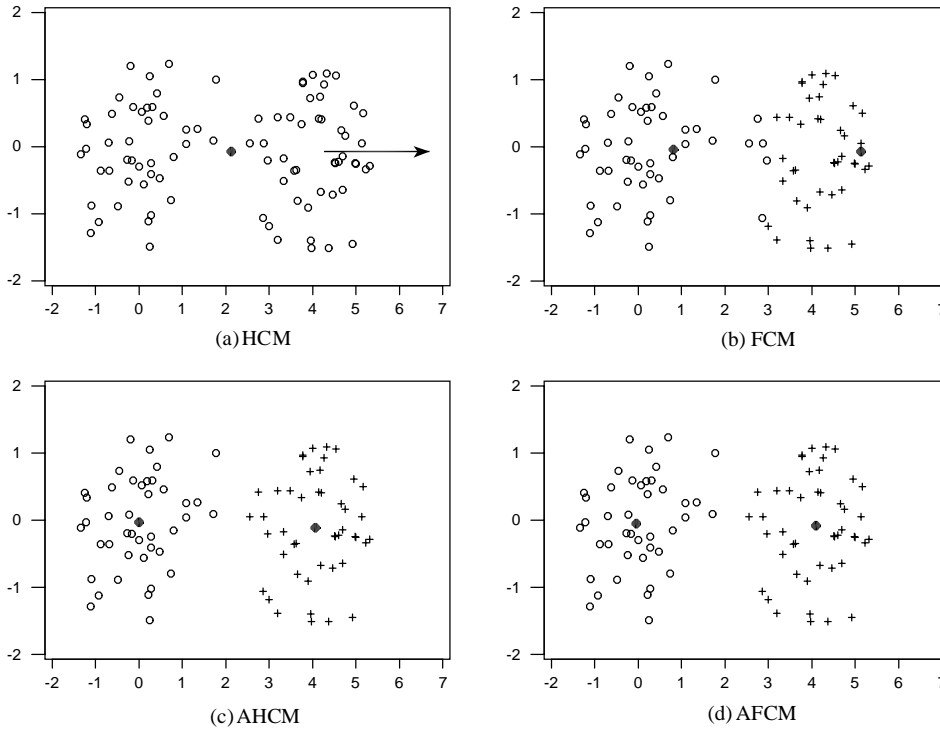
Fig. 3. Clustering results for HCM, FCM, AHCM and AFCM with added outlier coordinate $(100, 0)$.

are affected by this outlier in which the outlier becomes one center alone and the original two clusters become coincident. Note that this outlier center resulting from HCM is not visible in Fig. 3(a) so that we use the arrow to present this phenomenon. However, under the same initials and stopping conditions, the results of AHCM are robust with well two partitions not affected by the outlier which is shown in Fig. 3(c).

Example 2. This is a well-known clustering problem (see Ref. [1]) in that there are two great differences in numbers of sample clusters. The large size cluster will be split because this reduces the sum-of-squared-error criterion (21). The results is shown in Fig. 4(a). Although there are many procedures that can improve this weakness in HCM, AHCM can solve this problem simply as shown in Fig. 4(c).

4. Alternative fuzzy c-means clustering

Since Zadeh [4] introduced the concept of fuzzy sets, the research on fuzzy clustering has been widely investigated (see Ref. [5]). In the literature on fuzzy clustering, the fuzzy c-means (FCM) clustering algorithm defined by Dunn [7] and generated by Bezdek [6] is the best-known and most powerful method in cluster analysis. In this section we will

propose an alternative fuzzy c-means (AFCM) clustering algorithm based on the proposed new metric. The connection and comparisons between AFCM and FCM will be made in this section.

Let $X = \{x_1, \ldots, x_n\}$ be a data set of n data points. Let c be a positive integer greater than one. A partition of X into c parts can be presented by mutually disjoint sets $X_1, \ldots, X_c$ such that $X_1 \cup \cdots \cup X_c = X$ or equivalently by the indicator functions $\mu_1, \ldots, \mu_c$ such that $\mu_i(x) = 1$ if $x \in X_i$ and $\mu_i(x) = 0$ if $x \notin X_i$ for all $i = 1, \ldots, c$. The set of indicator functions $\{\mu_1, \ldots, \mu_c\}$ is called a hard c-partition of clustering X into c clusters. Thus, the hard c-means (HCM) objective function J_{HCM} of (21) can be rewritten as

$$J_{HCM} = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} ||x_j - z_i||^2.$$

Now consider an extension to allow $\mu_i(x)$ to be membership functions of fuzzy sets μ_i on X assuming values in the interval $[0,1]$ such that $\sum_{i=1}^{c} \mu_i(x) = 1$ for all x in X. In this fuzzy extension, $\{\mu_1, \ldots, \mu_c\}$ is called a fuzzy c-partition of X. Thus, the well-known FCM objective function J_{FCM} is an extension of J_{HCM} as

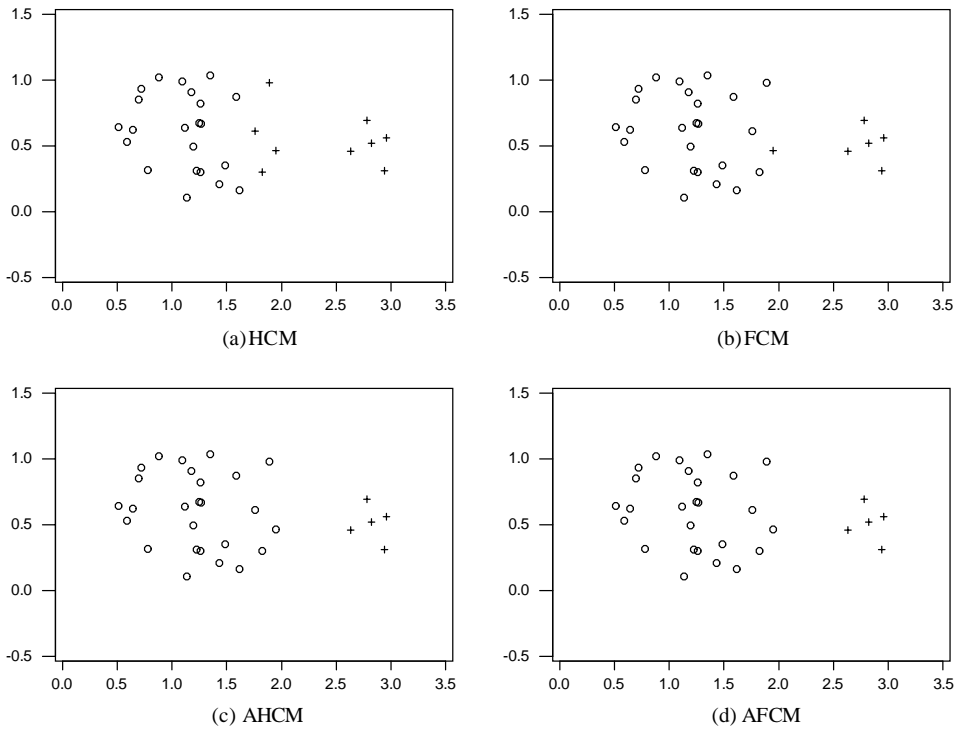$$J_{FCM} = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m ||x_j - z_i||^2,$$

Fig. 4. Two different sample sized clusters. (a), (b), (c) and (d) are results for HCM, FCM, AHCM and AFCM clustering algorithms, respectively.

where $m > 1$ is the index of fuzziness. The necessary conditions for a minimizer (μ, z) of J_{FCM} are the following updating equations:

$$\mu_{ij} = \frac{[1/||x_j - z_i||^2]^{1/(m-1)}}{\sum_{k=1}^{c} [1/||x_j - z_k||^2]^{1/(m-1)}} \tag{28}$$

and

$$z_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m x_j}{\sum_{j=1}^{n} (\mu_{ij})^m}. \tag{29}$$

Using the new metric, we propose a new alternative fuzzy c-means (AFCM) objective function as

$$J_{AFCM} = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m \{1 - \exp(-\beta ||x_j - z_i||^2)\} \tag{30}$$

with $m > 1$ and the restriction $\sum_{i=1}^{c} \mu_{ij} = 1$, $j = 1, \ldots, n$. Here the parameter β is also defined as Eq. (24). We then have the necessary conditions for minimizing J_{AFCM} as follows:

$$\mu_{ij} = \frac{[1/(1 - \exp(-\beta ||x_j - z_i||^2))]^{1/(m-1)}}{\sum_{k=1}^{c} [1/(1 - \exp(-\beta ||x_j - z_k||^2))]^{1/(m-1)}} \tag{31}$$

and

$$z_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m \exp(-\beta ||x_j - z_i||^2) x_j}{\sum_{j=1}^{n} (\mu_{ij})^m \exp(-\beta ||x_j - z_i||^2)}. \tag{32}$$

Similar to AHCM, we use the fixed-point iterative method to solve z_i in Eq. (32). Thus, if we solve μ_{ij} using Eq. (31) then we must use an iterative method to solve z_i before getting next μ_{ij}. We also use a one step method to approximate z_i in the AFCM algorithm.

AFCM Algorithm. Let $g(z_i)$ be the right term of Eq. (32) and set the iteration counter $l = 0$ and choose the initial values $z_i^{(0)}, i = 1, \ldots, c$. Give $\varepsilon > 0$

Step 1. Find $\mu_{ij}^{(l)}$ by Eq. (31);
Step 2. Find $z_i^{(l+1)}$ by $z_i^{(l+1)} = g(z_i^{(l)})$;
IF $\max_i ||z_i^{(l+1)} - z_i^{(l)}|| < \varepsilon$, THEN stop;
ELSE $l = l + 1$ and go to step 1.

The fuzzifier m in AFCM plays the same role as FCM which has been discussed in Ref. [6]. Fig. 5 shows different membership curves for different parameter β in the case of two clusters with one center of 0 and the other center of 2. The curve with circle points presents the FCM memberships. All membership curves are made with $m = 2$ for AFCM and FCM. They are all nonconvex fuzzy sets. If β is very small then the AFCM membership curve with this parameter β will be very close to the FCM membership curve which presents well for fuzzy boundary. If β is very large, the AFCM membership curve with this parameter β
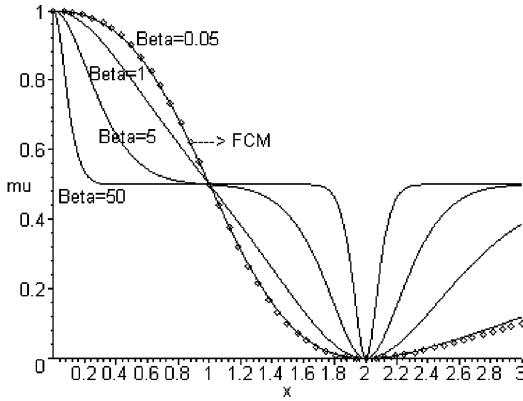
Fig. 5. The AFCM membership curves for different parameter *Beta* based on the data set with one center of 0 and the other center of 2. Circle points present the FCM membership function.

present well for the characteristic of separation. This reveals that the inverse of sample covariance of form Eq. (24) is a good estimate for β. That the sample covariance is small means two clusters are compact and well separated around two cluster centers 0 and 2 in which the estimate of β is large. The membership curve of AFCM with large β in Fig. 5 actually presents this goodness of fit for the data set. That is, if the data set with separated or distinguishable clusters (i.e. small sample covariance and large β) then the AFCM also presents well-separated memberships. However, if the data set with fuzzy or undistinguished clusters (i.e. large sample covariance and small β) then the AFCM actually presents fuzzy boundary memberships which is close to FCM memberships. Thus, the proposed AFCM with the estimate of β by the inverse of sample coveriance is a good clustering algorithm for different types of data set because it can be adjusted by the estimate of β with the inverse of sample covariance.

Suppose we set up S_{ij} with

$$S_{ij} = \exp(-\beta||x_j - z_i||^2), \tag{33}$$

the AFCM prototype updating Eq. (22) can be rewritten as

$$z_i = \frac{\sum_{j=1}^n (\mu_{ij})^m S_{ij} x_j}{\sum_{j=1}^n (\mu_{ij})^m S_{ij}}. \tag{34}$$

Comparing Eq. (34) with Eq. (29), we find that the AFCM prototype updating equation has an extra weighted item S_{ij} which decreases monotonically in the distance measure $||x_j - z_i||$ with $S_{ij} = 1$ as $||x_j - z_i|| = 0$ and $S_{ij} = 0$ as $||x_j - z_i||$ tends to infinity. The FCM prototype updating equation is the weighted sum of x_j with the weights μ_{ij}^m.

Under the constrain that the memberships of a data point across clusters sum to one, FCM assigns almost equal memberships to those points that are far away from each cluster even if they have a different degree of belonging to each

cluster. Therefore, the prototype updating equation also gives the same weight to those data. That is why FCM is troublesome in a noisy environment. Although the AFCM membership functions use the same probabilistic constraint, the prototype updating Eq. (34) gives those noise and outliers different weights $(\mu_{ij}^m)S_{ij}$.

Since S_{ij} decreases monotonically in $||x_j - z_i||$, Eq. (34) will reasonably assign a suitable weight to each data point. Thus, using prototype updating Eq. (34) will be more robust than using Eq. (29) and then AFCM shall be more robust than FCM. In Fig. 3, the same data set used in HCM and AHCM is also tested using FCM and AFCM and the results are shown in Figs. 3(b) and (d), respectively. Although the effect of the outlier for FCM is smaller than that for HCM, FCM does not produce a good clustering boundary. However, AFCM can work well in this noisy environment. AFCM also has the ability to detect unequal size clusters, as shown in Fig. 4(d). FCM cannot derive this target as shown in Fig. 4(b). Next, we discuss the connection between AFCM and FCM.

First we claim that when β tends to zero the AFCM and FCM membership will coincide. Suppose $||x_j - z_i||^2 = a_i$, we divide Eq. (31) by Eq. (28) and obtain the equation by numerator minus denominator as

$$\left(\frac{1}{1 - \exp(-\beta a_i)}\right)^{1/(m-1)} \left(\sum_{k=1}^c \left(\frac{1}{a_k}\right)^{1/(m-1)}\right)$$

$$- \left(\frac{1}{a_i}\right)^{1/(m-1)} \left(\sum_{k=1}^c \left(\frac{1}{1 - \exp(-\beta a_k)}\right)^{1/(m-1)}\right)$$

$$= \sum_{k=1}^c \left\{ \left(\frac{1}{a_k(1 - \exp(-\beta a_i))}\right)^{1/(m-1)} \right.$$

$$\left. - \left(\frac{1}{a_i(1 - \exp(-\beta a_k))}\right)^{1/(m-1)} \right\}. \tag{35}$$

Since

$$a_k(1 - \exp(-\beta a_i)) - a_i(1 - \exp(-\beta a_k))$$

$$= a_k - a_i - (a_k \exp(-\beta a_i) - a_i \exp(-\beta a_k))$$

will tend to zero as β tends to zero. Thus, Eq. (35) will tend to zero as β tends to zero and hence the AFCM membership (31) will tend to the FCM membership (28). This leads to the AFCM prototype updating Eq. (34) will also tend to the FCM prototype updating Eq. (29) as β tends to zero.

Fig. 6 shows the connection between HCM, AHCM, FCM and AFCM. The results from HCM and AHCM (or FCM and AFCM) will be coincident under a small β value that can also be experimented with numerical comparisons. HCM and AHCM are cases of FCM and AFCM, respectively, when the fuzzifier m tends to 1. AHCM and AFCM are extensions of HCM and FCM, respectively, under metric transformation. Finally, we compare these four algorithms simultaneously in the following three numerical examples
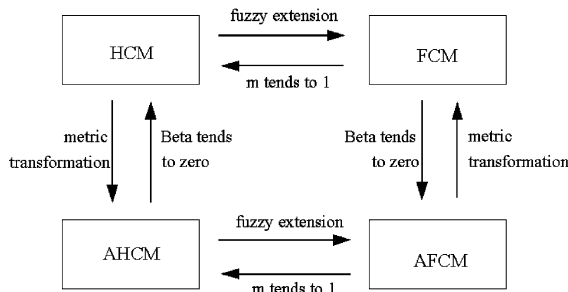
Fig. 6. The connection between HCM, AHCM, FCM and AFCM.

including the *Iris* data. All algorithms were started with the same initial values and stopped under the same stopping conditions in the same examples.

Example 3. We add more points to the right cluster. The HCM, FCM, AHCM and AFCM results are shown in Figs. 7a–d, respectively. The HCM and FCM results have many misclassified data. Under the proposed metric, AHCM has just only one misclassified data point and AFCM classifies these two clusters correctly.

Example 4. We provided uniform data on two planes in which one is zero-section (i.e. $z = 0$) plane and another

one-section (i.e. $z = 1$) plane as shown in Fig. 8. In Fig. 8(a), HCM found centers (the solid circles) on the $z = 0.5$ plane. Clearly, no data points surrounded these two centers. The AHCM and FCM results are quite similar, where the centers are closer to the data planes than HCM, as shown in Figs. 8(b) and (c), respectively. There are still many misclassified data for AHCM and FCM. However, AFCM classified these two clusters correctly and one center is located on the zero-section plane and another center is located on one-section plane. Results are shown in Fig. 8(d). In this example, we see that AHCM seems to correct HCM, but it is not good enough, which is similar to FCM. However, AFCM gives a perfect clustering result in this example.

Example 5. The three-dimensional plot for the four-dimensional *Iris* data set is shown in Fig. 9(a). The plus, solid circle and circle symbols represent the *Iris Setosa*, *Iris Versicolor* and *Iris Virginica*, respectively. HCM and AHCM have coincident clustering results, as shown in Fig. 9(b). The results for FCM are shown in Fig. 9(c). The clustering result shown in Fig. 9(d) for AFCM seems to have fewer misclassified data than other algorithms and can be confirmed in Table 1. The total error counts for HCM, FCM and AHCM are all 16. However, the total error count for AFCM is only 13. Thus, AFCM has the best clustering performance for the *Iris* data in these four clustering algorithms.
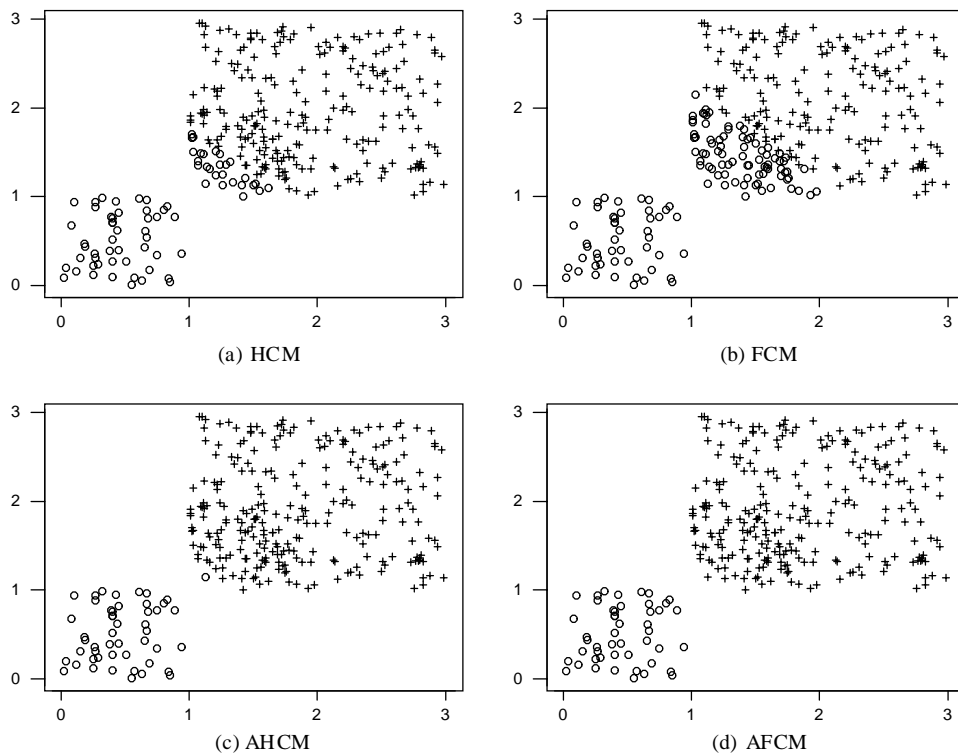

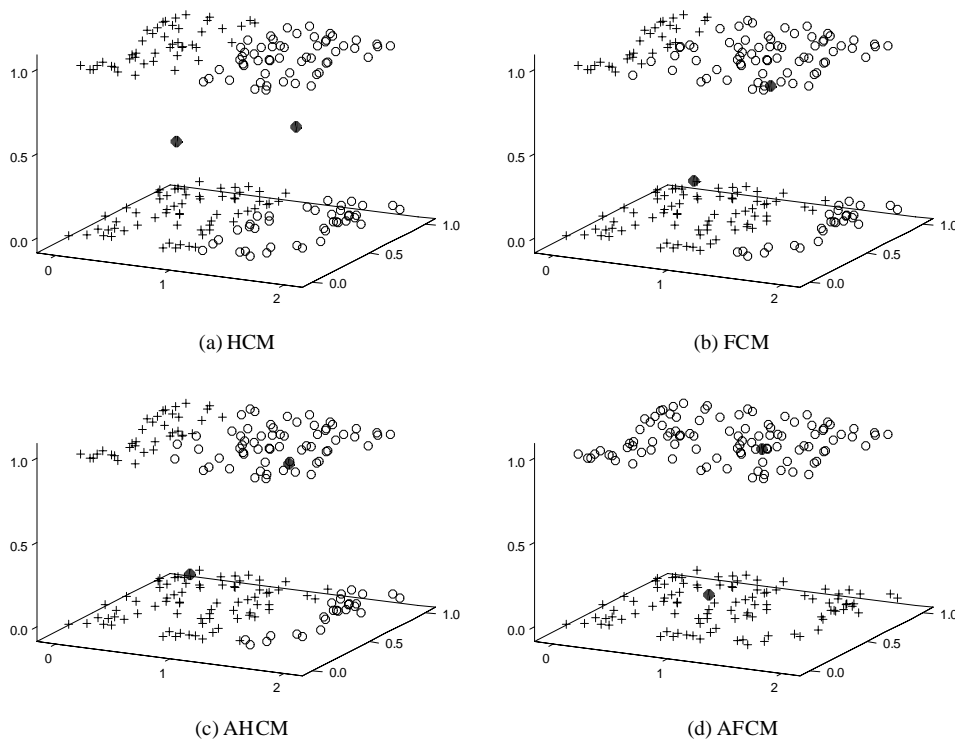Fig. 7. Clustering results for HCM, FCM, AHCM and AFCM in Example 3.

(a) HCM

(b) FCM

(c) AHCM

(d) AFCM

Fig. 8. 3D clustering results for HCM, FCM, AHCM and AFCM.

Example 6 (see Yang et al. [14]). The FCM and AFCM algorithms has recently been applied in a real case for magnetic resonance image (MRI) segmentation to differentiate normal and abnormal tissues in Ophthalmology by Yang et al. [14]. The patient is a 2-yr old girl. She was diagnosed with Retinoblastoma, an inborn malignant neoplasm of the retina with frequent metastasis beyond the lacrimal cribrosa. The MR image showed an intra muscle cone tumor mass with high T1-weight signal images and low signals on T2-weight images were noted in the left eyeball. The tumor measured 20 mm in diameter and nearly occupied the entire vitreous cavity. There was a shady signal abnormality along the optic nerve to the level of the optic chiasma. A negative study of the right eye and the intracranial cavity included the pineal region.

Accurate classification of the tissue edge and tumor volume is important for early treatment with radiotherapy and surgery. The FCM and AFCM clustering algorithms are used to determine the actual symptoms of Ratinoblastoma from MRI and determine if further treatment is necessary. Since Glioblastoma is hereditary, children whose families have a history of this problem can undergo periodic diagnostic tests with MRI using the FCM and AFCM algorithms to detect the smallest sign symptoms for early detection and treatment. From these experiments in Ref. [14], both FCM and AFCM segmentation techniques usually provide useful information and good results. However, the AFCM has better

detection of abnormal tissues than FCM based on a window selection, especially for blurred or distorted MR image.

For the purpose of early treatment with radiotherapy and surgery in the real cases, the physicians are highly recommended to use window segmentation to isolate the problem area and enlarge the image and then separate the tissues using three tissue groups (connective, nervous and tumor) and apply the AFCM algorithm to the image for every Retinoblastoma patient.

5. Conclusions

In this paper we proposed a new metric to replace the Euclidean norm in c-means clustering algorithms and then created the so-called AHCM and AFCM clustering algorithms. Many clustering methods under Euclidean norm construction may have problems with different sizes and cluster shape and inaccuracy in a noisy environment. However, the proposed AHCM and AFCM are robust to noise and outliers and also tolerate unequal sized clusters.

The HCM prototype updating equation gives equal weight to those x_j which belong to the same cluster even if they should have different weights. However, the AHCM prototype updating equation gives reasonable weights to each x_j even though the algorithm classifies them to the same cluster in each iteration. Thus, AHCM is more robust than HCM. As
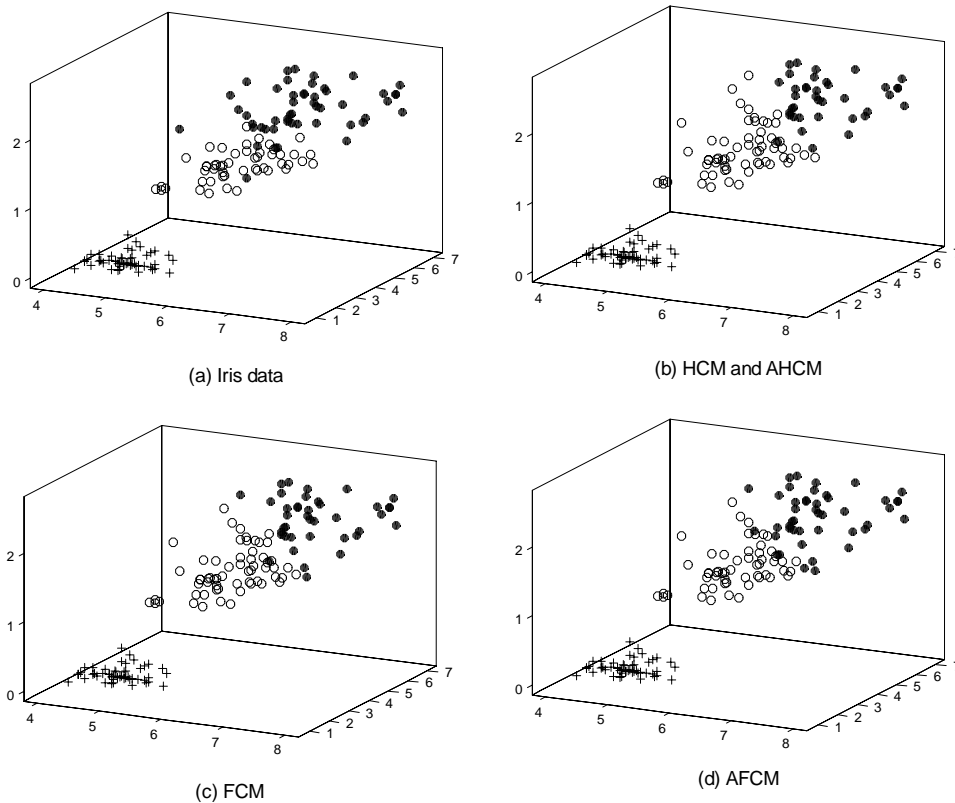
Fig. 9. Clustering results for the *Iris* data.

Table 1
Error count for *Iris*

	Setosa	*Versicolor*	*Virginica*	Total
HCM	0	2	14	16
FCM	0	3	13	16
AHCM	0	2	14	16
AFCM	0	4	9	13

we know, FCM is a fuzzy extension of HCM. Although the FCM prototype updating equation can give different weights to each x_j, it gives the same weights to noise and outliers even they should have different degrees of belonging. Under the same probabilistic constraint, however, the AFCM prototype updating equation gives different and reasonable weights to all data points including noise and outliers via the extra term S_{ij}.

The connection between HCM (or FCM) and AHCM (or AFCM) was also studied. The numerical examples show that both AHCM and AFCM actually work better than HCM and FCM. In addition, AFCM has the best performance among these algorithms in the *Iris* data set. We recommended those concerned with applications in cluster analysis to try to use the AFCM clustering algorithm.

Finally, Yang et al. [14] have applied FCM and AFCM algorithms to magnetic resonance image (MRI) segmentation in Ophthalmology. They present a case study of Retinoblastoma patients who were diagnosed using MRI in the ophthalmology field. The results show that the AFCM has better detection of abnormal tissues from normal tissues than FCM. The AFCM algorithm is actually a better choice in this real case study. Furthermore, we may also consider AFCM in a possibilistic type of memberships which as proposed by Krishnapuram and Keller [15].

Acknowledgements

References

[1] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[2] A.K. Jain, R.C. Bubes, Algorithm for Clustering Data, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[3] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York, 1990.

[4] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.

[5] M.S. Yang, A survey of fuzzy clustering, Math. Comput. Modelling 18 (1993) 1–16.

[6] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.

[7] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters, J. Cybernet. 3 (1974) 32–57.

[8] I. Gath, A.B. Geva, Unsupervised optimal fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 773–781.

[9] R.N. Dave, Fuzzy-shell clustering and applications to circle detection in digital images, Int. J. General Syst. 16 (1990) 343–355.

[10] M.S. Yang, C.H. Ko, On cluster-wise fuzzy regression analysis, IEEE Trans. Systems, Man, Cybern. 27 (1997) 1–13.

[11] W. Rudin, Principles of Mathematical Analysis, McGraw-Hill Book Company, New York, 1976.

[12] P.J. Huber, Robust Statistics, Wiley, New York, 1981.

[13] P.J. Huber, Robust estimation of a location parameter, Ann. Math. Statist. 35 (1964) 73–101.

[14] M.S. Yang, Y.J. Hu, K.C.R. Lin, C.C.L. Lin, Segmentation techniques for tissue differentiation in MRI of Ophthalmology using fuzzy clustering algorithms, Magnetic Resonance Imaging (2001) (in the first revision).

[15] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Systems 1 (1993) 98–110.

About the Author—KUO-LUNG WU received the B.S. degree in mathematics in 1997 and the M.S. degree in Applied Mathematics in 2000 from the Chung Yuan Christian University, Chung-Li, Taiwan, ROC. He is a Ph.D. student and a part-time Instructor in the Department of Mathematics, Chung Yuan Christian University. His research interests include cluster analysis, pattern recognition and neural networks.

About the Author—MIIN-SHEN YANG received the B.S. degree in mathematics from the Chung Yuan Christian University, Chung-Li, Taiwan, ROC, in 1977, the M.S. degree in Applied Mathematics from the National Chiao-Tung University, Hsinchu, Taiwan, in 1980, and Ph.D. degree in statistics from the University of South Carolina, Columbia, in 1989. He is a Professor in the Department of Mathematics, Chung Yuan Christian University. His current research interests include applications of statistics, fuzzy cluster analysis, pattern recognition and neural network systems.