FCM & FPCM Algorithm Based on Unsupervised Mahalanobis Distances with Better Initial Values and Separable Criterion

JENG-MING YIH, YUAN-HORNG LIN, HSIANG-CHUAN LIU and Chi-Fu Yih

***Abstract*—**The fuzzy partition clustering algorithms are most based on Euclidean distance function, which can only be used to detect spherical structural clusters. Gustafson-Kessel (GK) clustering algorithm and Gath-Geva (GG) clustering algorithm, were developed to detect non-spherical structural clusters, but both of them based on semi-supervised Mahalanobis distance needed additional prior information. An improved Fuzzy C-Mean algorithm based on unsupervised Mahalanobis distance, FCM-M, was proposed by our previous work, but it didn't consider the relationships between cluster centers in the objective function. In this paper, we proposed an improved Fuzzy C-Mean algorithm, FCM-MS, which is not only based on unsupervised Mahalanobis distance, but also considering the relationships between cluster centers, and the relationships between the center of all points and the cluster centers in the objective function, the singular and the initial values problems were also solved. Three real data sets was applied to prove that the performance of the FCM-MS algorithm gave more accurate clustering results than the FCM and FCM-M methods, and the ratio method which is proposed by us is the better of the two methods for selecting the initial values.

***Keywords*—**FCM-MS, FCM-M, GK algorithms, GG algorithms, Mahalanobis distance.

I. Introduction

Clustering plays an important role in data analysis and interpretation. It groups the data into classes or clusters so that the data objects within a cluster have high similarity in comparison to one another, but are very dissimilar to those data objects in other clusters.

Fuzzy partition clustering is a branch in cluster analysis, it is widely used in pattern recognition field. The well known ones, such as, C. Bezdek's "Fuzzy C-Mean (FCM)" [1], are all based on Euclidean distance function, which can only be used to detect the data classes with same super spherical shapes.

J. M. Yih is with the Graduate Institute of Educational Measurement and Statistics, and Department of Mathematics Education, National Taichung University 140 Min-Sheng Rd., Taichung City 403, Taiwan (corresponding author to provide phone: 86-913-465279; fax: 86-4-22183500; e-mail: yih@mail.ntcu.edu.tw).

Y. H. Lin., was with Department of Mathematics Education National Taichung University. He is now leader of Department of Mathematics Education National Taichung University. (e-mail: lyh@mail.ntcu.edu.tw).

H. C. Liu is with Department of Bioinformatics, Asia University,Taiwan).

C. F. Yih is with Department of Computer Science and Information Engineering National Chi Nan University , Nantou, Taiwan. (e-mail: aaroncat1758@gmail.ntcu.edu.tw)..

Extending Euclidean distance to Mahalanobis distance, the well known fuzzy partition clustering algorithms, Gustafson-Kessel (GK) clustering algorithm [3] and Gath-Geva (GG) clustering algorithm [2] were developed to detect non-spherical structural clusters, but these two algorithms fail to consider the relationships between cluster centers in the objective function, GK algorithm must have prior information of shape volume in each data class, otherwise, it can only be considered to detect the data classes with same volume. GG algorithm must have prior probabilities of the clusters.

On the other hand, When any dimension of a class is greater than the sample size of which class, the estimated covariance matrix of which class may not be full rank, it induces the singular problem of the inverse covariance matrix, it is an important issue without generally consider in above algorithms. Focusing the above two faults, we added a regulating factor of covariance matrix to each class in objective function, and deleted the constraint of the determinants of covariance matrices in GK Algorithm, An improved algorithm, based on Mahalanobis distance, "Fuzzy C-Mean based on Mahalanobis distance (FCM-M)" is proposed by our previous work [4].Yin et al. [5] described an extended objective function consisting of a fuzzy within-cluster scatter matrix and a new between-cluster centers scattering matrix. The corresponding fuzzy clustering algorithm assures the compactness between data points and cluster centers and also strengthens the separation between cluster centers based on the separation criterion. Then clustering algorithm solved the relationships between cluster centers question, but they did not consider the distance between the center of all points and the center of each cluster. This problem was also solved and presented in this paper. Moreover, In this paper, an improved fuzzy clustering algorithm, denoted FCM-MS, was developed based on FCM-M to obtain better quality clustering results with new separable criterion and better initial value. The improved equations for the membership and the cluster center were derived from the alternating optimization algorithm. The distance between the center of all points and the center of each cluster was considered by the authors of this paper, the singular problem was also solved. A real data set was applied to prove that the performance of the FCM-MS algorithm gave more accurate clustering results than the

FCM-M and FCM methods, and the ratio method which is proposed by us is the better of the two methods for selecting the initial values.

II. Literature Review

Fuzzy C-Mean Algorithm (FCM) is the most popular objective function based fuzzy clustering algorithm, it is first developed by Dunn [6] and improved by Bezdek [1].The algorithm of fuzzy C-Means Algorithm are the foundations of this study. The algorithm will be discussed as follows.

A. Fuzzy C-Mean Algorithm

The objective function used in FCM is given by Equation (1).

$$J_{FCM}^{m}\left(U,A,X\right)=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}d_{ij}^{2} = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left\|\underline{x}_j-\underline{a}_i\right\|^2 \qquad (1)$$

$\mu_{ij}\in[0,1]$ is the membership degree of data object $\underline{x}_j$ in cluster C_i and it satisfies the following constraint given by Equation (2).

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,...,n \qquad (2)$$

C is the number of clusters, m is the fuzzifier, m>1, which controls the fuzziness of the method. They are both parameters and need to be specified before running the algorithm. $d_{ij}^2=\left\|\underline{x}_j-\underline{a}_i\right\|^2$ is the square Euclidean distance between data object $\underline{x}_j$ to center $\underline{a}_i$ .Minimizing objective function (1) with constraint (2) is a non-trivial constraint nonlinear optimization problem with continuous parameters $\underline{a}_i$ and discrete parameters μ_{ij} **.** So there is no obvious analytical solution. Therefore an alternating optimization scheme, alternatively optimizing one set of parameters while the other set of parameters are considered as fixed, is used here. Then the updating function for $\underline{a}_i$ and μ_{ij} is obtained as (3) ~ (5).

Step 1: Determining the number of cluster; c and m-value (let m=2), given converging error, $\varepsilon>0$ (such as $\varepsilon=0.001$), randomly choose the initial membership matrix, such that the memberships are not all equal

$$U^{(0)}=\begin{bmatrix}\mu_{11}^{(0)}\ \mu_{12}^{(0)}\cdots\mu_{1n}^{(0)}\\ \mu_{21}^{(0)}\ \mu_{22}^{(0)}\cdots\mu_{2n}^{(0)}\\ \cdots\ \ \cdots\ \ \cdots\\ \mu_{c1}^{(0)}\ \mu_{c2}^{(0)}\cdots\mu_{cn}^{(0)}\end{bmatrix}=\begin{bmatrix}\mu_{1}^{(0)}(\underline{x}_1)\ \mu_{1}^{(0)}(\underline{x}_2)\ldots\mu_{1}^{(0)}(\underline{x}_n)\\ \mu_{2}^{(0)}(\underline{x}_1)\ \mu_{2}^{(0)}(\underline{x}_2)\ldots\mu_{2}^{(0)}(\underline{x}_n)\\ \cdots\ \ \cdots\ \ \cdots\\ \mu_{c}^{(0)}(\underline{x}_1)\ \mu_{c}^{(0)}(\underline{x}_2)\ldots\mu_{c}^{(0)}(\underline{x}_n)\end{bmatrix} \qquad (3)$$

Step 2: Find

$$\underline{a}_i^{(k)}=\frac{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m\underline{x}_j}{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m}\quad i=1,2,\ldots,c \qquad (4)$$

$$\mu_{ij}^{(k)}=\left[\sum_{l=1}^{c}\left[\frac{\left(\underline{x}_j-\underline{a}_i^{(k)}\right)'\left(\underline{x}_j-\underline{a}_i^{(k)}\right)}{\left(\underline{x}_j-\underline{a}_l^{(k)}\right)'\left(\underline{x}_j-\underline{a}_l^{(k)}\right)}\right]^{\frac{1}{m-1}}\right]^{-1} \qquad (5)$$

Step 3: Increment k; until $\max_{1\le i\le c}\left\|\underline{a}_i^{(k)}-\underline{a}_i^{(k-1)}\right\|<\varepsilon$

B. FCM-M Algorithm

Mahalanobis, an Indian statistician , introduced this distance in the 1930s. The Mahalanobis distance is a distance using the inverse of the covariance matrix as the metric. It is a distance in the geometrical sense because the covariance matrices as well as its inverse are positive definite matrices. [9]

We call clusters using the Mahalanobis distance as covariance clusters. The metric defined by the covariance matrix provides a normalization of the data relative to their spread. Using the Mahalanobis distance is done as follows: **1.**The covariance matrix of the measured quantities, V, is determined over a calibrating set. **2.**One compute the inverse of the covariance matrix, V^{-1}. **3.**The distance of a new object to the calibrating set is estimated using equation $d_M^2=(x-\bar{x})^T V^{-1}(x-\bar{x})$. ; if the distance is smaller than a given threshold value, the new object is considered as belonging to the same set.

One interesting property of the Mahalanobis distance is that it is normalized. Thus, it is not necessary to normalize the data , provided rounding errors is inverting the covariance matrix are kept under control. If the data are roughly distributed according to a normal distribution, the threshold for accepting whether an object belong to the calibrating set can be determined from the χ^2 distribution. The Mahalanobis distance can be applied in all problems in which measurements must be classified**.**

A good example is the detection of coins in a vending machine. When a coins is inserted into the machine, a series of sensors gives several measurements, between a handful and a dozen. The detector can be calibrated using a set of good coins forming a calibration set. The coin detector can differentiate good coins from the fake coins using the Mahalanobis distance computed on the covariance matrix of the calibration set, reference the following Figure 1.

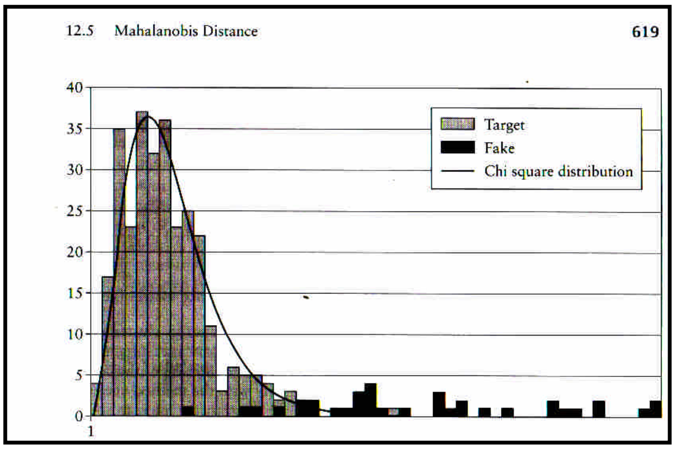

FIG.1. Use the Mahalanobis distance to detect the fake coins.
(Convert from Besset D. H. p619, FIG. 12.2)

Another field of application is the determination of cancer cells from a biopsy.Parameters of cells can be measured automatically and expressed in numbers. The covariance matrix can be determined using either measurements of healthy cells or measurements of malignant cells. Identification of cancerous cells can be automated using the Mahalanobis distance.

The final goal of the object implementing the Mahalanobis distance is to compute the square Mahalanobis distance as defined in equation $d_M^2 = (x-\bar{x})^T V^{-1}(x-\bar{x})$.

Implementation of the Mahalanobis distance is dictated by its future reuse in cluster analysis. There, we need to be able to accumulate measurements while using the result of a preceding accumulation .Thus, computation of the center and the inverse covariance matrix must be done , see the figure 2.explicitly with the method computer Parameters. There are two ways of creating a new instance. One is to specify the dimension of the vectors that will be accumulated into the object. The second supplies a vector as the tentative center.

The normalizing properties of the Mahalanobis distance make it ideal for this task. When Euclidean distance is used , the metric remains the same in all directions. Thus, the extent of each cluster has more or less circular shapes. With the Mahalanobis distance, the covariance matrix is unique for each cluster. Thus, covariance clusters can have different shapes since the metric adapts itself to the shape of each cluster. As the algorithm progresses, The metric changes dynamically.[10]

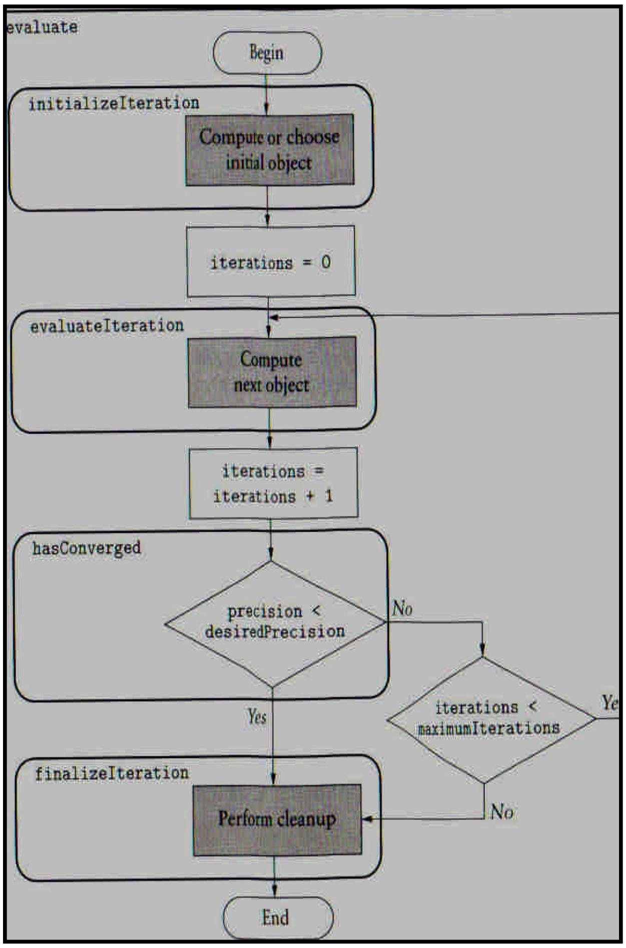

FIG.2. Method for successive approximation.
(Convert from Besset D. H. p118, FIG. 4.4)

For improving the above two problems, our previous work [4] proposed the improved algorithm FCM-M which added $-\ln\left|+\Sigma_i^{-1}\right|$ a regulating factor of covariance matrix to each class in objective function, and deleted $|M_i| = \rho_i$ the constraint of the determinant of covariance matrices in GK Algorithm as the objective function (6).

Using the Lagrange multiplier method, We can minimize the objective function (6). Constraint (7) with respect to the parameters $\underline{a}_i$, μ_{ij}, and Σ_i, we can obtain the solutions as (10), (11), and(13).

We want to avoid the singular problem and to select the better initial membership matrix, the updating functions for $\underline{a}_i$, μ_{ij} and Σ_i are obtained as (8) ~ (3-8). Both of FCM and FCM-M can not exploit all of the memberships with the same value. FCM is a special case of FCM-M, when covariance matrices equal to identity matrices by our previous work [8].

$$J^m_{FCM-M}(U,A,\Sigma,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|\right] \tag{6}$$

Constraints: membership,

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,\dots,n \tag{7}$$

$\Sigma=\{\Sigma_1,\Sigma_2,\dots,\Sigma_c\}$ is the set of covariance of cluster.

Step 1: Determining the number of cluster; c and m-value (let m=2), given converging error, $\varepsilon>0$ (such as $\varepsilon=0.001$).

Method 1: choose the result membership matrix of FCM algorithm as the initial one.

Method 2: let $\underline{a}_i^{(0)}, i=1,2,\dots,c$ be the result centers of k-mean algorithm, And $d_{ij}=\left\|\underline{x}_j-\underline{a}_i^{(0)}\right\|$ be distances between data object $\underline{x}_j$ to center $\underline{a}_i^{(0)}$.

$$\begin{aligned}d_i^M&=\max_{1\le j\le n}d_{ij}=\max_{1\le j\le n}\left\|\underline{x}_j-\underline{a}_i^{(0)}\right\|,\\ d_i^m&=\min_{1\le j\le n}d_{ij}=\min_{1\le j\le n}\left\|\underline{x}_j-\underline{a}_i^{(0)}\right\|,\end{aligned} \tag{8}$$

$$u_{ij}^{(0)}=\frac{d_i^M-d_{ij}}{d_i^M-d_i^m},\ i=1,2,\dots,c,\ j=1,2,\dots,n$$

$$U^{(0)}=\begin{bmatrix}u_{11}^{(0)}\,u_{12}^{(0)}\dots u_{1n}^{(0)}\\ u_{21}^{(0)}\,u_{22}^{(0)}\dots u_{2n}^{(0)}\\ \dots\ \dots\ \dots\\ u_{c1}^{(0)}\,u_{c2}^{(0)}\dots u_{cn}^{(0)}\end{bmatrix}=\begin{bmatrix}u_1^{(0)}(\underline{x}_1)\,u_1^{(0)}(\underline{x}_2)\dots u_1^{(0)}(\underline{x}_n)\\ u_2^{(0)}(\underline{x}_1)\,u_2^{(0)}(\underline{x}_2)\dots u_2^{(0)}(\underline{x}_n)\\ \dots\ \dots\ \dots\\ u_c^{(0)}(\underline{x}_1)\,u_c^{(0)}(\underline{x}_2)\dots u_c^{(0)}(\underline{x}_n)\end{bmatrix} \tag{9}$$

Step 2: Find

$$\Sigma_i^{(k)}=\frac{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m\left(\underline{x}_j-\underline{a}_i^{(k)}\right)\left(\underline{x}_j-\underline{a}_i^{(k)}\right)'}{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m} \tag{10}$$

$$\Sigma_i^{(k)}=\sum_{s=1}^{p}\lambda_{si}^{(k)}\Gamma_{si}^{(k)}\left(\Gamma_{si}^{(k)}\right)',$$

$$\left[\lambda_{si}^{(-1)}\right]^{(k)}=\begin{cases}\left[\lambda_{si}^{(k)}\right]^{-1} & if\ \lambda_{si}^{(k)}>0\\ 0 & if\ \lambda_{si}^{(k)}=0\end{cases}$$

$$\left[\Sigma_i^{-1}\right]^{(k)}=\sum_{s=1}^{p}\left[\lambda_{si}^{(-1)}\right]^{(k)}\Gamma_{si}^{(k)}\left(\Gamma_{si}^{(k)}\right)' \tag{11}$$

$$\left|\Sigma_i^{-1}\right|^{(k)}=\prod_{1\le s\le p,\lambda_{si}^{(k)}>0}\left[\lambda_{si}^{(-1)}\right]^{(k)} \tag{12}$$

$$\mu_{ij}^{(k)}=\left[\sum_{s=1}^{c}\left[\frac{\mathrm{w_i}'\left[\Sigma_i^{-1}\right]^{(k)}\mathrm{w_i}-\ln\left\|\left[\Sigma_i^{-1}\right]^{(k)}\right\|}{\mathrm{w_s}'\left[\Sigma_s^{-1}\right]^{(k)}\mathrm{w_s}-\ln\left\|\left[\Sigma_s^{-1}\right]^{(k)}\right\|}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{13}$$

$$where\ \mathrm{w_i}=\left(\underline{x}_j-\underline{a}_i^{(k)}\right)$$

Step 3: Increment k; until $\max_{1\le i\le c}\left\|\underline{a}_i^{(k)}-\underline{a}_i^{(k-1)}\right\|<\varepsilon$.

C. FCM-MS Algorithm[13]

The clustering optimization was based on objective functions. The choice of an appropriate objective function is the point to the success of the cluster analysis.[14] In FCM-M algorithm, it didn't consider the relationships between cluster centers in the objective function, now, we proposed an improved Fuzzy C-Mean algorithm, FCM-MS, which is not only based on unsupervised Mahalanobis distance, but also considering the relationships between cluster centers, and the relationships between the center of all points and the cluster centers in the objective function, the singular and the initial values problems were also solved. Let {x1, x2, x3,…, xn} be a set of n data points represented by p-dimensional feature vectors $\mathrm{x_j}=(\mathrm{x_{1j}},\mathrm{x_{2j}},\dots,\mathrm{x_{pj}})'\in\mathrm{R^p}$. The p×n data matrix Z has the cluster center matrix A=[a1, …, ac] , 1<c<n and the membership matrix $U=[\mu_{ij}]_{cxn}$, where μ_{ij} is the membership value of xj belonging to a_i. $V=[v_{ik}]_{cxc}$ express the weighting matrix, and v_{ik} is the weighting value between vi and vk. The fuzzy exponent *m* is greater than 1 [7]. Thus, the proposed objective function is

$$\begin{aligned}&J^m_{FCM-MS}(U,V,A,\Sigma,X)\\ &=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|\right]\\ &-\frac{1}{c(c-1)}\sum_{i=1}^{c}\sum_{l=1}^{c}v_{il}^m\left\|\underline{a}_i-\underline{a}_l\right\|^2\end{aligned} \tag{14}$$

Such that

$$\mu_{ij}\in[0,1],\sum_{i=1}^{c}\mu_{ij}=1,\forall j,\ 0<\sum_{j=1}^{n}\mu_{ij}<n,\forall i, \tag{15}$$

where v_{ij} is defined as

$$v_{il}=\frac{[\mathrm{y_i}+\mathrm{y_l}]-\min_{1\le r,s\le c}[\mathrm{y_r}+y_s]}{\max_{1\le r,s\le c}[\mathrm{y_r}+y_s]-\min_{1\le r,s\le c}[\mathrm{y_r}+y_s]}\quad where\ \mathrm{y_i}=\left\|\underline{a}-\underline{a}_i\right\|^2 \tag{16}$$

The goal of the clustering algorithm is to identify the cluster centers and the membership values by solving an optimization problem. Alternating optimization is a popular mathematical tool for the regular objective function-based fuzzy clustering algorithms.

The optimal update equations can be obtained using the Lagrange method by setting the partial derivative of the Lagrange with respect to v_i and with respect to μ_{ij} equal to zero. Setting $\partial \bar{J} / \partial \mu_{ij}$ equal to zero gives the update equation for μ_{ij}.

The new fuzzy clustering algorithm can be summarized in the following steps:

Step 1: Determining the number of cluster; c and m-value (let m=2), given converging error, $\varepsilon > 0$ (such as $\varepsilon = 0.001$).

Method 1: choose the result membership matrix of FCM algorithm as the initial one.

Method 2: let $\underline{a}_i^{(0)}, i=1,2,\ldots,c$ be the result centers of k-mean algorithm, and $d_{ij} = \left\| \underline{x}_j - \underline{a}_i^{(0)} \right\|$ be distances between data object $\underline{x}_j$ to center $\underline{a}_i^{(0)}$.

$$d_i^M = \max_{1 \le j \le n} d_{ij} = \max_{1 \le j \le n} \left\| \underline{x}_j - \underline{a}_i^{(0)} \right\|, \; d_i^m = \min_{1 \le j \le n} d_{ij} = \min_{1 \le j \le n} \left\| \underline{x}_j - \underline{a}_i^{(0)} \right\|,$$

$$u_{ij}^{(0)} = \frac{d_i^M - d_{ij}}{d_i^M - d_i^m}, \; i = 1, 2, \ldots, c, \; j = 1, 2, \ldots, n \tag{17}$$

$$\underline{a}_i^{(0)} = \left(\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right] \underline{x}_j\right)\left(\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]\right)^{-1}, \quad i = 1, 2, \ldots, c \tag{18}$$

$$\Sigma_i^{(0)} = \left(\sum_{j=1}^{n} \left[\mu_{ij}^{(0)} \right]^m \left(\underline{x}_j - \underline{a}_i^{(0)} \right)\left(\underline{x}_j - \underline{a}_i^{(0)} \right)'\right)\left(\sum_{j=1}^{n} \left[\mu_{ij}^{(0)} \right]^m\right)^{-1} \tag{19}$$

Step 2: Find

$$v_{il}^{(k)} = \frac{w_{il}^{k-1} - \min_{1 \le r,s \le c} w_{rs}^{k-1}}{\max_{1 \le r,s \le c} w_{rs}^{k-1} - \min_{1 \le r,s \le c} w_{rs}^{k-1}}, \tag{20}$$

$$\text{where} \quad w_{rs}^{k-1} = \left\| \underline{a} - \underline{a}_r^{(k-1)} \right\|^2 + \left\| \underline{a} - \underline{a}_s^{(k-1)} \right\|^2$$

$$a_i^{(k)} = \left[\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m \left[\Sigma_i^{(k-1)} \right]^{-1} - \frac{1}{c(c-1)} \sum_{l=1}^{c} \left[v_{il}^{(k)} \right]^m I \right]^{-1}$$
$$\left[\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m \left[\Sigma_i^{(k-1)} \right]^{-1} x_j - \frac{1}{c(c-1)} \sum_{l=1}^{c} \left[v_{il}^{(k)} \right]^m \underline{a}_l^{(k-1)} \right] \tag{21}$$
$$, i = 1, 2, \ldots, c$$

$$\Sigma_i^{(k)} = \left(\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m \left(\underline{x}_j - \underline{a}_i^{(k)} \right)\left(\underline{x}_j - \underline{a}_i^{(k)} \right)'\right)$$
$$\left(\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m\right)^{-1}, \tag{22}$$

$$\Sigma_i^{(k)} = \sum_{s=1}^{p} \lambda_{si}^{(k)} \Gamma_{si}^{(k)} \left(\Gamma_{si}^{(k)} \right)' \tag{23}$$

$$\left[\lambda_{si}^{(-1)} \right]^{(k)} = \left[\lambda_{si}^{(k)} \right]^{-1} \quad if \;\; \lambda_{si}^{(k)} > 0,$$
$$\left[\lambda_{si}^{(-1)} \right]^{(k)} = 0 \qquad if \;\; \lambda_{si}^{(k)} = 0,$$
$$\left[\Sigma_i^{-1} \right]^{(k)} = \sum_{s=1}^{p} \left[\lambda_{si}^{(-1)} \right]^{(k)} \Gamma_{si}^{(k)} \left(\Gamma_{si}^{(k)} \right)', \tag{24}$$
$$\left| \Sigma_i^{-1} \right|^{(k)} = \prod_{1 \le s \le p, \lambda_{si}^{(k)} > 0} \left[\lambda_{si}^{(-1)} \right]^{(k)}.$$

$$\mu_{ij}^{(k)} = \left[\sum_{s=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' \left[\Sigma_i^{-1} \right]^{(k)} \left(\underline{x}_j - \underline{a}_i^{(k)} \right) - \ln \left| \left[\Sigma_i^{-1} \right] \right|^{(k)}}{\left(\underline{x}_j - \underline{a}_s^{(k)} \right)' \left[\Sigma_s^{-1} \right]^{(k)} \left(\underline{x}_j - \underline{a}_s^{(k)} \right) - \ln \left| \left[\Sigma_s^{-1} \right] \right|^{(k)}} \right]^{\frac{1}{m-1}} \right]^{-1} \tag{25}$$

Step 3: Increment k; until $\max_{1 \le i \le c} \left\| \underline{a}_i^{(k)} - \underline{a}_i^{(k-1)} \right\| < \varepsilon$

D. 2.4 Fuzzy Possibilit y C-M ean Algorithm[12]

Combining FCM and PCM, the improved fuzzy partition clustering algorithms "Fuzzy Possibility C-Mean (FPCM)", was proposed

$$J_{FPCM}^m (U, T, A, X) = \sum_{i=1}^{c} \sum_{j=1}^{n} \left(\mu_{ij}^m + t_{ij}^\delta \right) \left\| \underline{x}_j - \underline{a}_i \right\|^2 \tag{26}$$

constraints：membership

$$\sum_{i=1}^{c} \mu_{ij} = 1, \; \forall j = 1, 2, \ldots, n, \tag{27}$$

$$\text{typicality} \; \sum_{j=1}^{n} t_{ij} = 1, \; \forall i = 1, 2, \ldots, c \tag{28}$$

Step 1: Determining the number of cluster; c and m-value (let m=2), $\delta = 3$ given converging error, $\varepsilon > 0$ (such as $\varepsilon = 0.001$) choose the result membership matrix of FCM algorithm as the initial one and the result typicality matrix of PCM algorithm as the initial one respectively;

$$U^{(0)} = \begin{bmatrix} \mu_{11}^{(0)} \; \mu_{12}^{(0)} \ldots \mu_{1n}^{(0)} \\ \mu_{21}^{(0)} \; \mu_{22}^{(0)} \ldots \mu_{2n}^{(0)} \\ \ldots \quad \ldots \quad \ldots \\ \mu_{c1}^{(0)} \; \mu_{c2}^{(0)} \ldots \mu_{cn}^{(0)} \end{bmatrix} = \begin{bmatrix} \mu_1^{(0)}(\underline{x}_1) \; \mu_1^{(0)}(\underline{x}_2) \ldots \mu_1^{(0)}(\underline{x}_n) \\ \mu_2^{(0)}(\underline{x}_1) \; \mu_2^{(0)}(\underline{x}_2) \ldots \mu_2^{(0)}(\underline{x}_n) \\ \ldots \quad \ldots \quad \ldots \\ \mu_c^{(0)}(\underline{x}_1) \; \mu_c^{(0)}(\underline{x}_2) \ldots \mu_c^{(0)}(\underline{x}_n) \end{bmatrix} \tag{29}$$

$$T^{(0)} = \begin{bmatrix} t_{11}^{(0)} \; t_{12}^{(0)} \ldots t_{1n}^{(0)} \\ t_{21}^{(0)} \; t_{22}^{(0)} \ldots t_{2n}^{(0)} \\ \ldots \quad \ldots \quad \ldots \\ t_{c1}^{(0)} \; t_{c2}^{(0)} \ldots t_{cn}^{(0)} \end{bmatrix} = \begin{bmatrix} t_1^{(0)}(\underline{x}_1) \; t_1^{(0)}(\underline{x}_2) \ldots t_1^{(0)}(\underline{x}_n) \\ t_2^{(0)}(\underline{x}_1) \; t_2^{(0)}(\underline{x}_2) \ldots t_2^{(0)}(\underline{x}_n) \\ \ldots \quad \ldots \quad \ldots \\ t_c^{(0)}(\underline{x}_1) \; t_c^{(0)}(\underline{x}_2) \ldots t_c^{(0)}(\underline{x}_n) \end{bmatrix} \tag{30}$$

Step 2: Find

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k)} \right]^m + \left[t_{ij}^{(k)} \right]^\delta \right) \underline{x}_j}{\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k)} \right]^m + \left[t_{ij}^{(k)} \right]^\delta \right)}, \tag{31}$$
$$i = 1, 2, \ldots, c$$

$$\mu_{ij}^{(k)} = \left[\sum_{l=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' \left(\underline{x}_j - \underline{a}_i^{(k)} \right)}{\left(\underline{x}_j - \underline{a}_l^{(k)} \right)' \left(\underline{x}_j - \underline{a}_l^{(k)} \right)} \right]^{\frac{1}{m-1}} \right]^{-1}, \tag{32}$$
$$i = 1, 2, \ldots, c, \; j = 1, 2, \ldots, n$$

$$t_{ij}^{(k)} = \left[\sum_{l=1}^{n} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' \left(\underline{x}_j - \underline{a}_i^{(k)} \right)}{\left(\underline{x}_j - \underline{a}_l^{(k)} \right)' \left(\underline{x}_j - \underline{a}_l^{(k)} \right)} \right]^{\frac{1}{\delta-1}} \right]^{-1} \tag{33}$$
$$i = 1, 2, \ldots, c. \; j = 1, 2, \ldots, n$$

Step 3: Increment k; until $\max_{1 \le i \le c} \left\| \underline{a}_i^{(k)} - \underline{a}_i^{(k-1)} \right\| < \varepsilon$

E. FPCM-M Algorithm

Now, for improving the above problems, we added a regulating factor of covariance matrix, $-\ln\left|+\Sigma_i^{-1}\right|$, to each class in objective function. The improved new algorithm, "Fuzzy Possibility C-Mean based on Mahalanobis distance (FPCM-M)", is obtained. Using the Lagrange multiplier method, to minimize the objective function (34) with constraint (35) respect to parameters $\underline{a}_i$, μ_{ij}, t_{ij}, Σ_i, we can obtain the solutions as (38), (39), (41), and(42), To avoid the singular problem and to select the better initial membership matrix, the updating functions for $\underline{a}_i$, μ_{ij}, and Σ_i are obtained as (36)~ (42).Note,

(i) All of the fuzzy partition clustering algorithms can not exploit all of the memberships with the same value

(ii) FPCM is a special case of FPCM-M ,when Adding a regulating factor of each clusters covarince in objective function, we proposed the fuzzy possibility c-mean based on mahalanobios distance (FPCM-M) as following.

$$J_{FPCM-M}^{m}(U,T,\Sigma,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|\right] \quad (34)$$

constraints：menbership

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,\dots,n,$$

typicality
$$\sum_{j=1}^{n}t_{ij}=1,\ \forall i=1,2,\dots,c \quad (35)$$

Step 1: Determining the number of cluster; c and m-value (let m=2), $\delta=3$,

Given converging error $\varepsilon>0$ (such as $\varepsilon=0.001$) choose the result membership matrix of FPCM-M algorithm as the initial one and the normalized result typicality matrix of PCM-M algorithm as the initial one respectively;

$$U^{(0)}=\begin{bmatrix}\mu_{11}^{(0)}\ \mu_{12}^{(0)}\dots\mu_{1n}^{(0)}\\ \mu_{21}^{(0)}\ \mu_{22}^{(0)}\dots\mu_{2n}^{(0)}\\ \dots\ \dots\ \dots\\ \mu_{c1}^{(0)}\ \mu_{c2}^{(0)}\dots\mu_{cn}^{(0)}\end{bmatrix}=\begin{bmatrix}\mu_1^{(0)}(\underline{x}_1)\ \mu_1^{(0)}(\underline{x}_2)\dots\mu_1^{(0)}(\underline{x}_n)\\ \mu_2^{(0)}(\underline{x}_1)\ \mu_2^{(0)}(\underline{x}_2)\dots\mu_2^{(0)}(\underline{x}_n)\\ \dots\ \dots\ \dots\\ \mu_c^{(0)}(\underline{x}_1)\ \mu_c^{(0)}(\underline{x}_2)\dots\mu_c^{(0)}(\underline{x}_n)\end{bmatrix} \quad (36)$$

$$T^{(0)}=\begin{bmatrix}t_{11}^{(0)}\ t_{12}^{(0)}\dots t_{1n}^{(0)}\\ t_{21}^{(0)}\ t_{22}^{(0)}\dots t_{2n}^{(0)}\\ \dots\ \dots\ \dots\\ t_{c1}^{(0)}\ t_{c2}^{(0)}\dots t_{cn}^{(0)}\end{bmatrix}=\begin{bmatrix}t_1^{(0)}(\underline{x}_1)\ t_1^{(0)}(\underline{x}_2)\dots t_1^{(0)}(\underline{x}_n)\\ t_2^{(0)}(\underline{x}_1)\ t_2^{(0)}(\underline{x}_2)\dots t_2^{(0)}(\underline{x}_n)\\ \dots\ \dots\ \dots\\ t_c^{(0)}(\underline{x}_1)\ t_c^{(0)}(\underline{x}_2)\dots t_c^{(0)}(\underline{x}_n)\end{bmatrix} \quad (37)$$

Step 2: Find

$$\underline{a}_i^{(k)}=\frac{\sum_{j=1}^{n}\left(\left[\mu_{ij}^{(k)}\right]^m+\left[t_{ij}^{(k)}\right]^\delta\right)\underline{x}_j}{\sum_{j=1}^{n}\left(\left[\mu_{ij}^{(k)}\right]^m+\left[t_{ij}^{(k)}\right]^\delta\right)}, \quad (38)$$

$i=1,2,\dots,c$

$$\Sigma_i^{(k)}=\frac{\sum_{j=1}^{n}\left(\left[\mu_{ij}^{(k)}\right]^m+\left[t_{ij}^{(k)}\right]^\delta\right)\left(\underline{x}_j-\underline{a}_i^{(k)}\right)\left(\underline{x}_j-\underline{a}_i^{(k)}\right)'}{\sum_{j=1}^{n}\left(\left[\mu_{ij}^{(k)}\right]^m+\left[t_{ij}^{(k)}\right]^\delta\right)} \quad (39)$$

$i=1,2,\dots,c$

$$\Sigma_i^{(k)}=\sum_{s=1}^{p}\lambda_{si}^{(k)}\Gamma_{si}^{(k)}\left(\Gamma_{si}^{(k)}\right)',$$

$$\left[\lambda_{si}^{(-1)}\right]^{(k)}=\begin{cases}\left[\lambda_{si}^{(k)}\right]^{-1} & if\ \lambda_{si}^{(k)}>0\\ 0 & if\ \lambda_{si}^{(k)}=0\end{cases} \quad (40)$$

$$\left[\Sigma_i^{-1}\right]^{(k)}=\sum_{s=1}^{p}\left[\lambda_{si}^{(-1)}\right]^{(k)}\Gamma_{si}^{(k)}\left(\Gamma_{si}^{(k)}\right)'$$

$$\left|\Sigma_i^{-1}\right|^{(k)}=\prod_{1\le s\le p,\ \lambda_{si}^{(k)}>0}\left[\lambda_{si}^{(-1)}\right]^{(k)}$$

$$\mu_{ij}^{(k)}=\left[\sum_{s=1}^{c}\left[\frac{\left(\underline{x}_j-\underline{a}_i^{(k)}\right)'\left[\Sigma_i^{-1}\right]^{(k)}\left(\underline{x}_j-\underline{a}_i^{(k)}\right)-\ln\left|\Sigma_i^{-1}\right|^{(k)}}{\left(\underline{x}_j-\underline{a}_s^{(k)}\right)'\left[\Sigma_s^{-1}\right]^{(k)}\left(\underline{x}_j-\underline{a}_s^{(k)}\right)-\ln\left|\Sigma_i^{-1}\right|^{(k)}}\right]^{\frac{1}{m-1}}\right]^{-1} \quad (41)$$

$i=1,2,\dots,c,\ j=1,2,\dots,n$

$$t_{ij}^{(k)}=\left[\sum_{s=1}^{n}\left[\frac{\left(\underline{x}_j-\underline{a}_i^{(k)}\right)'\left[\Sigma_i^{-1}\right]^{(k)}\left(\underline{x}_j-\underline{a}_i^{(k)}\right)-\ln\left|\Sigma_i^{-1}\right|^{(k)}}{\left(\underline{x}_s-\underline{a}_i^{(k)}\right)'\left[\Sigma_i^{-1}\right]^{(k)}\left(\underline{x}_s-\underline{a}_i^{(k)}\right)-\ln\left|\Sigma_i^{-1}\right|^{(k)}}\right]^{\frac{1}{\delta-1}}\right]^{-1}$$

$i=1,2,\dots,c.\ j=1,2,\dots,n$ (42)

Step 3: Increment k; until $\max_{1\le i\le c}\left\|\underline{a}_i^{(k)}-\underline{a}_i^{(k-1)}\right\|<\varepsilon$

F. The G-K Algorithm

The well-known Gustafson & Kessel algorithm (G-K algorithm) was proposed by Gustafson & Kessel (1979). It is a fuzzy partition clustering algorithms based on Mahalanobis distance and an extension of the fuzzy c-means algorithm on an adaptive norm, which will provide information about the clusters of various shapes in a data set. Each cluster is characterized by its normalization matrix $M_i\in M$. The matrix M_i is applied as the optimization of variables in the c-means functional. Each cluster is able to adapt its own norm, in accordance with a topology data of a specific region. The objective function is defined as the equation of (43).

$$J_{G-K}^{m}(U,A,M,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}\right)\left(\underline{x}_j-\underline{a}_i\right)'M_i\left(\underline{x}_j-\underline{a}_i\right) \quad (43)$$

constraints：membership

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,\dots,n, \quad (44)$$

Each groups of the determinent of standardization covariance matrix of cluster I,

$$\left|M_i\right|=\rho_i,\ \forall i=1,2,\dots,c \quad (45)$$

If there is no prior information about ρ_i , then $\rho_i = 1, \forall i = 1, 2, ..., c$. The algorithm is described as follows.

Step 1: Determining the number of cluster c, m-value (let m=2), and the converging error, $\varepsilon > 0$ (such as $\varepsilon = 0.001$), and choosing the initial membership matrix.

$$U^{(0)} = \begin{bmatrix} \mu_{11}^{(0)} \ \mu_{12}^{(0)} \cdots \mu_{1n}^{(0)} \\ \mu_{21}^{(0)} \ \mu_{22}^{(0)} \cdots \mu_{2n}^{(0)} \\ \cdots \quad \cdots \quad \cdots \\ \mu_{c1}^{(0)} \ \mu_{c2}^{(0)} \cdots \mu_{cn}^{(0)} \end{bmatrix} \quad (46)$$

Step 2: To calculate

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m \underline{x}_j}{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m} \quad i = 1, 2, ..., c \quad (47)$$

$$F_i^{(k)} = \frac{\sum_{j=1}^{n} \left(\mu_{ij}^{(k-1)} \right)^m \left(\underline{x}_j - \underline{a}_i^{(k)} \right) \left(\underline{x}_j - \underline{a}_i^{(k)} \right)'}{\sum_{j=1}^{n} \left(\mu_{ij}^{(k)} \right)^m} \quad (48)$$

$$M_i^{(k)} = \left[\rho_i \det\left(F_i^{(k)} \right) \right]^{\frac{1}{p}} \left(F_i^{(k)} \right)^{-1} \quad (49)$$

$$\mu_{ij}^{(k)} = \left[\sum_{l=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' M_i^{(k)} \left(\underline{x}_j - \underline{a}_i^{(k)} \right)}{\left(\underline{x}_j - \underline{a}_l^{(k)} \right)' M_i^{(k)} \left(\underline{x}_j - \underline{a}_l^{(k)} \right)} \right]^{\frac{1}{m-1}} \right]^{-1} \quad (50)$$

Step 3: Increment k; until $\max_{1 \le i \le c} \left\| \underline{a}_i^{(k)} - \underline{a}_i^{(k-1)} \right\| < \varepsilon$.

III. Data Resource

We have two real data sets was applied to prove that the performance of the FCM-MS algorithm gave more accurate clustering results than the FCM and FCM-M methods, and the ratio method which is proposed by us is the better of the two methods for selecting the initial values.

A. *Experiment of Mathematics Teaching Data*

A real data set of students with sample size 493 from elementary schools was selected. These data included the independent variables, test scores of four mathematics concepts (division, ordering, multiplication, and place value) and 10 questions. The samples were assigned to 4 clusters. The results were shown in Table 1.

Table 1. The Characteristics of 4 Clusters

Cluster	Sample size	Mathematics concepts	Average distance from points to center
1	115	division	1.2576760
2	128	ordering	1.2968550
3	168	multiplication	1.1244569
4	82	place value	1.7861002

Each 15 sample points were randomly drawn from Cluster 1, cluster 2, and cluster 3, respectively, and 5 from cluster 4. How to select the better initial value to improve the cluster accuracy is an important issue. In order to test the FCM-M algorithm, developed by the authors of this paper, the four .25 were selected as initial value. After calculating, the results were found that the memberships were all equal to .25 too. This evidence displayed that the FCM-M algorithm was work correctly. There were 2 methods (*Ratio, Random*) to calculate the Normalized initial number. which satisfied the Equation (2) .

The steps of ***Ratio Method*** were as follows.

Step 1: The distance between observing value and every cluster center of every Point, say d. Compute the average distance of clustering result marking group.

$$cd_i^{(0)} = \sum_{j=1}^{n_i} d_j n_i^{-1} \ n_i, number\ of\ \text{Result}\ Marking\ Group\ i$$

Step 2: Compute the difference of d and the average distance of clustering result marking group

$$l_{ij} = \left| d_j - cd_i \right| \ j = 1, 2, ..., n_i, \ i = 1, 2, ..., c$$

Step 3: Find the values of maximum and minimum

$$M = \max\left\{ l_{ij} \ j = 1, 2, ..., n_i, \ i = 1, 2, ..., c \right\},$$

$$m = \min\left\{ l_{ij} \ j = 1, 2, ..., n_i, \ i = 1, 2, ..., c \right\}$$

Step 4: Compute the initial membership Difference of every Point $\mu_{ij} = (M - l_{ij})(M - m)^{-1}$.

The steps of ***Random Method*** were as follows. Choose any 4 random numbers, r_1, r_2, r_3, r_4 .

Table2.Classification Accuracies of Testing Samples

Choosing the initial membership	Computing distance	Classification Accuracies (%)
Ratio	FCM-MS	54
	FCM-M	50
	FCM	40
Random	FCM-MS	50
	FCM-M	30
	FCM	26

From the data of Table 2, we found that the FCM-MS algorithm presented the best clustering accuracies, up to 54% and the Ratio method of FCM-MS could obtain the better results.

B. *Experiment of Teaching Fraction Data*

Another real data set of students with sample size 146 from elementary schools was selected. The main factors of the data were calculated by using factor analysis. According to the main factors, the samples were assigned to 4 clusters based on the clustering analysis. The results were shown in Table 4.

From the data of Table 5, we found that the Ratio method could obtain the best results. A real data set was applied to prove that the performance of the FCM-MS algorithm gave more accurate clustering results than the FCM-M and FCM methods.

Table 4. the characteristics of 4 clusters

Cluster	Sample size	mathematics concepts	average distance of the points from center of cluster
1	36	Partition	-.14984
2	89	Unit	.21161
3	16	Fraction	-.30416
4	5	Unknown unit	-.74490

Each 15 sample points were randomly drawn from Cluster 1, cluster 2, and cluster 3, respectively, and 5 from cluster 4.

The classification accuracies of testing samples were shown in Table 5

Table5.Classification Accuracies of Testing Samples

Choosing the initial membership	Computing distance	Classification Accuracies (%)
Ratio	FCM-MS	56
	FCM-M	38
	FCM	36
Random	FCM-MS	44
	FCM-M	30
	FCM	24

From the data of Table 5, we found that the FCM-MS algorithm presented the best clustering accuracies, up to 56% and the Ratio method of FCM-MS could obtain the better results.

C. *Experiment of Teaching Geometry Data*

A real data set of sample size 968 students from elementary schools was selected. These data included the 10 mathematics questions.

At first, the main factors of 968 data were calculated by using factor analysis. Next, according to the main factors, the samples were assigned to 4 clusters based on the clustering analysis using the k-mean clustering of SPSS for Windows 10.0. The results were shown in Table 6.

Table 6 The characteristics of 4 clusters

Cluster	Samples size	Grade	average distance of the points from center of cluster
1	220	2	2.082132
2	435	4	1.433158
3	275	3	2.032674
4	56	1	2.356698

From Cluster 1, 15 samples randomly were selected, 15 from cluster 2, 15 from cluster 3, and 5 from cluster 4.The combination the method of choosing the initial membership with distinct computing distance was shown in Table 7.

Table 7 Data Cluster and Sample sized

Cluster	Number of Samples
1	15
2	15
3	15
4	5

From the data of Table 8, we found that the algorithm based on unsupervised Mahalanobis distance of FCM-M is better classification accuracies than based on Euclidean distance of FCM , up to 52%. Similarly, Presented the best classification accuracies 58% is also based on unsupervised Mahalanobis distance of FPCM-M[13], up to 58%.

Table 8 Classification accuracies of testing samples.

Computing distance	Classification Accuracies (%)
FCM	32
FCM-M	52
FPCM	30
FPCM-M	58

IV. Conclusions

Extending Euclidean distance to Mahalanobis distance, Gustafson-Kessel (GK) clustering algorithm and Gath-Geva (GG) clustering algorithm, are developed to detect non-spherical structural clusters, but both of them based on semi-supervised Mahalanobis distance, these two algorithms fail to consider the relationships between cluster centers in the objective function, needing additional prior information.

When some training cluster size is small than its dimensionality, it induces the singular problem of the inverse covariance matrix. It is an important issue. The other important issue is how to select the better initial value to improve the cluster accuracy[15]. In this paper, focusing attention to above two problems, an improved new fuzzy clustering algorithm, FCM-MS, is developed to obtain better quality of fuzzy clustering results. The objective function includes a fuzzy within-cluster scatter matrix, a new between-prototypes scatter matrix, the regulating terms about the covariance matrices, and the regulating terms about the relationships between cluster centers, the relationships between the center of all points and the cluster centers. The update equations for the memberships and the cluster centers and the covariance matrices are directly derived from the Lagrange's method, which is different from the GK and GG algorithms.

The singular problem and the selecting initial values problem are improved by the Eigenvalue method and the Ratio method. Finally, a numerical example shows that the new fuzzy clustering algorithm FCM-MS gives more accurate clustering results than the FCM and FCM-M algorithms for a real data set, the ratio method which is proposed by us is the better of the two methods for selecting the initial values.

APPENDIX

Proof the initial memberships of FCM-M Algorithm and FCM Algorithm can not be all equal.

[Proposition] The initial memberships of FCM-M Algorithm and FCM Algorithm can not be all equal.

[proof：] (i) In FCM-M Algorithm, Let

$$u_{ij}^{(0)} = c^{-1}\ ,\ i = 1, 2, \ldots, c\ ,\quad j = 1, 2, \ldots, n$$

We get

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m \underline{x}_j}{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m} = \frac{\sum_{j=1}^{n}\left[c^{-1}\right]^m \underline{x}_j}{\sum_{j=1}^{n}\left[c^{-1}\right]^m} = \frac{\sum_{j=1}^{n}\underline{x}_j}{n} = \underline{a} \quad i = 1, 2, \ldots, c$$

$$\Sigma_i^{(k)} = \frac{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m \left(\underline{x}_j - \underline{a}_i^{(k)}\right)\left(\underline{x}_j - \underline{a}_i^{(k)}\right)'}{\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m} = \frac{\sum_{j=1}^{n}\left[c^{-1}\right]^m \left(\underline{x}_j - \underline{a}\right)\left(\underline{x}_j - \underline{a}\right)'}{\sum_{j=1}^{n}\left[c^{-1}\right]^m}$$

$$\Sigma_i^{(k)} = \frac{\sum_{j=1}^{n}\left(\underline{x}_j - \underline{a}\right)\left(\underline{x}_j - \underline{a}\right)'}{n} = \Sigma$$

$$\mu_{ij}^{(k)} = \left[\sum_{s=1}^{c}\left[\frac{\left(\underline{x}_j - \underline{a}\right)'\left[\Sigma^{-1}\right]\left(\underline{x}_j - \underline{a}\right) - \ln\left|\left[\Sigma^{-1}\right]\right|}{\left(\underline{x}_j - \underline{a}\right)'\left[\Sigma^{-1}\right]\left(\underline{x}_j - \underline{a}\right) - \ln\left|\left[\Sigma^{-1}\right]\right|}\right]^{\frac{1}{m-1}}\right]^{-1} = \frac{1}{c}$$

The proof is completed.

(ii) In FCM Algorithm, The proof is similar as above.

ACKNOWLEDGMENT

This paper is partially supported by the National Science Council grant (NSC-97 -2511-S -142-001).

REFERENCES

[1] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, NY, 1981.

[2] I. Gath, and A. B. Geva, Unsupervised Optimal Fuzzy Clustering, IEEE Trans. Pattern Anal. Machine Intell., 1989, pp. 773-781.

[3] D. E. Gustafson and W. C. Kessel, Fuzzy Clustering with a Fuzzy Covariance Matrix, Proc. IEEE Conf. Decision Contr., San Diego, CA, 1979.

[4] H. C. Liu, J. M. Yih and S. W. Liu, Fuzzy C-mean Algorithm Based On Mahalanobis Distances and Better initial values, 12th International Conference on Fuzzy Theory & Technology, JCIS, Salt Lake City , Utah, 2007.

[5] Z. H. Yin, Y. G. Tang Yuangang, F. C. Sun, and Z. Q. Sun, Fuzzy Clustering with Novel Separable Criterion, Tsinghua Science and Technology, Vol. 11, 2006, pp. 50-53.

[6] J. C. Dunn, A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact, Well-separated Clusters, J. Cybern, Vol. 3, 1973, pp. 32-57.

[7] J. Yu, Q. Cheng, and H. Huang, Analysis of the Weighting Exponent in the FCM, IEEE Trans. Systems, Man and Cybernetics, Part B, Vol. 34, 2004, pp 634-639.

[8] H. C. Liu, J. M. Yih, T.W. Sheu and S.W. Liu, A New Fuzzy Possibility Clustering Algorithms Based On Unsupervised Mahalanobis Distances, 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, 2007, pp.3939-3944.

[9] Somesh Das Gupta., "The evolution of the D2-statistics of Mahalanobis", Indian J. Pure Appl. Math.,26(1995), no.6,485-501.

[10] Besset, Didier H., "Object Oriented Imple-mentation of Numerical Methods".Morgan Kaufmann Publisfers(2001). U. S. A. ISBN:1-55860-679-3.

[11] H. C. Liu, J. M. Yih, Tian-Wei Sheu, Shin-Wu Liu (2007). A new Fuzzy Possibility Clustering Algorithms Based on Unsupervised Mahalanobis Distances. Proceedings of International Conference on Machine Learning and Cybernetics 2007 (ICMLC 2007), Vol.7. No.7, pp. 3939-3944, ISBN 1-4244-0973X. EI級論文 [Hong Kong, China, August 19-22, 2007].

[12] J. M. Yih, Yuan-Horng Lin, Hsiang-Chuan Liu (2008). FCM Algorithm Besed on Unsupervised Mahalanobis Distances with Better Initial Values and Separable Criterion. Proceedings of the 8th WSEAS International Conference on APPLIED COMPUTER SCIENCE (ACS'08), pp. 326-331, ISSN: 1790-5109/ ISBN: 978-960-474-028-4. [Venice, Italy, November 21-23, 2008].

[13] N. R. Pal, K. Pal and J.C. Bezdek, "A mixed c-mean clustering model", Proceedings of the Sixth IEEE Internationa l conference on Fuzzy Systems, Vol. 1, pp. 11-21, Jul. 1997.

[14] Seyed Amir Hadi Minoofam, Azam Bastanfard, (2008) A Novel Algorithm for Generating Muhammad Pattern Based on Cellular Automata Pro ceedings of the 13th WSEAS International Conference on APPLIED MATHEMATICS (MATH'08),pp339-344, ISSN: 1790-2769, ISBN: 978-960-474-034-5.

[15] J. M. Yih, Y.H. Lin, Hsiang-Chuan Liu (2008). Clustering Analysis Method Besed on Fuzzy C-Means Algorithm of PSO and PPSO with Application in Image Data. Proceedings of the 8th WSEAS International Conference on APPLIED COMPUTER SCIENCE (ACS'08), pp. 54-59, ISSN: 1790-5109/ ISBN: 978-960-474-028-4. [Venice, Italy, November 21-23, 2008].

[16] J. M. Yih, Y.H. Lin, Hung, W. L.(2007). Fuzzy Approach Method for Concept Structure Analysis Based on FLMP and ISM with Application in Cognition Diagnosis of Linear Algebra. Information Sciences 2007 Proceedings of The 10th Joint Conference & The 12th International Conference on Fuzzy Theory & Technology (FTT 2007), pp. 1356-1362, EI-paper [Salt Lake City, Utah, U. S. A., July 18-24, 2007].

[17] J. M. Yih, Y.H. Lin (2007). A Fuzzy Basis on Knowledge Structure Analysis for Cognition Diagnosis and Application on Geometry Concepts for Pupils. The Third International Conference on Intelligence

Information Hiding and Multimedia Signal Processing. (IIH-MSP 2007) Vol.2, pp. 183-186, ISBN: 0-7695-2994-1. EI-paper [Kaohsiung City, Taiwan, November 26-28, 2007].
[18] J. M. Yih, Y.H. Lin (2007). An Integration of Fuzzy Theory and ISM for Concept Structure Analysis with Application of Learning MATLAB. The Third International Conference on Intelligence Information Hiding and Multimedia Signal Processing. (IIH-MSP 2007) Vol.2, pp. 187-190, ISBN: 0-7695-2994-1. EI-paper [Kaohsiung City, Taiwan, November 26-28, 2007].
[19] Y.H. Lin, J. M. Yih (2008). Fuzzy Logic Approach on Cognition Diagnosis with Application on Number Concept for Pupils. Proceedings of 2008 International Conference on Machine Learning and Cybernetics. pp. 3575-3580, ISBN: 978-1-4244-2095-7. EI-paper [Grand Park Hotel, Kunming, China, July 12-15, 2008]
[20] Y.H. Lin, J. M. Yih, Min-Yen Chen (2008). Polytomous IRS with Application in Concepts Diagnosis and Clustering on Fraction Subtraction for Pupils. Second International Symposium on Intelligentv Information Technology Application. pp. 463-467, ISBN: 978-0-7695-3497-8.EI-paper[Shanghai, China, December 21-22, 2008]
[21] Y.H. Lin, J. M. Yih (2008). An Integration of Concept Structure Analysis and S-P Chart with Application in Equality Axiom Concepts Diagnosis. Second International Symposium on Intelligentv Information Technology Application. pp. 468-472, ISBN: 978-0-7695-3497-8.EI-paper[Shanghai, China, December

J. M. Yih is with the Graduate Institute of Educational Measurement and Statistics, and Department of Mathematics Education, National Taichung University 140 Min-Sheng Rd., Taichung City 403, Taiwan and the Dean of General Affairs of National Taichung University

In the future, We will consider Fuzzy Approach Method for Concept Structure Analysis[16]-[21].

Hsiang-Chuan Liu Department of Bioinformatics
Asia University, Taichung, 41354, Taiwan

Chi-Fu Yih Department of Computer Science and Information Engineering National Chi Nan University , Nantou, Taiwan. Email: aaroncat1758@gmail.ntcu.edu.tw.

Hsiang-Chuan Liu received the Ph.D. degree in Statistics from National Taiwan University, Taiwan. He is a professor at the Department of Bioinformatics, Asia University, Taiwan since August 2001 and also a honored professor at the Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taiwan. He was the President of National Taichung University , Taiwan from 1993 to 2000. He has funded research and published articles in the areas of Biostatistics, Bioinformatics, Fuzzy Theory, Educational Measurement, and E-Learning.

Jeng-Ming Yih received the B.S. and M.S. degrees from the National Taiwan Normal University, Taipei, Taiwan, R. O. C. in 1983 and 1986, respectively.He is currently a Asociate Professor with Department of Mathematics Education at Taichung University, Taichung, Taiwan, R.O.C. His research interests include pattern recognition of teaching linear algebra and fuzzy clustering.