International Journal of Computer and Information Technology (ISSN: 2279 – 0764)
Volume 05 – Issue 04, July 2016

Graphic Representation on Linear Algebra based on Clustering Approach

Fong-Jhu Yih
National Chung Cheng University
Tainan , Taiwan

Jeng-Ming Yih *
Min-Hwei Junior College of Health Care Management
Tainan, Taiwan
Email: yih [AT] mail.mhchcm.edu.tw

Abstract—**Euclidean distance function based fuzzy clustering algorithms can only be used to detect spherical structural clusters. The purpose of this study is improved Fuzzy C-Means algorithm based on Mahalanobis distance to identify concept structure for Linear Algebra. In addition, Concept structure analysis (CSA) could provide individualized knowledge structure. CSA algorithm is the major methodology and it is based on fuzzy logic model of perception (FLMP) and interpretive structural modeling (ISM). CSA could display individualized knowledge structure and clearly represent hierarchies and linkage among concepts for each examinee. Each cluster of data can easily describe features of knowledge structures. The results show that there are five clusters and each cluster has its own cognitive characteristics. In this study, the authors provide the empirical data for concepts of linear algebra from university students. To sum up, the methodology can improve knowledge management in classroom more feasible. Finally, the result shows that Algorithm based on Mahalanobis distance has better performance than Fuzzy C-Means algorithm.**

Keywords-Mahalanobis Distances; fuzzy clustering algorithms; interpretive structural modeling; ordering theory

I. INTRODUCTION

The well-known ones, such as Bezdek's Fuzzy C-Means (FCM)[1], FCM algorithm was based on Euclidean distance function, which can only be used to detect spherical structural clusters. To overcome the drawback due to Euclidean distance, we could try to extend the distance measure to Mahalanobis distance (MD). However, Krishnapuram and Kim (1999) [2] pointed out that the Mahalanobis distance can not be used directly in clustering algorithm. Gustafson-Kessel (GK) clustering algorithm [3] and Gath-Geva (GG) clustering algorithm [4] were developed to detect non-spherical structural clusters. In GK-algorithm, a modified Mahalanobis distance with preserved volume was used. However, the added fuzzy covariance matrices in their distance measure were not directly derived from the objective function. In GG algorithm, the Gaussian distance can only be used for the data with multivariate normal distribution.We know Gustafson-Kessel clustering algorithm and Gath-Geva clustering algorithm, were 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. Added a regulating factor of covariance matrix, , to each class in objective function, the fuzzy covariance matrices in the Mahalanobis distance can be directly derived by minimizing the objective function, but the clustering results of this algorithm is still not stable enough. For improving the stability of the clustering results, we replace all of the covariance matrices with the same common covariance matrix in the objective function in the FCM-M algorithm, and then, an improve fuzzy clustering method, called the Fuzzy C-Means algorithm based on Normalized Mahalanobis distance (FCM-NM), is proposed. Zadeh developed fuzzy theory and it flourishes methodologies in many fields [5] [6]. One of these fields is cognition diagnosis and it help represent knowledge structure [7] [8] [9]. It is a common viewpoint that human knowledge is stored in the form of structural relationship among concepts and their subordinate relationship is fuzzy, not crisp. There are some methodologies for concept structure analysis but little is known about methodologies of individualized concept structure [10] [11] [12] [13]. Therefore, the development for methodology of individualized concept structure is an important issue and it is essential for cognition diagnosis and pedagogy [14]. In this study, the integrated method of individualized concept structure based on fuzzy logic model of perception (FLMP) and interpretive structural modeling (ISM) will be developed [15] [16] [17]. An example of empirical test data of linear algebra concept for students of learning deficiencies will also be analyzed and discussed. For the feasibility of remedial instruction based on the cognition diagnosis, clustering method is needed so that students within the same cluster own similar knowledge structures and students among different clusters have the most variance on knowledge structures [18] [19]. In addition, cognitive diagnosis is essential for educational environment. As to cognitive diagnosis, clustering technique is useful to classify students and then features of concept structures from each cluster could reveal constructive information for cognitive diagnosis. Therefore, remedial instruction will be more feasible [20]. For the feasibility of remedial instruction based on the cognition diagnosis, clustering method is needed so that students within the same cluster own similar knowledge structures and students among different clusters have the most variance on knowledge structures [21,22].

II. FRAMEWORK

A. Fuzzy C-Mean Algorithm

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

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

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

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

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

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

Step 2: Find

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

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

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

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. [23]

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.

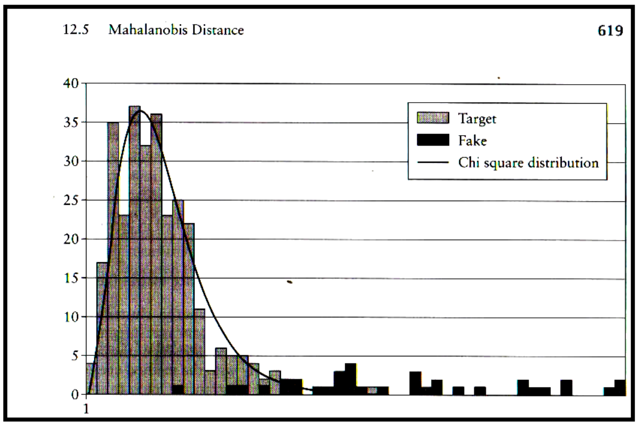

Figure 2. Use the Mahalanobis distance to detect the fake coins (Convert from Besset D. H. p619, FIG. 12.2) .

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.

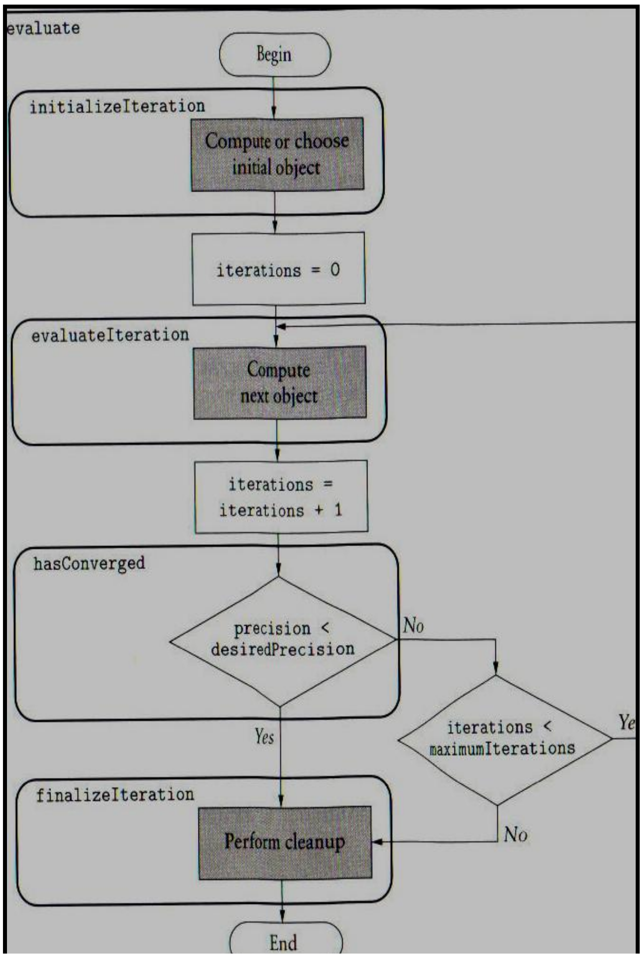

Figure 2. Method for successive approximation (Convert from Besset D. H. p118, FIG. 4.4) . .

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 [24] .

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 .

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

Constraints: membership,

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,\ldots,n \quad (7)$$

$\Sigma = \{\Sigma_1, \Sigma_2, \ldots, \Sigma_c\}$ 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 $\underline{a}_i^{(0)}, i = 1,2,\ldots,c$ 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 $\underline{a}_i^{(0)}$.

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

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

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

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

Step 2: Find

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

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

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

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

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

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

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

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

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 $x_j = (x_{1j}, x_{2j}, \ldots, x_{pj})' \in R^p$. 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

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

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

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

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 $\underline{a}_i^{(0)}, i = 1,2,\dots,c$ 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 $\underline{a}_i^{(0)}$.

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

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

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

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

Step 2: Find

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

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

$$\underline{a}_i^{(k)} = \left[\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m \left[\Sigma_i^{(k-1)} \right]^{-1} - \frac{1}{c(c-1)} \sum_{l=1}^{c} \left[v_{il}^{(k)} \right]^m I \right]^{-1}$$

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

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

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

$$(\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m)^{-1}, \quad (22)$$

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

$$\left[\lambda_{si}^{(-1)} \right]^{(k)} = \left[\lambda_{si}^{(k)} \right]^{-1} \quad \textit{if} \ \lambda_{si}^{(k)} > 0,$$

$$\left[\lambda_{si}^{(-1)} \right]^{(k)} = 0 \qquad \textit{if} \ \lambda_{si}^{(k)} = 0,$$

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

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

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

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

D. FCM-NM Algorithm

In this paper, not only z-score normalizing for each feature in the objective function in the FCM-CM algorithm, but also replacing the threshold D where

$$D = \sum_{i=1}^{c} \sum_{j=1}^{n} \left[\mu_{ij}^{(0)} \right]^m \left[\left(\underline{x}_j - \underline{a}_i^{(0)} \right)' \left(\underline{x}_j - \underline{a}_i^{(0)} \right) \right] > 0 \quad (26)$$

With the determinant value of the crisp correlation matrix, and then, the new fuzzy clustering method, called the Fuzzy C-Means algorithm based on normalized Mahalanobis distance (FCM-NM) is proposed. We can obtain the objective function of FCM-NM as following:

$$J_{FCM-NM}^{m}(U, A, R, Z) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} d^2 \left(z_j, \underline{a}_i \right) \quad (27)$$

Ωηερε $\quad X = \left[\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n \right], \underline{x}_j \in R^p, j = 1,2,\dots,n \quad (28)$

$$\underline{z}_j = \left(z_{1j}, z_{2j}, \dots, z_{pj} \right)', z_{tj} = \frac{x_{tj} - \overline{x}_t}{s_t}, j = 1,2,\dots,n, t = 1,2,\dots,p \quad (29)$$

$$\overline{x}_t = \frac{1}{n} \sum_{j=1}^{n} x_{tj}, s_t = \frac{1}{n} \sum_{j=1}^{n} \left(x_{tj} - \overline{x}_t \right)^2, \quad t = 1,2,\dots,p \quad (30)$$

Conditions for FCM-NM are

$$m \in [1,\infty); U = \left[\mu_{ij} \right]_{c \times n}; \mu_{ij} \in \left[0,1 \right], i = 1,2,\dots,c, j = 1,2,\dots,n$$

$$\sum_{i=1}^{c} \mu_{ij} = 1, j = 1,2,\dots,n, 0 < \sum_{j=1}^{n} \mu_{ij} < n, i = 1,2,\dots,c \quad (31)$$

$$d^2\left(\underline{z}_j, \underline{a}_i \right) = \begin{cases} \left(\underline{z}_j - \underline{a}_i \right)' R^{-1} \left(\underline{z}_j - \underline{a}_i \right) - \ln \left| \Sigma^{-1} \right| & \textit{if} \left(\underline{z}_j - \underline{a}_i \right)' R^{-1} \left(\underline{z}_j - \underline{a}_i \right) - \ln \left| R^{-1} \right| \ge 0 \\ 0 & \textit{if} \left(\underline{z}_j - \underline{a}_i \right)' R^{-1} \left(\underline{z}_j - \underline{a}_i \right) - \ln \left| R^{-1} \right| < 0 \end{cases} \quad (32)$$

the threshold of FCM-NM is a dynamic value rather than a constant, it is different from which of FCM-SM in our previous work [25], and the convergent process is different from all of before mentioned algorithms.

E. Fuzzy Logic Model of Perception

Suppose there be a combination of two factors and . There are levels and levels for factor and respectively. The fuzzy true values are expressed as and . Fuzzy truth value and express the degree that the combination of and will support prototype [26] [27]. The probability that the combination of could be viewed as prototype can be expressed as follows [13] [28].

$$p(c_i, o_j) = (c_i o_j)[c_i o_j + (1 - c_i)(1 - o_j)]^{-1} \quad (33)$$

F. Integrated model and Procedure

The procedure of the integrated model is depicted in Figure 1. Firstly, FCM is to classify examinee based on their response pattern. Secondly, concept structure analysis (CSA) will analyze individualized knowledge structures. CSA includes three algorithms, which are AMC (algorithm for mastery of concept), ASC (algorithm for subordination of concepts) and AFISM (algorithm for fuzzy ISM). By the integrated procedure, Examinee within the same cluster represent similar knowledge structures and remedial instruction could be feasible based on the information of cognition diagnosis within the same cluste.

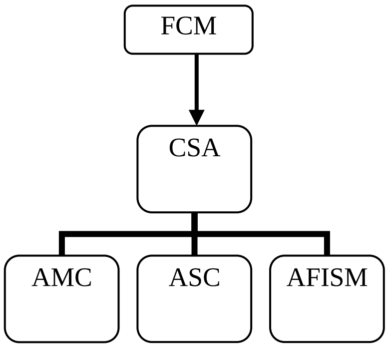

Figure 3. Procedure of the Integrated Model.

Three algorithms, AMC, ASC and AFISM, are combined in order to analyze individualized knowledge structure. Some basic definitions are as follows [29].

G. Fuzzy Clustering

Proper clustering number will be decided in advance and one student will be randomly selected from each cluster to describe features of knowledge structures. The proper number of cluster is 3.In this paper, we use the best performance of clustering Algorithm FCM-CM in data analysis and interpretation. It groups data into 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 clustering is widely used in the pattern recognition field. Hence each cluster of data can easily describe features of knowledge structures. Manage the knowledge structures of Mathematics Concepts to construct the model of features in the pattern recognition completely.

H. S-P chart and OT

S-P Chart is related to the issue of ideal response pattern (Guttman scale). Two indices are generated from S-P chart for items and students respectively [30]. There are N ($i=1,2,\cdots,N$) students and M ($j=1,2,\cdots,M$) dichotomous items. $Y=(y_{ij})_{N\times M}$ is the response matrix.

III. METHODOLOGY AND EMPIRICAL ANALYSIS

Linear algebra test for university students is designed by author. The data set used in the experimental study is an educational data from university students in Taiwan.

A. Fuzzy Clustering Algorithms

Attributes of concepts are depicted in Table 1.

TABLE I. THE DETAILS OF USED DATA SETS

Data sets	Attributes	Classes	Sample Size
Educational data	19	6	831

The educational data was collected from an achievement test of linear algebra of university students in Taiwan, which includes three concepts. According to their scores of achievement test, we can group the students into three mastery types or concepts, the concepts are shown in Table 2.

TABLE II. THE CHARACTERISTICS OF LINEAR ALGEBRA CONCEPTS

Cluster	Concepts of linear algebra	Sample Size
1	Matrix oprations	291
2	System of linear equations	104
3	Determinants	116
4	Vector space and the properties of R^n	141
5	Geometric properties of linear algebra	109
6	Eigen-value and Eigen-vector	70

The clustering performances of each data set are calculated by applying five fuzzy clustering algorithms, as mentioned above, with the same fuzzier m=2. In these experiments, the mean clustering accuracies of 100 different initial value set were calculated and compared for this data set. From Table 3, we can find that FCM Algorithm with Mahalanobis Distances has the better performance.

TABLE III. THE ACCURACIES OF FOUR ALGORITHMS

Algorithms	Accuracies(%)
FCM	69.53
FCM-M	75.56
FCM-MS	76.23
FCM-NM	79.76

The Mean clustering Accuracies of 100 different initial value sets of FCM and FCM-NM for the Dataset was shown in TABLE 4. From this table, we can find that the Accuracies of FCM is worse than the FCM Algorithm with Mahalanobis Distances in the dataset.

B. Method of Fuzzy Approach on Concept Structure Analysis

The data set used in the experimental study is an educational data from university students in Taiwan. As shown from Figure 4 to Figure 9, one student is randomly selected from each cluster respectively.

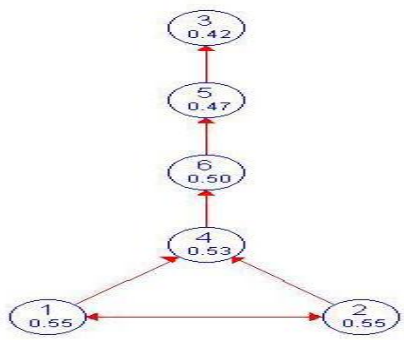

Figure 4. knowledge Structure of Student 25 in Cluster 1

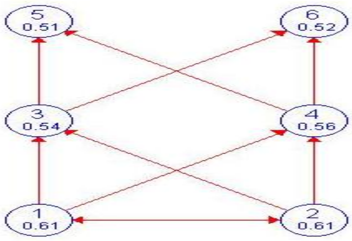

Figure 5. knowledge Structure of Student 98 in Cluster 2.

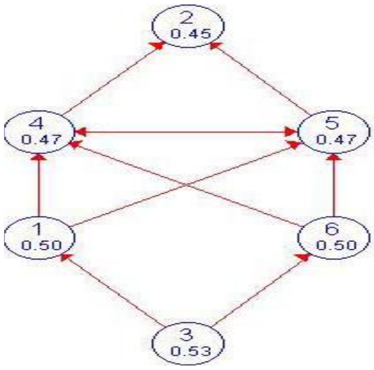

Figure 6. knowledge Structure of Student 225 in Cluster 3

As to student 25 in cluster 1, mastery of concept 1 is 0.55. Concept 1 and 2 are the basis for concept 4, 6, 5, 3. For student 98 in cluster 2, mastery of concept 1 and 2 is 0.61. Concept 1 and 2 are also the basis of the other concepts. As to student 225 in cluster 3, mastery of concept 3 is 0.53. Concept 3 are the basis for concept 1,6,4, 5.

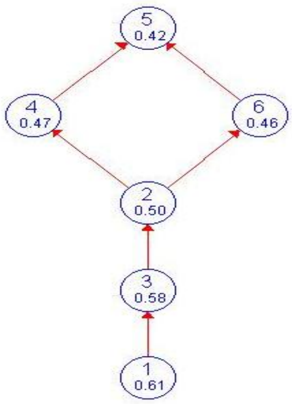

Figure 7. knowledge Structure of Student 342 in Cluster 4

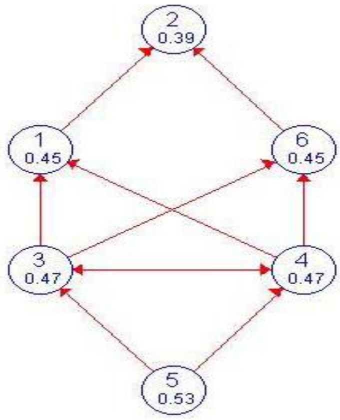

Figure 8. knowledge Structure of Student 568 in Cluster 5

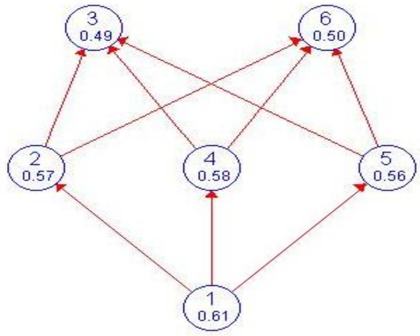

Figure 9. knowledge Structure of Student 738 in Cluster 6

As to student 342 in cluster 4, mastery of concept 1 is 0.61. Concept 1 is the basis for concept 3,4. As to student 568 in cluster 5, mastery of concept 5 is 0.53. Concept 5 are the basis for concept 3,4,1,6,and 2. For student 738 in cluster 6, mastery of concept 1 is 0.61. Concepts 2,4 and 5 are also the basis of concept 3 and concept 6. One is concluded that master of concepts for these six students are different. It is also clearly understood that concept structures of these six students vary a lot.

IV. CONCLUSION

A Fuzzy C-Means algorithm based on Mahalanobis distance (FCM-M) was proposed to improve those limitations of above two algorithms, but it is not stable enough when some of its covariance matrices are not equal. An improved Fuzzy C-Means algorithm based on Normalized Mahalanobis distance (FCM-NM) is proposed. The experimental results of two real data sets consistently show that the performance of our proposed FCM-NM algorithm is better than those of the FCM algorithms.

Each cluster of data can easily describe features of knowledge structures. We can manage the knowledge structures of linear algebra concepts to construct the model of features in the pattern recognition completely. An integrated method of FLMP and ISM for analyzing individualized concept structure is provided. With this integrated algorithm, the graphs of concept structures will display the characteristics of knowledge structure. This result corresponds with foundation of cognition diagnosis in psychometrics [13]. This study investigates an integrated methodology to display knowledge structures based on fuzzy clustering with Mahalanobis Distances. In addition, empirical test data of abstract algebra for university students are discussed. It shows that knowledge structures will be feasible for remedial instruction [31-33]. This procedure will also useful for cognition diagnosis. To sum up, this integrated algorithm could improve the assessment methodology of cognition diagnosis and manage the knowledge structures of linear algebra concepts easily. In addition, empirical test data of linear algebra for university students are discussed. It shows that knowledge structures will be feasible for remedial instruction. This procedure will also useful for cognition diagnosis.

ACKNOWLEDGMENT

This paper is partially supported by Min-Hwei Junior College of Health Care Management of Taiwan (MHCHCM105012) and also thanks the managing editors and reviewers for continued guidelines and untiring comments for the context of the paper.

REFERENCES

[1] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum press, (1981).pp. 65-70, N.Y.

[2] R. Krishnapuram and J. Kim, A note on the Gustafson-Kessel and adaptive fuzzy clustering algorithm, IEEE Transactions on Fuzzy Systems (1999). vol. 7 no. 4 August.

[3] D. E. Gustafson and W. C. Kessel, Proc. IEEE Conf. Decision Contr. San Diego, CA, 761 ,1979.

[4] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, Fuzzy Cluster Analysis(1999). John Wiley and Sons,.

[5] G. J. Klir, and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall (1995).New York, NY,.

[6] L. A. Zadeh, Fuzzy sets, Information and Control, 1965, Vol. 8, pp. 338-353.

[7] M. Smithson, and J. Verkuilen, Fuzzy Set Theory: Applications in the Social Sciences, Sage Publications (2006). Thousand Oaks, CA.

[8] R. Coppi, P. Giordani and P. D'Urso, Component Models for Fuzzy Data, Psychometrika (2006). Vol. 71, pp. 733-761.

[9] T. Sato, The S-P Chart and The Caution Index, NEC Educational Information Bulletin 80-1, C&C Systems Research Laboratories (1980). Nippon Electic Co., Ltd,. Tokyo, Japan.

[10] J. P. Doignon and J. C. Falmagne, Knowledge Space(1999). Springer-Verlag.

[11] K. VanLehn, Journal of the Learning Sciences(1999). Vol.8, p.71.

[12] R. W. Schvaneveldt, Pathfinder Associative Networks (1991). Ablex.

[13] W. P. Jr. Fisher, Rasch Measurement Transactions (1995). Vol.9, p. 442.

[14] R. J. Mislevy and N. Verhelst, Psychometrika (1990). Vol. 55, p.195.

[15] J. N. Warfield, Societal Systems Planning(1976). Policy and Complexity, Wiley.

[16] J. N. Warfield, Interpretive Structural Modeling (ISM). In S. A. Olsen (Eds.), Group Planning & Problem Solving Methods in Engineering (1982).pp.155-201, Wiley,.

[17] L. A. Zadeh, Information and Control (1965). Vol. 8, p.338.

[18] Y. H. Lin, M. W. Bart, and K. J. Huang, Generalized Polytomous Ordering Theory(2006). [manual and software], National Taichung University, Taiwan.

[19] T. Sato, Introduction to S-P Curve Theory Analysis and Evaluation (1985). Tokyo, Meiji Tosho.

[20] Y. H. Lin, W. M. Bart and K. J. Huang: WPIRS Software [Manual and Sftware for Gneralized Soring of Item Relational Structure] (2006).

[21] Y. H. Lin, M. W. Bart, and K. J. Huang, Generalized Polytomous Ordering Theory [manual and software], National Taichung University, Taiwan, 2006.

[22] L. F. Blixt, and T.E. Dinero, An Initial Book Look at The Validity of Diagnoses Based on Sato's Caution Index, Education and Psychological Measurement, Vol. 45, 1985, pp. 55-61.T.

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

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

[25] H.-C. Liu, B.-C. Jeng, J.-M. Yih, and Y.-K. Yu, Fuzzy C-means algorithm based on standard mahalanobis distances. Proceedings of the 2009 International Symposium on Information Processing (ISIP'09), ISBN 928-952-5726-02-2.(2009). pp.422-427.

[26] G. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic, Theory and Applications(1995). Prentice Hall.

[27] D. W. Massaro and D. Friedman, Psychological Review(1990). Vol. 97, p.225 .

[28] C. S. Crowther, W. H. Batchelder and X. Hu, Psychological Review (1995). Vol.102, p.396.

[29] Y. H. Lin, and S. M. Chen, The Integrated Analysis of S-P Chart and Ordering Theory on Equality Axiom Concepts Test for Sixth Graders, WSEAS Transactions on Mathematics, Vol. 5, 2006, pp. 1303-1308.

[30] A. Piu and C. Fregola: Simulation and Gaming for Mathematical Education: Epistemology and Teaching Strategies. Hershey, PA: IGI Global. (2010).

[31] K. K. Tatsuoka, and M. M. Tatsuoka, Computerized Cognitive Diagnostic Adaptive Testing: Effect on Remedial Instruction as Empirical Validation, Journal of Educational Measurement(1997).Vol. 34, , pp. 3-20.

[32] J. C. Falmagne: Journal of Mathematical Psychology Vol. 37 (1993) pp. 489-512.

[33] J.- M.Yih, Y.-H Lin, & W.- L Hung,.(2007). Fuzzy Approach Method for Concept Structure Analysis based on FLMP and ISM with Application in Cognition Diagnosis of Linear Algebra. Information Sciences 2007, pp.1356-1362.