Fuzzy Partition Clustering Algorithms Based on Alternative Mahalanobis Distances

Hsiang-Chuan Liu

Department of Bioinformatics & Psychology,
Asia University

Abstract

The well-known fuzzy partition clustering algorithms are mainly based on Euclidean distance measure for partitioning, which can only be used for the clusters in the data set with the same super-spherical shape distribution. Instead of using Euclidean distance measure, Gustafson & Kessel (1979) proposed the G-K algorithm which employs the Mahalanobis distance. It is a fuzzy partition clustering algorithm which can be used for the clusters in the data set with different geometrical shapes. However, without the prior information of the shape volume for each class, the G-K algorithm can only be utilized for the clusters with the same volume in the data set. In other words, if any dimension of a class is greater than the number of samples in the class, the estimated covariance matrix of that class may not be fully ranked. Hence, the algorithm will induce the singular problem for the inverse covariance matrix. This is an important issue need be addressed when we use the G-K algorithm for clustering. To overcome the issues, a new solution is proposed. A regulating factor of the covariance matrix for each class and the alternative global scatter matrix are added in the objective function, besides, the constraint of the determinant of the covariance matrices used in the G-K algorithm is removed. This new proposed algorithm is called Liu-algorithm. Based on the proposed Liu-algorithm, three well known fuzzy partition clustering algorithms using Euclidean distance measure; the Fuzzy C-Means (FCM), the Possibility C-Means (PCM), and the Fuzzy Possibility C-Means (FPCM), are extended by using the local and global Mahalanobis distance. They will be called the Fuzzy C-Means based on Alternative Mahalanobis distances (FCM-AM), the Possibility C-Means based on Alternative Mahalanobis distances (PCM-AM), the Fuzzy Possibility C-Means based on Alternative Mahalanobis distances (FPCM-AM), respectively.

Keywords:G-K algorithm, Liu-algorithm, FCM-AM, PCM-AM, FPCM-AM

基於選擇性馬氏距離之模糊分割聚類演算法

劉湘川

亞洲大學生物資訊系暨心理系

摘要

眾所周知之模糊分割聚類演算法大都為基於歐式距離之方法，只能辨識同為超球形之數據類，Gustafson & Kessel (1979) 推廣歐式距離至馬氏距離，提出 G-K 演算法是基於馬氏距離之模糊分割聚類演算法，兼可搜索超橢球形數據類等，但是各聚類之超體積須有先驗訊息，否則只能考慮各聚類之超體積為相同之情況，且當任一聚類之維度大於該聚類樣本點數時，該聚類之估計模糊共變數矩陣可能為非滿秩，其逆矩陣會產生奇異值問題，也是 G-K 演算法所未考慮之議題，本文針對上述兩種缺失，在目標函數中引進各聚類共變數矩陣之調整因子及一選擇性全域分散矩陣，刪除 G-K 演算法之共變數矩陣行列式之限制條件，並採用特徵向量減維法，而得改善之推廣新演算法，簡稱「Liu 演算法」，進而應用「Liu 演算法」，將基於歐式距離之三種常用模糊分割聚類演算法；「模糊 C 平均法（FCM）」，「可能性 C 平均法（PCM）」，及「模糊可能性 C 平均法（FPCM）」，分別推廣至基於不同選擇性馬式距離之三種模糊分割聚類演算法，對應簡記為「選擇性馬氏模糊 C 平均法（FCM-AM）」，「選擇性馬氏可能性 C 平均法（PCM-AM）」，「選擇性馬氏模糊可能性 C 平均法（FPCM-AM）」。

關鍵字：G-K 演算法、Liu 演算法、FCM-AM、PCM-AM、FPCM-AM

I、Introduction

Data 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 the high similarity, but are very dissimilar to those data objects in other clusters. Fuzzy partition clustering is a branch in cluster analysis and it is widely used in pattern recognition. Among many well- known fuzzy partition clustering algorithms, Bezdek's Fuzzy C-Means (FCM) (Bezdek, 1981), Pal, Pal, & Bezdek's Possibility C-Means (PCM) (Pal, Pal, & Bezdek, 1993), and Pal, Pal, & Bezdek's Fuzzy Possibility C-Means (FPCM) (Pal, Pal, & Bezdek, 1997) are all based on Euclidean distance measure for clustering. Hence, those fuzzy partition clustering algorithms can only be used for the data set with the same super spherical shape for each class.

Instead of using Euclidean distance measure, Gustafson & Kessel (1979) proposed the G-K algorithm which employs the Mahalanobis distance. It is a fuzzy partition clustering algorithm which can be used for the classes with different geometrical shapes in the data set. However, without the prior information of the shape volume for each class, the G-K algorithm can only be utilized for the classes with the same volume. In other words, if any dimension of a class is greater than the number of samples in the class, the estimated covariance matrix of the class may not be fully ranked. Hence, the algorithm will induce the singular problem for the inverse covariance matrix. This is an important issue need be addressed when we use the G-K algorithm for clustering.

To overcome the issues, a new solution is proposed. A regulating factor of the covariance matrix is added to each class in the objective function, and the constraint of the determinant of the covariance matrices defined in the G-K algorithm is removed. This new proposed algorithm is called Liu- algorithm. Based on the Liu-algorithm, three well known fuzzy partition clustering algorithms based on Euclidean distance; FCM, PCM, and FPCM, are extended by using the local and global Mahalanobis distances. They will be called the Fuzzy C-Means based on Alternative Mahalanobis distances (FCM-AM) algorithms, the Possibility C-Means based on Alternative Mahalanobis distances (PCM-AM) algorithms, the Fuzzy Possibility C-Means based on Alternative Mahalanobis distances (FPCM-AM) algorithms, respectively. Furthermore, the FCM-AM algorithms included two algorithms, FCM-M and FCM-CM, proposed by our previous works (Liu, H. C., Yih J. M., & Liu S. W., 2007), the PCM-AM algorithms included two new algorithms, PCM-M and PCM-CM, the FPCM-AM algorithms included two new algorithms, FPCM-M and FPCM-CM, where FPCM-M is proposed by our previous work, (Liu, H. C., Yih J. M., Sheu T. W., & Liu, S. W., 2007).

The paper is organized as follows. In section II we review three fuzzy partition clustering algorithms; namely, FCM, PCM, and FPCM. In section III the G-K algorithm is briefly described. Section IV introduces the proposed Liu-algorithm and other three extended algorithms; FCM-AM, PCM-AM, and FPCM-AM. Then the conclusion follows.

II、Fuzzy Partition Clustering Algorithms Based on Euclidean Distance

A、Background

We assume that the data set has n data objects $x = \{\underline{x}_1, \underline{x}_2, ..., \underline{x}_n\}$. Each data object has p features which are represented by a p-dimensional feature vector, $\underline{x}_k = \{x_{k1}, x_{k2}, ..., x_{kp}\}' \in R^p$, so the data set can be written as an $n \times p$ matrix as following,

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1P} \\ x_{21} & x_{22} & \dots & x_{2P} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nP} \end{bmatrix} \tag{1}$$

A fuzzy clustering algorithm will partition the data set into c fuzzy clusters with $c < n$. The fuzzy partition is represented as a fuzzy membership matrix U, each element $\mu_{ij} \in [0,1]$ in matrix U denotes the membership degree assigned for cluster C_i to each data object $\underline{x}_j$. Then each cluster is associated with a p-dimensional prototype vector $\underline{a}_i = (a_{i1}, a_{i2}, ..., a_{ip})' \in R^p$ in matrix A which denotes the center of the cluster i.

The objective of a fuzzy clustering algorithm is to partition the data into clusters so that the similarity of data objects within each cluster is maximized and the similarity of data objects among clusters is minimized. In the objective function based methods, the objective function is a function of data matrix, membership matrix and prototypes of clusters. It measures the overall dissimilarity of data objects within each cluster. Hence, by minimizing the objective function, we can obtain the best partition of the data set.

B、The Fuzzy C-Means Algorithm

The Fuzzy C-Means algorithm is the most popular objective function based fuzzy clustering algorithm. The FCM was first developed by Dunn (1973) and improved by Bezdek (1981) . The objective function used in FCM is given by Equation (2)

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

where $\mu_{ij} \in [0,1]$ is the membership degree of data object $\underline{x}_j$ in cluster C_i, and it satisfies the following constraint given by Equation (3)

$$\sum_{i=1}^{c} \mu_{ij} = 1, \forall j = 1, 2, ..., n \tag{3}$$

C is the number of clusters, m is the fuzzifier, $m > 1$, which controls the fuzziness of the algorithm. These two parameters need be specified before running the algorithm. The measure $d_{ij}^2 = \left\| \underline{x}_j - \underline{a}_i \right\|^2$ is the square Euclidean distance between data object $\underline{x}_j$ to center $\underline{a}_i$.

Minimizing the objective function (2) with the constraint (3) 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 alternative optimization scheme, which optimizes 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 shown in equations (4), (5), and (6). The algorithm is described in the following.

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:

$$U^{(0)} = \begin{bmatrix} \mu_{11}^{(0)} & \mu_{12}^{(0)} & \cdots & \mu_{1n}^{(0)} \\ \mu_{21}^{(0)} & \mu_{22}^{(0)} & \cdots & \mu_{2n}^{(0)} \\ \cdots & \cdots & & \cdots \\ \mu_{c1}^{(0)} & \mu_{c2}^{(0)} & \cdots & \mu_{cn}^{(0)} \end{bmatrix} \tag{4}$$

Step 2: To calculate

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m \underline{x}_j}{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m} \quad i = 1, 2, \ldots, c \tag{5}$$

$$\mu_{ij}^{(k)} = \left[\sum_{l=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' \left(\underline{x}_j - \underline{a}_i^{(k)} \right)}{\left(\underline{x}_j - \underline{a}_l^{(k)} \right)' \left(\underline{x}_j - \underline{a}_l^{(k)} \right)} \right]^{\frac{1}{m-1}} \right]^{-1} \tag{6}$$

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

C、The Possibility C-Means Algorithm

While the Fuzzy C-Means algorithm is sensitive to the outlier and noise, an improved fuzzy partition clustering algorithms based on Euclidean distance, called Possibility C-Means, was proposed by Pal, Pal, & Bezdek (1993). The objective function used in PCM is given by Equation (7)

$$J_{PCM}^{m}\left(T, A, \Sigma\right) = \sum_{i=1}^{c} \sum_{j=1}^{n} t_{ij}^{m} \left\| \underline{x}_j - \underline{a}_i \right\|^2 + \sum_{i=1}^{c} w_i \sum_{j=1}^{n} \left(1 - t_{ij}\right)^m \tag{7}$$

where $t_{ij} \ge 0$ is the possibility (typicality) degree of data object $\underline{x}_j$ in cluster C_i, and there is no constraint like Equation (3)

$\underline{w} = \left(w_1, w_2, \ldots, w_c \right)'$, $w_i \in R^+$ are the weight of cluster i.

Unconstrained optimization of the first term in Equation (2-6) will lead to the trivial solution.

$$t_{ij} = 0 \text{ , } \forall i, j$$

The second term in Equation (7) acts as a penalty which tries to bring typicality values towards 1. The algorithm is described in the following.

Step 1: Determining the number of cluster, c, m-value (let $m = 3$), and the converging error, $\varepsilon > 0$ (such as $\varepsilon = 0.001$) and choosing the initial typicality matrix:

$$T^{(0)} = \begin{bmatrix} t_{11}^{(0)} & t_{12}^{(0)} & \dots & t_{1n}^{(0)} \\ t_{21}^{(0)} & t_{22}^{(0)} & \dots & t_{2n}^{(0)} \\ \dots & \dots & & \dots \\ t_{c1}^{(0)} & t_{c2}^{(0)} & \dots & t_{cn}^{(0)} \end{bmatrix} \tag{8}$$

Step 2: To calculate

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n} \left[t_{ij}^{(k-1)} \right]^m \underline{x}_j}{\sum_{j=1}^{n} \left[t_{ij}^{(k-1)} \right]^m} \quad i = 1, 2, \dots, c \tag{9}$$

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

$$t_{ij}^{(k)} = \left[\frac{w_i^{(k)}}{w_i^{(k)} + \left(\underline{x}_j - \underline{a}_i^{(k)} \right)' \left(\underline{x}_j - \underline{a}_i^{(k)} \right)} \right]^{\frac{1}{m-1}} \tag{11}$$

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

D、The Fuzzy Possibility C-Means Algorithm

Combining the FCM and PCM, the new fuzzy partition clustering algorithms based on Euclidean distance, called Fuzzy Possibility C-Means, was proposed by Pal, Pal, & Bezdek (1997)

$$J_{FPCM}^{m}\left(U, T, A, X\right) = \sum_{i=1}^{c} \sum_{j=1}^{n} \left(\mu_{ij}^{m} + t_{ij}^{\delta} \right) \left\| \underline{x}_j - \underline{a}_i \right\|^2 \tag{12}$$

Constraints: membership $\sum_{i=1}^{c} \mu_{ij} = 1$, $\forall j = 1, 2, \dots, n$, (13)

typicality $\sum_{j=1}^{n} t_{ij} = 1$, $\forall i = 1, 2, \dots, c$ (14)

The algorithm is described below.

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

$$U^{(0)} = \begin{bmatrix} \mu_{11}^{(0)} \mu_{12}^{(0)} \dots \mu_{1n}^{(0)} \\ \mu_{21}^{(0)} \mu_{22}^{(0)} \dots \mu_{2n}^{(0)} \\ \dots \quad \dots \quad \dots \\ \mu_{c1}^{(0)} \mu_{c2}^{(0)} \dots \mu_{cn}^{(0)} \end{bmatrix}, \; T^{(0)} = \begin{bmatrix} t_{11}^{(0)} t_{12}^{(0)} \dots t_{1n}^{(0)} \\ t_{21}^{(0)} t_{22}^{(0)} \dots t_{2n}^{(0)} \\ \dots \quad \dots \quad \dots \\ t_{c1}^{(0)} t_{c2}^{(0)} \dots t_{cn}^{(0)} \end{bmatrix} \tag{15}$$

Step 2: To calculate

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k)} \right]^m + \left[t_{ij}^{(k)} \right]^\delta \right) \underline{x}_j}{\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k)} \right]^m + \left[t_{ij}^{(k)} \right]^\delta \right)}, \tag{16}$$

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

$$\mu_{ij}^{(k)} = \left[\sum_{l=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' \left(\underline{x}_j - \underline{a}_i^{(k)} \right)}{\left(\underline{x}_j - \underline{a}_l^{(k)} \right)' \left(\underline{x}_j - \underline{a}_l^{(k)} \right)} \right]^{\frac{1}{m-1}} \right]^{-1}, \tag{17}$$

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

$$t_{ij}^{(k)} = \left[\sum_{l=1}^{n} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' \left(\underline{x}_j - \underline{a}_i^{(k)} \right)}{\left(\underline{x}_j - \underline{a}_l^{(k)} \right)' \left(\underline{x}_j - \underline{a}_l^{(k)} \right)} \right]^{\frac{1}{\delta-1}} \right]^{-1}, \tag{18}$$

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

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

III、The G-K Algorithm

The well-known 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:

$$J_{G-K}^m \left(U, A, M, X \right) = \sum_{i=1}^{c} \sum_{j=1}^{n} \left(\mu_{ij}^m \right) \left(\underline{x}_j - \underline{a}_i \right)' M_i \left(\underline{x}_j - \underline{a}_i \right) \tag{19}$$

constraints：membership $\sum_{i=1}^{c} \mu_{ij} = 1, \forall j = 1,2,\dots,n$, (20)

determinent of sandardization covariance matrix of cluster i

$$\left| M_i \right| = \rho_i, \; \forall i = 1,2,\dots,c \tag{21}$$

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

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:

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

Step 2: To calculate

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m \underline{x}_j}{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)} \right]^m} \quad i = 1, 2, \ldots, c \tag{23}$$

$$F_i^{(k)} = \frac{\sum_{j=1}^{n} \left(\mu_{ij}^{(k-1)} \right)^m \left(\underline{x}_j - \underline{a}_i^{(k)} \right) \left(\underline{x}_j - \underline{a}_i^{(k)} \right)'}{\sum_{j=1}^{n} \left(\mu_{ij}^{(k)} \right)^m} \tag{24}$$

$$M_i^{(k)} = \left[\rho_i \det\left(F_i^{(k)} \right) \right]^{\frac{1}{p}} \left(F_i^{(k)} \right)^{-1} \tag{25}$$

$$\mu_{ij}^{(k)} = \left[\sum_{l=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)} \right)' M_i^{(k)} \left(\underline{x}_j - \underline{a}_i^{(k)} \right)}{\left(\underline{x}_j - \underline{a}_l^{(k)} \right)' M_i^{(k)} \left(\underline{x}_j - \underline{a}_l^{(k)} \right)} \right]^{\frac{1}{m-1}} \right]^{-1} \tag{26}$$

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

IV、The Proposed Liu-Algorithm and Three Extended Algorithms

A、Liu-Algorithm

The G-K algorithm must have some prior information of shape volume for each data class, $|M_i| = \rho_i$, $\forall i = 1, 2, ..., c$. Otherwise, let $\rho_i = 1$, $\forall i = 1, 2, ..., c$ and it can only be considered to cluster the data classes with the same volume 1. On the other hand, when any dimension of a class is greater than that of the sample size of that class, the estimated covariance matrix of the class may not be fully ranked; it will induce the singular problem of the inverse covariance matrix. This is an important issue need be addressed when we use the G-K algorithm for clustering.

Now, we add a regulating factor of covariance matrix, $-\ln\left|{}_{+}\Sigma_i^{-1}\right|$, to each class and a global scatter matrix in the objective function, and delete the constraint of the determinant of covariance matrices, $|M_i| = \rho_i$, in the G-K algorithm. This new proposed algorithm, called the Liu- algorithm. Based on this new algorithm, we can obtain three kinds of new fuzzy partition clustering algorithms based on alternative Mahalanobis distances; FCM-AM (including FCM-M and FCM-CM) , PCM-AM (in-

cluding PCM-M and PCM-CM) and FPCM-CM (including FPCM-M and FPCM-CM). These new algorithms are derived in the following.

B、The FCM-AM Algorithm and Its Special Cases

Using the Liu-algorithm, we can obtain the objective function of FCM-AM as following

$$J_{FCM-CM}^{m}\left(U,A,\Sigma,X\right)=$$
$$\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right] \tag{27}$$

constraints：membership, $\sum_{i=1}^{c}\mu_{ij}=1$, $\forall j=1,2,\ldots,n$, (28)

where $h=0,1$

$\Sigma=\left\{\Sigma_1,\Sigma_2,\ldots,\Sigma_c\right\}$ is the set of covariance of clusters.

$$\underline{a}_t=\frac{1}{n}\sum_{i=1}^{n}\underline{x}_i,\ \Sigma_t=\frac{1}{n}\sum_{i=1}^{n}\left(\underline{x}_i-\underline{a}_t\right)\left(\underline{x}_i-\underline{a}_t\right)' \tag{29}$$

1. If $h=0$ and $\Sigma_1=\Sigma_2=\ldots=\Sigma_c=\mathrm{I}$ then FCM-AM algorithm is just the well known FCM algorithm.
2. If $h=0$ but $\Sigma_1,\Sigma_2,\ldots,\Sigma_c$ not all equal to I, then FCM-AM algorithm is called FCM-M algorithm (Liu, H. C., et al, 2007).
3. If $h=1$ and $\Sigma_1,\Sigma_2,\ldots,\Sigma_c$ not all equal to I, then FCM-AM algorithm is called FCM-CM algorithm (Liu, H. C., et al, 2008).

Let

$$J=J_{FCM-CM}^{m}\left(U,A,\Sigma,X\right)=$$
$$\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$
$$-\sum_{j=1}^{n}\alpha_j\left(\sum_{i=1}^{c}\mu_{ij}-1\right) \tag{30}$$

$$\frac{\partial J}{\partial\underline{a}_i}=\frac{\partial}{\partial\underline{a}_i}\sum_{j=1}^{n}\mu_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$
$$=-2\sum_{j=1}^{n}\mu_{ij}^{m}\left[\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)+h\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$
$$=-2\sum_{j=1}^{n}\mu_{ij}^{m}\left[\left(\Sigma_i^{-1}\underline{x}_j-h\Sigma_t^{-1}\underline{a}_t\right)-\left(\Sigma_i^{-1}-h\Sigma_t^{-1}\right)\underline{a}_i\right]=0 \tag{31}$$

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

$$\frac{\partial J}{\partial \mu_{ij}} = \frac{\partial}{\partial \mu_{ij}}\left[\mu_{ij}^{m}\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)-\alpha_j\mu_{ij}\right]$$

$$= m\mu_{ij}^{m-1}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]-\alpha_j \triangleq 0$$

$$\Rightarrow \alpha_j = m\mu_{ij}^{m-1}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$

$$\Rightarrow \mu_{ij} = \alpha_j^{\frac{1}{m-1}}\left[m\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]\right]^{\frac{-1}{m-1}}$$

$$\frac{\partial J}{\partial \alpha_j} = \left(\sum_{i=1}^{c}\mu_{ij}-1\right) \triangleq 0 \Rightarrow \sum_{i=1}^{c}\mu_{ij} = 1,\ \forall j = 1,2,\ldots,n$$

$$\Rightarrow \sum_{i=1}^{c}\mu_{ij} = 1 = \alpha_j^{\frac{1}{m-1}}\sum_{i=1}^{c}\left[m\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]\right]^{\frac{-1}{m-1}}$$

$$\Rightarrow \alpha_j^{\frac{-1}{m-1}} = \sum_{i=1}^{c}\left[m\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]\right]^{\frac{-1}{m-1}}$$

$$, j = 1,2,\ldots,n$$

$$\mu_{ij} = \frac{\left[m\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]\right]^{\frac{-1}{m-1}}}{\alpha_j^{\frac{-1}{m-1}}}$$

$$\Rightarrow \mu_{ij} = \left[\sum_{s=1}^{c}\left[\frac{\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)}{\left(\underline{x}_j-\underline{a}_s\right)'\Sigma_s^{-1}\left(\underline{x}_j-\underline{a}_s\right)-\ln\left|\Sigma_s^{-1}\right|-h\left(\underline{a}_s-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_s-\underline{a}_t\right)}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{32}$$

Using the Theorem $\frac{\partial |X|}{\partial X} = |X|\left(X^{-1}\right)'$, $\frac{\partial \underline{a}'X\underline{a}}{\partial X} = \underline{a}\,\underline{a}'$, we get

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

$$= \sum_{j=1}^{n}\mu_{ij}^{m}\left(\underline{x}_j-\underline{a}_i\right)\left(\underline{x}_j-\underline{a}_i\right)'-\sum_{j=1}^{n}\mu_{ij}^{m}\frac{1}{\left|\Sigma_i^{-1}\right|}\left|\Sigma_i^{-1}\right|\Sigma_i' \triangleq 0 \tag{33}$$

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

$$\Sigma_i = \sum_{s=1}^{p} \lambda_{si} \Gamma_{si} \Gamma'_{si} \quad i = 1, 2, \ldots, c \frac{n!}{r!(n-r)!}$$
$$\Rightarrow \Sigma_i^{-1} = \sum_{s=1}^{p} \left(\lambda_{si}^{-1}\right)^{+} \Gamma_{si} \Gamma'_{si}, \left(\lambda_{si}^{-1}\right)^{+} = \begin{cases} \lambda_{si}^{-1} & if\ \lambda_{si} > 0 \\ 0 & if\ \lambda_{si} \le 0 \end{cases} \tag{34}$$

The algorithm is described below.

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:

$$U^{(0)} = \begin{bmatrix} \mu_{11}^{(0)} & \mu_{12}^{(0)} & \ldots & \mu_{1n}^{(0)} \\ \mu_{21}^{(0)} & \mu_{22}^{(0)} & \ldots & \mu_{2n}^{(0)} \\ \ldots & \ldots & & \ldots \\ \mu_{c1}^{(0)} & \mu_{c2}^{(0)} & \ldots & \mu_{cn}^{(0)} \end{bmatrix} \tag{35}$$

Step 2: To calculate

$$\underline{a}_i^{(k)} = \left[\sum_{j=1}^{n} \left(\mu_{ij}^{(k-1)}\right)^m \left(\left(\Sigma_i^{(k-1)}\right)^{-1} - h\Sigma_t^{-1}\right)\right]^{-1} \left[\sum_{j=1}^{n} \left(\mu_{ij}^{(k-1)}\right)^m \left(\left(\Sigma_i^{(k-1)}\right)^{-1} \underline{x}_j - h\Sigma_t^{-1} \underline{a}_t\right)\right] \tag{36}$$

$$\Sigma_i^{(k)} = \frac{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)}\right]^m \left(\underline{x}_j - \underline{a}_i^{(k)}\right)\left(\underline{x}_j - \underline{a}_i^{(k)}\right)'}{\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)}\right]^m} \tag{37}$$

$$\mu_{ij}^{(k)} = \left[\sum_{s=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)}\right)' \left(\Sigma_i^{(k)}\right)^{-1} \left(\underline{x}_j - \underline{a}_i^{(k)}\right) - \ln\left|\left(\Sigma_i^{(k)}\right)^{-1}\right| - h\left(\underline{a}_i^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i^{(k)} - \underline{a}_t\right)}{\left(\underline{x}_j - \underline{a}_s^{(k)}\right)' \left(\Sigma_s^{(k)}\right)^{-1} \left(\underline{x}_j - \underline{a}_s^{(k)}\right) - \ln\left|\left(\Sigma_s^{(k)}\right)^{-1}\right| - h\left(\underline{a}_s^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_s^{(k)} - \underline{a}_t\right)}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{38}$$

Step 3: Increment k until $\max_{1 \le i \le c} \left\|\underline{a}_i^{(k)} - \underline{a}_i^{(k-1)}\right\| < \varepsilon$ (39)

C、The PCM-AM Algorithm and Its Special Cases

Using the Liu algorithm, we can obtain the objective function of PCM-AM as following

$$J = J_{pcm-m}^{m}(T, A, \Sigma, X)$$
$$= \sum_{i=1}^{c} \sum_{j=1}^{n} t_{ij}^m \left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right)\right] \tag{40}$$
$$+ \sum_{i=1}^{c} w_i \sum_{j=1}^{n} \left(1 - t_{ij}\right)^m$$

where $h = 0, 1$

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

$$\underline{a}_t = \frac{1}{n} \sum_{i=1}^{n} \underline{x}_i, \Sigma_t = \frac{1}{n} \sum_{i=1}^{n} \left(\underline{x}_i - \underline{a}_t\right)\left(\underline{x}_i - \underline{a}_t\right)' \tag{41}$$

Unconstrained optimization of the first term in the objective function J;

$$\sum_{i=1}^{c}\sum_{j=1}^{n} t_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right] \quad ,$$

will lead to the trivial solution, $t_{ij}=0, \forall i=1,2,...,c, \forall j=1,2,...,n,$

$$\text{Let } w_i=\frac{\sum_{j=1}^{n} t_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]}{\sum_{j=1}^{n} t_{ij}^{m}} \tag{42}$$

be proportion to the average within and between clusters fluctuation.

The second term, $\sum_{i=1}^{c} w_i \sum_{j=1}^{n}\left(1-t_{ij}\right)^m$, acts as a penalty which tries to bring typicality value towards 1.

1. If $h = 0$ and $\Sigma_1=\Sigma_2=\ldots=\Sigma_c=\mathrm{I}$ then PCM-AM algorithm is just the well known PCM algorithm.
2. If $h = 0$ but $\Sigma_1, \Sigma_2, \ldots, \Sigma_c$ not all equal to I, then PCM-AM algorithm is called PCM-M algorithm.
3. If $h = 1$ and $\Sigma_1, \Sigma_2, \ldots, \Sigma_c$ not all equal to I, then PCM-AM algorithm is called PCM-CM algorithm.

$$\begin{aligned}
\frac{\partial J}{\partial \underline{a}_i} &= \frac{\partial}{\partial \underline{a}_i}\sum_{j=1}^{n} t_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]\\
&= -2\sum_{j=1}^{n} t_{ij}^{m}\left[\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)+h\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]\\
&= -2\sum_{j=1}^{n} t_{ij}^{m}\left[-\left[\Sigma_i^{-1}-h\Sigma_t^{-1}\right]\underline{a}_i+\left[\Sigma_i^{-1}\underline{x}_j-h\Sigma_t^{-1}\underline{a}_t\right]\right]
\end{aligned} \tag{43}$$

$$\sum_{j=1}^{n} t_{ij}^{m}\left[\Sigma_i^{-1}-h\Sigma_t^{-1}\right]\underline{a}_i=\sum_{j=1}^{n} t_{ij}^{m}\left[\Sigma_i^{-1}\underline{x}_j-h\Sigma_t^{-1}\underline{a}_t\right]$$

$$\Rightarrow \underline{a}_i=\left[\sum_{j=1}^{n} t_{ij}^{m}\left[\Sigma_i^{-1}-h\Sigma_t^{-1}\right]\right]^{-1}\sum_{j=1}^{n} t_{ij}^{m}\left[\Sigma_i^{-1}\underline{x}_j-h\Sigma_t^{-1}\underline{a}_t\right]$$

$$\begin{aligned}
\frac{\partial J}{\partial t_{ij}} &= \frac{\partial}{\partial t_{ij}}\left[t_{ij}^{m}\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)+w_i\left(1-t_{ij}\right)^m\right]\\
&= m t_{ij}^{m-1}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]-w_i m\left(1-t_{ij}\right)^{m-1}\triangleq 0
\end{aligned}$$

$$\Rightarrow w_i=\left(\frac{t_{ij}}{1-t_{ij}}\right)^{m-1}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$

$$t_{ij}=\left[\frac{w_i}{w_i+\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)}\right]^{\frac{1}{m-1}} \tag{44}$$

Using the Theorem $\frac{\partial |X|}{\partial X} = |X|\left(X^{-1}\right)'$, $\frac{\partial \underline{a}'X\underline{a}}{\partial X} = \underline{a}\,\underline{a}'$, we get

$$\frac{\partial J}{\partial \Sigma_i^{-1}} = \frac{\partial}{\partial \Sigma_i^{-1}} \sum_{j=1}^{n} t_{ij}^m \left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| \right]$$

$$= \sum_{j=1}^{n} t_{ij}^m \left[\left(\underline{x}_j - \underline{a}_i\right)\left(\underline{x}_j - \underline{a}_i\right)' - \frac{1}{\left|\Sigma_i^{-1}\right|}\left|\Sigma_i^{-1}\right|\Sigma_i' \right] \triangleq 0 \tag{45}$$

$$\Rightarrow \Sigma_i = \frac{\sum_{j=1}^{n} t_{ij}^m \left(\underline{x}_j - \underline{a}_i\right)\left(\underline{x}_j - \underline{a}_i\right)'}{\sum_{j=1}^{n} t_{ij}^m}$$

$$\Sigma_i = \sum_{s=1}^{p} \lambda_{si} \Gamma_{si} \Gamma_{si}' \qquad i = 1, 2, \ldots, c$$

$$\Rightarrow \Sigma_i^{-1} = \sum_{s=1}^{p} \left(\lambda_{si}^{-1}\right)^+ \Gamma_{si} \Gamma_{si}' \ , \left(\lambda_{si}^{-1}\right)^+ = \begin{cases} \lambda_{si}^{-1} & if\ \lambda_{si} > 0 \\ 0 & if\ \lambda_{si} \le 0 \end{cases} \tag{46}$$

The algorithm is described below.

Step 1: Determining the number of cluster c, m-value (let $m = 3$), and the converging error, $\varepsilon > 0$ (such as $\varepsilon = 0.001$) , and choosing the initial typicality matrix:

$$t^{(0)} = \begin{bmatrix} t_{11}^{(0)} & t_{12}^{(0)} & \ldots & t_{1n}^{(0)} \\ t_{21}^{(0)} & t_{22}^{(0)} & \ldots & t_{2n}^{(0)} \\ \ldots & \ldots & & \ldots \\ t_{c1}^{(0)} & t_{c2}^{(0)} & \ldots & t_{cn}^{(0)} \end{bmatrix} \tag{47}$$

Step 2: To calculate

$$\underline{a}_i = \left[\sum_{j=1}^{n} t_{ij}^m \left[\Sigma_i^{-1} - h\Sigma_t^{-1} \right] \right]^{-1} \sum_{j=1}^{n} t_{ij}^m \left[\Sigma_i^{-1} \underline{x}_j - h\Sigma_t^{-1} \underline{a}_t \right] \tag{48}$$

$$\Sigma_i^{(k)} = \frac{\sum_{j=1}^{n} \left[t_{ij}^{(k-1)} \right]^m \left(\underline{x}_j - \underline{a}_i^{(k)}\right)\left(\underline{x}_j - \underline{a}_i^{(k)}\right)'}{\sum_{j=1}^{n} \left[t_{ij}^{(k-1)} \right]^m} \tag{49}$$

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

$$\left[\lambda_{si}^{(-1)}\right]^{(k)} = \begin{cases} \left[\lambda_{si}^{(k)}\right]^{-1} & if \ \lambda_{si}^{(k)} > 0 \\ 0 & if \ \lambda_{si}^{(k)} = 0 \end{cases} \tag{50}$$

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

$$w_i^{(k)} = \frac{\sum_{j=1}^{n} \left[t_{ij}^{(k-1)}\right]^m \left[\left(\underline{x}_j - \underline{a}_i^{(k)}\right)' \left[\Sigma_i^{-1}\right]^{(k)} \left(\underline{x}_j - \underline{a}_i^{(k)}\right) - h\left(\underline{a}_i^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i^{(k)} - \underline{a}_t\right)\right]}{\sum_{j=1}^{n} \left[t_{ij}^{(k-1)}\right]^m} \tag{51}$$

$$t_{ij}^{(k)} = \left[\frac{w_i^{(k)}}{w_i^{(k)} + \left(\underline{x}_j - \underline{a}_i^{(k)}\right)' \left[\Sigma_i^{-1}\right]^{(k)} \left(\underline{x}_j - \underline{a}_i^{(k)}\right) - \ln\left|\left[\Sigma_i^{-1}\right]^{(k)} - h\left(\underline{a}_i^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i^{(k)} - \underline{a}_t\right)\right|}\right]^{\frac{1}{m-1}} \tag{52}$$

Step 3: Increment k until $\max_{1 \le i \le c} \left\| \underline{a}_i^{(k)} - \underline{a}_i^{(k-1)} \right\| < \varepsilon$. (53)

D、The FPCM-AM Algorithm and Its Special Cases

Using the Liu-algorithm, we can obtain the objective function of FPCM-AM as following

$$J_{FPCM-M}^{m}\left(U,T,\Sigma,X\right) = \sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m} + t_{ij}^{\delta}\right)\left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1}\left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1}\left(\underline{a}_i - \underline{a}_t\right)\right] \tag{54}$$

constraints：membership $\sum_{i=1}^{c} \mu_{ij} = 1$, $\forall j = 1, 2, ..., n$, (55)

typicality $\sum_{j=1}^{n} t_{ij} = 1$, $\forall i = 1, 2, ..., c$ (56)

where $h = 0, 1$

$\Sigma = \left\{\Sigma_1, \Sigma_2, ..., \Sigma_c\right\}$ is the set of covariance of clusters.

$$\underline{a}_t = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i , \Sigma_t = \frac{1}{n}\sum_{i=1}^{n}\left(\underline{x}_i - \underline{a}_t\right)\left(\underline{x}_i - \underline{a}_t\right)' \tag{57}$$

1. If $h = 0$ and $\Sigma_1 = \Sigma_2 = ... = \Sigma_c = \mathrm{I}$ then FPCM-AM algorithm is just the well known FPCM algorithm.
2. If $h = 0$ but $\Sigma_1, \Sigma_2, ..., \Sigma_c$ not all equal to I, then FPCM-AM algorithm is called FPCM-M algorithm (Liu H. C., et al, 2007).
3. If $h = 1$ and $\Sigma_1, \Sigma_2, ..., \Sigma_c$ not all equal to I, then FPCM-AM algorithm is called FPCM-CM algorithm.

Let

$$J=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$
$$+\sum_{j=1}^{n}\alpha_j\left(1-\sum_{i=1}^{c}\mu_{ij}\right)+\sum_{i=1}^{c}\beta_i\left(1-\sum_{j=1}^{n}t_{ij}\right) \tag{58}$$

$$\frac{\partial J}{\partial \underline{a}_i}$$
$$=\frac{\partial}{\partial \underline{a}_i}\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$
$$=-2\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)+h\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$
$$=2\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\left[\Sigma_i^{-1}-h\Sigma_t^{-1}\right]\underline{a}_i-\left[\Sigma_i^{-1}\underline{x}_j-h\Sigma_t^{-1}\underline{a}_t\right]\right]$$
$$\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\Sigma_i^{-1}-h\Sigma_t^{-1}\right]\underline{a}_i=\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\Sigma_i^{-1}\underline{x}_j-h\Sigma_t^{-1}\underline{a}_t\right]$$
$$\underline{a}_i=\left[\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\Sigma_i^{-1}-h\Sigma_t^{-1}\right]\right]^{-1}\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left[\Sigma_i^{-1}\underline{x}_j-h\Sigma_t^{-1}\underline{a}_t\right] \tag{59}$$

$$\frac{\partial J}{\partial \alpha_j}=\frac{\partial}{\partial \alpha_j}\alpha_j\left(1-\sum_{i=1}^{c}\mu_{ij}\right)\triangleq 0\Rightarrow\sum_{i=1}^{c}\mu_{ij}=1 \tag{60}$$
$$j=1,2,\ldots,n$$

$$\frac{\partial J}{\partial \beta_i}=\frac{\partial}{\partial \beta_i}\beta_i\left(1-\sum_{j=1}^{n}t_{ij}\right)\triangleq 0\Rightarrow\sum_{j=1}^{n}t_{ij}=1 \tag{61}$$
$$i=1,2,\ldots,c$$

$$\frac{\partial J}{\partial \mu_{ij}}=\frac{\partial}{\partial \mu_{ij}}\left[\mu_{ij}^{m}\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)-\alpha_j\mu_{ij}\right]$$
$$=m\mu_{ij}^{m-1}\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)-\alpha_j\triangleq 0$$
$$\Rightarrow\alpha_j=m\mu_{ij}^{m-1}\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)$$

$$\Rightarrow\mu_{ij}=\alpha_j^{\frac{1}{m-1}}\left[m\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)\right]^{\frac{-1}{m-1}}$$

$$\Rightarrow 1=\sum_{i=1}^{c}\mu_{ij}=\alpha_j^{\frac{1}{m-1}}\sum_{i=1}^{c}\left[m\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)\right]^{\frac{-1}{m-1}}$$

$$\Rightarrow\alpha_j^{\frac{-1}{m-1}}=\sum_{i=1}^{c}\left[m\left(\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-h\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right)\right]^{\frac{-1}{m-1}}$$

$$\Rightarrow \mu_{ij} = \left[\sum_{s=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right)}{\left(\underline{x}_j - \underline{a}_s\right)' \Sigma_s^{-1} \left(\underline{x}_j - \underline{a}_s\right) - \ln\left|\Sigma_s^{-1}\right| - h\left(\underline{a}_s - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_s - \underline{a}_t\right)} \right]^{\frac{1}{m-1}} \right]^{-1} \tag{62}$$

$$\frac{\partial J}{\partial t_{ij}} = \frac{\partial}{\partial t_{ij}} \left[t_{ij}^{\delta} \left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right) \right] + \sum_{i=1}^{c} \beta_i \left(1 - \sum_{j=1}^{n} t_{ij} \right) \right]$$

$$= \delta t_{ij}^{\delta-1} \left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right) \right] - \beta_i \triangleq 0$$

$$\Rightarrow \beta_i = \delta t_{ij}^{\delta-1} \left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right) \right]$$

$$t_{ij} = \delta^{\frac{-1}{\delta-1}} \left[\frac{\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right)}{\beta_i} \right]^{\frac{-1}{\delta-1}}$$

$$1 = \sum_{j=1}^{n} t_{ij} = \delta^{\frac{-1}{\delta-1}} \sum_{j=1}^{n} \left[\frac{\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right)}{\beta_i} \right]^{\frac{-1}{\delta-1}}$$

$$= \beta_i^{\frac{1}{\delta-1}} \sum_{j=1}^{n} \delta^{\frac{-1}{\delta-1}} \left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right) \right]^{\frac{-1}{\delta-1}}$$

$$\Rightarrow \beta_i^{\frac{-1}{\delta-1}} = \sum_{j=1}^{n} \delta^{\frac{-1}{\delta-1}} \left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right) \right]^{\frac{-1}{\delta-1}}$$

$$\Rightarrow t_{ij} = \left[\sum_{s=1}^{n} \left[\frac{\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right)}{\left(\underline{x}_j - \underline{a}_s\right)' \Sigma_s^{-1} \left(\underline{x}_j - \underline{a}_s\right) - \ln\left|\Sigma_s^{-1} - h\left(\underline{a}_s - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_s - \underline{a}_t\right)\right|} \right]^{\frac{1}{\delta-1}} \right]^{-1} \tag{63}$$

$$\frac{\partial J}{\partial \Sigma_i^{-1}} =$$

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

$$= \sum_{j=1}^{n} \left(\mu_{ij}^{m} + t_{ij}^{\delta} \right) \left[\left(\underline{x}_j - \underline{a}_i\right)\left(\underline{x}_j - \underline{a}_i\right)' - \frac{1}{\left|\Sigma_i^{-1}\right|} \left|\Sigma_i^{-1}\right| \Sigma_i' \right] \triangleq 0$$

$$\Rightarrow \Sigma_i = \frac{\sum_{j=1}^{n} \left(\mu_{ij}^{m} + t_{ij}^{\delta} \right)\left(\underline{x}_j - \underline{a}_i\right)\left(\underline{x}_j - \underline{a}_i\right)'}{\sum_{j=1}^{n} \left(\mu_{ij}^{m} + t_{ij}^{\delta} \right)} \tag{64}$$

$$\Sigma_i = \sum_{s=1}^{p} \lambda_{si} \Gamma_{si} \Gamma'_{si} \quad i = 1, 2, \ldots, c$$

$$\Rightarrow \Sigma_i^{-1} = \sum_{s=1}^{p} \left(\lambda_{si}^{-1}\right)^{+} \Gamma_{si} \Gamma'_{si}, \left(\lambda_{si}^{-1}\right)^{+} = \begin{cases} \lambda_{si}^{-1} & if\ \lambda_{si} > 0 \\ 0 & if\ \lambda_{si} \le 0 \end{cases}$$

The algorithm is described below.

Step 1: Determining the number of cluster c, m-value (let $m = 2$), $\delta = 3$, and the converging error $\varepsilon > 0$ (such as $\varepsilon = 0.001$), and choosing the result membership matrix, U, of the FPCM algorithm as the initial membership matrix and the result typicality matrix , T, of the FPCM algorithm as the initial typicality matrix, respectively:

$$U^{(0)} = \begin{bmatrix} \mu_{11}^{(0)} \mu_{12}^{(0)} \ldots \mu_{1n}^{(0)} \\ \mu_{21}^{(0)} \mu_{22}^{(0)} \ldots \mu_{2n}^{(0)} \\ \ldots \quad \ldots \quad \ldots \\ \mu_{c1}^{(0)} \mu_{c2}^{(0)} \ldots \mu_{cn}^{(0)} \end{bmatrix}, T^{(0)} = \begin{bmatrix} t_{11}^{(0)} t_{12}^{(0)} \ldots t_{1n}^{(0)} \\ t_{21}^{(0)} t_{22}^{(0)} \ldots t_{2n}^{(0)} \\ \ldots \ldots \quad \ldots \\ t_{c1}^{(0)} t_{c2}^{(0)} \ldots t_{cn}^{(0)} \end{bmatrix} \tag{65}$$

Step 2: To calculate:

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

$$\Sigma_i^{(k)} = \frac{\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k)}\right]^m + \left[t_{ij}^{(k)}\right]^{\delta} \right) \left(\underline{x}_j - \underline{a}_i^{(k)}\right) \left(\underline{x}_j - \underline{a}_i^{(k)}\right)'}{\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k)}\right]^m + \left[t_{ij}^{(k)}\right]^{\delta} \right)} \tag{66}$$

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

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

$$\left[\lambda_{si}^{(-1)}\right]^{(k)} = \begin{cases} \left[\lambda_{si}^{(k)}\right]^{-1} & if\ \lambda_{si}^{(k)} > 0 \\ 0 & if\ \lambda_{si}^{(k)} = 0 \end{cases} \tag{67}$$

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

$$\mu_{ij}^{(k)} = \left[\sum_{s=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)}\right)' \left[\Sigma_i^{-1}\right]^{(k)} \left(\underline{x}_j - \underline{a}_i^{(k)}\right) - \ln\left|\left[\Sigma_i^{-1}\right]^{(k)}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right)}{\left(\underline{x}_j - \underline{a}_s^{(k)}\right)' \left[\Sigma_s^{-1}\right]^{(k)} \left(\underline{x}_j - \underline{a}_s^{(k)}\right) - \ln\left|\left[\Sigma_s^{-1}\right]^{(k)}\right| - h\left(\underline{a}_s^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_s^{(k)} - \underline{a}_t\right)} \right]^{\frac{1}{m-1}} \right]^{-1} \tag{68}$$

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

$$t_{ij}^{(k)} = \left[\sum_{s=1}^{n} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)}\right)' \left[\Sigma_i^{-1}\right]^{(k)} \left(\underline{x}_j - \underline{a}_i^{(k)}\right) - \ln\left|\left[\Sigma_i^{-1}\right]^{(k)}\right| - h\left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t\right)}{\left(\underline{x}_j - \underline{a}_s^{(k)}\right)' \left[\Sigma_s^{-1}\right]^{(k)} \left(\underline{x}_j - \underline{a}_s^{(k)}\right) - \ln\left|\left[\Sigma_s^{-1}\right]^{(k)}\right| - h\left(\underline{a}_s^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_s^{(k)} - \underline{a}_t\right)} \right]^{\frac{1}{\delta-1}} \right]^{-1} \tag{69}$$

$$i = \mathbf{1, 2, ..., } c,\ j = \mathbf{1, 2, ..., } n$$

Step 3: Increment k until $\max_{1 \le i \le c} \left\| \underline{a}_i^{(k)} - \underline{a}_i^{(k-1)} \right\| < \varepsilon$. (70)

V、Conclusion

In this paper, three well-known fuzzy partition clustering algorithms based on Euclidean distance measure, FCM, PCM, and FPCM are reviewed. The well-known fuzzy partition clustering algorithms based on Mahalanobis distance proposed by Gustafson & Kessel (1979), the G-K algorithm, is also examined. We pointed out two important issues; (1) even though the G-K algorithm can be used for data clustering with different geometrical shapes associated with each class, there must have some prior information of shape volume for each class, otherwise, it can only be considered for the data set with the same volume for each class, and (2) when any dimension of a class is greater than the number of samples in the class, the estimated covariance matrix of the class may not be fully ranked. Hence, it induces the singular problem of the inverse covariance matrix. These two important issues were not considered in the G-K algorithm.

To solve these two issues, we added a regulating factor of covariance matrix to each class in the objective function, and deleted the constraint of the determinant of covariance matrices used in the G-K algorithm. This new proposed algorithm is called the L algorithm. Furthermore, based on this new algorithm, the following three well-known fuzzy partition clustering algorithms based on Euclidean distance; FCM, PCM, and FPCM can be extended to FCM-AM, PCM-AM, and FPCM-AM, respectively. The above three kinds of new fuzzy partition clustering algorithms were derived in this paper. Furthermore, each kind of above algorithms included two fuzzy partition clustering algorithms based on deferent alternative Mahalanobis distances, the FCM-AM includes FCM-M and FCM-CM, the PCM-AM includes PCM-M and PCM-CM, and FPCM-AM includes FPCM-M and FPCM-CM.

References

Bezdek, J. C. (1981) . *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, NY, 1981.

Dunn, J. C. (1973) . A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters, *J. Cybern, 3(3)*, 32-57, 1973.

Gustafson, D. E., & Kessel, W. C. (1979) . Fuzzy clustering with a fuzzy covariance matrix. Proc. *IEEE Conf. Decision Contr*., San Diego, CA, 761-766, 1979.

Liu, H. C., Yih, J. M., & Liu, S. W. (2007) . Fuzzy c-mean algorithm based on Mahalanobis distances and better initial values, *INFORMATION SCIENCES 2007, Proceedings of the 10th joint conferences, Salt Lake City, Utah, USA*, 18-24 July, 2007, 1398-1404.

Liu, H. C., Yih, J. M., Sheu, T. W., & Liu, S. W. (2007) . A new fuzzy possibility c-mean algorithm based on unsuppervise Mahalanobis distances , *Proceedings of International conference on Machine Learning and Cybernetics 2007, 7(7)*, 3939-3944, ISBN:978-1-4244-0973-0.

Liu, H. C., Yih J. M., Wu, B. B., & Liu S. W. (2008) . Fuzzy c-mean algorithm based on complete Mahalanobis distances, *2008 International Conferences on Machine Learning and Cybernetics*, Kuming, China, 12-15, July,2008 (accepted).

Pal, N. R., Pal, K., & Bezdek, J. C. (1993) . A possibilistic approach to clustering *IEEE Transactions on Fuzzy Systems*, *1(2)*, 98-110, May, 1993.

Pal, N. R., Pal, K., & Bezdek, J. C. (1997) . A mixed c-mean clustering model, *Proceedings of the Sixth IEEE International conference on Fuzzy System , 1,* 11-21, July. 1997.