Volume 1, Issue 12, pp. 32-37, 2017.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

Knowledge Management Based on Fuzzy Clustering Algorithm with Picard Iteration

Jui-Chiao Tseng
Department of Child Education Min-Hwei Junior College of Health Care Management

Abstract*—Knowledge management of concepts was essential in educational environment. The popular fuzzy c-means algorithm (FCM) converges to a local minimum of the objective function. Fuzzy clustering algorithms are based on Euclidean distance function, which can only be used to detect spherical structural clusters. A Fuzzy C-Means algorithm based on Mahalanobis distance (FCM-M) was proposed to improve those limitations of GG and GK algorithms, but it is not stable enough when some of its covariance matrices are not equal. Hence, different initializations may lead to different results. The important issue is how to avoid getting a bad local minimum value to improve the cluster accuracy. The particle swarm optimization (PSO) is a popular and robust strategy for optimization problems. But the main difficulty in applying PSO to real-world applications is that PSO usually need a large number of fitness evaluations before a satisfying result can be obtained. A new improved Fuzzy Clustering Algorithm with Picard Iteration is proposed. In this paper, the improved new algorithm, Use the best performance of clustering algorithm in data analysis and interpretation. Each cluster of data can easily describe features of knowledge structures. Manage the knowledge structures of Concepts to construct the model of features in the pattern recognition completely.*

Keywords*—Fuzzy Clustering Algorithm, Picard Iteration, Mahalanobis Distance.*

I. INTRODUCTION

In the 1930s, as an Indian statistician, Mahalanobis developed the distance, so called “Mahalanobis distance” which is a distance by using the inverse of the covariance matrix as the metric. Mahalanobis distance is a distance in the geometrical sense because the covariance matrices as well as its inverse are positive definite matrices [1], [2].

As we known, the 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. 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 with recursive process iteratively. However, Krishnapuram and Kim (1999) [3] pointed out that the Mahalanobis distance can not be used directly in clustering algorithm. Gustafson-Kessel (GK) clustering algorithm [4] and Gath-Geva(GG) clustering algorithm[5] 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.

The popular fuzzy c-means algorithm (FCM) is developed by using Picard Iteration through the first-order conditions for stationary points of the objective function. It converges to a local minimum of the objective function. Hence, different initializations may lead to different results. The important issue is how to avoid getting a bad local minimum value to improve the cluster accuracy. The particle swarm optimization (PSO) is a popular and robust strategy for optimization problems. A new improved Fuzzy Clustering Algorithm with Picard Iteration is proposed.

II. LITERATURE REVIEW AND NEW ALGORITHMS.

The new fuzzy clustering algorithm based on normalized Mahalanobis distance which use the homogenous correlation matrix for each cluster, called the Fuzzy C-Means algorithm based on Homogenous correlation matrix. The Euclidean distance based fuzzy clustering algorithms, such as Bezdek’s fuzzy clustering algorithms which can only be used to detect the data classes with the same super spherical shapes. Use alternative distance instead of Euclidean distance as a distance measurement. The shortcoming of Gustafson-Kessel algorithm is different fuzzy covariant matrix corresponding to different geometric, but it can't change its volume. To overcome the drawback due to Euclidean distance, we could try to extend the distance measure to alternative distance with recursive process iteratively. The experimental results of real data sets show that our proposed new algorithms get the better performance if data distribution approach to overlapping too dense or the data structure is not almost spherical.

A. Fuzzy c-Mean Algorithm

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

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

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

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

C is the number of clusters, m is the fuzzier, m>1, which controls the fuzziness of the method. They are both parameters and need to be specified before running the algorithm.

All rights reserved

International Journal of Scientific Engineering and Science
Volume 1, Issue 12, pp. 32-37, 2017. ISSN (Online): 2456-7361

$d_{ij}^2 = \|\underline{x}_j - \underline{a}_i\|^2$ is the square of the Euclidean distance between data object $\underline{x}_j$ to center $\underline{a}_i$. Minimizing objective function Eq. (1) with constraint Eq. (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 Eq. (3) ~ (4).

The steps of the FCM are as follows.

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 $u^{(0)}{}_{ij} = 1, i = 1,2,...,c, j = 1,2,...,n$ are not all equal;

Step 2: Find

$$\underline{a}_i^{(k)} = (\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m \underline{x}_j)(\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m)^{-1}, \quad i = 1,2,...,c \tag{3}$$

$$\mu_{ij}^{(k)} = \left[\sum_{l=1}^{c}\left[((\underline{x}_j - \underline{a}_i^{(k)})'(\underline{x}_j - \underline{a}_i^{(k)}))((\underline{x}_j - \underline{a}_l^{(k)})'(\underline{x}_j - \underline{a}_l^{(k)}))^{-1}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{4}$$

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

Now, for improving the above two problems, we added a regulating factor of covariance matrix, $-\ln\left|{}_{+}\Sigma_i^{-1}\right|$, to each class in objective function, and deleted the constraint of the determinant of covariance matrices, $|M_i| = \rho_i$, in GK Algorithm as the objective function (5). We can obtain the objective function of Fuzzy c-Mean based on adaptive Mahalanobis distance (FCM-M) as following:

$$J_{FCM-M}^{m}(U, A, \Sigma, X) = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m\left[(\underline{x}_j - \underline{a}_i)'\Sigma_i^{-1}(\underline{x}_j - \underline{a}_i) - \ln\left|\Sigma_i^{-1}\right|\right] \tag{5}$$

Proposed by our previous study [9], the algorithm "Fuzzy c-Mean based on adaptive Mahalanobis distance (FCM-M)", is obtained by using the Lagrange multiplier method and the following two theorems, Thm 1 and Thm 2 to minimize the objective function (5) with constraint (2) respect to parameters $\underline{a}_i$, α_j, μ_{ij}, Σ_i.

$$Thm\,1.\ \underline{a}: p\times 1, X: p\times p \Rightarrow \frac{\partial \underline{a}'X\underline{a}}{\partial X} = \underline{a}\underline{a}' \tag{6}$$

$$Thm\,2.\ \ X: p\times p \Rightarrow \frac{\partial \ln|X|}{\partial X} = (X^{-1})' \tag{7}$$

Let

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij}^m)\left[(\underline{x}_j - \underline{a}_i)'\Sigma_i^{-1}(\underline{x}_j - \underline{a}_i) - \ln\left|\Sigma_i^{-1}\right|\right] + \sum_{j=1}^{n}\alpha_j\left(1 - \sum_{i=1}^{c}\mu_{ij}\right),$$
$$where\ 0 \le \alpha_j \le 1,\ j = 1,2,...,n$$

and $\frac{\partial J}{\partial \underline{a}_i} = 0, \frac{\partial J}{\partial \alpha_j} = 0, \frac{\partial J}{\partial \mu_{ij}} = 0, and\ \frac{\partial J}{\partial \Sigma_i} = 0$ (8)

We can get the updating equations as follows.

$$\underline{a}_i = \left[\sum_{j=1}^{n}\mu_{ij}^m\right]^{-1}\left[\sum_{j=1}^{n}\mu_{ij}^m\underline{x}_j\right], i = 1,2,...,c \tag{9}$$

$$\mu_{ij} = \left[\sum_{s=1}^{c}\left[\frac{(\underline{x}_j - \underline{a}_i)'\Sigma_i^{-1}(\underline{x}_j - \underline{a}_i)}{(\underline{x}_j - \underline{a}_s)'\Sigma_i^{-1}(\underline{x}_j - \underline{a}_s)}\right]^{\frac{1}{m-1}}\right]^{-1}, \tag{10}$$
$$i = 1,2,...,c,\ j = 1,2,...,n$$

$$\Sigma_i = \frac{\sum_{j=1}^{n}\mu_{ij}^m(\underline{x}_j - \underline{a}_i)(\underline{x}_j - \underline{a}_i)'}{\sum_{j=1}^{n}\mu_{ij}^m}, \quad i = 1,2,..,c \tag{11}$$

The steps of the FCM-M are as follows. To avoid the singular problem, we let

$$\begin{aligned}
&\Sigma_i^{(k)} = (\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m(\underline{x}_j - \underline{a}_i^{(k)})(\underline{x}_j - \underline{a}_i^{(k)})')(\sum_{j=1}^{n}\left[\mu_{ij}^{(k-1)}\right]^m)^{-1},\\
&\Sigma_i^{(k)} = \sum_{s=1}^{p}\lambda_{si}^{(k)}\Gamma_{si}^{(k)}(\Gamma_{si}^{(k)})'\\
&\Sigma_i = \sum_{s=1}^{p}\lambda_{si}\Gamma_{si}\Gamma_{si}' \quad i = 1,2,...,c\\
&{}_{+}\Sigma_i^{-1} = \sum_{s=1}^{p}(\lambda_{si}^{-1})^{+}\Gamma_{si}\Gamma_{si}',\\
&(\lambda_{si}^{-1})^{+} = \begin{cases}\lambda_{si}^{-1} & if\ \lambda_{si} > 0\\ 0 & if\ \lambda_{si} \le 0\end{cases}\\
&\left|{}_{+}\Sigma_i^{-1}\right| = \prod_{1\le s\le p, \lambda_{si}>0}\lambda_{si}^{-1}
\end{aligned} \tag{12}$$

and we select the better initial membership matrix, the updating functions for $\underline{a}_i$, μ_{ij}, and Σ_i are obtained as Eq. (17), (18), and (21).

The steps of the FCM-M are as follows.

Step 1: Determining the number of cluster; c and m-value (let m=2), given converge 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 $\underline{a}_i^{(0)}, i = 1,2,...,c$ be the result centers of k-mean algorithm.

$d_{ij} = \left\|\underline{x}_j - \underline{a}_i^{(0)}\right\|$ be distances between data object $\underline{x}_j$ to center $\underline{a}_i^{(0)}$.

http://ijses.com/
All rights reserved

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

$$\mu_{ij}^{(0)} = \frac{\left(d_i^M - d_{ij}\right)}{\sum_{k=1}^{c}\left(d_i^M - d_{ij}\right)}, \quad i=1,2,\ldots,c,\ j=1,2,\ldots,n \tag{13}$$

$$\underline{a}_i^{(1)} = (\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right])^{-1}(\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right]\underline{x}_j), i=1,2,\ldots,c \tag{14}$$

$$\Sigma_i^{(1)} = \frac{\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right]\left(\underline{x}_j - \underline{a}_i^{(1)}\right)\left(\underline{x}_j - \underline{a}_i^{(1)}\right)'}{\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right]} \tag{15}$$

$$\left[\Sigma_i^{-1}\right]^{(1)} = \left[0.5\left(\Sigma_i^{(1)} + diag\Sigma_i^{(1)}\right)\right]^{-1} \tag{16}$$

Step 2: Find

$$\underline{a}_i^{(k)} = \frac{\sum_{j=1}^{n}\left(\mu_{ij}^{(k-1)}\right)^m \underline{x}_j}{\sum_{j=1}^{n}\left(\mu_{ij}^{(k-1)}\right)^m}, \ i=1,2,\ldots,c, k\in N \tag{17}$$

$$\left[\Sigma_i^{-1}\right]^{(k)} = \sum_{s=1}^{p}\left[\left(\lambda_{si}^{-1}\right)^+\right]^{(k)}\Gamma_{si}\Gamma_{si}', \tag{18}$$

$$\left(\lambda_{si}^{-1}\right)^+ = \begin{cases}\lambda_{si}^{-1} & if\ \lambda_{si} > 0\\ 0 & if\ \lambda_{si}\le 0\end{cases} \tag{19}$$

$$\left|\Sigma_i^{-1}\right|^{(k)} = \prod_{1\le s\le p,\lambda_{si}>0}\left[\lambda_{si}^{-1}\right]^{(k)} \tag{20}$$

$$\mu_{ij}^{(k)} = \left[\sum_{s=1}^{c}\left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)}\right)'\left[\Sigma_i^{-1}\right]^{(k)}\left(\underline{x}_j - \underline{a}_i^{(k)}\right) - \ln\left|\left[\Sigma_i^{-1}\right]\right|^{(k)}}{\left(\underline{x}_j - \underline{a}_s^{(k)}\right)'\left[\Sigma_s^{-1}\right]^{(k)}\left(\underline{x}_j - \underline{a}_s^{(k)}\right) - \ln\left|\left[\Sigma_s^{-1}\right]\right|^{(k)}}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{21}$$

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

C. New Algorithm Fuzzy Clustering Algorithm with Picard Iteration

Particle Swarm Optimization (PSO) is a quite convenient method for optimizing hard numerical function on metaphor of social behavior of flocks of birds and schools of fish [8]. A swarm consist M individuals, called particles, which change their position over time. Each particle represents a potential solution to the problem of optimization. In FCM, the problem of optimization is to minimize the value of the objective function. Let the particle k in a D-dimension space (D=nc) be represented as

(1) Let the particle k in a D-dimension space (D=nc) be represented as

$$\begin{aligned}\mu_k &= \left(\mu_{k1}, \mu_{k2}, \ldots, \mu_{kD}\right)\\ &= \left(\mu_{k11}, \mu_{k12}, \ldots, \mu_{k1n}, \mu_{k21},\right.\\ &\quad \left.\mu_{k22}, \ldots, \mu_{k2n}, \ldots, \mu_{kc1}, \mu_{kc2}, \ldots, \mu_{kcn}\right)\\ k &= 1,2,\ldots,M\end{aligned} \tag{22}$$

(2) Let the objective function of FCM be the fitness function as follows,

$$\begin{aligned}J_{FCM}^m(U,A,X) &= \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m d_{ij}^2\\ &= \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m\left\|\underline{x}_j - \underline{a}_i\right\|^2\end{aligned} \tag{23}$$

where $\underline{a}_i = (\sum_{j=1}^{n}\left[\mu_{ij}\right]^m)^{-1}(\sum_{j=1}^{n}\left[\mu_{ij}\right]^m \underline{x}_j), \quad i=1,2,\ldots,c$ (24)

(3) Let the particle k in a D-dimension space (D=nc) be represented as

$$\begin{aligned}\mu_k &= \left(\mu_{k1}, \mu_{k2}, \ldots, \mu_{kD}\right)\\ &= \left(\mu_{k11}, \mu_{k12}, \ldots, \mu_{k1n}, \mu_{k21},\right.\\ &\quad \left.\mu_{k22}, \ldots, \mu_{k2n}, \ldots, \mu_{kc1}, \mu_{kc2}, \ldots, \mu_{kcn}\right)\\ k &= 1,2,\ldots,M\end{aligned} \tag{25}$$

(4) Let the objective function of FCM be the fitness function as follows,

$$\begin{aligned}J_{FCM}^m(U,A,X) &= \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m d_{ij}^2\\ &= \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m\left\|\underline{x}_j - \underline{a}_i\right\|^2\end{aligned} \tag{26}$$

where $\underline{a}_i = (\sum_{j=1}^{n}\left[\mu_{ij}\right]^m)^{-1}(\sum_{j=1}^{n}\left[\mu_{ij}\right]^m \underline{x}_j), \quad i=1,2,\ldots,c$ (27)

(5) The best previous position (which possesses the best fitness value) of particle k is denoted by $p_k = \left(p_{k1}, p_{k2}, \ldots, p_{kD)}\right)$, which is also called p_{best}.

(6) The index of the best p_{best} among all the particles is denoted by the symbol g. The location $p_g = \left(p_{g1}, p_{g2}, \ldots, p_{gD)}\right)$ is also called g_{best}.

The velocity for the particle k is represented as $v_k = \left(v_{k1}, v_{k2}, \ldots, v_{kD}\right)$.

(7) p_{best} ,and g_{best} location for iteration t according to following two formulas,

$$\begin{aligned}v_k(t+1) &= wv_k(t) + c_1r_1\left(p_{best}(t) - \mu_k(t)\right)\\ &\quad + c_2r_2\left(g_{best}(t) - \mu_k(t)\right)\end{aligned}$$

$\mu_k(t+1) = normalized\left[\mu_k(t) + v_k(t+1)\right]$, satisfying

$$\mu_{kij}(t+1) == \frac{\mu_{kij}'(t+1)}{\sum_{i=1}^{c}\mu_{kij}'(t+1)}, \quad \forall i=1,2,\ldots,c,\ j=1,2,\ldots,n,\ k=1,2,\ldots,m \tag{28}$$

where

$$\begin{aligned}\mu_{kij}'(t+1) &= \frac{U_{kij}(t) - \min_{1\le j\le n}U_{kij}(t)}{\max_{1\le j\le n}U_{kij}(t) - \min_{1\le j\le n}U_{kij}(t)},\\ U_{kij}(t) &= \mu_{kij}(t) + v_{kij}(t+1)\\ k &= 1,2,\ldots,m,\ i=1,2,\ldots,c,\ j=1,2,\ldots,n\end{aligned} \tag{29}$$

Where w is the inertia coefficient which is a constant in the interval [0,1], and can be adjusted in the direction of linear

http://ijses.com/
All rights reserved

decrease, (In this paper w=0.75); c_1 and c_2 are learning rates which are nonnegative constants(In this paper, $c_1 = c_2 = 2$); r_1 an r_2 are generated randomly in the interval [0,1].

The termination criterion for iterations is determined according to whether the maximum generation or a designated value of the fitness is reached. In this paper, the given converging error is $\varepsilon = 0.001$

$$\max_{1 \le i \le c} \left\| \underline{a}_i(t+1) - \underline{a}_i(t) \right\| < \varepsilon = 0.001$$

where

$$\underline{a}_i(t+1) = \frac{\sum_{j=1}^{n} \left[\mu_{ij}(t)\right]^m \underline{x}_j}{\sum_{j=1}^{n} \left[\mu_{ij}(t)\right]^m}, i = 1, 2, \ldots, c \tag{30}$$

III. Empirical Analysis and Results

The Fuzzy C-Means algorithm (FCM) is the most popular objective function based fuzzy clustering algorithm.The performances of clustering Algorithm FCM, FCM-M, and new algorithm Fuzzy Clustering Algorithm with Picard Iteration all with the fuzzier m=2, are compared in these experiments. The results of FCM, GK, and GG are obtained by applying the Matlab toolbox developed by [9].

A. Equations Style

The mean clustering accuracies of 100 different initial value sets of three algorithms for the Datasets. The performances of clustering Algorithm all with m=2, are compared in these experiments.

TABLE I. The characteristics of 3 clusters.

Cluster	Sample size	Concepts	Average distance of center
1	50	Setosa	0.481705
2	50	Versicolor	0.706870
3	50	Virginica	0.819339

The balance Iris Data [10] with sample size 150 which features of the Iris data contains Length of Sepal, Width of Sepal, Length of Petal, and Width of Petal. The samples were assigned the original 3 clusters based on the clustering analysis. The Iris data with sample size 150 is used as first example. The features of the Iris data contain Length of Sepal, Width of Sepal, Length of Petal, and Width of Petal. The samples were assigned the original 3 clusters based on the clustering analysis. The results were shown in table I.

The results were shown in table II.

TABLE II. Classification accuracies of testing samples.

Algorithm	Accuracies (%)
FCM	89.33
FCM-M	90.00
New Algorithm	92.67

From table II, we find that the New Algorithm Fuzzy Clustering Algorithm with Picard Iteration has the best result, up to 92.67%.

The balance Wdbc Data with sample size 569 which features of the Wdbc data contains 30 attributes. The samples were assigned the original 2 clusters based on the clustering analysis. Data set comes from the University of California at Irvine (UCI) Machine Learning Repository was used in the empirical study. The mean clustering accuracies of 100 different initial value sets of FCM, FCM-M, FCM-NEW for the Datasets. From table III, we find that the New Algorithm Fuzzy Clustering Algorithm with Picard Iteration has the best result, up to 89.97.

TABLE III. The Accuracies of three Algorithms for Wdbc data

Algorithm	Accuracies (%)
FCM	79.78
FCM-M	79.89
New Algorithm	89.97

The data set was collected from Min-Hwei Junior College of Health Care Management in Taiwan is an achievement test on medicine and technology of nursing with 345 task-takers. About the data set contained 20 dichotomous items which measured five attributes. According to their scores of achievement test, we can group the students into 3 concepts.

TABLE IV. The Accuracies of three Algorithms for 345 students data

Algorithm	Accuracies (%)
FCM	69.78
FCM-M	69.89
New Algorithm	79.78

The mean clustering accuracies of 100 different initial value sets of FCM, FCM-M, FCM-NEW for the Datasets. From table III, we find that the New Algorithm Fuzzy Clustering Algorithm with Picard Iteration has the best result, up to 79.78.

An improved new fuzzy clustering algorithm is developed to obtain better quality of fuzzy clustering result. The objective function includes the regulating terms about the covariance matrices. The equations for the memberships and the cluster centers and the covariance matrices are directly derived from the Lagrange's method. The fuzzy c-mean algorithm is different from the FCM and FCM-M algorithms. The singular problem and detecting the local extreme value problem are improved by the Eigenvalue method and the algorithm of Particle Swarm Optimization. Finally, four numerical examples showed that the new fuzzy clustering algorithm with Picard Iteration gave more accurate clustering results than that of FCM and FCM-M algorithms.

IV. Conclusion

The clustering analysis 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. FCM is based on Euclidean distance function, which can only be used to detect spherical structural clusters. GK algorithm and GG algorithm were developed to detect non-spherical structural clusters. However, GK algorithm needs added constraint of fuzzy covariance matrix, GG algorithm can only be used for

All rights reserved

International Journal of Scientific Engineering and Science
Volume 1, Issue 12, pp. 32-37, 2017. ISSN (Online): 2456-7361

the data with multivariate Gaussian distribution.

To overcome the drawback due to Euclidean distance, we could try to extend the distance measure to Mahalanobis distance. However, Krishnapuram and Kim (1999) pointed out that the Mahalanobis distance can't be used directly in clustering algorithm.

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. To add a regulating factor of each covariance matrix to each class in the objective function, and deleted the constraint of the determinants of covariance matrices in the GK algorithm.

Pal, Pal and Bezdek's Fuzzy Possibility C-Means (FPCM) (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 and 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. Furthermore, An improved fuzzy new algorithm with Picard Iteration can improve the stability of the clustering results is proposed [11, 12]. We know that applying the fuzzy covariance matrices to compute the value from the objective function in Mahalanobis distances. Now, the fuzzy covariance matrices were not directly derived from the whole date in Gath-Geva clustering algorithm. So the correct rate of grouping is less performance. However Gath-Geva clustering algorithm can only be used for the data with multivariate Gaussian distribution. Hence our goal is to improve those limitations, and has better performance.

The Gustafson-Kessel algorithm cannot adjustment the fuzzy covariant matrix corresponding to different geometric, So the disadvantages of Gustafson-Kessel algorithm is that it can't change its volume. The experimental results of six real data sets consistently show that the performance of our proposed new algorithm is better than those of the GK and GG algorithms. For benchmark data sets; Iris data and Wdbc data, those have shown that the improved fuzzy partition algorithms, new algorithm are better than FCM and FCM-M [13-17]. Apply new algorithm to identify the master concepts for comparing the performances of other two partition algorithms, FCM and FCM-M. The result also shows that new fuzzy clustering algorithm with Picard Iteration is better than other two fuzzy partition algorithms.

APPENDIX

The mean clustering accuracies of 100 different initial value sets of three algorithms for the Datasets. The performances of clustering Algorithm all with m=2, are compared in these experiments.

TABLE V. Clustering accuracies of experiment for 100 initial value sets.

FCM	FCM-M	NEW Algorithm
63.33%	69.89%	84.43%
65.83%	79.98%	84.43%
64.17%	78.78%	81.97%
65.00%	77.78%	82.95%
64.17%	79.78%	79.70%
73.13%	78.78%	88.55%
69.13%	69.98%	86.07%
51.67%	79.78%	84.43%
65.08%	73.61%	84.43%
45.08%	73.77%	84.43%
70.00%	77.21%	84.43%
65.08%	78.85%	78.69%
65.78%	70.49%	86.43%
6408%	73.77%	84.43%
62.30%	74.59%	87.70%
69.18%	73.77%	84.43%
65.08%	78.85%	78.69%
69.18%	73.77%	84.43%
65.08%	73.77%	84.43%
55.08%	79.51%	87.43%
69.51%	81.15%	88.03%
55.08%	80.33%	84.43%
55.08%	73.77%	84.43%
66.39%	72.13%	86.43%
63.11%	75.41%	89.43%
63.77%	76.23%	84.43%
72.95%	74.59%	87.70%
56.72%	65.57%	84.43%
52.62%	73.77%	84.43%
63.77%	67.21%	84.43%
65.57%	73.77%	84.43%
55.08%	76.23%	84.43%
65.08%	73.77%	83.61%
69.95%	73.77%	80.49%
67.87%	68.03%	84.43%
61.31%	73.77%	81.97%
55.08%	73.77%	82.79%
65.08%	73.77%	84.43%
61.31%	73.77%	81.97%
55.08%	73.77%	82.79%
55.74%	73.77%	78.20%
55.08%	73.77%	84.43%
59.84%	81.97%	84.43%
45.08%	73.77%	84.43%
73.77%	81.15%	84.43%
65.08%	73.77%	84.43%
63.77%	76.23%	84.43%
72.95%	74.59%	87.70%
56.72%	65.57%	84.43%
52.62%	73.77%	84.43%
63.77%	67.21%	84.43%
65.57%	73.77%	84.43%
55.08%	76.23%	84.43%
65.08%	73.77%	83.61%
69.95%	73.77%	80.49%

http://ijses.com/
All rights reserved

FCM	FCM-M	NEW Algorithm
61.31%	78.77%	81.97%
55.08%	75.97%	82.79%
65.08%	73.87%	84.43%
65.08%	77.79%	84.43%
55.08%	79.51%	87.43%
69.51%	81.15%	88.03%

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, NY, 1981.
[2] 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, no. 1, pp. 634-639, 2004.
[3] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98-110, 1993.
[4] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," *Proc. IEEE Conf. Decision Contr., San Diego*, CA, 761-766, 1979.
[5] I. Gath, and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell*, pp. 773-781, 1989.
[6] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters," *J. Cybern*, vol. 3, issue 3, pp. 32-57, 1973.
[7] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, NY. 1981.
[8] J. Kennedy, and R. C. Eberhart, "Particle swarm optimization," *Proc. of the IEEE International Conference on Neural Networks,* vol. IV, pp. 1942-1948, 1995.
[9] B. Balasko, J. Abonyi and B. Feil "Fuzzy clustering and data analysis toolbox for use with matlab," From http://www.mathworks.com/ matlab central/fileexchange/7473.
[10] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Human genetics*, vol. 7, issue 2, pp. 179-188, 1936.
[11] C. C. Hung, A. Fahsi, and T. Colem, "An Improvements of multispectral image classification using contextual information," *International Journal of Remote Sensing*, (Revised- comments from editors and resubmitted September 2003).
[12] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis*, John Wiley and Sons, 1999. http://as.wiley.com/WileyCDA/WileyTitle/productCd-0471988642.html.
[13] C. C Hung, and G. Germany, "K-means and iterative selection algorithms in image segmentation," in *Proceedings of IEEE Southeast Conference*, Session 1: Software Development, Jamaica, West Indies, April 4-6, 2003.
[14] N. R. Pal and J. K. Bezdek, "A mixed C-mean clustering model," *Proceedings of the Sixth IEEE International conference on Fuzzy Systems*, vol. 1, pp. 11-22, 1997.
[15] H. C. Liu, T. W, Sheu, and S. W. Liu, "A new fuzzy possibility clustering algorithms based on unsupervised mahalanobis distances," *Proceedings of International Conference on Machine Learning and Cybernetics*, vol. 7, issue 7, pp. 3939-3944, 2007.
[16] Balasko, J. Abonyi, and B. Feil, "Fuzzy clustering and data analysis toolbox for use with matlab," pp. 15, 2008.
[17] W. M. Bart and D. J. Krus, "An ordering-theoretic method to determine hierarchies among items," *Educational and Psychological Measurement*, Sage Publications, vol. 33, pp. 291-300, 1973.
[18] H.-C. Liu, J.-M. Yih, D.-B. Wu, and S.-W. Liu, "Fuzzy c-means algorithm based on adaptive mahalanobis distances," *Information Sciences 2007, Proceedings of the 10th Joint Conferences*, Salt Lake City, Utah, USA. 18-24 pp. 1398-1404, July, 2007.

All rights reserved