The IEEE International Conference on
Industrial Informatics (INDIN 2008)
DCC, Daejeon, Korea July 13-16, 2008

Objective Function of Semi-Supervised Fuzzy C-Means Clustering Algorithm

Chunfang Li[1,3], Lianzhong Liu[2] and Wenli Jiang[2]
School of Automation and Electrical Engineering[1], School of Computer Science and Engineering[2]
Beijing University of Aeronautics and Astronautics, Beijing, 100083 China
Network Center, Hebei Institute of Physical Education[3], Shijiazhuang, 050041 China
lcf1974@126.com, lz_liu@buaa.edu.cn, jwljiangwenli@sina.com

***Abstract*-Analyzed here is the physical interpretation of objective function of semi-supervised Fuzzy C-Means (SS-FCM) algorithm and its coefficient *a*. A conclusion-Stutz's modification to the objective function of Pedrycz is much clearer: unlabeled samples involves in unsupervised learning of FCM, labeled samples involves in unsupervised learning with coefficient (1-*a*) and participate in supervised learning with *a*, and when *a*=1 or 0, the SS-FCM degrades to FCM-is illustrated. The corresponding alternately optimizing algorithm of SS-FCM with fuzzy covariance is provided. The experimental results show that: 1) Modified algorithm has the same semi-supervised role and has much clearer physical interpretation. 2) Using FCM algorithm to assign membership for labeled samples is better than using random number. 3) SS-FCM with fuzzy covariance and a small number of well-selected labeled samples can effectively improve the accuracy and convergence speed.**

I. Introduction

Semi-supervised learning (SSL) aims at to develop classifiers that can utilize both labeled and unlabeled samples. New SSL methods have significantly improved machine learning in various applications, including business intelligence, text categorization, computer vision, image processing, and bioinformatics. SSL can be divided into: semi-supervised classification and semi-supervised clustering. In supervised classification, there is a known, fixed set of categories and category-labeled training data are used to induce a classification function. In semi-supervised classification, training also exploits additional unlabeled data, frequently resulting in a more accurate classification function. In unsupervised clustering, an unlabeled dataset is partitioned into groups of similar examples. In semi-supervised clustering, some labeled samples are used along with the unlabeled samples to obtain a better clustering, such as much quickly convergence speed and higher accuracy. Unlike semi-supervised classification, semi-supervised clustering can group data using the categories in the initial labeled data, as well as extend and modify the existing set of categories as needed to reflect other regularities in the dataset [1].

Several semi-supervised algorithms have been proposed. They can be categorized based on the computational model used [2] [3]. The most known models are: the seeding model [1], the probabilistic model, the objective function optimization model, genetic algorithms, support vector machines, and graph-based model [4].

Semi-supervised Fuzzy C-Means (SS-FCM) algorithm investigated by Pedrycz [2][5][6][7] is typical example of the objective function optimization model, which is introduced the supervision of labeled samples in the objective function of Fuzzy C-Means algorithm.

This paper analyzed the composition and physical interpretation of objective function of SS-FCM, coefficient *a,* and initialization of membership values of labeled data. We have modified the objective function of SS-FCM to make its physical interpretation much clearer and provide new iterative procedure to calculate membership values of labeled and unlabeled samples respectively. More important we illustrate the base principle of SS-FCM, which helps to produce more new semi-supervised clustering algorithm.

This paper organized as follows: First, in section II, we give a brief introduction on FCM algorithm and FCM with fuzzy covariance (CFCM). In section III, the objective function of SS-FCM is discussed and a new form of SS-FCM with fuzzy covariance (SS-CFCM) is induced. In section IV, experimental results of FCM, SS-FCM CFCM and SS-CFCM are compared with each other and the interpretation of each result is discussed. We end with some useful conclusions in section V. SS-FCM in this paper sometimes refers to the generalized semi-supervised FCM, i.e., including SS-FCM and SS-CFCM, please distinguish them in the context.

II. A Brief Introduction On FCM And CFCM

In clustering algorithms with objective function, FCM is rather perfect in theory and is the most widely used. FCM is first proposed by Dunn, and then is extended by Bezdek [8], and etc.. Dunn applied fuzzy means and fuzzy partition theory to K-Means algorithm and induced FCM, and Bezdek generalized FCM to infinite family by introducing a coefficient m that controls the fuzzy degree of clustering.

A. Objective Function of FCM

Let $X=\{x_1, x_2,\dots, x_n\}$ be a finite dataset: each $x_i=(x_{i1},x_{i2},\dots,x_{if})$ has f features, $X\in R^{n\times f}$. To classify X into c ($2\le c<n$) categories, let $V=(v_1, v_2, \dots, v_c)^{\mathrm{T}}$ be the cluster centers or prototypes matrix, $V\in R^{c\times f}$. In order to obtain the optimization fuzzy classification, a fuzzy matrix U is selected, where $U\in M_{fc}$.

$$M_{fc}=\left\{U\in R^{c\times n}\mid u_{ik}\in[0,1]\forall i,k;\sum_{i=1}^{c}u_{ik}=1;\sum_{k=1}^{n}u_{ik}>0\forall i\right\} \quad (1)$$

Minimize the objective function J_m to compute U and V,

$$J_m(U,V)=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m\left\|x_k-v_i\right\|^2 \tag{2}$$

Here $m>1$, $\|x_k-v_i\|$ is any inner product induced norm on R^f.

Use Lagrange multiplier to minimize J_m, only if

$$v_i^{(l)}=\sum_{k=1}^{n}(u_{ik}^{(l)})^m x_k \Big/ \sum_{k=1}^{n}(u_{ik}^{(l)})^m \tag{3}$$

$$u_{ik}^{(l+1)}=\left(\sum_{j=1}^{c}\left(\frac{\left\|x_k-v_i^{(l)}\right\|}{\left\|x_k-v_j^{(l)}\right\|}\right)^{\frac{2}{m-1}}\right)^{-1} \tag{4}$$

(k=1,2,…,n, j=1,2,…,c).

In particular, it is shown in [8] that (U,V) may be a local solution for (2) when $1<m<\infty$, given c, initial fuzzy partition matrix $U^{(0)}$ and iterative accuracy e.

B. FCM with Fuzzy Covariance Matrices

Euclidean distance assumes that the overall features are equally important, so FCM with Euclidean distance fit for clustering of hyper-sphere dataset. In real-world dataset, features can be measured against different scales. Discrepancies resulting from the difference in the domain of the features can distort the distance calculations. Therefore, it is advisable to normalize the values of features before calculating Euclidean distance.

Besides different scales of feathers, if there are some correlations in the features of dataset that subject to Normal or Gaussian distribution, we can directly use Mahalanobis distance instead of normalization. The Mahalanobis distance of two vectors x and y is calculated as following:

$$d(x,y)=(x,y)S^{-1}(x,y)^T. \tag{5}$$

S^{-1} is inverse matrix of the covariance matrix. Especially if S^{-1} is identity matrix, Mahalanobis distance equals to Euclidean distance. S is a real symmetric matrix, and S_{ij} is the covariance of ith (vector x) and jth (vector y) feature.

$$\text{covariance(x, y)} = s_{xy} = 1\Big/(n-1)\sum\nolimits_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y}) \tag{6}$$

Mahalanobis distance eliminates the different scale and correlation in features by weighted with S^{-1}.

Gustafson [9] investigated FCM with covariance matrix and proposed the conception of fuzzy covariance. Prototype v_i is fuzzy mean of all samples belonging to ith category, so he uses v_i instead of ordinary mean to calculate the fuzzy covariance matrices P_i and it's inverse matrices M_i.

$$P_i=\frac{\sum_{k=1}^{n}u_{ik}^m(x_k-v_i)(x_k-v_i)}{\sum_{k=1}^{n}u_{ik}^m} \quad (i=1,2,\ldots,c) \tag{7}$$

$$M_i^{-1}=\left[\frac{1}{\rho_i\det(P_i)}\right]^{\frac{1}{f}}P_i \quad (i=1,2,\ldots,c) \tag{8}$$

Where $\rho_i=\det(M_i)$, typically ρ_i=1, f is the size of features. So compute distances using (9)

$$d(x_k,v_i)=(x_k,v_i)M_i(x_k,v_i)^T \tag{9}$$

FCM with fuzzy covariance matrices can discover hyper-ellipsoid structure of dataset, and so CFCM algorithm greatly improves the accuracy of clustering in various applications.

III. Semi-Supervised FCM Algorithms

The notion of good clustering is strictly related to application's area and the user's perspectives. Traditional clustering methods fail leading to meaningless results in the case of high-dimensional dataset. Semi-supervised algorithm use user input to guide clustering to a meaningful solution. User input can be classified into: constraints and labeled data [1]. There are two types of constraints, must-link (two samples have to be together in the same cluster) and cannot-link (two samples have to be in different clusters), and they are used in the clustering process to generate a partition that satisfies the given constraints as many as possible. In some cases, constraints can be converted into labeled data, or vice versa. User input assumes to be labeled data in SS-FCM, and you will find constraints easy to change into labeled data from SS-FCM algorithm itself.

A. Pedrycz 's Semi-Supervised FCM Algorithm

The core of SS-FCM is how to use labeled data to guide the iterative optimization. Pedrycz gives the objective function (10) of SS-FCM in [5], J_2^s denotes fuzzy coefficient m=2, superscript s denotes semi-supervised.

$$J_2^s(U,V;X,F)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^2d_{ik}^2+\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}-f_{ik})^2d_{ik}^2 \tag{10}$$

To distinguish the labeled and unlabeled samples, he [6] introduces a two-value vector $\boldsymbol{b}=[b_k]$(k=1,2,…n)

$$b_k=\begin{cases}1 & \text{if sample } \mathrm{x_k} \text{ is labeled}\\ 0 & \text{otherwise}\end{cases} \tag{11}$$

Membership values of the labeled samples are arranged in a matrix $F=[f_{ik}]$ (i=1,2,…,c, k=1,2,…,n). The component of supervised learning encapsulated in $\boldsymbol{b}$ and F contributes additively to the objective function

$$J_{m,a}^s(U,V;X,F)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m d_{ik}^2+a\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}-f_{ik}b_k)^m d_{ik}^2 \tag{12}$$

Here coefficient a denotes a scaling factor whose role is to maintain a balance between the supervised and unsupervised component within the optimization mechanism. Pedrycz suggests that $a\propto n/L$, L denotes the size of labeled samples. He also rewrite (12) in the form

$$J_{m,a}^s(U,V;X,F)=J_m+a\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}-f_{ik}b_k)^m d_{ik}^2 \tag{13}$$

Here J_m is the objective function of standard FCM algorithm.

Stutz [10] adopts the idea of Pedrycz and modifies the form of objective function

$$J^s_{m,a}(U,V;X,F)=(1-a)\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m d_{ik}^2+a\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}-f_{ik})^m d_{ik}^2 \quad (14)$$

Here $a \in [0,1]$ is a weighting factor that is chosen corresponding to the reliability of the labeled samples: The higher the reliability, the higher a.

Pedrycz's algorithms assume that the number of clusters c determined by the clustering algorithm should be the same as the number of categories (denoted by H) reflected by labeled samples. He also points SS-FCM can deal with the case of part of clusters having no labeled samples, that is, SS-FCM can discover hidden data structure by complete unsupervised learning.

In many real-world situations the available labeled samples do not reflect the whole structure of the dataset, but they reflect some user conditions. Bouchachia [2] proposes a SS-FCM that can solve the case of $c>H$. He assumes that a class can be partitioned into several clusters. Each class h contains a number of clusters c_h.

$$J^s_{m,a}(U,V;X,F)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m d_{ik}^2+a\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}-\tilde{u}_{ik})^m d_{ik}^2 \quad (15)$$

Here $\tilde{u}_{ik}$ denotes the membership of labeled sample, and compute it through minimizing another objective function Q

$$\tilde{u}_{ik}^{(r)}=\tilde{u}_{ik}^{(r-1)}-d\frac{\partial Q(F,\tilde{U})}{\partial \tilde{u}_{ik}} \quad (16)$$

Where d is a positive parameter.

$$Q(F,\tilde{U})=\sum_{h=1}^{H}\sum_{k=1}^{n}p_k(f_{hk}-\sum_{i\in T_h}\tilde{u}_{ik})^2 \quad (17)$$

Here p_k=1 if the sample k is labeled, 0 otherwise; f_{hk}=1 if sample x_k belongs to class h, otherwise 0; T_h is the set of clusters belonging to class h.

The interpretation of Q: only to compute the labeled samples, if x_k belongs to class h, compute the sum of it's membership values belonging to classes except h ; if x_k doesn't belong to class h, compute it's membership value belonging to h. To minimize Q is to maximize the x_k who belongs to class h, and minimize x_j who doesn't belong to h.

Note about Bouchachia's SS-FCM algorithm: 1) Apply the standard FCM on the whole dataset (both labeled and unlabeled samples) to get the partition matrix $U^{(0)}$. 2) Determine the set T_h of each class. 3) Compute $\tilde{u}_{ik}$ by minimizing Q. 4) Compute partition matrix U by minimizing (15). 5) According to the maximum of sum of the membership degree of clusters of the same class to get the partition, or the maximum of maximum of the membership degree of clusters of the same class get final partition.

This solution succeeds to deal with the case of assigning some clusters into one class to satisfy some of the user's requirements.

B. Modification to SS-FCM

The objective function of standard FCM is the function of within-group weighted sum of squared error. The second item of (10) (12) (13) (15) includes labeled and unlabeled samples and unlabeled samples is weighted by FCM membership value minus 0, but labeled samples is weighted by FCM algorithm membership value minus known membership value. In fact, the additive unlabeled data is unnecessary and makes the meaning of objective function unclear.

In this paper, let the former L samples be labeled, and others be unlabeled. We rewrite (14) in the form (18) because of the known membership degree of unlabeled being 0.

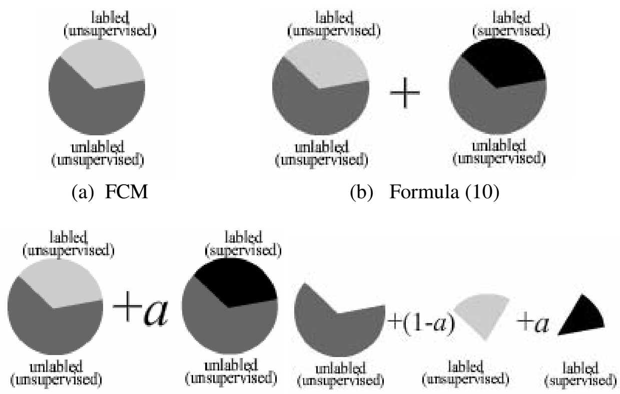
$$J^s_{m,a}(U,V;X,F)=(1-a)\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m d_{ik}^2+ a\sum_{i=1}^{c}\sum_{k=1}^{L}(u_{ik}-f_{ik})^m d_{ik}^2+a\sum_{i=1}^{c}\sum_{k=L+1}^{n}u_{ik}^m d_{ik}^2 \quad (18)$$

To merge unlabeled samples, (18) is another form :

$$J^s_{m,a}(U,V;X,F)=\sum_{i=1}^{c}\sum_{k=L+1}^{n}u_{ik}^m d_{ik}^2+ (1-a)\sum_{i=1}^{c}\sum_{k=1}^{L}u_{ik}^m d_{ik}^2+a\sum_{i=1}^{c}\sum_{k=1}^{L}(u_{ik}-f_{ik})^m d_{ik}^2 \quad . (19)$$

As show in Fig. 1: FCM doesn't distinguish labeled and unlabeled, all samples participate in unsupervised learning; the composition of (10) is added part of unlabeled samples' unsupervised clustering and labeled samples' supervised grouping; because if the labeled sample is well-selected, their membership values are very closer to membership of FCM, the supervised role of labeled is very unobvious, so coefficient a is introduced to augment the role of supervision in (12) (13) (15), but meanwhile the role of unsupervised of unlabeled is also expanded; In (14) (18) (19), unlabeled samples involves in unsupervised clustering, but labeled samples involves in unsupervised clustering with factor (1-a) and participate in supervised classification with factor a.

According to Pedrycz's suggestion, we recommend a=1-L/n when L is far less than n. If a=0, then L=n, all samples are labeled, but they participate in unsupervised learning; If a=1, then L=0, all samples are unlabeled, they participate in



(a) FCM (b) Formula (10)

(c) Formula (12)(13)(15) (d) Formula (14)(18)(19)

Fig. 1. Composition of objective function of FCM and various SS-FCM

unsupervised learning. That is, when a=1 or 0, the SS-FCM all degrades to FCM. In fact, when L is slightly less than and near to n and a is viewed as reliability of labeled samples, a can be selected near to 1 to augment the role of supervision.

C. Alternately Iterative Optimization Procedure of Modified SS-FCM with m=2

1) Given is the number of clusters c, labeled and unlabeled samples $\boldsymbol{b}$ and known membership values of labeled F. Initialize partition matrix U.

2) Calculate centers (prototypes) of the clusters and the fuzzy covariance matrices

$$v_i^{(l)} = \sum_{k=1}^{n} (u_{ik}^{(l)})^2 x_k \Big/ \sum_{k=1}^{n} (u_{ik}^{(l)})^2 \tag{20}$$

$$P_i = \frac{\sum_{k=1}^{n} u_{ik}^m (x_k - v_i)(x_k - v_i)}{\sum_{k=1}^{n} u_{ik}^m} \quad (i=1,2,\ldots,c) \tag{21}$$

$$M_i^{-1} = \left[\frac{1}{\rho_i \det(P_i)}\right]^{\frac{1}{f}} P_i \quad (i=1,2,\ldots,c) \tag{22}$$

where $\rho_i = \det(M_i)$, typically ρ_i=1

$$d(x_k, v_i) = (x_k, v_i) M_i (x_k, v_i)^T \tag{23}$$

3) Update partition matrix U'.

If unlabeled, $$u_{ik} = 1 \Big/ \sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^2 = u_{ik}^{FCM} \tag{24}$$

If labeled,

$$u_{ik} = (1-a)\left(1 \Big/ \sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^2\right) + af_{ik} = (1-a)u_{ik}^{FCM} + af_{ik} \tag{25}$$

4) Compare U' to U, if $\|U - U'\| < e$ (with e being iterative accuracy) then stop, else go to *2)* with $U=U'$.

IV. Simulation Studies

A misclassified sample is defined as the assigned label is different from most of other samples coming from original same class. The performance accuracy rate R is:

$$R = \frac{\text{Number of correctly assigned samples}}{\text{Size of the testing dataset}} \tag{26}$$

There are two methods to assign known membership values for labeled samples: 1) random number; 2) FCM algorithm. Randomly generated number should satisfy the condition that according to fuzzy partition theory labeled samples are correctly classified; Membership values calculated by FCM algorithm should also ensure that according to fuzzy partition theory labeled samples are correctly classified. Which samples are used as labeled is selected randomly, so that we can observe the stability of SS-FCM. And the size of labeled samples is the same number in each cluster. Following results are average of 20 times experiments.

A. Experiment on artificial XOR dataset

2-demensional dataset shown in Fig.2 (a) has been generated manually in Matlab. It consists of 4 overlapping clusters-two of them are ellipsoidal while the remaining ones form a visible cross of points resembling the standard exclusive-or problem. Each cluster has 50 samples. The obtained classification results vary significantly between the 4 algorithms. Fig. 2 (b)-(f) shows misclassified samples. Notice that SS-CFCM generated prototypes that were most representative of each class. The effect of semi-supervised learning is significantly improving the outcomes of clustering.

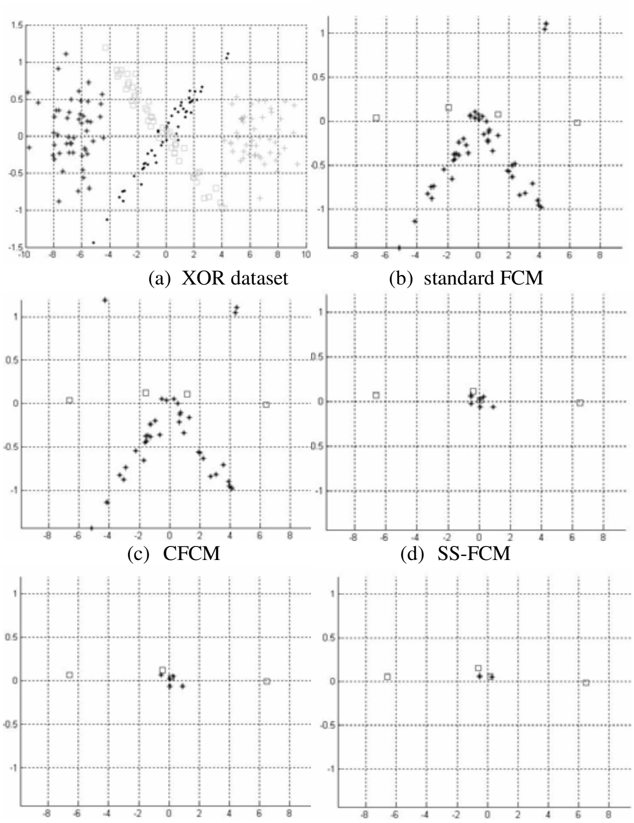
The classification results are illustrated numerically in Table I. Convergence performance of FCM algorithm is relatively stable but low accuracy. The accuracy of SS-FCM with 20% labeled samples is increased slightly, but there are 2 clustering failures because of algorithm falling into local extreme of objective function. Accuracy of CFCM is greatly improved compared to FCM, especially the classes of cross-structure; Accuracy under stable convergence achieves to 96.6%, but there are 11 clustering failures. Convergence speed of SS-CFCM with random initialization of labeled is slightly improved, and there is one clustering failure. When SS-CFCM is initialized by FCM's membership values of labeled, the accuracy does not increase, but iteration reduces to 11 times and no clustering failure. SS-CFCM* is the algorithm of Pedrycz's objective function, and because of adjustment of a, clustering results are generally the same to SS-CFCM. SS-CFCM with the labeled samples of 60%, a = 0.4, convergence speed is increased slightly. When a is explained for the reliability of the labeled, the labeled is 90%, a = 0.9, the misclassified reduced to three.

B. Experiment on Iris

Iris dataset is commonly viewed as a standard benchmark for testing supervised and unsupervised learning algorithms. Iris has 4 feathers and 3 classes. Two features of Iris plot in (a) of Fig. 3. Notice that (c) and (d) in Fig. 3, the misclassified samples of SS-FCM and of CFCM are greatly different, as is mainly because fuzzy covariance matrices eliminate the differences between features of dataset.

The performance of SS-CFCM on Iris is similar with on XOR, as shown in Table II. Membership degree of the labeled by random number is unstable in convergence, and there are 10 clustering failures in 20 tests. It is worth mentioning that when labeled samples is 60% and 90% respectively, a = 0.9, there are three times completely correct clustering.

In fact, users always hope that a small number of samples have better supervisory role. Using well-selected labeled samples, clustering results of SS-CFCM are shown in Table III. With 10% labeled samples, the misclassified reduce to 3 samples, so it reflects that SS-CFCM achieves the purpose of supervised clustering only with a small number of labeled samples. With 6% labeled samples, there are 9 times clustering failure, because the size of labeled samples is too small, and they are not enough to represent the whole dataset.

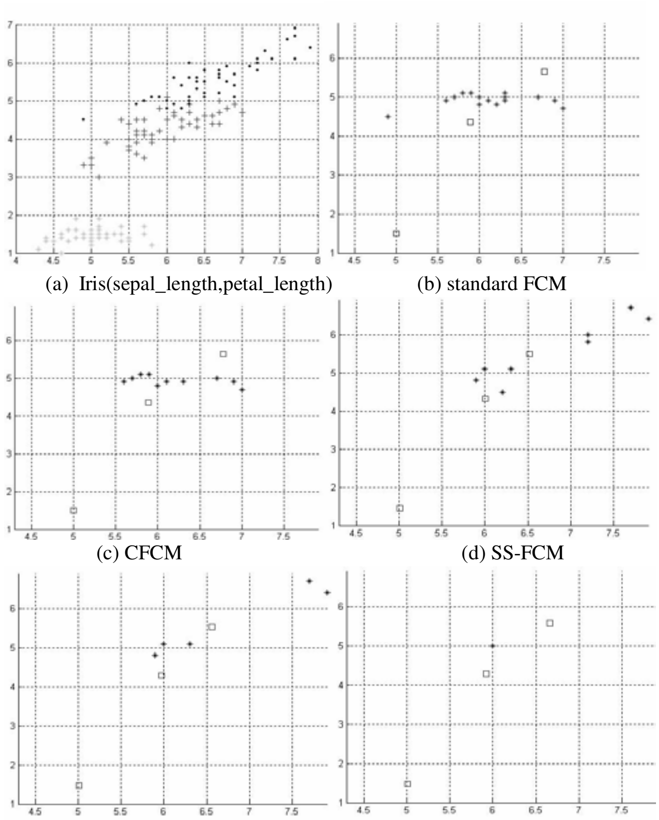


(a) XOR dataset (b) standard FCM

(c) CFCM (d) SS-FCM

(e) SS-CFCM with L=40 and α=0.8 (f) SS-CFCM with L=180 and α=0.9

Fig. 2. XOR dataset and misclassified samples
(□ centers of clusters, * misclassified samples)



(a) Iris(sepal_length,petal_length) (b) standard FCM

(c) CFCM (d) SS-FCM

(e) SS-CFCM with L=40 and α=0.8 (f) SS-CFCM with L=180 and α=0.9

Fig. 3. Iris dataset and misclassified samples
(□ centers of clusters, * misclassified samples)

TABLE I
CLASSIFICATION STATISTICS FOR XOR DATASET.

Algorithm	L'(%)	a	Labeled init.	Misclassified	Accuracy rate	Iterations	Failure
FCM	0	0		45	77.5%	26	0
SS-FCM	20%	0.8	FCM	39	80.5%	17	2
CFCM	0	0		7	96.5%	20	11
SS-CFCM	20%	0.8	random	7	96.5%	16	1
SS-CFCM	20%	0.8	FCM	7	96.5%	11	0
SS-CFCM*	20%	5	FCM	7	96.5%	11	0
SS-CFCM	60%	0.4	FCM	7	96.5%	9	0
SS-CFCM	60%	0.9	FCM	5	97.5%	9	0
SS-CFCM	90%	0.9	FCM	3	98.5%	8	0

TABLE II
CLASSIFICATION STATISTICS FOR IRIS DATASET

Algorithm	L'(%)	a	Labeled init.	Misclassified	Accuracy rate	Iterations	Failure
FCM	0	0		16	89.3%	21	0
SS-FCM	20%	0.8	FCM	13	91.3%	17	0
CFCM	0	0		8	94.7%	56	0
SS-CFCM	20%	0.8	random	8	94.7%	59	10
SS-CFCM	20%	0.8	FCM	4	97.3%	32	0
SS-CFCM*	20%	5	FCM	5	96.7%	53	0
SS-CFCM	60%	0.4	FCM	4	97.3%	24	0
SS-CFCM	60%	0.9	FCM	2	98.7%	20	0
SS-CFCM	90%	0.9	FCM	2	98.7%	15	0

TABLE III
EFFECT OF THE SIZE OF LABELED

L'	a	Misclassified	Accuracy Rate	Iterations	Failure
6%	0.94	8	94.7%	37	9
10%	0.9	3	98%	27	0
20%	0.8	2	98.7%	22	0

V. CONCLUSIONS

This paper clarified the physical meaning of the objective function of semi-supervised FCM algorithm, and we modify SS-FCM to make it easy to understand and apply in engineering. The simulation experiments show that this modified algorithm has the same semi-supervised role and has clearer physical interpretation compared to the original.

Samples will be divided into unlabeled and labeled:

unlabeled samples involves in unsupervised learning of FCM, labeled samples involves in unsupervised learning with coefficient (1-a) and participate in supervised learning with a. A recommendation of selecting a is a = 1-L/n, where L denotes the size of labeled samples. a can be regarded as reliability of labeled samples, selecting its value closer to 1 to increase the role of supervision when L is slightly less than and near to n. When the number of labeled samples is smaller in dataset, 1-L/n is consistent with that labeled samples should be higher reliability for a better clustering result .

SS-FCM clustering with a small number of representative labeled samples and their membership degree computed by FCM and with fuzzy covariance matrices to calculate distance, the clustering performance effectively improves the accuracy and the speed of convergence, and meanwhile greatly reduces the opportunities of falling into local minimum of objective function. On the contrary, if selecting labeled samples is not typical in clusters, or random numbers is assigned as the known membership of labeled samples, SS-FCM algorithm will lead to a large number of clustering failures.

ACKNOWLEDGMENT

This work is partly supported by the National High-Tech Research and Development Plan of China under grant No. 2004AA113040 and partly by Project of Crime Analysis with Contract No. HDFJ-SD-051201.

REFERENCES

[1] S. Basu, A. Banerjee, and Mooney, "Semi-supervised clustering by seeding," *Proceedings of the International Conference on Machine Learning*, pp.19-26, July 2002.

[2] A. Bouchachia and W. Pedrycz, "Data clustering with partial supervision," *Data Mining and Knowledge Discovery*, vol.12, pp. 47-78, January 2006.

[3] A. Bouchachia, "Learning with hybrid data," *Proceedings of the 5th International IEEE Conference on Hybrid Intelligent Systems*, pp. 193-198, 2005.

[4] X. J. Zhu, "Semi-supervised learning with graphs," Pittsburgh: School of Computer Science, Carnegie Mellon University.2005.

[5] W. Pedrycz, "Algorithms of fuzzy clustering with partial supervision," *Pattern Recognition Letter*, vol.3, pp.13-20, 1985.

[6] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Trans. on Syst., Man, Cyber.,* vol. 27, pp.787-795, May 1997.

[7] W. Pedrycz and J. Waletzky, "Neural-network front ends in unsupervised learning," *IEEE Trans. on Neural Networks*, vol.8, pp.390-401, March 1997.

[8] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.

[9] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," *IEEE CDC in Proc.*, pp. 761-766, 1978.

[10] C. Stutz and T. A. Runkler, "Classification and prediction of road traffic using application-specific fuzzy clustering," *IEEE Trans. on Fuzzy Systems*, vol.10, pp. 297-308, March 2002.