Cluster Analysis Algorithms Based on the Gradient Descent Procedure of a Fuzzy Objective Function

Hyun-Sook Rhee and Kyung-Whan Oh

Abstract

Fuzzy clustering has been playing an important role in solving many problems. Fuzzy c-Means(FCM) algorithm is most frequently used for fuzzy clustering. But some fixed point of FCM algorithm, known as Tucker's counter example, is not a reasonable solution. Moreover, FCM algorithm is impossible to perform the on-line learning since it is basically a batch learning scheme. This paper presents unsupervised learning networks as an attempt to improve shortcomings of the conventional clustering algorithm. This model integrates optimization function of FCM algorithm into unsupervised learning networks. The learning rule of the proposed scheme is a result of formal derivation based on the gradient descent procedure of a fuzzy objective function. Using the result of formal derivation, two algorithms for fuzzy cluster analysis, the batch learning version and on- line learning version, are devised. They are tested on several data sets and compared with FCM. The experimental results show that the proposed algorithms find out the reasonable solution on Tucker's counter example.

I. Introduction

Cluster analysis is based on partitioning a collection of data points into a number of clusters, where the data points inside a cluster have a certain degree of similarity. It has been playing an important role in solving many problems in pattern recognition and image processing as a fundamental process of data analysis. Many algorithms depending on distance criteria have been developed accomplish this[1] Babu94 The principal goal of the neural network research for cluster analysis well known as self organizing feature map(SOFM)[3] is to discover crisp clusters of overlapping input. It assigns each input vector to one and only one of the clusters, assuming well defined boundaries between the clusters. This clustering method by hard partitioning acts as a winner take all fashion. So this method often does not reflect the description of real data in the area such as speech understanding and image recognition, where boundaries might be fuzzy.

To have a more useful information, the representation of clusters by fuzzy sets may be more appropriate. In this case a pattern does not necessarily belong to just one cluster, but there is a certain degree of possibility that the pattern might belong to each one of the clusters. Bezdek[4] developed a fuzzy clustering algorithm, fuzzy c-means(FCM), based on fuzzy extension of the least square error. As an attempt to improve the SOFM, the researches which integrate FCM model into SOFM have been performed[5] Tsao94. Huntsberger and Ajjimarangsee[7] attempted to establish a connection between SOFM and fuzzy clustering by modifying the learning rule. This hybrid learning scheme was a first attempt to merge SOFM and fuzzy clustering, but it required choosing several parameters such as learning rate, size of an update neighborhood, and a strategy to alter these two parameters during learning. These parameters must be varied from one data set to another to achieve useful results. Moreover their algorithms are heuristic procedures that are not tied to the optimization of an objective function, and their termination procedure does not guarantee estimates of points having any well defined properties connected to cluster substructure. Bezdek et al. proposed a batch learning scheme, a fuzzy Kohonen clustering(FKCN) algorithm[6], as a alternative to Hunsberger and Ajjimarangsee's attempt. They proved that FKCN with fixed m is equal to FCM. Thus FKCN can be considered as a neural network type implementation of FCM. But this learning scheme was not the result of formal derivation.

In FCM algorithm, the point $u_{ij}=\frac{1}{c}$ for all i, j, one fixed point of the algorithm, is not a meaningful result. It is the counter example of the basic convergence theorem of FCM algorithm found by Tucker[11]. Moreover, since FCM is basically a batch learning scheme, it is impossible to perform the on-line learning. In this paper, a hybrid scheme to improve shortcomings of FCM and SOFM will be presented. This model integrates the optimization function of FCM algorithm into unsupervised learning

Manuscript received September 15, 1997; accepted November 19, 1997.

H. S. Rhee is with the department of Computer Science and Management Technology, Dongyang Technology Junior College, Seoul, Korea.

K. W. Oh is with the department of Computer Science, Sogang University, Seoul Korea.

network. The learning rule of the proposed scheme is a result of formal derivation based on gradient descent procedure of fuzzy objective function. Using the result of formal derivation, two algorithms for fuzzy cluster analysis, the batch learning version and on-line learning version, are devised. They are tested on several data sets and compared with FCM. The experimental results show that the proposed scheme finds out the reasonable solution on Tucker's counter example.

This paper is organized as follows : Section II presents an introduction to FCM algorithm which is most frequently used for fuzzy clustering and Tucker's counter example. In Section III, we describe the process of formal derivation based on gradient descent procedure and two learning rules for batch learning version and on-line learning version. In Section IV, we describe the experimental results on two data sets to show the validity of this model. Finally, Section V is for the conclusions and summary of this research.

II. The FCM Algorithm

Fuzzy theory was first introduced by Zadeh in 1965[8]. Fuzzy concepts can yield more accurate representations of data structures. Bezdek developed the fuzzy c-means algorithm(FCM) in 1980 based on iterative optimization of the least square error functional J_m[4]. The fuzzy c-means functional J_m[4] is as follows:

$$J_m=\frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{c}(u_{ij})^m(d_{ij})^2 \qquad (1)$$

where U is a fuzzy c-partition of a data set, and the entry u_{ij} is the fuzzy membership value of data j in cluster i, $1\le i\le c$, $1\le j\le n$, which satisfy

$$u_{ij}\in[0,1],\ i\le i\le c,\ i\le j\le n$$

$$\sum_{i=1}^{c}u_{ij}=1;\ 1\le j\le n$$

$$\sum_{j=1}^{n}>0;\ 1\le i\le c$$

$d_{ij}^2=\|x_j-v_i\|$, where $\|\cdot\|$ any inner product norm metric, and weighting exponent $m\in[1,\infty)$. This leads to infinite families of fuzzy clustering algorithms and it is broadly applied to optimal fuzzy partition, pattern classification and image segmentation etc. The FCM algorithm, via iterative optimization J_m, produces a fuzzy c-partition of the data set $X=\{x_1,\dots,x_n\}$. The basic steps of the algorithm are given as follows.

step 1. Fix the number of clusters c, $2\le c\le n$, where n is the number of data items. Fix m, $i< m<\infty$. Choose any inner product introduced norm metric $\|\cdot\|$. Initialize $U^{(0)}$ in fuzzy c-partition space. Then at step b, b = $0, 1, 2, \dots$,

step 2. Calculate the c cluster centers (v,(b)) with U(b) and the formula for the ith cluster center :

$$v_{il}=\frac{\sum_{j=1}^{n}(u_{ui})^m x_{jl}}{\sum_{j=1}^{n}(u_{ij})^m},\ l=1,2,\dots,k. \qquad (2)$$

step 3. Update $U^{(b)}$: calculate the memberships in $U^{(b+1)}$ as follows

For $j=1$ to n,

a) Calculate I_j and $\overline{I_j}$:

$I_j=\{i\mid 1\le i\le c, d_{ij}=\|x_j-v_i\|=0\}$,

$\overline{I_j}=\{1,2,\dots,c\}-I_k$;

b) For data item j, compute new membership values:

i) If $I_j=\varnothing$,

$$u_{ij}=\frac{1}{\sum_{s=1}^{c}\left(\frac{d_{ij}}{d_{sj}}\right)^{2/(m-1)}} \qquad (3)$$

ii) Else $u_{ij}=0$ for all $i\in\overline{I_j}$ and $\sum_{i\in I_i}=1$;

step 4. Compare $U^{(b)}$ and $U^{(b+1)}$ in convenient matrix norm; if $\|U^{(b)}-U^{(b+1)}\|<\varepsilon$ Stop; otherwise, set $b=b+1$, and go to step 2.

Use of the FCM algorithm requires determination of several parameters, i.e., c,ε,m, and the set of $U^{(0)}$ of initial cluster centers must be defined. Particularly in case of the point $u_{ij}^{(b)}=\frac{1}{c}$ for all i, j, we have a number of interesting properties. To begin with,

$$v_i=\frac{\sum_{j=1}^{n}(1/c)^m x_j}{\sum_{j=1}^{n}(1/c)^m}=\frac{\sum_{j=1}^{n}x_j}{n}=\bar{x},\ \text{for } all\ i$$

so that $d_{ij}^2=\|x_j-vi\|^2=\|x_j-\bar{x}\|^2=d_j^2$, say. Then,

$$u_{ij}^{(b+1)}=1/\sum_{l=1}^{c}\left(\frac{d_{lj}}{d_{ij}}\right)^{2/(m-1)}=1/\sum_{i=1}^{c}\left(\frac{d_j}{d_j}\right)^{2/(m-1)}=1/c$$

and thus the point meats the necessary conditions for a local minimum of J_m, and it is a fixed point of the FCM algorithm. But under certain conditions, it is not a local minimum. It is the counter example of the basic convergence theorem of FCM algorithm found by Tucker[11].

III. Fuzzy clustering Algorithms Based On Gradient Descent Procedure

Let R be the set of reals, R_k the set of k dimensions of reals, weight vector V_{ck} the set of real $c\times k$ matrices, where c is the number of clusters, input vector X_{nk} the set of $x=(x_{j1},\dots,x_{jk})\in R_k, j=1,\dots n$, where n is the number of input data and fuzzy c-partition space U_{cn} the set of membership degrees in each cluster of each input vector.

The learning rule of the proposed scheme is derived by gradient descent method on the objective functional, J_m of Eq.(1). The basic procedure of the gradient descent method is that starting from an initial center vector v, the gradient $\triangle J_m$ of objective function is computed. The next value of v is obtained by moving in the direction of the negative gradient along the multidimensional error surface. This procedure is summarized as follows :

$$\triangle v_{i,t} = -\eta_i \frac{\delta j_m}{\delta v_{i,t}}$$

For fixed j, $j=1,\dots,n$, let $d_{ij}^2 = w_i$ and

$$(u_{ij}{}^m) = \frac{1}{\sum_{s=1}^{c}\left(\frac{w_i}{w_s}\right)^{m/(m-1)}} = F_i(w_1,\dots,w_c)$$

Then,

$$J_m = \frac{1}{2}\sum_{j=1}^{n}\{(u_{ij})^m d_{ij}^2 + \cdots + (u_{cj})^m d_{cj}^2\}$$
$$= \frac{1}{2}\sum_{j=1}^{n}\{F_1(w_1,\cdots,w_c)w_1 + \cdots + F_c(w_1,\cdots,w_c)w_c\}$$

Let $F_{il}(w_1,\cdots,w_c)$ is the $l$the partial derivative of $F_i(w_1,\cdots,w_c)$. It is computed by following.

$$F_{il}(w_1,\cdots,w_c) = \frac{\delta F_i(w_1,\cdots,w_c)}{\delta w_l}$$
$$= m(u_{ij})^{m-1}\frac{\delta u_{ij}}{\delta w_l}$$
$$= m(u_{ij})^{m-1}\left[-\frac{\frac{\delta}{\delta w_l}\left(\sum_{s=1}^{c}\left(\frac{w_i}{w_s}\right)^{\frac{1}{m-1}}\right)}{\left(\sum_{s=1}^{c}\left(\frac{w_i}{w_s}\right)^{\frac{1}{m-1}}\right)^2}\right]$$
$$= m(u_{ij})^{m-1}(u_{ij})^2\left[-\frac{\delta}{\delta w_l}\sum_{s=1}^{c}\left(\frac{w_i}{w_s}\right)^{\frac{1}{m-1}}\right] \quad (4)$$

$$\frac{\partial}{\partial\omega_l}\sum_{s=1}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}} = \frac{\partial}{\partial\omega_l}\left|\left(\frac{\omega_i}{\omega_1}\right)^{\frac{1}{m-1}} + \cdots + \left(\frac{\omega_i}{\omega_l}\right)^{\frac{1}{m-1}} + \cdots + \left(\frac{\omega_i}{\omega_c}\right)^{\frac{1}{m-1}}\right| \quad (5)$$

If $l=i$

$$\frac{\partial}{\partial\omega_l}\sum_{s=1}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}} = 1l\,\frac{!d}{m-1}\sum_{s=1,s\neq i}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}-1}\left(\frac{1}{\omega_s}\right) \quad (6)$$

If $l \neq i$

$$\frac{\partial}{\partial\omega_l}\sum_{s=1}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}} = \frac{\partial}{\partial\omega_l}\left(\frac{\omega_i}{\omega_l}\right)^{\frac{1}{m-1}}$$
$$= \frac{1}{m-1}\left(\frac{\omega_i}{\omega_l}\right)^{\frac{1}{m-1}-1}\left(-\frac{\omega_i}{\omega_l^2}\right) \quad (7)$$

From Eq. (5),(6),(7), We obtain

$$F_{il} = \frac{m}{m-1}(u_{ij})^{m+1}\left(\frac{\omega_i}{\omega_l}\right)^{\frac{1}{m-1}}\left(\frac{1}{\omega_l}\right) : l \neq i$$

$$\frac{m}{m-1}(u_{ij})^{m+1}\sum_{s=1,s\neq i}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}}\left(\frac{1}{\omega_l}\right) : l = i \quad (8)$$

$$\frac{\partial\omega_i}{\partial V_{i,t}} = -2(x_j - v_{i,t}) \quad (9)$$

Therefore,

$$\frac{\partial J_m}{\partial V_{i,t}} = \frac{1}{2}\sum_{j=1}^{n}\{\omega_1 F_{1i}(\omega_1,\cdots,\omega_c) + \cdots + \omega_i F_{ii}(\omega_1,\cdots,\omega_c)$$
$$+ F_i(\omega_1,\cdots,\omega_c) + \cdots + \omega_c F_{ci}(\omega_1,\cdots,\omega_c)\frac{\partial\omega_i}{\partial V_{i,t}}$$
$$= \frac{1}{2}\sum_{j=1}^{n}\frac{m}{m-1}\left\{(u_{1j})^{m+1}\left(\frac{\omega_1}{\omega_i}\right)^{\frac{1}{m-1}+1} + \cdots +\right.$$
$$(u_{ij})^{m+1}(-1)\sum_{s=1,s\neq i}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}}$$
$$\left. + (u_{1j})^m + \cdots + (u_{cj})^{m+1}\left(\frac{\omega_c}{\omega_i}\right)^{\frac{1}{m-1}+1}\right\}\frac{\partial\omega_i}{\partial V_{i,t}} \quad (10)$$

Then,

$$(u_{ij})^{m+1}(-1)\sum_{s=1,s\neq i}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}} + (u_{ij})^m$$
$$= (u_{ij})^m(1-u_{ij})\sum_{s=1,s\neq i}^{c}\left(\frac{\omega_i}{\omega_s}\right)^{\frac{1}{m-1}}$$
$$= (u_{ij})^m\left(1-u_{ij}\left(\frac{1}{u_{ij}}-1\right)\right) = (u_{ij})^{m+1}$$

Thus

$$\frac{\partial J_m}{\partial v_{i,t}} = \frac{1}{2}\frac{m}{m-1}\sum_{j=1}^{n}\left\{(u_{1j}\left(\frac{w_1}{w_i}\right)^{\frac{m}{m-1}} + \dots +\right.$$
$$\left.(u_{ij})^{m+1} + \dots + (u_{cj})^{m+1}\left(\frac{w_c}{w_i}\right)^{\frac{m}{m-1}}\right\}\frac{\partial w_i}{\partial v_{i,t}}$$
$$= -\frac{m}{m-1}\sum_{j=1}^{n}\sum_{l=1}^{c}\left\{(u_{lj})^{m+1}\left(\frac{w_l}{w_i}\right)^{\frac{m}{m-1}}\right\}(x_j - v_i) \quad (11)$$

Let

$$a_{ij} = \frac{m}{m-1}\left\{\sum_{l=1}^{c}(u_{lj})^{m+1}\left(\frac{w_l}{w_i}\right)^{\frac{m}{m-1}}\right\} \quad (12)$$

Thus

$$\triangle v_{i,t} = \eta_i\sum_{j=1}^{n}a_{ij}(x_j - v_{i,t}) \quad (13)$$

Based on this derivation, we devise the batch learning algorithm, which updates the weight with respect to all inputs, called by FCGD-B. This algorithm can be summarized as follows.

step 1. Set the value of c, m and ε. Prepare input data set.

step 2. Initialize weight vector, $v_0 = (v_{1,0},\dots,v_{c,0}) \in R_{ck}$ as random numbers. Set t=0.

step 3. Compute the fuzzy membership value u_{ij} for $1 \le i \le c$, $1 \le j \le n$, using the method of the FCM algorithm.

step 4. Compute a_{ij} with Eq. (12) for $1 \le i \le c, 1 \le j \le n$ and set
$$\eta_i = \frac{1}{\sum_{j=1}^{n}a_{ij}} \text{ for } 1 \le i \le c.$$

step 5. Update weight vectors $v_{i,t}$ for $i=1,\dots,c$ *with*

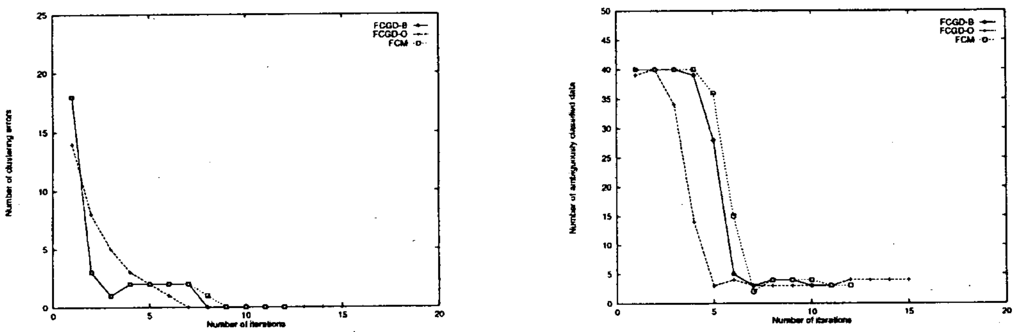


Fig. 1. Clustering the Chiou's data using FCGD-B, FCGD-O and FCM algorithm (a) Number of clustering errors (b) Number of ambiguously classified data.

$$\begin{aligned} v_{i,t+1} &= v_{i,t} + \Delta v_{i,t} \\ &= v_{i,t} + \eta_i \sum_{j=1}^{n} a_{ij}(x_j - v_{i,t}). \end{aligned}$$

step 6. Compute $diff = \sum_{i=1}^{c} \| v_{i,t+1} - v_{i,t} \|^2$.

step 7. If $diff < \varepsilon$, stop.
Else $t \leftarrow t+1$ and go to step 3

The learning rule derived by this method is summarized as follows:

$$\begin{aligned} v_{i,t+1} &= v_{i,t} + \eta_i \sum_{j=1}^{n} a_{ij}(x_j - v_{i,t}) \\ &= (1 - \eta_i \sum_{j=1}^{n} a_{ij}) v_{i,t} + \eta_i \sum_{j=1}^{n} a_{ij} x_j \\ &= \frac{\sum_{j=1}^{n} a_{ij} x_j}{\sum_{j=1}^{n} a_{ij}} \end{aligned} \quad (14)$$

Compared with center update equation in the iterative process of FCM, we notice that $(u_{ij})^m$ of Eq. (3) replaced by a_{ij} in this method.

Since this scheme is basically a batch learning scheme such as FCM, it is impossible to perform the on-line learning. For this, we device the on-line learning version using the above result, called FCGD-O, which update the weight with respect to each data. This algorithm can be summarized as follows.

step 1. Set the value of c, m and ε. Prepare input data set.

step 2. Initialize weight vector, $v_0 = (v_{1,0}, \cdots, v_{c,0}) \in R_{ck}$ as random numbers. Set t=1.

step 3. Save the current weight $v_{i,t}$ to $v_{i,t-1}$.

step 4. For each input $x_j, j=1, \cdots, n$,

a) Compute the fuzzy membership value u_{ij} for $1 \le i \le c$, $1 \le j \le n$, using the method of the FCM algorithm.

b) Compute a_{ij} with Eq.(12), for $i=1, \cdots, c$.

c) Update weight vectors $v_{i,t}$ for $i=1, \cdots, c$ using the rule:

$v_{i,t} = v_{i,t-1} + \eta a_{ij}(x_j - v_{i,t-1})$,

where $\eta = \frac{1}{n}$.

step 5. Compute $diff = \sum_{i=1}^{c} \| v_{i,t} - v_{i,t-1} \|^2$

step 6. If $diff < \varepsilon$, stop. Else $t \leftarrow t+1$ and go to step 3.

Ⅳ. Experimental Studies

To show the validity of the proposed method, we prepare Chiou's data, Anderson's Iris data, which is commonly used to test clustering algorithms. The performance of the algorithm is evaluated by counting the number of crisp clustering errors and the number of data points to be considered ambiguous. The formula to decide whether data point x, is considered ambiguous is as follows:
Let

$$u_{lj} = \max_i \{ u_{ij} \}, \quad u_{sj} = \max_{i \neq l} \{ u_{ij} \}$$

If

$$\frac{u_{lj} - u_{sj}}{u_{lj}} \le \frac{1}{c}, \quad (16)$$

then the classification for that point is considered ambiguous. That is, if the difference between the largest and the second largest of fuzzy membership values $u_{ij}, i=1, \cdots, c$ for data point x, is not enough, the data point x, is considered to be ambiguously classified. The proposed algorithm are compared with those of FCM algorithm that is most frequently used for fuzzy clustering. In this experiments, we set $\varepsilon = 0.001$ and $m=2$.

Chiou's data set consists of 40 two dimensional vectors. It has two clusters, but the classification in the border of the clusters in not quite clear. The results of running the three algorithms, FCGD-B, FCGD-O, and FCM, appear in Figure 1. We note that three algorithm generate the similar curve and they terminate rapidly at a very predictable solution(no errors). But it show that FCGD-B converges a little faster than FCM.

Anderson's Iris dta set contains three classes of 50 four dimensional vectors, where each class refers to a type of iris plant. One class is linearly separable from the others and the latter are not linearly separable from each other. This data set has been used

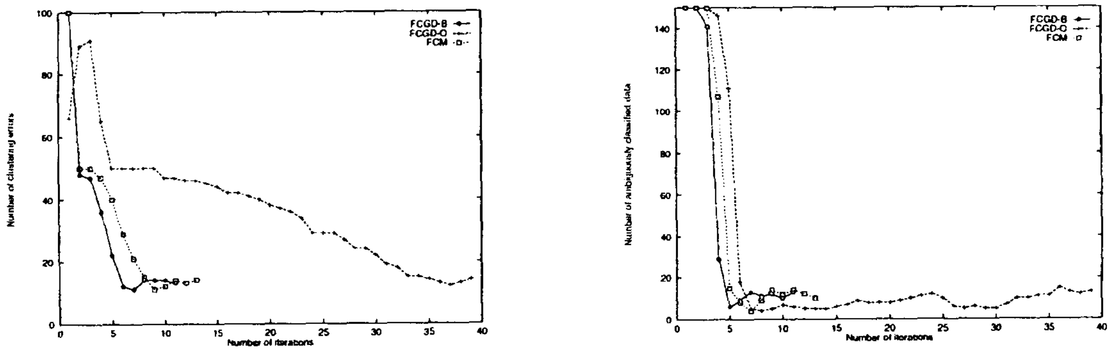


Fig. 2. Clustering the Iris data using FCGD-B, FCGD-O and FCM algorithm (a) Number of clustering errors (b) Number of ambiguously classified data.

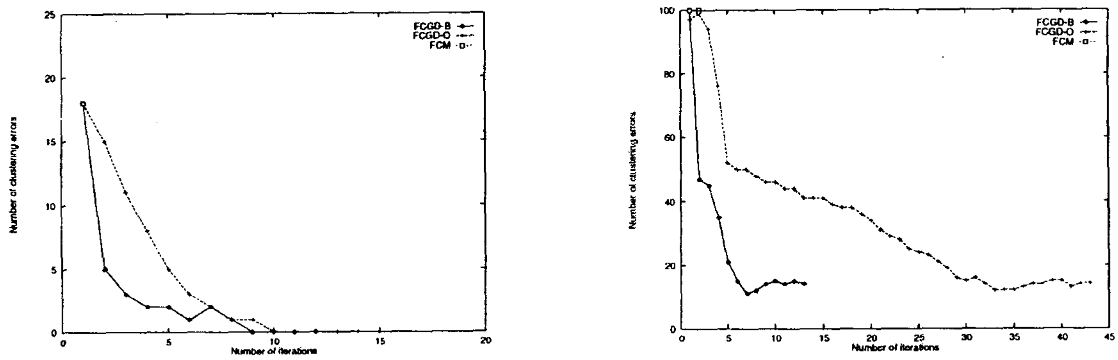


Fig. 3. Clustering the Tucker's counter example (a) Number of clustering errors for Chiou's data (b) Number of clustering errors for Iris data.

many papers to illustrate various unsupervised clustering and supervised classifier designs. Typical error rates for supervised designs are 0.5-mistakes around 15 mistakes. Figure 2 show the results for data set. We note that FCGD-B and FCM generate the similar curve and they terminate in 20 iterations at a very predictable solutions(10-15 errors). But it shows that FCGD-B converges faster than FCM. We also note that FCGD-O generates the slower curve than other algorithms and it terminates around 40 iterations.

To show the advantages of the proposed method over FCM algorithm, we test them with the counter example to the convergence theorem for the FCM algorithm found by Tucker[11]. Figure 3 shows the experimental result for this case using Choiu's data and Iris data. We note that FCGD-B and FCGD-O work correctly and they have similar result to the general case in Figure 1-2. But FCM terminates at an initial state in two iterations.

Ⅴ. Conclusions

Fuzzy clustering has been playing an important role in solving many problems in the area such as speech understanding and image recognition, where boundaries might be fuzzy. In this paper we have presented an unsupervised learning network for the fuzzy clustering based on the gradient descent procedure of a fuzzy objective function. We have described the process of the formal derivation and devised two learning rules for batch learning version and on-line learning version using the result of formal derivation. The proposed scheme is not a heuristic procedure since it is tied to optimization criterion of FCM algorithm. So it does not suffer from several problems of conventional learning network devised on the basis of intuitive arguments (e.g. termination is not guaranteed, no model is optimized by the learning strategy). We have tested two versions of this scheme, FCGD-B and FCGD-O, and compared them with FCM algorithm. In the experiment on Tucker's counter example, we have shown that the proposed algorithms have similar results to the general case, but FCM terminates at an initial state in two iterations.

Acknowledgement

This research was supported by MOST Soft Science Research Project entitled A Study on Soft Inference Systems for Large Scale Knowledge-based Agents.

References

[1] R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[2] Babu GP, Murty MN, "Clustering with Evolution Strategies,"

Pattern Recognition, Vol. 27, No. 2, pp. 321-329, 1994.

[3] T. Kohonen, Self-Organizing Maps Berlin, Springer-Verlag, 1995.

[4] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum New York, 1981.

[5] N. B. Karayiannis and P. I. Pai PI, "A Fuzzy Algorithm for Learning Vector Quantization", it Proc. 1994 IEEE Int. Conf. on System, Man and Cybernetics, Texas, pp. 126-131, 1994.

[6] E. C. Tsao, J. C. Bezdek and N. R. Pal, "Fuzzy Kohonen Clustering Networks", Pattern Recognition, Vol. 27, No. 5, pp. 757-764, 1994.

[7] Huntsberger TL, Ajjimarangsee P, "Parallel Self-Organizing Feature Maps for Unsupervised Pattern Recognition", Int'l, J. General Systems, Vol. 16, No. 4, pp. 357-372,1990.

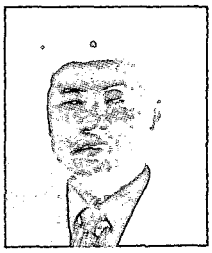
[8] L. A. Zadeh, "Fuzzy Sets", Information and Control. Vol. 8, 1965.

[9] T. L. Huntsberger, C. L. Jacobs and R. L. Cannon, "Iterative Fuzzy Image Segmentation", Pattern Recognition, Vol. 18, No. 2, 1985.

[10] R. L. Cannon, J. V. Dave, J. C. Bezdek, "Efficient Implementation of the Fuzzy c-Means Clustering Algorithms", IEEE Trans. on Pattern Analysis and machine Intelligence, Vol. PAMI-8, No. 2, pp. 248-255, 1986.

[11] J. C. Bezdek, R. J. Hathaway, M. J. Sabin and W. T. Tucker, "Convergence Theory for Fuzzy c-Means: Counter examples and Repairs", IEEE Trans. Syst., Man, Cybern., Vol. 17, No. 5, 1987.

Hyun-Sook Rhee received her B.S. degree in Computer Science from Sogang University in 1989, the M.S. degree in Computer Science from Pohang Technology University in 1991 and the Ph. D. degree in Computer Science from Sogang University in 1997. From 1991 to 1993, she worked as a member of technical staff at ETRI, Korea. Currently, she is an assistant professor at the department of Computer Science and Management Technology, Dongyang Technology Junior College, Seoul, Korea. Her major research interests include Image Information Processing, Fuzzy Neuro Modelling and Software Development Methodologies.



Kyung-Whan Oh received his B.S. degree in Mathematics from Sogang University, Seoul, Korea, in 1978. He did the M.S. degree and Ph.D. in Computer Science from Florida State University, U.S.A., 1985 and 1988, respectively. Since 1989, he has worked as a professor at Computer Science Department in Sogang University, Seoul, Korea. His major research interests include Fuzzy Expert Systems, Intelligent Agent Systems, Neural Networks, and Face Recognition.