Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

Post-supervised Fuzzy c-Means Classifier with Hard Clustering

Hidetomo Ichihashi, Katsuhiro Honda, Naho Kuwamoto, and Takao Hattori
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531 Japan

***Abstract*— A fuzzy c-means classifier (FCMC) based on a generalized fuzzy c-means clustering with iteratively reweighted least square technique (IRLS-FCM) has been proposed. In this paper, we derive a generalized hard c-means (HCM-g) clustering algorithm by defuzzifying IRLS-FCM. Many hard clustering results are obtained from local minima of the HCM-g objective function. Although HCM-g is not a fuzzy clustering algorithm, it is applied to a fuzzy classifier and the best values of the parameters such as the fuzzifiers were chosen by using golden section search method.**

Whereas the goal of FCMC is to minimize classification error rate on unseen new test data, the proposed classifier aims at minimizing resubstitution error rate by using only a small number of clusters. The proposed classifier with two clusters for each class achieves low resubstitution error rate on several benchmark data sets.

I. Introduction

In this paper we first generalize the standard fuzzy c-means clustering objective function [1] a little further, then propose a defuzzified hard clustering and apply it to a post-supervised classifier with fuzzy classification functions. The proposed classifier is based on the hard clustering, though the classification is done using fuzzy membership function.

There are four types of basic ideas representing clusters, i.e., crisp, probabilistic, fuzzy, and possibilistic. Examples of alternating optimization algorithms of clustering that can generate memberships to clusters as well as a set of cluster centroids from unlabeled object data are hard c-means (HCM), Gaussian mixture models (GMM) or normal mixture [2], fuzzy c-means (FCM) [1], and possibilistic c-means [3].

The entropy regularized FCM (FCM-e) [4] has a close relationship to GMM or deterministic annealing by Rose [5]. The difference between the standard FCM (FCM-s) and FCM-e comes from the difference of their membership functions. The aim of the generalized FCM (FCM-g) [6] is to alleviate the singularity of the membership function and to equip both of the properties of FCM-s and FCM-e.

Miyamoto *et al.* [7] proposed a generalized hard c-means clustering (HCM-g) by introducing Mahalanobis distances. The approach is originated from the FCM clustering with regularization by KL-information (FCM-K). FCM-K is a special case of FCM-e. Though the FCM-K is similar to the statistical clustering method known as GMM, its representation of the objective function is rather simple and does not strictly follow the EM algorithm and Bayes' rule. This reinterpretation of the statistical clustering method may lead to general FCM objective function, but it is still limited to a few models of fuzzy clustering.

Various membership functions different from those in FCM-s and FCM-e can be used in an FCM clustering algorithm with the iteratively reweighted least square (IRLS) technique [8]. Cluster memberships are defined by a function of Mahalanobis distances or Euclidean distances between data vectors and cluster centers. The algorithms of GMM, FCM-e and FCM-K are the special cases of IRLS fuzzy c-means clustering (IRLS-FCM). The algorithm is applied to a classifier design [9], [10], [11] and is called FCM classifier (FCMC). The revised algorithm with deterministic initialization in [12] is also abbreviated to FCMC. The classifier improved the classification performance in terms of the generalization ability (classification accuracy on test sets) and the receiver operating characteristics [9], [10], [11], [12] for several benchmark data sets.

In supervised classifier design, a data set is usually crisply partitioned into a training set and a test set. Testing a classifier designed with the training set means finding its misclassification rate. The standard method for doing this is to submit the test set to the classifier and count errors. This yields the performance index by which the classifier is judged because it measures the extent to which the classifier generalizes to the test data. When the test set is equal to the training set, the error rate is called the resubstitution error rate. This error rate is not reliable for assessing the generalization ability of the classifier, but this is not an impediment to using as a basis for comparison of different designs. If training set is large enough and its substructure is well delineated, and if the number of clusters used in the classifier is small, we expect classifiers trained with it to yield good generalization ability or it may not deteriorate.

The clustering phase of IRLS-FCM is replaced with HCM-g, and the cluster centers and the covariance matrices are determined by the HCM-g. In this post-supervised design, the clustering is implemented by using the data from one class at a time, i.e., the clustering is done on a per class basis. When working with the data class by class, the prototypes (cluster centers) that are found for each labeled class already have the assigned physical labels. HCM-g is implemented in the unsupervised phase, and then the parameters in the membership function such as the fuzzifiers are chosen so that the resubstitution error rate attains minimum in the supervised phase. Whereas the goal of FCMC is to optimize generalization ability of the classifier, the proposed HCM-g classifier (HCMGC) aims at minimizing resubstitution error rate. Clustering is known as a combinatorial optimization problem and the HCM-g algorithm produces many local minima, from which we can choose to minimize resubstitution errors. The strategy for classification is not just based on the hard clustering results but also on fuzzy memberships.

The classification algorithms using Mahalanobis distances should include steps to check that the covariance matrices are nonsingular and hence invertible. The way of handling singular covariance matrices in the mixture of probabilistic principal component analysis (MPCA) [13] or character recognition [14] is employed to prevent unexpected termination of HCM-g and improve classification accuracy of the algorithm.

In this paper we focus our discussion to the resubstitution classification error rate and the data set compression ratio as performance criteria. Please refer to [12] for more details of FCMC and its generalization ability. Other techniques such as feature selection issues are beyond our consideration here.

The trained classifiers are tested on the benchmark data sets from the UCI ML repository (http://www.ics.uci.edu/˜mlearn/) [15]. HCMGC with small number of clusters shows relatively low classification error rates on several data sets. Also concerning storage requirements and classification speed, the HCM-g classifier gives a good performance and efficiency.

The paper is organized as follows. Section II gives a brief description of the generalized FCM clustering and proposes a post-supervised classifier. Section III provides the results of numerical experiments. By using graphs of the classification function, knowledge acquisition from the learned classifier is described in Section IV. Section V concludes the paper.

II. Post-supervised Classifier with Hard c-Means Clustering and Fuzzy Memberships

FCM clustering partitions data set by introducing memberships to fuzzy clusters. The clustering criterion used to define good clusters for fuzzy c-means partitions is the FCM objective function.

A. A Generalization of FCM Clustering

The clustering is used as an unsupervised phase of the classifier design. Let r dimensional vector $\boldsymbol{v}_i$ denote prototype parameter (i.e., cluster centroid). u_{ik} denotes the membership of k-th object data $\boldsymbol{x}_k \in \mathcal{R}^r$ to i-th cluster.

The objective function of the standard method is:

$$J_{\text{fcm}} = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^{\lambda} d_{ik}^2, \quad (\lambda > 1), \tag{1}$$

where c denotes the number of clusters. d_{ik}^2 denotes the squared distance between $\boldsymbol{x}_k$ and $\boldsymbol{v}_i$, so the standard FCM objective function is the weighted sum of squared distances. Taking the objective function for the entropy-based method and the quadratic-term-based method [16] into account, we can generalize the standard objective function a little further as:

$$J_{\text{gfc}} = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ki})^{\lambda} d_{ik}^2 + \eta \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^{\lambda}. \tag{2}$$

where $\eta > 0, \lambda > 1$. From the necessary condition for optimality, we have

$$u_{ik} = \left[\sum_{j=1}^{c}\left(\frac{\eta + d_{ik}^2}{\eta + d_{jk}^2}\right)^{\frac{1}{\lambda-1}}\right]^{-1}. \tag{3}$$

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{n}(u_{ki})^{\lambda}\boldsymbol{x}_k}{\sum_{k=1}^{n}(u_{ki})^{\lambda}}. \tag{4}$$

For more detail descriptions of the derivation and properties, see [6].

The objective function of the entropy term based method [4], [17] is the only case where covariance matrices (A_i) can be taken into account. Although Gustafson and Kessel's modified FCM [18] is derived from an objective function with fuzzifier λ, we need to specify the values of determinant $|A_i|$ for all i.

B. IRLS FCM Clustering

In order to deal with covariance structure within the scope of fuzzy c-means clustering, we need some simplifications based on the IRLS technique. Runkler and Bezdek's [19] fuzzy clustering scheme called alternating cluster estimation (ACE) is this kind of simplification.

Now we employ a technique from the robust M-estimation [8], [20]. The M-estimators try to reduce the effect of outliers by replacing the squared residuals with ρ-function, which is chosen to be less increasing than square. Instead of solving directly this problem, we can implement it as the IRLS. While the IRLS approach does not guarantee the convergence to a global minimum, experimental results have shown reasonable convergence points. If one is concerned about local minima, the algorithm can be run multiple times with different initial conditions.

Let the objective function of the IRLS-FCM be

$$J_{\text{ifc}} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}\left(d_{ik}^2 + \log|A_i|\right). \tag{5}$$

where

$$d_{ik}^2 = (\boldsymbol{x}_k - \boldsymbol{v}_i)^{\top} A_i^{-1}(\boldsymbol{x}_k - \boldsymbol{v}_i) \tag{6}$$

is squared Mahalanobis distance from $\boldsymbol{x}_k$ to i-th cluster centroid. A_i is a covariance matrix of data samples of the i-th cluster, which is derived from (5) as:

$$A_i = \frac{\sum_{k=1}^{n} u_{ik}(\boldsymbol{x}_k - \boldsymbol{v}_i)(\boldsymbol{x}_k - \boldsymbol{v}_i)^{\top}}{\sum_{k=1}^{n} u_{ik}}. \tag{7}$$

and $\boldsymbol{v}_i$ is as:

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{n}(u_{ki})^{\lambda}\boldsymbol{x}_k}{\sum_{k=1}^{n}(u_{ki})^{\lambda}}. \tag{8}$$

To facilitate competitive movements of cluster centroids, we need to define the membership function to be normalized as:

$$u_{ik} = \frac{u_{ik}^*}{\sum_{l=1}^{c} u_{lk}^*}. \tag{9}$$

We confine our discussion to the membership function

$$u_{ik}^* = \frac{\pi_i |A_i|^{-1/\gamma}}{(\eta + d_{ik}^2/0.1)^{1/\lambda}}, \tag{10}$$

then, by (9), u_{ik} in (3) is modified as:

$$u_{ik} = \pi_i |A_i|^{-1/\gamma} \left[\sum_{j=1}^{c} \left(\frac{\eta + d_{ik}^2/0.1}{\eta + d_{ik}^2/0.1} \right)^{\frac{1}{\lambda}} \pi_j |A_j|^{-1/\gamma} \right]^{-1}. \tag{11}$$

u^* is a modified and parameterized multivariational version of Cauchy's weight function in M-estimator or of the probability density function (PDF) of Cauchy distribution.

$$\pi_i = \frac{\sum_{k=1}^{n} u_{ik}}{\sum_{j=1}^{c} \sum_{k=1}^{n} u_{jk}} = \frac{1}{n} \sum_{k=1}^{n} u_{ik}. \tag{12}$$

C. FCM Classifier

After completing the IRLS-FCM clustering for each class, the classification is performed by computing class memberships of unseen test data. The FCM classifier is abbreviated to FCMC. Let α_q denote the mixing proportion of class q, i.e., the $a\ priori$ probability of class q. The class membership of k-th data $\boldsymbol{x}_k$ to class q is computed as:

$$u_{qjk}^* = \pi_{qj} |A_{qj}|^{-1/\gamma} / (\eta + d_{qjk}^2/0.1)^{1/\lambda}, \tag{13}$$

$$\tilde{u}_{qk} = \alpha_q \sum_{j=1}^{c} u_{qjk}^* \Big/ \sum_{s=1}^{Q} \alpha_s \sum_{j=1}^{c} u_{sjk}^*, \tag{14}$$

where c denotes the number of clusters of each class. The denominator in (14) can be disregarded when applied solely for classification.

From objective function of FCM-K, Miyamoto *et al.* [7] derived the generalized hard clustering HCM-g by setting $\lambda = 0$. The HCM-g is a defuzzified clustering algorithm of FCM-K. Similarly, we can derive the same hard clustering algorithm from (11), and we call it HCM-g.

The modification of covariance matrices in the mixture of probabilistic principal component analysis (MPCA) [13] is applied in the IRLS-FCM classifier for preventing singular matrices. Let A_i' denotes an approximation of covariance matrix A_i in (10) as:

$$A_i' = P_i^p (\Delta_i^p - \sigma_i I_p) P_i^{p\top} + P_i (\sigma_i I_r) P_i^\top, \tag{15}$$

where P_i is an $r \times r$ matrix of eigenvectors of A_i. Δ_i =diag$(\delta_{i1}, ..., \delta_{ir})$ is an $r \times r$ diagonal matrix of eigenvalues. r equals the dimensionality of input samples. P_i^p is an $r \times p$ matrix of eigenvectors corresponding to the p largest eigenvalues, where $p < r - 1$. P_i^p is an $r \times p$ matrix and Δ_i^p is a $p \times p$ diagonal matrix. p is chosen so that all A_i's are nonsingular and the classifier maximizes its classification performance. $\sigma_i = (\mathrm{trace}(A_i) - \Sigma_{l=1}^{p} \delta_{il})/(r - p)$ and $P_i(\sigma_i I_r) P_i^\top = \sigma_i I_r$.

When p=0, A_i is reduced to a unit matrix and d_{ik}^2 in (6) is reduced to Euclidean distance. This modification can be used for both the fuzzy and hard clustering to compute distances in (6).

FCMC is a fuzzy approach and post-supervised, and the IRLS clustering phase can be replaced by a hard clustering algorithm. Although the main thesis of Miyamoto *et al.* [7] is the sequential hard clustering algorithm, for simplicity's sake we confine our discussion to its simple batch algorithm of hard clustering. The objective function of the HCM-g is (2) with $\lambda = 1$ or (5) with $\lambda = 0$. The simple HCM classifier uses A_i of unit matrix, and thus, d_{ik} in (6) is reduced to Euclidean distance.

An alternating optimization algorithm of HCM-g [7] is the repetition of (7) through (12) and

$$u_{ik} = \begin{cases} 1 & ;\ i = \arg\min_{1 \le j \le c} d_{jk}^2 + \log|A_j| \\ 0 & ;\ \text{otherwise} \end{cases} \tag{16}$$

The modification of covariance matrices by (15) is not enough for preventing singular matrices when the number of instances included in a cluster is very small or zero. When the number becomes too small and an A_i results in a singular matrix, for increasing the number we modify (16) as:

$$u_{ik} = \begin{cases} 0.9 & ;\ i = \arg\min_{1 \le j \le c} d_{jk}^2 + \log|A_j| \\ \frac{0.1}{c-1} & ;\ \text{otherwise} \end{cases} \tag{17}$$

By this fuzzification of membership, even the smallest cluster may include some instances with small membership values, and the centroids come somewhat near to the global center of the class. This fuzzification is used for a benchmark data set in section III.

Figs.1-2 show clustering results of artificial 2-D data. HCM-g produces many different results for a nonseparate data set as shown in Fig.1. Five different clustering results are obtained by 10 trials of HCM-g, while the result similar to the one shown in Fig.2 was obtained 9 times out of 10 trials by GMM. Fig.3 show the result obtained 9 times out of 10 trials by IRLS-FCM with $\lambda = 0.6, \gamma = 0.5$ and $\eta = 1$. As we apply the classifier to data with more than one class, we usually have many more local minima of the clustering criterion of HCM-g. Convergence speed by HCM-g is much faster than GMM and FCM-g. HCM-g needs only around 10 iterations, while GMM and FCM-g usually need around 50.

Our proposed classifier is of post-supervised and, thus, the optimum clustering result with respect to the objective function does not guarantee the minimum classification error. Our strategy is to select the best one in terms of classification error from many local minima of the clustering criterion of HCM-g. Parameter values used for HCMGC are chosen by the golden section search [21], which is applied to λ, η, and γ one after another with random initializations.

HCMGC algorithm with golden section search method used in the next section is as follows:

Algorithm: HCMGC

Step 1: Initialize $\boldsymbol{v}_i$'s by choosing data vectors randomly for each class.

Step 2: Partition the training set by HCM-g and fix A_i and $\boldsymbol{v}_i$ of each cluster for each class.

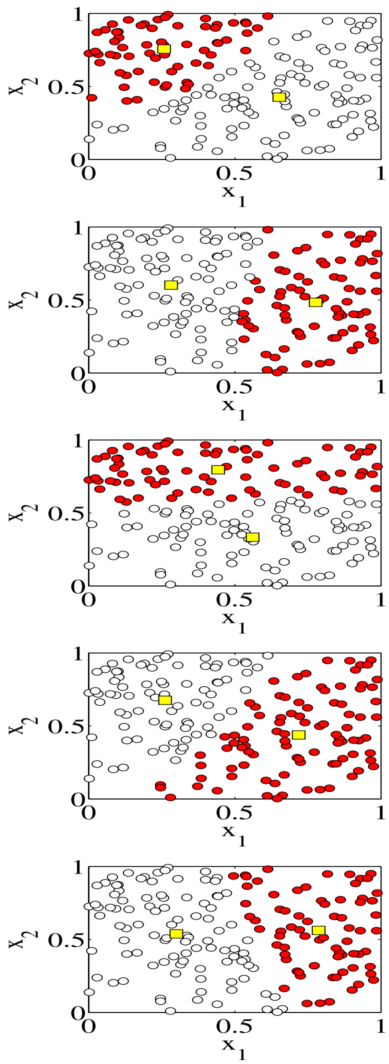

Fig. 1. Five different clustering results observed by 10 trials of HCM-g

Step 3: Choose γ and η randomly from interval [0.01 5].
Step 4: Optimize λ by the golden section search in interval [0.01 5].
Step 5: Optimize γ by the golden section search in interval [0.01 5].
Step 6: Optimize η by the golden section search in interval [0.01 5].
Step 7: If iteration $t < 500$, $t := t + 1$, go to *Step 1* else terminate.

III. Numerical Comparisons

We used 8 data sets of Iris plant, Wisconsin breast cancer, Ionosphere, Glass, Liver disorder, Pima Indian diabetes, Sonar and Wine as shown in Table I. These data sets are available from the UCI ML repository (http://www.ics.uci.edu/~ mlearn/) and were used in [22] to compare the generalization ability of various prototype-based classifiers such as k-nearest neighbor (k-NN), hard c-means, and learning vector quantization (LVQ) [23]. Incomplete samples in the breast cancer data set were eliminated. All attribute values of each data set were normalized to zero mean and unit variance.

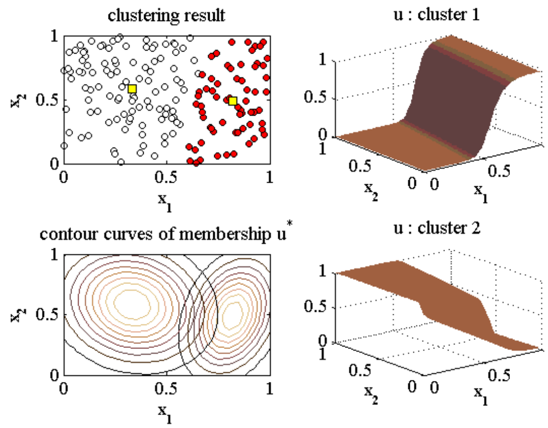

Fig. 2. Result observed 9 times out of 10 trials by GMM

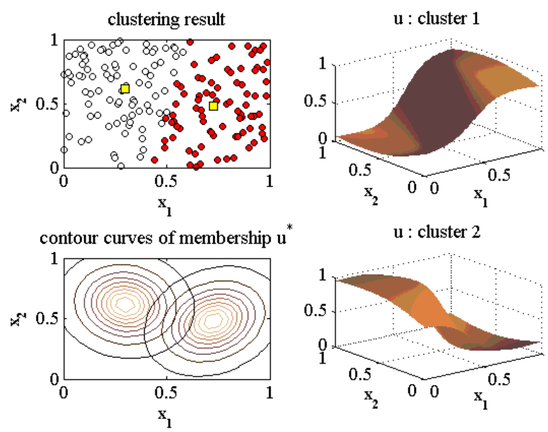

Fig. 3. Result observed 9 times out of 10 trials by IRLS-FCM

Generalization ability of the revised FCMC algorithm is compared with the well established classifiers in [12]. Classification error rates by 10-fold cross validation (10-CV) with a default partition are shown in Table II. The standard deviation is displayed with $\pm$ for LVQ since the classifier is tested by 10 complete runs of 10-CV with random initializations. Initializations for FCMC and k-NN is deterministic.

For the parameters of k-NN (k) and LVQ (c), we tested all integer values from 1 to 50. Parameters of FCMC are optimized by 10-CV with a default partition and the golden section search. FCMC outperforms k-NN and LVQ on the benchmark data sets. For more details of FCMC, and comparisons with the support vector machine (SVM) [24], [25] by a MATLAB interface to SVM^{light} [26] and the decision tree approach C4.5 reported in [27], [28], [29], see [12].

In Table III, “HCMGC” column shows the best resubstitution error rates on training sets from a 500 trials of

TABLE I

DATA SETS USED IN THE EXPERIMENTS

	features	instances	classes
Iris	4	150	3
Breast	9	683	2
Ionosphere	33	351	2
Glass	9	214	6
Liver	6	345	2
Pima	8	768	2
Sonar	60	208	2
Wine	13	178	3

TABLE II

CLASSIFICATION ERROR RATES BY 10-FOLD CV WITH A DEFAULT PARTITION.

	FCMC	k-NN	LVQ1
Iris	1.33	2.67	5.40 $\pm$ 0.87
Breast	2.79	2.65	3.16 $\pm$ 0.16
Ionosphere	3.43	13.43	10.60 $\pm$ 0.35
Glass	28.10	27.62	30.10 $\pm$ 1.38
Liver	27.94	32.65	35.24 $\pm$ 1.44
Pima	22.50	23.42	24.11 $\pm$ 0.52
Sonar	10.50	13.50	14.85 $\pm$ 2.07
Wine	0.00	1.76	2.35 $\pm$ 0.00

TABLE III

BEST RESUBSTITUTION ERROR RATES FROM A 500 TRIALS BY FCMC, HCMGC AND LVQ

	FCMC	**HCMGC**			LVQ1	
	M c=2	**M** c**=2**	E c=5	E c=10	E c=5	E c=10
Iris	0.67	**0.67**	1.33	0.67	2.0	1.33
Breast	1.90	**1.76**	2.78	2.05	2.34	1.61
Ionosphere	3.13	**0.85**	5.13	3.70	7.41	5.70
Glass	18.69	**9.81**	18.69	13.08	20.56	18.22
Liver	23.19	**18.84**	25.22	23.48	27.54	21.45
Pima	20.18	**19.40**	20.44	19.14	20.18	18.88
Sonar	5.29	**0**	6.25	0.48	4.81	1.92
Wine	0.00	**0**	0	0	0	0

TABLE IV

PARAMETER VALUES USED FOR HCMGC WITH MAHALANOBIS DISTANCES (c=2)

	p	λ	γ	η
Iris	1	0.29	3.31	4.96
Breast	7	0.05	0.78	2.26
Ionosphere	20	0.04	1.29	1.96
Glass	9	0.01	0.25	4.96
Liver	4	0.27	4.06	2.82
Pima	4	0.16	1.90	4.36
Sonar	9	0.03	1.16	4.96
Wine	1	0.04	1.21	4.96

clustering by HCM-g and the golden section search. “M” and “E” indicate that Mahalanobis and Euclid distances are used respectively. LVQ result is also the best one from 500 trials on each set with random initializations. Initial value of LVQ learning rate β was set as 0.3 and was changed as in [22], i.e., $\beta(t+1) = \beta(t) \times 0.8$ where t (=1, ..., 100) denotes iteration number. The resubstitution error rates of FCMC is the best results from 10 runs of clustering with different λ^* and 50 runs of the golden section search for each clustering result. Since FCMC uses IRLS-FCM, which is not a hard clustering, 10 runs of clustering seem enough.

For $c > 2$, we set $p = 0$, then HCM-g is a simple hard clustering with Euclidean distances. Naturally, as the number c is increased, the resubstitution error rate decreases and for example when $c = 50$ the rate is 1.17% for Breast cancer data. Since Glass data have 6 classes, when c=2 and (16) is used, all trials unexpectedly terminate due to the lack of instances resulting in a singular covariance matrix. By using (17) the algorithm successfully converged.

Despite the continuous increase in computer memory capacity and CPU speed, especially in data mining, storage and efficiency issues become even more and more prevalent. For this reason we also measured the compression ratios of the trained classifiers in Table VI. The ratio is defined as $Ratio$=$(p + 1) \times c \times$ number of classes $\div$ number of instances. The ratios for HCMGC ($c > 2$) and LVQ are the same. For HCMGC with Mahalanobis distances and c=2, the compression ratios of Ionosphere and Glass are high, though the error rate is small in Table III. When p=3 and c=2, the best error rate for Ionosphere is 2.85% and the compression ratio is 4.56%. The error rate for the glass data is 10.28% and the compression ratio is 33.6% when p=5 and c=2. HCMGC demonstrates relatively low compression ratios. Parameter values of HCMGC chosen by the golden section search method are shown in Table IV. HCMGC with Mahalanobis distance and c=2 attains the lowest error rate when $c \leq 5$ as indicated by boldface letters in Table III. The compression ratios of HCMGC is not so good for Ionosphere, Glass and Sonar, though we can conjecture from the results of FCMC that the generalization ability will not deteriorate largely since only two clusters for each class are used.

IV. KNOWLEDGE ACQUISITION FROM THE TRAINED CLASSIFIER

Rule extraction or knowledge acquisition from the obtained fuzzy classifier is a subsidiary goal of our research. We can graphically display rules for classification in an easily understandable form. Fig.4 shows the classification functions for the Iris-Versicolor data in each feature variable around the cluster centroids. Iris-Versicolor is a binary classification problem and the task is to discriminate Versicolor form other two Iris subspecies. All feature values are mean corrected with unit variance and classification functions on each feature variable are listed in order from top to bottom. The locations of the cluster centroids are indicated by the vertical lines in blue. The centroids of cluster 1 and 2 of the iris versicolor subspecies (class 1) are located near the mean of each feature value. Those of the clusters of the other subspecies (i.e., setosa and verginica subspecies) are clearly on the negative or positive sides. Breast cancer data has 9 features and the four clusters are scattered in a high dimensional space as shown in Fig.5. Cluster centroids of the diseased patients (class 1) are all on the negative side and those of the disease free are on the positive side. We can see from Fig.5 that perturbation

TABLE V

PARAMETER VALUES USED FOR HCMGC WITH EUCLIDEAN DISTANCES (c=5, 10)

	c=5, p=0		c=10, p=0	
	λ	η	λ	η
Iris	1.11	0.01	1.08	0.01
Breast	1.16	1.92	1.04	1.92
Ionosphere	1.02	1.92	1.08	1.92
Glass	1.43	0.01	1.15	1.92
Liver	4.85	5.00	1.19	1.92
Pima	1.49	3.82	1.89	1.19
Sonar	1.02	1.92	1.04	1.92
Wine	1.08	1.92	1.23	1.92

TABLE VI

COMPRESSION RATIO (%)

	M c=2	E c=5	E c=10
Iris	8.0	10	20
Breast	4.7	1.5	2.9
Ionosphere	23.9	2.8	5.7
Glass	56.1	14.0	28.0
Liver	5.8	2.9	5.8
Pima	2.6	1.3	2.6
Sonar	19.2	4.8	9.6
Wine	6.7	8.42	16.9

of single feature value does not affect the classification decisions. Fig.6 demonstrates the classification decisions of four patients of breast cancer who are misclassified. Though from the 4-th to 6-th feature variables assume large positive values, patient 1 is classified as disease free. We can see from the graphs that if the value of the 6-th variable is smaller then the decision becomes correct. In the same way, if the 8-th variable for patient 2 is smaller, the patient is classified as diseased.

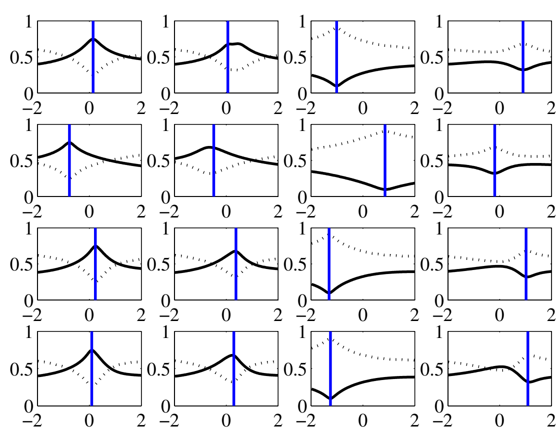

Fig. 4. Classification functions on the Iris-Vc data around each of the cluster centroids. All feature variables are listed in order from top to bottom.

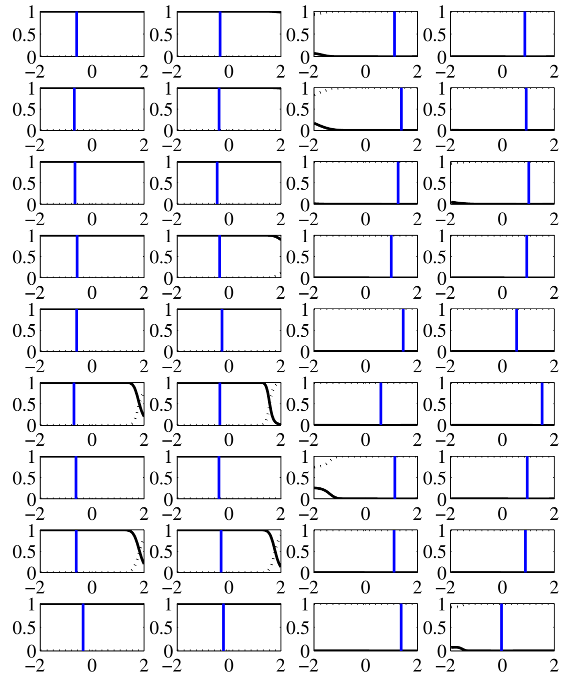

Fig. 5. Classification functions on Breast cancer data around each of the cluster centroids. The graphs on each feature variable are listed in order from top to bottom.

V. CONCLUDING REMARKS

We have applied the generalized hard clustering algorithm with covariance structure to a post-supervised classifier to improve resubstitution error rate by choosing best clustering results from local minima of the clustering criterion. The low resubstitution error rates and data set compression ratios are achieved on several benchmark data sets by HCMGC with c=2. The golden section search is not necessarily the best way for parameter optimization, and our quest for more efficient algorithms continues.

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.

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

[3] R. Krishnapuram and J. Keller, "A Possibilistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems* vol.1, 1993, pp.98-110.

[4] Z.-Q. Liu and S. Miyamoto (Eds.), *Softcomputing and Human-Centered Machines*, Springer-Verlag, 2000.

[5] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proc. of the IEEE*, vol.86, no.11, pp.2210-2239, 1998.

[6] H. Ichihashi, K. Honda, A. Notsu and T. Hattori, "Aggregation of Standard and Entropy Based Fuzzy c-Means Clustering by a Modified Objective Function," *Proc. of the 2007 IEEE Symposium on Foundations of Computational Intelligence*, Hawaii, April, 2007 (to appear).

[7] S. Miyamoto, T. Yasukochi, and R. Inokuchi, "A family of fuzzy and defuzzified c-means algorithms." *Proc. of the International Conference on Computational Intelligence for Modelling, Control and Automation*, Vienna, Austria, pp.170-176, 2005.

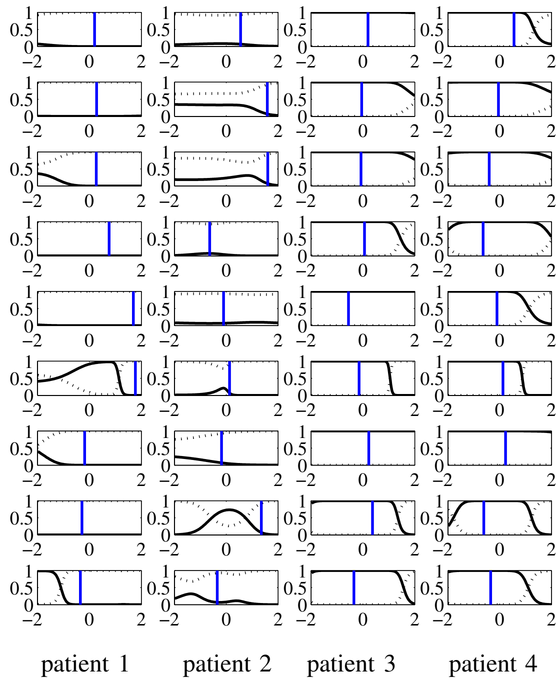

Fig. 6. Classification functions on Breast cancer data. Examples of misclassified patients are indicated by the vertical lines.

[8] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares, "*Communications in Statistics*, vol.A6, no.9, pp. 813-827, 1977.

[9] H. Ichihashi and K. Honda, "Fuzzy *c*-means classifier for incomplete data sets with outliers and missing values," *Proc. of the International Conference on Computational Intelligence for Modelling, Control and Automation*, Vienna, November, pp.457-564, 2005.

[10] H. Ichihashi, K. Honda, and T. Hattori, "Regularized discriminant in the setting of fuzzy *c*-means classifier," *Proc. of the IEEE World Congress on Computational Intelligence*, Vancouver, Canada, pp.4266-4271, 2006.

[11] H. Ichihashi, K. Honda, and F. Matsuura, "ROC analysis of FCM classifier with Cauchy weight," *Proc. of the 3rd International Conference on Soft Computing and Intelligent Systems*, Tokyo, Japan, pp.1912-1917, 2006.

[12] H. Ichihashi, K. Honda, A. Notsu and T. Yagi, "Fuzzy *c*-Means classifier with deterministic initialization and missing value imputation," *Proc. of the 2007 IEEE Symposium on Foundations of Computational Intelligence*, Hawaii, April, 2007 (to appear).

[13] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analysers, "*Neural Computation*, Vol.11, pp.443-482, 1999.

[14] F. Sun, S .Omachi, and H. Aso, "Precise selection of candidates for hand written character recognition," *IEICE Trans. Information and Systems*, vol.E79-D, no.3, pp.510-515, 1996

[15] C.L. Blake and C.J. Merz. UCI repository of machine learning databases, 1998.

[16] S. Miyamoto, D. Suizu, and O. Takata, "Methods of fuzzy *c*-means and possibilistic clustering using a quadratic term, *Scientiae Mathematicae Japonicae* vol.60, no.2, pp.217-233, 2004.

[17] H. Ichihashi, K. Miyagishi, and K. Honda, "Fuzzy c-means clustering with regularization by K-L Information." *Proc. of 10th IEEE International Conference on Fuzzy Systems*, vol. 3, Melbourne, Australia, pp.924-927, 2001.

[18] E. E. Gustafson, and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," *IEEE CDC*, San Diego, California, pp.761-766, 1979.

[19] T. A. Runkler and J. C. Bezdek, "Alternating cluster estimation: a new tool for clustering and function approximation," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 4, pp. 377-393, 1999.

[20] P. J. Huber, *Robust Statistics*. New York:Wiley, first edition, 1981.

[21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing, second edition*, Cambridge University Press, Cambridge, 1999.

[22] C. J. Veenman and M. J. T. Reinders, "The nearest sub-class classifier: a compromise between the nearest mean and nearest neighbor classifier," *IEEE Transactions on PAMI*, vol.27, no.9, pp.1417-1429, 2005.

[23] T. Kohonen, *Self-Organization and Associative Memory*, 3rd Edn. Springer, Berlin, 1989.

[24] V.N. Vapnik, *Estimation of Dependences Based on Empirical Data*. Springer-Verlag, Berlin, 1982.

[25] C. Cortes and V. Vapnik, "Support-vector network." *Machine Learning*, vol.20, pp.273-297, 1995.

[26] T. Joachims, "Making large-scale SVM learning practical," *In: B. Scholkopf et al.(ed.), Advances in Kernel Methods: Support Vector Learning*, MIT Press, Cambridge, pp.169-184, 1999.

[27] J. R. Quinlan, *C4.5, Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 1993.

[28] T. Elomaa and J. Rousu, "General and efficient multisplitting of numerical attributes,"*Machine Learning*, vol. 36, no. 3, pp. 201-244, 1999.

[29] J. R. Quinlan, "Bagging, Boosting, and C4.5,"*Proc. 13th National Conference on Artificial Intelligence, AAAI/IAAI*, vol. 1, pp. 725-730, Portland, Oregon, August, 1996.