Fuzzy Clustering to the Detection of Defects from Nondestructive Testing

Hasanzadeh R.P.R. *, Moradi M.H. ** and Sadeghi S.H.H. *

*Electrical Engineering Department, Amirkabir University of Technology, Hafez Ave., Tehran, 15914, IRAN.
hasanzadeh@aut.ac.ir
sadeghi@aut.ac.ir

** Biomedical Engineering Department, Amirkabir University of Technology, Hafez Ave., Tehran, 15914, IRAN.
mhmoradi@aut.ac.ir

Abstract: In NDT images, defect, generally contains aggregation of special brightness level in its position. In considerable cases, these brightness levels have no notable contrast relative to background of image. Hence detection and reconstruction of defect using automatic or handle threshold levels cause to misfit the result. In order to solve this problem, we used density and size of defect for classification. In this paper, Gath-Geva fuzzy clustering algorithm is obtained for clustering and reconstruction of defects relative to noise and background of images and its result compared with two most popular fuzzy clustering algorithms: fuzzy c-means and fuzzy Gustafson-Kessel algorithms.

Key words: Clustering, Fuzzy, NDT, Defect.

1. Introduction

Today, digital signal processing is an important solution to detect and reconstruct of defects in NDT applications. Several kind of application in neural networks, fuzzy logic and image processing for clustering and segmenting of NDT images have been more attention than the other kind of digital signal processing methods. Upon having knowledge howsoever little of grabbed data, we can use neural network. For instance, perceptron and back-propagation is utilized in order to cluster of defect features. In this method, preprocessing stage is applied for reduction of data dimensions using image processing and self organizing map neural network. Also pattern recognition technique is applied to classify two kinds of defects: crack and delamination, in ultrasonic images (Kim et al., 2001). Several kinds of classes of eddy current data are detected using modeled input signal trained by a neural network, but analytical inverse models are possible only if strong simplifying assumptions are considered (udpa et al., 1991).

In addition fuzzy logic applications are expanded in NDT category. Weld image of x-ray line by line is processed which three features, width, peak intensity and mean square error between the object and its Gaussian are extracted. Extracted features were classified by fuzzy K-NN algorithm (Liao et al., 1997). On the other hand they used trial and error curve fitting for noise reduction and feature extraction. Welding flaw in xray image was detected by fuzzy c-means clustering algorithm (Liao et al., 1999). Fuzzy inference is proposed to enhance contrast of NDT images. This method based on knowledge of intensity and propagation delay in scanned data which is obtained from pulse-radar (Matsumoto et al., 2000). Although wavelet-based texture segmentation is applied for ultrasonic images, ultimately percentage of damage is computed by segmentation using ISODATA algorithm which is crisp kind of fuzzy c-means algorithm (Kim et al., 1998). All of the above methods extracted a number of features or applied methods that were based on knowledge howsoever little of input signals. A glance review of above methods shows that only improving of algorithms can not help us to increasing of efficiency. Searching and finding new features or increasing of knowledge of features about test and defect can help us to enhancing of efficiency. In noisy images and low contrast images, density and size of defect are adequate for classification. By using density and size of clusters, defect, speckle noise and background, various classes can be separated.

2. The Gath-Geva Fuzzy Clustering

Gath-Geva (GG) fuzzy clustering algorithm is an
extension of the Gustafson-Kessel (GK) fuzzy clustering algorithm that also takes the size and density of the clusters into account (Hoppner et al., 1999).

GK algorithm is produced by replacing the Euclidean distance in fuzzy c-means using a positive definite and symmetric matrix that ellipsoidal clusters could also be recognized, instead of only spherical ones. Also in GG algorithm, the kind of clustering such as GK algorithm could also be recognized in ellipsoidal clusters. We discuss about GK and c-means algorithms later. Probabilistic interpretation of GG clustering is shown by equation (1):

\[ p(X|\eta) = \sum_{i=1}^{c} p(X|\eta_i) = \sum_{i=1}^{c} p(X|\eta_i)p(\eta_i) \]  

(1)

GG algorithm, as a previous discussed, also take the density and size of clusters for classification. Hence, it has better behaviour for irregular features. Gath and Geva assumed that the normal distribution with expected value \( \mu \) and covariance matrix \( \Sigma \) (3).

\[ \text{the normal distribution in equation (2) should be given memberships to all probabilistic cluster partitions} \]

2 is chosen for generating a datum with a priori probability \( p \) (Gath & Geva 1989). A is chosen for indirect proportionality to equation (2) distance function, which is posteriori probability (likelihood) function. A is assumed to be approximately the same size and spherical clusters in p-dimensional space. The clusters resulting of fuzzy c-means algorithm recognizes spherical clusters in p-dimensional space. The clusters are assumed to be approximately the same size and each cluster is represented by its center. Also this cluster corresponds to the noise cluster increases the number of clusters to c+1. The additional cluster is defined as one minus the sum of the memberships to all other clusters. In the other hand we produce an ambiguity rejection region which increases the number of clusters to c+1. The additional cluster corresponds to the noise cluster \( \nu_{noise} \) for which \( d_{noise,k}(X_j,\nu_{noise}) = \delta \).

\[ m \in [1, \infty); \quad U \in \mathbb{R}^{c \times N}; \quad U_{i,j} \in [0,1], \forall i, j; \]

\[ \sum_{i=1}^{c} u_{i,k} = 1, \forall j; \quad 0 < \sum_{i=1}^{N} u_{i,j} < N, \forall i \]

(4)

In order to approximately minimize \( J \) with respect to all probabilistic cluster partitions \( (\nu_1, A_1, p) \) and given memberships \( u_{i,j} \), the parameters \( (\nu_1, A_1, p) \) of the normal distribution in equation (2) should be chosen as follows (l is the number of iterations):

\[ v_l^1 = \frac{\sum_{j=1}^{N} (u_{i,j})^m X_j}{\sum_{j=1}^{N} (u_{i,j})^m} \]  

(5)

\[ A_l^1 = \frac{\sum_{j=1}^{N} (u_{i,j})^m (X_j - v_l^1)(X_j - v_l^1)^T}{\sum_{j=1}^{N} (u_{i,j})^m} \]  

(6)

\[ p_l^1 = \frac{\sum_{j=1}^{N} (u_{i,j})^m X_j}{\sum_{j=1}^{N} \sum_{k=1}^{c} (u_{i,j})^m} \]  

(7)

We iteratively compute the parameters in (5) and applied them for calculation the distance function and memberships as follows equations (8), (9):

\[ d_{l,k}^2(X_j,\nu_l^1) = \frac{2\pi^p |\Sigma_l|^{1/2}}{p_l^1} \exp\left(\frac{1}{2}(X_j - \nu_l^1)^T(\Sigma_l)^{-1}(X_j - \nu_l^1)\right) \]  

(8)

\[ u_{l,k} = \frac{1}{\sum_{i=1}^{c} (d_{l,i}(X_j,\nu_l^1)/d_{l,k}(X_j,\nu_l^1))^{2/(m-1)} + (d_{l,i}(X_j,\nu_l^1)/\delta)^{2/(m-1)}} \]  

(9)

For reduction the noise, we can use memberships which are result of Dave’s objective function (Dave 1991):

\[ d_{noise,k}(X_j,\nu_{noise}) = \delta \]

4. The Fuzzy C-means Algorithm

Hard c-means or hard ISODATA algorithm was developed by Duda and Hart (Duda & Hart 1973) then Dunn (Dunn 1974) and Bezdek (Bezdek 1973) introduced a fuzzy version of this algorithm. The resulting of fuzzy c-means algorithm recognizes spherical clusters in p-dimensional space. The clusters are assumed to be approximately the same size and each cluster is represented by its center. Also this algorithm iteratively tries to minimize the equation (3) but the distance function obtains from equation (11):

\[ d_{l,k}(X_j,\nu_l^1) = ||X_j - \nu_l^1||^2 \]  

(11)

The equations (6) and (7) have no effect in fuzzy c-means algorithm.
5. The Gustafson-Kessel Algorithm

This algorithm designed by Gustafson and Kessel (Gustafson & Kessel 1979). In comparison with the fuzzy c-means algorithm, in addition to cluster centers each cluster is characterized by a symmetric and positive definite matrix A. thus, only the cluster shapes are variable now, but not the cluster’s sizes.

Also in this algorithm, the equation (3) should be iteratively minimized. Distance matrix obtains from equation (12):

\[ d_{ik}(X_j, v_k^l) = (X_j - v_k^l)^T A_k^l (X_j - v_k^l) \]  

(12)

The covariance matrix can be computed from equation (13) and (14):

\[ A_k^l = \frac{1}{|S_k^l|} \text{det}(S_k^l)^{-1} \]  

(13)

\[ S_k^l = \sum_{i=1}^{N} u_{ki}^l (X_j - v_k^l) (X_j - v_k^l)^T \]  

(14)

6. Experimental Results

Figure (1-a) shows the original ultrasonic image which was obtained from NDE (Moysan et al. 1996). The image is a three dimensional data which is normalized in its x and y axes and in its intensity. Using any kind of fuzzy clustering just for extraction the intensity of image leads to several thresholds, the number of which is equal to the number of clusters. Irregardless of fuzzy clustering techniques, defects tend to the clusters, the threshold levels of which are similar to its intensity. Hence, the defects with different intensity may be led to different classes. In order to avoid of above problems, we use density and size of clusters which are in x and y axes of image.

Now we evaluate result of the Gath-Geva algorithm. In primary stage, we applied this algorithm for two-class case: defect-background and in secondary stage, we applied it for three-class case: defect-speckle noise-background. Figures (1- a, b) show the result of two-class GG fuzzy clustering algorithm using equation (9). Image of defect is shown in figure (1-a) and image of background is shown in figure (1-b). On the other hand in figures (2- a, b and c) we used three-class GG algorithm using equation (9). We compare the results of two last stages. Three-class case separates speckle noise from defect which cause better contrast of defect’s image of three-class case relative to defect’s image of two-class case, but computing time of three-class case is more than of two-class case. Figures (3- a, b and c) show the result of two-class GG algorithm using equation (10) with \( \delta = 3 \). This noise clustering method uses an ambiguity rejection region to separate speckle noise from defect and background. The result of defect’s image in figure (3-b) is better than defect’s image in figure (1-b) and a few less than defect’s image in figure (2-b) but the computing time is similar to two-class GG algorithm. Figures (4- a, b and c) show that increasing of \( \delta \) reduces the speckle noise in defect’s image but it moves part of defect to ambiguity rejection class. Table (1) shows the ability of each method. In all figures, the images of background are approximately similar. Usually the velocity of three-class GG algorithm is 67% of two-class GG algorithm.

We applied noise clustering such as two-class GG algorithm but ambiguity rejection causes a third class with same velocity. Figures (5- a, b) show the result of two-class c-means algorithm. This algorithm can not correctly classify the defects. This problem repeated for GK algorithm in figures (6- a, b), the result of which is same. Figures (7- a, b and c) are for three-class c-means algorithm and figures (8- a, b and c) are for three-class GK algorithm. Both of them such as previous can not correctly classify the defects.

Table 1 Compare the abilities of methods

<table>
<thead>
<tr>
<th>Method</th>
<th>2 Class GG</th>
<th>3 Class GG</th>
<th>2 Class with GG Noise Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing Time</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Detection of Defect</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Noise Reduction</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

7. Conclusion

C-means and GK algorithms couldn’t correctly classify defects as a result of neglecting the size and density. We used Gath-Geva fuzzy clustering algorithm that was sensitive to density and size of clusters which applied on three dimensional of image containing area and intensity. By using an ultrasonic image of defects, we classified the speckle noise, defect and background in separated clusters. Quality and time of clustering depends to the number of classes. Noise clustering reduces the number of classes and the time of clustering. This method applied without any image processing technique such as removing particle, morphology and etc. but this method lonely could remove noise and segment defects of image. We obtain it on ultrasonic image but it can be applied on each NDT image.

References


Fig. 5. Fuzzy c-means clustering using two classes

Fig. 6. GK clustering using two classes

Fig. 7. Fuzzy c-means clustering using three classes

Fig. 8. GK clustering using three classes