Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Vol. 4
Lipo Wang, Jagath C. Rajapakse, Kunihiko Fukushima, Soo-Young Lee, and Xin Yao (Editors)

ACOUSTIC EMMISION SIGNAL CLASSIFIATION USING FUZZY C-MEANS CLUSTERING

S. N. Omkar, S. Suresh, T. R. Raghavendra, and V. Mani
Aerospace Engineering Department
Indian Institute of Science
Bangalore - 560012

ABSTRACT

Fuzzy C-means (FCM) clustering is used to classify the Acoustic Emission (AE) signal to different sources of signals. FCM has the ability to discover the cluster among the data, even when the boundaries between the subgroup are overlapping. FCM based technique has an advantage over conventional statistical technique like maximum likelihood estimate, nearest neighbor classifier etc, because they are distribution free (i.e.) no knowledge is required about the distribution of data. AE test is carried out using pulse, pencil and spark signal source on the surface of solid steel block. Four parameters - Event duration (E_d), Peak amplitude (P_a), Rise time (R_t) and Ring down count (R_d) are measured using AET 5000 system. These data's are used to train and validate the FCM based classification.

1. INTRODUCTION

Acoustic Emission (AE) signal refers to the stress or pressure wave produced due to transient energy release caused by irreversible deformation process in material. Superficial similarities between signals produced by different source may belie difference in features hidden in them. Hence, the fundamental step in tackling the problem is to extract these features to uniquely characterize signal and noise. Because of statistical variation in source characteristics or time varying source characteristics, classification of these signals from noise is difficult. Recently, some researchers have reported the use of pattern recognition in AE signal classification [1-4]. Most of these studies are based on simple problems and they are not generalized to all situations of signal analysis.

Different parameters of AE signal are measured to classify the signal to corresponding source. Hence the problem of classification boils down to multicategory pattern classification. The problems associated with AE signal classification are varied in nature because of the complexity and diversity of the phenomena. Hence, we have to use some intelligent mechanisms to classify such a complex AE signals. Traditionally maximum likelihood classifier (MLC) and neural networks (NN) are used to classify the multicategory pattern classification problems [5].

In MLC approach, first mean vectors and covariance matrices are computed, later the likelihood values are computed for each class for a given vector. The given vector is assigned to the class with maximum likelihood value [5]. The basic drawback of MLC classification is that a distance-based approach for classification is adapted and a normal distribution is assumed for the input data.

In the NN approach, a multi layer neural network with m inputs (features) and n output (classes) is trained with the training set. Subsequently, the given vector is applied to the network, and an n-dimensional output vector is obtained. The given vector is assigned the class of maximum output [6]. In NN approach, the basic draw back is that optimal configuration of the network is not known *apriori*. Moreover training time is very high, and knowledge represented in the weights is often opaque [6].

In multicategory pattern classification, apart from classification, there is a need to discover the underlying cluster in the data. Fuzzy based clustering is to discover the cluster in unsupervised manner, and express the hypothetical center for each cluster [7]. The Fuzzy C - Means (FCM) is gaining attention due to its ability to classify the data even in the boundaries between the subgroup is fuzzy [8]. AE signal consist of overlapping between the classes because of the superficial similarities in AE signal source and noise from the environment. Hence, in this work, FCM is used to classify the AE signals.

The paper is organized as follows; section 2 gives brief description about acoustic emission signals. Section 3 deals about fuzzy C-means clustering algorithm. Simulations and the results are discussed in Section 4.

2. ACOUSTIC EMISSION SIGNAL

Acoustic emission signal (AE) refers to the stress or pressure waves produced due to transient energy release caused by irreversible deformation process in a material [1-3]. Acoustic emission signals are electrical version of these waves produced by sensitive transducers. These signals can be analyzed to characterize the source of their emission.

In real Life, because of the presence of ambient noise and pseudo acoustic emission signals, it is very difficult to classify the sources of acoustic emission signals. We can eliminate the spurious signals by frequency filtering and amplitude threshold [4]. Even in noise free condition, signals from more than one source complicate the issue. Superficial similarities between the acoustic emission signals produced by different sources may belie differences in features hidden in them. Thus we have to extract the fundamental features to classify acoustic emission signal.

Basically, there are two types of acoustic emission signals,

- **Burst Type**: Amplitude of the signal rises sharply and decays gradually.
- **Continuous**: A continuous, sustained signal

Acoustic emission signals are rich in information about their source mechanisms. This information is measured using appropriate waveform parameters. In case of burst type, the following are the typical acoustic emission signal parameters,

Event Duration (E_d): The beginning of signals are marked when the envelope of the signals crosses the threshold value V_T and end is marked when the threshold value falls below V_T. *Ed* is the time distance between beginning and end of the mark.

Peak Amplitude (P_a): Highest peak obtained by a signal in an event.

Rise Time (R_t): Time taken to reach its peak after it first crosses the threshold value (V_T).

Ring down count (R_d): Number of time the signal crosses the threshold value (V_T).

Event gap: Time interval between two successive events

Full time, energy, RMS value, and dominant spectral frequency are also parameters of burst type acoustic emission signal. Most of the acoustic emission signals are of burst type. In our study, we assume noise free burst type acoustic emission signal from the metal surface. Three methods are commonly used for simulating acoustic emission on a test specimen,

(a)Breaking a fixed length of 2H pencil lead
(b)Feeding by a pulse generator to AE transducer mounted on the surface
(c)Generating spark by a spark gun close to the surface

All these sources give rise to burst-type AE signals. However, due to inherent differences in the mechanism, the signals from three sources differ in one or more of the signal parameters. AE signals are simulated by the three sources on the surface of a 6"-long, 4"-wide and 2"- thick steel block. The receiving transducer is kept at a fixed location on one side of the steel block. Events are simulated at random and signal parameters like, E_d, P_a, R_t, and R_a are measured for classification purpose. These parameters are also affected by a noise source. In order to avoid the noise from the signal, the transducer outputs are preprocessed and measured using AET2000 [4] system. We have 199 events, with four field of data item on each record corresponding to measured parameters. Out of 199 records, 40 belong to the pencil source, 45 to the pulse source and 39 to the spark source and the rest to noise (unknown source). The signals generated using the three source and unknown sources are used for fuzzy c-means clustering.

3. FUZZY BASED CLUSTER CLASSIFICATION:

Clustering refers to identifying the number of subclass of c clusters in a data set, consisting of n samples, and partitioning them in to c clusters ($2 \le c \le n$). Note that $c = 1$ denotes rejection of hypothesis that there are clusters in data set and $c = n$ constitutes the trivial case where each sample is in and clusters by it self. Cluster analysis is based on partitioning of a collection of data set (n - samples) into number of sub-group 'c', where the object inside a cluster (a subgroup) show certain degree of closeness or similarity. There are two kinds of partitioning the dataset,

(a)Hard (crisp)
(b)Soft (fuzzy)

Hard clustering assigns each data point (features) to one and only one of the clusters, with a degree of membership equal to one, assuming well-defined boundaries between the clusters. This model does not reflect the description of real data, where boundaries between the subgroups may be fuzzy. A family of clustering algorithms is developed based on fuzzy extension (soft clustering) using the least square error criterion [7] and many other clustering algorithms have been proposed for distinguishing sub structure in high dimensional data [8]. Bezdek developed an extremely powerful classification method to accommodate fuzzy sets, which is an extension of hard c - means clustering algorithm.

3.1 Fuzzy C- means Clustering (FCM):

Consider a n set sample data to classify in c classes, $X=\{x_1, x_2, \ldots x_n\}$. Each data sample x_i defined by m features, i.e., $x_i = \{x_{i1}, x_{i2}, \ldots x_{im}\}$, where x_i in the set X (X is a m - dimensional space). Since the m features all can have different units in general, we have to normalize each of the features to unified scale before classification.

Objective function approach is utilized for clustering n - data points to c - clusters. In this approach, each cluster is considered as one hyper spherical shape with hypothetical geometric cluster center. The main aim of the objective function is to minimize the euclidian distance between each data point in the cluster d and its cluster center, and maximize the euclidian distance between other cluster centers

We define a family of fuzzy sets $\{A_i, i = \{1, 2, ?.., c\}$ as a fuzzy c - partition on a dataset X in FCM method. Because fuzzy sets allow for degree of membership we can extend the crisp classification idea in to fuzzy classification notion. We can assign membership to the various data sets in each fuzzy set (fuzzy class, fuzzy cluster) Here, a single point x_k can have partial membership value, i.e., k^{th} data point in i^{th} class, is represented as

$$\mu_{ik} = \mu_{Ai}(x_k) \in [0,1]$$

With the restriction (as with crisp) that the sum of all membership values for a single point in all the classes has to be unity

$$\sum_{i=1}^{c} \mu_{ik} = 1 \quad \forall \text{ k} = 1, 2, 3, ..., \text{n}$$

There can be no class, contain empty set and there can be no class that contains all the data point. Which is represented as below,

$$0 < \sum_{k=1}^{n} \mu_{ik} < 1$$

In case of fuzzy classification, each data point can have partial membership in more than one class,

$$\mu_{ik} \wedge \mu_{jk} \neq 0$$

Now we can define a family of fuzzy partition matrixes M_{fc}, for the classification involving c classes and n data points.

$$M_{fc} = \{U | \mu_{ik} \in [0,1]\}$$
$$\forall \text{ i} = 1, 2, \ldots, \text{c and } \text{k} = 1, 2, \ldots, \text{n}$$

Any $\mu \in M_{fc}$, is a fuzzy c-partition, and it follows from the overlapping character of the classes and the infinite number of membership values possible for describing the class membership. The objective function used for fuzzy c - clustering is,

$$J_m(u,v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^{m'} (d_{ik})^2$$

where

d_{ik} - distance between the point x_k and the center v_i,

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij}))^2 \right]^{\frac{1}{2}}$$

μ_{ik} - membership of the k^{th} data point in i^{th} class

m' -weighting parameter $m' \in [1, \infty]$

The weighting parameter m' controls the amount of fuzziness in the classification process; v_i is the cluster center of i^{th} class,

$$v_{ij} = \frac{\sum_{k=1}^{n} \mu_{ik}^{m'} \cdot x_{kj}}{\sum_{k=1}^{n} \mu_{ik}^{m'}} \qquad j = 1, 2, \ldots, m$$

The objective of the clustering is to minimize the objective function with respect to the partition matrix and cluster center

$$J_m^*(U^*, V^*) = \min_{M_{fc}} J(u,v)$$

This may not give optimal solution (global minimum). In order to achieve global optimal interactive optimization technique (Iterative FCM algorithm) are used to obtain partition matrix. The iterative FCM algorithm is given below,

3.2 Iterative FCM Algorithm:

1. Fix c $(2 \le c < n)$ and select a value for m' and also for $\in_L$.
2. Initialize the partition matrix $U^{(0)}$, set r = 0.
3. Calculate the centers $v_i^{(r)}$
4. Update the partition matrix

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}} \right)^{\frac{2}{m'-1}} \right]^{-1}$$

$I_K = \Phi$ (null set) or $\mu_{ik}^{(r+1)} = 0, \forall \ i \in \bar{I}_k$

Where

$$\bar{I}_k = \{i/2 \le c \le n; \quad d_{ik}^{m'} = 0\}$$

and

$$\bar{I}_k = \{1, 2, \ldots, c\} - I_K$$

$$\sum_{i \in I_k} \mu_{ik}^{(r+1)} = 1$$

5. If $\| U(r+1) - U(r) \| \le \in_L$ then stop, else go to step 3

The main advantage of fuzzy clustering is partition matrix converges faster even with quite poor guess.

4. SIMULATIONS AND DISCUSSION

In our example each data point has four input features namely Event duration (Ed), Peak amplitude (Pa), Rise time (Rt) and Ring down count (Ra) [4]. There are 199 samples are available [4]. The total set divided in to Training set and testing set, 63 data points are used to calculate the cluster center and remaining 136 data points used for testing.

There are 4 classes in the data sets, namely *A* - pencil source, *B* - noise (unknown source) *C* - pulse and *D* - spark. There is lot of interference in features between the noise and spark signals. Hence we first look for *three* clusters in the data set (*A*, *C*, and *B* + *D*). In training set, we have 14 - pencil source 15 - pulse source and 34 - (17 + 17) unknown source and spark source. Cluster centers are computed from the training data. After running our iterative FCM program, 13 points are classified as pencil source in cluster no 1, 15 points are classified as pulse in cluster no 2 and 35 points classified as unknown and spark source in cluster no 3. In the Training set only one pattern belongs to class-*A* is misclassified as unknown or spark source. The results are shown in Table 1.

All the 35 data points classified as class (*A-C* and *B-D*) obtained from the iterative FCM program is selected for second run. The number of cluster in the second run is assumed as *five*. After the FCM second run, we found that four clusters (No. 1-4) belong to spark signal source and one cluster (No. 5) belong to unknown source. Out of 35 data point 17 data points are classified as unknown source and 18 data points are classified as spark signal source. The results of the FCM second run program are shown in Table 2. From Table 2, we can observe 9-data points are misclassified, in which 5 - spark signals are misclassified as unknown source; and 3 - unknown source classified as spark signal and 1 - pencil is classified as spark source.

The classification accuracy, for all these FCM runs are obtained using the testing data set. The classification accuracy obtained for all these FCM runs are shown in Table 3 and Table 4. We can see from Table 3 and 4, that most of the samples are classified accurately. In case of testing/validation the pattern is classified into different sources depends on the distance between the centers of the clusters obtained from the first FCM run. If the pattern is very close to center of class-*A* then the new pattern is classified as pencil source i.e., depends on the distance from center, the pattern is classified to class-*A*, or class-*C* or class *B* and *D*. If the testing pattern is classified as class *B* and *D* then, we have to calculate the distance between the pattern and cluster center of second FCM run. Distance between the testing pattern and any of the 4 clusters (No. 1-4) is less than the pattern is classified as spark signal. Otherwise it is classified as unknown source. Because of close relationship between the spark source and unknown sources, the classification accuracy of class *D* and B is poor when compared to other classes.

5.CONCLUSION

The Fuzzy C-means clustering algorithm has been successfully applied for classification of acoustic emission signals. The FCM classifies the signal generated by pencil and spark signal source with a high degree of accuracy (> 93%). However, the signals originated from noise and pulse source shows some form of overlapping and as such the accuracy of their classification stands at around 80%. The study shows that FCM can be used as an effective algorithm for unsupervised classification of signal in Non-Destructive Testing (NDT).

REFERENCE:

[1]. R. K. Elsley, L. J. Graham, "Pattern Recognition Technique Applied to Sorting Acoustic Emission signals", *Prod. IEEE Ultrasonic Symp.*, pp. 147-150, 1976.

[2]. R. K. Elsley, L. J. Graham, "AE Source Identification by Frequency Spectral analyis for Aircraft Monitoring Applications", *Journal of Acoustic Emission*, vol. 4, No. 4, pp. 115-123, 1987.

[3]. C. R. L. Murthy, B. Dattaguru, and A. K. Rao, "Application of Pattern Recognition Technique in Acoustic Emission Signal analysis", *Journal of Acoustic Emission*, Vol. 6, No 1, pp. 19-28, 1987.

[4]. M. Abdul Majeed, "Acoustic Emission Signal Analysis: An Unsupervised Pattern Recognition approach", Ph.D Thesis, Indian Institute of Science, Bangalore, 1991.

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

[6]. S. Haykin, *Neural Networks – A Comprehensive Foundation*, New York: Macmillan College, 1994.

[7]. J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithm, *Plenum*, New York.

[8]. T. J. Ross, Fuzzy Logic with Engineering Applications, *McGraw-Hill*, New York, 1997.

Cluster No.	Actual class	No of points in each class	Class *A*	Class *C*	Class *B* and *D*	% Classification accuracy	% Total efficiency
1	*A*	14	13	-	1	92.8	97.6
2	*C*	15	-	15	-	100	
3	*B* and *D*	34	-	-	34	100	

Table 1. Training results from first FCM run

Cluster No	Actual class	No of points in each class	Class *A*	Class *B*	Class *D*	% Classification accuracy	% Total efficiency
5	*B*	17	-	14	3	82.35	76.465
1-4	*D*	17	1*	5	12	70.58	

Table 2. Training results from second FCM Run

Cluster No	Actual class	No of points in each class	Class *A*	Class *C*	Class *B* and *D*	% Classification accuracy	% Total efficiency
1	*A*	26	22	1	3	84.61	93.34
2	*A*	30	-	29	1	96.66	
3	*B* and *D*	80	-	1	79	98.75	

Table 3. Testing/validation for first FCM run

Cluster No	Actual class	No of points in each class	Class *A*	Class *B*	Class *D*	% Classification accuracy	% Total efficiency
5	*B*	58	-	54	4	93.10	80.64
1-4	*D*	22	3	4	15	68.18	

Table 4. Testing / validation result for second FCM run.

*- Misclassified pattern from first FCM run