Influence of the choice of histogram parameters at Fuzzy Pattern Matching performance

MOAMAR SAYED MOUCHAWEH, PATRICE BILLAUDEL
Laboratoire d'Automatique et de Microélectronique
IFTS
7 Boulevard Jean Delautre
08000 Charleville-Mézières
FRANCE

Abstract: - Fuzzy Pattern Matching (FPM) is a supervised classification method, which uses a histogram, for each attribute of each class, to obtain a probability density function and a transformation probability-possibility to have a possibilistic membership function. A histogram is not unique for given data set, it depends upon two parameters : the number of bins and the histogram width. Their sound choice determines the quality of the histogram and consequently the quality of the corresponding possibilistic membership function, which influences the performance of FPM. In the literature, there exist only a few explicit guidelines, which are based on statistical theory, for choosing the number of bins. These guidelines give some formulas for the optimal number of histogram bins that minimizes an error function. Since in FPM the probability density function is unknown, it is not clear how one should apply this minimization in practice. Moreover, these formulas do not take into consideration the problem of training sample size and the overall optimal value of the number of bins for several histograms. In this paper, we will study the influence of the choice of histogram parameters on FPM performance and we will propose a method to well determine them.

Key-Words : - Histogram, Fuzzy Pattern Matching, Possibility theory, Supervised learning, Class separability

1 Introduction

Pattern recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decision about the category of the pattern [1]. This recognition is considered as a classification or categorisation task and it is done in using a classifier.

The statistical pattern recognition is one of the best known approaches for pattern recognition. It represents each pattern in terms of α features and it views this pattern as a point in a α dimensional space which is called the feature space.

The classification is done by means of a discriminate function which gives, for a new point, a membership degree to each class. The new point is assigned to the class for which it has the highest membership degree.

Our team of research, Diagnosis of Industrial Processes, uses Fuzzy Pattern Matching (FPM) as a method of classification for its simplicity and its low calculation time [2]. It is a supervised classification method, which uses a histogram, for each feature or attribute of each class, to obtain a probability density function (PDF) and a transformation probability-possibility in order to have a possibilistic membership function.

The number of bins h determines the quality of the histogram and consequently the quality of the corresponding possibilistic membership function, which influences the performances of FPM.

In the literature, there exist only a few explicit guidelines, which are based on statistical theory, for choosing the number of bins to use in the histogram [3]. These guidelines give some formulas for the optimal number of histogram bins that minimizes an error function [3, 4, 5].

We wish finding the optimal values of histogram parameters to obtain the best performance of FPM. Since in FPM the probability density function is unknown, it is not clear how one should apply this minimization in practice. Moreover, these formulas do not take into consideration the problem of training sample size and the overall optimal value of the number of bins for several histograms. Using the cross validation methods [5] entails a large sampling variance which is a real problem when the training sample size is small. Additionally, its computation time is expensive. The histogram limits are usually the minimal and maximal values of the training set for the considered attribute.

In the literature, we could not find any study about the influence of histogram parameters or how they can be chosen to optimise FPM performance. In this paper, we will propose a method to well determine these parameters for FPM.

2 Histogram parameters

The histogram is the most important graphical tool for exploring the shape of data distributions. It gives an idea of how frequently data in each class occur in the training data set. We are considering the decision problem where a histogram is an estimate of an unknown probability density function.

In FPM, a histogram is constructed for each feature of each class. We will consider a single feature in a single class. The treatment is then extended for α features of c classes.
The histogram is computed in the following manner : an interval $(x_2 - x_1)$ of a feature is divided into h subintervals of equal length, each subinterval is called bin. The bin width thus is defined by :

$$b = \frac{x_2 - x_1}{h} \quad (1)$$

The height of bin m is determined in calculating the number n_m of occurrences of the data patterns within the interval of this bin. The probability p_m assigned to the bin m is the ratio of bin height to the total number n of patterns :

$$p_m = \frac{n_m}{n} \quad (2)$$

The most important parameter that need to be specified when constructing a histogram is the number of bins h. It controls the trade-off between presenting a data distribution with too much detail or too little detail with respect to the true distribution. Indeed if too few or too many bins are used, the histogram can be misleading. Despite its importance, there is no criterion to estimate the optimal value of h especially in the case where the probability density function is unknown [3, 4, 5, 6, 7, 8].
The width of a feature $(x_2 – x_1)$ defines the variability of a process according to this feature. In the literature, if we do not know the PDF, x_1 et x_2 are determined either as the minimal and maximal values of the data set according to each feature [3]. If the PDF is known, the hypothesis that every bin should have at least two occupancies is used [8].

3 Fuzzy Pattern Matching

Fuzzy Pattern Matching [9, 10, 11] is a classification method which has been developed in the framework of fuzzy set and possibility theory to take into account the imprecision and the uncertainty of the data [11]. The histograms of the data are transformed into histograms of probability in using (2). Then two bins are added to each histogram, one at the beginning and the other at the end of the histogram. These two additional bins have a probability value equal to zero. The probability densities are constructed in linking linearly the bin centres. The probability distributions are transformed into possibility distributions π in using a probability-possibility transformation. We had chosen the transformation of Dubois and Prade :

$$\pi_m = \sum_{k=1}^{l=h+2} \min(p_m, p_k), \quad m = 1 \,..\, h+2 \quad (3)$$

These possibility distributions are transformed into density ones by linear linking between each two bin centres.
The classification of a new sample y whose values of the different attributes are $y_1, \ldots, y_\alpha$, is made in three steps [9]:

- determination of the possibility membership value of y for each attribute of each class by linear interpolation,
- fusion of all the possibility membership values concerning class i, into a single one by the operator minimum. The result of this fusion represents the possibility that the new sample y belongs to the class i,
- finally, y is assigned to the class for which it has the highest membership degree.

4 Overlap degree

The performance of a classification system is dependent upon the data presented to the system. If these data are not sufficiently separable, then the classification performance of the system will be insufficient, regardless of the classification method used [13].
There is large number of class separability measures in the literature [14, 15, 16, 17, 18]. All these measures are calculated in using all the samples. This causes a large computation time and needs a high memory size especially in big sample size cases with high dimension. In this section, we propose to use another indication to measure the class overlap degree for FPM.
Let I_{ij}^k be the overlap degree between the class i and the class j according to the attribute k, and C be the set of all the possible subsets of two classes. I_{ij}^k is then a mapping :

$$I_{ij}^k : C \to [0\ 1],\ i,j = 1 \,..\, c,\ k = 1 \,..\, \alpha \quad (4)$$

Separability degree between two classes is simply :

$$S_{ij}^k = 1 – I_{ij}^k \quad (5)$$

$I_{ij}^k = 1$ means that the class i covers completely the class j according to the attribute k while $I_{ij}^k = 0$ denotes that the class i is completely separated from the class j. $I_{ij}^k \in [0\ 1]$ means that the class i covers partially the class j with the degree I_{ij}^k. The overlap degree I_{ii}^k is equal to 0 because it is not used by the method.
The overlap degree for attribute k is the following matrix of dimension c x c :

$$I_{c,c}^k = \begin{pmatrix} 0 & I_{12}^k & \ldots & I_{1c}^k \\ I_{21}^k & 0 & \ldots & I_{2c}^k \\ \ldots & \ldots & \ldots & \ldots \\ I_{C1}^K & I_{c2}^k & \ldots & 0 \end{pmatrix} \quad (6)$$

In FPM, each probability density for each attribute k of each class i has an active interval $[x_{1i}^k\ x_{2i}^k]$ where a new point can have a membership value according to this class. Additionally, a bin m of a histogram of the attribute k of the class i starts at x_{1im}^k and finishes at x_{2im}^k as it is explained in Fig.1.

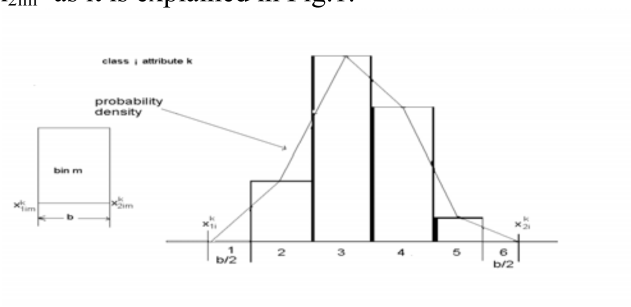

Fig.1. Active interval of probability histogram

The overlap degree I_{ij}^k between class i and class j according to the attribute k is :

$$I_{ij}^k = \sum_{m=1}^{h} I_{jm}^k \text{ where:}$$

$$I_{jm}^k = p_{jm}^k \quad \text{if } x_{1jm}^k \geq x_{1i}^k \text{ and } x_{2jm}^k \leq x_{2i}^k \text{, otherwise:}$$

$$I_{jm}^k = \left(\frac{x_{2jm}^k - x_{1i}^k}{b}\right).p_{jm}^k \quad \text{if } x_{1jm}^k \prec x_{1i}^k \text{ and } x_{2i}^k > x_{2jm}^k > x_{1i}^k$$

otherwise:

$$I_{jm}^k = \left(\frac{x_{1jm}^k - x_{2i}^k}{b}\right).p_{jm}^k \quad \text{if } x_{2jm}^k \succ x_{2i}^k \text{ and } x_{1i}^k < x_{1jm}^k < x_{2i}^k \tag{7}$$

otherwise:

$$I_{jm}^k = \left(\frac{x_{2i}^k - x_{1i}^k}{b}\right).p_{jm}^k \quad \text{if } x_{1jm}^k \leq x_{1i}^k \text{ and } x_{2jm}^k \geq x_{2i}^k$$

otherwise:

$$I_{jm}^k = 0$$

The Fig.2 shows how we calculate I_{ij}^k .

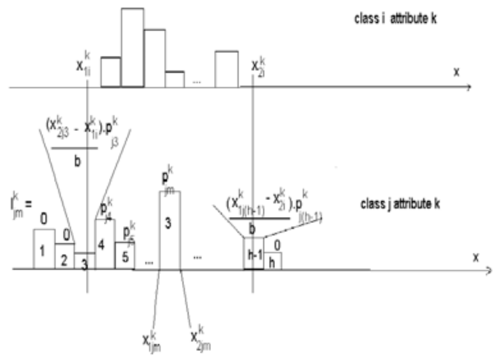

Fig.2. Calculation of the overlap degree

These matrixes are not symmetric thus to make them symmetric, we calculate the mean value of overlap degrees between the classes i and j and between the classes j and i :

$$I^k_{ij} = I^k_{ji} = \text{mean}(I^k_{ij}, I^k_{ji}) \tag{8}$$

To discriminate two classes, it is sufficient that they are separated by at least one attribute. Thus we will aggregate the overlap degrees matrixes for the different attributes in one matrix in using the minimum operator :

$$I_{c,c} = \min(I_{c,c}^1, I_{c,c}^2, \ldots, I_{c,c}^\alpha) \tag{9}$$

The overlap degree for each class i is calculated in using the maximum operator :

$$od_i = \max(I_{ij} : j = 1..c) \tag{10}$$

The different overlap degree values, od_i : i = 1 .. c, are aggregated to give one value which evaluate the overall overlap degree for all the classes :

$$od = \frac{\sum_{i=1}^{c} od_i}{c} \tag{11}$$

The overlap degree gives the upper envelope of the misclassification rate; in other words it gives the worst case of misclassification in considering all the points which are located in the overlap area as misclassified points.

5 rejection gaps number

The overall overlap degree od must be calculated for different values of h in order to choose the one which yields to the least od. But when h increases, the histogram gives too much detail, which leads it to see the gaps, or spaces, between samples. This fact is reflected in possibility densities as zero values. They entail the rejection of samples, which are located inside the class. Each gap is represented by a null bin inside the histogram. The number of rejection gaps is calculated by :

r_g = (Σi, $p_i = 0$, m < i < n : p_m and p_n are, respectively, the first and last bins which their heights are not equal to zero) $0 \leq r_g \leq h - 2$ (12)

The Fig.3 shows an example of the calculation of r_g .

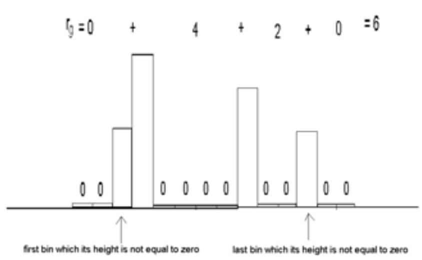

Fig.3. Calculation of the number of rejection gaps

6 Application

6.1 Washing machine data

This example corresponds to the detection of unbalance failures in a washing machine [20]. The lateral and frontal amplitudes of the movements of the machine define the feature space. The unbalance failures make to appear four classes in this space. One of these classes corresponds to the good functioning and the three other ones correspond to different types of unbalance failures. The Fig.4 shows these classes in the feature space.

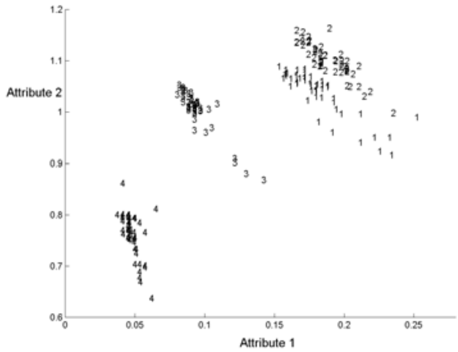

Fig.4. The 4 classes of the washing machine data

The Fig.5 shows the overlap degree and the misclassification rate for different values of h. h = 14 is the best compromise value which gives the best separation between the classes and does not cause the formation of rejection gaps. The overlap degrees for each class are : od_1 = 0.72, od_2 = 0.72, od_3 = 0.03 and od_4 = 0.01. Thus the problem of separation is due to the overlap between the classes 1 and 2

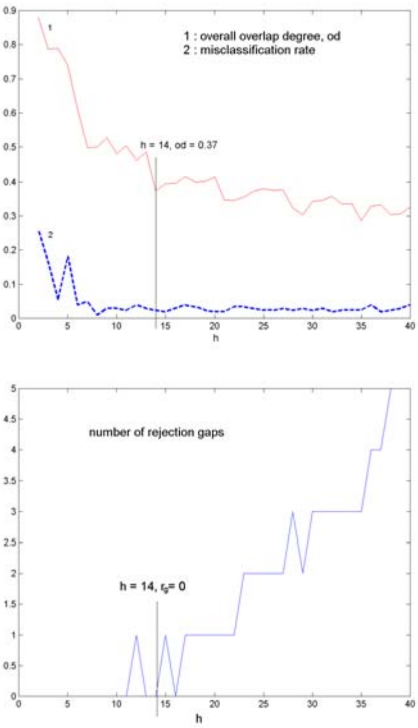

Fig.5. Comparison between the overall overlap degree, the misclassification rate, and the number of rejection gaps for different values of h for the washing machine.

6.2 Plastic injection data

This example concerns the diagnosis of the quality of a plastic injection moulding process [9]. The data are divided into 5 classes in a feature space of 3 attributes : maintenance time, final position of mattress, and the barrel temperature. The classes 1 and 2 present the good quality products and the other classes present different kinds of production faults. The Fig.6 shows these classes and the Fig.7 shows the comparison between the overall overlap degree, the misclassification rate, and the number of rejection gaps for different h. We can find that h = 9 gives overlap degree equal to zero, and avoid the formation of rejection gaps. The overlap degrees for the classes are : od_1 = 0, od_2 = 0, od_3 = 0, od_4 = 0 and od_5 = 0.

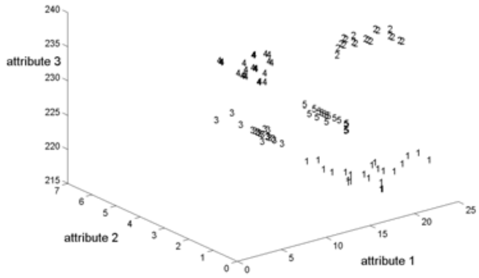

Fig.6. The 5 classes of the plastic injection moulding process

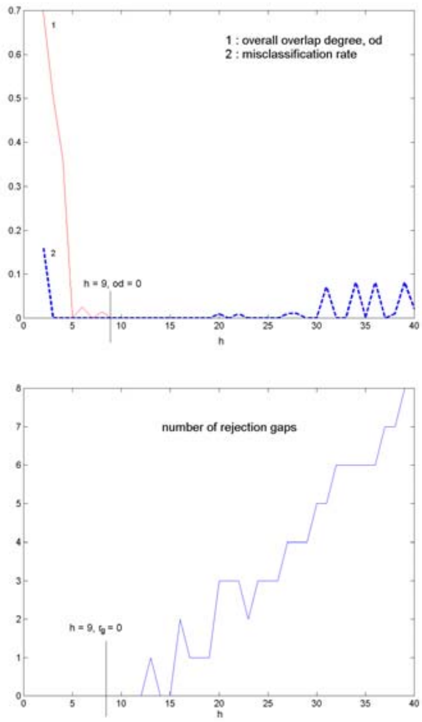

Fig.7. Comparison between the overall overlap degree, the misclassification rate and the number of rejection gaps for plastic injection moulding.

Indeed, for the high values of h, we can notice the misclassification rate is bigger than the overall overlap degree. This fact is due to the formation of rejection gaps which causes the rejection of samples inside classes.

A high h, even if it does not cause the formation of rejection gaps, increases the computing time which makes the classification of new point and the updating of possibility densities hard in real time. Therefore, for the choice of h, we must add a third condition which is the computation time. In addition, a too big value of h makes the classification system sensible to the local noise. Thus, the expert must choose a suitable value of h even if the misclassification rate increases.

7 Influence of histogram limits location

Terrell and Scott [19] showed that the sample range may be used if the interval $[x_1 \;\; x_2]$ is unknown or even if $x_2 - x_1 = \infty$ but the tail is not too heavy. Indeed, Scott [7] considered the histogram bin origin as a nuisance and he suggested deleting it in averaging several histograms which have the same bin width but different histogram origins. The number of histogram origins, m, must not be too big in order to keep the computational efficiency of the histogram. For Scott, x_2 has an infinite value since the histogram has an infinite number of bins.

In FPM, the number of bins is finite so we will study the influence of both origin and upper limit of histograms. To do that, h will be fixed and both x_1 and x_2 will be changed starting from the data range. In considering x_{min} and x_{max} the least and the greatest values of the data according to each attribute, x_1 and x_2 will be changed as the following manner :

$$x_1 = 0, \frac{x_{min}}{m}, \frac{2.x_{min}}{m}, \ldots, x_{min},$$

$$x_2 = x_{max}, x_{max} + \frac{x_{max}}{m}, x_{max} + \frac{2.x_{max}}{m}, \ldots, \tag{13}$$

$$x_{max} + x_{max}$$

thus the first pair $[x_1 \;\; x_2]$ is the data range $[x_{min} \;\; x_{max}]$.

The overlap degree and the number of rejection gaps will be calculated according to the difference $x_2 - x_1$ for a given h. The Fig.8 and Fig.9 show the relationship between $x_2 - x_1$ and the overall overlap degree for the previous two examples. We have chosen the values of h which were determined before to give a suitable compromise between overlap degree and rejection gaps.

These figures show that :

- origin and upper limit of a histogram influence the overlap degree and consequently the performance of FPM,
- the range of leaning data set gives the best performance for a given h.

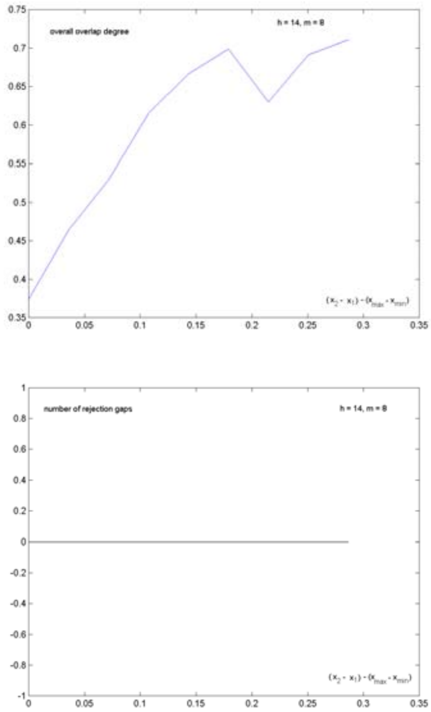

Fig.8. Relationship between origin and upper limit of histograms with overlap degree and number of rejection gaps for washing machine example

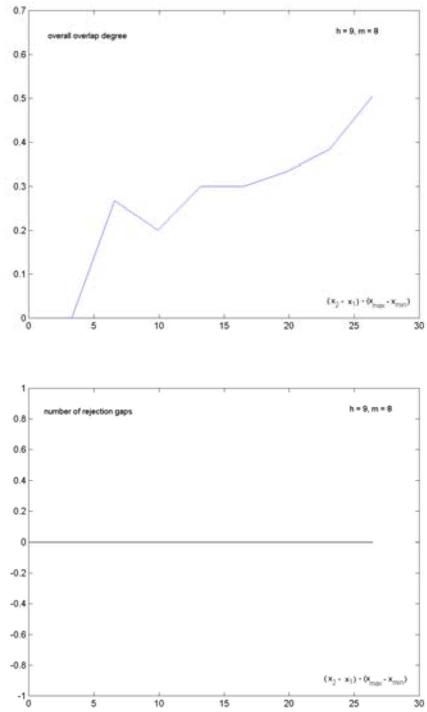

Fig.9. Relationship between origin and upper limit of histograms with overlap degree and number of rejection gaps for plastic injection moulding example

8 Conclusion

A histogram is not unique for given data, it depends upon two parameters : the number of bins and the histogram width. Despite its importance, there is no criterion in the literature to estimate the optimal value of these parameters especially when the probability density function is unknown which is the case of the classification method Fuzzy Pattern Matching. In this paper, we have showed how we can determine the optimal values of histogram parameters in order to maximise as possible the performance of FPM.

The performance of a classification system is dependent upon the data presented to the system. If these data are not sufficiently separable, then the classification performance of the system will be insufficient, regardless of the classification method used. There is large number of class separability measures in the literature. All these measures are calculated in using all the samples. This causes a large computation time and needs a high memory size especially in big sample size cases with high dimension. For this reason, we have proposed a new class overlap measure which is independent of the sample size and is adapted for FPM. The optimal values of histogram parameters are chosen to minimize the overlap degree of the classes.

The overall overlap degree gives the maximal value of the misclassification rate because it takes the worst case in considering all the samples located in the overlap zone as misclassified points. Thus as the Bayes error defines the minimal value of the error rate, the overall overlap degree defines the maximal value of the error rate in using FPM.

If the sample size is insufficient to determine the overlap degree between classes, we need to take benefit of the information carried by the new classified points. Since the overlap degree, proposed here, is independent of the sample size, the update of the overlap degree can be done in a fixed time which makes its use for real time application totally possible.

References:

[1] Anil K.J., Robert P.W., Mao J., Statistical Pattern Recognition: A review, IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, No. 1, 2000

[2] Billaudel P., Performance evaluation of a fuzzy classification methods designed for real time application, International Journal of Approximating 20, 1999, 1-20.

[3] He K., Meeden G., Selecting the number of bins in a histogram: A decision theoretic approach, Journal of Statistical Planning and inference, Vol. 61, 1997

[4] Scott D. W., On optimal and data-based histograms, Biometrika, Vol. 66, 1979, pp. 605-610.

[5] Rudemo M., Empirical choice of histograms and kernel density estimates. Scandinavian Journal of Statistics, 9, 1982, pp. 65-78.

[6] Izenman, A. J., Recent developments in nonparametric density estimation, Journal of the American Statistical Association, 86(413), 1991, pp. 205-224.

[7] Scott D. W., Multivariate density estimation, Wiley, New York, 1992

[8] Otnes R. K., Enochson L., Digital time series analysis, Wiley-Interscience Publication, New York, 1972

[9] Devillez A., Billaudel P., Villermain Lecolier G., Use of the Fuzzy Pattern Matching in a diagnosis of a plastic injection moulding process, European Control Conference ECC'99, Germany, 1999

[10] Grabish M., Sugeno M., Multi-attribute classification using fuzzy integral, Proc. of fuzzy IEEE, 1992, 47-54.

[11] Dubois D., Prade H., Testemale C., Weighted Fuzzy Pattern Matching, Fuzzy Sets Systems, 1988, 313-331.

[12] Dubois D., Prade H., On possibility/probability transformations, Fuzzy Logic, 1993, pp. 103-112.

[13] Sancho J. L. et al., Class separability estimation and incremental learning using boundary methods, Neurocomputing 35, 2000, pp. 3-26.

[14] ZADEH L. A., Fuzzy sets, Informations and control 8, 1965, pp. 338-353.

[15] Chen C. H., On information and distance measures, error bounds and feature selection, Inform. Sci. 10, 1976, pp. 159-173.

[16] BEZDEK J. C., HARRIS J. D., Fuzzy relations and partitions : an axiomatic basis for clustering, Fuzzy Sets and Systems 1, 1978, pp. 11-27.

[17] Bezdek J. C., Pattern recognition with fuzzy objective function algorithms, Plenum Press, New York, 1981

[18] Frigui H., Krishnapuram R., A robust algorithm for automatic extraction of an unknown number of clusters from noisy data, Pattern Recognition Letters 17, 1996, pp. 1223-1232.

[19] Terrel G. R., Scott D. W., Oversmoothed non parametric density estimates, Journal American Statistics Association, Vol. 80, 1985, pp. 209-214.

[20] BILLAUDEL P., DEVILLEZ A., VILLERMAIN LECOLIER G., Identification of the unbalance faults in washing machines by a possibilistic classification method, International Conference on Artificial and Computational Intelligence For Decision, Control and Automation, ACIDCA'2000, Tunisia, 2000