# Incremental possibilistic approach for online clustering and classification

Moamar Sayed Mouchaweh, Bernard Riera

Université de Reims Champagne-Ardenne, Centre de Recherche en STIC (URCA-CReSTIC)
Moulin de la Housse, BP 1039, 51687 Reims Cedex, France
Email: {moamar.sayed-mouchaweh,bernard.riera}@univ-reims.fr

***Abstract***— *In this paper, we propose to develop the supervised classification method Fuzzy Pattern Matching to be in addition a non supervised one. The goal is to monitor dynamic systems with a limited prior knowledge about their functioning. The detection of the occurrence of new states as well as the reinforcement of the estimation of their membership functions are performed online thanks to the combination of supervised and non supervised classification modes. No information in advance about the shape of classes or their number is required to achieve this detection and estimation reinforcement.*

***Keywords***— Classification, clustering, Sequential learning, Fuzzy Pattern Matching.

## 1 Introduction

Pattern Recognition (PR) [7] is the study of how machines can learn from experience to make sound decisions about the classes of patterns of interest. PR involves two stages: preprocessing and classification. The preprocessing includes feature extraction [1] and selection [12]. The classification stage is a mapping of a pattern from the feature space into the decision one. The latter is defined by a set of predefined classes. This mapping is done using a classifier. The latter is a method or algorithm which generates a class membership function in order to classify unlabelled incoming patterns into one of the predefined classes. Depending on the information available for classifier training, one can distinguish between supervised [10] and unsupervised [2] learning. In the first case, called also classification, there exists a set of patterns with their class assignment or label, called learning set. The goal of supervised learning is to learn a set of membership functions that allows the classification of new patterns into one of existing classes. The problem of unsupervised learning, also called clustering, arises if clusters' memberships of available patterns, and perhaps even the number of clusters, are unknown. In such cases, a classifier is learned based on similar properties of patterns. Hence, the clustering aims to partition a given set of patterns into clusters based on their similarity.

Semi supervised learning techniques use small or limited labelled patterns to estimate the classes' membership functions and the unlabelled ones to detect the occurrence of new classes and refine their membership functions estimation. Examples of semi supervised methods can be found in [4, 9, 11, 13] and the references therein. These methods are based either on the use of the Expectation Maximisation algorithm for maximum likelihood based parameters estimation [9], on the integration of an incremental algorithm for the update of classifiers' parameters [13], on the optimisation of an objective or learnable distance function [4] or on a classifier retraining to integrate new labelled points [11]. The popularity of these methods can be attributed to the fact that new information can be incorporated resulting better estimation of classes' membership functions and thus more prediction accuracies thanks to the unlabelled patterns. However, the representativeness of the labelled data is of crucial importance especially for small ratios of labelled to unlabelled patterns [9]. This is due to the fact that the clustering is guided by the labelled patterns.

One of the applications of PR methods is the monitoring of dynamic systems. System states, normal or faulty, are characterized by classes in the feature space. The performance of statistical PR methods depends on the prior knowledge, or learning set, about system behavior. The number of available learning patterns is often limited and small compared to the dimension of the feature space. Thus, it becomes hard to estimate the class membership function leading to a large variance in parameter estimates and thus higher classification error rates. Moreover, the behavior of a dynamic system can assume different operating states in the course of time. The learning set cannot contain patterns about all these states especially the faulty ones. Thus, the occurrence of these missing states must be anticipated online and integrated in the data base. In this paper, we propose a solution for these problems. This solution is based on the use of the supervised classification method Fuzzy Pattern Matching [3]. This method presents the advantage to process data with a low and constant classification time according to the size of data base. We propose to develop FPM as an unsupervised classification method. The goal is to combine the supervised and unsupervised learning strategies within a single algorithm leading to a semi supervised version of FPM.

The paper is organized as follows. Firstly, the functioning of FPM is illustrated. In the next section, the proposed solution to perform supervised and unsupervised classification using FPM is detailed. The performance of semi supervised FPM is illustrated and tested using some simulated examples.

ISBN: 978-989-95079-6-8

## 2 Supervised Fuzzy Pattern Matching

### 2.1 Learning phase

In the learning phase, the probability histograms are constructed for each class according to each attribute. The number of bins $h$ for a histogram is experimentally determined. This number has an important influence on the performances of FPM [14]. The histogram upper and lower borders can be determined either as the maximal and minimal learning data coordinates or by experts. The height of each bin $b_k^j, k \in \{1,2,..,h\}$, according to each attribute $j$ is the number of learning patterns $n_{ib_k^j}$ of the class $C_i$ located in this bin. The probability distribution $\{ p_i^j(y_{b_k^j}), k \in \{1,2,..,h\}, j \in \{1,2,..,d\}\}$, of the class $C_i$, $i \in \{1,2,..,c\}$, according to the attribute $j$ is obtained by dividing the height of each bin by the total number $N_i$ of learning patterns belonging to the same class $C_i$. These probabilities are assigned to the bins' centres $y_{b_k^j}, k \in \{1,2,..,h\}$:

$$p_i^j(y_{b_k^j}) = \frac{n_{ib_k^j}}{N_i} \tag{1}$$

The Probability Density Function (PDF) is obtained by a linear linking between bins heights centres.

In order to take into account the uncertainty and the imprecision contained in the data, the probability distribution is converted into possibility one $\left\{\pi_i^j(y_{b_k^j}), k \in \{1,2,..,h\}, j \in \{1,2,..,d\}\right\}$. The conversion is performed using the transformation of Dubois and Prade [6] defined by:

$$\pi_i^j(y_{b_k^j}) = \sum_{f=1}^{h} \min(p_i^j(y_{b_k^j}), p_i^j(y_{b_f^j})) \tag{2}$$

A linear linking between bins heights centres converts the distribution of possibilities into density one $\Pi_i^j$. This operation is repeated for all the attributes of each class.

### 2.2 Classification Phase

The membership function for each class $C_i$ according to each attribute $j$ is considered to be numerically equivalent to the possibility distribution [15]. Thus, the classification of a new pattern $x$, whose values of the different attributes are $x^1$, .., $x^j$, .., $x^d$, is made in two steps :

Determination of the possibility membership value $\pi_i^j(x^j)$ of $x^j$ to each class $C_i$ according to the attribute $j$ by projection on the corresponding possibility density $\Pi_i^j$,

Merging all the possibility values $\pi_i^1(x^1)$, $\pi_i^2(x^2)$,.., $\pi_i^d(x^d)$ concerning the class $C_i$, into a single one by the aggregation operator "minimum" :

$$\pi_i(x) = \min(\pi_i^1(x^1), \pi_i^2(x^2),.., \pi_i^d(x^d)) \tag{3}$$

The result $\pi_i$ of this fusion corresponds to the global possibility value that $x$ belongs to the class $C_i$. Finally, $x$ is assigned to the class for which it has the maximum membership value.

## 3 Semi Supervised Fuzzy Pattern Matching

The proposed semi supervised FPM has an agglomerative characteristic. Thus, it does not require any prior information about the number of classes. The classes' membership functions are constructed sequentially with the patterns' arrival. According to the ratio $r = \frac{L}{UL + L}$ of the number $L$ of labelled points to the one $UL$ of unlabelled points, the proposed method can be totally supervised, $r$ = 1, or totally unsupervised, $r$ = 0. Let $r_i = \frac{L_i}{UL_i + L_i}$ be the ratio of labelled points $L_i$ belonging to the class $C_i$ to the unlabelled ones $UL_i$ which will be assigned to $C_i$. In the case that $0 \le r_i < 1$, the benefit of semi supervised FPM is to enhance the quality of class's membership estimation thanks to the incorporation of the unlabelled points in this class. This enhancement is performed online thanks to the use of an incremental approach as we can see later. While if $r_i$ = 0, the benefit of semi supervised FPM is to detect this new class and to learn its membership function online. Thus, semi supervised FPM presents benefits in both classification and clustering.

In the case of $r_i$ = 0, the first incoming unlabelled pattern is considered as the point prototype of a new class and its possibilistic membership function according to each attribute is computed as in supervised FPM based on this only pattern. The next unlabelled pattern is either classified in this created class, if it has a membership value according to this class, or considered as a point prototype of a new class. After the classification of each new pattern, the membership function of the corresponding class is updated online using an incremental algorithm. Due to the initialisation, created classes may need to be merged. This merging is performed using a similarity measure. The functioning of semi supervised FPM involves the following two steps.

### 3.1 Classes detection and local adaptation step

Let $x = (x^1, x^2,.., x^d) \in IR^d$ be a given pattern vector in a feature space constituted of $d$ parameters or attributes. There is no learning set containing labelled patterns, nor a prior information about classes' probability density shape or their number. Each attribute is divided into equal intervals defining the bins of the histogram according to this attribute. This histogram is used to estimate the conditional probability density for the class that $x$ is driven from. Let $X_{\min}^j$ and $X_{\max}^j$ be respectively the lower and upper borders of the histogram according to the attribute $j$. These borders can be determined by expert as the minimal and maximal values that an attribute can reach. Let $h$ be the number of histogram's bins, then each bin according to the attribute $j$ has the width :

ISBN: 978-989-95079-6-8

$$\Delta^j = \frac{X^j_{\max} - X^j_{\min}}{h}, \; j \in \{1,2,..,d\} \tag{4}$$

Thus the limits of these bins are defined as follows :

$$b^j_1 = [X^j_{\min}, X^j_{\min} + \Delta^j],\; b^j_2 = [X^j_{\min} + \Delta^j, X^j_{\min} + 2\Delta^j], \\ .., b^j_h = [X^j_{\min} + (h-1)\Delta^j, X^j_{\max}],\; j \in \{1,2,..,d\} \tag{5}$$

The classes detection and local adaptation step involves two strategies : detection of new classes and local adaptation of their membership functions. The local adaptation strategy is based on an update of classes' possibility densities after the classification of each new pattern so that classifier can follow online gradual temporal, or local, changes of classes' membership functions. This online update requires a recursive representation of classes' possibility densities. However the incremental updating cannot detect abrupt changes as changes in the number of clusters. This abrupt change is followed up by the detection strategy which is based on the fact that each new rejected pattern by all the learned classes is considered as a point prototype of a new class. The detection strategy is a mechanism for adjusting the number of clusters online, which is incremented after the detection of each new cluster or class.

*A) Detection of new classes strategy*

The first rejected pattern $x$ according to all the known $c$ classes is considered as the point prototype of the first new class: $C_c \leftarrow x, c \leftarrow c+1$. The PDF is obtained as in supervised FPM. If $x$ is located in the bin $b^j_k, k \in \{1,2,..,h\}$, then the probability histogram of $C_c$ according to the attribute $j$ is : $p^j_c = \{p^j_{c1} = 0, p^j_{c2} = 0,.., p^j_{ck} = 1,.., p^j_{ch} = 0\}$. The possibility histogram will then be computed using (2). Since there is just one pattern, the possibility histogram is equal to the probability one. The possibility density of the class $C_c$ is obtained by a linear linking between the centre of the bin $b^j_k$ , which has the height 1, and the ones of its left $b^j_{k-1}$ and right $b^j_{k+1}$ neighbours, which have both at present the height 0. Generally, if $C = \{C_1, C_2, .., C_c\}$ is the set of learned classes at present, $x$ a new pattern which is rejected by all the learned classes. The detection strategy is defined as follows :

$$\pi_i(x) = 0, \forall i \in \{1,2,..,c\} \Rightarrow c \leftarrow c+1, \\ C_c = \{x\}, \pi_c = \{\pi^1_c,.., \pi^j_c,.., \pi^d_c\} \tag{6}$$

*B) Local adaptation strategy*

For a next pattern $x'$, the membership value to each class $C_i$, $\forall i \in \{1,2,..,c\}$, will be obtained by a projection on its possibility density $\Pi^j_i$ according to each attribute $j$ and then merging the values according to all the attributes using the aggregation operator "minimum" as in supervised FPM. If the membership value $\pi_i(x')$ of $x'$ to the class $C_i$ is different of zero, then this pattern will be assigned to the class $C_i$ and the possibility densities of this class according to each attribute will be incrementally updated. To establish an incremental update of possibility densities, let $p^j_i = \{p^j_{i1}, p^j_{i2},.., p^j_{ik},.., p^j_{ih}\}$ and $\pi^j_i = \{\pi^j_{i1}, \pi^j_{i2},.., \pi^j_{ik},.., \pi^j_{ih}\}$ define respectively the probability and possibility histograms of the class $C_i$ according to the attribute $j$. Let $p'^j_i = \{p'^j_{i1}, p'^j_{i2},.., p'^j_{ik},.., p'^j_{ih}\}$ and $\pi'^j_i = \{\pi'^j_{i1}, \pi'^j_{i2},.., \pi'^j_{ik},.., \pi'^j_{ih}\}$ define respectively the updated probability and possibility histograms of the class $C_i$ according to the attribute $j$ after the assignment of $x'$ to the class $C_i$. Let suppose for the simplicity that : $p^j_{ih} < p^j_{i(h-1)} < .. < p^j_{i1}$, then these new probabilities can be computed incrementally by [14] :

$$x'^j \in b^j_k, \forall k \in \{1,..,h\} \Rightarrow p'^j_{ik} = p^j_{ik} * \frac{N_i}{N_i+1} + \frac{1}{N_i+1}, \\ p'^j_{iz} = p^j_{iz} * \frac{N_i}{N_i+1}, \forall z \in \{1,..,h\}, z \neq k \tag{7}$$

Then the new possibilities can be computed using Dubois and Prade transformation defined by (2). Thus, the local adaptation strategy is defined as follows :

$$\pi_i(x') = \max_{z \in \{1,..,c\}} (\pi_z(x')) \Rightarrow C_i \leftarrow \{C_i, x'\}, \\ \pi'_i = \{\pi'^1_i, \pi'^2_i,.., \pi'^j_i,.., \pi'^d_i\} \tag{8}$$

### 3.2 Classes merging step

The occurrence order of incoming patterns influences the final constructed clusters. This entails the possibility to obtain several different partitions or number of clusters. Thus, several clusters can represent the same class. These clusters must be merged into one cluster to obtain one partition and one membership function. This fusion can be done either by expert or by a merging measure. The later measures the overlap or closeness between constructed clusters. There are different measures for merging clusters in the literature. Most of them are based on a similarity measure between clusters, which takes into account either the degree of overlapping of clusters or the distance between clusters' centres. The clusters overlapping degree is based on the number of ambiguous patterns, belonging to several clusters, and their membership values to these clusters. If the number of these ambiguous patterns is large enough and their membership values to several clusters are high then these clusters cannot be considered as heterogeneous anymore and must be merged. An interesting similarity criterion which takes into account at the same time the number of ambiguous patterns as well as their membership values is defined by (Frigui *el al.* 1996) :

$$\delta_{iz} = 1 - \frac{\sum_{x \in C_i \text{ or } x \in C_z} |\pi_i(x) - \pi_z(x)|}{\sum_{x \in C_i} \pi_i(x) + \sum_{x \in C_z} \pi_z(x)} \tag{9}$$

$\delta_{iz}$ is the fuzzy similarity measure between the classes $C_i$ and $C_z$. More this measure is close to one, more the two clusters are overlapped. We adopt this measure for the merging step of semi supervised FPM.

ISBN: 978-989-95079-6-8

The merging criterion can be applied offline or online. In the first case, several iterations of clustering with a variable number of clusters start with a high over-specified number (an upper bound). Then, the clusters number is reduced gradually until an appropriate number is found. In each iteration, similar clusters are merged and this procedure is repeated until no more clusters can be merged and finished with the optimal number of clusters. In dynamic applications, the merging operation must be done online because the whole unlabelled patterns are not available *a priori*. Thus, the merging criterion can be tested either after the classification of each pattern or in each time window. The problem of the testing after each classification pattern is the calculation complexity which depends on the cardinality of clusters to be merged. However, the clusters are merged online when this measure reaches a predefined threshold. While in the other case, the merging can be delayed according to the size of the time window, but the calculation complexity is less than the one of the first case. We propose to update the fuzzy similarity measure within a time window. This update requires the calculation of the new membership values for all the patterns of all the classes inside which new patterns have been assigned. Indeed, if a new incoming pattern $x$ is assigned to the cluster $C_i$, then the fuzzy similarity measure must be computed only between $C_i$ and the other clusters $C_z$, $z \in \{1,..,c\}, z \neq i$.

When two clusters are merged, their membership functions must also be merged. We propose to merge the membership functions online using an incremental approach. Let $p_i^j = \{p_{i1}^j, p_{i2}^j,.., p_{ik}^j,.., p_{ih}^j\}$, $p_z^j = \{p_{z1}^j, p_{z2}^j,.., p_{zk}^j,.., p_{zh}^j\}$ be respectively the probability distributions of the two clusters $C_i$ and $C_z$ to be merged. Each bin probability is computed as (1) : $p_{ik}^j = \frac{n_{ib_k^j}}{N_i}$, $p_{zk}^j = \frac{n_{zb_k^j}}{N_z}, \forall k \in \{1,..,h\}$. The probability of the bin $b_k^j$ after the merging of the patterns of the two clusters $C_i$ and $C_z$ is equal to :

$$p_{izk}^j = \frac{n_{ib_k^j} + n_{zb_k^j}}{N_i + N_z} \tag{10}$$

Where $n_{ib_k^j}$ and $n_{zb_k^j}$ are respectively the number of patterns of the classes $C_i$ and $C_z$ located in the bin $b_k^j$ according to the attribute $j$. $N_i$ and $N_z$ are respectively the number of patterns of the classes $C_i$ and $C_z$. We can rewrite (10) as follows:

$$\begin{aligned} p_{izk}^j &= \frac{n_{ib_k^j}}{N_i} * \frac{N_i}{N_i + N_z} + \frac{n_{zb_k^j}}{N_z} * \frac{N_z}{N_i + N_z} = \\ & p_{ik}^j * \frac{N_i}{N_i + N_z} + p_{zk}^j * \frac{N_z}{N_i + N_z} \end{aligned} \tag{11}$$

Using (11), the probability distribution of the class after merging is obtained incrementally according to each attribute. Based on (2), the corresponding possibility distributions defining the membership functions can be obtained.

## 4 Experimental Results

Figure 1 presents a simulated data base of two classes in a feature space of two attributes. The two classes are of different sizes, the first has 200 patterns and the second has 100 patterns. The distribution of each class has two independent normal variables with different standard deviations and means. Anyway, the two classes are not overlapped.

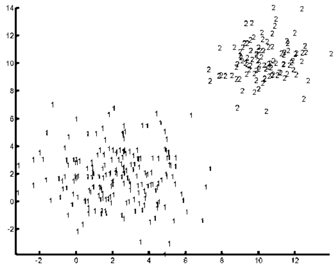

Figure 1: Example of two classes with different sizes, standard deviations and means.

We apply the proposed semi supervised FPM by selecting the patterns with a complete random order. The goal is to test the robustness of our algorithm against the initialisation problem. The experience is repeated several times with a different random pattern's occurrences at each time. We start the experience with a time window which is equal to the size of the data set, i.e., the merging step is applied after the reception of all the available patterns in the data base. Semi supervised FPM detects between 2 to 5 clusters according to the initialisation. Figure 2 shows the result of our algorithm in the case of the detection of 5 clusters.

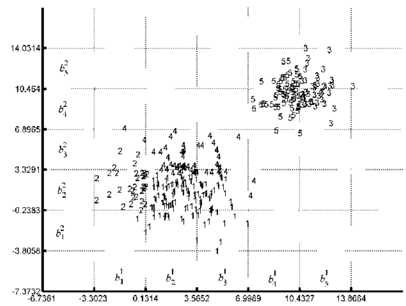


Figure 2: Results of the detection and local adaptation step of semi supervised FPM for the example of Figure 1.

The fuzzy similarity measures between the classes in this case are :

$$\delta_{12} = 0.23,\ \delta_{13} = 0,\ \delta_{14} = 0.69,\ \delta_{15} = 0,\ \delta_{23} = 0,$$
$$\delta_{24} = 0.42,\ \delta_{34} = 0.04,\ \delta_{35} = 0.52,\ \delta_{45} = 0.04$$

The merging threshold $\lambda_{\min} = 0.23$ is sufficient to merge the classes $C_1$, $C_2$ and $C_4$ into one cluster and $C_3$ with $C_5$ into one another cluster in order to obtain at the end the two necessary clusters. Based on the experimentation of several time windows' sizes, a time window of 30 patterns is sufficient to

ISBN: 978-989-95079-6-8

well estimate the correct number of clusters. Anyway, the width of the time window depends of the application: its dynamic, and the initialisation. If this width is too large, then the clusters merging will be delayed, while if it is too small the fuzzy similarity measure value may not be enough to validate the merging.

### *4.1 Classes of non convex shape*

Supervised FPM works well if the classes are separated by at least one attribute. This is due to the fact that the classification of a pattern by FPM is based on a selection of one attribute. Another consequence of the selection of one attribute is that FPM does not respect the shape of classes if this shape is not convex, which is the case of the majority of real applications. Indeed, FPM provides always rectangular membership level curves for all the classes. Figure 3 presents a case for which the classes are not separated by at least one attribute. In addition, the classes $C_1$ and $C_3$ are of non convex shape. Figure 4 presents the results of the application of semi supervised FPM on this data, with a histogram containing 8 bins. Indeed, a higher number of histogram's bins is necessary when the classes are not separated by at least one attribute. A threshold equal to 0.14 is the required one for the merging measure to obtain the three classes.

Figure 5 shows the membership level curves obtained for this data set after the merging of clusters. Here the application of (11) leads to obtain one membership function for each class according to each attribute equivalent to the one resulting by the application of supervised FPM on each class after merging. We can see that these curves do not respect the shape non convex of classes $C_1$ and $C_3$. This, as we said before, is due to the classification decision based on the selection of one attribute. Inspired of the multi-prototypes approach [5] used in the literature to respect the shape of classes, we propose to merge the membership functions of the classes as follows :

$$\pi_{iz} = \min(1, \pi_i + \pi_z) \qquad (12)$$

Where $\pi_{iz}$ is the membership function after the fusion of classes $C_i$ and $C_z$.

Using (12) means that each class is composed of several subclasses. Each subclass keeps its membership function. The application of (12) provides the membership level curves of Figure 6. We can see that these curves respect the classes shape. In addition, no supplementary computation is required to obtain this fusion. When the number of bins $h$ increases, the number of subclasses increases also. This leads to obtain membership level curves which respect more precisely the classes' shape. However, increasing too much $h$ entails the appearance of some membership peaks in the centre of classes.

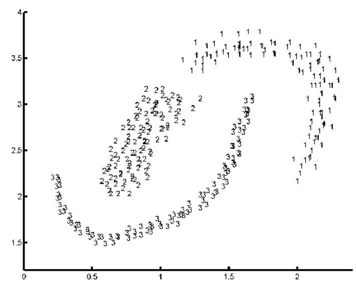

Figure 3: Classes which are not separated by at least one attribute and their shape is not convex (classes $C_1$ and $C_3$).

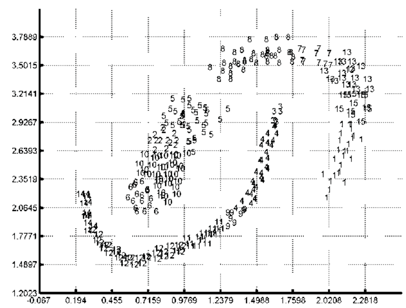

Figure 4: Clusters obtained by the application of semi supervised FPM on the data set of Figure 3.

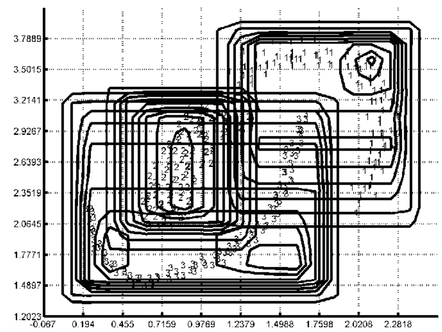

Figure 5: Membership level curves for the example of Figure 3. The membership functions of classes are obtained using (11).

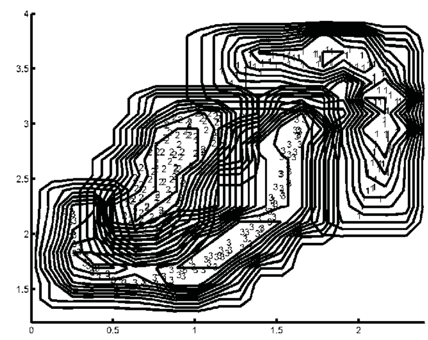

Figure 6: Membership level curves for the example of Figure 3 using (12).

ISBN: 978-989-95079-6-8

### 4.2 *Overlapped classes*

In the case of overlapped classes, the similarity measure becomes useless since the classes are not well separated. In addition, some points can be misclassified according to the ratio of labelled points to the total number of points in data sets. To allow testing the performance of semi supervised FPM in the case of overlapped classes, we take the following artificial 2-dimensioal dataset (Figure 7) available at http://www.stats.ox.ac.uk/pub/PRNN/. It is a normal mixtures data set. The training data consists of two classes with 125 patterns in each class. Each of the two classes has bimodal distribution. The testing set is an independent set of 1000 patterns drawn from the same distribution. The reported misclassification error is based on this testing set. The patterns were selected randomly which means that labelled patterns about some classes may not be presented at all. The experience is repeated 50 times to take into account the effect of initialisation. Table 1 presents the obtained results using semi supervised FPM. We can see that when $r$ increases the misclassification error decreases thanks to the existence in advances of labelled patterns about some classes.

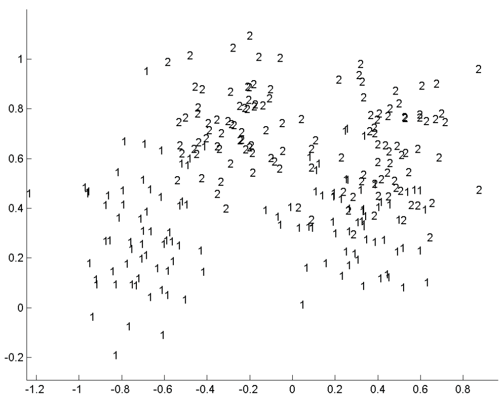


Figure 7: Normal mixtures data set.

Table 1: Misclassification Rate (*MR*) and its STandard Deviation (*STD*) in % according to different values of $r$ for the normal mixture data set.

| $r$ % | 0% | 10% | 20% | 30% | 40% |
|---|---|---|---|---|---|
| *MR* %<br>*STD* % | 13.49<br>12.52 | 11.89<br>3.71 | 9.91<br>1.27 | 9.39<br>1.03 | 9.19<br>0.62 |
| 50% | 60% | 70% | 80% | 90% | 100% |
| 8.84<br>0.39 | 8.61<br>0.32 | 8.43<br>0.30 | 8.26<br>0.27 | 8.16<br>0.2 | 8.10<br>0 |

## 5 Conclusions

In this paper, the supervised classification method Fuzzy Pattern Matching (FPM) is developed to be also an unsupervised classification one. The goal is to obtain a semi supervised classification method adapted to dynamic systems for which a limited prior knowledge is available. Since the unsupervised learning technique is not based on a distance measure, the proposed algorithm will not favour the smaller sized clusters. In addition, it can start with no prior information. Finally, the membership functions can be adapted to elongated and non convex clusters.

We are developing FPM to be operant in the case of non stationary data. Indeed, in many practical situations, the environment changes. A learning data set used to construct the membership functions will be no more valid after a certain time. Thus, the classification method must be able to forget the information which is no more valid or representative of classes and adapt the membership functions based only on the recent and useful one.

## Acknowledgements

This work is integrated in the regional project MOSYP (Performances Measurements and Optimization of Production Systems). The authors thank the region Champagne-Ardennes within this project (CPER ICOS).

**References**

[1] Benediktsson, J. A., J. R. Sveinsson, Feature extraction for multisource data classification with artificial neural networks, Int. journal of remote sensing 18(4) (1997) 727-740.

[2] Bezdek, J. C., N. R. Pal, Some new indexes of cluster validity, IEEE Tran. on Systems, Man and Cybernetics 28(3) (1998) 301-315.

[3] Cayrol, M., H. Farreny, H. Prade, Fuzzy Pattern Matching, Kybernetes 11, (1982) 103-116.

[4] Demiriz, K. P. B., M. J. Embrechts, Semi-supervised clustering using genetic algorithms, Artificial Neural Networks in Engineering, (1999) 809–814.

[5] Devillez, A., M. Sayed Mouchaweh, P. Billaudel, A process monitoring module based on fuzzy logic and pattern recognition, International Journal of Approximate Reasoning 37 (2004) 43-70.

[6] Dubois, D., H. Prade, On possibility/probability transformations, Fuzzy Logic (1993) 103-112.

[7] Duda, R. O., P. E. Hart, D. E. Stork, Pattern Classification second edition, Wiley, New York, (2001).

[8] Frigui, H., R. Krishnapuram, Robust algorithm for automatic extraction of an unknown number of clusters from noisy data, Pattern Recognition Letters 17 (1996) 1223-1232.

[9] Gabrys, B., L. Petrakieva, Combining labelled and unlabelled data in the design of pattern classification systems, International Journal of Approximate Reasoning 35(3), (2004) 251-273.

[10] Jain, A. K., P. Devijver, J. Kittler, Advances in statistical pattern recognition, Pattern Recognition Theory and Applications, Springer-Verlag, Berlin, Heidelberg, (1987) 1-19.

[11] Kate, R. J., R. J. Mooney, Semi-Supervised Learning for Semantic Parsing using Support Vector Machines, Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, NY, (2007) 81-84.

[12] Liu, H., L. Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Trans. on knowledge and data engineering 17 (4) (2005).

[13] Marsili-Libelli, S., Adaptive Fuzzy Monitoring and Fault Detection.Int. Journal of COMADEM 1(3) (1998) 31-37.

[14] Sayed Mouchaweh, M., A. Devillez, G. V. Lecolier, P. Billaudel, Incremental learning in real time using fuzzy pattern matching, Fuzzy Sets and Systems 132(1) (2002) 49-62.

[15] Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility, Fuzzy sets and systems 1(3-2) (1978).

ISBN: 978-989-95079-6-8