

# Multiple Classifiers Approach based on Dynamic Selection to Maximize Classification Performance

O. Ayad and M. Syed-Mouchaweh

**Abstract**—In the past decade, many researchers have employed various methodologies to combine decisions of multiple classifiers in order to achieve high pattern recognition performance. However, two main strategies of combination are possible. The first strategy uses the different opinions of classifiers to make the final decision; it corresponds to classifiers fusion. The second strategy uses the decisions of one or more better classifiers in a specific region of feature space; it corresponds to the selection of classifiers. In this paper, we propose a dynamic multiple classifiers selection system organized in two levels of decision. Two classification methods are used: Semi-Supervised Fuzzy Pattern Matching (SSFPM) and Support Vector Machines (SVM). SSFPM is used to determine the ambiguous regions. Then, the patterns located in these regions are classified by SVM. The detection of the occurrence of new classes and the learning of their membership functions are achieved online using SSFPM. This combination helps to overcome the drawbacks of the both methods by gathering their advantages leading to increase the classification performance.

**Index Terms**— Classifiers Selection, Multiple Classifier Systems, Pattern Recognition, Semi-Supervised Fuzzy Pattern Matching, Support Vector Machine

## I. INTRODUCTION

Multiple classifier combination is a technique that combines the decisions of different classifiers that are trained to solve the same problem, but make independent errors. Many studies of combination systems have been attempted in different applications of Pattern-Recognition (PR) as handwritten character recognition [17], speaker recognition [18], face identification [26], etc. Recently, the concept of Multiple Classifier Systems (MCSs) was proposed as an alternative to obtain high classification performance, i.e., in the sense of a high classification rate based on the decisions of a set of classifiers. Indeed, each individual classifier, i.e. classification method, has its advantages but also its drawbacks. Combining several classifiers can lead to exploit their complementarities and thus to increase their classification performance.

There are two main approaches for the design of MCSs: (1) classifier fusion [10]; and (2) classifier selection [21]. Classifier fusion is based on the assumption that all classifiers make independent errors. Thus, combining the decisions of the different classifiers may lead to increase the overall performance of the system. The classifiers decisions

can be merged using one of the following architectures [29]. The first one is the serial or sequential fusion architecture. In this latter, the fusion is organized in succession stages of decision. One classifier is placed in each stage and it will be activated by the decision of the precedent classifier. The philosophy behind the serial fusion is to use the decision of one classifier to guide one or more other classifiers of the next level. This alternative reduces progressively the final decision ambiguity. The second fusion architecture proposed in the literature is the parallel one. In this latter, classifiers work independently. Then, the classifiers decisions are merged using combination rules. Many combination rules have been proposed based on different principles as the ones based on majority-voting [8], on Bayesian theory [29], on fuzzy rules [29], on belief functions and Dempster-Shafer theory of evidence [4]. In addition, this fusion can be achieved by an aggregation operator as sum, product, maximum and minimum [16]. Finally, the third fusion architecture is the hybrid or mixed one [29]. This latter operates using both serial and parallel architectures. Nonetheless, the condition of classifiers independence cannot be assured. In this case, the combination of classifiers may not lead to improve the final classification performance.

The other approach for the design of MCSs is the Classifier selection. It is based on the assumption that each classifier is an expert in some local regions of the feature space [21]. Thus, no need to satisfy the classifiers error independence assumption. The most locally accurate classifier is selected to estimate the class of each incoming pattern. Classifier selection can be static or dynamic. In the first case, regions of competence are defined during the training phase. While in the second case, they are defined during the classification phase taking into account the characteristics of the new patterns to be classified. However, there are common drawbacks to both selection strategies [11]. When the local expert does not classify the test pattern correctly, there is no way to avoid the misclassification [27]. Moreover, these approaches, often involve high computing complexity as a result of estimating regions of competence.

In this paper, we propose a dynamic classifiers selection approach based on the use of one of the two classification methods: Semi-Supervised Fuzzy Pattern Matching (SSFPM) [22] and Support Vector Machines (SVMs) [6,28]. SSFPM is well adapted to learn online the classes membership functions with a limited training patterns thanks to the use of an incremental learning algorithm. In addition, the occurrence of new classes can be detected and their membership functions can be learned online. However, SSFPM is a marginal classification method. Thus, it is not

Manuscript received March 23, 2011. All authors are with Université de Reims Champagne-Ardenne, Centre de Recherche en STIC (URCA-CReSTIC) Moulin de la Housse, BP 1039, 51687Reims Cedex, France

{omar.ayad,moamar.sayed-mouchaweh}@univ-reims.fr

well adapted for the classes requiring a non linear discrimination. Therefore, we have selected SVM as a second classification method since it is a powerful solution for this problem.

In the training phase, SSFPM is employed to construct the membership function for each class. The goal of this phase is to divide the initial feature space into two regions of competence: non-ambiguous (region of competence of SSFPM) and ambiguous (region of competence of SVM) ones. The training data patterns located in the ambiguous regions are used by SVM to learn the suitable discrimination function leading to separate the classes in these regions. In the classification phase, the selection of SVM or SSFPM depends on the complexity of classification of a pattern. This complexity is represented by the existence of a pattern into an ambiguity area. The final decision of the proposed approach is thus a collaborative one between SSFPM and SVM.

The paper is organized as follows. In section 2, the principals as well as the advantages and drawbacks of SSFPM and SVM are presented. Then in section 3, the proposed dynamic classifier selection approach is detailed. Next in section 4, the performance of the proposed approach is evaluated using simulated and real examples. The last section concludes the paper and presents the future work.

## II. CLASSIFICATION METHODS

### A. Support Vector Machine

The goal of this subsection is to briefly describe Support Vector Machines (SVM). This one is a supervised classification method based on the statistical learning theory [28]. It is initially designed for binary classification problems by constructing a maximum-margin hyperplane that separates two classes of data points. This hyperplane can be generalized later to multi-classification problems.

In the first time, we start with the simplest case; linear separable data. Next, we study the nonlinear separable case.

#### 1) Linear separable case

Let  $(X, C)$  be the set of training patterns defined as follows:

$$X = \{(x_1, C_1), (x_2, C_2), \dots, (x_n, C_n)\}, x \in \mathfrak{R}^d, C \in \{-1, +1\} \quad (1)$$

where  $x_i \in \mathfrak{R}^d$ , is the input vector corresponding to the  $i^{th}$  pattern belonging to one of two classes labeled as  $C = \{-1, +1\}$ , and  $n$  is the number of training patterns. For a linearly separable training set, SVMs look to define a large margin classifier. This margin is defined as the maximum distance between the training patterns of the positive and negative classes. The linear hyperplane is given by:

$$f(x) = wx + b = 0 \quad (2)$$

Sometimes, due to the noise or mixture of classes introduced during data, slack variables are used to reduce the effects of misclassification. Then the equation (2) can be written as

$$C_m(w, x_m) + b \geq 1 - \xi_m, \forall m \in \{1, \dots, n\} \quad (3)$$

where  $w$  determines the orientation of discrimination hyperplane,  $b$  is a bias and  $\xi$  the slack variable. Among the whole possible solutions, SVM choose the hyperplane which maximizes the distance  $2/\|w\|$  between classes, i.e., the optimal hyperplane (Fig 1), that leads to the problem of optimization (4):

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + P \sum_{m=1}^n \xi_m \\ C_m(wx_m + b) \geq 1 - \xi_m \quad \forall m = 1 \dots n, \\ \xi_m \geq 0 \end{cases} \quad (4)$$

This writing is called primal or original problem of optimization. The resolution of (4) requires the adjustment of  $n+1$  parameters. Then, according to the theory of optimization, a problem of optimization has a dual form if it is strictly convex. Since  $\|w\|^2$  is a convex criterion, the solution of (4) can be found while using Lagrange multipliers and its dual problem (5):

$$\begin{cases} \min_{\alpha} \frac{1}{2} \sum_{m,p=1}^n \alpha_m \alpha_p C_m C_p \langle x_m, x_p \rangle - \sum_{m=1}^n \alpha_m \\ \sum_{m=1}^n \alpha_m C_m = 0, m \in \{1, \dots, n\} \\ 0 \leq \alpha_m \leq G \end{cases} \quad (5)$$

$\alpha$ : Lagrange multipliers

Finally, the optimal hyperplane which separate the positive from the negative examples is given by:

$$f(x) = \sum_{m=1}^n \langle w^*, x_m \rangle + b \quad (6)$$

$$w^* = \sum_{m=1}^n \alpha^* x_m C_m \quad (7)$$

$w^*, \alpha^*$ : are the optimal parameters.

The class of a new pattern is given by the sign of the function  $f$ .

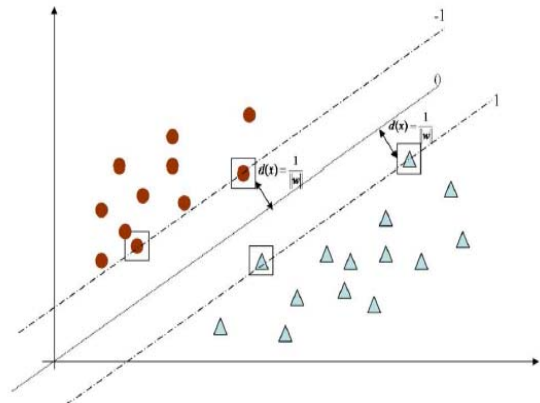


Fig.1. Example of two linearly separable classes in two dimensional feature space.

#### 2) Non linear separable case

In the previous subsection, we supposed that the entire training set example is linearly separable by an optimal

hyperplane. But in the majority of practical problems of classification, the data is non-linearly separable in the initial feature space. For this reason, SVM maps the data points into a higher dimensional feature space in order to achieve a non linear separation between classes using a linear hyperplane. To overcome the curse of dimensionality of the new feature space, SVM uses a Kernel function. SVM of the previous subsection can be generalized to solve non-linear classification problems by substituting the product by a symmetric kernel function  $\Phi(x)$ . The optimization method for the linear case remains valid while replacing  $\langle x_i, x_j \rangle$  by  $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$  in (5). In this case, the separating function is given by:

$$f(x) = \sum_{m=1}^n \alpha_m^* C_m k(x, x) + b^* \quad (8)$$

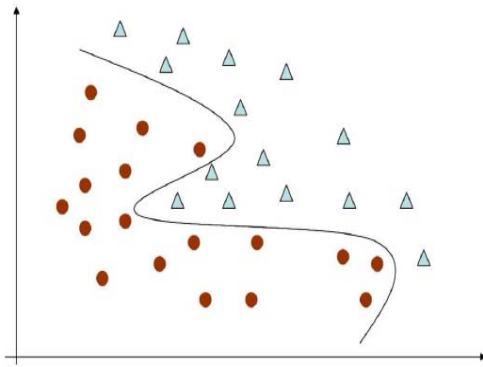


Fig. 2. Example of two non linearly separable classes in two dimensional feature space.

### B. Semi-Supervised Pattern Matching

The Semi-Supervised FPM (SSFPM) 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  $0 \leq r_i < 1$ , 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 initialization, created classes may need to be merged. This merging is performed using a similarity measure. The functioning of semi-supervised FPM is detailed in the next sections. It involves the following steps: classes detection and local adaptation of their membership functions, classes merging and their validation steps.

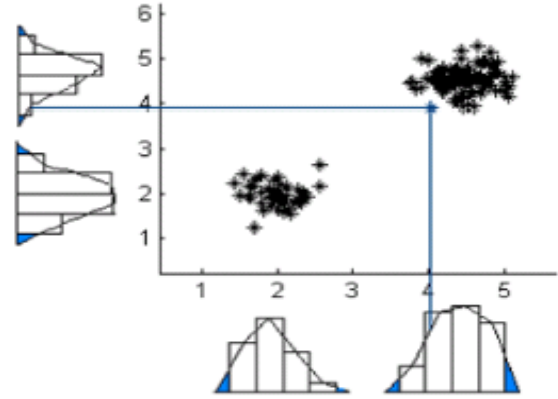


Fig. 3. Projection of a point on the densities of possibility obtained by SSFPM for an example of two classes in a feature space composed of two attributes.

#### 1) Classes detection and local adaptation step

Let  $x = (x^1, x^2, \dots, x^d) \in \mathbb{R}^d$  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:

$$\Delta^j = \frac{X_{\max}^j - X_{\min}^j}{h}, \quad j \in \{1, 2, \dots, d\} \quad (9)$$

Thus the limits of these bins are defined as follows:

$$b_1^j = [X_{\min}^j, X_{\min}^j + \Delta^j], \quad b_2^j = [X_{\min}^j + \Delta^j, X_{\min}^j + 2\Delta^j], \quad \dots, \quad b_h^j = [X_{\min}^j + (h-1)\Delta^j, X_{\max}^j], \quad j \in \{1, 2, \dots, d\} \quad (10)$$

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.

### 2) 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  $b_k^j, k \in \{1, 2, \dots, h\}$ , then the probability histogram of  $C_c$  according to the attribute  $j$  is:  $p_c^j = \{p_{c1}^j = 0, p_{c2}^j = 0, \dots, p_{ck}^j = 1, \dots, p_{ch}^j = 0\}$ . The possibility histogram will then be computed using (10). 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  $b_k^j$ , which has the height 1, and the ones of its left  $b_{k-1}^j$  and right  $b_{k+1}^j$  neighbours, which have both at present the height 0. Generally, if  $C = \{C_1, C_2, \dots, C_c\}$  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:

$$\begin{aligned} \pi_i(x) = 0, \forall i \in \{1, 2, \dots, c\} &\Rightarrow c \leftarrow c+1, \\ C_c = \{x\}, \pi_c = \{\pi_c^1, \dots, \pi_c^j, \dots, \pi_c^d\} \end{aligned} \quad (11)$$

### 3) Local adaptation strategy

For a next pattern  $x'$ , the membership value to each class  $C_i, \forall i \in \{1, 2, \dots, c\}$ , will be obtained by a projection on its possibility density  $\Pi_i^j$  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.

## C. Classes merging and classes validity steps

### 1) 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.

### 2) Classes validity step

After the merging step, the obtained clusters must be validated. This validation is achieved by a validity criterion. The later aims evaluating the degree to which the obtained partition, e.g. classes, approximates the real structure of the data set.

We need to define a validity measure  $V$  in order to evaluate the quality of clusters after the merging step. This measure is based on the cardinality of clusters and their compactness:

$$V = \sum_{i=1}^c \left( \frac{N_i}{N} \left( \sum_{k=1}^{N_i} \pi_i(x_k) \right) \right) \quad (12)$$

$N = \sum_{i=1}^c N_i$  is the number of all the classified patterns.

This measure can give an idea to the expert, at the present moment, about the validity of the obtained clusters after the merging according to the one of clusters before merging. If the validity measure increases after the merging, this means that the obtained clusters are better compact. However, if there is no sufficient number of patterns assigned or absorbed by the cluster, two cases can be arisen. In the first case, the cardinality of the cluster remains insufficient with the course of time. The cluster is a stray or outliers and it must be eliminated. In the second case, the cluster absorbs new patterns through next time windows, leading to increase its compactness, until obtaining sufficient assigned patterns to represent a distinguish functioning mode.

## III. PROPOSED DYNAMIC MULTI-CLASSIFIERS SELECTION APPROACH

The proposed approach is composed of the following steps: (1) learning of membership functions for each class using SSFPM (2) determination of competence regions between classes (3) learning of separation function between classes in ambiguous region using SVMs (4) assignment of new incoming patterns based on the selection of one of the classifiers (SSFPM or SVMs). The steps (1), (2) and (3), constitute the learning phase, while step (4) represents the classification phase. Moreover, this approach is organized in two stages of decision, where everyone adopts one classifier. The two methods engaged by our approach are; Semi-Supervised Fuzzy Pattern Matching and Support Vector Machines. This choice is motivated by the fact; SSFPM is a dynamic and incremental method. But, if the data require a nonlinear separation, SSFPM is not well adapting for this type of data. This observation motivates the search for a suitable method which can properly classify this type of data. Support Vector machines is one of the appropriate methods in this case.

A. Learning phase

Let  $X$  be the set of training patterns  $x$  belonging to  $c$  classes described in a feature space of  $d$  attributes. In this phase (Fig. 4), SSFPM is employed to construct the membership function  $\pi_i$  for each class  $C_i, i \in [1, \dots, c]$ . The goal of this phase is to divide the feature space into two regions of competence: non ambiguous (regions of competence of SSFPM) and ambiguous (regions of competence of SVMs). SSFPM determines these regions as flows:

- estimating of the conditional density of probability of each class according to each attribute by constructing the corresponding probability histogram,
- converting the probability histograms into possibility ones using Dubois-Prade transformation [9]. The advantage of the use of possibility histograms is to take into account both the imprecision and the uncertainty contained in the data [9],
- The membership function  $\pi_i^k$  of each class  $C_i$  according to each attribute  $k \in [1, \dots, d]$  is considered to be the corresponding possibility histogram,
- The possibility membership value  $\pi_i(x)$  of a pattern  $x$  to a class  $C_i$  is calculated as the minimal value of all possibility values of this pattern according to all attributes.
- 

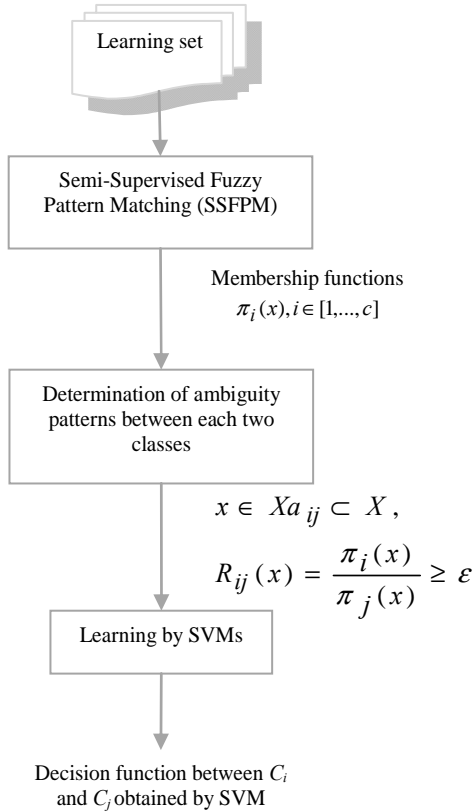


Fig. 4. Architecture of the learning phase in the proposed multi-classifiers system.

The patterns  $x \in Xa_{ij}$  having close membership values  $\pi_i(x), \pi_j(x)$  to two classes  $C_i$  and  $C_j$  are considered to be inside an ambiguous region between these two classes if the following holds:

$$x \in Xa_{ij} \subset X, \pi_i(x) > \pi_j(x), \pi_j(x) \neq 0 \Rightarrow R_{ij}(x) = \frac{\pi_i(x)}{\pi_j(x)} \geq \epsilon, \tag{13}$$

where  $\epsilon$  is a predefined ambiguity threshold. While the other ones are considered as located in non-ambiguous regions. After building the regions of competence, the training patterns located in the ambiguous regions are used to train SVMs.

B. Classification phase

As given at the beginning of this section, the proposed approach requires a classification phase. The architecture of this phase is shown in Fig. 5.

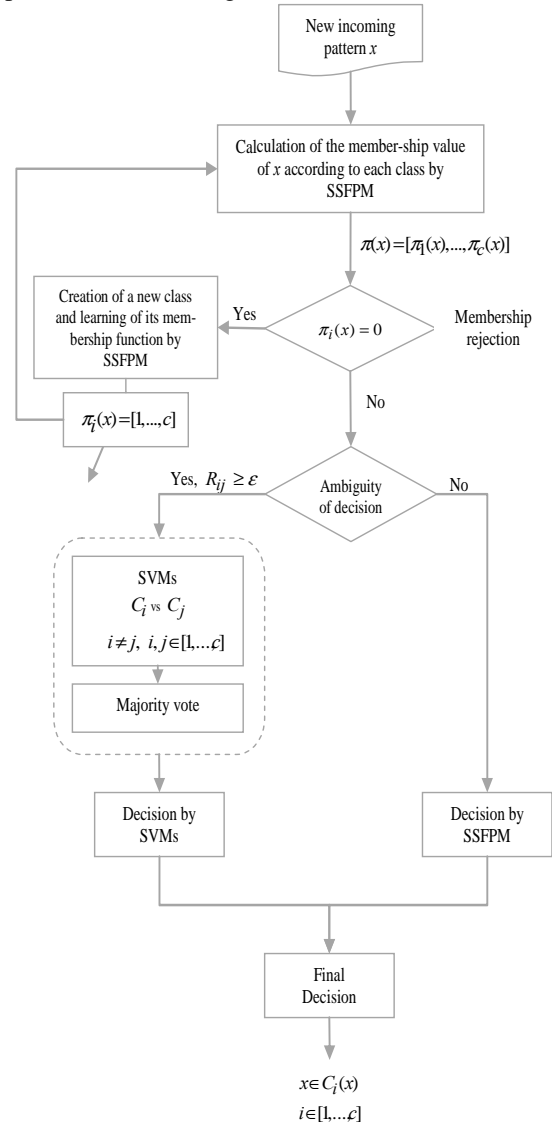


Fig. 5. Architecture of the classification phase in the proposed multi-classifiers system.

For each new incoming pattern, the collaborative classifiers system looks for the proper class of  $x$ . Two cases are possible. The first case corresponds to the one when  $x$  does not belong to any class among the known classes. Thus,  $x$  will be considered as the prototype of a new class. The membership function of this new class will be learned online using SSFPM and the number of classes will be incremented. When a new pattern is assigned to this class, its membership

function will be updated online. In the second case, the pattern has a membership value  $\pi_i(x)$  to several classes  $i \in [1, \dots, c]$ . The classification of this pattern is achieved by SVMs or SSFPM. The selection of SVMs or SSFPM depends on the complexity of classification of a pattern. This complexity is represented by the existence of a pattern into an ambiguity region. The final decision of the proposed approach is thus a collaborative one between SSFPM and SVMs. The classification of this pattern  $x$  is achieved as follows. If  $R_{ij}(x) = \frac{\pi_i(x)}{\pi_j(x)} \geq \epsilon, i \neq j \in [1, \dots, c]$ , then  $x$  will be classified by SVMs. Otherwise, the pattern will be classified using SSFPM. This classification is achieved by using the maximal decision rule:

$$C(x) = C_i \Rightarrow \pi_i(x) = \max[\pi_1(x), \pi_2(x), \dots, \pi_c(x)] \quad (14)$$

#### IV. EXPERIMENTS AND RESULTS

To evaluate the performances of the proposed multi-classifiers system, we have conducted experiments on both synthetic and real datasets. In the first time two type of synthetic data set are used, these datasets are spiral and banana shape datasets. In the second time, to evaluate the performance of the proposed approach, we have used five real datasets.

##### A. Experiments with synthetic data

In this subsection we tested the proposed approach with synthetic data sets. Two types of problems are treated to evaluate the proposed multi-classifiers system. The first data set treat the case of non convex data when the learning dataset and the testing dataset are generated randomly. The goal of this test is to shown the complementarity between SSFPM and SVM. Fig.6 is an example of two classes of non-convex shape with 200 patterns each. The discrimination of these two classes requires a non-linear classifier. In Fig. 6 the 0.1 level membership curves obtained by SSFPM are represented. These curves do not respect the non-convex shape form of the classes. On the other hand, as shown in Fig. 7 SVMs has a boundary of decision well adapted to the form of the classes that permit to build a non-linear separation function.

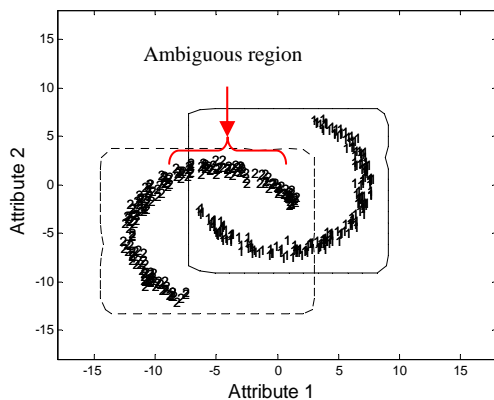


Fig. 6. Banana shape dataset (200 patterns in each class) and 0.1 level membership curves obtained by SSFPM.

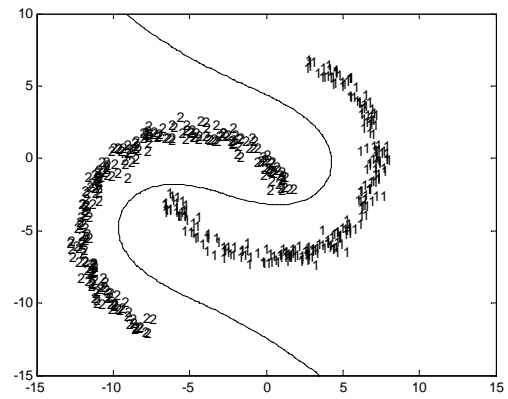


Fig. 7. Separation function calculated by Support Vector Machines

After, to classify this dataset by the proposed architecture two phases are necessary to give a final decision. The first phase consists constructing competence regions of each classifier. We apply the learning set to train SSFPM to construct them. Then this one builds the membership function in the aim to localize the ambiguous zone (Fig.6). The ambiguous region concern data with nearest membership value to the classes  $C_1$  and  $C_2$ . Next all patterns assigned to ambiguous region are used to train SVMs. This one constructs the separation function that separate and maximize the distance between the classes  $C_1$  and  $C_2$  of the ambiguous region. In the second phase, we classify the test dataset. In the beginning SSFPM calculate the membership possibility of the new pattern. If this one is equal to zero, this later will be rejected and it's considered as prototype of possible new class. If not, the proposed approach selects SVMs or SSFPM according to the position of this pattern compared to the ambiguous or non-ambiguous regions. If the new incoming pattern is located in the ambiguous zone, the multiple classifier system selects SVM to classify it. However, if the new incoming pattern is localized in a non ambiguous region, it will be classified by SSFPM. The obtained result is given in the Fig. 8. In Figure 8 the 0.1 level membership curves obtained by the proposed approach are represented. We can notice that these curves respect the non convex shape of classes.

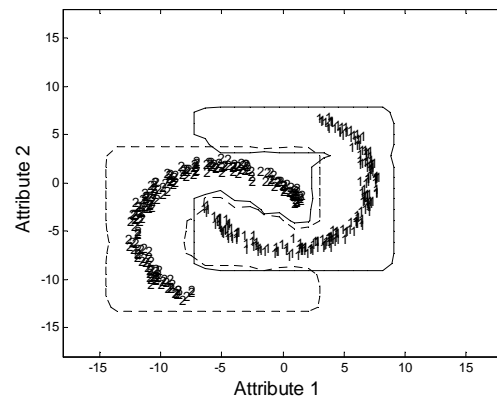


Fig. 8. The optimal nonlinear membership functions that maximize the distance and respect the non-convex shape of the classes generated by the proposed multi-classifiers system.

The occurrence of new classes is another problem tested on the proposed multiple classifier system. The goal of this task is to show the performance of the proposed approach when new classes appear in the course of time. We started by testing the proposed approach on the synthetic example shown in Fig. 6. But we considered the occurrence of a third class. This later creates another ambiguous zone with one class of the preceding example (Fig. 9). As described in the previous section, in the beginning, SSFPM detect the new class and learn its membership functions. Next, the proposed approach evaluates its position in the feature space compared to the known classes. As result of this evaluation, tow case are possible; in the first one there is no ambiguity between the known classes and the occurred one. In the second one, we detect a new ambiguity zone between the new class and one or more other known classes (figure 9). Then, for each new ambiguous region, one SVMs classifier is considered to classify the elements located in these regions. In Figure 10 the 0.1 level membership curves obtained by the proposed approach are represented. We can notice that the occurred class is well detected and their membership function learned online by SSFPM, we can also notice that these curves respect the non convex shape of classes.

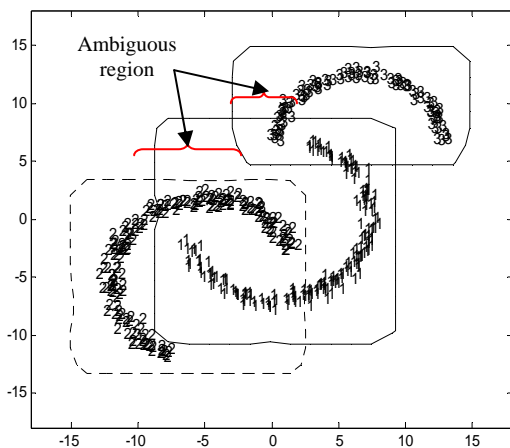


Fig.9. Detection and classification of classes occurred in the course of time by the proposed multi-classifier system.

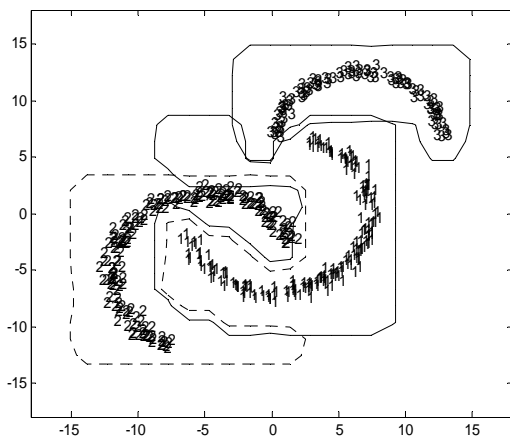


Fig. 10. Decision obtained by the proposed multi-classifier system.

The second type tested concern the case when the data are totally overlapped and require a nonlinear discrimination. We have chosen the 2D spiral data (Fig. 11) to evaluate the proposed approach. This pattern recognition problem is interesting for several reasons: i. the problem is almost impossible to solve using a linear method; ii. Require long training times; iii. the data have some temporal characteristics; vi. several real time applications involving similar data. The data points for the two classes  $C_1$  and  $C_2$  spiral around each other for a total of 151 patterns in each class figure 11.

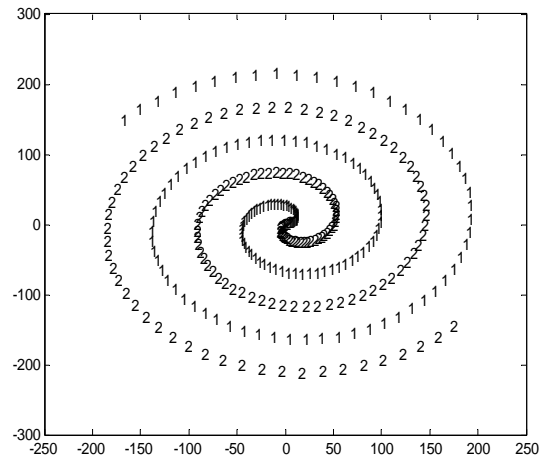


Fig. 11. A two-spiral classification problem with the two classes indicated by '1' and '2' 151 training data for each class.

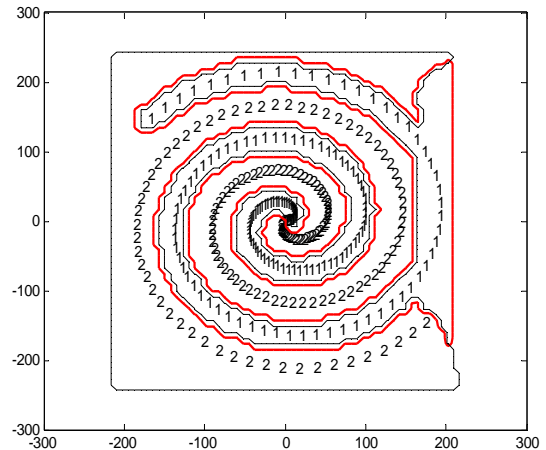


Fig. 12. The membership function achieved by the proposed multi-classifier system.

As shown in the Fig.12, the proposed approach classified the 2D spiral data with success. The success rate is measured in the terms of the test patterns that are correctly predicted to belong to their appropriate class using the training data to make the membership function. Moreover, the leave-one-out method is used in the procedure of classification.

### B. Experiments with real data bases

After the synthetic data, we evaluate in this subsection the proposed approaches on the real data. Table 1 summarizes the main characteristics of each dataset: number of classes, attributes, and patterns, used to evaluate the proposed

approach. The three first datasets are taken from the UCI Machine Learning Database Repository (<http://archive.ics.uci.edu/ml/-datasets.html>). TEP dataset is related to the test and evaluation of the functioning of Tennessee Eastman Process [14]. XOR data are a classical example used in literature to show the correlation between attributes.

TABLE I. DATA SETS USED FOR THE TEST

| Dataset    | Classes | Dim | Per class     |
|------------|---------|-----|---------------|
| Ionosphere | 2       | 34  | {126,76}      |
| Pima       | 2       | 2   | {268,500}     |
| Iris       | 3       | 4   | {50,50,50}    |
| WBC        | 2       | 9   | {444,239}     |
| TEP        | 3       | 52  | {480,480,480} |
| XOR        | 2       | 2   | {400,400}     |

Table 2 recapitulates the results obtained on these datasets. We have used the leave-one-out test strategy to calculate the Misclassification Rate (MR) because it gives a pessimistic and unbiased estimation of MR. We can see that the best result is obtained by the proposed approach.

TABLE II. MISCLASSIFICATION RATE (MR) FOR EACH DATASET USING SVMs, SSFPM AND THE PROPOSED COLLABORATIVE CLASSIFIER APPROACH

| DATA       | METHOD            | MR %  |
|------------|-------------------|-------|
| Ionosphere | SVMs              | 2.01  |
|            | SSFPM             | 7.38  |
|            | Proposed approach | 1.34  |
| Pima       | SVMs              | 27.99 |
|            | SSFPM             | 36.72 |
|            | Proposed approach | 14.84 |
| Iris       | SVMs              | 4     |
|            | SSFPM             | 4.67  |
|            | Proposed approach | 3     |
| WBC        | SVMs              | 20.21 |
|            | SSFPM             | 4.83  |
|            | Proposed approach | 3.22  |
| TEP        | SVMs              | 5.85  |
|            | SSFPM             | 9.46  |
|            | Proposed approach | 4.83  |
| XOR        | SVMs              | 0     |
|            | SSFPM             | 44.12 |
|            | Proposed approach | 0     |

Figure 13 shows the relationship between the ambiguity threshold  $\epsilon$  and the misclassification rate for Pima dataset. We can notice that an ambiguity threshold  $\epsilon = 0.3$  ensures the best classification performance. Based on the experimentations of all the other datasets, the ambiguity threshold must be defined in the interval  $\epsilon \in [0.2, \dots, 0.4]$  to ensure the best classification performance of the proposed approach.

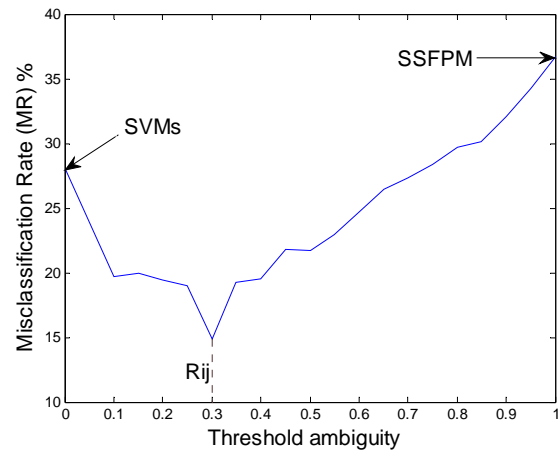


Fig. 13. Relationship between the ambiguity threshold  $\epsilon$  and the misclassification rate for Pima dataset.

### V. CONCLUSIONS

The combination of classifiers is a promising way of research. It is a very effective technique to enhance the performances of classification without increasing the complexity of the methods of classification. The adaptation to the variety and the complexity, of a great number of problems of classifications and for different contexts from application, is another advantage of this idea.

The main contribution of this paper is that proposed a novel multiple classifier selection system to increase the recognition rate. This approach combines SSFPM and SVM methods. Two rejoins of competence are defined one for the non ambiguous data and another for the ambiguous data. These regions are obtained by fuzzy partition based on the membership value calculated with SSFPM. The performance of the proposed method was tested on synthetic data and six well-know real datasets. The detection of the occurrence of new classes is the second task confide to SSFPM method. Our idea marries the advantages of SSFPM and SVM methods to enhance the performances. The second originality of the proposed approach resides in the use of SSFPM to detect and learn the apparition of new class.

We are developing this approach to be adapted to non-stationary data. Moreover, we aim to integrate the active learning in order to take into account the a priori knowledge provided by experts.

### REFERENCES

- [1] A. Al-Ani, M. Deriche, A New Technique for Combining Multiple Classifiers using The Dempster-Shafer Theory of Evidence. *Journal of Artificial Intelligence Research*, 2002, pp. 333-361.
- [2] A. Giani, Parallel cooperative classifier systems. *Ph. D. Thesis*, Université de Pise, Italie.1999.
- [3] Bezdek, J. C., N. R. Pal, Some new indexes of cluster validity, *IEEE Transaction on Systems, Man and Cybernetics*, 1998, pp 301-315.
- [4] B. Quost, M.H. Masson T. Denoeux, 2008. Refined classifier combination using belief functions. *11<sup>th</sup> International Conference on Information Fusion, Allemagne*.2008.
- [5] C.C.Chang and C.J.Lin, LIBSVM: a library for support vector machines. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>. 2001



- [6] C.J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 1998, pp. 121–167.
- [7] C.L. Liu, Classifier combination based on confidence transformation. *Pattern Recognition*, 2005, pp. 11 – 28.
- [8] D. Bahler, L. Navarro, Methods for combining heterogeneous sets of classifiers. *17th Natl. Conf. Artificial Intelligence (AAAI 2000)*, 2000.
- [9] D. Dubois, H.Prade, On possibility/probability transformations. *International Journal of Fuzzy Logic*, 1993, pp. 103-112.
- [10] D. Ruta, B.Gabrys, An overview of classifier fusion methods. *Computing and Information Systems*, 2000, pp. 1-10.
- [11] E.M. Dos Santosa, R. Sabourina, 2008 .A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. *Pattern Recognition*, 2008, pp. 2993 – 3009.
- [12] Frelicot, C., L. Mascarilla, M. Beithier, 2006.A New Cluster Validity Index for Fuzzy Clustering based on Combination of Dual Triples, *IEEE International Conference on Fuzzy Systems*, Vancouver, BC, Canada, 2006, pp 42-47.
- [13] Frigui, H., R. Krishnapuram, Robust algorithm for automatic extraction of an unknown number of clusters from noisy data, *Pattern Recognition Letter*, 1996, pp 1223-1232.
- [14] F.Roli, Semi-supervised multiple classifier systems: Background and research direction. Multiple classifier systems. *International workshop No6*, Seaside CA, ETATS-UNIS, 2005.
- [15] G. Giacinto, F. Roli, Methods for dynamic classifier selection. ICIAP'99, *10th Int. Conf. Image Analysis and Processing*, Venice, Italy, 1999.
- [16] J.Kittler, M.Hatef, R.P.W.Duin, J.Matas, On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell*, 1998, pp. 226-239.
- [17] J.Milgram, R.Sabourin, M.Chretien, Combining Model-based and discriminative approaches in a modular tow-stage classification system: application to isolated handwritten digit recognition. *Electronic Letters on Computer Vision and Image Analysis*, 2005, pp. 1-15.
- [18] K.Chen, L.Wang, H.Chi, Methods of combining multiple classifiers with different features and their application to text-independent speaker identification. *International Journal of Pattern Recognition and Artificial Intelligence*, 1997, pp. 417-445.
- [19] Lu, F. Yamaoka. Fuzzy integration of classification results. *Pattern Recognition*, 1997.
- [20] L. Xu, A. Krzyzak, C. Y .SUEN, Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE transactions on systems, man, and cybernetics*, 1992, pp. 418-435.
- [21] M. Aksela. Comparison of classifier selection methods for improving committee performance. *In Lecture Notes in Computer Science, 4<sup>th</sup> International Workshop, Multiple Classifier Systems*, 2003, pp. 84-93, June, Guildford, UK.
- [22] M.Sayed-Mouchaweh. Semi-supervised classification method for dynamic applications. *Fuzzy Sets and Systems*, 2010, pp. 544–563.
- [23] M. Sayed Mouchaweh, A. Devillez, G. V. Lecolier, P. Billaudel. Incremental learning in real time using fuzzy pattern matching. *Fuzzy Sets and Systems*, 2002, pp. 49-62.
- [24] M.Y. Nam, S. SEDAI, P.K. Rhee. Cascade of fusion for adaptive classifier combination using context-awareness. *In Lecture notes in computer science*, 2006, pp. 252-261.
- [25] P.Gancarski, C.Wemmert. Collaborative multi-step mono level multi-strategy classification. *Journal on Multimedia Tools and Applications*, 2007, pp. 1-27.
- [26] R. Brunelli, D. Falavigna. Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, pp. 955–966.
- [27] R. Liu, B. Yuan. Multiple classifiers combination by clustering and selection. *Information fusion*, 2000, pp. 163-168.
- [28] V.N. Vapnik. The nature of statistical learning theory, Springer-Verlag. New York, 1995.
- [29] V.Gunes, M.Menard, P.Loonis, and S. Petit-renaud. “Combination, Cooperation and Selection of Classifiers: A State of the Art,” *International Journal of Pattern Recognition and Artificial Intelligence*, 2003 ,pp. 1303–1324.
- [30] X. Zhu, X. Wu, Y. Yang. Dynamic selection for effective mining from noisydata streams, in: *Proceedings of Fourth IEEE International Conference on Data Mining*.2004, pp. 305-312.