

Sélection dynamique des classifieurs pour l'amélioration du taux de classification dans les zones d'ambiguïté

Omar AYAD, Moamar SAYED-MOUCHAWEH, Patrice BILLAUDEL

Université de Reims Champagne-Ardenne (URCA), CReSTIC,
Moulin de la Housse, BP 1039, 51687 Reims Cedex, France
omar.ayad@univ-reims.fr
moamar.sayed-mouchaweh@univ-reims.fr
patrice.billaudel@univ-reims.fr

Résumé— Dans ce papier, nous proposons un système de sélection dynamique des classifieurs pour améliorer le taux de classification dans les zones d'ambiguïté. Ce système combine les méthodes de classification Fuzzy Pattern Matching Semi Supervisée (FPMSS) et Séparateurs à Vaste Marge (SVM). Ce choix est justifié par le fait que FPMSS est une méthode semi supervisée, incrémentale et dynamique ce qui permet de détecter en ligne les nouvelles classes, d'apprendre et ensuite d'adapter leur fonctions de décision. Cependant, FPMSS n'est pas adaptée aux problèmes nécessitant une séparation non linéaire entre les classes. SVM est une méthode performante et adaptée pour la séparation non linéaire. La sélection d'une de ces méthodes est dynamique et elle dépend de la complexité de la classification d'un nouveau point. Cette complexité est représentée par l'existence du nouveau point dans une zone d'ambiguïté. Ainsi, cette combinaison permet d'obtenir un classifieur dynamique non linéaire et semi-supervisé.

Mots-clés— Reconnaissance des formes, Combinaison des classifieurs, Fuzzy Pattern Matching Semi Supervisée, Séparateur à Vaste Marge.

I. INTRODUCTION

La première application des systèmes multi-classifieurs n'est apparue que dans les années 90. Cette technique est devenue l'une des voies de recherche les plus prometteuses et les plus soutenues dans le domaine de la reconnaissance des formes. Le but de ces combinaisons est d'améliorer les performances (précision, fiabilité, robustesse...) des systèmes de classification en exploitant la complémentarité qui peut exister entre les classifieurs sans augmenter la complexité des méthodes utilisées. Par conséquent, l'utilisation simultanée de plusieurs méthodes peut, éventuellement permettre d'en cumuler les avantages sans en cumuler les inconvénients.

Le problème de l'association de plusieurs classifieurs peut être abordé selon deux approches. La première approche cherche à fusionner les résultats de plusieurs classifieurs afin d'obtenir un meilleur taux de reconnaissance. La seconde approche consiste à choisir les meilleurs classifieurs parmi un ensemble de classifieurs. De là, dérivent les deux problématiques de la sélection [20] et de la fusion [2].

- La fusion des classifieurs : consiste à réaliser une association des classifieurs en parallèle ou en cascade [14]. Dans la première architecture, les classifieurs sont appliqués indépendamment dans un premier temps, puis on fusionne leurs sorties pour atteindre un consensus. Cette fusion peut être soit démocratique, dans le sens

où elle ne favorise aucun classifieur par rapport à un autre, soit au contraire dirigée et, dans ce cas, on attribue à la réponse de chaque classifieur un poids en fonction de ses performances. Plusieurs lois de combinaison ont été proposées dans la littérature. Parmi ces lois, on trouve le vote majoritaire [3], la théorie Bayésienne [3], la théorie de Dempster-Shafer [1], les règles fixes comme le maximum, le minimum, la somme, la moyenne et le produit [6]. La deuxième architecture appelée combinaison séquentielle ou série, est organisée en niveaux successifs de décision permettant de réduire progressivement le nombre de classes possibles. L'objectif de cette technique est d'utiliser la décision d'un classifieur placé en amont pour mieux guider un ou plusieurs autres classifieurs placés en aval.

- La sélection des classifieurs : cette approche considère chaque classifieur comme un expert dans une région spécifique de l'espace de représentation [2]. Ces méthodes cherchent à prédire le classifieur le mieux adapté à un nouveau point à classer. Seule la sortie du classifieur sélectionné est prise en compte dans la décision finale. La sélection des classifieurs peut être statique [2] ou dynamique [11]. Dans le premier cas, le classifieur engagé par le système de classification est celui défini dans la phase d'apprentissage. Par contre, la classification dynamique choisit le classifieur le plus adapté à un point à classer pendant la phase de classification.

Dans ce papier, nous proposons un système de sélection dynamique des classifieurs pour améliorer le taux de classification dans les zones d'ambiguïté. Ce système combine la méthode Fuzzy Pattern Matching Semi Supervisée (FPMSS) [21] et la méthode des Séparateurs à Vaste Marge (SVM) [23]. Ce choix est justifié par le fait que FPMSS est une méthode semi supervisée, incrémentale et dynamique ce qui permet de faire l'apprentissage en ligne. Par contre FPMSS n'est pas adaptée aux problèmes nécessitant une séparation non linéaire. SVM est l'une des méthodes qui ont montré de grandes capacités à résoudre ce type de problème. La sélection d'une de ces méthodes est dynamique et elle dépend de la complexité de classification d'un nouveau point. Cette complexité est représentée par l'existence du nouveau point dans une zone d'ambiguïté. Pour cela, deux régions de compétences sont construites à par-

tir de l'espace de représentation. La première contient les points non ambigus (région de compétence pour FPMSS) et la deuxième sera réservée aux points ambigus (région de compétence pour SVM). Ainsi, cette combinaison permet d'obtenir un classifieur dynamique non linéaire et semi-supervisé.

Ce papier est organisé comme suit : dans la deuxième section, nous décrivons les principes généraux des méthodes utilisées. Ensuite, nous développons notre approche de combinaison dans la troisième section. Dans la section 4, nous évaluons cette approche sur un exemple d'illustration académique et sur cinq bases de données réelles. Enfin, la conclusion et les perspectives terminent ce papier.

II. MÉTHODES DE CLASSIFICATION UTILISÉES

A. Séparateurs à Vaste Marges (SVMs)

Parmi les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik[22],[4], la méthode SVM constitue la forme la plus connue. SVM est une méthode de classification par apprentissage supervisé, elle fut introduite par Vapnik en 1995 [22]. Puisque c'est une méthode de classification supervisée, SVM fait appel à un jeu de données d'apprentissage, où chaque point est associé à sa classe, pour apprendre les paramètres du modèle afin de trouver le séparateur optimal entre les différentes classes.

A.1 SVM linéaire

A partir d'un ensemble d'apprentissage (X, C) où $X \in \mathbb{R}^n$ ensemble des données et $C \in \{-1, 1\}$, celui des indexes des classes. Chaque classe comporte n points ou formes. Le SVM consiste à trouver l'Hyperplan Séparateur Optimal qui maximise la distance entre l'hyperplan et les deux classes. Cette distance est appelée marge.

L'hyperplan est défini par $w \cdot x + b = 0$ où (w, b) désignent les paramètres de l'hyperplan (respectivement un vecteur normal au plan et le biais). Le classifieur est donnée par :

$$f : X \in \mathbb{R}^n \mapsto \text{sign}(w \cdot x + b) \in \{-1, +1\}, \forall x \in X \quad (1)$$

Néanmoins, ce dernier doit satisfaire :

$$\begin{cases} w \cdot x + b \geq 0, \text{ si } C = +1 \\ w \cdot x + b \leq 0, \text{ si } C = -1 \end{cases} \quad (2)$$

De plus, l'hyperplan doit maximiser la marge : $2/\|w\|$. La solution est obtenue en résolvant le problème de maximisation sous les contraintes précédentes, ce qui mène au problème d'optimisation quadratique suivant :

$$\begin{cases} \min_{(w,b)} : \frac{\|w\|^2}{2} \\ C_m(w \cdot x_m + b) \geq 1, m \in \{1, \dots, n\} \end{cases} \quad (3)$$

En introduisant les multiplicateurs de Lagrange, on obtient le problème dual :

$$\begin{cases} \max_{\alpha} w(\alpha) = \frac{1}{2} \sum_{m,t} \alpha_m \alpha_t y_m y_t \langle x_m, x_t \rangle - \sum_m \alpha_m \\ \sum_m \alpha_m C_m = 0, \alpha_m \geq 0, m \in \{1, \dots, n\} \end{cases} \quad (4)$$

Qui est un problème d'optimisation quadratique convexe sous contraintes linéaires.

L'hyperplan optimal est obtenu par $w^* = \sum_{m=1}^n \alpha_m^* x_m C_m$ et $b^* = C_m - w \cdot x_m$. La fonction de classification vaut donc :

$$f(x) = w^* \cdot x_m + b^* = \sum_{m=1}^n \alpha_m^* C_m x_m \cdot x + b^* \quad (5)$$

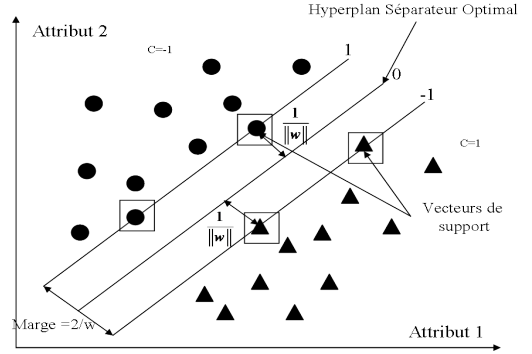


Fig. 1. Hyperplan Séparateur Optimal

Pour des données non linéairement séparables ou même dans le cas linéairement séparables, il peut être préférable de violer quelques contraintes pour agrandir la marge. Ainsi, on améliore la robustesse au détriment de l'erreur d'apprentissage. Pour ce faire, on introduit les variables de relaxation ou relâchement et le paramètre de régularisation ζ . Le problème d'optimisation primal (3) devient :

$$\begin{cases} \min_{(w,b)} : \frac{\|w\|^2}{2} + \zeta \sum_{m=1}^n \xi_m \\ C_m(w \cdot x_m + b) \geq 1 - \xi_m, \xi_m \geq 0, m \in \{1, \dots, n\} \end{cases} \quad (6)$$

Ce problème d'optimisation à l'aide des multiplicateurs de Lagrange et devient :

$$\begin{cases} \max_{\alpha} w(\alpha) = \frac{1}{2} \sum_{m,t} \alpha_m \alpha_t y_m y_t \langle x_m, x_t \rangle - \sum_m \alpha_m \\ \sum_m \alpha_m C_m = 0, 0 \leq \alpha_m \leq \zeta, m \in \{1, \dots, n\} \end{cases} \quad (7)$$

A.2 SVM non linéaire

Dans la plupart des problèmes de classification, les données sont non linéairement séparables. Donc, l'hyperplan optimal qui sépare les différentes classes doit pouvoir prendre une forme plus compliquée. La méthode du noyau est un moyen élégant et efficace pour traiter ce problème : on transfère les données de l'ensemble de départ $X \in \mathbb{R}^n$ vers un ensemble de dimension supérieure $\Phi(X) \in \mathbb{R}^M$ dans lequel le problème devient séparable linéairement.

Le nouveau problème est obtenu en remplaçant les vecteurs de l'ensemble des données qui interviennent seulement dans le produit scalaire $\langle x_m, x_t \rangle$ par leurs correspondants dans le nouvel espace caractéristique $\langle \Phi(x_m), \Phi(x_t) \rangle$. Ainsi, on introduit la fonction noyau dans l'équation (1) : $K(x_m, x_t) = \langle \Phi(x_m), \Phi(x_t) \rangle$ et la nouvelle fonction de classification est donnée par :

$$f(x) = \sum_{m=1}^n \alpha_m^* y_m K(x_m \cdot x) + b^* \quad (8)$$

Il n'est cependant pas nécessaire de connaître explicitement la fonction Φ : toute fonction qui satisfait la condition de Mercer [4] peut être utilisée comme noyau. Dans les problèmes de classification les noyaux les plus souvent utilisés sont :

1. le noyau polynomial : $K(x, x') = ((x \cdot x') + b)^q$
2. le noyau gaussien : $K(x, x') = \exp(-\frac{\|x-x'\|^2}{2\sigma^2})$

B. Fuzzy Pattern Matching Semi Supervisée (FPMSS)

La méthode FPM Semi Supervisée (FPMSS) [21] est une version développée de FPM [5]. En fonction du rapport r entre le nombre de points étiquetés et le nombre de points qui seront classifiés, la méthode FPM peut être totalement non supervisée si $r = 0$, semi supervisée si $r < 1$ et supervisée si $r = 1$. Le fonctionnement de FPMSS passe par une phase d'apprentissage, puis par une phase de détection et d'adaptation, et enfin par une phase de fusion.

B.1 Phase d'apprentissage

Soit X_i l'ensemble des points d'apprentissage appartenant à la classe C_i . Chaque classe C_i comporte N_i points, ou formes, dans un espace de représentation, $w \in \mathbb{R}^d$, de d attributs. La phase d'apprentissage consiste en l'estimation des probabilités conditionnelles de chaque classe par rapport à chaque attribut via l'établissement d'histogrammes. Les bornes supérieures et inférieures pour chaque attribut j sont fixées initialement par un expert. De la même façon, le nombre h , de barres b_k^j , $k \in \{1, 2, \dots, h\}$, d'un histogramme est déterminé expérimentalement. La largeur d'une barre Δ^j selon l'attribut j est alors définie par :

$$\Delta^j = \frac{(\max^j - \min^j)}{h} \quad (9)$$

L'histogramme ou la distribution de probabilité $\{p_i^j(b_k^j), i \in \{1, 2, \dots, c\}, j \in \{1, 2, \dots, d\}, k \in \{1, 2, \dots, h\}\}$ pour la classe C_i par rapport à l'attribut j est déterminé en calculant la probabilité $p_i^j(b_k^j)$ de chaque barre b_k^j comme suit :

$$p_i^j(b_k^j) = \frac{n_{ik}^j}{N_i} \quad (10)$$

Où n_{ik}^j est le nombre de points d'apprentissage de la classe C_i se trouvant dans la barre et N_i est le nombre total des points de la classe C_i .

Ensuite, l'histogramme de probabilité est converti en distribution de possibilité $\{\pi_i^j(b_k^j), i \in \{1, 2, \dots, c\}, j \in \{1, 2, \dots, d\}, k \in \{1, 2, \dots, h\}\}$ en utilisant la transformation de probabilité en possibilité de Dubois et Prade [9] définie par :

$$\pi_i^j(b_k) = \sum_{z=1}^h \min(p_i(b_z), p_i(b_k)) \quad (11)$$

Finalement, la densité de possibilité Π_i^j de la classe C_i par rapport à l'attribut j est déduite par l'interpolation linéaire entre les centres des barres de l'histogramme de possibilité.

Les histogrammes des classes sont mis à jour séquentiellement avec l'arrivée de chaque nouvelle forme. L'insertion d'une forme x dans une classe C_i connue ou dans une nouvelle classe se fait en trois étapes en fonction de sa valeur d'appartenance :

1. détermination de la possibilité d'appartenance de x à la classe C_i selon l'attribut j . Cette possibilité est calculée par simple projection de x_j sur la densité de possibilité Π_i^j (Fig. 2).

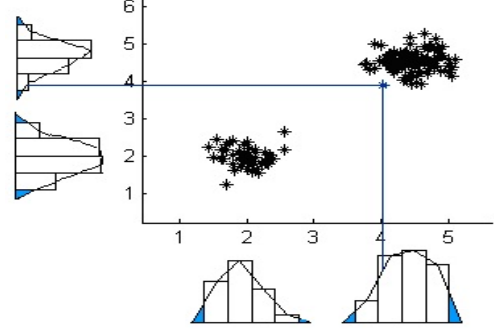


Fig. 2. Projection d'un point sur les densités de possibilité calculées par FPM pour un exemple de deux classes dans un espace de représentation de deux attributs.

2. fusion, pour chaque classe C_i , de toutes les valeurs de possibilité d'appartenance $\pi_i^1, \pi_i^2, \dots, \pi_i^d$ par l'opérateur d'agrégation "min". Le résultat de cette fusion représente la possibilité d'appartenance π_i de x à la classe C_i .

3. affectation du point x à la classe pour laquelle il a la valeur de possibilité d'appartenance la plus élevée et mise à jour des densités de possibilité de manière incrémentale. Si le point est situé dans la barre b_k^j , la nouvelle probabilité $p_i^j(b_k^j)$ de cette barre est :

$$p_i^j(b_k^j) = \frac{n_{ik}^j}{N_i} \times \frac{N_i}{N_i + 1} + \frac{1}{N_i + 1} = p_i^j(b_k^j) \times \frac{N_i}{N_i + 1} + \frac{1}{N_i + 1} \quad (12)$$

Pour les autres barres, la nouvelle probabilité est :

$$p_i^j(b_z^j) = \frac{n_{iz}^j}{N_i} \times \frac{N_i}{N_i + 1} = p_i^j(b_z^j) \times \frac{N_i}{N_i + 1}, z = 1, \dots, h, z \neq k \quad (13)$$

B.2 Phase de détection et adaptation

Lorsque le point a une valeur d'appartenance nulle par rapport à toutes les classes, ce point est alors considéré comme le prototype d'une nouvelle classe et la fonction d'appartenance de cette classe est calculée selon chaque attribut en se basant seulement sur ce point. Si le point suivant est classé dans cette nouvelle classe, parce que son degré de possibilité d'appartenance est le plus élevé pour cette classe, la fonction d'appartenance de la classe sera alors mise à jour de manière incrémentale, en utilisant (12) et (13), afin d'affiner son estimation.

Supposons que x a été rejeté par rapport à toutes les classes connues : $\pi_i(x) = 0, \forall i \in \{1, 2, \dots, c\}$. Une nouvelle classe sera créée pour x et le nombre de classes connues est incrémenté : $c = c + 1$. Si x se trouve dans la barre

$b_k^i, k \in \{1, 2, \dots, h\}$, alors, l'histogramme de probabilité, calculé en utilisant le seul point x , pour cette nouvelle classe par rapport à chaque attribut j est :

$$p_i^j = \left\{ p_i^j(b_1^j) = 0, p_i^j(b_k^j) = 1, \dots, p_i^j(b_h^j) = 0 \right\} \quad (14)$$

L'histogramme de possibilité est ensuite calculé par (11).

B.3 Phase de fusion des classes

Plusieurs partitions peuvent être obtenues selon l'ordre de présentation des nouveaux points. Une classe peut donc être représentée par plusieurs classes proches les unes des autres. Dans ce cas, ces classes doivent être fusionnées afin d'obtenir un nombre de classes occupant des régions distinctes dans l'espace de représentation. Cette fusion peut être réalisée en utilisant la mesure de similarité entre les classes suivante :

$$\delta_{iz} = 1 - \frac{\sum_{x \in C_i \text{ ou } x \in C_z} |\pi_i(x) - \pi_z(x)|}{\sum_{x \in C_i} \pi_i(x) - \sum_{x \in C_z} \pi_z(x)} \quad (15)$$

Où $\pi_i(x)$. et $\pi_z(x)$. ont respectivement les valeurs d'appartenance de x à la classe C_i et C_z . δ_{iz} est la mesure de similarité entre ces deux classes. Plus la valeur de cette mesure est proche de 1, plus les deux classes sont similaires et doivent donc être fusionnées.

III. SYSTÈME DE SÉLECTION DES CLASSIFIEURS

Dans ce travail, nous proposons un système de sélection dynamique des classifieurs. Cette approche combine deux méthodes de classification. Les méthodes engagées par notre approche sont : la méthode Fuzzy Pattern Matching Semi Supervisée et la méthode des Séparateur à Vaste Marge. La philosophie de la méthode d'agrégation que nous proposons, consiste à dire que les méthodes ont intérêt à coopérer entre eux dans l'espoir de minimiser le plus possible l'erreur de classification.

Le principal but de notre méthode est de trouver l'ensemble des données la plus adapté à chacun des classifieurs pour obtenir le meilleur taux de classification possible. Dans ce travail on distingue deux types de données. Les données dite ambiguës et les données non ambiguës.

L'approche proposée pour la sélection dynamique des classifieurs se compose de quatre étapes : (1) apprentissage des fonctions d'appartenance pour chaque classe (2) construction des régions de compétence pour chaque classifieur (3) apprentissage de l'hyperplan qui sépare les données ambiguës par SVM (4) classification basé sur la sélection dynamique des classifieurs (FPMSS ou SVM). Les étapes (1), (2) et (3) constituent la phase d'apprentissage (Fig 3), cependant l'étape (4) représente la phase de classification pour l'approche proposée (Fig. 4). En plus, cette approche est organisée en deux niveaux de décisions. Chaque niveau adopte un classifieur (FPMSS ou SVM). Le choix de FPMSS et SVM est motivé par le fait que FPMSS est une méthode incrémentale, adaptative et multiclasse. En plus dans le cas des données linéairement séparables, FPMSS donne des bons résultats. Cependant, FPMSS n'est pas opérante pour le cas des classes non linéairement séparables. Cette remarque nous a poussés

à la recherche d'une autre méthode complémentaire avec FPMSS qui prendra le relais dans les zones d'ambiguïté (données généralement non linéaires). Parmi les méthodes existantes, SVM est l'une des méthodes qui ont montré de grande capacité à résoudre ce type de problème.

A. Phase d'apprentissage

A partir de l'ensemble des données initiales, nous construisons deux sous ensemble, une pour les données ambiguës et une autre pour les données non ambiguës (figure 3). On entraîne FPMSS sur la base de données initiale. FPMSS construit les densités de possibilité pour chaque classe connue. En fonction de cette dernière, les régions de compétences seront délimitées. Ensuite, nous définissons le rapport des possibilités d'appartenance entre chaque deux classe $Ra(x) = \pi_i(x)/\pi_j(x)$, $ij \in [1, \dots, c]$ et $i \neq j$. Le rapport Ra sera comparé avec un seuil de tolérance prédéfini ϵ . Cependant, un point est dit ambigu si ses possibilités d'appartenance à la classe C_i et C_j sont proches. Le but de cette phase est de diviser l'espace de représentation initial en deux régions de compétences, une pour les données ambiguës (région de compétence de SVM) et une autre pour les données non ambiguës (région de compétence de FPMSS). Nous utilisons les données qui forment la zone d'ambiguïté pour entraîner SVM. Cette dernière calcule l'hyperplan qui sépare les classes.

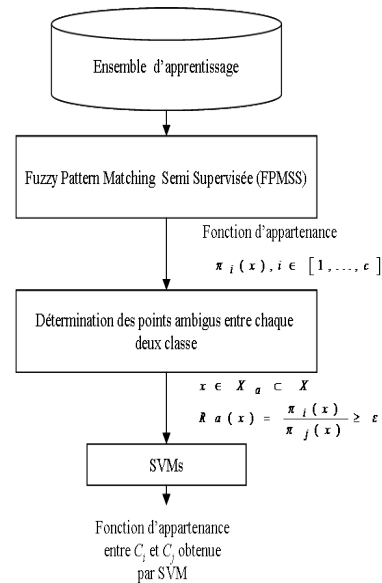


Fig. 3. Phase d'apprentissage

B. Phase de classification

Comme chaque système de classification, l'approche proposée requiert une phase de classification. L'architecture de cette phase est donnée dans la figure 4. Pour chaque nouveau point x on calcule son degré d'appartenance par rapport à toutes les classes connues $\pi_i(x), \forall i \in \{1, 2, \dots, c\}$. Deux cas sont possibles selon la valeur de $\pi_i(x)$. Si le point a une valeur d'appartenance nulle par rapport à toutes les classes, ce point est alors considéré comme le prototype d'une nouvelle classe et la fonction d'appartenance de cette classe est calculée en ligne selon chaque attribut en se basant seulement sur ce point. Par contre, si $\pi_i(x) \neq 0$ on

calcule le rapport d'ambiguïté Ra qui est défini comme le rapport entre les possibilités d'appartenance de x aux classes C_i et C_j avec $i \neq j$. Ce dernier x sera comparé à un seuil de tolérance ε . Si $Ra < \varepsilon$, la classification de x sera assurée par la méthode FPMSS, sinon la classification est effectuée par SVM.

Dans le cas d'apparition d'une nouvelle classe qui rentre en ambiguïté avec une autre classe déjà connue, un apprentissage hors ligne sera nécessaire pour la méthode SVM.

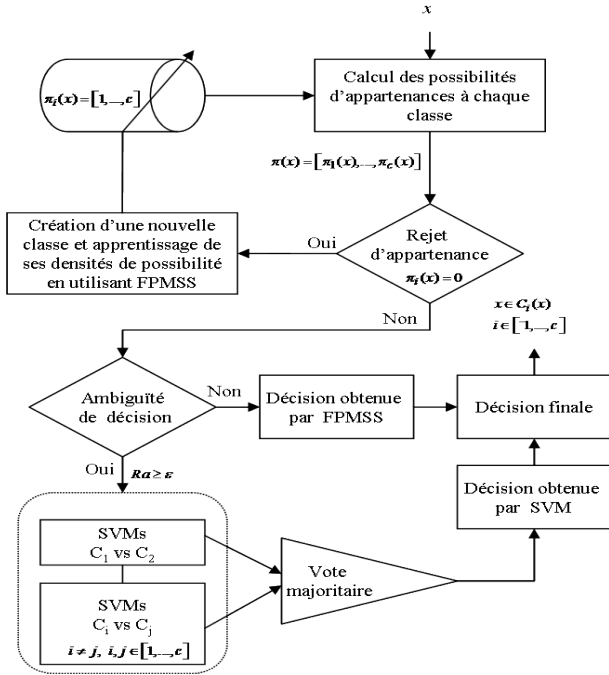


Fig. 4. Phase de classification

IV. RÉSULTATS

Nous avons testé l'approche proposée sur 7 jeux de données d'illustrations académiques et réels. Dans un premier lieu, l'approche proposée a été testée sur un ensemble de données synthétique dans R^2 . L'ensemble d'apprentissage X se compose dans un premier temps de deux classes.

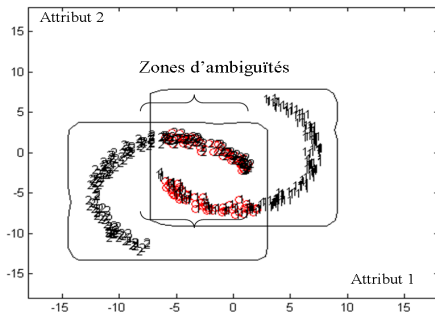


Fig. 5. Classification par FPMSS, localisation des régions des compétences

Deux étapes sont nécessaires pour la classification de cet ensemble de données. La première concerne la localisation de la zone d'ambiguïté (Figure 6) et la deuxième est liée à la classification des points de cette région (Figure 9).

Afin de montrer l'intérêt de la combinaison, nous avons appliqué l'ensemble d'apprentissage initial sur les deux méthodes. Les résultats sont donnés dans les Figures 5 et 6. La Figure 5 montre que les courbes des niveaux d'appartenances calculées par FPMSS ne respectent pas la forme non convexe des classes. Par contre, SVM a une frontière de décision adaptée à la séparation non linéaire de ces deux classes (Figure 6).

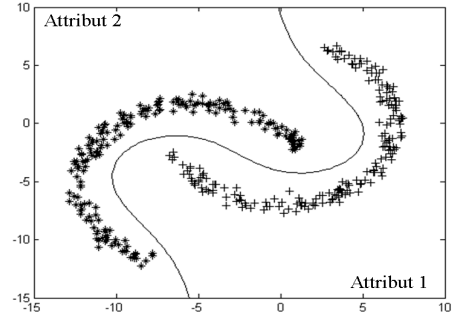


Fig. 6. Classification par SVM

En suite, nous avons appliqué ces données sur le système proposé de sélection des classifieurs. Les résultats de classification sont montrés dans la Figure 7. Nous remarquons que les courbes des niveaux d'appartenance respectent la forme non convexe des classes.

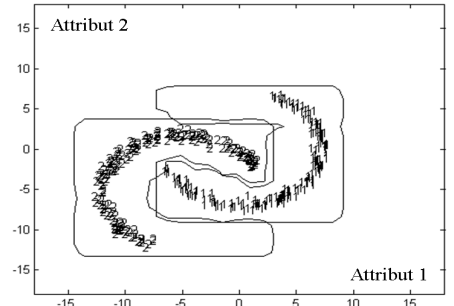


Fig. 7. Classification par le système de sélection dynamique des classifieurs

L'apparition d'une nouvelle classe (Figure 8) est une autre problématique qui a été testée sur le système proposé de sélection des classifieurs.

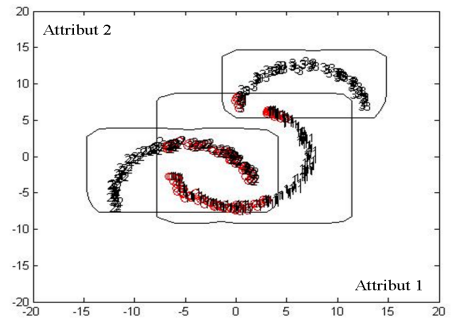


Fig. 8. Cas d'apparition d'une nouvelle classe

Ce test nous a permis de voir comment ce système réagit dans le cas d'apparition d'une nouvelle classe. Nous avons

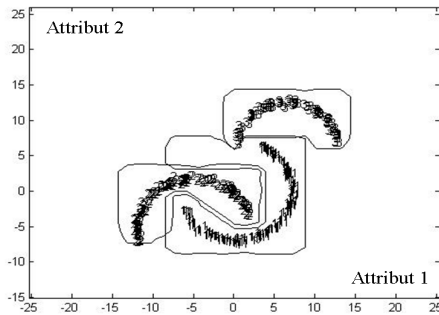


Fig. 9. Classification par l'approche proposée

suivi les mêmes étapes que dans l'exemple précédent. Les résultats obtenus (Figure 9) montrent que le système a bien classé les trois classes.

Enfin, nous avons appliqué notre système sur des bases de données réelles. Il s'agit de 5 jeux de données classiques en classification : Ionosphere, Pima, Iris, Breast Cancer Wisconsin et Pima Indian Diabetes (WBC) dont la description complète est donnée à <http://www.ics.uci.edu/ml/MLRepository.html>. Ainsi que la base des données TEP (Tennessee Estman Process)[8].

D'après le tableau suivant, nous remarquons, le système de sélection des classificateurs donne le meilleur taux de classification par rapport à FPMSS et SVM.

TABLE I
ETENDUE DES MESURES ET PRÉCISIONS

Données	Méthode	Erreur %
TEP	SVM	5,85
	FPMSS	9,46
	Approche proposée	4,83
Ionosphere	SVM	2,01
	FPMSS	7,38
	Approche proposée	1,34
Pima	SVM	27,99
	FPMSS	36,72
	Approche proposée	14,84
WBC	SVM	20,20
	FPMSS	4,83
	Approche proposée	3,22
Iris	SVM	4
	FPMSS	4,67
	Approche proposée	3

V. CONCLUSION

La combinaison de classificateurs est une voie de recherche prometteuse. C'est une technique efficace pour améliorer les performances de classification sans augmenter la complexité des méthodes utilisées. L'adaptation à la variété et la complexité d'un grand nombre de problèmes de classification et pour différents contextes d'application est un autre atout des méthodes combinaisons de classificateurs.

Nous avons présenté dans cet article un système de sélection des classificateurs qui combine la méthode Fuzzy Pattern Matching Semi Supervisée (FPMSS) et la méthode des

Séparateurs à Vaste Marge (SVM) pour améliorer les performances de classification dans les zones d'ambiguïté. La combinaison que nous avons proposée combine les avantages des deux méthodes FPMSS (incrémentale, adaptative, dynamique) et SVM (hyperplan optimal). Nous nous sommes intéressés aux données de formes complexes pour valider notre approche. L'autre avantage de notre approche réside dans l'utilisation de la méthode FPMSS pour la détection des nouvelles classes qui peuvent se présenter pendant la durée de vie du système.

RÉFÉRENCES

- [1] Al-Ani A., Deriche M. A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence. *Journal of Artificial Intelligence Research*, vol. 17, pp. 333-361, 2002.
- [2] Aksela M. Comparison of classifier selection methods for improving committee performance. *Lecture notes in computer science*, vol. 2709, pp. 84-93, 2003.
- [3] Bahler D., Navarro L. Methods for combining heterogeneous sets of classifiers. 17th national Conference on Artificial Intelligence, Texas, July 30-August 3, 2000.
- [4] Burges C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, vol. 2, n°2, pp. 121-167, 1998.
- [5] Cayrol M., Farreny H., Prade H. Fuzzy Pattern Matching. *Kybernetes*, vol. 11, n° 2, pp. 103-116, 1982.
- [6] Chen K., Wang L., Chi H. Methods of combining multiple classifiers with different features and their application to text independent speaker identification. *International Journal of Pattern Recognition and Artificial Intelligence*, pp 417-445, 1997.
- [7] Dos Santosa E.M., Sabourina R. A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. *Pattern Recognition*, Vol 41, pp. 2993-3009, 2008.
- [8] Downs J.J., Vogel E.F. Plant-wide industrial process control problem. *Computers and Chemical Engineering*, vol. 17, n° 3, pp. 245-255, 1993.
- [9] Dubois D., Prade H., Sandri S. On Possibility/Probability Transformations. *Fuzzy Logic*, pp. 103-112, 1993.
- [10] Gancarski P., Wemmert C. Collaborative multi-step mono-level multi-strategy classification. *Journal on Multimedia Tools and Applications*, vol. 35, n° 1, pp. 1-27, 2007.
- [11] Giacinto G., Roli F. Methods for dynamic classifier selection. 10th international Conference on Image Analysis and Processing, Venice, 27-29 September, 1999.
- [12] Giani A. Parallel cooperative classifier systems. Thèse de doctorat de l'université de Pise, Italie, 1999.
- [13] Grabisch M., Bienvenu G., Grandin J.F., Lemer A., Moruzzi M. A Formal Comparison of Probabilistic and Possibilistic. 7th IFSA World Congress, Prague 117-122, 1997.
- [14] Gunes V., Menard M., Loonis P., and Petit-renaud S. Combination, Cooperation and Selection of Classifiers : A State of the Art. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, n° 8, pp. 1303-1324, 2003.
- [15] <http://archive.ics.uci.edu/ml/>
- [16] Kittler J., Hatef M., Duin R.P.W., Matas J. On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, n°3, pp. 226-239, 1998.
- [17] Last M., Bunke H., Kandel A. A feature-based serial approach to classifier combination. *Pattern analysis and applications*, vol. 5, n° 4, pp. 385-398, 2002.
- [18] Lu Y., Yamaoka F. Fuzzy integration of classification results. *Pattern recognition*, vol. 30, n° 11, pp. 1877-1891, 1997.
- [19] Mandler E., Schuermann J. Combining the classification results of independent classifiers based on the demster theory of evidences. *International Journal of pattern Recognition and Artificial Intelligence*, pp 381-393, 1988.
- [20] Ruta D., Gabrys B. An overview of classifier fusion methods. *Computing and Information Systems*, pp. 1-10, 2000.
- [21] Sayed Mouchaweh M. Semi-supervised classification method for dynamic applications. *Fuzzy Sets and Systems*, vol. 161, n° 4, pp. 544-563, 2010.
- [22] Vapnik N. An Overview of statistical Learning Theory. *IEEE transactions on neural networks*, vol. 10, n° 5, pp. 988-999, 1999.
- [23] Zhu X., Wu X., Yang Y. Dynamic selection for effective mining from noisy data streams. 4th IEEE International Conference on Data Mining, Brighton, November 01-04, 2004.