HAL
open science

Coalescence floue avec ambiguïté pour la fusion multi-classifieurs : coopération classification-classement.

Veyis Gunes, Michel Menard, Pierre Loonis

To cite this version:

Veyis Gunes, Michel Menard, Pierre Loonis. Coalescence floue avec ambiguïté pour la fusion multi-classifieurs : coopération classification-classement.. 12e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle., Feb 2000, Paris, France. hal-02092853

HAL Id: hal-02092853

https://hal.science/hal-02092853

Submitted on 8 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coalescence floue avec ambiguïté pour la fusion multi-classifieurs : coopération classification-classement

Fuzzy clustering with ambiguity for multi-classifiers fusion : clustering-classification cooperation

V. Gunes M. Ménard P. Loonis

Laboratoire d'Informatique et d'Imagerie Industrielle

Université de La Rochelle
Avenue Marillac
17 042 La Rochelle Cedex 1
E-mail : vgunes@univ-lr.fr

Résumé

*Le principal objectif de cet article est de montrer les performances de la fusion multi-classifieurs fondée sur une coalescence (clustering) floue avec ambiguïté. Le problème est considéré d'un point de vue multi-décisionnel (i.e., plusieurs modules de classement). Chaque module de classement est spécialisé sur une région particulière de l'espace de représentation. Ces régions sont obtenues par coalescence floue et leur union constitue l'ensemble des données originelles. L'objectif de la coalescence est de réduire l'ensemble des données en regroupant les individus similaires en régions ou en classes de coalescence. L'algorithme de coalescence choisi permet d'introduire la notion d'*ambiguïté *qui concerne les individus proches des frontières des classes de coalescence. Ensuite, un classifieur est associé à chaque classe. L'ensemble d'apprentissage de chaque classifieur est alors défini sur une classe de coalescence et les classes de rejet d'ambiguïté qui lui sont associées. Le système global est parallèle puisque chaque classifieur travaille avec son propre ensemble d'apprentis. L'algorithme permet de faire une sélection adaptée de classifieurs dans le sens où la coalescence floue avec ambiguïté détermine les régions d'apprentissage de l'espace de représentation. La prise de décision est le résultat de la fusion des sorties des classifieurs voisins.*

Mots Clef

Coalescence floue, ambiguïté, fusion, reconnaissance de formes.

Abstract

The main aim of this paper is to demonstrate the performance of multi-classifiers fusion based on fuzzy clustering with ambiguity. The problem is seen from the multi-decision point of view (i.e. several classification modules). Each classification module is specialized on a particular region of the features space. These regions are obtained by fuzzy clustering and constitute the original data set by union. The goal is to reduce the amount of data by merging similar patterns together into regions or clusters. The chosen clustering algorithm allows to introduce the ambiguity *notion, which deals with patterns lying near the clusters boundaries. A classifier is associated with each cluster. The training set for each classifier is then defined on the cluster and its associated ambiguous clusters. The overall system is parallel, since different classifiers work with their own training data clusters. The algorithm makes it possible the adaptive classifier selection in the sense that the fuzzy clustering with ambiguity gives adapted training data regions of the features space. The decision making is the fusion of outputs from neighbour classifiers.*

Keywords

Fuzzy clustering, ambiguity, fusion, pattern recognition.

1 Introduction

Selon l'approche reconnaissance des formes, les p paramètres observés dans une forme sont utilisés pour construire le vecteur de forme. La reconnaissance est liée aux classes d'individus à discriminer dans l'espace de représentation à p-dimensions. La plupart des problèmes réels sont considérés complexes : un seul classifieur, gérant toutes les classes et paramètres, est souvent mal adapté ; l'ensemble d'apprentissage est alors trop grand pour un apprentissage correct (i.e. distribution de probabilités *a priori* sous jacente multi-modales). Pour la combinaison de classifieurs multiples, un algorithme prend généralement en compte une des deux approches suivantes :

- la fusion des classifieurs. Dans ce cas, chaque classifieur est appliqué en parallèle et leurs sorties sont combinées pour atteindre un consensus ; ces approches ont pour cadre la théorie du vote [13] [21], le consensus unanime [21] [10], les méthodes de vote utilisant des règles de décision heuristiques [12], la théorie Bayesienne [21], les théories de l'incertain [19] [21], la théorie possibiliste [6], et le classement hiérarchique [11]. Ces méthodes fournissent des techniques sub-divisant l'espace de représentation en un nombre convenable de sous-espaces, permettant la coopération de différentes techniques de classement. Même si le comportement global est amélioré grâce à la détermination des frontières de méta-décision, la spécification de chaque module sur un sous-espace de paramètres décroit, généralement, le taux de reconnaissance de chaque classifieur [14].
- la sélection adaptée de classifieurs. Chaque module de classement est spécialisé sur une région spécifique de l'espace de représentation. Ces méthodes tentent de prédire le classifieur qui semble le plus adapté à l'échantillon à classer. Ainsi, seule la sortie du classifieur sélectionné est prise en compte dans la phase de décision. Plusieurs articles présentent des méthodes de sélection d'un classifieur utilisant une estimation des performances locales dans de petites régions de l'espace de représentation entourant un échantillon de test inconnu [20]. Cependant, c'est uniquement la sortie du meilleur classifieur pour cette donnée qui est utilisée pour la prise de décision finale. Ces algorithmes sont souvent basés sur des méthodes d'optimisation, comme par exemple, les algorithmes génétiques et les réseaux de neurones [18].

Cet article porte sur la combinaison de ces deux approches en utilisant des méthodes déterminant une partition de l'espace des caractéristiques en un nombre convenable de régions, permettant ainsi la coopération de plusieurs modules de classement. Chacun de ces modules est spécialisé sur une région particulière de l'espace de représentation. Ces régions sont obtenues par coalescence floue et leur union constitue l'ensemble des données originelles. L'objectif de la coalescence est de réduire l'ensemble des données en regroupant les individus similaires en régions. De plus, la notion d'ambiguïté qui concerne les individus proches des frontières de classes de coalescence est introduite dans l'algorithme de coalescence. Le système global est parallèle puisque les différents classifieurs travaillent chacun avec leur propre ensemble d'apprentis.

De plus, dans le but d'éviter une gestion complexe de conflit entre classifieurs, la phase de fusion doit être réalisée avec seulement un sous-ensemble de classifieurs. L'algorithme permet de faire une sélection adaptée de classifieurs dans le sens où la coalescence floue avec ambiguïté détermine les régions d'apprentissage de l'espace de représentation. La prise de décision est le résultat de la fusion des sorties des classifieurs voisins (i.e., dont les régions définies par les ensembles d'apprentissage de ces classifieurs incluent l'individu à classer).

Dans ce qui suit, nous appellons *Classe de Coalescence* CC, un ensemble d'individus obtenu par une méthode de coalescence ou de catégorisation, et *Classe de Coalescence de rejet d'Ambiguïté* CCA, un ensemble d'individus ambigus obtenu par une méthode de coalescence avec rejet d'ambiguïté. Nous appellons *classe*, un ensemble d'individus appartenant, selon l'expert ou selon un classifieur, à une même classe d'interprétation.

Cet article est organisé de la manière suivante. Le principe de la méthode est décrit en section 2 sur un exemple de 3 classifieurs. Dans la section 3, la détermination des CCs et CCAs par la méthode des (c+2) moyennes floues (Fc+2M) est présentée. Le but de la coalescence est de déterminer l'ensemble d'apprentissage pour chaque classifieur. L'intérêt de l'ambiguïté est d'accroître la redondance entre classifieurs et rendre la fusion possible. La section 4 contient la description de la fusion multi-classifieurs adoptée. Cette fusion est parallèle et adaptée. Lors du classement, un individu inconnu est assigné à une CC ou à une CCA. S'il s'agit d'une CCA, les sorties des classifieurs voisins (classifieurs ayant appris sur cette CCA) sont fusionnées en utilisant la règle de combinaison de Dempster. Enfin, une application est présentée dans la section 5, elle est suivie d'une conclusion.

2 Principe de la méthode

La méthode est définie de la manière suivante :
- Soit $\Phi = (\phi_i)_{i\in[1,l]}$ une famille de classes ; La reconnaissance est liée aux classes d'individus à discriminer dans l'espace de représentation à p-dimensions ;
- Soit $X = (x_k)_{k\in[1,n]}$ l'ensemble des individus où

$$x_k = (x_{k1}, x_{k2}, \ldots, x_{kp})^t$$

  est un individu décrit par p caractéristiques (i.e. $x_k \in R^p$). Soit $X^{lr} \subset X$, un ensemble d'individus d'apprentissage étiquetés dans Φ.

Un algorithme de coalescence génère une nouvelle partition de X^{lr}, composée de CCs et de CCAs. Cette partition (cf. Figure 1) est obtenue en utilisant l'algorithme des Fc+2M [15] (c-moyennes floues avec rejet d'ambiguïté, cf. section 3). Un classifieur est alors associé à chaque CC. L'ensemble d'apprentissage de chaque classifieur est alors défini sur sa CC et ses CCAs associés (cf. Figure 2). Chaque classifieur apprend séparément. La prise de décision, pour un individu inconnu, est le résultat de la fusion des sorties des classifieurs *voisins*.

3 Détermination des CCs et CCAs : coalescence avec rejet d'ambiguïté

Pour réduire un taux d'erreur excessif dû au bruit et aux autres facteurs d'incertitudes inhérents aux mesures réelles, la coalescence avec rejet d'ambiguïté apporte une solution intéressante. Dans la littérature, les règles permettant de rejeter un individu situé sur la frontière de

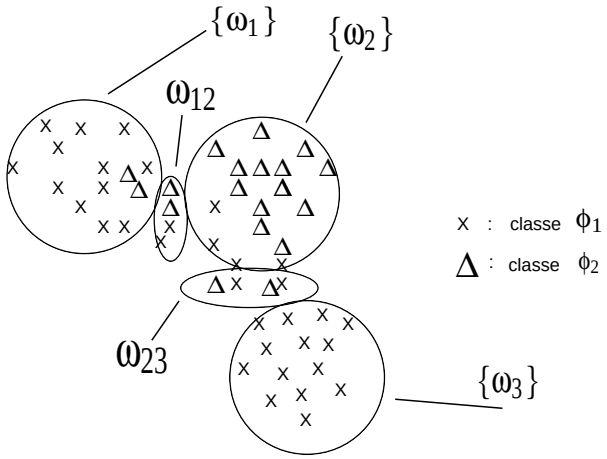

FIG. 1 – La partition des ensembles d'individus d'apprentissage, $X^{lr} \subset X$, lorsque le rejet d'ambiguïté est exercé. $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$: CCs ; $\omega_{12} = \{\omega_1, \omega_2\}$, $\omega_{23} = \{\omega_2, \omega_3\}$: CCAs.

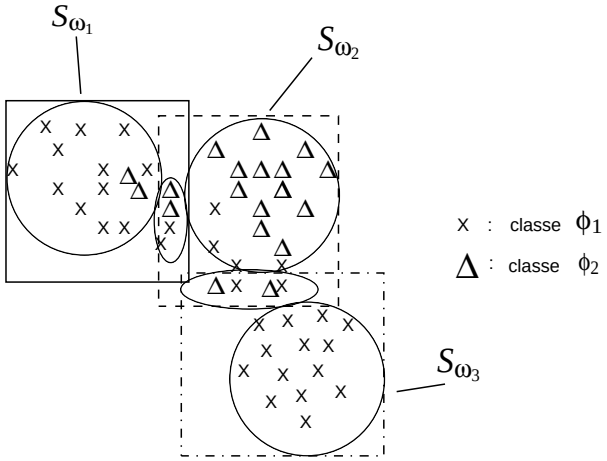

FIG. 2 – Un classifieur est associé à chaque CC. L'ensemble d'apprentissage, S_{ω_i} du classifieur e_i, est défini sur ω_i et ses CCAs associés.

deux CC sont fondées sur des valeurs de seuils dans la phase de décision et non dans la phase de coalescence ou d'apprentissage. Pour spécifier cette prise de décision, il est courant de caractériser un individu x_k avec un concept d'ambiguïté.

Le rejet d'ambiguïté a été introduit dans la phase de décision par Chow ([4], [5]). Le but est de mesurer la proximité de x_k aux frontières de décision. Dans le cas de la règle de Bayes, le problème est de définir une nouvelle classe de coalescence ω_0, appelée *classe de rejet*, associée à un coût de rejet constant Cr :

$$\begin{cases} x_k \rightarrow \omega_i \\ \text{ si } P(\omega_i/x_k) = \max_{j\in[1,c]}(P(\omega_j/x_k)) > 1 - Cr \\ x_k \rightarrow \omega_0 \text{ sinon} \end{cases}$$

où $(P(\omega_j/x_k))_{j\in[1,c]}$ sont les probabilités *a posteriori*. Le terme C_r, qui contrîle le taux de rejet, doit être choisi entre 0 et $\frac{c-1}{c}$.

Cette règle partage l'espace de représentation en $(c+1)$ classes de coalescence. Dans [7], Dubuisson affirme que ce rejet est un rejet d'ambiguïté puisque la région correspondante dans l'espace de représentation est toujours située entre les CCs. Le principal inconvénient de ces définitions et de leur utilisation est que ce rejet consiste à fixer un seuil *a priori*, sans tenir compte de la "forme" des classes d'apprentissage. Une extension de ce rejet simple est le rejet sélectif introduit par Ha [9].

Dans cette section, nous construisons une fonction objective appropriée dans le but de prendre en compte explicitement le rejet d'ambiguïté dans la coalescence floue.

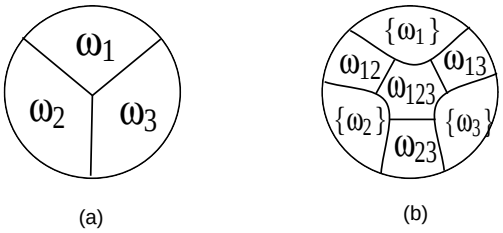

FIG. 3 – La figure (a) illustre la partition de l'espace de représentation en trois CCs. La figure (b) illustre la partition de l'espace de représentation lorsque le concept d'ambiguâté est appliqué. Nous associons une CCA à chacun des sous-ensembles $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i\in[1,3]}\}$.

Soit $\Omega = (\omega_i)_{i\in[1,c]}$, une famille de CCs. Afin de tenir compte de l'ambiguïté entre CCs, des fonctions d'appartenances sont affectées à tous les éléments de 2^Ω, au lieu des éléments de Ω seulement. Puisqu'il y a 2^c sous-ensembles dans un ensemble de c éléments, nous cherchons à partitionner l'espace des représentations en $2^c - 1$ classes simples et composées, excluant l'ensemble vide, dans un problème à c classes. Il y a donc c CCs, $(\{\omega_i\})_{i\in[1,c]}$, et pour déterminer le rejet d'ambiguïté, nous associons une CCA à chaque sous-ensemble $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i\in[1,c]}\}$. Dans la Figure 3, il y a $(2^3 - 1)$ CCs et CCAs correspondants aux sous-ensembles :

$\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$,
$\omega_{12} = \{\omega_1, \omega_2\}$,
$\omega_{13} = \{\omega_1, \omega_3\}$,
$\omega_{23} = \{\omega_2, \omega_3\}$ et
$\omega_{123} = \{\omega_1, \omega_2, \omega_3\}$.

Definition 1 *Soit $\Omega = \{\omega_1, ..., \omega_c\}$ l'ensemble des classes de coalescence ω_i. Nous associons une CCA à chacun des sous-ensembles de classes $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i\in[1,c]}\}$. Nous considérons qu'il existe un ensemble de $\sum_{k>1}\binom{c}{k}$ CCAs disjointes dans un problème à c classes.*

Une autre manière de caractériser les CCAs est d'exprimer les propriétés que doit vérifier la distance d_{Ak} (cette distance est définie plus loin, cf. partie 3.2) d'un individu x_k à une classe de coalescence de rejet d'ambiguïté A : un individu est considéré comme ambiguß s'il appartient à la frontière de deux ou plus CCs. La Figure 4 illustre les propriétés de cette distance dans le cas de trois CCs. Les lignes en pointillées représentent le lieu d'ambiguïté totale vérifiant :

$$min_{A,|A|>1} d_{Ak} \leq d_{ik} \ \forall i \in \{1,..,c\}$$

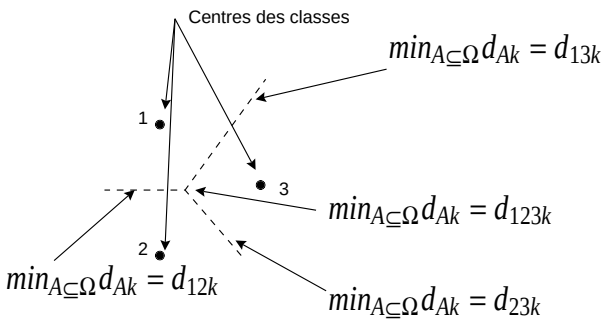

FIG. 4 – Lieux où la distance aux classes d'ambiguïté est minimale.

3.1 Le concept d'ambiguåté : frontières et régularisations dans les problèmes de coalescence

Cette section concerne la recherche de frontières de plusieurs CC dans un objectif de régularisation en classification. Nous faisons la distinction entre deux types de frontière : le premier est net et est défini dans le cadre de la théorie des ensembles approximatifs, le second est obtenu en généralisant ce concept dans le cadre de la théorie des sous-ensembles flous. Le premier utilise les relations d'indiscernabilité, le second, les fonctions à valeurs réelles. A partir de ces définitions, nous étendons le concept de frontière au concept de frontières floues de CCs d'une partition floue. Nous utilisons la théorie des ensembles approximatifs de Pawlak [17] pour préciser le concept d'ambiguåté et définir le critère de coalescence à minimiser. Cette théorie modèlise l'ambiguïté des connaissances, attribuée à la granularité des connaissances, i.e., à l'indiscernabilité et aux approximations.

Un système d'information est une paire $\mathcal{A} = (U, A)$, où U est un ensemble fini, non vide appelé l'univers et A, un ensemble fini, non vide d'attributs, i.e. $a \ : U \to V_a \ \forall a \in A$, où V_a est l'ensemble des valeurs de a. Par V nous notons l'ensemble $\cup\{V_a \ : a \in A\}$. A chaque sous-ensemble d'attributs $B \subseteq A$, une relation appelée B-indiscernable notée $IND_{\mathcal{A}}B$ est associée et est définie par :

$$IND_{\mathcal{A}}B = \{(u, u') \in U^2 \ : a(u) = a(u') \ \forall a \in B\}$$

x_k	B	décision : ω
1	oui	$\{\omega_1\}$
2	non	$\{\omega_2\}$
3	non	$\{\omega_3\}$
4	oui	ω_{12}
5	oui	ω_{13}
6	non	ω_{23}
7	oui	ω_{123}

TAB. 1 – Exemple de Système d'Information pour le concept "appartenant à ω_1".

Les objets u et u' satisfaisant la relation $IND(B)$ sont indiscernables selon B car ils ont les mêmes valeurs relativement aux attributs de B. On appelle ensemble élémentaire (par rapport à B) un ensemble $[u]_B$ d'objets qui sont tous indiscernables par rapport à B.
Si $\mathcal{A} = (U, A)$ est un système d'information, $B \subseteq A$ est un ensemble d'attributs et $X \subseteq U$ est un ensemble d'objets, alors les ensembles :

$$\{u \in U : [u]_B \subseteq X\}$$

et

$$\{u \in U : [u]_B \cap X \neq \emptyset\}$$

sont appelés respectivement la plus basse et la plus haute approximation de X dans $\mathcal{A}$, et sont notées $\underline{B}X$ et $\overline{B}X$, respectivement. L'ensemble $BN_B(X) = \overline{B}X - \underline{B}X$ est appelé frontière de X. L'ensemble $\underline{B}X$ est l'ensemble de tous les éléments de U qui peuvent être classés de façon certaine comme élément de X, avec la connaissance représentée par les attributs de B, tandis que l'approximation haute de X, $\overline{B}X$, est l'ensemble de tous les éléments pouvant appartenir à X. L'ensemble $BN_B(X)$ est constitué des éléments que l'on ne peut classer ni dans X, ni dans $-X$, avec la connaissance de B.

Les informations sur les individus sont souvent disponibles sous forme de tables. Supposons que l'on ait les individus x_k, avec $k \in [1,..7]$, comme indiqués dans la Table 1. Soit $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Chaque individu x_k sera d'abord supposé posséder l'étiquette d'une classe de coalescence $I \subseteq \{1, 2, 3\}$ indiquant avec certitude son appartenance à un sous-ensemble de 2^Ω. Ainsi, il y a 4 classes correspondants aux CCAs $\omega_{12}, \omega_{13}, \omega_{23}, \omega_{123}$ et 3 CCs $\{\omega_1\}$, $\{\omega_2\}$ et $\{\omega_3\}$. L'attribut *décision* se trouve dans la troisième colonne. Pour $x_k \in \omega_I$, l'attribut B est "oui" si $1 \in I$ et "non" sinon. Dans la Table 1 les individus x_1, x_4, x_5 *et* x_7 sont indicernables selon l'attribut B. L'attribut B génère deux ensembles élémentaires $\{x_1, x_4, x_5, x_7\}$ et $\{x_2, x_3, x_6\}$. x_1 possède les informations qui nous permettent de le classer, avec certitude comme appartenant à la CC ω_1. Ainsi, la plus basse approximation de l'ensemble des individus appartenant à la classe de coalescence ω_1 est l'ensemble $\{x_1\} = \{x_k \in \{\omega_1\}\}$ et la

plus haute approximation de cet ensemble est l'ensemble :

$$\{x_1, x_4, x_5, x_7\} = \{x_k \in \omega_I \text{ où } \omega_I \in 2^\Omega \setminus \emptyset \text{ avec } 1 \in I\}$$

La frontière du concept "appartenant à ω_1" est l'ensemble :

$$\{x_4, x_5, x_7\} = \{x_k \in \omega_I \text{ où } \omega_I \in 2^\Omega \setminus \{\emptyset, \{\omega_1\}\} \text{ avec } 1 \in I\}$$

Cet exemple donne une définition simple de la frontière d'une classe.

Definition 2 *Soit $\Omega = (\omega_i)_{i\in[1,c]}$, un ensemble de CCs. Soit $X = (x_k)_{k\in[1,n]}$, l'ensemble des individus où $x_k = (x_{k1}, x_{k2}, \ldots, x_{kp})^t$ est un individu décrit par p caractéristiques (i.e. $x_k \in R^p$). Nous définissons la frontière de la classe de coalescence ω_i avec $i \in [1, c]$ comme : $F_{\omega_i} = \{x_k \in A$ où $A \in 2^\Omega \setminus \{\emptyset, \{\omega_i\}\}$ avec $\omega_i \in A\}$*

Dans une partition floue, chaque individu x_k peut appartenir avec un certain degré aux frontières de chaque CC. Aussi, en nous appuyant sur la définition précédente, nous proposons une mesure de la taille de la frontière floue de la classe de coalescence ω_i :

$$\sum_{k=1}^{n} \sum_{A\in 2^\Omega\setminus\{\emptyset,\{\omega_i\}\}\ ,\ \omega_i\in A} \mu_{Ak}^m d_{Ak}^2 =$$

$$\sum_{k=1}^{n} \sum_{A\in 2^\Omega\setminus\{\emptyset,(\{\omega_j\})_{j\in[1,c]}\}\ ,\ \omega_i\in A} \mu_{Ak}^m d_{Ak}^2$$

où d_{Ak} est la distance de l'individu x_k à la *CCA* A : $d_{Ak} = d(x_k, v_A)$, m est un facteur de flou et μ_{Ak} est le degré d'appartenance de l'individu x_k à la *CCA A*. Le choix de d_{Ak} sera discuté plus loin. Pour des grandes valeurs de d_{Ak}, si μ_{Ak} prend des valeurs importantes alors la taille de la frontière floue de la i^{eme} classe doit être importante. Du fait qu'une frontière est commune à plusieurs CC, la taille totale des frontières floues, dans une partition floue, est définie par :

$$J_m^a(U, v) = \sum_{k=1}^{n} \sum_{A\in 2^\Omega\setminus\{(\{\omega_j\})_{j\in[1,c]},\emptyset\}} \mu_{Ak}{}^m d_{Ak}^2$$

Du point de vue de la régularisation, le problème de la coalescence est dit "mal posé" s'il ne satisfait pas un ou plusieurs des critères suivants : la solution existe, est unique et dépend continuement des données. Des hypothèses à priori additionnelles doivent être imposées sur la solution pour convertir un problème mal posé en un problème bien posé. Lorsque l'on utilise la théorie des champs de Markov en traitement d'images, l'hypothèse de lissage est incorporée dans la fonction d'énergie par laquelle le coût de la solution est défini. Par exemple, Pavlidis propose dans [16] une fonction d'énergie dans laquelle est prise en compte, la frontière (i.e. $\sum_{x\in O_i} L_i(x)$) des régions $(O_i)_{i\in[1,c]}$:

$$E_\lambda = \sum_{i=1}^{c} (\sum_{x\in O_i} (1 - L_i(x))(u_i - g(x))^2 + \lambda L_i(x))$$

où $g(x)$ est le niveau de gris de x, c est le nombre de régions et u_i est l'étiquette associée à la région O_i. Chaque discontinuité est pénalisée par le coût λ. Le concept de régularisation peut être étendu au problème de la coalescence floue. Aussi, nous définissons le critère de coalescence comme :

$$J_m^1(U, v)) = J_m^0(U, v) + J_m^a(U, v) =$$

$$\sum_{k=1}^{n}\sum_{i=1}^{c} \mu_{ij}^m ||x_k - v_i||^2 + \sum_{k=1}^{n} \sum_{A\in 2^\Omega\setminus\{(\{\omega_j\})_{j\in[1,c]},\emptyset\}} \mu_{Ak}{}^m d_{Ak}^2$$

$(u_i, g(x))$ dans l'expression de E_λ correspondent à (v_i, x_k) dans $J_m^0(U, v)$ avec $\mu_{ik} = 1$ ou 0 (classification nette). $J_m^0(U, v)$ représente la fonction objective utilisée dans l'algorithme des c-moyennes floues FcM [2]. Les discontinuités $\sum_{i=1}^{c} \sum_{x\in O_i} \lambda L_i \ \ \forall O_i$ correspondent à :

$$J_m^a(U, v) = \sum_{k=1}^{n} \sum_{A\in 2^\Omega\setminus\{(\{\omega_j\})_{j\in[1,c]},\emptyset\}} \mu_{Ak}{}^m d_{Ak}^2$$

.

3.2 Discontinuités

Dans le problème de la coalescence floue, les régions ambiguës, situées entre les CCs, peuvent être considérées comme des dicontinuités des fonctions d'appartenance définies sur les CCs. Pour détecter ces discontinuités, nous définissons la distance d'un individu x_k à la *CCA* $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_j\})_{j\in[1,c]}\}$ par :

$$d_{Ak} = \frac{\alpha(\sum_{\omega_i\in A} d_{ik})^2}{|A|^2(\prod_{\omega_i\in\Omega} d_{ik})^{\frac{1}{c}}}$$

où, le terme α est un nombre positif ajusté de manière à détecter les discontinuités. d_{ik} est la distance classique utilisée dans l'algorithme des FcM [2]. d_{Ak} joue le rîle de l'amplitude du gradient utilisée dans les champs de Markov, pour détecter les discontinuités. Par analogie avec les modèles des membranes [3], le terme α^{-1} représente le facteur de discontinuité et joue sur l'interdépendance des fonctions d'appartenance. Nous avons développé un algorithme c-moyennes floues étendu [15], où les notions d'ambiguïtés précédemment décrites sont introduites.

Les Figures 5 et 6 montrent la forme de la fonction d'appartenance maximale des trois CCs (respectivement, des quatre CCAs) dans le plan factoriel principal pour les données IRIS [8] qui ont été souvent utilisées pour évaluer les performances des algorithmes de classifications. Une des trois classes est clairement séparée des deux autres, tandis que les deux dernières admettent un chevauchement important entre-elles. Cet exemple est fourni dans le but d'illustrer la partition de l'espace de représentation en 3 CCs et 4 CCAs.

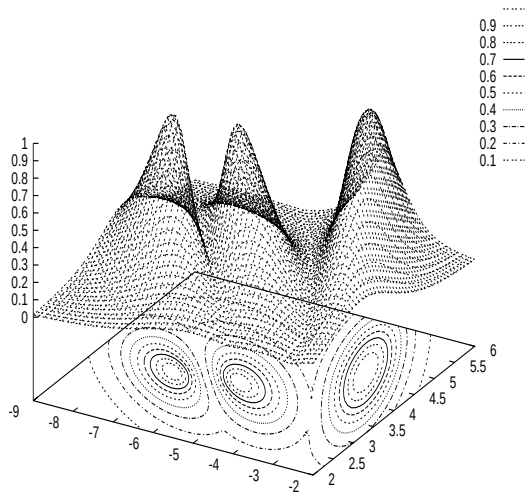

FIG. 5 – La fonction d'appartenance maximale dans le plan factoriel principal des données IRIS, parmis les trois CCs, avec $m = 1.2$ et $\alpha = 0.8$. Norme de Mahalanobis.

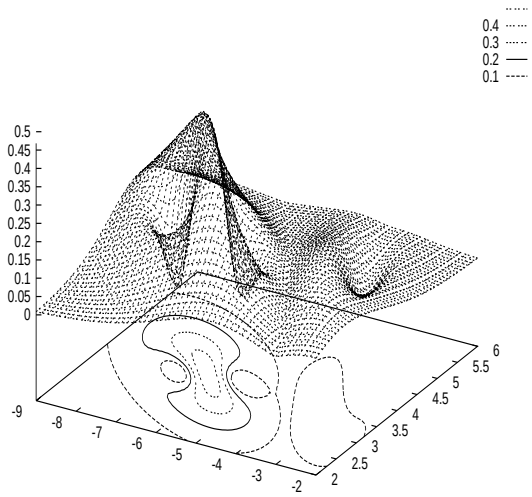

FIG. 6 – La fonction d'appartenance maximale dans le plan factoriel principal des données IRIS, parmis les quatre CCAs, avec $m = 1.2$ et $\alpha = 0.8$. Norme de Mahalanobis.

4 Une fusion multi-classifieurs parallèle et adaptée

Le système global comporte quatre composants principaux : un ensemble de c classifieurs $(e_k)_{k\in[1,c]}$ qui déterminent la possibilité qu'un individu de test particulier appartienne ou non à la classe ϕ_i, $i \in [1, l]$, un module de fusion $\mathcal{F}$ qui combine les sorties des sous-ensembles de classifieurs, un module de décision $\mathcal{D}$ qui affecte l'individu de test à l'une des l classes et un module de sélection qui choisit, pour un individu donné, le sous-ensemble de classifieurs à partir de l'ensemble des classifieurs.

4.1 La procédure de coalescence : Module de sélection

Soit X^{lr}, un ensemble d'individus d'apprentissage étiquetés selon les classes connues. L'algorithme des Fc+2M génére une nouvelle partition $\mathcal{P}$ de X^{lr}, composée de CCs et de CCAs. Le nombre de CCs, c, est généralement choisi plus grand que le nombre de classes, l. L'utilisation de la coalescence sur des données complexes et bruitées permet de spécifier les ensembles d'apprentissage et le voisinage des classifieurs. Cette spécification autorise la coopération entre classifieurs, permettant ainsi la gestion de la décision à points de vue multiples. L'idée principale est que la procédure de coalescence permet de gérer la sélection des classifieurs les plus adaptés.

4.2 Les classifieurs

Le classifieur Bayesien est retenu pour cette étude. La règle de Bayes est une approche statistique bien connue. L'hypothèse principale est la connaissance des lois de distribution dont les caractéristiques sont déterminées d'après l'ensemble d'apprentissage, X^{lr}. Un classifieur est associé à chaque CC. L'ensemble d'apprentissage, S_{ω_i} pour le classifieur e_i, est défini sur ω_i et ses CCAs associés : $S_{\omega_i} = \{x_k \in A \text{ où } A \in 2^\Omega \text{ avec } \omega_i \in A\}$. Les lois de distribution de probabilités *a priori* , pour le classifieur e_i, sont notées $P_i(x_k/\phi_j)_{k\in[1,l]}$. L'approche Bayesienne conduit typiquement à évaluer la probabilité *a posteriori* $P_i(\phi_j/x_k)$ pour chaque classe ϕ_j à l'aide du théorème de Bayes.

Definition 3 *Soit 2^Ω, l'ensemble des classes CCs et CCAs. Nous associons un classifieur, e_i, à chaque classe de coalescence ω_i. L'ensemble d'apprentissage, S_{ω_i}, pour le classifieur e_i, est défini sur $S_{\omega_i} = \{x_k \in A \text{ où } A \in 2^\Omega \text{ avec } \omega_i \in A\}$. Les classes de décision, pour le classifieur e_i, sont définies par $\{\phi_j \in \Phi \ : \phi_j \ni x_k \ \forall x_k \in S_{\omega_i}\}$.*

Definition 4 *Soit $(e_k)_{k\in[1,c]}$, l'ensemble des classifieurs. L'ensemble E_i des classifieurs voisins, du classifieur e_i, est défini par : $E_i = \{e_k \ : S_{\omega_i} \cap S_{\omega_k} \neq \emptyset\}$*

4.3 Module de fusion

Soit $X^t \subset X$ $(avec X^t \cap X^{lr} = \emptyset)$, l'ensemble des individus inconnus. Lors du classement, un individu inconnu $x_k \in X^t$ est affecté à la partition $\mathcal{P}$. Le classement se fait de la façon suivante :

- si x_k est situé dans la CC , ω_i, la probabilité *a posteriori* $P_i(\phi_j/x_k)$ donnée par le classifieur e_i, associé à cette classe, est utilisée pour prendre la décision finale ;
- si x_k est situé dans la CCA, $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_j\})_{j\in[1,c]}\}$, la prise de décision est donnée par le résultat de la fusion de l'ensemble des sorties des classifieurs voisins (i.e., classifieurs ayant appris sur la CCA à laquelle appartient

x_k, cf. Figure 7) tels que :

$$\{e_j \ : \omega_j \in A\}$$

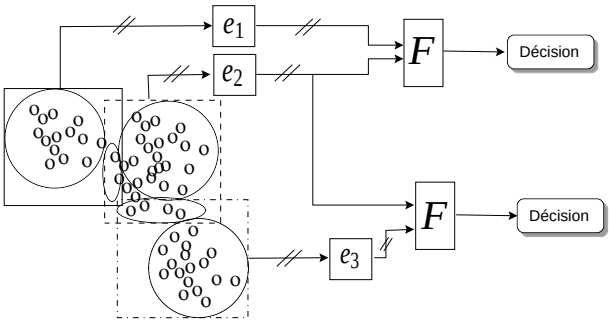

FIG. 7 – La prise de décision est le résultat de la fusion des sorties des classifieurs voisins.

Plusieurs méthodes de combinaison des entités d'informations issues de plusieurs sources, basées sur des facteurs de certitude comme dans l'algorithme MYCIN, ou sur des mesures de crédibilité et de plausibilité comme dans la théorie de l'évidence ont été proposées dans la littérature. Le module de fusion que nous proposons est fondé sur une combinaison des sorties de classifieurs par la règle de Dempster. Nous pouvons combiner les probabilités *a posteriori* par cette règle de combinaison, puisque nous émettons l'hypothèse que les classes sont équiprobables *a priori* (cf. section 5). Cependant, dans le cas de classes non équiprobables, il est nécessaire d'utiliser d'autres modèles de fusion, par exemple ceux proposés par Appriou [1].

5 Résultats

La méthode ci-dessus, que nous nommerons 3C (Coopération Classification-Classement), est testée sur un ensemble de données synthétiques dans R^2. L'ensemble d'apprentissage X est composé de 210 individus et est représenté sur la Figure 8. L'écart-type est $\sigma = 2$.

- Φ_1 : classe composée de 100 individus provenant de la distribution Gaussienne :

$$m_1 = \begin{pmatrix} 10 & 20 \end{pmatrix}^t, \Sigma_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

- Φ_2 : classe multi-modes composée de 110 individus générés de la manière suivante :
  - Deux populations Gaussiennes générées par :

$$\Sigma_2 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

$$m_{2a} = \begin{pmatrix} 10 & 10 \end{pmatrix}^t \ 50 \ individus$$

pour la première population,

$$m_{2b} = \begin{pmatrix} 10 & 30 \end{pmatrix}^t \ 50 \ individus$$

pour la seconde population

$\nearrow$	ω_1	ω_2	ω_{12}
Φ_1	48.0	15.0	37.0
Φ_2	50.91	49.09	0

TAB. 2 – Partition nette de X^{lr} obtenue par l'algorithme Fc+2M (pourcentages) pour $c = 2$: Φ_1 et Φ_2 sont les classes réelles ; ω_1 et ω_2 sont les CCs ; ω_{12} est la CCA.

$\nearrow$	ω_1	ω_2	ω_3	ω_{12}	ω_{13}	ω_{23}	ω_{123}
Φ_1	67.0	0	0	17.0	16.0	0	0
Φ_2	3.6	36.4	45.5	11.8	2.7	0	0

TAB. 3 – Partition nette de X^{lr} obtenue par l'algorithme Fc+2M pour $c = 3$: Il y a trois CCs et 4 CCAs possibles (2 d'entre-elles, ω_{12} et ω_{13}, existent réellement).

  - et 10 individus reliant ces deux modes entre eux (à droite de Φ_1).

L'ensemble de test est généré dans les mêmes conditions (cf. Figure 9). La mesure de distance est faite selon une métrique Euclidienne. Notons que, pendant la phase de classification (Fc+2M), le rejet de distance n'est pas prise en compte. Les résultats obtenus sont montrés dans la Figure 10 (c=2).

Les paramètres Fc+2M sont : $m = 1.5$ et $\alpha = 0.1$. La partition obtenue par l'algorithme Fc+2M est représentée dans la Table 2 pour $c = 2$ et dans la Table 3 pour $c = 3$. Toutes les valeurs de ces tables sont exprimées en pourcentages. Les résultats après fusion sont montrés dans la Table 4 (pour $c = 2$).

Une comparaison avec les résultats de la théorie bayesienne de la décision appliquée à l'ensemble des données est effectuée. La Table 5 montre les résultats obtenus avec les trois méthodes (bayesien, algorithme 3C avec $c = 2$ et $c = 3$). La meilleure performance est obtenue pour la méthode 3C (c=2, lors de la coalescence par l'algorithme Fc+2M).

6 Conclusion

Dans cet article, nous montrons l'utilité du rejet d'ambiguïté dans la phase de coalescence pour le partage des données entre plusieurs classifieurs. Chaque classifieur est spécialisé sur une région particulière de l'espace de représentation, ce qui lui permet de mieux décider l'attribution d'un individu à une classe. Lorsqu'il y a ambiguïté, la décision est prise suite à la fusion des sorties des classifieurs voisins à l'aide de la règle de combinaison

$\nearrow$	Φ_1	Φ_2
Φ_1	95.0	5.0
Φ_2	0.91	99.09

TAB. 4 – Algorithme 3C ; Matrice de confusion pour $c = 2$.

	Bayes	3C (c=2)	3C (c=3)
Succès	88.6	97.1	96.7
Confusion	11.4	2.9	3.3

TAB. 5 – Comparaison avec le classifieur bayesien : le mauvais résultat du classifieur Bayesien vient du fait que la deuxième classe est bi-modale

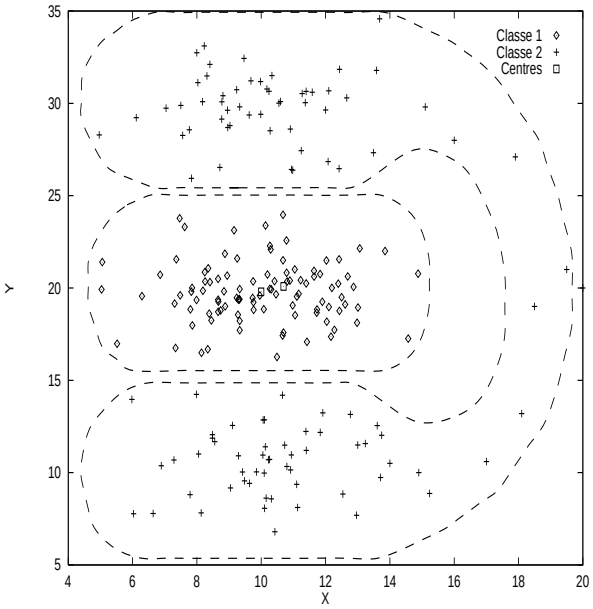

FIG. 8 – *Ensemble d'apprentissage :* $X^{lr} \subset X$.

de Dempster. Une des améliorations de notre algorithme peut venir de l'utilisation des modèles de fusion proposés par Appriou. En prenant en compte également le rejet de distance, cette méthode nous permettra de traiter l'évolution des classes et permettre ainsi l'apparition de nouvelles classes et la disparition de classes existantes.

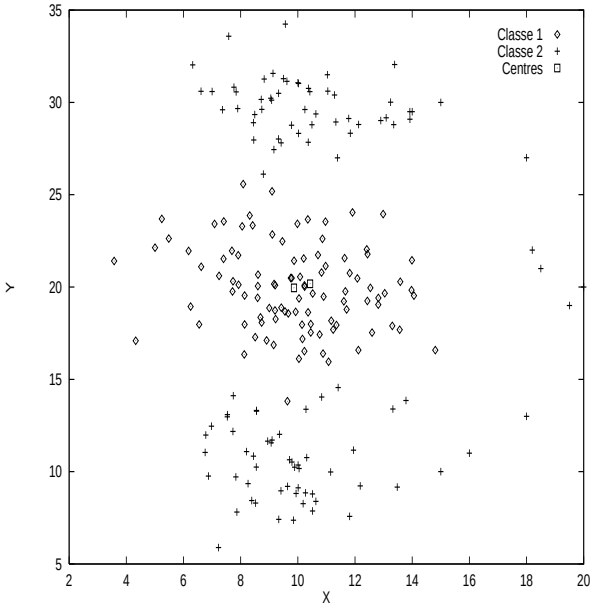

FIG. 9 – *Ensemble de test :* $X^t \subset X$.

Références

[1] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, pages 27–40, Jan. 1991.

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

[3] A. Black and A. Zisserman. Visual reconstruction. In *Cambridge*, MA :MIT Press, 1987.

[4] C.K.Chow. On optimum recognition error and reject tradeoff. *IEEE Trans. on Information Theory*, 16 :41–46, January 1990.

[5] C.K.Chow. Recognition error and reject trade-off. In *Proc. Third Ann. Symp. on Document Analysis and Information Retrieval*, volume 13, pages 1–8, Univ.of Nevada, Las Vegas, USA, April 1994.

[6] D. Dubois and H. Prade. Possibility theory. *Plenum Press. New York*, 1988.

[7] B. Dubuisson. Decision with reject option. In *European Signal Processing Conference*, Barcelone, SPAIN, 1990.

[8] R. A. Fisher. The statistical utilisation of multiple measurements. *Ann. Eugen.*, 8, 1938.

[9] T.M. Ha. The optimum class-selective rejection rule. In *IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.19, No.6*, pages 608–614, Yokohama, Japan, March 1997.

[10] T. K. Ho, J.J. Hull, and S.N. Srihari. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 :66–75, jan 1994.

[11] Y.S. Huang and Suen C.Y. A method of combining multiple classifiers - a neural network approach. *Proc. 12th Int'l Conf. Pattern Recognition and Computer Vision. Jerusalem*, pages 473–475, 1994.

[12] F. Kimura and M. Shridar. Handwritten numerical recognition based on multiple algorithms. *Pattern recognition*, 24(10) :969–983, 1991.

[13] L. Lam and C.Y. Suen. A theoritical analysis of the application of majority voting to pattern recognition. *Int'l Conf. pattern Recognition, Jerusalem*, pages 418–420, 1994.

[14] P. Loonis and M. Ménard. z_{01}, a reliability limit in pattern recognition : Application in exterior scenes area and automatic fish sorting. In *Proceedings of IPMU*, Paris, july 1998.

[15] M. Ménard, C. Demko, and P. Loonis. The fuzzy c+2-means : Solving the ambiguity reject in clustering. *Pattern recognition*, 33(7) :1219–1237, 2000.

[16] T. Pavlidis. A critical survey of image analysis methods. In *Proceedings of ICPR*, pages 502–511, 1986.

[17] Z. Pawlak. Rough sets. theorical aspects of reasoning about data. *Kluwer Academic Publishers. Dordrecht, Boston, London*, 1991.

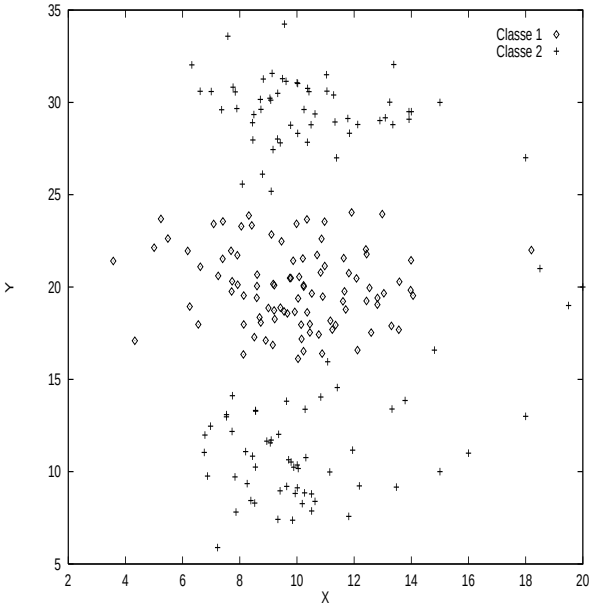

FIG. 10 – *Résultats obtenus avec notre méthode de coopération (3C).*

[18] P.Loonis, M.Ménard, and C.Demko. A new genetic algorithm for the multi-classifiers fusion optimization. In *Proceedings of IPMU*, Granada, Spain, July 1996.

[19] G. Shafer. A mathematical theory of evidence. *Princeton University Press,Princeton, N.J.*, 1976.

[20] K. Woods, W.P. Kegelmeyer, and JR.K. Bowyer. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4), 1997.

[21] L. Xu, A. Krzyzak, and C.Y. Suen. Methods of combining multiple classifiers and their applications to handwritting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3) :418–435, 1992.