

Combinaison de classification supervisée et non-supervisée par la théorie des fonctions de croyance

Fatma Karem*, Mounir Dhibi*
Arnaud Martin**

*Unité de Recherche PMI 09/UR/13-0
Campus Universitaire Zarouk Gafsa 2112, Tunisie
fatoumacy@yahoo.fr, mounir.dhibi@ensta-bretagne.fr,
**Université de Rennes 1, UMR 6074 IRISA
Rue Edouard Branly BP 30219, 22302 Lannion Cedex, France
Arnaud.Martin@univ-rennes1.fr

Résumé. Nous proposons dans cet article une nouvelle approche de classification fondée sur la théorie des fonctions de croyance. Cette méthode repose sur la fusion entre la classification supervisée et la classification non supervisée. En effet, nous sommes face à un problème de manque de données d'apprentissage pour des applications dont les résultats de classification supervisée et non supervisée sont très variables selon les classificateurs employés. Les résultats ainsi obtenus sont par conséquent considérés comme incertains.

Notre approche se propose de combiner les résultats des deux types de classification en exploitant leur complémentarité *via* la théorie des fonctions de croyance. Celle-ci permet de tenir compte de l'aspect d'incertitude et d'imprécision. Après avoir dressé les différentes étapes de notre nouveau schéma de classification, nous détaillons la fusion de classificateurs. Cette nouvelle approche est appliquée sur des données génériques, issues d'une vingtaine de bases de données. Les résultats obtenus ont montré l'efficacité de l'approche proposée.

1 Introduction

La classification est un moyen utile d'organisation et de hiérarchisation des données. Le but de la classification non-supervisée est de trouver des groupes compacts et bien séparés dans un ensemble de données et donc d'affecter à chaque observation une étiquette de classe qui matérialise l'appartenance de celle-ci aux classes dégagées. De plus on souhaite pouvoir également affecter à toute nouvelle observation une étiquette. Cette situation n'est pas rencontrée en classification supervisée puisque les observations disponibles sont déjà étiquetées le but à atteindre est d'affecter à une nouvelle observation, une classe préexistante et apprise sur les données d'apprentissage. Le problème dans le contexte de la classification non-supervisée est plus difficile puisqu'aucune information n'est fournie sur l'appartenance des données à telle ou telle classe. Cette appartenance est généralement déduite à partir de la répartition spatiale des points expliquée par Campedel (2005). Indépendamment du type de classification, le choix de

la méthode se pose toujours pour un problème donné. En effet, il n'existe pas d'approche optimale pour n'importe quel type de données. De plus, une fois l'approche choisie, les résultats sont très dépendants du paramétrage, souvent nécessaire, de la méthode. Dans le cas supervisé, les données d'apprentissage ne représentent pas toujours parfaitement la réalité des données en question, ce qui peut dégrader les résultats de la classification. En particulier, les classes retenues pour l'apprentissage de la classification supervisée, ne sont pas toujours pertinentes. En effet, il peut arriver que des classes soient oubliées, ou encore que l'étiquetage ait été réalisé dans de mauvaises conditions ne permettant pas la distinction de classes intéressantes. La classification non-supervisée ne se positionnant pas sur ce type de problème, puisqu'elle détermine les classes, peut venir en aide à une classification supervisée, pour confirmer ou infirmer le choix des classes initiales, et proposer une classification avec des classes ayant plus de sens. Une solution envisagée est alors de fusionner les résultats donnés par la classification supervisée et non-supervisée en vue de tirer profit des deux types de méthodes et répondre à un tel problème. Les résultats des deux types de classification peuvent ainsi être vus dans un système de fusion comme des données hétérogènes, imprécises et incertaines. La combinaison offre par conséquent à l'utilisateur un compromis entre les deux méthodes.

Les précédents travaux réalisés sur la fusion de classifications concernent l'un des deux types c'est-à-dire la fusion de classifications non-supervisées (par exemple proposée par Forestier et Gañarski (2008); Gañarski et Wemmert (2005); Wemmert et Gañarski (2002); Masson et Denœux (2004, 2011)); ou supervisées (par exemple proposée par Xu et al. (1992); Martin (2010)). Les approches de fusion de classifications non-supervisées restent plus délicates vu l'absence d'information sur les labels des classes. On remarque aussi que la fusion entre non-supervisé et supervisé a été réalisée non pas dans un but propre de classification dans le sens où le non-supervisé était déployé pour faire l'apprentissage du supervisé par Guijarro et Pajares (2009); Urszula et Switek (2009); Prudent et Ennaji (2004).

Cet article propose donc une approche de fusion en vue d'améliorer les résultats de classification supervisée. Elle est fondée sur la théorie des fonctions de croyance. Celle-ci offre un cadre théorique solide pour modéliser les aspects d'incertitude et d'imprécision liés à notre problème de classification et permet de gérer les conflits entre les approches supervisées et non-supervisées. En effet, elle modélise la croyance en un événement par une fonction appelée fonction de masse et travaille sur un espace étendu d'hypothèses qui englobe toutes les disjonctions et les intersections possibles du cadre de discernement retenu. Dans la section 2, on présentera la classification supervisée et non-supervisée. Dans la section 3 nous décrivons la fusion d'informations dans le cadre de la théorie des fonctions de croyance. Dans la section 4 nous mettons l'accent sur l'approche proposée et on finira par l'étude expérimentale dans la section 5.

2 Classification

L'objectif de la classification est d'identifier les classes auxquelles appartiennent des objets à partir de traits descriptifs (attributs, caractéristiques, etc.). On distingue essentiellement deux types de classification : supervisée et non-supervisée.

Classification non-supervisée

Cette classification est aussi appelée "classification automatique", "clustering" ou encore "regroupement". Dans ce type de classification on est amené à identifier les populations d'un ensemble de données. On suppose qu'on dispose d'un ensemble d'objets que l'on note par $X = \{x_1, x_2, \dots, x_N\}$ caractérisé par un ensemble de descripteurs D , l'objectif du clustering est de trouver les groupes auxquels appartiennent chaque objet x qu'on note par $C = \{C_1, C_2, \dots, C_n\}$. Ce qui revient à déterminer une fonction notée Y_s qui associe à chaque élément de X un ou plusieurs éléments de C . Il faut pouvoir affecter une nouvelle observation à une classe. Les observations disponibles ne sont pas initialement identifiées comme appartenant à telle ou telle population. L'absence d'étiquette de classe est un lourd handicap qui n'est que très partiellement surmontable. Seule l'analyse de la répartition spatiale des observations peut permettre de "deviner" où sont les véritables classes. Les deux difficultés essentielles que rencontre la classification non supervisée sont les suivantes :

- S'il est naturel de reconnaître comme "appartenant à une même classe" des observations regroupées dans une même zone de forte densité, il n'en est pas de même dans des zones de faible densité. En particulier, on peut s'attendre à ce que la définition de frontières entre les classes soit sujette à caution, et pour le moins hasardeuse.
- L'œil humain est un extraordinaire outil de classification non supervisée. Malheureusement, il n'est opérationnel que pour des données bidimensionnelles, alors que les données que rencontre l'analyste sont couramment décrites par des dizaines de variables ou plus. Il s'avère que reproduire les performances de l'œil humain dans des espaces de grande dimension est un exploit aujourd'hui hors d'atteinte des machines.

Parmi les méthodes non-supervisées les plus utilisées, citons deux types d'approches : les centres mobiles (k -means) et la classification hiérarchique.

Classification supervisée

Dans le contexte supervisé on dispose déjà d'exemples dont la classe est connue et étiquetée. Les données sont donc associées à des labels des classes notés $\Theta = \{q_1, q_2, \dots, q_n\}$. L'objectif est alors d'apprendre à l'aide d'un modèle d'apprentissage des règles qui permettent de prédire la classe des nouvelles observations ce qui revient à déterminer une fonction Cl qui à partir des descripteurs (D) de l'objet associe une classe q_i et de pouvoir aussi affecter toute nouvelle observation à une classe parmi les classes disponibles. Ceci revient à la fin à trouver une fonction qu'on note Y_s qui associe chaque élément de X un élément de Q . On construit alors un modèle en vue de classer les nouvelles données. Parmi les méthodes supervisées on cite : les k -plus proches voisins, les arbres de décision, les réseaux de neurones, les machines à support de vecteurs (SVM) et les classificateurs de Bayes.

Quelque soit le type de la classification, on est confronté à différents problèmes. Dans le cas supervisé, un problème important peut être le manque de données pour réaliser l'apprentissage ou la disponibilité de données inadéquates par exemple incertaines et imprécises ce qui empêche la construction d'un modèle correct. Pour la classification non-supervisée, la délimitation des frontières entre les classes n'est pas toujours franche et reconnaissable. Indépendamment du type de classification, les données multi dimensionnelles, ou encore la dépendance des méthodes de classification aux paramètres initiaux comme le nombre de classes peuvent poser problèmes. Afin de mesurer les performance des approches, on cherche à trouver des critères

sur la qualité des résultats. Généralement nous avons recours à des indices qu'on appelle indices de validité. Il faut choisir l'indice le plus adéquat aux données ; il n'y a pas d'indice standard.

Dans cet article nous proposons une nouvelle approche afin de surmonter les problèmes de classification. Celle-ci est fondée sur la fusion entre la classification supervisée et non supervisée. Le cadre de cette fusion est fondé sur la théorie des fonctions de croyance que nous présentons dans la section suivante.

3 Fusion d'informations dans le cadre de la théorie des fonctions de croyance

La fusion de données ou d'informations peut être faite selon 3 niveaux liés au processus de la classification : données, caractéristiques et décisions. Nous nous concentrons sur le troisième niveau. Les approches de fusion d'informations sont essentiellement fondées sur les théories de l'incertain afin de modéliser finement les imprécisions et incertitudes. Parmi celles-ci, la théorie des probabilités offre une modélisation de l'incertitude, la théorie des sous-ensembles flous des imprécisions et la théorie des fonctions de croyance, qui nous intéresse le plus dans ce travail, permet une modélisation des deux. La théorie des fonctions de croyance (ou théorie de Dempster-Shafer) permet de représenter à la fois l'imprécision et l'incertitude à travers deux fonctions : la fonction de crédibilité et la fonction de plausibilité. Ces deux fonctions sont dérivées des fonctions de masses qui sont définies sur des sous-ensembles et non des singletons comme dans la théorie des probabilités, permettant de tenir compte des imprécisions. On désigne par m_j la fonction de masse relative à la source S_j définie sur 2^Θ , à valeur dans $[0, 1]$ et vérifiant la contrainte suivante :

$$\sum_{A \in 2^\Theta} m_j(A) = 1 \quad (1)$$

2^Θ est l'ensemble des disjonctions possibles des décisions q_i ou classes q_i dans notre contexte de classification par exemple : $2^\Theta = \{\emptyset, \{q_1\}, \{q_2\}, \{q_1 \cup q_2\}, \dots, \Theta\}$.

Les décisions d_i doivent être exclusives mais pas nécessairement exhaustives. Le choix de la fonction de masse est délicat. Plusieurs approches ont été proposées nous citons deux approches une fondée sur un modèle probabiliste de Appriou (2002) et l'autre sur une transformation en distance de Denœux (2008). La fonction de crédibilité est donnée par :

$$Cr_j(X) = \sum_{Y \subset X, X \neq \emptyset} m(Y) \quad (2)$$

Les fonctions de plausibilité représentent l'intensité avec laquelle on ne doute pas en un élément donnée pour tout X par :

$$Pl_j(X) = \sum_{Y \in 2^\Theta, Y \cap X \neq \emptyset} m_j(Y) \quad (3)$$

$$= Cr_j(\Theta) - Cr_j(X^c) \quad (4)$$

$$= 1 - m_j(\emptyset) - Cr_j(X^c) \quad (5)$$

où X^c est le complémentaire de X .

Plusieurs modes de combinaison ont été développés dans le cadre de la théorie des fonctions de croyance. Les plus importants sont : la combinaison conjonctive et la combinaison disjonctive qui ont été déclinées en un grand nombre d'opérateurs de combinaison dont des combinaisons mixtes. On donne l'exemple de la combinaison conjonctive introduite par Dempster et reprise par Shafer, elle combine les fonctions de masse en considérant les intersections des éléments de 2^Θ comme rappelé par Martin (2010). Elle s'écrit d'une façon générale pour M fonctions de masses par :

$$m(A) = (m_1 \oplus m_2, \dots \oplus m_M)(A) \quad (6)$$

$$= \sum_{B_1 \cap B_2, \dots \cap B_M = A} \prod_{j=1}^M m_j(B_j) \quad (7)$$

On obtient à la fin les masses relatives à chaque élément du cadre de discernement issue de la combinaison des différentes sources. Il est alors aisé de calculer les fonctions de croyance et de plausibilité. La dernière étape qui reste est celle de prise de décision c'est-à-dire le choix de la décision q_i ou d'une disjonction. Ce choix peut se faire par la maximisation d'un critère. Les critères sont multiples. Ils se fondent essentiellement sur les fonctions de plausibilité et de crédibilité. Nous citons trois critères : le maximum de plausibilité, le maximum de crédibilité et la probabilité pignistique. Pour le premier critère on prend le singleton q_i donnant le maximum de plausibilité. Pour une observation x , nous décidons q_i si :

$$Pl_j(q_i)(x) = \max_{1 \leq k \leq n} Pl(q_k)(x) \quad (8)$$

Ce critère est très optimiste puisque la plausibilité d'un singleton représente la croyance que nous aurions si toutes les masses des disjonctions l'incluant étaient focalisées sur ce singleton. Le deuxième critère choisit q_i pour une observation x si elle donne le maximum de crédibilité :

$$Cr_j(q_i)(x) = \max_{1 \leq k \leq n} Cr(q_k)(x) \quad (9)$$

Ce critère est applicable seulement si la combinaison porte seulement sur les singletons ce qui est rarement rencontré. Il est plus sélectif que le précédent. En effet la fonction de crédibilité fournit la croyance minimale en une décision c'est-à-dire celle obtenue si toutes les masses des disjonctions l'incluant n'étaient pas focalisées sur cette décision mais sur les autres. Un compromis entre le maximum de plausibilité et le maximum de crédibilité est celui de la probabilité pignistique. Elle approche le couple crédibilité et plausibilité en équirépartissant les masses placées sur les éléments différents d'un singleton sur les singletons qui les composent. Ainsi pour une décision q_i , la probabilité pignistique est définie par :

$$bet(q_i) = \sum_{A \in 2^\Theta, q_i \in A} \frac{m(A)}{|A|(1 - m(q_i))} \quad (10)$$

où $|A|$ représente le cardinal de A . Ainsi le critère du maximum de probabilité pignistique revient à décider q_i pour l'observation x si :

$$bet(q_i)(x) = \max_{1 \leq k \leq n} bet(q_k)(x) \quad (11)$$

Ce critère est peu conforme à la notion de masse il s'emploie dans un contexte probabiliste. Dans ce qui suit nous présentons l'approche menée dans cet article.

4 Fusion crédibiliste de la classification supervisée et non-supervisée

La plupart des travaux déjà réalisés sur la fusion concernent la classification non-supervisée. Par exemple, dans Forestier et Gançarski (2008) la combinaison est faite à travers une collaboration des classifieurs. Le processus se compose de trois étapes : exécution initiale et parallèle des classifieurs, raffinement des résultats puis unification des résultats. Dans la seconde étape on tend à converger les différentes classifications par résolution des conflits existants. La convergence est faite grâce à la recherche de conflits entre les différentes classifications. La recherche est faite deux à deux. Pour deux résultats on cherche la correspondance entre les classes (*clusters*) via une mesure de similarité. S'il y a un conflit entre deux résultats au niveau de quelques clusters on le résout soit par fusion ou division. Les résolutions des conflits peuvent être néfastes sur le résultat global c'est-à-dire qu'on peut converger vers un résultat qui contient une seule classe qui englobe tous les pixels de l'image par exemple. Par conséquent on ne choisit que les résolutions qui améliorent le résultat global. Une fois le raffinement terminé. On fusionne les classifications par méthode de vote.

Pour ce qui est de la fusion de classification supervisée on cite le travail fait dans Martin (2005) où une comparaison est faite entre plusieurs méthodes de fusion pour la classification d'images sonar décrites par des méthodes d'analyse de texture. Les fonds marins présentent des zones de sédiments homogènes ou non qui peuvent s'apparenter à des textures. La caractérisation des fonds marins est un problème difficile en soi vu l'inexactitude des appareils de mesures et la multitude des méthodes d'analyse. Une solution choisie était de fusionner les classifieurs. Ces derniers sont composés d'un perceptron multicouche pour 4 approches d'extraction de texture. La fusion d'informations est de haut niveau elle se fait soit au niveau des sorties numériques des perceptrons (6 sorties correspondant aux 6 classes en question à valeur dans $[0,1]$) soit au niveau des sorties symboliques représentant les classes affectées. La fusion était testée avec la méthode de vote, une approche possibiliste et une approche crédibiliste testée avec les modèles de distance et de probabilité pour le calcul des masses. Ce qu'il faut mentionner est que la théorie des fonctions de croyance donne de meilleurs résultats.

Concernant la fusion de deux types de classification supervisée et non-supervisée, les approches développées sont essentiellement utilisées en vue de déployer le non-supervisé pour faire l'apprentissage du supervisé Guijarro et Pajares (2009); Urszula et Switek (2009). Dans Guijarro et Pajares (2009), la combinaison est faite grâce à une approche de décision floue multi-critère (MCDM : Fuzzy multicriteria decision making approach). Le fonctionnement du nouveau classifieur se fait selon 2 étapes : apprentissage et classification. Durant l'apprentissage, une partition optimale est construite à partir des données d'apprentissage. Le partitionnement se fait grâce au clustering flou (FC). Le nombre de clusters est mis à jour automatiquement jusqu'à tomber sur le nombre optimal de clusters. La validation de la partition se fait grâce à un critère : l'inertie intra-classe ou la somme des erreurs carrées. Le critère en question

a été normalisé afin d'obtenir une valeur comprise entre 0 et 1. Une fois la partition optimale trouvée (le nombre optimal de classes et les sous-ensembles S_i), on passe à l'estimation des paramètres des classifieurs restants. Chacun des classifieurs reçoit les centres initiaux obtenus pour la partition validée. Puis il les met à jour tout en tenant compte du nombre optimal de clusters trouvé. Tous les paramètres estimés de tous les classifieurs sont stockés. Ensuite on passe à l'estimation des compétences de chacun. On calcule la somme normalisée des erreurs carrées pour certains. Et on calcule pour d'autres le critère de variance minimale relative (VC) (Related minimum variance criteria). On obtient par conséquent à la fin les compétences de chaque classifieur tout en tenant compte des sous-ensembles S_i avec leur nombre estimé de clusters. Une fois finie l'étape d'apprentissage on passe à la classification. Il faut chercher à quelles classes appartiennent les nouveaux vecteurs x . La décision est prise grâce à la combinaison des supports fournis par chacun des classifieurs et leurs compétences à travers une méthodologie floue multi-critères (MCDM). On garde toujours le même nombre de classes. L'approche en question considère deux critères : bénéfice Critère1 et côte Critère2 s'appliquant chacun à certains classifieurs. Critère1 tient compte des degrés flous d'appartenance de x aux clusters et les probabilités de x sachant les clusters existants en utilisant les fonctions de densité de probabilité. Le critère (Critère2) utilise la distance euclidienne de x à chacun des centres. Une fois calculés ces deux critères pour chacun ils seront pondérés par des poids qui tiennent compte des performances des classifieurs calculés dans l'étape d'apprentissage. On procède après à une construction de table de décisions de performances normalisées qui tient compte des critères Critère1 et Critère2 calculés pour les classifieurs en question et les classes dégagées lors de la phase d'apprentissage. On sélectionne la meilleure alternative pour le nouveau vecteur (une classe parmi celles dégagées). L'approche était appliquée sur des images naturelles texturées ce sont des images multi-spectrales prises d'une région en Espagne.

L'approche menée dans cet article fusionne les deux classifications afin d'améliorer son résultat. Elle se compose de deux étapes principales. La première consiste à appliquer la classification non-supervisée et supervisée séparément sur les données. La seconde étape consiste à fusionner les résultats issus de deux méthodes. Ayant comme entrées à notre processus de fusion deux résultats issus de deux sources différentes. On essaie de mesurer la précision de chaque résultat via la théorie des fonctions de croyance. Il faut calculer les fonctions de masses associées à chacun. Un premier obstacle rencontré est le choix du modèle adéquat pour les fonctions de masses. On adoptera essentiellement le modèle probabiliste d'Appriou. Ce choix est fait du à la simplicité d'usage des probabilités pour une fusion de décisions (singletons). En effet si le choix est porté sur un modèle de distance on rencontrait le problème du choix de la meilleure métrique. On dispose donc de deux sources fournissant deux fonctions de masses. Chacune donne ses mesures relativement à chaque observation ou objet x pour chacun des éléments du cadre de discernement $\Theta = \{q_i, i = 1, \dots, n\}$. L'ignorance, Θ est composée d'un ensemble de n classes. Pour la construction des fonctions de masse du classifieur supervisé nous utilisons le modèle d'Appriou.

Si la classification supervisée donne la label q_j à l'observation x , nous avons n fonctions de masse suivantes avec les 3 éléments focaux :

$$m_i(q_j)(x) = \frac{\alpha_{ij} R_i p(q_i|q_j)}{1 + R_i p(q_i|q_j)} \quad (12)$$

$$m_i(q_j^c)(x) = \frac{\alpha_{ij}}{1 + R_i p(q_i|q_j)} \quad (13)$$

Combinaison crédibiliste de classification supervisée et non-supervisée

$$m_i(\Theta)(x) = 1 - \alpha_{ij} \quad (14)$$

On désigne par : q_i la classe réelle et α_{ij} le coefficient de fiabilité de la classification supervisée pour la classe q_j . Les probabilités conditionnelles sont estimées à partir de la matrice de confusion :

$$\alpha_{ij} = \max p(q_i|q_j) (i = 1, \dots, n) \quad (15)$$

$$R_i = \max_{q_l} (p(q_i|q_l))^{-1} \quad (16)$$

Pour ce qui est du côté non-supervisé, on est censé octroyer des masses aux classes en question qu'on ignore (du côté non-supervisé). Les sorties du clustering sont les clusters. On octroie les masses alors en mesurant les similarités entre les clusters et les classes. La similarité est estimée non pas par calcul des distances entre les clusters et les classes mais par calcul de recouvrement entre les deux. En effet, la distance entre les deux ne renseigne pas parfaitement sur la similarité entre les deux. On peut avoir des couples de classes (cluster et classe) ayant la même distance qui les séparent mais ne sont pas séparées de la même façon. Généralement pour le calcul on mesure la distance séparant les centres des classes. Pour surmonter ce problème on calcule le recouvrement. Une classe est considérée comme similaire à un cluster et réciproquement si elle est recouverte en totalité par celui-ci. Plus elle a d'éléments en commun plus elle est similaire. Pour se faire on cherchera les proportions des classes q_1, \dots, q_n dans chaque cluster (Gançarski et Wemmer (2005); Forestier et Gançarski (2008)). $\forall x \in C_i$ avec c regroupements trouvés. Nous définissons c fonctions de masse par :

$$m_i(q_1)(x) = \frac{|C_i \cap q_1|}{|C_i|} \quad (17)$$

$$m_i(q_j)(x) = \frac{|C_i \cap q_j|}{|C_i|} \quad (18)$$

$$m_i(q_n)(x) = \frac{|C_i \cap q_M|}{|C_i|} \quad (19)$$

On affaiblit par la suite les masses tel que $\forall A \in 2^\Theta$:

$$m_i^{\alpha_i}(A) = \alpha_i m_i(A) \quad (20)$$

$$m_i^{\alpha_i}(\Theta) = 1 - \alpha_i(1 - m_i(\Theta)) \quad (21)$$

Le coefficient d'affaiblissement α_i dépend des points. Ainsi on peut affaiblir de la même façon pour tous les points. Un point situé au centre d'un cluster se considère comme plus représentant du cluster en question qu'un point situé à la frontière, α_i est défini par (v_i est le centre du cluster C_i) :

$$\alpha_i = e^{-\|x-v_i\|^2} \quad (22)$$

Une fois les masses calculées on peut fusionner à l'aide de la règle de Dempster et on adopte comme critère de décision le maximum de la probabilité pignistique. On cherche dans notre problème des singletons qui sont déjà connus vu l'emploi de la classification supervisée. Chaque pixel est affecté à une classe précise.

5 Etude expérimentale

Nous donnons dans cette section les résultats obtenus pour notre approche de fusion entre une classification supervisée et une classification non-supervisée. Pour cela, nous avons réalisé notre étude expérimentale sur différentes données issues des bases de données génériques. L'objectif est de montrer la performance de la méthode proposée et par conséquent l'apport de la fusion dans la classification. Notre expérience est fondée sur trois méthodes non-supervisées : *C*-means, FCM et modèle de mélange et une méthode supervisée : *k*-plus proches voisins testée pour différentes valeurs de *k* allant de 1 à 7. Nous commençons par présenter les bases puis nous montrons dans les tableaux 1 et 2 les taux de bonne classification trouvés pour les bases de données avant et après fusion respectivement pour FCM et K-PPV, Kmeans et K-PPV et modèle de mélange et K-PPV.

5.1 Description des bases de données

Dans ce travail nous avons utilisés plusieurs types de données.

- Iris : C'est la base de données la plus populaire dans le domaine de reconnaissance de formes. Elle contient 3 classes de 50 échantillons chacun. Chaque classe se réfère à un type de la plante iris. Une classe est linéairement séparable des deux autres. Les deux autres ne sont pas linéairement séparables.
- Haberman : La base contient des cas pris d'une étude conduite entre 1958 et 1970 à l'université de Chicago. Elle vise l'étude de la survie des patients qui ont subi une opération chirurgicale sur le cancer du sein. La base contient 306 instances. L'objectif cherché est de trouver le status de survie du patient soit il va vivre pour 5 ans ou plus ou il va mourir durant les cinq ans.
- sensor-readings-4 : La base contient des mesures de capteurs suivant le mur de navigation d'un robot mobile. L'objectif est de dégager les différentes directions prises par le robot : aller tout droit ou tourner à gauche ou à droite. Les échantillons sont distribués comme suit : première classe contient 2205 échantillons (40.41%), deuxième classe contient 826 échantillons (15.13%), troisième classe : 2097 échantillons (38.43%) et quatrième classe : 328 échantillons (6.01%).

5.2 Expérimentation

La fixation du nombre de clusters testé est soit égal au nombre de classes donné par le supervisé(K-PPV) ou dépend de l'utilisateur qui peut le refixer. Nous validons notre travail avec la méthode de validation croisée. Nous itérons le processus 10 fois variant à chaque fois l'ensemble d'apprentissage et l'ensemble de test. L'effet de la fusion est remarquable au niveau du premier tableau. En effet on a atteint un taux de 100% pour les données : iris et sensor-readings4 et un taux supérieur à 66% pour la donnée Haberman et abalone. Dans le tableau 2 et 3, on a un taux de 100% pour iris et sensor-readings4 et un taux proche de 70% ou même plus pour les autres.

Combinaison crédibiliste de classification supervisée et non-supervisée

Données	NbC	NbCl	NbA	TC-AF	TC-ApF			
					k=2	k=3	k=5	k=7
Iris	3	3	5	96.00	100.00	100.00	100.00	100.00
Sensor-readings-4	4	4	5	98.70	96.63	100.00	100.00	100.00
Habeman	2	2	4	68.63	80.00	77.67	70.65	70.67
Abalone	2	2	8	49.53	77.07	76.91	68.92	67.03

TAB. 1 – Résultats obtenus avec FCM et KPPV. NbC : nombre de classes, NbCl : nombre de clusters testé, NbA : Nombre d'attributs, TC-AF : Taux de bonne classification avant fusion, TC-ApF Taux de bonne classification après fusion

Données	NbC	NbCl	NbA	TC-AF	TC-ApF			
					k=2	k=3	k=5	k=7
Iris	3	3	5	96.00	100.00	100.00	100.00	100.00
Sensor-readings-4	4	4	5	98.70	100.00	100.00	100.00	100.00
Habeman	2	2	4	68.63	80.32	77.67	70.32	70.99
Abalone	2	2	8	49.53	72.64	77.20	69.95	67.03

TAB. 2 – Résultats obtenus avec Kmeans et KPPV. NbC : nombre de classes, NbCl : nombre de clusters testé, NbA : Nombre d'attributs, TC-AF : Taux de bonne classification avant fusion, TC-ApF Taux de bonne classification après fusion

6 Conclusion

Cet article propose une nouvelle approche permettant la fusion des résultats de classification supervisée et non-supervisée. Cette approche originale fondée sur la théorie des fonctions de croyance permet de lever certaines ambiguïtés et de gérer les conflits. Chacune des deux classifications supervisée et non-supervisée montrent des limites que ce soient au niveau des méthodes ou au niveau des données testées. L'approche proposée a montré des résultats encourageants sur des données génériques. Le présent travail est fait sur des bases de données complètes. Celui-ci peut être étendu par l'élargissement de la base de données. On envisagerait faire le test sur des bases comprenant des erreurs d'étiquetage. Nous envisagerons aussi le test sur des images réelles telles que les images sonar et/ou des images médicales ainsi que l'approfondissement du processus de fusion avec d'autres méthodes.

Références

- Appriou, A. (2002). *Décision et Reconnaissance des formes en signal, chapitre Discrimination multisignal par la théorie de l'évidence*. France, FR : Hermes Science Publication.
- Campedel, M. (2005). Classification supervisée. Technical report, Telecom Paris.
- Denœux, T. (2008). A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man and Cybernetics* 25, 904–913.

Données	NbC	NbCl	NbA	TC-AF	TC-ApF			
					k=2	k=3	k=5	k=7
Iris	3	3	5	96.00	100.00	100.00	100.00	100.00
Sensor-readings-4	4	4	5	98.70	100.00	100.00	100.00	100.00
Haberman	2	2	4	68.63	80.33	77.33	70.98	70.63
Abalone	2	2	8	49.53	71.71	75.60	71.22	66.69

TAB. 3 – Résultats obtenus avec Modèle de mélange et KPPV. NbC : nombre de classes, NbCl : nombre de clusters testé, NbA : Nombre d'attributs, TC-AF : Taux de bonne classification avant fusion, TC-ApF Taux de bonne classification après fusion

- Forestier, G. Wemmert, G. et P. Gañarski (2008). Multisource images analysis using collaborative clustering. *EURASIP Journal on Advances in Signal Processing* 11, 374–384.
- Gañarski, P. et G. Wemmert (2005). Collaborative multi-strategy classification: application to per-pixel analysis of images. In *In Proceedings of the 6th international workshop on Multimedia data mining: mining integrated media and complex data*, Volume 6, pp. 595–608.
- Guijarro, M. et G. Pajares (2009). On combining classifiers through a fuzzy multi-criteria decision making approach: Applied to natural textured images. *Expert Systems with Application* 39, 7262–7269.
- Martin, A. (2005). Comparative study of information fusion methods for sonar images classification. In *In Proceeding of the 8th International Conference on Information Fusion*, Volume 2, pp. 657–666.
- Martin, A. (2010). Le conflit dans la théorie des fonctions de croyance. In *10ème journées Francophones : Extraction et Gestion des Connaissances*.
- Masson, M. et T. Dencœur (2004). Clustering interval-valued proximity data using belief functions. *Pattern Recognition Clustering interval-valued proximity data using belief functions* 25, 163–171.
- Masson, M. et T. Dencœur (2011). Ensemble clustering in the belief functions framework. *International Journal of Approximate Reasoning* 52, 92–109.
- Prudent, Y. et A. Ennaji (2004). Clustering incrémental pour un apprentissage distribué : vers un système évolutif et robuste. In *In Conférence CAP*, Volume 3287, pp. 446–453.
- Urszula, M. K. et T. Switek (2009). Combined unsupervised-supervised classification method. In *In Proceedings of the 13th International Conference on Knowledge Based and Intelligent Information and Engineering Systems: Part II*, Volume 13, pp. 861–868.
- Wemmert, C. et P. Gañarski (2002). A multi-view voting method to combine unsupervised classifications. In *In Proceedings of the 2nd IASTED International Conference on Artificial Intelligence and Applications*, Volume 2, pp. 447–453.
- Xu, A., L. Krzyzak, et C. Y. Suen (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics* 22, 418–435.

Summary

In this paper we propose a new classification approach. The latter is based on fusion between clustering and classification. In fact both types present limits such as dependance on parameters, lack of learning data and availability of uncertain data. All the mentioned drawbacks leads to uncertain results. Our approach aims to combine the results of both classification types by using their complementarity throw belief functions theory. The last one treats very well both aspects: uncertainty and imprecision. We present in this paper our classification scheme. After that we analyse the fusion process. This approach is applied on generic data issued from twenty databases. The results show the usefulness of the proposed method.