Combinaison de Classifieurs Binaires dans le cadre du Modèle des Croyances Transférables

Pairwise Classifier Combination in the Transferable Belief Model Framework

Benjamin Quost[1] Thierry Denœux[1] Mylène Masson[2]

[1,2] HeuDiaSyC – UMR UTC CNRS 6599

[1] Université de Technologie de Compiègne

BP 20529 - F-60205 Compiègne cedex – France

[2] Université de Picardie Jules Verne

{bquost, tdenoeux, mmasson}@hds.utc.fr

Résumé :

La combinaison de classifieurs constitue une approche intéressante pour la résolution de problèmes de discrimination multi-classes. Nous proposons d'effectuer cette combinaison dans le cadre de la théorie des fonctions de croyance. La méthode, similaire à celle proposée par Hastie et Tibshirani dans un cadre probabiliste, est tout d'abord présentée ; puis les performances obtenues sur plusieurs jeux de données sont analysées ; enfin, les perspectives ouvertes par l'utilisation de ce formalisme pour la résolution de problèmes de discrimination sont évoquées en conclusion.

Mots-clés :

Fonctions de croyance, théorie de Dempster-Shafer, Reconnaissance des Formes, Classification.

Abstract:

Classifier combination constitutes an interesting approach when solving multi-class classification problems. We propose to carry out this combination in the belief functions framework. The method, similar to the one proposed by Hastie and Tibshirani in a probabilistic framework, is first presented ; then, the performances obtained on various datasets are analyzed ; the perspectives held by the use of this formalism in classification problems are eventually mentioned as a conclusion.

Keywords:

Belief functions, Dempster-Shafer theory, Pattern Recognition, Classification.

1 Introduction

La résolution d'un problème de discrimination peut être vue de la manière suivante : soit un *ensemble d'apprentissage* constitué de données $X = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$, associées à des étiquettes $Y = \{y_1, \ldots, y_n\}$, dénotant leur appartenance à une classe $\omega_k \in \Omega = \{\omega_1, \ldots, \omega_K\}$. On cherche à construire, sur la base de cet ensemble de données étiquetées, un processus automatique ou *classifieur* permettant de déterminer la classe d'un nouvel individu $\mathbf{x}$.

Il est nécessaire de construire des classifieurs d'architecture adaptée à la complexité du problème traité : plus la situation considérée est complexe, plus les classifieurs construits le sont. Leur *coût d'apprentissage* peut donc devenir élevé en temps et en nombre de données d'apprentissage requises.

De plus, le *cadre théorique* de représentation des connaissances limite parfois la précision ou la robustesse du classifieur, en particulier s'il ne permet pas une modélisation suffisamment riche de l'appartenance des individus aux classes ; cela est notamment le cas lorsque les connaissances dont on dispose pour construire le classifieur sont imprécises ou incertaines.

Dans cet article, nous proposons une méthode pour construire des classifieurs de complexité adaptée, dans un cadre de représentation des connaissances riche et flexible. Nous présentons tout d'abord les motivations de notre approche, puis nous décrivons le *Modèle des Croyances Transférables* et les possibilités qu'il offre dans le cadre de la présente problématique, pour enfin exposer les résultats de cette méthode, avant de conclure sur les perspectives ouvertes par notre approche.

2 Combinaison de classifieurs binaires

2.1 Les avantages de la combinaison

Nous considérons ici le cas le plus général d'un problème de discrimination, où l'on cherche à déterminer la classe d'un individu $\mathbf{x}$ parmi un ensemble de $K > 2$ classes en compétition (problème dit *polychotomique*).

Approche directe. L'approche la plus couramment utilisée (approche *directe*) consiste à construire un classifieur reconnaissant l'ensemble des classes : la détermination de *frontières de décision* entre ces différentes classes permet de choisir une classe parmi les K lors de l'évaluation d'un nouvel individu $\mathbf{x}$. Pour certaines méthodes d'apprentissage, cette approche peut être coûteuse lors de la construction du classifieur.

Approche binaire. Une approche alternative, dite approche *binaire*, consiste à séparer l'ensemble des K classes en paires de classes distinctes, et à entraîner pour chaque paire un *classifieur binaire* pouvant séparer les deux classes. Un individu $\mathbf{x}$ est évalué par chacun des classifieurs binaires, et affecté à une classe suivant une *règle de décision* tenant compte de ces différentes évaluations. Cette approche permet de spécialiser les classifieurs utilisés : une seule frontière de décision doit être tracée pour chaque paire de classes.

Les paires peuvent être constituées en opposant chaque classe à toutes les autres (approche dite *une contre toutes*), ou chaque classe à chaque autre (approche *une contre une*). Dans le premier cas, chacun des K classifieurs binaires est entraîné à partir de l'ensemble des données ; dans le second cas, chacun des $K(K-1)/2$ classifieurs binaires est entraîné à partir d'un sous-ensemble de données. De ce fait, bien que le nombre de classifieurs soit moins important dans le cas *une contre toutes*, la complexité globale est moindre dans le cas *une contre une* [6].

Les *codes correcteurs d'erreur* permettent également de décomposer un problème multiclasses en problèmes binaires. On associe à chaque classe un *mot-code* de N *bits* ; à chaque bit correspond une fonction binaire. La détermination arbitraire de la taille des mots-codes et le choix des fonctions permettent de contrôler le nombre de problèmes binaires, et donc la robustesse et la complexité de la décomposition. L'évaluation des fonctions binaires associe à chaque individu $\mathbf{x}$ un code qui permet d'affecter $\mathbf{x}$ à la classe dont le mot-code en est le plus proche.

Nous nous sommes intéressés aux différentes manières de combiner les $K(K-1)/2$ classifieurs binaires, dans le cas de l'approche *une contre une*.

2.2 Quelques méthodes de combinaison

Dans le cadre d'une binarisation *une contre une*, il existe plusieurs méthodes de combinaison des classifieurs binaires, dont quelques-unes sont présentées ci-dessous :

Règle du vote [5]. Chaque classifieur prend une décision (vote), l'individu $\mathbf{x}$ étant affecté à la classe $\widehat{\omega}$ remportant le plus de votes.

Recherche des probabilités *a posteriori* consistantes [7, 8, 16]. Soient les probabilités *a posteriori* de chacune des classes $p_i = \mathbb{P}(\omega_i|\mathbf{x})$, les probabilités conditionnelles $\mu_{ij} = \mathbb{P}(\omega_i|\omega_i \textit{ ou } \omega_j, \mathbf{x})$, n_{ij} les effectifs des paires $\{\omega_i, \omega_j\}$; avec ces notations, $\mu_{ij} = p_i/(p_i + p_j)$. Supposons en outre que les classifieurs binaires fournissent des estimations r_{ij} des μ_{ij}.

Rechercher directement les p_i telles que $\sum p_i = 1$, $\mu_{ij} = r_{ij}$ est un problème à $K-1$ inconnues et $(K-1)K/2$ contraintes, qui n'a en général pas de solution exacte ; on peut donc chercher à estimer les p_i de telle sorte que les μ_{ij} soient proches des r_{ij} dans un sens à définir.

Dans [7], il est proposé de déterminer de manière itérative la distribution de probabilité

$\hat{\mathbf{p}}$ minimisant la distance de Kullback-Leibler entre les r_{ij} et les μ_{ij} :

$$\hat{\mathbf{p}} = \arg\min \mathcal{L}(\mathbf{p}), \tag{1}$$

$$\mathcal{L}(\mathbf{p}) = \sum_{i<j} n_{ij} \left(r_{ij} \log \frac{r_{ij}}{\mu_{ij}} + (1 - r_{ij}) \log \frac{1 - r_{ij}}{1 - \mu_{ij}} \right),$$

où $\mathbf{p}$ représente le vecteur des p_i. Des probabilités $\tilde{p_i}$, respectant le même ordonnancement que les $\hat{p_i}$, peuvent également être approchées de manière non-itérative :

$$\tilde{p}_i = \frac{2}{K} \frac{\sum_{j \neq i} r_{ij}}{(K-1)}. \tag{2}$$

Dans [16], deux méthodes d'estimation des probabilités p_i, par résolution de systèmes d'équations, sont proposées :

$$p_i = \sum_{j:j \neq i} \frac{p_i + p_j}{K-1} r_{ij}, \quad \forall i \tag{3}$$

sous les contraintes $\sum_{i=1}^{K} p_i = 1, p_i \geq 0, \forall i$, et :

$$\mathbf{p} = \arg\min_{\mathbf{p}} \frac{1}{2} \sum_{i=1}^{K} \sum_{j:j \neq i} (r_{ji} p_i - r_{ij} p_j)^2 \tag{4}$$

sous les contraintes $\sum_{i=1}^{K} p_i = 1, p_i \geq 0, \forall i$.

Remarquons que dans tous les cas, l'évaluation d'un individu $\mathbf{x}$ requiert une optimisation ou la résolution un système d'équations.

Limitations de la combinaison des probabilités. Chaque classe apparaît dans l'ensemble d'apprentissage de $K-1$ classifieurs sur les $K(K-1)/2$. Lors d'une évaluation, on espère que les classifieurs dont le domaine de validité comprend la classe recherchée, minoritaires, fourniront des informations cohérentes, compensant celles, *a priori* non pertinentes, fournies par les autres.

Dans [8], il est proposé d'entraîner des classifieurs supplémentaires, dits "correcteurs", permettant de déterminer si l'individu évalué $\mathbf{x}$ se trouve dans l'une des deux classes ω_i et ω_j considérées. Ces classifieurs sont entraînés en prenant comme classe positive $\{\omega_i, \omega_j\}$ et comme classe négative $\{\omega_k, k \notin \{i, j\}\}$, et estiment les probabilités $q_{ij} = \hat{\mathbb{P}}(\{\omega_i, \omega_j\}|\mathbf{x})$. Les p_i sont approchées à partir des produits $r_{ij} q_{ij}$, par la procédure non-itérative de l'équation (2).

3 Le Modèle des Croyances Transférables

La nécessité de représenter différentes formes d'imprécision et d'incertitude a conduit, au cours des trente dernières années, à la définition de cadres de représentation autres que le *cadre probabiliste* couramment utilisé, parmi lesquels la *théorie des fonctions de croyance* [12]. Plusieurs déclinaisons de cette dernière, dont le Modèle des Croyances Transférables (MCT) [13], ont été proposées.

3.1 Représentation des connaissances

Tout comme le modèle probabiliste, le Modèle des Croyances Transférables a pour but la modélisation de la connaissance partielle d'un agent rationnel quant à la valeur prise par une variable y. Cette valeur appartient à un ensemble $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$, ou *cadre de discernement*, d'éléments singletons ou *atomes*. Pour un problème de discrimination, y représente la classe d'un individu $\mathbf{x}$ évalué.

La croyance en des éléments de Ω peut ainsi être quantifiée par une *fonction de masse de croyance* $m^\Omega : 2^\Omega \rightarrow [0; 1]$ telle que $\sum_{A \subseteq \Omega} m(A) = 1$, notée m lorsqu'il n'y a pas d'ambiguïté sur le cadre de discernement.

On peut considérer que la valeur de y ne correspond à aucun des éléments composant Ω, et allouer une part de croyance à l'élément $\emptyset$ (hypothèse du monde ouvert). Si l'on suppose au contraire que la valeur de y se trouve dans Ω (hypothèse du monde fermé), on normalise systématiquement les masses m :

$$m^*(A) = \frac{m(A)}{1 - m(\emptyset)}, \quad \forall A \subseteq \Omega, A \neq \emptyset.$$

3.2 Manipulation des connaissances

Combinaison de fonctions de croyances. Deux fonctions de croyance m_1 et m_2 peuvent être combinées, simulant le recoupement de diverses informations lors d'un processus cognitif. Deux opérateurs de combinaison ont été définis : les *sommes* ou *combinaisons conjonctive* et *disjonctive*, symbolisées respectivement par $\textcircled{\cap}$ et $\textcircled{\cup}$:

$$m_1 \textcircled{\cap} m_2(A) = \sum_{X \cap Y = A} m_1(X) m_2(Y), \forall A \subseteq \Omega;$$
$$m_1 \textcircled{\cup} m_2(A) = \sum_{X \cup Y = A} m_1(X) m_2(Y), \forall A \subseteq \Omega.$$

Fonctions de croyance conditionnelles. Tout comme dans le cadre probabiliste, il est possible d'exprimer des croyances *conditionnelles,* valides à condition qu'une hypothèse soit vérifiée. Étant donné une fonction de masse m et une hypothèse $B \subseteq \Omega$, on définit une fonction de masse conditionnelle $m[B](A)$ par :

$$m[B] = m \textcircled{\cap} m_B,$$

où m_B est la fonction de masse catégorique définie par $m_B(B) = 1$.

La part de croyance conditionnelle $m[B](\emptyset)$ représente le degré de conflit entre ces deux masses, donc la quantité de croyance distribuée par la masse m à des éléments focaux incompatibles avec B.

3.3 Utilisation des connaissances : prise d'une décision

Dans le MCT, la prise de décision passe par la transformation des mesures de croyance en mesures de probabilités. La *transformation pignistique* [13] permet de redistribuer aux atomes les parts de croyance allouées aux autres éléments focaux.

La masse attribuée à tout élément focal $A \subseteq \Omega$ est redistribuée à parts égales entre ses éléments, après normalisation :

$$BetP(\omega_k) = \sum_{A:\omega_k \in A} \frac{m^*(A)}{|A|}.$$

4 Utilisation du MCT pour combiner les classifieurs binaires

4.1 Justification de l'approche crédibiliste

Les classifieurs binaires fournissent des croyances conditionnelles. Le problème de la combinaison de classifieurs binaires se présente naturellement dans le cadre du Modèle des Croyances Transférables : en effet, on peut considérer, comme dans le cas probabiliste, que les classifieurs binaires prennent une décision dans un cadre de discernement ou *contexte* restreint $\Omega_{ij} = \{\omega_i, \omega_j\}$. Les valeurs fournies par un classifieur peuvent alors être considérées comme étant des mesures conditionnelles à ce contexte restreint.

Recherche d'une fonction de croyance consistante. Supposons que l'on dispose, pour chaque Ω_{ij}, d'une masse de croyance m_{ij} traduisant notre croyance quant à l'appartenance d'un individu évalué $\mathbf{x}$. En combinant ces différentes mesures, il est possible de déterminer une masse m^Ω en accord avec l'ensemble des informations fournies par les différentes sources, c'est-à-dire dont le conditionnement par rapport à chaque Ω_{ij} donnera une fonction de masse aussi proche que possible de m_{ij}.

Remarquons que chaque classe $\omega_k \in \Omega$ apparaît dans l'ensemble d'apprentissage de $K-1$ classifieurs ; les m_{ij} estimées par ces derniers ne sont donc pas indépendantes, et il est nécessaire d'en tenir compte lors de leur combinaison.

Détermination de la validité d'un classifieur binaire. Dans un contexte Ω_{ij}, il est possible de déterminer si l'individu $\mathbf{x}$ évalué appartient à l'une des deux classes ou non, donc de *valider* les résultats fournis par chaque classifieur binaire évaluant $\mathbf{x}$. Nous avons ainsi déterminé le domaine de validité de chaque classifieur en utilisant une méthode d'estimation de densité. Contrairement à la solution proposée dans [8], nous préférons effectuer cette validation à partir des paires de classes $\{\omega_i, \omega_j\}$ uniquement.

4.2 Formalisation

Traduction des connaissances fournies par les classifieurs en masses de croyance. Considérons le cas le plus répandu où les classifieurs fournissent des résultats sous la forme de probabilités *a posteriori*. On dispose donc :

– d'une masse $m_{ij}^{(2)}$, conditionnelle à Ω_{ij}, quantifiant l'appartenance de $\mathbf{x}$ à ω_i ou ω_j ;
– d'une masse $m_{ij}^{(1)}$, définie sur Ω, quantifiant l'appartenance de $\mathbf{x}$ à $\{\omega_i, \omega_j\}$.

Soit $r_{ij} = \hat{\mathbb{P}}(\omega_i | \omega_i \cup \omega_j, \mathbf{x})$ la probabilité estimée par un classifieur binaire. On pose :

$$\begin{aligned} m_{ij}^{(2)}(\{\omega_i\}) &= r_{ij} \\ m_{ij}^{(2)}(\{\omega_j\}) &= 1 - r_{ij}. \end{aligned}$$

Soit q_{ij} la densité conditionnelle estimée des classes ω_i et ω_j en $\mathbf{x}$:

$$q_{ij} = \min\left(1, \frac{\hat{f}(\mathbf{x}|\{\omega_i, \omega_j\})}{\sup_{\mathbf{x'} \in \{\omega_\mathbf{i}, \omega_\mathbf{j}\}} \hat{f}(\mathbf{x'}|\{\omega_i, \omega_j\})}\right).$$

Une valeur de q_{ij} faible traduit la non-appartenance de $\mathbf{x}$ à Ω_{ij}. Cependant, la densité étant conditionnelle, une forte valeur ne permet pas de conclure à l'appartenance à Ω_{ij}. On pose donc :

$$\begin{aligned} m_{ij}^{(1)}(\Omega) &= q_{ij} \\ m_{ij}^{(1)}(\overline{\Omega_{ij}}) &= 1 - q_{ij}. \end{aligned}$$

$m_{ij}^{(2)}$ et $m_{ij}^{(1)}$ sont définies sur des cadres de discernement différents. Pour les combiner, il faut les ramener sur le même ; on conditionne donc $m_{ij}^{(1)}$ par rapport à Ω_{ij} :

$$\begin{aligned} m_{ij}^{(1)}[\Omega_{ij}](\Omega_{ij}) &= q_{ij} \\ m_{ij}^{(1)}[\Omega_{ij}](\emptyset) &= 1 - q_{ij}. \end{aligned}$$

Cette répartition de $m_{ij}^{(1)}$ traduit le fait que l'estimation de densité permet de quantifier l'appartenance de $\mathbf{x}$ à l'une des classes du contexte Ω_{ij} ou à une autre classe ne lui appartenant pas.

On pose ensuite $m_{ij} = m_{ij}^{(2)} \bigcirc\!\!\!\!\cap\, m_{ij}^{(1)}[\Omega_{ij}]$; on en déduit :

$$\begin{aligned} m_{ij}(\{\omega_i\}) &= r_{ij} q_{ij} \\ m_{ij}(\{\omega_j\}) &= (1 - r_{ij}) q_{ij} \\ m_{ij}(\{\emptyset\}) &= (1 - q_{ij}) \end{aligned}$$

Combinaison des connaissances fournies par les classifieurs. Lors de l'évaluation d'un individu $\mathbf{x}$, on dispose de masses m_{ij} traduisant la connaissance quant à l'appartenance de $\mathbf{x}$ aux différentes classes des contextes Ω_{ij}.

Les informations à l'origine de ces différentes masses n'étant pas indépendantes, on propose de déterminer la fonction de masse $\widehat{m}^\Omega$ la plus consistante possible avec les m_{ij}, d'une manière similaire à celle proposée dans le paragraphe 2.2, équ. (1).

On propose pour cela de résoudre le problème d'optimisation suivant :

$$\widehat{\mathbf{m}}^\Omega = \arg\min_{\mathbf{m}^\Omega} \sum_{j>i} \|\Gamma_{ij} \mathbf{m}^\Omega - m_{ij}\|^2 \qquad (5)$$

sous les contraintes :

$$\begin{cases} 0 \leq \widehat{m}^\Omega(A) \leq 1 & \forall A \subseteq \Omega \\ \sum_A \widehat{m}^\Omega(A) = 1 \end{cases},$$

où $\mathbf{m}^\Omega$ représente un vecteur de masses $m^\Omega(A)$, et Γ_{ij} la matrice de conditionnement par rapport à Ω_{ij} [14].

Évatuation d'un nouvel individu $\mathbf{x}$. L'évaluation d'un nouvel individu $\mathbf{x}$ correspond donc à la séquence d'opérations suivante :

1. évaluation par les différents classifieurs : détermination des masses conditionnelles m_{ij} ;
2. recherche de la masse non conditionnelle la plus consistante avec les m_{ij}, par résolution de (5) ;
3. passage aux probabilités pignistiques pour l'affectation de l'individu à une classe.

Réduction du nombre d'éléments focaux. Pour un problème à K classes, le cadre de discernement comporte 2^K atomes ; lorsque le nombre de classes en compétition devient élevé, la complexité de l'approche crédibiliste la rend difficile à mettre en pratique.

L'analyse des masses de croyance issues de la combinaison permet de constater que la part de croyance allouée à un élément focal décroît lorsque le nombre d'atomes le composant augmente. Cela n'est guère surprenant dans la mesure où les éléments focaux des masses conditionnelles sont des singletons.

Afin de réduire la complexité des opérations effectuées, le nombre d'éléments focaux peut être limité : on contraint alors l'attribution de masse de croyance aux éléments composés d'un nombre d'atomes inférieur à un seuil. Ce faisant, la complexité du problème n'est plus fonction exponentielle mais polynomiale par rapport au nombre de classes.

5 Résultats

5.1 Jeux de données utilisés

Le tableau 1 présente brièvement les jeux de données utilisés (nombre de variables, nombre de classes). La répartition des individus dans les ensembles d'apprentissage et de test est présentée dans les tableaux 2, 3 et 4. Dans le cas du jeu de données *Satimage*, tous les individus n'ont pas été utilisés.

Ces quatre jeux de données sont disponibles sur le site du département d'Apprentissage Automatique de l'Université de Californie à Irvine : `http://www.ics.uci.edu/~mlearn/`.

5.2 Paramétrages

Nous avons employé les Séparateurs à Vaste Marge (SVM) [11, 15].

Les classifieurs binaires sont des 2-SVM (SVM à deux classes), à noyau gaussien ; les valeurs de largeur du noyau et du paramètre contrôlant

Tableau 1 – Description des jeux de données

jeu de données	nb. de variables	nb. de classes
Iris	4	3
Glass	10	11
Satimage	36	6
Vowel	10	11

Tableau 2 – Répartition des individus : Iris et Vowel

données	ensemble d'appr.	ensemble de test
Iris	30 ind. / cl.	20 ind. / cl.
Vowel	48 ind. / cl.	42 ind. / cl.

Tableau 3 – Répartition des individus : Glass

classe	ensemble d'appr.	ensemble de test
classe 1	46	24
classe 2	49	27
classe 3	11	6
classe 4	8	5
classe 5	6	3
classe 6	19	10

la tolérance aux erreurs ont été fixées manuellement. Une méthode permet de transformer les sorties des 2-SVM en probabilités ([9]).

Un 1-SVM (SVM à une classe) à noyau gaussien permet de déterminer le domaine de validité de chaque 2-SVM : le réglage des paramètres utilisés permet d'estimer la densité des classes en un point évalué. La largeur du noyau gaussien peut être déterminée de manière

Tableau 4 – Répartition des individus : Satimage

classe	ensemble d'appr.	ensemble de test
classe 1	700	200
classe 2	300	200
classe 3	600	200
classe 4	250	200
classe 5	300	200
classe 6	700	200

automatique ([4]). Pour les données des jeux "Satimage" et "Vowel", en dimension élevée, cette détermination étant rendue difficile, le paramètre a été fixé manuellement.

Pour réduire la complexité de la méthode, le nombre d'éléments focaux a été restreint pour le traitement des jeux de données réelles : leur cardinal a été limité à 4 atomes.

5.3 Résultats obtenus

Les méthodes suivantes de combinaison des informations binaires ont été comparées :

– TBM : recherche de la masse m consistante avec les $m_{ij}^{(2)}$;
– TBM + 1-SVM : recherche de la masse m consistante avec les m_{ij}, issues de la combinaison des $m_{ij}^{(2)}$ et des $m_{ij}^{(1)}$;
– P. Coupl. : couplage des probabilités conditionnelles [7] ;
– P. Est. 1, P. Est. 2 : deux méthodes d'estimation des probabilités à partir des probabilités conditionnelles [16] ;
– P. Est. Corr. : estimation des probabilités à partir des probabilités conditionnelles "corrigées" [8].

Les taux d'individus bien classés, pour chacune des méthodes testées sur les quatre jeux de données, sont présentés dans le tableau 5.

Tableau 5 – Taux de bonne classification pour les jeux de données traités

Méthode	Iris	Glass	Satim.	Vowel
TBM	96.7%	44 %	81.5 %	47.8 %
TBM + 1-SVM	98.3%	45.3 %	87.1 %	64.3 %
Pairwise Coupling	96.7%	44 %	81.5 %	47.8 %
Proba. Estim. 1	96.7%	42.7 %	82 %	47.4 %
Proba. Estim. 2	96.7%	42.7 %	82.6 %	48.1 %
Proba. Est. Corr.	96.7%	45.3 %	85.3 %	48.3 %

Équivalence des méthodes "TBM" et "Couplage des probabilités". Les résultats permettent de constater une équivalence entre les méthodes "TBM" et "Couplage des probabilités" ([7]). Cela peut s'expliquer par le fait que les masses $m_{ij}^{(2)}$ sont *bayésiennes* : elles ciblent, de même que des probabilités, des éléments focaux singletons ($m_{ij}^{(2)}(A) \neq 0 \Rightarrow |A| = 1$).

Une analyse détaillée permet de constater des différences entre les probabilités obtenues par ces deux méthodes. Cependant, de par la similarité du mode de combinaison de ces informations, les résultats obtenus sont de qualité similaire : on constate que l'ordonnancement des probabilités obtenues dans chacun des cas est le même, les deux méthodes conduisant donc à la même décision.

Amélioration des performances par l'évaluation du domaine de validité des classifieurs binaires. On constate en outre que les résultats obtenus par la méthode "TBM + 1-SVM" mettent en évidence la nette amélioration des résultats par l'utilisation de classifieurs valideurs.

Cette amélioration est plus marquée dans le cas de la méthode crédibiliste (méthode TBM) que dans le cas probabiliste ([8]). Rappelons en outre que la méthode proposée dans [8] pour construire les classifieurs correcteurs, à partir de l'ensemble d'apprentissage total, est très coûteuse.

6 Conclusions et perspectives

Dans cet article, nous avons proposé une méthode permettant de résoudre un problème de discrimination multi-classes, à partir de classifieurs binaires dont on détermine le domaine de validité. Cette méthode, qui s'appuie sur la théorie des fonctions de croyance, généralise une méthode préalablement proposée dans un cadre probabiliste [7]. Les résultats obtenus confirment l'apport de la détermination de la validité des classifieurs binaires, lors de l'évaluation d'un individu : cette précaution peut éviter un biais des résultats obtenus par la combinaison de classifieurs binaires, dont certains peuvent ne pas être pertinents.

Les perspectives quant au développement de cette méthode de combinaison de classifieurs sont nombreuses. Des procédures de rejets en distance et en ambiguïté peuvent être mises en œuvre de manière simple.

Comme dans [7], il est possible de pondérer chaque terme de l'expression (5) (pour chaque contexte Ω_{ij}) par les effectifs des classes correspondantes, pour donner plus de poids aux classifieurs construits à partir de classes comptant beaucoup d'individus.

La méthode présentée offre la possibilité de combiner un nombre indéterminé de classifieurs de natures diverses, fournissant des résultats concernant un nombre indéfini, variable, de classes non nécessairement binaires : cela permet l'utilisation de classifieurs aux qualités complémentaires. Il est probable que la combinaison de classifieurs crédaux [1, 2], fournissant des croyances conditionnelles plus riches, permette d'améliorer la qualité des résultats.

Références

[1] APPRIOU, Alain. « Probabilités et incertitude en fusion de données multi-senseurs », *Revue Scientifique et Technique de la Défense*, 11 :27-40, 1991.

[2] DENŒUX, Thierry. « A neural network classifier based on Dempster-Shafer theory », *IEEE Trans. on Systems, Man and Cybernetics*, 30 :131-150, 2000.

[3] DIETTERICH, Thomas & BAKIRI, Ghulum. « Solving Multiclass Learning Problems via Error-Correcting Output Codes », *Journal of Artificial Intelligence Research*, 2 :263-286, 1995.

[4] DUIN, Robert. « On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions », *IEEE Transactions on Computers*, 1175-1179, 1974.

[5] FRIEDMAN, Jerome. « Another Approach to Polychotomous Classification », *Technical Report, Stanford University*, 1996.

[6] FÜRNKRANZ, Johannes. « Round Robin Classification », *Journal of Machine Learning Research*, 2 :721-747, 2001.

[7] HASTIE, Trevor & TIBSHIRANI, Robert. « Classification by Pairwise Coupling », *Advances in Neural Information Processing Systems, Denver*, 10 :507-513, 1998.

[8] MOREIRA, Miguel & MAYORAZ, Eddy. « Improved Pairwise Coupling Classification with Correcting Classifiers », *European Conference on Machine Learning, Chemnitz*, 10 :160-171, 1998.

[9] PLATT, John. « Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods », *Microsoft Research*, 61-74, 1999.

[10] PLATT, John, CRISTIANINI, Nello & SHAWE-TAYLOR, John. « Large Margin DAGs for Multiclass Classification », *Advances in Neural Information Processing Systems*, 12 :547-553, 2000.

[11] SCHÖLKOPF, Bernhard & SMOLA, Alexander. « Learning with Kernels », *The MIT Press, Cambridge*, 2002.

[12] SHAFER, Glenn. « A Mathematical Theory of Evidence », *Princeton University Press, Princeton*, 1976.

[13] SMETS, Philippe & KENNES, Robert. « The Transferable Belief Model », *Artificial Intelligence*, 66 :191-234, 1994.

[14] SMETS, Philippe. « The Application of the Matrix Calculus to Belief Functions », *International Journal of Approximate Reasoning*, 31 :1-30, 2002.

[15] VAPNIK, Vladimir. « The Nature of Statistical Learning Theory », *Sprinter-Verlag, New York*, 1996.

[16] WU, Ting-Fan, LIN, Chih-Jen & WENG, Ruby. « Probability Estimates for Multi-class Classification by Pairwise Coupling », *Advances in Neural Information Processing Systems, Vancouver*, 16, 2003.