Reconnaissance de symboles graphiques par le biais de l'intégrale de Choquet

Laurent Wendling, Jan Rendek, Pascal Matsakis

Reconnaissance de symboles graphiques par le biais de l'intégrale de Choquet

Laurent Wendling[1], Jan Rendek[1], Pascal Matsakis[2]

LORIA - ESIAL[1]
Université Henri Poincaré
54506 Vandœuvre-lès-Nancy
wendling,rendek@loria.fr

CIS Dpt[2]
Université de Guelph
Guelph, ON N1G 2W1 Canada
matsakis@cis.uoguelph.ca

Résumé : *nous présentons dans cet article trois modèles pour extraire un sous-ensemble de règles de décision et leur agrégation en une règle de décision unique par le biais de l'intégrale de Choquet. Nous nous intéressons à des applications où l'on possède peu d'échantillons représentatifs par classe. Les approches que nous avons construites sont de type global, par classe ou bi-classes. Des applications sur des données réelles attestent de la robustesse de nos méthodes et leur adaptabilité.*

Mots-clés : reconnaissance de symboles, capacités, intégrale de Choquet, décision multicritères.

1 Introduction

Un système de reconnaissance des formes peut être grossièrement décomposé en trois étapes successives [JAI 00, SME 00]. Premièrement les objets sont extraits du fond. Cette étape de segmentation est souvent conditionnée par une connaissance *a priori* des documents à traiter. Puis une représentation est construite à partir des zones extraites. Celle-ci est, normalement, en concordance avec une mesure de (dis)similarité entre deux formes. Généralement, la structure correspond soit à une description sous la forme d'un vecteur caractéristique représentant un ensemble de mesures calculées sur les objets soit à une description symbolique correspondant à un ensemble de primitives simples reliées entre elles par des lois de composition. La troisième étape est axée sur la construction d'une règle de décision en conformité avec l'étape de représentation. Cette phase peut dépendre d'une connaissance experte sur les éléments à rechercher ou d'un apprentissage sur des échantillons représentatifs.

Les états de l'art sur les techniques de représentation des objets et leur classification ne permettent pas de conclure sur l'existence d'un ensemble de méthodes génériques fonctionnant efficacement sur n'importe quel type de documents [CHH 98, JAI 00, COR 00, LLA 02]. Pour pallier ce problème, de nombreuses techniques conditionnées par les domaines d'applications ont été développées. Pour un problème donné, le choix du type de représentation ou du modèle de décision revient à tester au mieux plusieurs combinaisons de méthodes. Il n'est pas évident pour un utilisateur de décider quelle est la technique à sa disposition la plus adéquate en fonction de l'application courante. Une des solutions revient à combiner plusieurs règles de décision, fondées sur des représentations et/ou des méthodes de classification variables au lieu d'en choisir une unique. L'idée est d'obtenir une décision finale plus robuste prenant en compte les aspects discriminants des règles de décision. Cette approche est intéressante dans le cas où les données sont insuffisantes pour déterminer de manière précise la meilleure méthode par des tests ou pour construire des règles de décision robustes par apprentissage.

De nombreux systèmes de combinaison ont été proposés et comparés dans la littérature [KIT 98, RUT 00, KUN 03, STE 05]. Nous nous focalisons ici sur des applications où l'on possède peu de données d'apprentissage. Ce qui est fréquemment le cas, dans le cadre de la reconnaissance de symboles techniques, sauf si ceux-ci sont réalisés à la main ou si l'on effectue des traitements artificiels de dégradation sur les données.

Nous nous sommes intéressés à l'intégrale de Choquet qui permet de prendre en compte de manière efficace les interactions entre règles de décision tout en offrant un modèle de décision robuste en présence de peu de données d'apprentissage pouvant être même inconsistantes. Les intégrales floues, et l'intégrale de Choquet en particulier, ont été utilisées avec succès comme opérateur de fusion dans de nombreuses applications [TAH 90, CHO 95, MIK 99] incluant le « CBIR » [CHO 04] et la reconnaissance de la parole [CHA 03].

Dans cet article nous présentons de manière succincte un ensemble de stratégies de sélection que nous avons définies (approche globale, par classe et bi-classes). Le lecteur intéressé pourra se référer aux références suivantes [REN 07, REN 08, SCH 08] pour avoir une description plus théorique de nos méthodes ainsi qu'une étude expérimentale plus conséquente (sur des données industrielles et sur des bases UCI) avec une comparaison avec d'autres approches (médiane, sélection par SVM, C4.5, SBFS,...).

2 Agrégation de règles de décision

2.1 Fusion de règles de décision

Nous considérons m classes, $C_1, \ldots, C_m$, et un ensemble X de n règles de décision (RDs) $X = \{D_1, ..., Dn\}$. Par règle de décision, nous associons ici un descripteur et un rapport de similarité (pour que les entrées soit normalisées entre 0 et 1). Soit une observation x^0, l'objectif est de calculer pour chaque règle de décision, le degré de confiance ϕ^i que l'on a

dans l'affirmation « selon D_j, x_0 appartient à la classe C_i ». Ainsi, la confiance globale associée à l'affirmation « x_0 appartient à la classe C_i » noté $\Phi(C_i|x^o)$, devient :

$$\Phi(C_i|x^o) = \mathcal{H}(\phi_1^i, \ldots, \phi_n^i)$$

Finalement, x^o est attribué à la classe la plus proche.

$$label(x^o) = \underset{i=1}{\overset{m}{\arg}}\ max\ \Phi(C_i|x^o)$$

L'étape suivante revient à combiner toutes ces RDs en choisissant un opérateur d'agrégation $\mathcal{H}$ approprié.

2.2 Intégrale de Choquet

Nous avons porté notre choix sur l'intégrale de Choquet qui généralise de nombreux opérateurs comme les OWA, la somme arithmétique, la somme pondérée, la médiane, les statistiques d'ordre n [GRA 95b]... L'intégrale de Choquet a été introduite en théorie des capacités [CHO 53]. Nous nous limiterons ici aux notions utiles pour notre étude et le lecteur intéressé par le concept pourra se référer aux articles complets [GRA 94, MAR 02, MUR 91]. Le calcul de l'intégrale de Choquet requiert la définition d'une capacité ou mesure floue. Une capacité sur X, est une fonction $\mu : 2^X \longrightarrow [0, 1]$ qui vérifie : $\mu(\emptyset) = 0$, $\mu(X) = 1$ et $\mu(A) \leqslant \mu(B)$ si $A \subseteq B$ (monotonie).

Soit μ une capacité sur X. L'intégrale de Choquet de $\vec{\phi} = [\phi_1, \ldots, \phi_n]^t$ notée $\mathcal{C}_\mu(\vec{\phi})$, est définie par :

$$\mathcal{C}_\mu(\vec{\phi}) = \sum_{j=1}^{n} \phi_{(j)}[\mu(A_{(j)}) - \mu(A_{(j+1)})]$$

avec $A_{(j)} = \{(j), \ldots, (n)\}$ correspond aux $[j..n]$ où $(.)$ est une permutation, telle que $(i) \leq (j) \Rightarrow \phi_{(i)} \leq \phi_{(j)}$.

2.3 Phase d'apprentissage

Il existe plusieurs méthodes pour déterminer la mesure floue la plus adéquate pour une application donnée [GRA 94]. Des algorithmes ont été développés pour intégrer au mieux les problèmes d'initialisation des treillis associés à la mesure floue (matrices mal conditionnées, convergence, temps de calcul). L'objectif est de trouver une approximation de la mesure floue en minimisant un critère d'erreur.

À notre connaissance, l'algorithme fournissant la meilleure approximation et le mieux adapté à notre problème, est celui proposé par M. Grabisch [GRA 95a]. Il part du principe qu'en l'absence d'information le modèle d'agrégation le plus raisonnable est la moyenne arithmétique. À partir d'un ensemble d'alternatives - ensemble d'échantillons caractéristiques et valeur de l'intégrale attendue - cet algorithme apprend la mesure floue en se fondant sur un principe de descente de gradient avec contraintes. L'idée est de minimiser l'erreur, au sens des moindres carrés, entre la valeur de l'intégrale calculée sur la mesure floue associée et la sortie attendue.

Par ailleurs, cette approche permet de conserver une cohérence même si des données sont manquantes pour traiter tous les chemins du treillis. Enfin, l'apprentissage est très rapide.

Schématiquement, le calcul de l'intégrale de Choquet est associé à un chemin dans le treillis. La figure 1 présente le treillis correspondant à 4 règles de décision. Soit $\phi^1 \leq \phi^4 \leq \phi^2 \leq \phi^3$, la valeur de l'intégrale de Choquet associée à $\Phi = (\phi^1, \phi^2, \phi^3, \phi^4)$ n'est calculée qu'à partir des valeurs du chemin $(\mu_0, \mu_3, \mu_{23}, \mu_{234}, \mu_{2341})$.

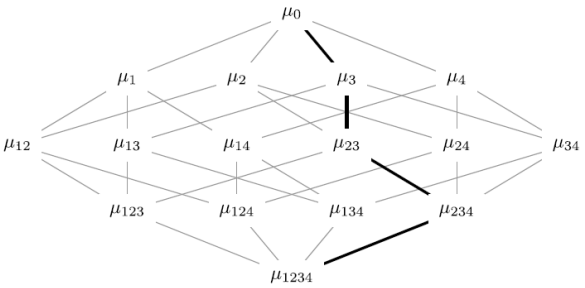

FIG. 1 – Exemple de chemin pour un treillis pour 4 RDs.

2.4 Calcul d'indices

Dès que la mesure floue est entraînée, il est possible d'interpréter la contribution de chaque règle de décision dans la décision finale. Plusieurs indices peuvent être extraits à partir de la mesure floue pour mieux analyser le comportement et l'influence des règles de décision.

Indice d'importance. L'indice d'importance est fondé sur la définition proposée par Shapley dans le cadre de la théorie des jeux [SHA 53] et replacé dans le contexte des mesures floues par Murofushi et Soneda [MUR 93]. Son expression est la suivante :

$$\sigma(\mu, i) = \frac{1}{n} \sum_{t=0}^{n-1} \frac{1}{\binom{n-1}{t}} \sum_{\substack{T \subseteq X \setminus i \\ |T|=t}} [\mu(T \cup i) - \mu(T)]$$

La valeur de Shapley peut être interprétée comme la valeur moyenne pondérée de la contribution $\mu(T \cup i) - \mu(T)$ de la règle de décision i parmi toutes les combinaisons. Une propriété intéressante est que la somme de toutes les valeurs relatives à toutes les règles de décision est égale à 1. En d'autres termes : $\Sigma_{i=1}^{n} \sigma(\mu, i) = 1$. Ainsi une règle de décision avec un degré d'importance plus petit que $1/n$ peut être interprétée comme une importance faible pour la décision finale.

Indice d'interaction. L'indice d'interaction introduit par Murofushi et Soneda représente le degré d'interaction positif ou négatif entre deux règles de décision. Si la mesure floue a un comportement monotone alors des RDs interagissent. La valeur de l'interaction entre i et j, conditionnée par la présence des éléments de la combinaison $T \subseteq X \setminus ij$ est donnée par :

$$(\Delta_{ij}\mu)(T) = \mu(T \cup ij) + \mu(T) - \mu(T \cup i) - \mu(T \cup j)$$

En étendant ce critère sur tous les sous-ensembles de $T \subseteq X \setminus ij$ on obtient une évaluation de l'interaction entre les RDs

i et j, comme suit :

$$I(\mu, ij) = \sum_{T \subseteq X \setminus ij} \frac{(n-t-2)!t!}{(n-1)!} (\Delta_{ij}\mu)(T)$$

Une interaction positive pour deux RDs i et j signifie que l'importance d'une règle de décision est renforcée par la seconde. En d'autres termes, les deux RDs sont complémentaires et leur combinaison va en améliorant la décision finale. Une interaction négative signifie que les RDs sont antagonistes et que leur combinaison réduit l'impact de la décision finale.

3 Méthode globale

3.1 Données d'apprentissage

Le but de cette étape est de déterminer les données d'apprentissage les plus adéquates pour prendre en compte la confusion pouvant exister entre les RDs.

Nous considérons ici une base d'apprentissage comportant un nombre faible d'éléments par classe et nous déterminons les matrices de confusion associées à chaque règle de manière indépendante. À partir de celles-ci nous déterminons le jeux d'échantillon d'apprentissage en considérant les valeurs de décisions moyennes. Pour chacun de ces échantillons une valeur, relative à l'appartenance à la classe ciblée, doit être assignée.

Une matrice de confusion globale est ensuite construite pour déterminer le coté discriminant de l'agrégation des RDs pour l'ensemble des classes. À partir de l'estimation de cette confusion, nous assignons une valeur de sortie telle que, plus la valeur de confusion est élevée, plus la valeur cible associée à l'échantillon est proche de zéro (cf. [REN 08] pour une description plus détaillée de l'approche).

3.2 Extraction automatique d'un sous-ensemble de règles de décision

Dès que le treillis est appris, nous analysons les performances individuelles de chaque règle de décision à partir des indices présentés dans le paragraphe 2.4. Notre objectif est de rechercher les RDs qui sont les moins importantes et qui interagissent le moins avec les autres. Nous partons du principe que de telles règles peuvent dégrader la décision finale. Nous avons réalisé un schéma de sélection en deux étapes pour retirer de telles RDs. Dans un premier temps, la valeur de Shapley est multipliée par le nombre de RDs n. Ainsi une règle de décision ayant une valeur de Shapley supérieure à 1 est jugée plus importante que la moyenne. Nous sélectionnons l'ensemble des règles S_L supposées les moins utiles pour l'application en cours :

$$S_L = \{k \mid n \cdot \sigma(\mu, k) < 1\}$$

Puis nous extrayons de S_L le sous-ensemble de règles de décision ayant le plus de synergies positives avec les autres. Pour chaque règle SL_i, nous déterminons leur interaction moyenne avec les autres pour donner une estimation globale de son influence. Finalement, le sous-ensemble de règles dont on doit tester la suppression, noté MS_L, est composé des règles de S_L ayant une valeur d'interaction plus petite que l'interaction moyenne associée à toutes les autres RDs de S_L :

$$MS_L = \{k \mid \sum_{j=1,n} I(\mu, kj) < m\}_{k \in S_L}$$

Avec la valeur moyenne d'interaction donnée par :

$$m = \frac{1}{|S_L|} \sum_{k \in S_L} \sum_{j=1,n} I(\mu, kj)$$

L'algorithme général est le suivant :

Étape 0 %Initialisation%
- Calcul des poids à partir des données d'apprentissage
- Initialisation de la capacité à la somme pondérée associée

Étape 1 %Définition des valeurs de sortie%
- Calcul de la matrice de confusion moyenne normalisée
- Affectation d'une valeur de sortie pour chaque échantillon

Étape 2 %Apprentissage de la capacité%
- Apprentissage avec l'algorithme de descente de gradient

Étape 3 %Sélection des RDs%
- Calcul des indices d'importance et d'interaction
- Suppression des règles faibles issues de ces indices

Étape 4 %Test de fin d'itération%
- **Si** la reconnaissance a été améliorée **alors** retour en 1 **sinon** « rollback » et stop

Nous avons montré expérimentalement que se limiter à l'indice de Shapley (de S_L) le plus faible était trop restrictif dans de nombreux cas. Par ailleurs nous avons testé toutes les autres combinaisons sur plusieurs bases (UCI, symboles et défauts industriels) et montré que la solution trouvée par notre approche était optimale sauf dans un cas (base industrielle très bruitée et inconsistante) où toutes les approches de sélections se sont arrêtées à 6 RDs sur 9 alors que le sous ensemble « optimal » était composé uniquement de deux règles. Naturellement même si les résultats obtenus sont probants, cette considération ne peut être qu'expérimentale et des tests plus conséquents, à partir de données synthétiques par exemple, seraient nécessaire pour donner les limites réelles des modèles.

3.3 Exemple d'application

Nous avons défini une petite base (cf. figure 2), composée de capitales extraites sur des lettrines fournies gracieusement par le Centre d'études Supérieures de la Renaissance (voir http ://l3iexp.univ-lr.fr/madonne/ pour avoir une idée du projet et des différents corpus).

Nous avons calculé plusieurs règles de décision (moments de Zernike, signatures, GFD, ...) que nous avons ensuite agrégées. Deux modèles ont été utilisés : $Ch1$ fondé sur une somme arithmétique et $CH2$ défini à partir d'une somme pondérée (cf. figure 3). Cet exemple est intéressant car uniquement un descripteur sort du lot et les méthodes d'agrégation classiques (y compris le vote) fournissent des résultats plus faibles que les moments de Zernike. La méthode de sélection permet d'endiguer ce phénomène de « noyade » après

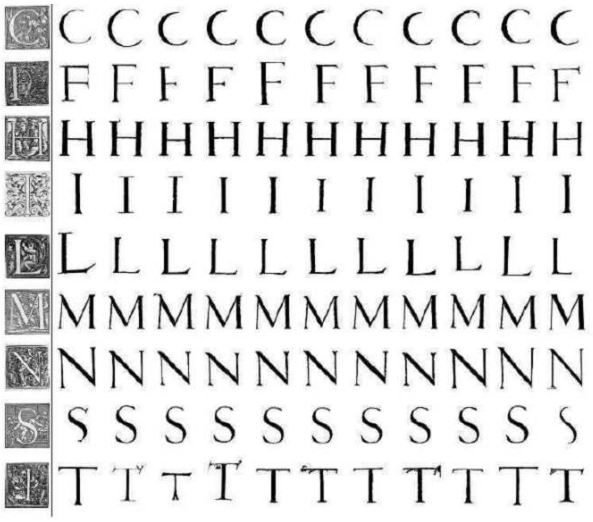

FIG. 2 – Échantillons (base « Madonne »).

la suppression de 4 règles de décision qui ont permis de passer de 71,6% à 89,3% de reconnaissance pour $Ch2$.

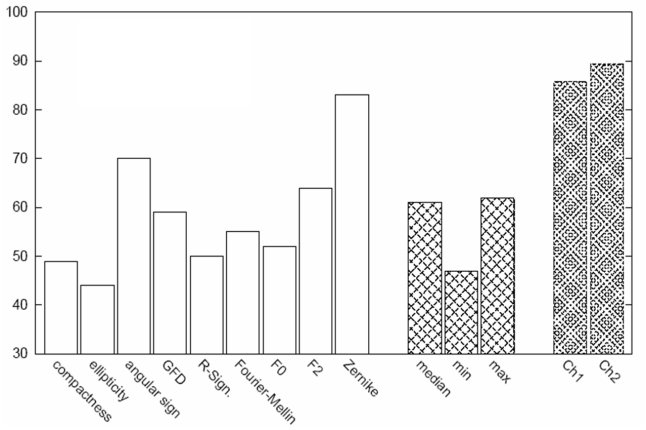

FIG. 3 – Resultats (base Madonne).

4 Modèles par classe

4.1 Une capacité par classe

Nous considérons ici une capacité par classe. L'objectif est classique car il s'agit d'apprendre à reconnaître une classe parmi d'autres. Les treillis sont initialisés à la somme arithmétique et l'apprentissage est réalisé de manière alternative sur une partie de la base (valeur de règles de décision calculées entre éléments de la classe considérée puis par rapport aux autres) pour éviter les problèmes de « pédagogie ».

Pour les techniques utilisant une mesure floue différente par classe, la sortie optimale minimisant l'erreur quadratique est connue dans le cas de deux classes [GRA 95a] :

$$E^2 = \sum_i \sum_j |\Psi(\Delta\Phi_{12}(X_k^j)) - 1|^2$$

avec X_k^j le $k^{\text{ème}}$ échantillon d'apprentissage de la classe j, Ψ une fonction sigmoïde. La valeur de $\Delta\Phi_{12}$ correspond à :

$$\Delta\Phi_{12} = \Phi_{\mu_1}(C_1|X_k^j) - \Phi_{\mu_2}(C_2|X_k^j)$$

Le calcul de l'erreur à propager dans l'algorithme de descente du gradient, dans un cas de deux classes, est donné par :

$$e = \Psi(\Delta\Phi_{12}(X_k^j)) - 1)$$

Nous avons modifié le critère pour traiter le cas N-classes :

$$\Delta\Phi_{qr_{r\in X-\{q\}}} = \Phi_{\mu_q}(C_q|X_k^j) - \bigotimes_{\{r\in X-\{q\}\}} \Phi_{\mu_r}(C_r|X_k^j)$$

avec q et r deux classes. L'opérateur $\bigotimes$ peut être une médiane, un $\min$... Dans notre contexte le $\max$ a été utilisé pour éloigner les classes les plus ambiguës et ainsi favoriser un comportement discriminant. Une valeur médiane est intéressante pour conserver des résultats acceptables même si certaines sources peuvent être contradictoires. Nous avons étudié le signe de $\Delta\Phi_{qr_{r\in X-\{q\}}}$ pour initialiser l'erreur à propager dans le treillis associé à la capacité :

$$sign(\Delta\Phi_{qr_{r\in X-\{q\}}}) = \begin{cases} +e = 1./f(\Delta\Phi_{qr_{r\in X-\{q\}}}) \\ -0,\ e = 0 \text{ pour } q,\ 1 \text{ autres.} \end{cases}$$

avec f une fonction croissante.

Chaque classe $C_1, \ldots, C_n$ est associée à une capacité. L'application de l'algorithme d'apprentissage, suivie de l'extraction des règles de décision les plus pertinentes, forme une itération de notre algorithme. Cette étape est proche de celle employée pour la méthode globale mais concerne un ensemble de classes dont les valeurs de Choquet, pour une observation donnée, sont ensuite dirigées vers un $argmax$ pour sélectionner la classe jugée la plus probable.

Comme il n'est pas évident d'évaluer chaque combinaison de capacités, nous les avons considérées ici comme indépendantes. Un algorithme glouton a ensuite été utilisé pour garantir une extraction continue des règles de décision jugées les plus faibles pour décrire les classes. À chaque itération, le descripteur le plus faible, au sens des indices, est enlevé et ainsi de suite tant qu'un gain (taux de reconnaissance, « ranking », ...) est apporté. L'algorithme général est présenté ci-dessous :

% Initialisation %
Pour Chaque C_i
– Apprendre μ_i
– Extraire le descripteur le plus faible
Fin Pour
% Principal %
TQ La minimisation est possible
– Remplacer l'ancienne capacité par la nouvelle
– Évaluer le gain (minimiser une fonction coût)
– Conserver la « meilleure » nouvelle capacité
– Extraire les caractéristiques de cette capacité
Fin TQ

4.2 Modèle bi-classes

Nous avons défini un modèle, dit bi-classes, qui rajoute un niveau hiérarchique en considérant des treillis limités à la distinction de deux classes et en les combinant. Les capacités ne sont apprises que pour comparer des symboles deux à deux (par exemple : un cercle et un triangle) et une

agrégation finale permet d'associer toutes ces capacités. En d'autres termes, le problème est retranscrit ici en associant à chaque couple de classes i et j une capacité $\mathcal{C}_{\mu_{ij}}$ apprise, comme précédemment, en fonction des règles de décision. Ainsi nous obtenons un ensemble de capacités spécialisées dans la reconnaissance d'un symbole par rapport à un autre. Si on considère une classe i, l'ensemble des capacités associées correspond à : $V^i = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{im}]$. En généralisant à l'ensemble des classes on obtient :

$$M = \begin{pmatrix} - & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & - & \cdots & \mu_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1} & \mu_{m2} & \cdots & - \end{pmatrix}$$

Considérons un symbole s. L'objectif est de définir sa classe d'appartenance. Pour chaque classe i, nous considérons chaque capacité μ_{ij} comme étant une nouvelle règle de décision. Nous déterminons, par apprentissage, la nouvelle capacité μ_{V_i} associée à i. Le nouveau jeu d'apprentissage est fourni par les valeurs de décision obtenues par chaque capacité en réinjectant les valeurs des règles de décision initiales calculées sur les échantillons.

La phase d'apprentissage par descente de gradient se fait de manière similaire au modèle à une capacité par classe. Enfin, l'$argmax$ est recherché pour déterminer la classe cible de s. Par ailleurs, pour limiter le nombre de capacités dans M et se focaliser sur les plus pertinentes, l'algorithme glouton défini précédemment a été appliqué pour retirer les μ_{ij} les moins pertinentes.

4.3 Exemple d'application

Nous avons testé ces modèles sur une base de 10 symboles fournie par le CVC Barcelone. Elle comporte 10 classes de 300 échantillons dessinés par 10 personnes avec un stylo « Anoto » (cf. exemples donnés figure 4).

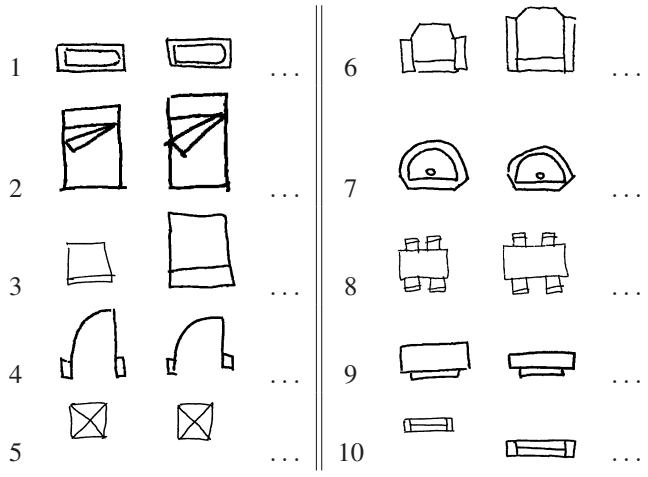

FIG. 4 – Exemples de symboles (base anoto).

La figure 5 présente le résultat de l'application de nos deux approches comparées à des règles de décision fondées sur des descripteurs classiques (ART, moment de Zernike, Yang, ...). Des gains de l'ordre de 8% (bi-classes) ou 9% (une capacité par classe) par rapport à la meilleure règle de décision ont été obtenus sur cette base.

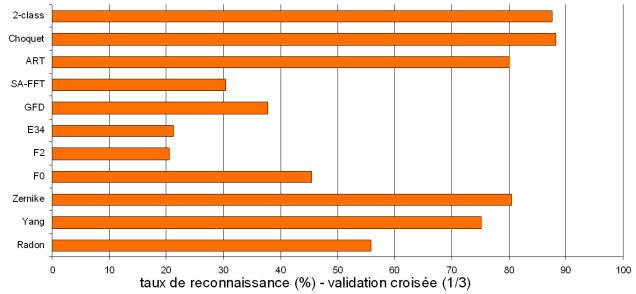

FIG. 5 – Resultats (base anoto).

5 Discussion sur les modèles

Le modèle global fournit des résultats très intéressants lorsque le système présente des mesures erronées ou ambiguës comme dans le cas des procédés industriels. Il permet de dresser une carte globale des descripteurs les plus pertinents (au sens des indices). Il est aussi intéressant de signaler que suivant le type d'application et la qualité des descripteurs, nous obtenons une réduction de l'ordre de 40% en moyenne (bases de symboles, de formes et industrielles) du nombre de règles de décision.

Dans des exemples classiques, le modèle par classe permet de dresser une cartographie des meilleurs descripteurs par classe et fournit des résultats plus performants que l'approche globale lorsque l'on augmente le nombre de classes à traiter. Par contre, le temps de calcul de l'apprentissage est beaucoup plus élevé et on peut se poser la question de la validité de manipuler des descripteurs statistiques différents par classe en fonction de la distribution des valeurs obtenues par les règles de décision. Cette remarque est valable pour la plupart des méthodes de sélection.

Le modèle bi-classes possède naturellement le temps d'apprentissage le plus élevé. Par contre, il fournit d'excellents résultats sur de petites bases (style Sharvit) et permet de mieux différencier les bases hétérogènes mélangeant symboles graphiques, lettres et formes par exemple. Il permet aussi de traiter des classes contradictoires sans perte conséquente de la robustesse. Enfin, cette décomposition du problème devrait permettre d'intégrer facilement des « experts » humains qui valident de manière sûre une ou plusieurs règles de décision (moins fiables au sens des indices de Shapley par exemple) pour améliorer la reconnaissance.

Ces modèles sont relativement faciles à implémenter et très rapides lors de la phase de décision (tri des valeurs et parcours d'un chemin du (ou des) treillis). Par ailleurs, ils ont été comparés à des méthodes classiques d'agrégation et de sélection et montré leur efficacité sur de nombreuses bases où peu d'échantillons d'apprentissage sont considérés. Il est évident que d'autres méthodes statistiques seront plus adaptées et plus performantes lorsque le nombre de données d'apprentissage sera plus conséquent (et les distributions plus gaussiennes...). Un point complexe (NP) serait de déterminer automatiquement à partir de quelle échelle de telles approches ne sont plus efficaces.

6 Conclusion et perspectives

Nous avons présenté dans cet article trois modèles fondés sur l'intégrale de Choquet pour extraire les règles de décision les plus adéquates, suivant des critères d'importance et d'interaction, en fonction de l'application en cours. Actuellement, nous étudions la possibilité de combiner nos modèles avec des mécanismes de retour de pertinence pour affiner nos modèles d'apprentissage. Un deuxième axe concerne l'extraction de règles, en fonction d'une analyse d'indices calculés sur le treillis, pour générer une description sémantique relative au domaine étudié. Enfin, les spécificités de l'intégrale de Choquet permettent d'envisager de mixer connaissances expertes et données numériques lors de la phase de décision.

Références

[CHA 03] CHANG S., GREENBERG S., Application of fuzzy-integration-based multiple-information aggregation in automatic speech recognition, *Proceedings of the IEEE Conference on Fuzzy Integration Processing, Beijing*, 2003.

[CHH 98] CHHABRA A. K., Graphic Symbol Recognition : An Overview, TOMBRE K., CHHABRA A. K., Eds., *Graphics Recognition—Algorithms and Systems*, vol. 1389 de *Lecture Notes in Computer Science*, pp. 68–79, Springer-Verlag, avril 1998.

[CHO 53] CHOQUET G., Theory of capacities, *Annales de l'Institut Fourier*, vol. 5, 1953, pp. 131–295.

[CHO 95] CHO S. B., KIM J. H., Combining Multiple Neural Networks by Fuzzy Integral for Robust Classification, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, n° 2, 1995, pp. 380–384.

[CHO 04] CHOI Y. S., KIM D., Relevance Feedback For Content-Based Image Retrieval Using The Choquet Integral, *IEEE trans. on Knowledge and Data Engineering*, vol. 16, n° 10, 2004, pp. 1185–1199.

[COR 00] CORDELLA L. P., VENTO M., Symbol recognition in documents : a collection of techniques ?, *International Journal on Document Analysis and Recognition*, vol. 3, n° 2, 2000, pp. 73–88.

[GRA 94] GRABISCH M., NICOLAS J. M., Classification by fuzzy integral - performance and tests, *Fuzzy Sets and Systems, Special Issue on Pattern Recognition*, vol. 65, 1994, pp. 255–271.

[GRA 95a] GRABISCH M., A new algorithm for identifying fuzzy measures and its application to pattern recognition, *4th IEEE International Conference on Fuzzy Systems*, 1995, pp. 145–150.

[GRA 95b] GRABISCH M., The application of fuzzy integral in multicriteria decision making, *European journal of operational research*, vol. 89, 1995, pp. 445–456.

[JAI 00] JAIN A. K., DUIN R. P. W., , MAO J., Statistical Pattern Recognition : A Review, *IEEE Transactions on PAMI*, vol. 22, n° 1, 2000, pp. 4–37.

[KIT 98] KITTLER J., HATEF M., DUIN R., , MATAS J., On combining classifiers, *IEEE Transactions on PAMI*, vol. 20, n° 3, 1998, pp. 226–239.

[KUN 03] KUNCHEVA L., WHITAKER C., Measures of diversity in classifier ensembles, *Machine Learning*, vol. 51, 2003, pp. 181–207.

[LLA 02] LLADÓS J., VALVENY E., SÁNCHEZ G., , MARTÍ E., Symbol Recognition : Current Advances and Perspectives, BLOSTEIN D., KWON Y.-B., Eds., *Graphics Recognition – Algorithms and Applications*, vol. 2390 de *Lecture Notes in Computer Science*, pp. 104–127, Springer-Verlag, 2002.

[MAR 02] MARICHAL J.-L., Aggregation of interacting criteria by means of the discrete Choquet integral, *Aggregation operators : new trends and applications*, pp. 224–244, Physica-Verlag GmbH, Heidelberg, Germany, 2002.

[MIK 99] MIKENINA L., ZIMMERMANN H.-J., Improved feature selection and classification by the 2-additive fuzzy measure, *Fuzzy Sets and Systems*, vol. 107, 1999, pp. 197-218.

[MUR 91] MUROFUSHI T., SUGENO M., A theory of fuzzy measures : representations, the Choquet integral, and null sets, *Journal of Mathematical Analysis and Applications*, vol. 159, 1991, pp. 532–549.

[MUR 93] MUROFUSHI T., SONEDA S., Techniques for reading fuzzy measures(III) : interaction index, *Proceedings of the 9th Fuzzy System Symposium, Sapporo, Japan,*, mai 1993, pp. 693–696, In Japanese.

[REN 07] RENDEK J., WENDLING L., Symbol Recognition using a 2-class hierarchical model of Choquet Integrals, *Proceedings of 9th International Conference on Document Analysis and Recognition*, 2007, pp. 207–232.

[REN 08] RENDEK J., WENDLING L., On Determining Suitable Subsets of Decision Rules using Choquet Integral, *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 22, n° 2, 2008, pp. 207–232.

[RUT 00] RUTA D., GABRYS B., An Overview of Classifier Fusion Methods, *Computing and Information System*, vol. 7, n° 1–10, 2000.

[SCH 08] SCHMITT E., BOMBARDIER V., , WENDLING L., Improving Inference Engine by Extracting Suitable Features from Capacities with respect to the Choquet Integral, *IEEE Transactions on Systems, Man, and Cybernetics – Part B : Cybernetics*, vol. à paraitre, 2008.

[SHA 53] SHAPLEY L., A value for n-person games, KHUN H., TUCKER A., Eds., *Contributions to the Theory of Games, Annals of Mathematics Studies*, Princeton University Press, 1953, pp. 307–317.

[SME 00] SMEULDERS A. W. M., WORRING M., SANTINI S., GUPTA A., , JAIN R., Content-Based Image Retrieval at the End of the Early Years, *IEEE Transactions on PAMI*, vol. 22, n° 12, 2000, pp. 1349–1380.

[STE 05] STEJIC Z., TAKAMA Y., , HIROTA K., Mathematical aggregation operators in image retrieval : effect on retrieval performance and role in relevance feedback, *Signal Processing*, vol. 85, n° 2, 2005, pp. 297–324.

[TAH 90] TAHANI H., KELLER J. M., Information Fusion in Computer Vision Using the Fuzzy Integral, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, n° 3, 1990.