

Extension de la méthode FRIFS pour l'automatisation de l'extraction de paramètres

EMMANUEL SCHMITT¹, VINCENT BOMBARDIER¹, LAURENT WENDLING²

¹CRAN – UMR 7039 Université Henri Poincaré
Campus scientifique, BP 239, 54506 Vandœuvre – France

²LORIA – UMR 7053 Université Henri Poincaré
Campus scientifique, BP 239, 54506 Vandœuvre – France

¹{Schmitt,bombardier}@cran.uhp-nancy.fr, ²wendling@loria.fr

Résumé - Une méthode itérative de sélection de paramètres (FRIFS) a été présentée dans un contexte de reconnaissance de défauts sur des tissus techniques. Elle combine une méthode générale de sélection de paramètres fondée sur l'intégrale de Choquet et un classificateur par règles linguistiques floues. Dans cet article nous proposons une extension de cette méthode pour traiter un nombre croissant de paramètres. L'étude expérimentale, portant sur une problématique industriel d'identification de défauts du bois, montre la pertinence et l'efficacité de notre approche de sélection automatique par rapport à des méthodes classiques.

Abstract - An iterative method to select suitable features for pattern recognition context has been proposed (FRIFS). It combines a global feature selection method based on the Choquet integral and a fuzzy linguistic rule classifier. In this paper, an enhancement of this method is presented which suppress the empirical selection of feature to be eliminated. The experimental study, made in a wood defect recognition context, shows the efficiency of the automatic selection.

1 Introduction

L'utilisation d'une méthode de sélection de paramètres s'intègre dans une problématique de reconnaissance de formes. Le domaine d'application concerne le contrôle qualité dans les industries du bois. L'objectif du système de vision est d'identifier les défauts du bois en continu pendant la production.

Ce contexte industriel induit de nombreuses contraintes, comme la nécessité de travailler avec des lots de données réduits (on peut ne posséder qu'un ou deux échantillons pour une classe de défauts). En outre, les défauts peuvent être « intrinsèquement flous », c'est à dire qu'il n'existe pas de frontière franche entre la zone saine et la zone « défectueuse » du bois. L'étape de segmentation fournit une R.O.I. stricte de la région « défectueuse », les paramètres calculés sont ainsi précis mais incertains. La méthode de reconnaissance de formes utilisée doit donc prendre en compte cette spécificité.

Une autre difficulté tient dans le respect des contraintes « temps réel » du système de production. Ainsi, la complexité du modèle de reconnaissance de formes doit être aussi réduite que possible. Ce travail se place donc dans un domaine à « échelle réduite » selon la définition donnée en [1], notamment en raison du faible nombre de caractéristiques utilisées.

Dans [2] nous avons proposé une méthode itérative originale de sélection globale de paramètres combinant un apprentissage des capacités associées à l'intégrale de

Choquet et un classificateur flou par règles linguistiques (FRIFS).

L'étude expérimentale a montré, sur des bases textiles, que les objectifs recherchés sont atteints : le nombre de paramètres utiles décroît tandis que les taux de reconnaissance augmentent. Le nombre de règles floues générées est également diminué ce qui permet une meilleure interprétabilité du Système.

La principale contrainte de la méthode concerne le nombre de paramètres qui est limité à une dizaine. Nous proposons ici d'étendre notre approche afin qu'elle puisse s'appliquer à un ensemble de paramètres plus important (FRIF-MS), ce qui est le cas pour l'application industrielle de ce cadre d'étude (ici 20). Une autre amélioration concerne l'automatisation de la sélection du paramètre le moins significatif à chaque itération de la méthode.

Après avoir rappelé le principe de la méthode de sélection de paramètres FRIFS, nous détaillons le nouveau mécanisme de sélection employé qui se fonde une recherche itérative du paramètre jugé le moins significatif. Ensuite, après avoir présenté le cadre applicatif, nous exposons les résultats obtenus et les comparons en terme de taux de reconnaissance à ceux fournis par la méthode FRIFS ainsi qu'à d'autres méthodes de références (SBFS, SFFS, SVM).

2 Description de la méthode FRIFS-MS

2.1 Méthode FRIFS

La méthode combine deux approches qui ont prouvé expérimentalement leur efficacité dans le contexte applicatif décrit en [2]. La méthode de sélection de paramètres fondée sur l'intégrale de Choquet fournit les paramètres jugés les plus pertinents à un classificateur à base de règles linguistiques floues (F.R.C.) [3], utilisé pour la reconnaissance. Le jeu de règles du FRC est généré par apprentissage. Ce couplage est itératif et vise à déterminer le sous ensemble de paramètres le plus pertinent en fonction d'un taux de reconnaissance désiré. A chaque itération, un sous-ensemble de paramètres (généralement 1, 2 ou 3) les moins pertinents est extrait à partir de l'analyse des indices d'importance et d'interaction. L'analyse des taux de reconnaissance obtenus avec le F.R.C permet de repérer le paramètre le moins discriminant qui est ensuite extrait.

2.2 Fuzzy Rules Classifier

Le classificateur utilisé (FRC : Fuzzy Rule Classifier) [3] est basé sur un mécanisme de règles linguistiques floues qui est adapté au contexte applicatif. En effet, il possède une bonne capacité de généralisation comme l'ont montré les comparaisons effectuées avec d'autres classificateurs (K-NN, RdN, SVM) [4]. Il peut de plus rendre compte d'une gradualité d'appartenance aux différentes classes de sortie. Cette méthode de reconnaissance floue est un mécanisme supervisé qui se décompose en trois parties : la fuzzification des paramètres du vecteurs caractéristiques, la génération des règles et l'ajustement du modèle.

L'étape de fuzzification vise à traduire une variable numérique en une variable linguistique. Elle détermine le nombre de termes à fournir aux prémisses des règles floues. Ces différents termes sont choisis en lien avec le vocabulaire de l'expert du domaine et sont généralement définis de façon empirique. Pour simplifier le réglage, les utilisateurs industriels de la méthode préfèrent une partition égale des termes ce qui conduit généralement à utiliser plus de termes qu'il n'en faut. Dans cet article, nous utilisons cette méthode avec 5 termes équirépartis pour chaque paramètre. Cependant, quand le nombre de termes augmente, le nombre de règles augmente également et donc la complexité du système. C'est pourquoi, pour pouvoir interpréter le système de règles obtenu, les termes doivent être adaptés aux données d'entrée. Des méthodes peuvent être utilisées pour effectuer une adaptation automatique de la fuzzification.

Les règles floues utilisées sont de type "SI... ALORS...". Chaque règle décrit la façon dont les défauts sont perçus par le système. Dans la présente étude, le choix des règles conjonctives a été fait car elles peuvent être obtenues à partir de lots de données numériques [5]. Chaque règle est activée en parallèle et un opérateur de disjonction est utilisé pour agréger les

conclusions partielles dans le but de fournir la décision finale

Le classificateur que nous avons utilisé est basé sur l'algorithme d'Ishibuchi qui propose une étape automatique de génération de règles [6] et a l'avantage d'être simple et très efficace. En outre, il correspond à un mécanisme d'inférence de type Larsen, qui est préférable à celui de Mamdani dans le cas où il y a plusieurs prémisses. La version itérative de l'algorithme est utilisée pour notre étude [7], les règles sont obtenues en favorisant la règle de réponse maximum, à partir d'un lot de défauts préparé par un expert du domaine.

2.3 Indices calculés

La première étape du système consiste à apprendre les poids associés à la capacité calculée à l'aide de l'intégrale de Choquet. Soient m classes, C_1, \dots, C_m , et un ensemble X de n Critères de Décision (DC) $X = \{D_1, \dots, D_n\}$. Par critère de décision, nous considérons ici un descripteur et la métrique associée. Soit une observation x_0 , l'objectif est de calculer pour chaque DC le degré de confiance de l'affirmation « Selon D_j , x_0 appartient à la classe C ».

A notre connaissance, l'algorithme fournissant la meilleure approximation et, le mieux adapté à notre problème, est celui proposé par M. Grabisch [8]. Il part du principe qu'en l'absence d'information le modèle d'agrégation le plus raisonnable est la moyenne arithmétique. Ce qui est intéressant pour notre problématique où nous possédons peu de données d'apprentissage par classe.

Dès que la mesure floue μ (ou capacité) est entraînée, il est possible d'interpréter la contribution de chaque critère de décision i dans la décision finale. Plusieurs indices peuvent être extraits à partir de la mesure floue pour mieux analyser le comportement. L'indice d'importance σ est fondé sur la définition proposée par Shapley dans le cadre de la théorie des jeux et replacée dans le contexte des mesures floues par Murofushi and Soneda [10]. Son expression est la suivante :

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

La valeur de Shapley σ peut être interprétée comme la valeur moyenne pondérée de la contribution $\mu(T \cup i) - \mu(T)$ du critère de décision i parmi toutes les combinaisons. L'indice d'interaction I introduit par Murofushi et Soneda [9] [10] représente le degré d'interaction positif ou négative entre deux DCs. Si la mesure floue a un comportement monotone alors des DCs interagissent. La valeur de l'interaction entre i et j est conditionnée par la présence des éléments de la combinaison $T \subseteq X \setminus \{i, j\}$

En étendant ce critère sur tous les sous ensembles de $T \subseteq X \setminus \{i, j\}$, on obtient une évaluation de l'interaction I entre les DCs i et j , comme suit :

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

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

Une interaction positive pour deux DCs i et j signifie que l'importance d'un DC est renforcée par la seconde. En d'autres termes, les deux DCs sont complémentaires et leur combinaison améliore la décision finale. Une interaction négative signifie que les DCs sont antagonistes et que leur combinaison réduit l'impact de la décision finale.

2.4 Algorithme de sélection fondé sur les indices.

2.4.1 Extraction de DC

Nous proposons de classer les règles de décision en combinant leur degré d'importance et leurs interactions $f_i(\sigma, I) \rightarrow [0, 1]$.

$$f_i(\sigma, I) = n \times \sigma(\mu, i) \times (\sum_{j=1, n} I(\mu, k_{ij}) - M) / K$$

avec :

$$K = \sum_{i=1, n} \sum_{j=1, n} |I(\mu, k_{ij})| \text{ et } M = \min \{ \sum_{j=1, n} I(\mu, k_{ij}) \}_{i=1, n}$$

La règle de décision ayant la plus faible importance et interagissant le moins avec les autres est supposée dégrader la décision finale.

2.4.2 Algorithme multi échelles

Il est bien connu que les méthodes d'agrégation multicritères ne permettent pas de gérer efficacement plus de $L=10$ critères. Nous proposons une approche combinant une méthode assimilable à un algorithme glouton et une étude des indices pour diminuer de manière itérative l'espace de paramètres jusqu'à un niveau d'interprétabilité réaliste. La première étape consiste à éclater aléatoirement X en N sous ensembles interprétables X_i tel que $X = \cup \{X_i\}_{i=1, N}$, $X_i \cap X_j = \emptyset$, $|X_i| \cong L$ and $N \geq 2$.

Pour chacun de ces sous-ensembles X_i , nous réalisons une phase d'apprentissage puis nous classons, par ordre croissant, les paramètres à partir d'une combinaison linéaire des deux principaux indices normalisés σ et I calculés sur toutes les capacités μ_i obtenues (cf. 2.4.1). Ensuite nous considérons aléatoirement des paires de sous-ensembles X_i et X_j pour lesquels nous permutons 2 à 2 les paramètres les plus forts et les plus faibles tout en conservant la moitié des paramètres initiaux sur chacun des sous-ensembles.

Puis nous effectuons un nouvel apprentissage et nous calculons le nouveau classement associé à ces nouvelles capacités. Si des paramètres jugés faibles à l'étape précédente sont à nouveau classés faibles, en considérant la moitié inférieure des sous-ensembles, ceux-ci sont directement extraits de X . Sinon, les paramètres faibles deviennent forts, alors on extrait le plus faible paramètre trouvé pour tous les sous-ensembles de X dans les deux étapes et ceci, pour garantir la convergence de la méthode. Le processus est

réitéré jusqu'à obtenir un modèle dit interprétable garantissant l'utilisation classique de FRIFS.

Étape 1: Partitionnement aléatoire de X en N sous ensembles X_i

Étape 2: **For Each** X_i ,
- Apprentissage des capacités (2.2)
- Tri des DC suivant $f_i(\sigma, I)$ (2.3)

End For all

Étape 3: Choix aléatoire des $(X_i, X_j)_{i \neq j}$

Étape 4: **For each** (X_i, X_j)
- Permutation 2 \Leftrightarrow 2 ($\text{argmin} \{f_i(\sigma, I)\}$ et $\text{argmax} \{f_j(\sigma, I)\}$)
- Apprentissage et tri / $f_i(\sigma, I)$, $f_j(\sigma, I)$

End for each

If DCs sont à nouveau faibles **then**

- Extraction de X de ces DCs

Else

(6) - Extraction de X de la DC la plus faible parmi toutes les configurations

End if

If non interprétable **Go back to** Étape 1

Étape 5: Appliquer FRIFS.

3 Etude expérimentale

3.1 Contexte Applicatif

Le contexte applicatif de cet article concerne l'amélioration d'un système de reconnaissance par vision de défauts sur des planches de bois. L'analyse des planches s'effectue en temps réel sur une ligne de production dont la vitesse peut atteindre 300 m/mn ; ce qui fixe les contraintes de temps du système pour l'identification d'un défaut. Le mécanisme de reconnaissance se base sur l'exploitation d'un vecteur caractéristique formé d'une vingtaine de paramètres calculés sur l'image des défauts (surface d'un défaut, orientation, couleur, ...).

3.2 Résultat / Comparaison

Les résultats ont été obtenus à partir d'une base de données composée de 877 échantillons répartis en 9 classes de défauts. La base de données d'apprentissage, sur laquelle la méthode de sélection est appliquée, est composée de 250 échantillons. La base de données utilisée pour tester et valider les sous-ensembles construits est composée de 627 échantillons. Le moteur d'inférence floue du FRC est composé d'une inférence unique où toutes les caractéristiques sont en entrée du modèle et toutes les classes à reconnaître en sortie. La fuzzyfication est faite en 5 termes équirépartis sur l'univers de discours pour chaque paramètre.

Les tests visent à réduire la dimension du vecteur caractéristique en supprimant les paramètres les moins pertinents jusqu'à obtenir le meilleur taux de reconnaissance possible. Nous utilisons 3 méthodes de référence (SBFS, SFFS et SVM) pour supprimer les 10

premiers paramètres que nous comparons à notre méthode automatique FRIF-MS. La méthode FRIF initiale ne peut être appliquée qu'à partir du moment où le nombre de paramètres est inférieur ou égale à 10. Cette première étape conduit à des ensembles de paramètres identiques pour les méthodes SBFS, SFFS et FRIF-MS. Pour la méthode SVM, deux paramètres diffèrent. Les taux de reconnaissance obtenus sont respectivement de 94% en mémorisation et 74,5% en généralisation (93,6% et 71% pour SVM).

Le tableau 1 donne les suppressions successives des paramètres en fonction des différentes méthodes.

Tableau 1: Comparaison des taux de reconnaissance en fonction des paramètres supprimés

paramètres / méthodes	FRIFS-MS	FRIFS	SBFS / SFFS	SVM	
10 paramètres	Tx App	94,00%	94,00%	94,00%	93,60%
	Tx Gene	74,48%	74,48%	74,48%	71,00%
9 paramètres	Param LR_RE	LR_RE	LR	CR3	
	Tx App	95,20%	95,20%	92,00%	93,60%
8 paramètres	Param SURF	SURF	G AXE	ORIENT	
	Tx App	95,20%	95,20%	92,00%	95,20%
7 paramètres	Param C3	C3	DX/DY	LR	
	Tx App	94,40%	94,40%	90,40%	87,60%
6 paramètres	Param C4	GD_AXE	C3	C3	
	Tx App	84,00%	93,60%	90,00%	88,00%
5 paramètres	Param GD_AXE	PT_AXE	SURF	SURF	
	Tx App	82,00%	92,40%	84,80%	88,00%
	Tx Gene	72,25%	76,24%	71,93%	76,40%
4 paramètres	Param PT_AXE	C4	PT_AXE	DX/DY	
	Tx App	80,80%	80,80%	74,80%	75,60%
	Tx Gene	72,41%	72,41%	69,86%	69,38%

Le taux de reconnaissance maximal correspond à un ensemble de 4 - 5 paramètres. Pour 5 paramètres, on peut s'apercevoir que la méthode FRIFS-MS délivre un résultat semblable en généralisation, aux méthodes de références (72%) et très comparable à la méthode FRIFS initiale (76,2%) qui nécessite une boucle d'itération supplémentaire. Pour 4 paramètres, on retrouve le même ensemble que celui délivré par la méthode FRIFS. Les ensembles obtenus à partir des méthodes SVM, SBFS ou SFFS diffèrent peu. Le lot de données du contexte applicatif est assez réduit et ne permet pas des analyses statistiques précises quant aux résultats. Néanmoins les essais confirment, sur un lot de données issu d'un domaine différent de celui pour lequel elle a été développée (bois vs textile), que la méthode FRIFS conduit à un ensemble de paramètres pertinents [2].

4 Conclusion

Nous présentons dans cet article une extension « Multi-Scale » de la méthode FRIFS pour la sélection de paramètres. Cette amélioration vise à automatiser l'étape de choix du paramètre le moins pertinent à éliminer en fonction du taux de reconnaissance global. Les essais réalisés sur des images industrielles, dans le cadre de l'identification de défauts du bois, montrent l'efficacité de FRIF-MS et confirme sa capacité à fournir un modèle de reconnaissance interprétable sur un nouveau lot de données. Nos prochains travaux concernent notamment l'application de cette méthode à une sélection de paramètres pertinents spécifiques à chaque classe à reconnaître.

5 Bibliographie

- [1] Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition* 33 (2000) 25–41.
- [2] Schmitt E., Bombardier V., Wendling L., Improving Fuzzy Rule Classifier by Extracting Suitable Features from Capacities with Respect to the Choquet Integral, *IEEE TSMC- part B*, Vol 38, N° 5, October 2008.
- [3] Schmitt, E., Mazaud, C., Bombardier V., Lhoste, P., A Fuzzy Reasoning Classification Method for Pattern Recognition. *Proc. 15th Int. Conf. on Fuzzy Systems, FUZZIEEE'06*. Vancouver, Canada, (2006) 5998-6005.
- [4] Bombardier V., Schmitt E., Charpentier P., A fuzzy sensor for color matching vision system, *Measurement*, Vol 42, n° 2, Feb. 2009, Pages 189-201.
- [5] Dubois D., Prade H., What are Fuzzy rules and how to use them?, *Fuzzy Sets and Systems* 84 (1996) 169-185.
- [6] Ishibuchi, H., Nozaki, K., Tanaka, H.: A Simple but powerful heuristic method for generating fuzzy rules from numeric data, *Fuzzy sets and systems* 86 (1997) 251-270.
- [7] Ishibuchi, H., Nozaki, K., Tanaka, H: Distributed representation of fuzzy rules and its application to pattern classification, *Fuzzy Sets and Systems* 52 (1992) 21-32.
- [8] Grabisch, M.: A new algorithm for identifying fuzzy measures and its application to pattern recognition in *Int. Joint Conf. 7th IEEE Int. Conf. on Fuzzy Systems*, Yokohama, Japan, (1995) 145-150.
- [9] Shapley, L.: A value for n-person games. *Contributions to the Theory of Games*, *Annals of Mathematics Studies*. Khun, H., Tucker, A., (ed.). Princeton Univ. Press (1953) 307-317.
- [10] Murofushi T., Sugeno M.: A theory of fuzzy measures: representations, the Choquet integral, and null sets. *J. Mathematical Analysis and Applications* 159 (1991) 532-549.