Une approche de type filtrage pour la sélection de variables. Application à la reconnaissance automatique de poissons

Dahbia Semani, Carl Frélicot, Pierre Courtellemont
Laboratoire d'Informatique – Image – Interaction (UPRES EA 2118),
Université de La Rochelle, avenue M. Crépeau, 17042 La Rochelle Cedex 1, France
{dahbia.semani,carl.frelicot,pierre.courtellemont}@univ-lr.fr

Résumé − Cet article aborde le problème de la sélection de variables dans le cadre de la classification supervisée. Nous présentons un nouveau critère permettant de mesurer la pertinence d'un sous-ensemble de variables. Ce critère repose sur une mesure d'ambiguïté fondée sur la combinaison d'étiquettes représentant le degré d'appartenance aux classes en présence. Des tests sont menés sur des jeux de données réels issus de la littérature. L'application traitée concerne un problème réel de reconnaissance des formes, à savoir la reconnaissance d'objets dans les séquences vidéo.

Abstract − This paper addresses the feature selection problem for supervised classification. Feature selection methods are based on a selection algorithm and a criterion function assessing how effective feature subsets are. We propose an ambiguity measure that allows to define a new evaluation criterion. It is based on a combination of labels representing the degree of typicality to the classes. The new criterion is compared to others found in the literature on various real data sets. It is validated on a real object recognition problem.

1 Introduction

La sélection de variables joue un rôle très important en classification lorsqu'un grand nombre p de variables sont disponibles, certaines pouvant être peu significatives, corrélées ou non pertinentes au regard de l'application considérée [4]. Elle consiste à sélectionner un sous-ensemble de q variables ($q < p$) sans que les performances de la règle de classement diminuent trop voire même augmentent. La sélection permet également de faciliter l'étape d'apprentissage et de réduire la complexité des algorithmes ainsi que les temps de calcul.

Une méthode de sélection repose principalement sur un algorithme de recherche et un critère d'évaluation pour mesurer la pertinence des sous-ensembles potentiels de variables. Nous nous intéressons aux critères d'évaluation. Ainsi, nous proposons un nouveau critère fondé sur une mesure d'ambiguïté. Cette mesure repose sur la combinaison d'étiquettes représentant le degré de spécificité ou d'appartenance des données aux classes en présence. Des opérateurs d'agrégation sont utilisés pour la combinaison de ces étiquettes.

L'article est organisé comme suit : Un bref état de l'art sur les algorithmes de sélection et les critères d'évaluation est dressé à la section 2. Le critère proposé ainsi que sa validation sur des jeux de données réels issus de la littérature sont présentés à la section 3. La section 4 est consacrée à la description du système de reconnaissance automatique de poissons. Ce système constitue un cadre d'application de notre approche de sélection de variables. Les résultats de sélection et les taux de reconnaissance sont présentés à la section 4.1.

2 État de l'art

Algorithmes de recherche

Il existe divers algorithmes de recherche et études comparant leurs avantages et inconvénients [3]. Les méthodes les plus populaires procèdent par ajouts et/ou suppressions séquentiels de variables. Elles évitent la recherche exhaustive du meilleur sous-ensemble et conduisent à une solution sous-optimale. Nous utilisons l'algorithme SFFS (*Sequential Forward Floating Search*) [7] permettant à chaque pas, d'ajouter une variable et d'en supprimer plusieurs tant que le sous-ensemble résultant améliore le critère d'évaluation. Il est considéré comme la méthode sous-optimale la plus efficace [3]. Les deux étapes (ajout / suppression) de l'algorithme sont alternées jusqu'à ce qu'une condition d'arrêt soit vérifiée. Parmi celles-ci, citons une borne sur q ou un seuil sur la valeur du critère d'évaluation.

Critères d'évaluation

Deux approches sont couramment utilisées pour évaluer la pertinence d'un sous-ensemble de variables sélectionnées [5] : l'approche de type *filtrage* (*filter approach*) et celle de type « *enveloppante* » (*wrapper approach*). Dans la première, les critères sont fondés uniquement sur les données et sont donc totalement indépendants du discriminateur utilisé. Les variables sont alors filtrées avant le processus d'apprentissage et de classification. Parmi les différentes fonctions utilisées, citons celles fondées sur des mesures de distance probabilistes (ex : Mahalanobis, Battacharyya), d'information (ex : entropie) ou de dépendance (ex : corrélation, information mutuelle). La seconde approche tient compte de la règle de classement dans le calcul du critère d'évaluation. Celui-ci est simplement la probabilité d'erreur estimée sur l'ensemble des données.

3 Un nouveau critère d'évaluation

Étiquetage

Considérons un point $\mathbf{x} = (x_1, ..., x_p)^t$ dans un espace de représentation de dimension p et un ensemble de c classes $\omega = \{\omega_1, ..., \omega_c\}$. On peut lui associer un vecteur d'étiquettes à l'aide d'une fonction : $\Re^p \rightarrow [0,1]^c$, $\mathbf{x} \mapsto \mu(\mathbf{x}) = (\mu_1(\mathbf{x}), ..., \mu_c(\mathbf{x}))^t$ où $\mu_i(\mathbf{x})$ représente le degré d'appartenance de $\mathbf{x}$ à la classe ω_i, par exemple :

$$\mu_i(\mathbf{x}) = (1 + d(\mathbf{x}, \omega_i))^{-1} \quad (1)$$

où $d(\mathbf{x}, \mathbf{p}_i)$ est une distance (ex : distance de Mahalanobis) entre $\mathbf{x}$ et le vecteur moyenne de la classe ω_i.

La mesure d'ambiguïté

En classification, une variable est d'autant moins discriminante que les projections des classes sur cette variable se chevauchent. Le chevauchement est révélateur d'une certaine ambiguïté. Nous voulons quantifier l'ambiguïté entre les classes en combinant les étiquettes μ_i. Pour cela, nous avons choisi d'utiliser les normes (et conormes) triangulaires ou *t-normes* (et *t-conormes*) qui sont des opérateurs de combinaison tout à fait adaptés à notre problème ; le lecteur intéressé peut en trouver une synthèse dans (Klir et Yuan 1995). Dans ce qui suit, $\top$ désigne une t-norme arbitraire et $\perp$ sa t-conorme duale.

Dans [8] nous avons introduit un nouvel opérateur d'agrégation dans le cadre de la classification supervisée avec double option de rejet et en particulier option de rejet d'ambiguïté. Nous l'avons baptisé *OU-2 flou*, noté $\overset{2}{\perp}$, défini par :

$$\overset{2}{\underset{i=1,c}{\perp}} \mu_i = \underset{i=1,c}{\top} \left(\underset{j=1,c;j\neq i}{\perp} \mu_j \right) \quad (2)$$

Nous avons également montré que dans le cas où la t-norme utilisée est le *min*, $\overset{2}{\perp}_{i=1,c} \mu_i$ est égal au deuxième plus grand des μ_i. Cet opérateur possède plusieurs propriétés mathématiques dont certaines découlent naturellement de celles des t-normes et t-conormes (ex : bornes, monotonie, continuité, symétrie, etc.) [8].

Un moyen assez naturel de mesurer l'ambiguïté entre les classes est d'effectuer le rapport entre le *deuxième plus grand* et le *plus grand* des μ_i. La mesure d'ambiguïté que nous proposons généralise ce rapport :

$$A(\mathbf{x}) = \overset{2}{\underset{i=1,c}{\perp}} \mu_i(\mathbf{x}) \Big/ \underset{i=1,c}{\perp} \mu_i(\mathbf{x}) \quad (3)$$

Le critère d'évaluation proposé

Nous proposons d'utiliser la mesure d'ambiguïté (3) pour définir le nouveau critère d'évaluation d'un sous-ensemble S_q de q variables, suivant :

$$J_A(S_q) = \sum_{\mathbf{x}} A^{[q]}(\mathbf{x}) \quad (4)$$

où l'exposant $^{[q]}$ indique que la mesure d'ambiguïté est définie à partir d'étiquettes $\mu_i^{[q]}$ représentant le degré d'appartenance, donné par exemple par (1), du point $\mathbf{x}$ à la classe ω_i dans l'espace $\mathbb{R}^q$. Il s'agit alors de sélectionner l'ensemble des q variables parmi les p d'origine qui minimisent le critère $J_A(S_q)$.

Validation du critère d'évaluation

Nous avons associé le critère J_A à l'algorithme SFFS pour définir une méthode de sélection, baptisée *SFFS–fOU2*. Les normes standard et d'Hamacher avec $\gamma = 0$ et $\gamma = 1$ ont été utilisées dans la définition de la mesure J_A. Bien que les résultats soient comparables, ceux reportés correspondent à chaque fois aux normes ayant donné les meilleures performances.

Nous avons testé la méthode *SFFS–fOU2* sur des jeux de données standard issus de la base de données UCI : *Iris*, *Pima Indian* et *Breast Cancer Wisconsin*. La méthode de sélection a permit de réduire, pour tous les jeux, le nombre de variables d'au moins la moitié. La comparaison des taux de bon classement obtenus avec les p variables d'origine et les q variables sélectionnées est faite pour deux discriminateurs classiques : la règle des *k-Plus Proches Voisins* (k-PPV) et la règle de *Bayes Quadratique*, sous hypothèse gaussienne (BQ). Rappelons que la règle BQ correspond à la règle du *Maximum A Posteriori* (MAP) en considérant une matrice de covariance Σ_i propre à chaque classe ω_i. L'estimation des taux a été réalisée selon une procédure 10-CV (*Validation Croisée*). Pour réduire le biais dû au caractère aléatoire de la construction des ensembles test et apprentissage par la procédure 10-CV, cette dernière est répétée en réalisant 10 essais indépendants. Nous disposons donc de dix taux estimés, nous avons reporté les intervalles de confiance à 95% sur la moyenne de ces taux dans le tableau 1. Les résultats reportés montrent que la méthode proposée est capable de sélectionner les variables pertinentes puisque aucune baisse des performances n'a été enregistrée. Les performances sont même augmentées dans certains cas.

Notre critère a été comparé avec d'autres critères issus de la littérature et appartenant à l'approche *filtrage* : la distance de Mahalanobis, l'entropie [6] et la distance floue [1]. Les résultats obtenus par notre critère et la distance de Mahalanobis sont comparables, quel que soit le discriminateur. Les résultats obtenus par notre critère sont cependant meilleurs que ceux obtenus par la mesure d'entropie et la distance floue. Nous avons reporté, à titre d'exemple, dans le tableau 2 les taux de bon classement obtenus en utilisant la règle BQ.

TAB. 1 – Moyennes et intervalles de confiance des taux de bon classement obtenus avec la procédure *10-VC*.

Jeux	Critère	BQ (%)	k-PPV (%)
Iris	$p = 4$	97.20±0.20	96.26±0.63
	$J_A(q = 2)$	97.13±0.32	96.27±0.56
Pima	$p = 8$	74.06±0.51	74.60±0.57
	$J_A(q = 3)$	75.57±0.30 +	75.75±0.51 +
Breast	$p = 9$	95.17±0.16	97.11±0.24
	$J_A(q = 3)$	96.27±0.11 +	96.82±0.20

Amélioration (+) ou dégradation (−) statistiquement significative (95%)

TAB. 2 – Taux de bon classement obtenus dans le cas de la règle BQ.

Jeux	p **variables**	J_A	$Maha$	DF	E
Iris	97.20±0.20	97.13±0.32	97.40±0.15	94.67±0.42−	97.07±0.33
Pima	74.06±0.51	75.57±0.30+	75.40±0.17+	74.39±0.25	74.73±0.22+
Breast	95.17±0.16	96.27±0.11+	95.99±0.18+	95.23±0.07	94.73±0.17−

Amélioration (+) ou dégradation (−) statistiquement significative (95%)

4 Application à la reconnaissance automatique de poissons

Nous décrivons, dans cette section, un système de reconnaissance automatique de poissons évoluant dans un aquarium. Ce travail s'inscrit dans le cadre du projet Aqu@thèque[1] dont l'objectif est de permettre au visiteur d'un aquarium de désigner un poisson sur un écran tactile diffusant les images acquises par une caméra vidéo, distante et fixe, orientée vers un bassin de l'aquarium. Le système de reconnaissance doit identifier automatiquement, et en temps réel, l'espèce correspondant au poisson désigné. Des informations multimedia (textuelles, graphiques, audio ou vidéo) à caractère pédagogique sur le poisson sont alors mises à disposition du visiteur. Le système de reconnaissance comporte les étapes suivantes :

1. **Segmentation** : Correspond à effectuer une partition de l'image en régions susceptibles d'être des poissons ou parties de poissons. L'objectif principal est donc de séparer les objets en mouvement du fond de la scène supposé statique. Pour cela, nous utilisons la méthode de *détection de changement temporels* d'Elgammal et al. [2].
2. **Extraction d'attributs** : Les régions extraites des images sont caractérisées par un ensemble d'attributs (ou variables) pour permettre leur identification. Différents attributs peuvent être calculés sur une région [9]. Nous avons extrait 85 variables réparties sur différents groupes [8] : attributs géométriques, photométriques, texture, moments couleur, moments de Hu, paramètres de déplacement. Le tableau 3 résume le nombre de variables par groupe.
3. **Sélection d'attributs** : Cette étape permet de ne retenir, parmi tous les attributs extraits, que ceux qui sont discriminants au regard de notre application.
4. **Reconnaissance** : L'identification de l'espèce d'un poisson se fait grâce à une procédure de classification supervisée, c'est-à-dire que seules les classes apprises seront reconnues.

4.1 Résultats expérimentaux

La base d'apprentissage
Un bassin de l'Aquarium de la Rochelle comprenant 12 espèces $\Omega = \{\omega_1, \omega_2, ..., \omega_{12}\}$ a été filmé pour obtenir les séquences d'apprentissage. La base d'apprentissage a été construite en étiquetant, suivant les 12 classes, 1900 régions extraites des images segmentées. Les régions sont décrites par les 85 attributs.

Sélection d'attributs
L'étape de la sélection d'attributs est appliquée à l'ensemble des données centrées réduites de la base d'apprentissage. Nous avons sélectionné les attributs les plus pertinents en utilisant deux approches différentes :

1. *Sélection globale* (*Sélection # 1*) : L'utilisation de la méthode *SFFS–fOU2*, permet de sélectionner 24 attributs, ce qui représente une réduction significative du nombre de variables d'origine (71.77% de réduction, voir le tableau 3).
2. *Sélection hiérarchique* (*Sélection # 2*) : En analysant les résultats de la classification obtenus avec les $p = 85$ variables d'origine, nous avons remarqué que certaines classes sont très confondues induisant ainsi une baisse des performances. La confusion provient des espèces appartenant en réalité à une même espèce de poissons. La ressemblance (visuelle) entre des espèces complètement différentes est également source de confusion. Il est alors raisonnable de penser qu'un seul ensemble de variables est insuffisant pour discriminer toutes les classes. Un ensemble de variables capable de discriminer une classe des autres n'est pas forcement capable de discriminer une autre classe des autres.

   Nous avons décider de fusionner certaines classes selon le principe suivant : 1) une première classification des objets, par rapport aux différentes espèces présentent dans le bassin, est effectuée dans l'espace des variables d'origine ($p = 85$ variables), 2) les espèces qui se confondent avec un taux d'erreur de classement supérieur à 10% sont fusionnées. Les classes ayant été discriminées avec un taux de reconnaissance supérieur à 70% sont gardées comme des classes singleton. Finalement, nous obtenons 19 nouvelles classes contenant des classes *fusionnées* et des classes *singleton*. Par exemple, si la classe ω_i est confondue avec la classe ω_j qui elle même se confond avec la classe ω_k, nous obtenons les deux classes *fusionnées* $\omega'_l = \{\omega_i, \omega_j\}$ et $\omega'_m = \{\omega_j, \omega_k\}$.

   La *sélection hiérarchique* s'effectue en deux étapes :
   – *Étape 1 :* nous considérons l'ensemble des 19 nouvelles classes et nous appliquons d'une manière globale la méthode *SFFS–fOU2* comme dans *Selection # 1*. Cette étape permet de réduire les 85 attributs en 27 attributs (voir tableau 3).

[1] Le projet Aqu@thèque est mené par le laboratoire L3i de l'Université de La Rochelle.

TAB. 3 – Résumé des attributs.

Nombre da variables	Avant sélection	*Sélection # 1*	*Sélection # 2*	
			Étape 1	Étape 2
Géométriques	10	2	3	5
Photométriques	37	11	10	30
Texture	18	7	9	15
Moments	13	4	4	7
Moments de Hu	4	0	1	3
Déplacement	3	0	0	0
Total	85	24	27	60
Taux de réduction		71.77%	68.23%	30%

– *Étape 2 :* les objets sont classés dans l'une des 19 nouvelles classes en utilisant le sous-ensemble des 27 variables sélectionnées lors de l'Étape 1. Les objets classés dans une classe fusionnée doivent être séparés en utilisant le sous-ensemble de variables approprié. L'objectif de cette étape est de sélectionner pour chaque classe fusionnée $\omega'_l = \{\omega_i, \omega_j\}$, le sous-ensemble qui discrimine au mieux les deux classes ω_i et ω_j. Une nouvelle sélection est donc effectuer par la méthode *SFFS–fOU2* au niveau de chaque classe fusionnée. Ceci permet d'obtenir différents sous-ensembles de variables. Au final, 60 variables ont été retenues dans cette étape, représentant un taux de réduction d'environ 30% (voir tableau 3).

Les attributs sélectionnés par les deux méthodes *Sélection # 1* et *Selection # 2* (voir tableau 3) sont essentiellement des moments de couleur (moments d'ordre 1 et 2, moments de chromaticité, moments généralisés couleur) et des attributs de texture (couleur et niveaux de gris).

Résultats de classification

Les taux de bon classement obtenus avec les p variables d'origine et les variables sélectionnées par les deux méthodes *Sélection # 1* et *Sélection # 2*, sont présentés dans le tableau 4. Les taux de bon classement sont estimés par une procédure *Holdout* [4] pour deux discriminateurs : BQ et MDAG. La règle de classement MDAG intègre deux options de *rejet* suivant la stratégie « *mélange d'abord généralisée* » que nous avons proposé dans [8] : le *rejet de distance* et le *rejet d'ambiguïté*. Le premier type de rejet permet de n'associer un point $\mathbf{x}$ à aucune des classes ; il concerne les $\mathbf{x}$ situés en général loin de toute classe. Le second consiste à associer $\mathbf{x}$ à plusieurs ou à toutes les classes ; il concerne en général les $\mathbf{x}$ se projetant dans des régions situées entre deux classes ou plus. L'introduction du rejet permet de réduire le risque de mauvais classement.

Les performances obtenues avec les attributs sélectionnés par la méthode de *sélection globale* sont meilleures que celles obtenues avec les variables d'origine, quel que soit le discriminateur. Des meilleures performances sont obtenues en utilisant la méthode de *sélection hiérarchique*. Le taux de bon classement est, dans ce cas, supérieur à 90% pour les deux discriminateurs.

TAB. 4 – Taux de bon classement obtenus avec les variables d'origine et les variables sélectionnées.

Méthode	BQ (%)	MDAG (%)
Avant sélection ($p = 85$)	63.73	73.79
Sélection # 1 ($q = 24$)	76.51	77.96
Sélection # 2 ($q = 60$)	**90.87**	**92.52**

5 Conclusion

Nous avons présenté, dans cet article, un nouveau critère permettant d'évaluer la pertinence d'un (sous-)ensemble de variables. Le critère proposé peut être associé à n'importe quel discriminateur. Ses performances ont été comparées avec d'autres critères issus de la littérature. Les tests menés sur des jeux de données réels ont montré que le critère proposé est capable de sélectionner les variables pertinentes et d'augmenter dans la plupart des cas les taux de bon classement. Nous avons validé la méthode de sélection de variables proposée sur un problème réel de reconnaissance des formes. L'application concerne un système de reconnaissance automatique de poissons évoluant dans un aquarium. La sélection de variables a permit une amélioration de plus de 14% du taux de reconnaissance.

Références

[1] T. E. Campos, I. Bloch, and R. M. Cesar Jr. *Feature Selection Based on Fuzzy Distances between Clusters : First Results on Simulated Data.* In : Lecture Notes in Comp. Sc. 2013 : Advances in Pattern Recognition, Springer-Verlag, 2001.

[2] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis. Background and foreground modeling using nonparametric kernel density for visual surveillance. *Proceedings of the IEEE*, 90(7) :1151–1163, July 2002.

[3] A. Jain and D. Zongker. Feature selection : Evaluation, application and small sample performance. *IEEE Trans. on PAMI*, 19(2) :153–158, 1997.

[4] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition : A review. *IEEE Trans. on PAMI*, 22(1) :4–37, 2000.

[5] P. Langley. Selection of relevant features in machine learning. *In AAAI Fall Symposium on Relevance*, pages 140–144, 1994.

[6] P. Mitra, C. A. Murthy, and S. K. Pal. Unsupervised feature selection using feature similarity. *IEEE Trans. on PAMI*, 24(3) :301–312, March 2002.

[7] P. Pudil, J. Novovicová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15 :1119–1125, 1994.

[8] D. Semani. *Une méthode supervisée de sélection et de discrimination avec rejet. Application au projet Aqu@thèque.* Thèse de doctorat, Université de La Rochelle, Mai 2004.

[9] S. Theodoridis and K. Koutroumbas. *Pattern Recognition.* Academic Press Inc., 1999.