

Régression floue et crédibiliste par SVM pour la classification des images sonar

Hicham Laanaya^{*,**}, Arnaud Martin^{**}
Driss Aboutajdine^{*} Ali Khenchaf^{**}

^{*}GSCM-LRIT, Université Mohammed V-Agdal, Faculté des sciences de Rabat, Maroc
aboutaj@fsr.ac.ma,

<http://www.fsr.ac.ma/GSCM/>

^{**}ENSIETA-E³I²-EA3876, 2, rue François Verny 29806 Brest cedex 9,
laanayhi, Arnaud.Martin, Ali.Khenchaf@ensieta.fr

<http://www.ensieta.fr/e3i2/>

Résumé. La classification des images sonar est d'une grande importance par exemple pour la navigation sous-marine ou pour la cartographie des fonds marins. En effet, le sonar offre des capacités d'imagerie plus performantes que les capteurs optiques en milieu sous-marin. La classification de ce type de données rencontre plusieurs difficultés en raison des imprécisions et incertitudes liées au capteur et au milieu. De nombreuses approches ont été proposées sans donner de bons résultats, celles-ci ne tenant pas compte des imperfections des données. Pour modéliser ce type de données, il est judicieux d'utiliser les théories de l'incertain comme la théorie des sous-ensembles flous ou la théorie des fonctions de croyance. Les machines à vecteurs de supports sont de plus en plus utilisées pour la classification automatique aux vues leur simplicité et leurs capacités de généralisation. Il est ainsi possible de proposer une approche qui tient compte de ces imprécisions et de ces incertitudes au cœur même de l'algorithme de classification. L'approche de la régression par SVM que nous avons introduite permet cette modélisation des imperfections. Nous proposons ici une application de cette nouvelle approche sur des données réelles particulièrement complexes, dans le cadre de la classification des images sonar.

1 Introduction

Les images sonar sont utilisées pour leur rapidité à imager de grandes zones là où l'optique ne peut le faire. On les retrouve ainsi dans de nombreuses applications telles que l'aide à la navigation sous-marine, ou la cartographie sous-marine.

Les images sonar sont entachées de plusieurs imprécisions et incertitudes dues à l'instrumentation utilisée (le capteur sonar) et au milieu marin. Les paramètres qui entrent en jeu pour la reconstruction de ces images (géométrie du dispositif, coordonnées du bateau, mouvements du sonar, . . .) sont aussi entachés des bruits de mesure. Il s'ajoute à ceci des interférences dues à des trajets multiples des signaux utilisés, à des bruits de chatolement ou encore à la faune et

la flore. Ces imperfections rendent la tâche difficile pour la caractérisation des fonds marins à partir de ce type de données. Il est donc nécessaire de proposer des algorithmes, robustes aux imperfections, pour la classification automatique des images sonar.

Plusieurs choix sont envisageable pour remédier aux problèmes d'imperfections : soit nous tentons de supprimer ces imperfections, ce qui nécessite une compréhension, souvent difficile, de la physique qui a conduit à ces imperfections ; soit nous cherchons à développer des processus de traitement robustes à ces imperfections ; soit nous cherchons à les modéliser.

Le cadre théorique des théories de l'incertain offre la possibilité de modéliser finement ces imperfections. Parmi elles, la théorie des ensembles flous et la théorie des fonctions de croyance permettent de tenir compte des incertitudes et imprécisions.

De nombreuses approches ont été proposées pour la classification des images sonar par exemple dans [Laanaya et al. \(2005b\)](#) et [Leblond et al. \(2005\)](#). Ces approches ne tiennent pas compte de l'incertitude de l'expert lors de la segmentation de ces images. Nous adopterons dans ce papier l'approche que nous avons proposée dans [Laanaya et al. \(2006\)](#) avec une résolution du problème d'optimisation adaptée à la classification automatique des images sonar. Cette approche a donné des résultats intéressants sur des données générées, nous montrerons ici son intérêt sur les données complexes que sont les images sonar.

Ainsi, nous présenterons une description rapide des fonctions d'appartenance et des fonctions de croyance utilisées par l'approche de la régression par SVM. Nous rappelons ensuite, l'approche de la régression par SVM après une brève introduction du principe des SVM. Cette approche est comparée au SVM classique et discutée dans une dernière partie à partir d'images sonar.

2 Théories de l'incertain

Nous avons vu dans [Martin \(2005\)](#) que les théories de l'incertain telles que la théorie des sous-ensembles flous introduite par [Zadeh \(1965\)](#), la théorie des possibilités de [Dubois et Prade \(1987\)](#) ou encore la théorie des fonctions de croyance de [Dempster \(1967\)](#) et [Shafer \(1976\)](#) permettent la modélisation de données incertaines et imprécises dans le cadre de la classification d'images sonar.

Ces théories sont fondées sur les fonctions d'appartenance pour les premières et sur les fonctions de croyance pour la dernière. Afin d'intégrer directement les contraintes liées à ces fonctions dans un algorithme de classification, nous rappelons ici les caractéristiques des fonctions d'appartenance de la théorie des sous-ensembles flous et des fonctions de croyance de Dempster et Shafer.

2.1 Les fonctions d'appartenance

Les fonctions d'appartenance permettent de décrire une appartenance floue à une classe. Ainsi l'appartenance d'une observation x à une classe C_i parmi N_c classes, est donnée par une fonction $\mu_i(x)$ telle que :

$$\begin{cases} \mu_i(x) \in [0, 1] \\ \sum_{i=1}^{N_c} \mu_i(x) = 1. \end{cases} \quad (1)$$

Dans ce cas, nous considérons les classes floues. Dans le cas de classes nettes, il est possible de considérer les distributions de possibilité. Typiquement x peut représenter une partie du fond marin et C_i le type de sédiment présent sur l'image x . Nous verrons au paragraphe 4.1.2 comment ces fonctions μ_i peuvent être choisies dans notre application.

2.2 Les fonctions de croyance

La théorie des fonctions de croyance est fondée sur la manipulation des fonctions de masse. Les fonctions de masse sont définies sur l'ensemble de toutes les disjonctions du cadre de discernement $\Theta = \{C_1, \dots, C_{N_c}\}$ et à valeurs dans $[0, 1]$, où C_i représente l'hypothèse "l'observation appartient à la classe i ". La contrainte de normalité couramment employée est ici donnée par :

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

où $m(\cdot)$ représente la fonction de masse. La première difficulté est donc de définir ces fonctions de masse selon le problème. Nous verrons comment il est possible de le faire pour notre application dans la section 4.1.2. A partir de ces fonctions de masse, d'autres fonctions de croyance peuvent être définies, telles que les fonctions de crédibilité, représentant l'intensité que toutes les sources croient en un élément, et telles que les fonctions de plausibilité représentant l'intensité avec laquelle on ne doute pas en un élément.

Afin de conserver un maximum d'informations, il est préférable de rester à un niveau crédal (*i.e.* de manipuler des fonctions de croyance) pendant l'étape de manipulation des informations pour prendre la décision sur les fonctions de croyance à l'issue de la manipulation de ces fonctions. Si la décision prise par le maximum de crédibilité peut être trop pessimiste, la décision issue du maximum de plausibilité est bien souvent trop optimiste. Le maximum de la probabilité pignistique, introduite par Smets (1990), reste le compromis le plus employé. La probabilité pignistique est donnée pour tout $X \in 2^\Theta$, avec $X \neq \emptyset$ par :

$$\text{betP}(X) = \sum_{Y \in 2^\Theta, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m(Y)}{1 - m(\emptyset)}. \quad (3)$$

2.3 Similitudes

Ainsi les fonctions d'appartenance et les fonctions de masse permettent une modélisation de l'incertitude et de l'imprécision à partir de points de vue différents.

Ces fonctions ont toutes deux la particularité d'être à valeurs dans $[0, 1]$ et d'avoir une contrainte de normalité équivalente. Nous allons voir dans la section suivante comment intégrer ces contraintes dans une régression linéaire multiple.

3 Régression floue et crédibiliste par SVM

Nous avons proposée dans Laanaya et al. (2006) une nouvelle approche pour la classification automatique fondée sur une régression effectuée à partir des SVM. Cette approche a

montré des performances remarquables sur des données générées. Afin d'assoir les notations utiles pour la suite, nous rappelons le principe des SVM sur laquelle s'appuie la régression floue et crédibiliste présentée ensuite.

3.1 Principe du classifieur SVM

Les machines à vecteurs de support initiées par [Vapnik \(1998\)](#), sont avant tout une approche de classification linéaire à deux classes. Elles tentent de séparer des individus issus de deux classes (+1 et -1) en cherchant l'hyperplan optimal qui sépare les deux ensembles, en garantissant une grande marge entre les deux classes. Un nombre réduit d'exemples pour la recherche de l'hyperplan est suffisant pour la description de cet hyperplan.

Dans le cas où les exemples sont linéairement séparables, on cherche l'hyperplan $y = w \cdot x + b$ qui maximise la marge entre les deux ensembles où $w \cdot x$ est le produit scalaire de w et x . Ainsi w est la solution du problème d'optimisation convexe :

$$\text{Min } \|w\|^2/2 \quad (4)$$

sous les contraintes :

$$y_t(w \cdot x_t + b) - 1 \geq 0 \quad \forall t = 1, \dots, l, \quad (5)$$

où les $x_t \in \mathbb{R}^d$ représentent les l données d'apprentissage, et $y_t \in \{-1, +1\}$ la classe. Ce problème d'optimisation se résout par la méthode du lagrangien.

Dans le cas où les données ne sont pas linéairement séparables, les contraintes (5) sont relâchées par l'introduction de termes positifs ξ_t . Nous cherchons alors à minimiser :

$$\frac{1}{2} \|w\|^2 + C \sum_{t=1}^l \xi_t, \quad (6)$$

sous les contraintes données pour tout t :

$$\begin{cases} y_t(w \cdot x_t + b) \geq 1 - \xi_t \\ \xi_t \geq 0 \end{cases} \quad (7)$$

où C est une constante choisie par l'utilisateur. Le problème se résout alors de manière similaire au cas linéairement séparable.

Afin de classer un nouvel élément x il suffit d'étudier la fonction de décision donnée par :

$$f(x) = \text{sign}\left(\sum_{t \in SV} y_t \alpha_t^0 x_t \cdot x - b_0\right), \quad (8)$$

où $SV = \{t ; \alpha_t^0 > 0\}$ pour le cas séparable et $SV = \{t ; 0 < \alpha_t^0 < C\}$ pour le cas non séparable, est l'ensemble des vecteurs de support, et $\alpha_t \geq 0$ sont les multiplicateurs de Lagrange.

Dans les cas non linéaire, le principe des SVM est de projeter, par une fonction noyau, les données de départ dans un espace de grande dimension (éventuellement infinie). Ainsi la classification d'un nouvel élément x est donnée par la fonction de décision :

$$f(x) = \text{sign}\left(\sum_{t \in SV} y_t \alpha_t^0 K(x, x_t) - b_0\right) \quad (9)$$

où K est la fonction noyau, dont les plus utilisées sont le noyau polynomial $K(x, x_t) = (x \cdot x_t + 1)^d$, $d \in \mathbb{N}$, et le noyau gaussien $K(x, x_t) = e^{-\gamma \|x - x_t\|^2}$, $\gamma \in \mathbb{R}^+$. Le choix du noyau et l'optimisation des paramètres de celui-ci reste délicat selon l'application.

3.2 Régression floue et crédibiliste par SVM

Nous avons situé cette approche dans la littérature [Laanaya et al. \(2006\)](#). Ainsi elle est novatrice par la prise en compte des contraintes similaires de normalisation des fonctions de croyance et d'appartenance dans le problème de régression multiple. De plus nous proposons ici d'employer une résolution du problème d'optimisation pouvant gérer de grande quantité de données.

Soient les vecteurs d'apprentissage $x_t \in \mathbb{R}^d$ et les fonctions associées $y_t \in \mathbb{R}^N$, où $N = N_c$ le nombre de classes dans le cas des fonctions d'appartenance et $N = 2^{N_c}$ dans le cas des fonctions de masse. Ainsi par la régression multiple linéaire, nous cherchons une fonctionnelle $f = (f_1, \dots, f_N)$ où les f_n sont linéaires, de forme $f_n(x) = w_n \cdot x + b_n$. Nous cherchons à déterminer cette fonctionnelle telle que pour les (x_t, y_t) de la base d'apprentissage $|y_{tn} - w_n \cdot x_t + b_n|$ ne dépasse pas un certain ϵ fixé pour tout n . Nous supposons ainsi que tous les points sont à l'intérieur du cylindre défini par ϵ . Afin de généraliser, nous associons un facteur C pour les points qui sont à l'extérieur du cylindre défini par ϵ . Le problème d'optimisation convexe revient donc à celui exposé dans la section 3.1, et le critère à minimiser est :

$$\frac{1}{2} \sum_{n=1}^N \|w_n\|^2 + C \sum_{n=1}^N \sum_{t=1}^l (\xi_{tn} + \xi_{tn}^*), \quad (10)$$

sous les contraintes données pour tout t et tout n :

$$\left\{ \begin{array}{l} y_{tn} - w_n \cdot x_t - b_n \leq \epsilon + \xi_{tn}, \\ w_n \cdot x_t + b_n - y_{tn} \leq \epsilon + \xi_{tn}^*, \\ \sum_{n=1}^N (w_n \cdot x_t + b_n) = 1, \\ w_n \cdot x_t + b_n \geq 0, \\ w_n \cdot x_t + b_n \leq 1, \\ \xi_{tn}, \xi_{tn}^* \geq 0. \end{array} \right. \quad (11)$$

Le lagrangien est donc donné par :

$$\begin{aligned} L &= \frac{1}{2} \sum_{n=1}^N \|w_n\|^2 + C \sum_{n=1}^N \sum_{t=1}^l (\xi_{tn} + \xi_{tn}^*) - \sum_{n=1}^N \sum_{t=1}^l (\eta_{tn} \xi_{tn} + \eta_{tn}^* \xi_{tn}^*) \\ &- \sum_{n=1}^N \sum_{t=1}^l \alpha_{tn} (\epsilon + \xi_{tn} - y_{tn} + w_n \cdot x_t + b_n) \\ &- \sum_{n=1}^N \sum_{t=1}^l \alpha_{tn}^* (\epsilon + \xi_{tn}^* + y_{tn} - w_n \cdot x_t - b_n) \\ &- \sum_{n=1}^N \sum_{t=1}^l \beta_{tn} (w_n \cdot x_t + b_n) - \sum_{n=1}^N \sum_{t=1}^l \beta_{tn}^* (1 - w_n \cdot x_t - b_n) \end{aligned} \quad (12)$$

Régression floue et crédibiliste par SVM pour la classification des images sonar

$$- \sum_{t=1}^l \gamma_t \left(1 - \sum_{n=1}^N (w_n \cdot x_t + b_n) \right)$$

où les η , α , β et γ sont les multiplicateurs de Lagrange et sont positifs.

Au point selle du lagrangien L , on a pour tout t et tout n , $\partial L / \partial b_n = 0$, $\partial L / \partial w_n = 0$, $\partial L / \partial \xi_{tn} = 0$ et $\partial L / \partial \xi_{tn}^* = 0$. Ainsi :

$$\begin{cases} \sum_{t=1}^l \sigma_{tn} = 0, \\ w_n = \sum_{t=1}^l \sigma_{tn} x_t, \\ \eta_{tn} = C - \alpha_{tn}, \\ \eta_{tn}^* = C - \alpha_{tn}^*, \end{cases} \quad (13)$$

avec $\sigma_{tn} = \alpha_{tn} - \alpha_{tn}^* + \beta_{tn} - \beta_{tn}^* - \gamma_t$.

En intégrant ces équations (13) dans le lagrangien (équation (12)), le problème revient à maximiser :

$$-\frac{1}{2} \sum_{n=1}^N \sum_{t,t'=1}^l \sigma_{tn} \sigma_{t'n} x_t \cdot x_{t'} - \sum_{n=1}^N \sum_{t=1}^l \beta_{tn}^* + \frac{\gamma_t}{N} - \sum_{n=1}^N \sum_{t=1}^l \alpha_{tn}^* (\epsilon + y_{tn}) - \sum_{n=1}^N \sum_{t=1}^l \alpha_{tn} (\epsilon - y_{tn})$$

sous les contraintes :

$$\begin{cases} \sum_{t=1}^l \sigma_{tn} = 0, \\ \alpha_{tn} \in [0, C], \\ \alpha_{tn}^* \in [0, C], \\ \beta_{tn} \geq 0, \\ \beta_{tn}^* \geq 0, \\ \gamma_t \geq 0. \end{cases}$$

Enfin, pour prédire la $n^{\text{ème}}$ sortie \tilde{y}_n , d'un nouvel élément x , on calcule :

$$\tilde{y}_n = \sum_{t=1}^l \sigma_{tn} x_t \cdot x + b_n,$$

où b_n est déduite des conditions de Kuhn, Karush et Tucker :

$$\begin{cases} \alpha_{tn} (\epsilon + \xi_{tn} - y_{tn} + w_n \cdot x_t + b_n) = 0, \\ \alpha_{tn}^* (\epsilon + \xi_{tn}^* + y_{tn} - w_n \cdot x_t - b_n) = 0, \\ (C - \alpha_{tn}) \xi_{tn} = 0, \\ (C - \alpha_{tn}^*) \xi_{tn}^* = 0, \\ \beta_{tn} (w_n \cdot x_t + b_n) = 0, \\ \beta_{tn}^* (1 - w_n \cdot x_t - b_n) = 0. \end{cases}$$

Si pour un t_0 , $\alpha_{t_0 n} \in]0, C[$ alors, $\xi_{t_0 n} = 0$, ainsi $b_n = y_{t_0 n} - w_n \cdot x_{t_0} - \epsilon$, un raisonnement identique sur α^* donne $b_n = y_{t_0 n} - w_n \cdot x_{t_0} + \epsilon$.

La résolution du système d'optimisation de la régression par SVM pour des problèmes de grande dimension nécessite des mémoires de stockage de grande taille. Ainsi, l'application des algorithmes d'optimisation classiques est difficile. Ces limites ont été constatées dans [Laanaya et al. \(2006\)](#). Une solution est d'utiliser des méthodes d'optimisation itératives, où on essaye de résoudre des sous-problèmes du problème principale. Nous avons adapté la résolution par SMO (*Sequential Minimal Optimization*) développée par [Platt \(1998\)](#) pour les machines à vecteurs de support, pour notre problème d'optimisation. Il résout des sous-problèmes de dimension deux d'une manière analytique. Nous pouvons ainsi résoudre des problèmes de grande taille avec une vitesse remarquable.

Si on suppose que la relation entre les x_t et les sorties \tilde{y}_t est non-linéaire, nous pouvons représenter les données de départ en utilisant un noyau. Ainsi le produit scalaire entre les données de la base d'apprentissage peut être donc substitué par un noyau : le produit scalaire $x \cdot x'$ devient $K(x, x')$. Une régression linéaire peut alors s'appliquer dans l'espace de représentation. Pour une observation x , la sortie \tilde{y} se prédit en considérant les N valeurs :

$$\tilde{y}_n = \sum_{t=1}^l \sigma_{tn} K(x, x_t) + b_n.$$

A partir de cette approche de régression sur les fonctions d'appartenance ou les fonctions de croyance, nous obtenons un classifieur en prenant la décision *via* le maximum des fonctions d'appartenance ou le maximum de la probabilité pignistique.

4 Expérimentations

Nous présentons ici l'application de notre approche pour la classification des images sonar. En effet, l'environnement sous-marin lui même est très incertain et les systèmes de mesure sont complexes et imprécis. Il est particulièrement important de classifier les fonds marins pour de nombreuses applications telles que la navigation et la cartographie sous-marine. Nous trouverons plusieurs études sur la classification de images sonar, citons par exemple [Martin et al. \(2004\)](#), [Laanaya et al. \(2005a\)](#), [Laanaya et al. \(2005b\)](#) et [Leblond et al. \(2005\)](#).

Les données à classifier sont ainsi entachées de nombreuses imperfections dues aux bruits de mesure, aux interférences des signaux utilisés pour l'acquisition, aux bruits de chatoiement et à la faune et la flore.

4.1 Base de données

La base de données est constituée de 42 images sonar fournies par GESMA (Groupe d'Etudes Sous-Marines de l'Atlantique) et ont été obtenues à partir d'un sonar Klein 5400 au large des côtes finistériennes. Ces images ont été labellisées à partir d'un logiciel développé spécialement en spécifiant le type de sédiment présent (sable, ride, vase, roche, cailloutis ou ombre) (voir figure 1) et le degré de certitude de l'expert (sûr, moyennement sûr ou non sûr). Parmi ces sédiments, nous avons considéré trois classes distinctes, particulièrement

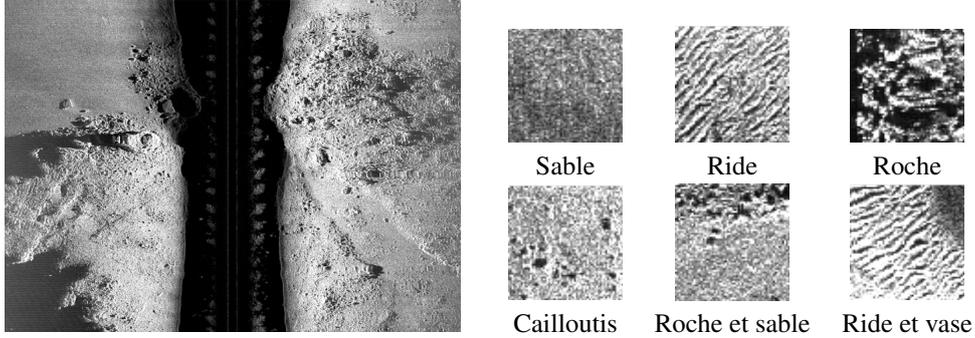


FIG. 1 – Exemple d’image sonar (fournit par le GESMA) et d’imagerie étiquetées.

importantes pour la navigation sous-marine et les sédimentologues. Ainsi la première classe regroupe roche et cailloutis, la deuxième classe les rides et la troisième le sable et les vases.

L’unité de classification retenue est l’imagerie de taille 32×32 pixels (soit environ 640×640 cm).

4.1.1 Extraction de paramètres

Afin de réduire les problèmes de représentativité des imagerie qui comportent plus d’un sédiment et les problèmes liés à l’évaluation (cf. [Martin et al. \(2006\)](#)), nous ne considérons ici que les imagerie homogènes (imagerie avec un seul type de sédiment). Nous avons ainsi 31957 imagerie.

Nous avons calculé sur ces imagerie six paramètres extraits à partir des matrices de co-occurrence calculés sur les imagerie [Martin et al. \(2004\)](#). Les matrices de cooccurrence C_d sont calculées en comptant les occurrences identiques de niveaux de gris entre deux pixels contigus dans une direction d donnée. Quatre directions sont considérées : 0, 45, 90 et 135 degrés. Dans ces quatre directions six paramètres d’Haralick sont calculés : l’homogénéité, le contraste, l’entropie, la corrélation et l’uniformité. L’homogénéité qui a une valeur élevée pour des images uniformes ou possédant une texture périodique dans la direction d est donnée par :

$$\sum_{i=1}^{N_G} \sum_{j=1}^{N_G} C_d^2(i, j) \quad (14)$$

où N_G est le niveau de gris des imagerie.

L’estimation du contraste est donnée par :

$$\frac{1}{N_G - 1} \sum_{k=0}^{N_G-1} k^2 \sum_{i, j=1, |i-j|=k}^{N_G} C_d(i, j) \quad (15)$$

L'entropie qui a de faibles valeurs s'il y a peu de probabilités de transition élevées dans C_d , est définie par :

$$1 - \sum_{i=1}^{N_G} \sum_{j=1}^{N_G} p(i, j) \log(C_d(i, j)) \quad (16)$$

La corrélation entre les lignes et les colonnes de la matrice est donnée par :

$$\sum_{i=1}^{N_G} \sum_{j=1}^{N_G} \frac{(i - \mu_x)(j - \mu_y)C_d(i, j)}{\sigma_x \sigma_y} \quad (17)$$

où $\mu_x, \sigma_x, \mu_y, \sigma_y$ représentent respectivement les moyennes et écart-types des distributions marginales des éléments de la matrice de cooccurrence.

La directivité qui définit l'existence d'une direction privilégiée de la texture est calculée par :

$$\sum_{i=1}^{N_G} C_d(i, i) \quad (18)$$

L'uniformité qui caractérise la proportion d'un même niveau de gris est donnée par :

$$\sum_{i=1}^{N_G} C_d(i, i)^2 \quad (19)$$

Nous avons moyenné ces paramètres selon les quatre directions d , ainsi chaque imagerie est représentée uniquement par six paramètres.

4.1.2 Modélisation des fonctions floues et crédibilistes

Nous avons utilisé l'approche de Keller et al. (1985) pour calculer la fonction d'appartenance des vecteurs d'apprentissage que nous utiliserons pour l'apprentissage du SVM flou, et l'approche de Denœux (1995) pour estimer les fonctions de masses que nous utiliserons pour l'apprentissage du SVM crédibiliste.

L'approche de Keller et al. (1985) est celle d'un k -plus proches voisins flou. Les fonctions d'appartenance d'un vecteur d'apprentissage x_t sont estimées dans un premier temps par :

$$\mu_i(x_t) = \frac{k_i(x_t)}{k_f}, \quad (20)$$

où k_f est le nombre de plus proches voisins choisi pour le voisinage flou V_{K_f} et $k_i(x_t) = |C_i \cap V_{K_f}(x_t)|$. Dans un second temps, nous calculons la fonction d'appartenance pour un vecteur x à classifier :

$$\mu_i(x) = \frac{\sum_{t=1}^l \frac{\mu_i(x_t)}{\|x - x_t\|^2}}{\sum_{t=1}^l \frac{1}{\|x - x_t\|^2}}. \quad (21)$$

La norme employée est ici la norme euclidienne.

La classe d'appartenance de x est ensuite décidée de manière classique comme la classe donnant le maximum des fonctions d'appartenance prédites par notre régression.

L'approche de [Dencœur \(1995\)](#) calcule une estimation des fonctions de masses à partir d'un modèle de distance :

$$\begin{cases} m_k(C_i|x^{(t,k)})(x) = \alpha_i e^{\gamma_i d^2(x, x^{(t,k)})} \\ m_k(\Theta|x^{(t,k)})(x) = 1 - \alpha_i e^{\gamma_i d^2(x, x^{(t,k)})} \end{cases} \quad (22)$$

où C_i est la classe associée à $x^{(t,k)}$, qui sont les k vecteurs d'apprentissage les plus proches de la valeur x et la distance employée est la distance euclidienne. α_i et γ_i sont des coefficients d'affaiblissement, et de normalisation. Les k fonctions de masse ainsi calculées pour chaque x sont combinées par la règle orthogonale normalisée de Dempster-Shafer. Cette règle est donnée pour deux experts et pour tout $A \in 2^\Theta$, $A \neq \emptyset$ par :

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B)m_2(C)}, \quad (23)$$

et $m(\emptyset) = 0$.

La décision est ensuite prise par le maximum sur les fonctions de masse prédites par notre régression. Dans ce cas, il est équivalent au maximum de probabilité pignistique car les seuls éléments focaux sont les singletons et l'ignorance.

4.2 Résultats

Nous avons effectué un tirage aléatoire de 3000 imagettes homogènes sur toute la base de données (31957 imagettes), ainsi la base d'apprentissage contient des effectifs différents pour les trois classes : 15.53% des imagettes contiennent du roche et cailloutis, 11.85% sont des imagettes rides et les 72.62% restants sont du sable et de la vase. La base de test est constituée de 1000 imagettes choisies de façon aléatoire. Nous avons répété cette opération 10 fois afin d'obtenir des estimations plus fiables des taux de classification.

Nous avons comparé la classification fondée sur les machines à vecteurs de support donnée par le logiciel *libSVM* de [Chang et Lin \(2001\)](#) et une version modifiée de ce dernier qu'on a développée pour intégrer notre approche.

Les matrices de confusion normalisées obtenues par le SVM classique (avec les paramètres par défaut de *libSVM* : noyau gaussien avec $\gamma = 1$ et $C=1$), SVM crédibiliste et SVM flou avec un noyau gaussien, avec $\gamma = 1$, $C = 1$ et $\epsilon = 0.1$, sont données ci-dessous.

SVM classique	SVM crédibiliste	SVM flou
$\begin{pmatrix} 61.62 & 25.26 & 13.12 \\ 23.08 & 62.59 & 14.32 \\ 4.99 & 10.97 & 84.04 \end{pmatrix}$	$\begin{pmatrix} 70.59 & 15.73 & 13.68 \\ 16.81 & 70.28 & 12.91 \\ 6.42 & 13.91 & 79.66 \end{pmatrix}$	$\begin{pmatrix} 68.99 & 16.60 & 14.41 \\ 16.18 & 66.09 & 17.74 \\ 5.85 & 7.59 & 86.56 \end{pmatrix}$

Nous avons ainsi obtenu un taux de $78.02 \pm 5.14\%$ pour un SVM classique et des taux de $77.14 \pm 5.22\%$ pour le SVM crédibiliste et $81.40 \pm 4.83\%$ pour le SVM flou. Dans le cas du

SVM classique nous avons obtenu un vecteur de bonne classification de $[61.62 \pm 4.85 \ 62.59 \pm 5.52 \ 84.04 \pm 1.69]$, $[70.59 \pm 4.54 \ 70.28 \pm 5.22 \ 79.66 \pm 1.83]$ par la régression crédibiliste par SVM et $[68.99 \pm 4.61 \ 66.09 \pm 5.6 \ 86.56 \pm 1.55]$ par la régression floue par SVM. Les vecteurs d'erreurs sont donnés par : $[26.21 \ 27.76 \ 14.84]$ pour le SVM classique, $[20.51 \ 22.27 \ 16.82]$ pour le SVM crédibiliste et $[21.01 \ 23.00 \ 14.76]$ pour le SVM flou. En comparaison avec l'approche classique des SVM, nous avons ainsi une amélioration significative de la classification des trois classes pour la régression floue issue des SVM et une amélioration significative avec la régression crédibiliste issue des SVM pour les deux premières classes (la classe des roches et des cailloutis, et la classe des rides) ces deux dernières classes sont particulièrement difficiles à classifier du fait de leur faible représentativité. Notons que c'est l'approche crédibiliste qui donne les meilleurs résultats pour ces deux classes. Ainsi la nouvelle approche a apporté une amélioration pour la classification des différentes classes des images sonar qui sont particulièrement difficile à caractérisées.

5 Conclusion

Nous avons proposé dans ce papier une nouvelle résolution de l'approche de régression floue et crédibiliste à partir de machines à vecteurs de support pour la classification précédemment introduite. Les résultats obtenus sur les images sonar ont montré l'intérêt de cette approche. En particulier l'approche crédibiliste donne de très bons résultats sur des données faiblement apprises, alors que l'approche floue permet d'avoir une meilleure classification pour chaque classe considérée.

Nous n'avons donné ici que des résultats en utilisant des valeurs empiriques pour les paramètres (C , ϵ et γ). Le réglage de ces paramètres peut se faire en utilisant les algorithmes génétiques par exemple, il est possible aussi d'intégrer l'optimisation de ces constantes dans le problème d'optimisation générale des SVM pour la régression.

Références

- Chang, C. C. et C. J. Lin (2001). Libsvm : a library for support vector machines. *Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>*.
- Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* 83, 325–339.
- Denœux, T. (1995). A k -nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans* 25(5), 804–813.
- Dubois, D. et H. Prade (1987). *Théorie des possibilités*. Masson.
- Keller, J. M., M. Gray, et J. Givens (1985). A fuzzy k -nn neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics* 15, 580–585.
- Laanaya, H., A. Martin, D. Aboutajdine, et A. Khenchaf (20-23 June 2005a). A new dimensionality reduction method for seabed characterization : Supervised curvilinear component analysis. *IEEE OCEANS'05 EUROPE, Brest, France*.

Régression floue et crédibiliste par SVM pour la classification des images sonar

- Laanaya, H., A. Martin, A. Khenchaf, et D. Aboutajdine (15-18 March 2005b). Feature selection using genetic algorithm for sonar images classification with support vector machines. *European Conference on Propagation and Systems, Brest, France*.
- Laanaya, H., A. Martin, A. Khenchaf, et D. Aboutajdine (19-20 Octobre 2006). Classification par régression floue et crédibiliste à base de machines à vecteurs de support. *LFA 2006, Toulouse, France*.
- Leblond, I., M. Legris, et B. Solaiman (20-23 June 2005). Use of classification and segmentation of sidescan sonar images for long term registration. *IEEE Oceans'05 Europe, Brest, France*.
- Martin, A. (Novembre 2005). Fusion de classifieurs pour la classification d'images sonar. *RNTI Extraction des connaissances : Etat et perspectives, 259-268*.
- Martin, A., H. Laanaya, et A. Arnold-Bos (2006). Evaluation for uncertain image classification and segmentation. *Pattern Recognition 39*.
- Martin, A., G. Sevellec, et I. Leblond (21-22 October 2004). Characteristics vs decision fusion for sea-bottom characterization. *Colloque Caractérisation in-situ des fonds marins, Brest, France*.
- Platt, J. (1998). Sequential minimal optimization : A fast algorithm for training support vector machines. *Microsoft Research Technical Report MSR-TR-98-14*.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.
- Smets, P. (1990). Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence 5, 29-39*.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wesley and Sons.
- Zadeh, L. A. (1965). Fuzzy sets. *Information Control 8, 338-353*.

Summary

The sonar image classification is of great importance, for underwater navigation or for seabed cartography. Indeed, the sonar is more suitable than optical captors for seabed imagery. The classification of such kind of data encounters several difficulties due to the imprecisions and uncertainties present on these data. Many approaches were proposed without giving good results, they do not take into account the data imperfections. To model this kind of data, it is judicious to use the uncertain theories like the fuzzy subsets theory or the belief function theory. The support vector machines are more and more used for automatic classification due to their simplicity and their generalization capacities. Thus, it is possible to propose an approach that take into account these imprecisions and uncertainties. The regression by SVM approach that we have proposed model these imperfections. We propose here an application of this new approach for particularly complex real data in the framework of sonar image classification.