

Approche bayésienne et fonctions de croyance continues pour la classification

Anthony Fiche et Arnaud Martin
Anthony.Fiche, Arnaud.Martin@ensieta.fr

ENSIETA - E3I2 EA3876
Brest, France

Motivation

Classification

Estimation des distributions

Classifieur bayésien

Théorie des fonctions de croyance

Fonctions de croyance discrètes

Fonctions de croyance continues

Résultats

Cas mono-dimensionnel

Cas multi-dimensionnel

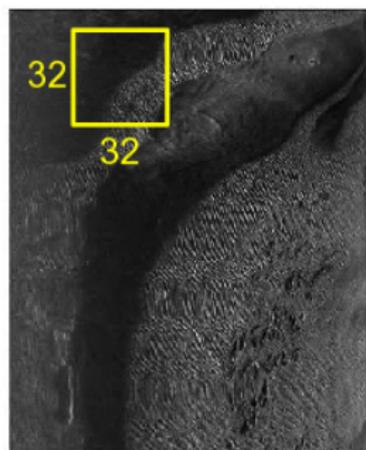
Motivation : Classification d'images sonar

Motivation

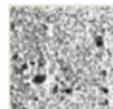
Classification

Théorie des fonctions de croyance

Résultats



Rides



Cailloutis



Roche

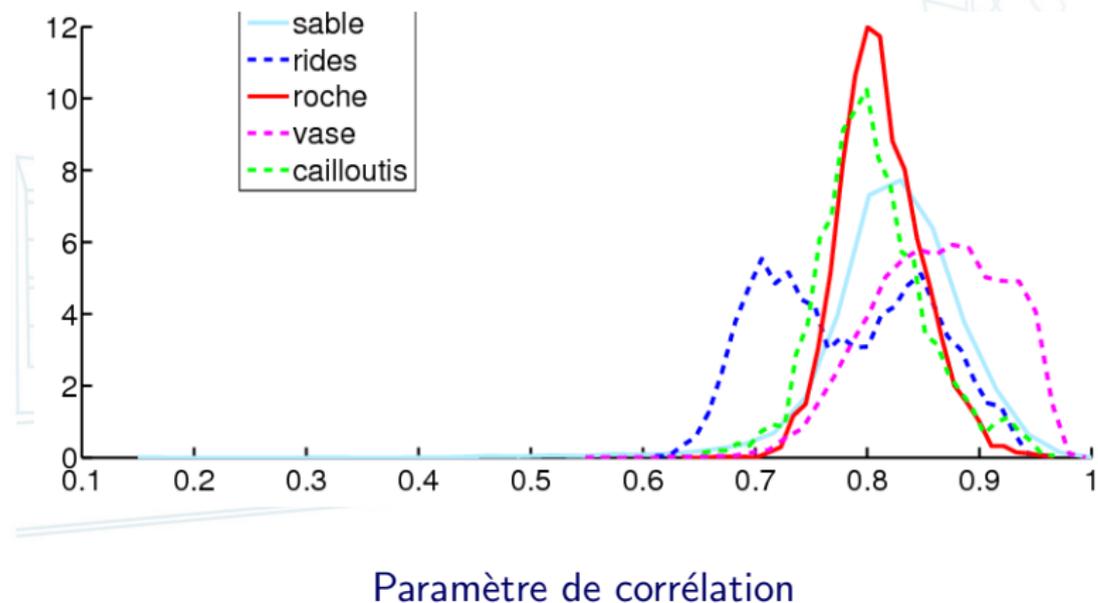


Sable



Vase

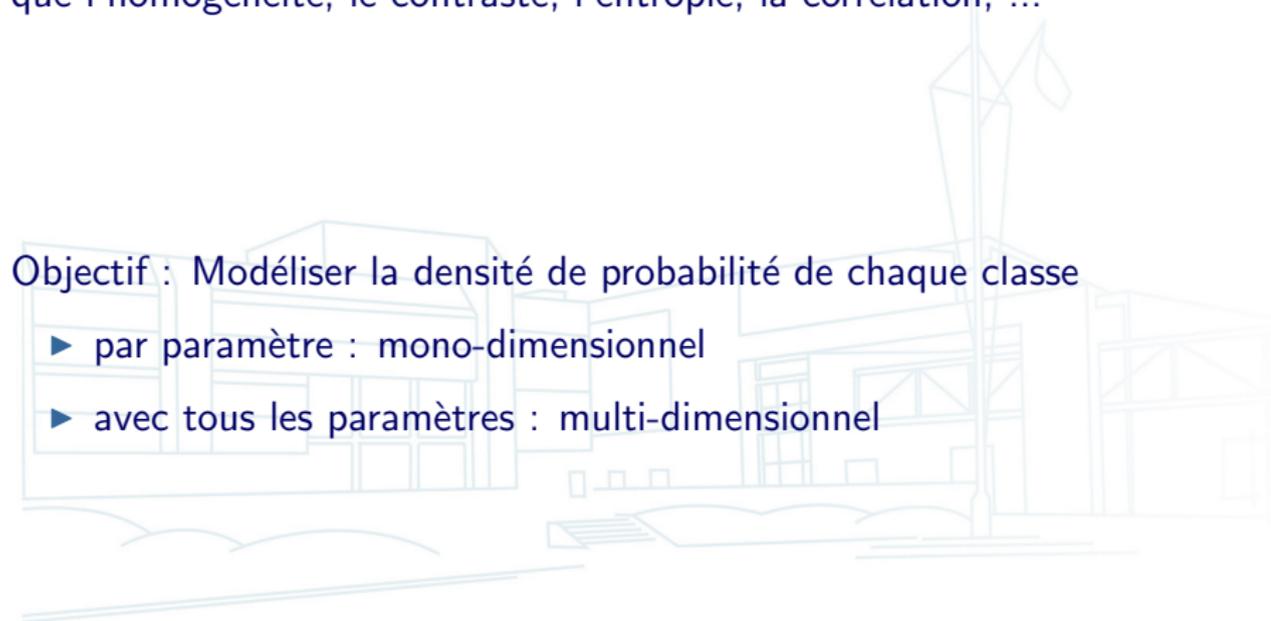
Description des imagerie par différents paramètres de texture tels que l'homogénéité, le contraste, l'entropie, la corrélation, ...



Description des imageries par différents paramètres de texture tels que l'homogénéité, le contraste, l'entropie, la corrélation, ...

Objectif : Modéliser la densité de probabilité de chaque classe

- ▶ par paramètre : mono-dimensionnel
- ▶ avec tous les paramètres : multi-dimensionnel



Chaque distribution peut être modélisée par une somme pondérée de gaussiennes

$$f_j(\mathbf{x}) = \sum_{k=1}^K W_k^j (2\pi)^{-\frac{d}{2}} \det(\Sigma_k^j)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_k^j)^T (\Sigma_k^j)^{-1} (\mathbf{x} - \mu_k^j)\right)$$

Estimation de W^j les fonctions poids, μ^j les moyennes et Σ^j les matrices de covariances par l'algorithme EM par classe C_j

$K=5$ fixé est généralement suffisant

- ▶ Estimation des probabilités *a priori* par :

$$p(\mathbf{x}/C_j) = f_j(x) \quad (1)$$

avec C_j représentant une classe j possible.

- ▶ Application du théorème de Bayes :

$$p(C_j/\mathbf{x}) = \frac{p(\mathbf{x}/C_j)p(C_j)}{\sum_{j=1}^n p(\mathbf{x}/C_j)p(C_j)} \quad (2)$$

- ▶ Décision donnée par $\operatorname{argmax}(p(C_j/\mathbf{x}))$.

- ▶ $\Theta = \{C_1, C_2, \dots, C_n\}$ correspondant à l'ensemble des classes possibles
- ▶ Les fonctions de masse sont définies sur 2^Θ et à valeurs dans $[0, 1]$
- ▶ Condition de normalisation : $\sum_{A \in 2^\Theta} m(A) = 1$
- ▶ Combinaison de Dempster :

$$m(X) = \frac{1}{1 - \sum_{Y_1 \cap Y_2 = \emptyset} m_1(Y_1)m_2(Y_2)} \sum_{Y_1 \cap Y_2 = X} m_1(Y_1)m_2(Y_2) \quad (3)$$

- ▶ Décision par la probabilité pignistique :

$$BetP(C_j) = \sum_{A \subseteq \Theta, C_j \in A} \frac{1}{|A|} \frac{m(A)}{(1 - m(\emptyset))} \quad (4)$$

Les fonctions de masse deviennent des densités de masse définies sur des intervalles $[a, b]$ de \mathbb{R} par $m([a, b]) = f(a, b)$, où f est une densité de probabilité de $\{(y_1, y_2) \in \mathbb{R}^2 | y_1 \leq y_2\}$ dans $[0, +\infty]$.
Le degré de plausibilité sur un intervalle $[a, b]$:

$$pl([a, b]) = \int_{y_1=-\infty}^{y_1=b} \int_{y_2=\max(a, y_1)}^{y_2=+\infty} f(y_1, y_2) dy_2 dy_1. \quad (5)$$

À une densité de masse correspond une densité de probabilité et donc une probabilité pignistique

À une probabilité pignistique correspond un ensemble de fonctions de masse

Un critère pour choisir : la fonction de masse de moindre engagement

À une distribution de probabilité multimodale modélisée par un mélange de gaussiennes, correspond une densité de masse caractérisée par sa fonction de plausibilité.

La fonction de plausibilité conditionnellement à la classe C_j au point $\mathbf{x} \in \mathbb{R}^d$ est alors définie par Caron *et al.*, 2006 par :

$$pl_j(\mathbf{x}) = 1 - \sum_{\kappa=1}^K W_{\kappa}^j \times F_{d+2} \left[{}^t(\mathbf{x} - \mu_{\kappa}^j)(\Sigma_{\kappa}^j)^{-1}(\mathbf{x} - \mu_{\kappa}^j) \right],$$

avec F_{d+2} la fonction de répartition de la loi du χ^2 avec $d+2$ degrés de liberté (d représentant la dimension du vecteur \mathbf{x}).

À partir des plausibilités, on détermine les fonctions de masse en appliquant le théorème généralisé de Bayes :

$$m[\mathbf{x}](A) = \prod_{C_j \in A} pl_j(\mathbf{x}) \prod_{C_j \in A^c} (1 - pl_j(\mathbf{x})) \quad (6)$$

où $m[\mathbf{x}](A)$ représente la fonction de masse connaissant l'observation \mathbf{x} .

La décision peut alors être prise par la probabilité pignistique.

8 paramètres de texture choisis ($d=8$)

Protocole

- ▶ Tirage aléatoire de 5000 imagettes sur 30279 données
- ▶ 1/2 pour la base d'apprentissage et 1/2 pour la base de test
- ▶ 10 tirages effectués

Puis variation de la proportion entre données d'apprentissage et de test

Estimation des distributions de chaque classe pour chaque paramètre avec combinaison

- ▶ des probabilités pour l'approche bayésienne

$$p(C_j/(x_1, \dots, x_d)) = \frac{\prod_{i=1}^d p(x_i/C_j)p(C_j)}{\sum_{j=1}^n \prod_{i=1}^d p(x_i/C_j)p(C_j)} \quad (7)$$

- ▶ des fonctions de croyance

$$m[x_i](A) = \prod_{C_j \in A} p_{lj}(x_i) \prod_{C_j \in A^c} (1 - p_{lj}(x_i)) \quad (8)$$

puis combinaison par la règle de Dempster et décision par la probabilité pignistique

- classifieur bayésien : 65.80 % avec un intervalle de confiance à 95 % de [63.94;67.66]

		Classe estimée				
		roche	ride	vase	cailloutis	sable
Classe réelle	roche	2583	3	9	1	2
	ride	0	1791	741	151	211
	vase	0	279	8096	550	2410
	cailloutis	0	120	189	508	336
	sable	0	434	2621	492	3473

- ▶ classifieur bayésien : 65.80 %, IC=[63.94;67.66]
- ▶ approche crédibiliste : 64.49 % avec un intervalle de confiance à 95 % de [62.61;66.37]

		Classe estimée				
		roche	ride	vase	cailloutis	sable
Classe réelle	roche	2581	7	7	1	2
	ride	0	1976	561	166	191
	vase	0	346	7375	707	2907
	cailloutis	0	134	134	556	329
	sable	0	485	2249	650	3636

- ▶ classifieur bayésien : 65.80 %, IC=[63.94;67.66]
- ▶ approche crédibiliste : 64.49 %, IC=[62.61;66.37]
- ▶ classifieur k-ppv crédibiliste (Denœux, 1995) : 77.09 % avec un intervalle de confiance à 95 % de [75.44;78.74]

		Classe estimée				
		roche	ride	vase	cailloutis	sable
Classe réelle	roche	2598	0	0	0	0
	ride	0	2578	238	17	61
	vase	0	61	9373	45	1856
	cailloutis	0	44	317	487	305
	sable	0	67	2684	32	4237

Confusion entre les classes vase et sable

Estimation des distributions de chaque classe dans l'espace des paramètres.

- Pour l'approche bayésienne

$$p(C_j/\mathbf{x}) = \frac{p(\mathbf{x}/C_j)p(C_j)}{\sum_{j=1}^n p(\mathbf{x}/C_j)p(C_j)} \quad (7)$$

- Pour les fonctions de croyance

$$m[\mathbf{x}](A) = \prod_{C_j \in A} p_{j}(\mathbf{x}) \prod_{C_j \in A^c} (1 - p_{j}(\mathbf{x})) \quad (8)$$

puis décision par la probabilité pignistique

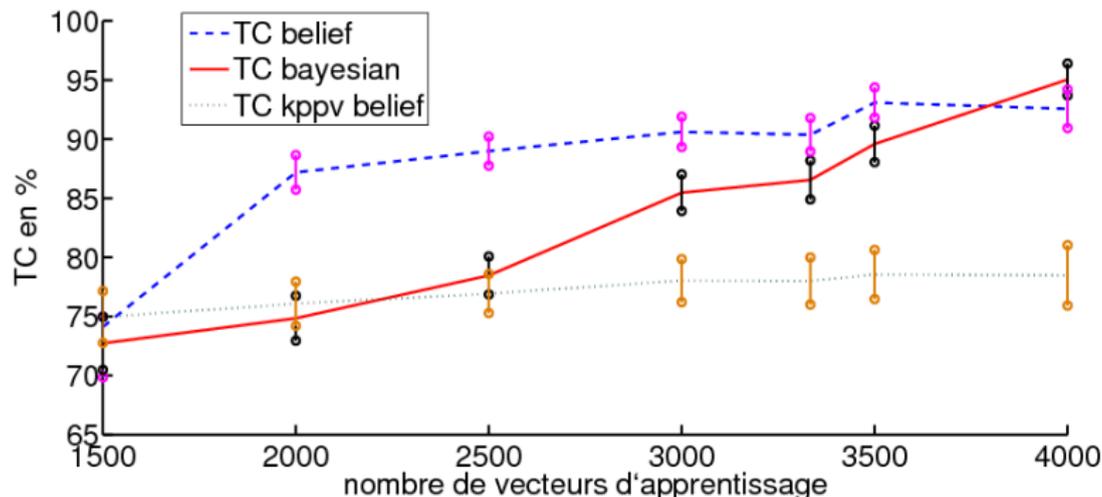
- classifieur bayésien : 83.57 % avec un intervalle de confiance à 95 % de [82.12;85.02]

		Classe estimée				
		roche	ride	vase	cailloutis	sable
Classe réelle	roche	2088	16	23	485	3
	ride	0	2626	4	320	2
	vase	0	1	9731	1104	526
	cailloutis	0	26	41	948	196
	sable	0	0	650	710	5500

- ▶ classifieur bayésien : 83.57 % IC=[82.12;85.02]
- ▶ approche crédibiliste : 87.98 % avec un intervalle de confiance à 95 % de [86.90;89.26]

		Classe estimée				
		roche	ride	vase	cailloutis	sable
Classe réelle	roche	2436	0	7	172	0
	ride	35	2661	5	247	4
	vase	61	1	9717	460	1123
	cailloutis	36	5	9	1132	29
	sable	35	0	514	261	6050

Moins de confusion entre les classes vase et sable



Approche bayésienne connue pour être optimale si les probabilités *a priori* sont connues

Entre 2000 et 3500 vecteurs d'apprentissage (mauvais apprentissage), l'approche avec les fonctions de croyance continues est significativement meilleure que les 2 autres approches.

- ▶ Estimation des distributions multi-dimensionnelles des caractéristiques de texture permet de considérer les corrélations entre ces caractéristiques.
- ▶ Approche avec les fonctions de croyance continues donne des résultats de classification significativement meilleurs que l'approche bayésienne pour la caractérisation des fonds marins à partir d'images sonar
- ▶ Approche avec les fonctions de croyance continues est plus robuste à un mauvais apprentissage
- ▶ Problème de l'estimation des fonctions de masse à partir de distributions multi-modales : voir l'article de Pierre-Emmanuel Doré présenté hier