A comparison between a Bayesian approach and a method based on continuous belief functions for pattern recognition

Anthony Fiche*, Arnaud Martin**, Jean-Christophe Cexus* and Ali Khenchaf*

Abstract The theory of belief functions in discrete domain has been employed with success for pattern recognition. However, the Bayesian approach performs well provided that once the probability density functions are well estimated. Recently, the theory of belief functions has been more and more developed to the continuous case. In this paper, we compare results obtained by a Bayesian approach and a method based on continuous belief functions to characterize seabed sediments. The probability density functions of each feature of seabed sediments are unimodal and estimated from a Gaussian model and compared with an α-stable model.

1 Introduction

The theory of belief functions, introduced by Dempster [4] and formalized by Shafer [13], has found in these recent years many applications especially in pattern recognition. The Bayesian approach performs well provided that once the probability density functions (pdfs) are well estimated. However, the Bayesian approach introduces the notion of prior probabilities. It is possible to avoid this problem by using the theory of belief functions. The theory of belief functions is often presented as an extension of the probability theory. However, the theory of belief functions is not often been used in problem of estimation. Recently, many papers [5, 16] have been proposed to extend the theory of belief functions in discrete domain to continuous domain. In [1, 11], the authors proposed solutions to solve problem of pattern recognition from continuous belief functions.

* Anthony Fiche, Jean-Christophe Cexus, Ali Khenchaf
ENSTA Bretagne, 2 rue François Verny, 29806 Brest Cedex 9, France, e-mail: {anthony.fiche,jean-chirstophe.cexus,ali.khenchaf}@ensta-bretagne.fr

** Arnaud Martin
UMR 6074 IRISA, Université de Rennes 1, rue Édouard Branly BP 30219, 22302 Lannion Cedex, France, e-mail: arnaud.martin@univ-rennes1.fr

We propose a supervised classification of seabed sediments based on a Bayesian approach and compared with a method based on the theory of continuous belief functions. The pdfs of each seabed sediment are bell-shaped [1]. Many distributions can have this property: Gaussian, Weibull, K .... However, the pdfs from seabed sediments have the properties of skewness and heavy tails. A distribution is said to have heavy tails if the tails decays slower than the tail of the Gaussian distribution. Therefore, the property of skewness means that it is impossible to find a mode where the curve is symmetric. It is possible to consider these contraints from α-stable distribution. Consequently, we use two models of estimation during the classification: Gaussian and α-stable distributions.

The remainder of this paper is organized in the following manner. In section 2, we introduce the theory of continuous belief functions. In section 3, we describe the data set, the model of estimation and compare results between the Bayesian approach and the method based on continuous belief functions.

2 Background on continuous belief functions

2.1 Basic belief density

Recently, Smets [16] extended the definition of belief functions to the set of reals $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and basic belief assignment (bbd) are only attributed to intervals of $\overline{\mathbb{R}}$. Let us consider $\mathscr{I} = \{[x,y],(x,y],[x,y),(x,y); x,y \in \overline{\mathbb{R}}\}$ as a set of closed, half-opened and opened intervals of $\overline{\mathbb{R}}$. A bbd $m^{\mathscr{I}}(x,y)$ linked to a specific *pdf* is a non negative function on $\mathscr{I}$ such that $m^{\mathscr{I}}(x,y) = 0$ if the interval defined by (x,y) is not closed in $\mathscr{I}$. The closed intervals $[x,y]$ which satisfy the relation $m^{\mathscr{I}}(x,y) > 0$ are called focal elements. From the definition of the bbd, it is possible to define others belief functions [16] as in the discrete case credibility function $bel^{\overline{\mathbb{R}}}$, plausibility function $pl^{\overline{\mathbb{R}}}$ and communality function $q^{\overline{\mathbb{R}}}$. A bbd is said to be "consonant" when focal elements are nested. Focal elements I_u can be labeled as an index u such that $I_u \subseteq I'_u$ with $u' > u$.

2.2 Least commitment bbd induced by an unimodal pdf

The definition of pignistic probability [14] for $a < b$ is:

$$Betf([a,b]) = \int_{x=-\infty}^{x=+\infty} \int_{y=x}^{y=+\infty} \frac{\min(y,b) - \max(x,a)}{y-x} m^{\mathscr{I}}(x,y)dxdy \qquad (1)$$

[1] *i.e.* the probability density function is unimodal with a mode μ, continuous and strictly monotonous increasing (decreasing) at left (right) of the mode

It is possible to calculate pignistic probabilities to have basic belief densities. However, many basic belief densities exist for one same pignistic probability. To resolve this issue, we can use the consonant basic belief density. This definition is used to apply the least commitment principle [15], which consists in choosing the least informative belief function when a belief function is not totally defined and is only known to belong a family of functions. The function $Betf$ can be induced by a set of isopignistic belief functions $\mathscr{B}\mathbf{iso}(Betf)$. Many papers [12, 16, 1] deal with the particular case of continuous belief functions with nested focal elements. The least commitment principle proposes to choose the least informative mass function, *i.e.* the mass functions must be ordered. An order relation is given in equation 2, but there are other order relations.

$$(\forall A \subseteq \overline{\mathbb{R}}, q_1^{\overline{\mathbb{R}}}(A) \leq q_2^{\overline{\mathbb{R}}}(A)) \Rightarrow (m_1^{\overline{\mathbb{R}}} \leq m_2^{\overline{\mathbb{R}}}) \tag{2}$$

For example, Smets [16] proved that the basic belief assignment $m^{\overline{\mathbb{R}}}$ attributed to an interval $I = [x, y]$ with $y > \mu$ related to a bell-shaped pignistic probability function with a mode μ is determined by [2]:

$$m^{\overline{\mathbb{R}}}([x,y]) = \theta(y)\delta(x - \gamma(y)) \tag{3}$$

with $x = \gamma(y)$ satisfying $Betf(\gamma(y)) = Betf(y)$ and $\theta(y)$:

$$\theta(y) = (\gamma(y) - y)\frac{dBetf(y)}{dy} \tag{4}$$

The build basic belief assignment $m^{\overline{\mathbb{R}}}$ is consonant and belongs to the set $\mathscr{B}\mathbf{iso}(Betf)$.

2.3 *Link between pignistic probability function and plausibility function in $\overline{\mathbb{R}}$*

The available information are the conditioned pignistic density $Betf[C_i]$ with $C_i \in \Theta$, where Θ is called the frame of discernement. The function $Betf[C_i]$ is supposed to be bell-shaped. The plausibility function from a bbd $m^{\overline{\mathbb{R}}}$ with $x > \mu$ is obtained by an integral of equation (4) between $[x, +\infty[$:

$$pl^{\overline{\mathbb{R}}}[C_i](I) = \int_x^{+\infty} (\gamma(t) - t)\frac{dBetf(t)}{dt}dt \tag{5}$$

By assuming that $Betf$ is symmetrical, an integration by parts can simplified the equation (5):

$$pl^{\overline{\mathbb{R}}}[C_i](I) = 2(x - \mu)Betf(x) + 2\int_x^{+\infty} Betf(t)dt \tag{6}$$

[2] δ refers to the Dirac's measure.

We can calculate $\int_x^{+\infty} Betf(t)dt$ in a particular case of symmetrical $Betf$ by using the Chasles' theorem. Consequently, the equation (6) can be simplified [7]:

$$pl^{\overline{\mathbb{R}}}[C_i](I) = 2(x-\mu)pdf(x) + 2(1-cdf(x)) \tag{7}$$

If $x < \mu$, we use the variable modification $x = 2\mu - y$. In the particular case of Gaussian pdf, Caron *et al.* [1] propose the plausibility function:

$$pl^{\overline{\mathbb{R}}}[C_i](I) = 1 - F_3((x-\mu)(\Sigma)^{-1}(x-\mu)) \tag{8}$$

The function F_{d+2} is a cumulative density function of the χ^2 distribution with 3 degrees of freedom, μ the mean and Σ the standard-deviation of a Gaussian pdf. It is difficult to generalize in the case of asymmetric pdf because the function $\gamma(y) = x$ satisfying $Betf(\gamma(y)) = Betf(y)$ is not trivial. The plausibility function related to an interval $I_1 = [x_1, y_1]$ is defined by the area defined under the α-cut such as $\alpha = Betf(x_1)$ (Figure 1):

$$pl^{\overline{\mathbb{R}}}[C_i](I_1) = \int_{-\infty}^{x_1} Betf(t)dt + (y_1 - x_1)Betf(x_1) + \int_{y_1}^{+\infty} Betf(t)dt \tag{9}$$

In general, we know only one point y_1. We estimate numerically x_1 such that $pdf(y_1) = pdf(x_1)$. Finally, the plausibility function related to the interval I_1 is:

$$pl^{\overline{\mathbb{R}}}[C_i](I_1) = 1 + cdf(x_1) - cdf(y_1) + (y_1 - x_1)pdf(x_1) \tag{10}$$

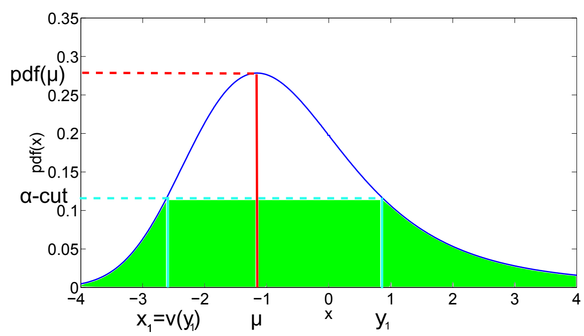

Fig. 1: Plausibility function in the case of asymmetric pdf.

In classification, we assume that we have several pdfs associated to a class C_i. We can calculate a plausibility function related to its pdfs by using the least commitment principle. Several plausibility functions can be combined by using the general Bayes theorem [15, 3] to calculate mass functions allocated to A of an interval I:

$$m^{\overline{\mathbb{R}}}[x](A) = \prod_{C_j \in A} pl_j(x) \prod_{C_j \in A^c} (1 - pl_j(x)) \tag{11}$$

3 Application to pattern recognition

3.1 Data set

The data set are picked up by the Service Hydrographique et Océanique de la Marine (SHOM) with the Daurade Autonomous Underwater Vehicle (AUV) from the Atlas DESO 35 mono-beam echo sounder in the Mediterranean Sea off the coast of Toulon. Raw data represents an echo signal amplitude according to time. These data are processed to obtain some features, which have been normalized between [0,1] (defined and used in the Quester Tangent Corporation (QTC) software [2]). The frame of discernment is $\Theta = \{\text{rock}, \text{sand}, \text{silt}\}$, with 6017 samples from rock, 7338 samples from sand and 4853 samples from silt. From the data, we choose the features called the “third quantile calculated on echo signal amplitude” and the “75th quantile calculated on cumulative energy”. The authors would like to thank the Service Hydrographique et Océanique de la Marine (SHOM) for the data and G. Le Chenadec for his advices about the data.

3.2 Models of estimation

We use two models of estimation: Gaussian and α-stable distributions. The Gaussian distribution is a particular case of α-stable distribution [10]. Several equivalent definitions have been suggested in the literature to parametrize an α-stable distribution from its characteristic function [17, 18]. Zolotarev [18] proposed the following:

$$\phi(t) = \begin{cases} \exp(it\nu - |\gamma t|^{\alpha}[1 + i\beta \tan(\dfrac{\pi\alpha}{2}) sign(t)(|t|^{1-\alpha} - 1)]) & \text{if } \alpha \neq 1 \\ \exp(it\nu - |\gamma t|[1 + i\beta \dfrac{2}{\pi} sign(t) \log|t|]) & \text{if } \alpha = 1 \end{cases} \tag{12}$$

with $\alpha \in]0,2]$ is the characteristic exponent, $\beta \in [-1,1]$ is the skewness parameter, $\gamma \in \mathbb{R}^{+*}$ represents the scale parameter and $\nu \in \mathbb{R}$ is the location parameter. In general, the notation $S_\alpha(\beta, \gamma, \nu)$ refers to α-stable distributions.
The α-stable pdf, noticed pdf_α, is obtained by calculating the Fourier transform of its characteristic function (*cf.* [9] for the implementation). An α-stable random variable can be estimated by using methods based on quantiles or moments. For the rest of the paper, we use a method based on moments developed by Koutrouvelis [8] in order to estimate the parameters α, β, γ and ν.

To implement the classification with the belief functions, we firstly need to estimate the parameters of distribution from the learning base. For each feature of

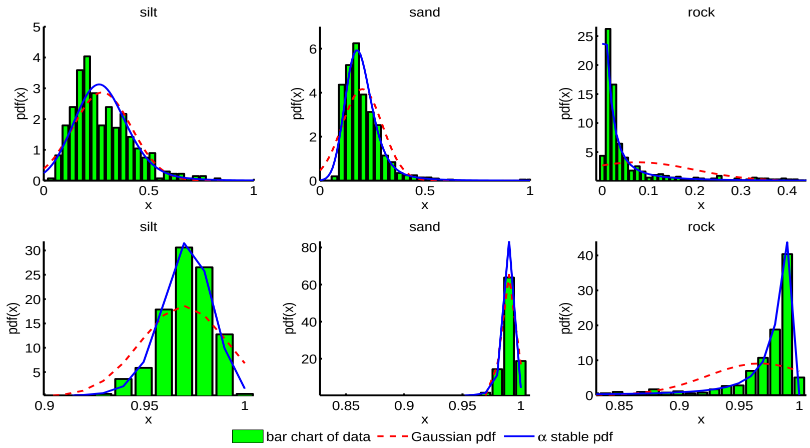

Fig. 2: Empirical pdfs and its estimations (The first row corresponds to the feature called "third quantile calculated on echo signal amplitude" and the second row corresponds to the feature called "25th quantile calculated on cumulative energy").

vectors belonging to the test base, the plausibility functions for each class are then calculated from equation (10). These plausibility functions are combined from equation (11) to obtain two mass functions. These two mass functions are combined by the conjunctive combination (we stay in open-world). Indeed, m_1 and m_2 and $\forall X \in 2^{\Theta}$:

$$m(X) = \sum_{Y_1 \cap Y_2 = X} m_1(Y_1) m_2(Y_2) \tag{13}$$

The decision is finally made by using the maximum of the pignistic probabilities.

3.3 Results

The two features are considered as a source of information. 5000 samples are randomly selected for the data set. Half the samples are used for the learning base and the rest for the test base. For the two approaches, the parameters of each model are estimated from the learning base. For the Bayesian approach, we need to estimate the prior probabilities $p(C_i)$ from the learning base approach. For each seabed sediment, the prior probabilities correspond to the proportion of seabed sediments in the learning base. The application of Bayes theorem gives posterior probabilities:

$$p(C_i/x) = \frac{p(x/C_i)p(C_i)}{\sum_{i=1}^{n} p(x/C_i)p(C_i)} \tag{14}$$

Finally, the decision is chosen by using the maximum of the posterior probabilities.

We can observe that the assumption of the α-stable model can easily accommodate the data compared to the Gaussian model (Figure 2). For each model and each method, we can observe that there is confusion between sand and silt (Table 1,2,3,4). Indeed, these sediments have similar properties. With the Gaussian models, we can observe that the theory of belief functions (Table 2) (classification accuracy of 70.92 %) give better results compared to the Bayesian approach (Table 1) (classification accuracy of 61.24 %). The belief functions take into account the imprecision of data introduced by the Gaussian model. The α-stable model gives better results compared to the Gaussian model because the α-stable can easily accommodate the data compared the Gaussian model. However, the Bayesian approach (Table 3) (classification accuracy of 82.68 %) gives better results than the belief functions (Table 4) (classification accuracy of 80.44 %) with the α-stable model but not significantly. We can explain these phenomena by the fact we introduce more information with the prior probability. The Bayesian approach performs well provided that once the probability density functions are well estimated. However, the probability density functions are poorly estimated. The theory of belief functions takes into account of imprecision/uncertainty during the learning step.

Table 1: Confusion matrix of seabed classification results based on the Bayesian approach with the Gaussian model

Ground truth seabed type	Predicted Seabed Type rock	sand	silt
rock	8.48 %	23.00 %	1.28 %
sand	0.00 %	37.32 %	2.80 %
silt	0.36 %	11.32 %	15.44 %

Table 2: Confusion matrix of seabed classification results based on the theory of belief functions with the Gaussian model

Ground truth seabed type	Predicted Seabed Type rock	sand	silt
rock	32.40 %	0.00 %	0.36 %
sand	12.44 %	20.92 %	6.76 %
silt	7.20 %	2.32 %	17.60 %

Table 3: Confusion matrix of seabed classification results based on the Bayesian approach with the α-stable model

Ground truth seabed type	Predicted Seabed Type rock	sand	silt
rock	28.28 %	0.04 %	4.44 %
sand	0.00 %	34.88 %	5.24 %
silt	0.84 %	6.76 %	19.52 %

Table 4: Confusion matrix of seabed classification results based on the theory of belief functions with the α-stable model

Ground truth seabed type	Predicted Seabed Type rock	sand	silt
rock	26.48 %	0.00 %	6.28 %
sand	0.00 %	29.84 %	10.28 %
silt	0.52 %	2.48 %	24.12 %

3.4 Conclusion

In this paper, we show the interest in using the theory of belief functions compared to a Bayesian approach in classification, especially to model imprecision of data. The problem with the Bayesian approach is that we introduce the *prior* probability

We show the interest to use the α-stable model compared to the Gaussian model to estimate data from a mono-beam echo sounder. However, the proposed approach is limited to the unimodal case. In [6], the authors deal with the problem of the belief functions linked to a multimodal pdf.

References

1. F. Caron, B. Ristic, E. Duflos and P. Vanheeghe. Least Committed basic belief density induced by a multivariate Gaussian pdf, International Journal of Approximate Reasoning. Vol. 48(2): 419-436, 2008.
2. D. Caughey, B. Prager and J. Klymak. Sea bottom classification from echo sounding data. Quester Tangent Corporation, Marine Technology Center, British Columbia, V8L 3S1, Canada, 1994.
3. F. Delmotte and Ph. Smets. Target identification based on the transferable belief model interpretation of Dempster–Shafer model, *IEEE Transactions on Systems, Man, and Cybernetics*, 34(4): 457–471, 2004.
4. A. Dempster. Upper and Lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*,38: 325-339, 1967.
5. T. Denœux. Extending stochastic ordering to belief functions on the real line, *Information Science*, 179 (9): 1362–1376, 2009.
6. P.E. Doré, A. Martin and A. Khenchaf. Constructing of a consonant belief function induced by a multimodal probability density function, COGnitive systems with Interactive Sensors (COGIS09), Paris, 2009.
7. A. Fiche, A. Martin, J.C. Cexus and A. Khenchaf. Continuous belief functions and α-stable distributions, *International Conference on Information Fusion*. Edinburgh, United Kingdom, July 2010.
8. I.A. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws, Communications in Statistics-Simulation and Computation, 10(1): 17–28, 1981.
9. J.P. Nolan. Numerical calculation of stable densities and distribution functions, *Communications in Statistics-Stochastic Models*, 13(4): 759–774, 1997.
10. J.P. Nolan. Stable Distributions - Models for Heavy Tailed Data, In progress, Chapter 1 online at academic2.american.edu/∼jpnolan, 2012.
11. B. Ristic and Ph. Smets. Belief function theory on the continuous space with an application to model based classification, Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU, 4–9, 2004
12. B.Ristic and Ph. Smets. Target classification approach based on the belief function theory, *IEEE Transactions on Aerospace and Electronic Systems*, 42(2): 574–583, 2005.
13. G. Shafer. *A mathematical theory of evidence*, Princeton University Press, 1976.
14. Ph. Smets. Constructing the pignistic probability function in a context of uncertainty, *Uncertainty in artificial intelligence*, 5: 29–39, 1990.
15. Ph. Smets. Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem, *International Journal of Approximate Reasoning*, 9(1): 1–35, 1993.
16. Ph. Smets. Belief functions on real numbers, *International journal of approximate reasoning*, 40(3): 181-223, 2005.
17. M.S. Taqqu and G. Samorodnisky. Stable non-gaussian random processes, *Chapman and Hall*, 1994.
18. V.M. Zolotarev. *One-dimensional stable distributions*, Volume 65 of Translations of Mathematical Monographs. American Mathematical Society. Translation from the original 1983 Russian edition, 1986.