INTRODUCTION OF SPATIAL INFORMATION WITHIN THE CONTEXT OF EVIDENCE THEORY

A.-S. Capelle, C. Fernandez-Maloigne

Laboratoire IRCOM-SIC
Université de Poitiers
Bd Marie et Pierre Curie, BP 30179
86962 Futuroscope Chasseneuil, France

O. Colot

Laboratoire I3D
Université des Sciences et de Technologie de Lille
Bât. P2
59655 Villeneuve D'Ascq Cedex, France

ABSTRACT

In this article, we propose a method to introduce spatial information within the context of pattern recognition by the mean of evidence theory. Indeed, we can consider that each neighbor brings some information useful to determined the class of a pattern to classify. We propose to introduce such information through the Dempster's combination rule. This combination, which takes into account the distance between neighbors, provides a more accurate modeling of the information and improves the classification process of the data. We illustrate the interest and the impact of this method through the problem of segmentation of multi-echo magnetic resonance (MR) images. In particular, we show that the segmentation results are more accurate and that some ambiguities of classification are resolved.

1. INTRODUCTION

Data fusion becomes an important field of research this last few years; its applications are numerous: multi-sensor fusion, classification, or target tracking [1]. Different theories are adapted to data fusion problematic: probabilistic fusion, Bayesian inference, fuzzy set theory, possibility theory or evidence theory. In this paper, we focus on the evidence theory.

The evidence theory, also called theory of belief functions was initially introduced by Dempster [2] and formalized by Shafer [3]. Such theory is particularly used in the context of pattern recognition of imprecise and uncertain data. It is based on the construction of belief functions which model the knowledge and the belief we place on the data. Different interpretations of evidence theory successively appeared. In [4], Smets and Kennes deviate from the initial probabilistic interpretation of the evidence theory with the *transferable belief model* (TBM) giving a clear and coherent interpretation of the underlying concept of the theory.

In this article, we are interested in the problem of the classification of spatial ordered data. Indeed, a lot of classification methods of uncertain and imprecise data do not take into account the neighborhood of the patterns to classify. However we can consider the neighborhood as a source of information and introduce these information within a fusion process.

In section 2, we quickly introduce the evidence theory background. In section 3, we propose a method to introduce the spatial information through the Demspter's combination rule. In section 4, the method is applied to the segmentation of MR images. We present the benefits of our method on the segmentation results. Finally, in section 5, we conclude about the interests of the proposed method.

2. EVIDENCE THEORY BACKGROUND

2.1. Belief structures

Within the context of evidence theory, imprecise and uncertain data are modelled thanks to belief structures. The existency and the definition of theses functions involve the definition of a frame of discernement Θ composed of the N exhaustive and exclusive hypothesis H_i of the classification problem:

$$\Theta = \{H_1, H_2, \ldots, H_N\}. \tag{1}$$

From the frame of discernment, we define 2^Θ the power set of 2^N propositions defined on Θ:

$$2^\Theta = \{\emptyset, H_1, H_2, \ldots, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \ldots, \Theta\}. \tag{2}$$

The piece of evidence brought by a source of information (sensor, agent...) on a proposition A (singleton or composed hypothesis of 2^Θ), is modelled by the belief structure m, called Basic Belief Assignment (bba), defined by:

$$m : 2^\Theta \rightarrow [0, 1], \tag{3}$$

and verifying:

$$m(\emptyset) = 0, \tag{4}$$

$$\sum_{A\subseteq\Theta} m(A) = 1. \tag{5}$$

From this function, two belief structures, the credibility (Bel) and the plausibility (Pl) can be derived by the following equations:

$$Bel(A) = \sum_{B\subseteq A} m(B), \tag{6}$$

$$Pl(A) = \sum_{A\cap B\neq\emptyset} m(B). \tag{7}$$

The degree of belief $Bel(A)$ can be interpreted as the total amount of belief in the proposition A. The plausibility $Pl(A)$ quantifies the maximum amount of belief potentially attributed to A. The credibility and the plausibility are thus dual notions: the plausibility is defined by $Pl(A) = Bel(\Theta) - Bel(\overline{A})$ where $\overline{A}$ is the complementary of A.

The main difficulty raises in correctly model the knowledge given by the different sources. Generally, the models depend on the classification problem. Therefore, we can distinguish two main approaches: the distance-based models initially proposed by Denœux [5] which take into account the neighborhood information and the models based on likelihood functions [3, 6].

2.2. Belief attenuation

As we said, the belief structure m models the piece of evidence brought by a source of information on the different hypothesis of 2^Θ. When this source is considered as imprecise or not completely reliable, the confidence in this source can be attenuated by a factor α and a derived belief structure m_α is defined by:

$$m_\alpha(A) = \alpha.m(A) \quad \forall A \in 2^\Theta, \tag{8}$$

$$m_\alpha(\Theta) = 1 - \alpha + \alpha.m(\Theta). \tag{9}$$

The difficulty lies then in the correct definition of the factor α [7].

2.3. Combination

Let denote $\{m_1, \ldots, m_N\}$ N bba associated to N independant sources $S_1, \ldots, S_N$ of information. The evidence theory provides an adapted framework to fusion or combine these N sources in a synthesized information. A common operator is the orthogonal sum also called the Dempser's combination. Thus, the fusioned bba $m_\oplus$ is defined by

$$m_\oplus = m_1 \oplus \ldots \oplus m_i \oplus \ldots m_N. \tag{10}$$

For two sources of information S_1 and S_2, the fusioned bba $m_\oplus$ is given by:

$$\forall A \subseteq \Theta \quad m_\oplus(A) = \frac{1}{1-k} \sum_{B\cap C=A} m_1(B).m_2(C), \tag{11}$$

where k is defined by:

$$k = \sum_{B\cap C=\emptyset} m(B).m(C). \tag{12}$$

The normalization term k, with $0 \leq k \leq 1$, can be interpreted as a measure of the conflict between the sources to combine. The Dempster's rule has been justified theoretically by several authors [8, 9]. However the normalization step was also criticized [7, 9]. It is very important to take into account the value of this term: when it is high ($k \approx 1$), combining the sources is a non-sense leading to incoherence [7] and involving counter-intuitive behaviors.

2.4. Decision

For most applications, a decision have to be taken generally in favor of a simple hypothese. Several rules are proposed, the most current being the rule of the maximum of *credibility* and the maximum of the *plausibility*. Within the context of the TBM, Smets [10] defines the *pignistic* probability distribution.

Thus if $\Upsilon(.)$ represents the credibility, the plausibility or the pignistic function, the decision function δ for a vector X to classify is given by:

$$\delta(X) = H_n \quad \text{with} \quad H_n = \arg[\max_{H_i\in\Theta} \Upsilon(H_i)]. \tag{13}$$

We denote that the decision can be taken including the concept of cost functions and the concept of rejection class [11].

3. INTRODUCTION OF SPATIAL INFORMATION

3.1. Motivations

The solving of pattern recognition problems by the evidence theory is usually divided into two parts. In a first time, the knowledge is modelled via bba functions. This part often includes an estimation of parametric models. In a second time, a decision is taken given the bba functions. In this article, we are particularly interested in the specific case of spatial ordered data.

We propose to include before the decision step a regularization by means of the introduction of spatial neighborhood information using belief functions. Indeed, when the data are spatially ordered, which is for example the case in image processing problems, the class of a pattern X to classify in a homogeneous region is often the same that the class of its neighbors. The pieces of evidence brought by the neighbors and represented by bba can thus be considered as some sources of information about the class X. Concequently, the use of this kwowledge increase our belief about X and lead to accurate solution.

3.2. Principle

Let X be the pattern to classify and $\partial(X) = \{X_1, \dots, X_k\}$ the set of the k spatial neighbors of X. We note m and $m_{\partial(X)} = \{m_{X_1}, \dots, m_{X_k}\}$ the bba respectively associated with X and $\partial(X)$. We propose to introduce the spatial information by modifying the bba m through a weighted Demspter's combination. Thus, we define the new bba associated with X by:

$$m' = m \oplus \alpha_1 m_{X_1} \oplus \ldots \oplus \alpha_k m_{X_k}, \tag{14}$$

where α_i, for $i = 1, \dots, k$, are attenuation factors given by:

$$\alpha_i = \phi(d(X, X_i)) \tag{15}$$

where $\phi(.)$ is a decreasing function depending of the distance $d(X, X_i)$ between X and its neighbor X_i. The number k of neighbors can be 8 in a 2-dimensional problem or 26 in 3-dimensional ones. However, it depends on the shape of the function ϕ.

The existence of the attenuation factors is very important: the more a neighbor is far, the less the information brought is reliable and the less it should be taken into account. Without these factors, the information about X could be noised by erroneous sources of information.

3.3. Consequences

The combination of the bbas taking into account the distance allows to increase the belief about X. The decision about the class of X is then more reliable.

The spatial information allows to reduce the main errors encountered in a classification: these are located on the frontiers of the different regions and/or are due to noise. The use of spatial information reduces such errors and consequently increases the quality of the result. These aspects are illustrated in section 4.

4. APPLICATION TO MULTI-ECHO MR BRAIN IMAGES SEGMENTATION

4.1. Problematic

Our problem is to segment into homogeneous regions some multi-echo MR brain volumes. The volumes we use present some pathologies such as lesions, tumors or edemas. The use of several echoes provides a numerous, redundant and complementary information. Moreover, these information are uncertain and imprecise. Thus, the evidence theory is well adapted to such problem.

4.2. Segmentation scheme

We suppose that the number of classes of our classification problem is constant, each class being associated to an anatomical region: white matter (WM), gray matter (GM), cerebrospineal fluid (CF), tumor (T) and edema (E). The segmentation scheme is divided into four parts as represented in figure 1.

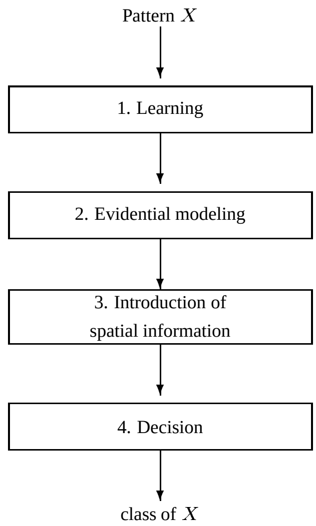


Fig. 1. Segmentation scheme

We suppose that each class can be represented by a gaussian process which is a classical hypothesis [12] with MR images. The learning thus consists to estimate the gaussian distribution. Next, each pattern X of the volume is modelled by a bba issued from the Appriou's model [6]. The spatial information is introduced using the eq. 14 and 15 with $\phi(d) = \exp\{(-d)^2\}$ and the use of 26 neighbors. Finally, the decisions are made using the pignistic function.

4.3. Effects of the spatial information

We will observe the effects of the spatial information on two aspects: firstly, on the segmentation results and secondly, on the localisation and the number of the rejected patterns.

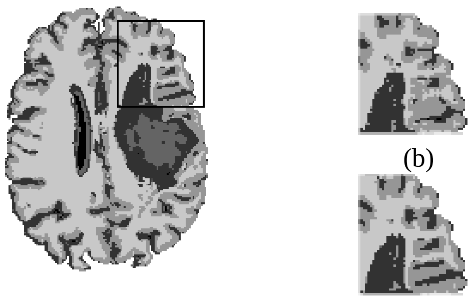
4.3.1. Segmentation results

The figure 2-a is the final segmentation obtained with the process described above. We observe that the main anatomical structures are well retrieved. The effects of the spatial information are observed on figures 2-b and 2-c obtained respectively without and with spatial information. With spatial information, we observe that some structures are better detected as we can notice in particular with the WM. Moreover the regions are smoother as we can see on the outlines of the edema which are more regular.

The spatial information thus accurate the segmentation results. The different regions are better detected, smoothed while keeping the small details.

4.3.2. Rejected patterns

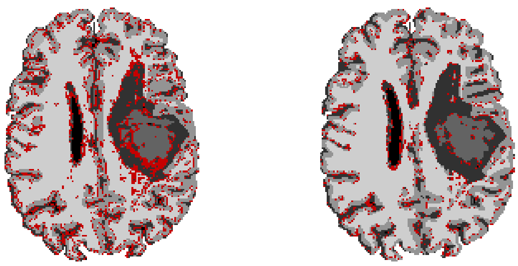
We study here the influence of the neighborhood information on the localisation and the number of *ambiguous* pattern. We call *ambiguous* patterns the patterns classified in



(a) (b) (c)

Fig. 2. Effects of the introduction of spatial information. From darker to lighter, we retrieve the CF, E, T, WM and GM.

a rejection class during the decision step. In this example, the rejected patterns are represented in red on figures 3. The rejected threshold is set to 0.4.
The figure 3-a is obtained without using spatial information. The rejected patterns are numerous and located both in the frontiers between regions and inside the regions themselves. In this last case, they mainly correspond to noise patterns. When spatial information are used the number of rejected pattern noticeably decreases. Most of the noise points are now classified. Moreover, some ambiguous points are still located on the frontiers. This behaviour is the expected one: on the frontiers, the information included by our methods are antagonistic and a reliable decision is not always possible.



(a) without spatial information (b) with spatial information

Fig. 3. Segmentation results obtained with a rejection threshold equal to 0.4 (in red, the rejected voxels).

5. CONCLUSION

We have proposed a method which introduces spatial information in pattern recognition process of spatial ordered data by means of the evidence theory. We consider that each neighbor of a pattern to classify is a source of information about the class of the pattern. Thus, the information about a pattern, modelled by belief functions, is completed by the neighborhood information. These information are included through a weighted Dempster's combination rule and provide a more reliable modelling of the local information and consequently improves the classification results.
The application of the method for the segmentation of multi-echo MR brain images shows that the segmentation results are more accurate. The anatomical regions are better segmented. Some classification ambiguities are resolved contributing to improve the well classification rate.
The proposed method can be used in many applications involving uncertain, imprecise and spatially ordered data. Moreover, the use of spatial information can be adjusted to different contexts by modifying the distance function and the number of neighbors used.

6. REFERENCES

[1] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Belief function combination and conflict management," *Information Fusion*, pp. 149–162, 2002.

[2] A. Dempster, "Upper and lower probabilities induced by multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.

[3] G. Shafer, *A Mathematical Theory of Evidence*, Princetown University Press, 1976, Princetown New Jersey.

[4] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.

[5] L.M. Zouhal and T. Denœux, "An adaptative k-NN rule based on Dempster-Shafer theory," in *6th International Conference on Computer Analysis of Images and Pattern*, Sept. 1995, pp. 310–317.

[6] A. Appriou, "Multisensor signal processing in the framework of the theory of evidence," *NATO/RTO - Lecture Series 216 on "Application of Mathematical Signal Processing Techniques to Mission Systems"*, pp. 5–31, Nov. 1999.

[7] L. A. Zadeh, *On the validity of Dempster's rule of Combination of Evidence*, University of California, Berkeley, 1979, ERL Memo M79/24.

[8] D. Dubois and H. Prade, "On the unicity of Dempster rule of combination," *International Journal of Intelligent System*, pp. 133–142, 1996.

[9] F.Voorbraak, "On the justification of Dempster's rule of combinations," *Artificial Intelligence*, pp. 171–197, 1991.

[10] P. Smets, *Constructing the pignistic probability function in a context of uncertainty*, M. Henrion and R.D. Schachter and L.N. Kanal and J.F. Lemmers, editors, Amsterdam, North-Holland, 1990.

[11] T. Denœux, "Analysis of evidence-theoretic decision rules for pattern classification," *Pattern Recognition*, vol. 30, no. 7, pp. 1095–1107, 1997.

[12] L.P. Clarke, R.P. Velthuizen, S. Phuphanich, J.D. Schellenberg, J.A. Arrington, and M. Silberger, "MRI stability of three supervised segmentation techniques," *Magnetic Resonance Imaging*, vol. 11, pp. 95–106, 1993.