Knowledge modeling methods in the framework of Evidence Theory An experimental comparison for melanoma detection

Eric Lefevre, Olivier Colot, Patrick Vannoorenberghe, Denis de Brucq
Laboratoire Perception, Systèmes Information (PSI), UPRES EA 2120
Université/INSA de Rouen

Abstract

The Dempster-Shafer's theory, or evidence theory, is used in different fields such as data fusion, regression or classification. Within the framework of this theory, the uncertain and imprecise data are represented using belief functions. Data fusion operators as well as the decision rule of this theory were largely developed and formalized. The aim of this paper is to present modeling methods of knowledge for the initialization of belief functions. Moreover, an experimental comparison of these different modeling on real data extracted from images of dermatological lesions is presented.

1 Introduction

Data analysis and processing are two important tasks in today's information society. The data management becomes essential when the informations are imperfect, that is to say imprecise and uncertain. Traditionally, probability theory, which is inadequate in some cases as well known, is used for dealing with imperfect data. In this paper, we deal with a classification method of imperfect data sets using evidence theory proposed by Shafer [1] and Smets [2]. In the framework of this theory, the knowledge modeling rests on the construction of belief functions. These last years, several methods to initialize these functions were developed. In this paper, we present and compare some of these methods. This paper is organized as follows. In section 2, we introduce notations allowing to describe the Dempter-Shafer's Theory of evidence. Section 3 introduces several methods of basic belief assignments. These methods are checked and compared on real data extracted from images of dermatological lesions (Section 4).

2 Dempster-Shafer's theory

The Dempster-Shafer's theory of evidence is based on the concept of lower and upper bounds for a set of compatible probability distributions introduced by Dempster [3]. On this basis, Shafer [1] showed the advantage of using belief functions for modeling the uncertain knowledge. The use of belief functions as an alternative to subjective probabilities was later justified axiomatically by Smets [2] who introduced the *Transferable Belief Model* providing a clear and consistent interpretation of the various concept underlying the theory.

2.1 Belief structure

Let Θ represents a finite set of N hypotheses. The set Θ is called *frame of discernment* and is defined by :

$$\Theta = \{H_1, \cdots, H_n, \cdots, H_N\}. \quad (1)$$

2^Θ represents the set of the $2^N - 1$ propositions $\mathcal{H}$ of Θ :

$$2^\Theta = \{\mathcal{H}/\mathcal{H} \subseteq \Theta\} = \{H_1, \cdots, H_N, H_1 \cup H_2, \cdots, \Theta\}. \quad (2)$$

A piece of evidence that influences our belief concerning the true value of a proposition $\mathcal{H}$ can be represented by a basic belief assignment (bba) $m(.)$. For each source S_j ($j = \{1, \ldots, J\}$), a mass function $m_j(.)$ is defined by :

$$m_j : 2^\Theta \rightarrow [0, 1] \quad (3)$$

and verifying the following properties :

$$m_j(\emptyset) = 0 \quad (4)$$

$$\sum_{\mathcal{H} \subseteq \Theta} m_j(\mathcal{H}) = 1. \quad (5)$$

From the basic belief assignment $m_j(.)$, a credibility function and plausibility function can be computed using the following equations :

$$Bel_j(\mathcal{H}) = \sum_{\mathcal{H}' \subseteq \mathcal{H}} m_j(\mathcal{H}') \quad (6)$$

$$Pl_j(\mathcal{H}) = \sum_{(\mathcal{H} \cap \mathcal{H}') \neq \emptyset} m_j(\mathcal{H}') = 1 - Bel_j(\overline{\mathcal{H}}) \quad (7)$$

where $\overline{\mathcal{H}}$ denotes the negation of $\mathcal{H}$. The value $Bel_j(\mathcal{H})$ quantifies the strength of the belief that event $\mathcal{H}$ occurs. The plausibility function $Pl_j(\mathcal{H})$ provides a measure of no doubt about the hypothesis $\mathcal{H}$. The plausibility of $\mathcal{H}$ is the sum of belief associated with subset $\mathcal{H}'$ no contradicting $\mathcal{H}$. The main difficulty consists in modeling knowledge to initialize the basic belief assignment $m_j(.)$. Many modeling methods have been proposed, which depend usually on the concerned application. They are detailed in section 3.

2.2 Belief function attenuation

An additional aspect of this theory concerns the attenuation of the basic belief assignment m_j by a coefficient α_j for a source S_j. The aim of this step is to introduce a reliable degree between all information sources. For each source S_j, the attenuated belief function, noted $m_{(\alpha,j)}$, can be written as it follows :

$$\forall\, \mathcal{H} \in 2^{\Theta} \qquad m_{(\alpha,j)}(\mathcal{H}) = \alpha_j.m_j(\mathcal{H}) \quad (8)$$

$$m_{(\alpha,j)}(\Theta) = 1 - \alpha_j + \alpha_j.m_j(\Theta). \quad (9)$$

The main difficulty consists in the adjustment of the attenuation factor [4, 5].

2.3 Combination operators

In addition, Dempster-Shafer's theory allows the fusion of several independent sources using the Dempster's combination rule. It is defined like the orthogonal sum (commutative and associative) following the equation :

$$m(\mathcal{H}) = m_1(\mathcal{H}) \oplus m_2(\mathcal{H}) \oplus .... \oplus m_J(\mathcal{H}). \quad (10)$$

For two sources S_i and S_j, the aggregation of evidence can be written :

$$\forall\, \mathcal{H} \subseteq \Theta \qquad m(\mathcal{H}) = \frac{1}{1-\mathcal{K}} \sum_{\mathcal{H}' \cap \mathcal{H}'' = \mathcal{H}} m_i(\mathcal{H}').m_j(\mathcal{H}'') \quad (11)$$

where $\mathcal{K}$ is defined by :

$$\mathcal{K} = \sum_{\mathcal{H}' \cap \mathcal{H}'' = \emptyset} m_i(\mathcal{H}').m_j(\mathcal{H}''). \quad (12)$$

The normalization coefficient $\mathcal{K}$ evaluates the conflict between two sources. This normalization step allows to ensure the properties of mass functions. In [6], Zadeh presents a situation where the normalization step used by Dempster's combination rule leads to surprising results. In order to cope with this problem, other combination operators have been proposed [7, 8, 9].

2.4 Decision

Once the resulting mass m thus obtained, the decision can then be taken. Various rules of decision were defined, most current being the rule of the maximum of plausibility and the rule of the maximum of credibility. From the functions of belief, Smets [10] defines a function of probability called *pignistic* probability distribution. In a general way, one defines the decision function δ for a vector X' to classify by :

$$\delta(X') = H_n \quad avec \quad H_n = arg \quad [\max_{H_i \in \Theta} \Upsilon(H_i)] \quad (13)$$

where $\Upsilon(.)$ is the credibility function, the plausibility function or the pignistic probability. An analysis of several decision rules including the concept of cost functions is presented in [11].

3 Basic Belief Assignment

In a classification problem, we consider a pattern X' to be classified. X' is a vector with J components : $X' = [x'_1, \ldots, x'_J]^t$. An important aspect of the discrimination concerns learning knowledge using data. In evidence theory, this problem leads to initialize the belief functions m_j. Two kind of methods to initialize belief functions were proposed. First one is based on the data analysis in the features space [12]. The second kind of methods analyses separately each feature of the pattern [4, 13]. The approach proposed by Denoeux uses a neighborhood information. Each nearest neighbor of a pattern to be classified is considered as an item of evidence. The resulting belief assignment is also defined as a function of the distance between the pattern and its neighborhood. The approach proposed by Appriou considers the belief structure must be compatible with several axioms leading to compatibility with the Bayesian approach. In fact, one of the two proposed models can be derived from the generalized Bayesian theorem [14]. These two methods are presented (Section 3.1 and 3.2), allowing us to introduce an original approach for the basic belief assignment (Section 3.3).

3.1 Denoeux's method

The presence of a training pattern X^s of the class H_n among the k nearest neighbors of a pattern X' to be classified is considered as a piece of evidence that influences our belief concerning the class membership of the entity under consideration. This information is represented by a bba m^s over the set Θ of classes. A fraction of the unit mass is assigned by m^s to the singleton $\{H_n\}$, and the rest is assigned to the whole frame of discernment Θ. The mass $m^s(\{H_n\})$ is defined as a decreasing function of the distance d^s between X' and X^s in the feature space :

$$m^s(\{H_n\}) = \alpha\phi_n(d^s) \quad (14)$$
$$m^s(\Theta) = 1 - \alpha\phi_n(d^s). \quad (15)$$

where $0 < \alpha < 1$ is a constant, and ϕ_n is a monotonically decreasing function verifying $\phi_n(0) = 1$ and $lim_{d\to\infty}\phi_n(d) = 0$. An exponential form can be postulated for ϕ_n :

$$\phi_n(d^s) = exp(-\gamma_n(d^s)^2) \quad (16)$$

where γ_n is a positive parameter associated to class H_n. The k nearest neighbors of X' can be regarded as k independent sources of information represented by a bba. These several piece of evidence can be aggregated by means of Dempster's combination rule representing our belief concerning the class membership of X'. This evidence-theoretic k-NN rule was shown to have good classification accuracy as compared to the voting and distance-weighted rule [12]. A learning algorithm was proposed by Zouhal and Denoeux [15] for determining the parameters γ_n in the equation (16) by optimizing an error criterion. This improvement was shown to yield further reduction of classification error in most cases.

3.2 Appriou's method

Dealing with the problem of imprecise and unreliable observations in the terms of evidence theory, Appriou proposes finding for each source S_j, a model of its N *a priori* probabilities $p(x'_j/H_n)$ and their N respective confidence factors q_{nj} in the form of a mass function $m_j(.)$. Since the source S_j are distinct, a global evaluation $m(.)$ can the be obtained by computation of the orthogonal sum of the $m_j(.)$. Appriou proposes to conduct an exhaustive and exact search of all the models that might satisfy three fundamental axioms :

- *Axiom 1* : Consistency with the Bayesian approach;
- *Axiom 2* : Separability of the evaluation of the hypotheses H_n;
- *Axiom 3* : Consistency with the probabilistic association of sources S_j.

The models satisfying these three axioms have been presented by Appriou in [13] by progressively restricting the set of possible models. It appears that only two models satisfy the three desired axioms. *Model 1* is particularized by :

$$m_{nj}(H_n) = 0 \quad (17)$$
$$m_{nj}(\overline{H_n}) = q_{nj} * \{1 - R_j * p(x'_j/H_n)\} \quad (18)$$
$$m_{nj}(\Theta) = 1 - q_{nj} + q_{nj} * R_j * p(x'_j/H_n) \quad (19)$$

and *Model 2* by :

$$m_{nj}(H_n) = q_{nj} * R_j * \frac{p(x'_j/H_n)}{1 + R_j * p(x'_j/H_n)} \quad (20)$$
$$m_{nj}(\overline{H_n}) = \frac{q_{nj}}{1 + R_j * p(x'_j/H_n)} \quad (21)$$
$$m_{nj}(\Theta) = 1 - q_{nj}. \quad (22)$$

In the both cases, the normalization factor R_j is simply constrained by :

$$R_j \in [0, \frac{1}{max_{x'_j,n}\{p(x'_j/H_n)\}}]. \quad (23)$$

The coefficients q_{nj} correspond to reliability factors computed for each information source S_j according to each hypothesis H_n. When the densities are perfectly representative of the training then the coefficients q_{nj} are equal to 1 and the belief functions are not discounted. Moreover, Appriou shows that the choice of the factor q_{nj} does not involve any problem. If the confidence into the density is high, q_{nj} is fixed to 1 and to 0.9 otherwise. Other authors [16, 17] proposed automatic methods to compute these confidence coefficients. A mass function $m(.)$ synthesizing all the evaluations is obtained by computing the orthogonal sum (see Section 2.3) of the different mass functions $m_{nj}(.)$:

$$m_j(.) = \bigoplus_n m_{nj}(.) \quad (24)$$
$$m(.) = \bigoplus_j m_j(.) \quad (25)$$

The decision rule consists in choosing H_n which maximizes the plausibility function.

3.3 Our method

In this section, we present a method proposed in [5, 18]. As for the modeling suggested by Appriou, each component of vector X' to classify is considered as an imprecise and uncertain information source intended to reinforce the membership of X' to a hypothesis H_n. On the contrary, we build for an information source only one belief structure instead of N. For a source S_j, we determine for each singleton hypothesis a not standardized mass noted $P_j(.)$ which is defined using the probabilities densities :

$$M_j(H_n) = p(x'_j/H_n) \tag{26}$$

From this mass, we define the masses accorded to disjunctions of hypothesis using the operator min. This operator will be applied to the hypothesis constituting the subset :

$$P_j(\mathcal{H}) = \min(p(x'_j/H_n), \ldots, p(x'_j/H_{n'})) \tag{27}$$

with $\mathcal{H} = \bigcup_{i \in [n,\ldots,n']} H_i$. Then, we build the belief structure as it follows :

$$\forall \mathcal{H} \in 2^\Theta \qquad m_j(\mathcal{H}) = R_j * P_j(\mathcal{H}) \tag{28}$$

where R_j is a normalization factor. It allows to verify the condition given by equation (5). It is defined by:

$$\forall \mathcal{H} \in 2^\Theta \qquad R_j = \frac{1}{\sum_{\mathcal{H} \in 2^\Theta} P_j(\mathcal{H})}. \tag{29}$$

An attenuation factor α_j can be introduced (see Section 2.2). This factor will intend to attenuate the belief function $m_j(.)$ relative to each source S_j according to its ability to distinguish the hypotheses. It will be given using the measure of confusion which is one of the uncertainty measures defined in [19]. The measure of confusion $Conf(.)$ of a belief structure m_j is defined by :

$$\alpha_j = Conf(m_j) = -\sum_{\mathcal{H} \in 2^\Theta} m_j(\mathcal{H}).log_2(Bel_j(\mathcal{H})) \tag{30}$$

The evolution of attenuation factor α_j determined with the measure of confusion according to classification capacity of sources is represented on the figures Fig. 1 and Fig. 2. In the first case (Fig. 1), the source allowed a good discrimination between the hypotheses. The reliability factor evolves according to the mass distributions. We can notice that the coefficient is minimal when ambiguity between the hypotheses is maximum. In the second case (Fig. 2), we can note that the attenuation zone (i.e. when the coefficient is close to zero) is more significant than in the first case.

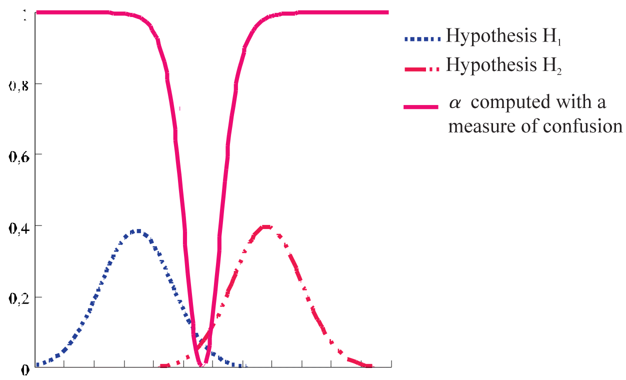

Figure 1: Evolution of α.

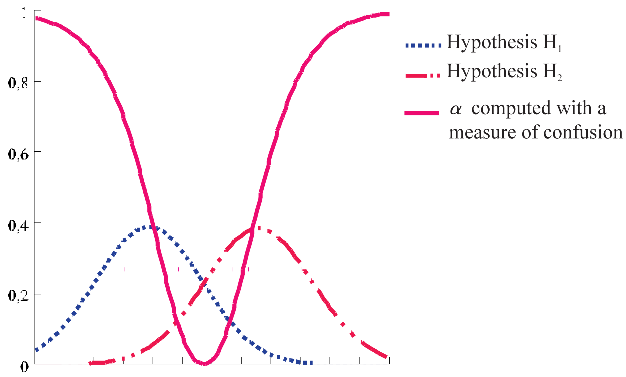

Figure 2: Evolution of α.

4 Results

We applied these methods of basic belief assignment to classification between benign and malignant lesions in dermatology in the aim to help the clinical practitioner for melanoma diagnosis. Melanoma is an increasing form of cancer in the world. It has increased twice for 15 years in Canada and it is 3% in the USA at this time. For the classification, we propose 19 characteristics (geometric as well as colorimetric) which are robust and relevant. Details concerning the features can be found in [20]. These characteristics show that the melanoma class is heterogeneous. We have applied the classification process on a training set of 81 lesions : 61 benign lesions (naevi) and 20 malignant lesions (melanoma) and a test set of 209 lesions : 191 naevi and 18 melanoma. The probability densities for each information source according to each hypothesis $p(x'_j/H_n)$, computed in the methods presented Section 3.2 and Section 3.3, are estimated by the well-known Parzen's windows method with a gaussian kernel. The number k of neighbors, used in the Denoeux's method, is fixed at 9. However, this method is used without optimized belief structure parameters. Final results obtained on the training set are presented in the following tables (Table 1, 2 and 3). The method we propose allows to obtain 98.8% of good classification (Table 3). This result is quite similar to the result obtained with Appriou's approach (97.5%). The result obtained with Denoeux's approach are less significant; the good classification rate is 91.35% (Table 1).

Table 1: Result on training set of Denoeux's method.

		Decision	
	%	H_1 **(naevus)**	H_2 **(melanoma)**
Nature	H_1	98.37	1.63
Nature	H_2	30	70

Table 2: Result on training set of Appriou's method.

		Decision	
	%	H_1 **(naevus)**	H_2 **(melanoma)**
Nature	H_1	100	0
Nature	H_2	10	90

Table 3: Result on training set of our method.

		Decision	
	%	H_1 **(naevus)**	H_2 **(melanoma)**
Nature	H_1	100	0
Nature	H_2	5	95

The results obtained with the test set are presented in the following tables (Table 4, 5 and 6). We can notice that the best result of good classification (85.2% of good classification) is obtained with the method proposed by Denoeux (Table 4). However, the percentage of non-detection of melanoma (72.22%) is more important for this method. For a medical diagnosis application, this result is not satisfactory. In this context, we search to minimize the non-detection and false alarm rates. This problem shows the limits of the distance based methods in the case of heterogeneous classes (the melanoma class). The two other methods of basic belief assignments give percentages of good classification of 74.65% for the method proposed by Appriou (Table 5) and 72.72% for the method presented in this paper (Table 6). As for the training set, the results obtained on the test set are quite similar.

Table 4: Result on test set of Denoeux's method.

		Decision	
	%	H_1 **(naevus)**	H_2 **(melanoma)**
Nature	H_1	90.6	9.4
Nature	H_2	72.22	27.78

Table 5: Result on test set of Appriou's method.

		Decision	
	%	H_1 **(naevus)**	H_2 **(melanoma)**
Nature	H_1	74.9	25.1
Nature	H_2	27.77	72.23

Table 6: Result on test set of our method.

		Decision	
	%	H_1 **(naevus)**	H_2 **(melanoma)**
Nature	H_1	72.26	27.74
Nature	H_2	22.22	77.78

5 Conclusions

In this paper, we have presented different methods to initialize belief functions using data extracted from a training set. We applied these methods to a dataset extracted from dermatological lesions, as a tool for help in the diagnosis of skin cancer (melanoma). According to the results obtained with this particular database, the distance based method (proposed by Denoeux) does not seem to be adapted for the classification of heterogeneous classes. In fact, the class corresponding to the malign lesions, is divided in several clusters in the feature space. The method suggested in this paper and the one presented by Appriou are more robust with respect to this kind of data. Future work is concerned with a global evaluation on these belief functions estimation methods on synthetic and real datasets.

References

[1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

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

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

[4] A. Appriou, "Multisensor signal processing in the framework of the theory of evidence," in *Application of Mathematical Signal Processing Techniques to Mission Systems*. November 1999, pp. (5–1)(5–31), Research and Technology Organization (Lecture Series 216).

[5] E. Lefevre, O. Colot, and P. Vannoorenberghe, "A classification method based on the Dempster-Shafer's theory and the information criteria," in *Proceedings of Fusion'99*, 1999, pp. 1179–1184.

[6] L. A. Zadeh, *On the Validity of Dempster'Rule of Combination of Evidence*, Univ. of California, Berkeley, 1979.

[7] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 447–458, 1990.

[8] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Comput. Intell.*, vol. 4, pp. 244–264, 1988.

[9] E. Lefevre, O. Colot, P. Vannoorenberghe, and D. De Brucq, "A generic framework for resolving the conflict in the combination of belief structures," in *FUSION'2000 (To appear)*, 2000.

[10] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," in *Uncertainty in Artificial Intelligence 5*, M. Henrion, R. D. Schachter, L.N. Kanal, and J.F. Lemmer, Eds., Amsterdam, 1990, pp. 29–40, North-Holland.

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

[12] T. Denoeux, "A k-nearest neighbour classification rule based on Dempster-Shafer theory," *IEEE Trans. Syst. Man. and Cyber.*, vol. 25, no. 5, pp. 804–813, 1995.

[13] A. Appriou, "Probabilités et incertitude en fusion de données multi-senseurs," *Revue Scientifique et Technique de la Défense*, vol. 11, pp. 27–40, 1991.

[14] P. Smets, "Belief functions: The disjonctive rule of combination and the generalized bayesian theorem," *Int. J. of Approximate Reasoning*, vol. 9, pp. 1–35, 1993.

[15] L. Zouhal and T. Denoeux, "An adaptative k-NN rule based on Dempster-Shafer theory," in *Proc. Of the 6th Intern. Conf. On Comput. Analys. Of Imag. and Pattern.* ICAIP'95, September 1995, pp. 310–317, Springer Verlag.

[16] E. Lefevre, O. Colot, P. Vannoorenberghe, and D de Brucq, "Contribution des mesures d'information à la modélisation crédibiliste des connaissances," *Traitement du Signal (To appear)*, 1999.

[17] S. Mathevet, L. Trassoudaine, P. Checchin, and J. Alizon, "Application de la théorie de l'évidence à la combination de segmentations en régions," in *Dix-Septième Colloque GRETSI (Vannes)*, Septembre 1999, pp. 635–638.

[18] E. Lefevre, P. Vannoorenberghe, and O. Colot, "Using information criteria in Dempster-Shafer's basic belief assignment," in *Proceeding of Fuzz'ieee 99*, 1999.

[19] G.J. Klir, *Measures of Uncertainty in the Dempster-Shafer Theory of Evidence*, pp. 35–49, Advances in the Dempster-Shafer Theory of Evidence. John Wiley and Sons, New York, R.R. Yager and M. Fedrizzi and J. Kacprzyk edition, 1994.

[20] O. Colot, R. Devinoy, A. Sombo, and D. de Brucq, *A Colour Image Processing Method for Melanoma Detection*, pp. 562–569, Lecture Notes in Computer Science, 1496. Springer-Verlag, October 1998.