

Combination rules for belief functions

Eric Lefevre¹, Patrick Vannoorenberghe², and Olivier Colot¹

¹ Laboratoire Perception Systèmes Information (PSI), Université/INSA de Rouen, UPRES EA 2120
Place Emile Blondel, B.P. 08, 76131 Mont-Saint-Aignan Cedex, France
(e-mail : Eric.Lefevre@insa-rouen.fr)

² Heudiasyc, Université de Technologie de Compiègne, UMR CNRS 6599
B.P. 20529, 60205 Compiègne Cedex, France
(e-mail : Patrick.Vannoorenberghe@hds.utc.fr)

Abstract. Within the evidence theory, the Dempster's combination rule is of particular importance for the aggregation of different belief functions. Some authors consider this rule to be the only reasonable solution to the combination. However this rule is subject to various controversies because it badly manages the conflict between different bodies of evidence at the normalization step. In this paper, we define a formalism to describe a family of combination operators. We propose to develop a generic framework in order to unify the traditional combination operators used within the framework of the evidence theory. Various operators of this family were tested on sets of synthetic and real belief functions.

1 Introduction

In the recent years, the Dempster-Shafer theory of evidence (Dempster (1967), Shafer (1976)) emerged as an adequate formal framework for handling uncertainty in the context of knowledge representation. In this theory, the data fusion rests on the building of a single belief mass by combination of several mass functions from distinct information sources. This combination, called Dempster's combination rule (or orthogonal sum), has several interesting mathematical properties like commutativity and associativity. The properties of belief functions requires a normalization step. Zadeh (1979) presents a situation where the step of normalization lead to counterintuitive results. In order to deal with this problem, other combination operators have been proposed (Yager (1987), Dubois and Prade (1988), Smets (1990)). However, these operators have more or less satisfactory behaviors. The idea, in this paper, is to distribute proportionally the conflict mass with weights which can be defined by an expert or using a cost function. So, we present a general framework for the combination of information sources which makes it possible to define a family of operators which gathers the various fusion operators of evidence theory. First, we introduce notations allowing to describe the basic concepts of evidence theory (Section 2). We present in section 3, the generic framework which enables to unify the traditional combination operators and lastly some methods for determining the weights. Tests on sets of synthetic and real belief functions are presented in section 4.

2 Background

The aim of this section is to clarify the notations of evidence theory, as used in the rest of the paper. The Dempster-Shafer theory is based on Dempster's original work on lower and upper bounds for a set of compatible probability distributions (Dempster (1967)). Shafer (1976) expanded Dempster's original concepts and produced what is now generally referred to as Dempster-Shafer theory. Among this work, a non probabilistic approach of evidence theory was presented by Smets and Kennes (1994).

Let Θ be a set of mutually exclusive and exhaustive hypotheses about some problem domain. This set is called *frame of discernment* and is defined by $\Theta = \{H_1, \dots, H_n, \dots, H_N\}$. The property of exhaustive assumption, called *close-world* assumption is in opposition to open-world presented by Smets (1990). A piece of evidence that influences our belief concerning the true value of a proposition A can be represented by a *basic belief assignment* or mass function $m(\cdot)$. For each source S_j for $j = \{1, \dots, J\}$, a mass function m_j is defined from 2^Θ to $[0, 1]$ verifying $m_j(\emptyset) = 0$ and $\sum_{A \subseteq \Theta} m_j(A) = 1$. Any subset A of Θ such as $m(A) > 0$ is called a *focal element* of m . We will denote by F_m the set of focal elements of m .

Let us now assume that two distinct pieces of evidence induce two belief function m_1 and m_2 . The *orthogonal sum* of m_1 and m_2 , denoted as $m_\oplus = m_1 \oplus m_2$ is defined as:

$$m_\oplus(A) = \frac{1}{1 - m(\emptyset)} \cdot m_\cap(A) \quad \forall A \subseteq \Theta \quad (1)$$

with the term $m_\cap(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C)$ corresponds to the conjunctive combination rule and the coefficient $m(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C)$ expresses the conflict. When this term is equal to 1, the conflict is total and the sources cannot be aggregated. Conversely, when $m(\emptyset)$ is equal to 0, the sources are in perfect harmony. This combination rule verifies some interesting properties and its use was justified theoretically by several authors (Dubois and Prade (1986), Voorbraak (1991)). However, in some situations, this operator cannot be used. It is the case when the independence constraint of the information sources is not respected or when the mass function building is imprecise. The normalization coefficient depends on this conflict and so results in high sensitivity of the combination rule to small imprecisions of the mass functions, as demonstrated by Zadeh (1979). To avoid the sensitivity problems of the Dempster's combination rule, some authors (Yager (1987), Dubois and Prade (1988), Smets (1990)) have proposed other solutions concerning the interpretation of the conflict. We propose here a generic framework in order to unify these different combination operators.

3 Generic framework

The aim of the proposed combination rules is to distribute the conflicting mass $m(\emptyset)$ on a set of propositions. Let us denote by \mathcal{P} the set of subsets A on which the conflicting mass is distributed. A part of the mass $m(\emptyset)$ is assigned to each subset A according to a weighting factor denoted $w(A, m_1, \dots, m_j, \dots, m_J)$. This weighting factor is a function of the considered subset A and belief functions which have involved the conflict. So, the total mass after aggregation for a proposition A is the sum of the two masses as follows:

$$m(A) = m_{\cap}(A) + m^c(A) \quad \forall A \subseteq \Theta. \quad (2)$$

In equation (2), the second term, denoted $m^c(\cdot)$, is the part of the conflicting mass assigned to subset A . This value can be written as follows:

$$\begin{cases} m^c(A) = w(A, m_1, \dots, m_J).m(\emptyset) & \forall A \in \mathcal{P} \\ m^c(A) = 0 & \text{otherwise} \end{cases} \quad (3)$$

with the constraint $\sum_{A \in \mathcal{P}} w(A, m_1, \dots, m_J) = 1$ so as to respect the property that the sum of mass functions has to be equal to 1. This generic framework allows to rewrite the operators proposed by Smets (1990), Yager (1987) and Dubois and Prade (1988). For each operator, one has just to define the set \mathcal{P} on which the conflicting mass will be distributed and the weighting factors $w(A, m_1, \dots, m_J)$ associated to each subset $A \in \mathcal{P}$.

3.1 Automatic learning by means of cost functions

In this section, we deal with a classification problem. Often, this problem is presented as follows. Let some pattern X has to be classified according to a set $\{H_1, \dots, H_N\}$ of N classes using a training set \mathcal{X} composed of I patterns of the form $e^i = \{X^i, u^i\}$, where X^i denotes a feature vector, and u^i a vector of membership indicator, i.e., $u_k^i = 1$ if the vector X^i belongs to the class k , and $u_n^i = 0$ otherwise. We propose a learning of the weighting factors based on the use of learning set and the minimization of an error criterion. This error criterion is defined as the mean square error between the pignistic probability $BetP$ defined by Smets (1994) in the TBM and the membership indicator to each hypothesis. The mean square error of the learning set vectors is defined as $E_{MS} = \sum_{i=1}^I \sum_{n=1}^N [BetP^{(i)}(H_n) - u_n^i]^2$ with $BetP^{(i)}$ represents the pignistic probability of a vector X^i in the learning set. We determine then the set of the weighting factors by minimization of the mean square error.

3.2 Conflict distribution based on expert valuation

The introduced combination formalism can be useful in case of an additional knowledge in order to solve the conflict. An specialist of the application domain can provide this knowledge. In medical fields, target tracking or obstacle detection, the non-detection rate has important consequences in decision

making. In these fields, the conflicting mass is assigned to the most cautious hypothesis. When no additional information can be provided by an expert, we can adopt a cautious strategy consisting in distributing uniformly the conflict or by learning the weighting factors.

4 Results

At first, we present a comparison between the Dempster's operator and our strategy in terms of resulting belief mass interpretation (Section 4.1). Finally, classification results with a weighting factor learning will be presented in section 4.2.

4.1 Resulting mass distribution

In this test, the method developed by Dencœux (2000) was used for basic belief assignment. This one lies on the computation of a distance function between the vector to be classified and the prototypes issued from a learning. The basic belief assignment is obtained with three prototypes per class, that is to say three belief functions by class. We considered three Gaussian hypotheses with means $\mu_1 = (0, 2)^t$, $\mu_2 = (-2, 0)^t$, $\mu_3 = (2, 0)^t$ and identity variance matrices Σ_1 , Σ_2 and Σ_3 . In Figure 1, one sees the evolution of the maximum value of the combined belief mass in the feature space according to the Dempster's operator and the proposed operator. In both figures, lighter areas correspond to higher maximum value of the belief mass function. The weighting factor used in the distribution process are equal to $1/3$. With its choice, we do not favour any hypothesis in the conflict distribution process. One can observe in

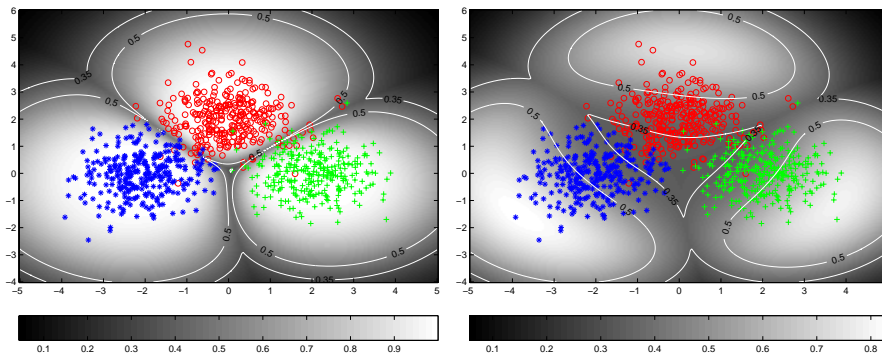


Fig. 1. Maximum value of the belief masses according to: the Dempster's combination rule (left) and our combination rule with $w(H_n, m_1, \dots, m_9) = 1/3$ (right).

the left part of Figure 1 that the transitions between the maximum values of

the masses are abrupt. In the right part of Figure 1, the variation of the belief mass functions is smoother. Indeed, there exists a flat area corresponding to the conflict area when one goes from one maximum of the belief mass to another one. The belief mass value for each hypothesis in this area is close to $1/3$ which emphasizes the conflict and leads to no decision. In conclusion, our approach allows to keep the information given by the conflict for the decision process at the opposite of the Dempster's combination rule.

4.2 Weighting factor learning

For this test, we used the method developed by Zouhal and Dencœux (1998) for the basic belief assignment. The number of neighbors is fixed to 5. We obtain 5 belief functions m_1, \dots, m_5 . We consider a bi-dimensional problem with two hypotheses. The available learning is given by normal probability distributions with means $\mu_1 = (2, 2)^t$, $\mu_2 = (4, 4)^t$ and identity variance matrices Σ_1, Σ_2 . The simulated measures are generated with Gaussian distribution with means $\mu'_1 = (2 + S, 2 + S)^t$, $\mu'_2 = (4 + S, 4 + S)^t$ and variance matrices $\Sigma_1 = \Sigma_2 = I$. This test can represent a classical case of time varying behaviour of the sensors (context evolution). In this example, the measures vary linearly according to a signal S . A validation set allowing to compute the values of the weighting factors is generated on the same model as the test set. The weighting factors are computed by learning with the validation set. The results in term of classification rate are given in Figure 2. According

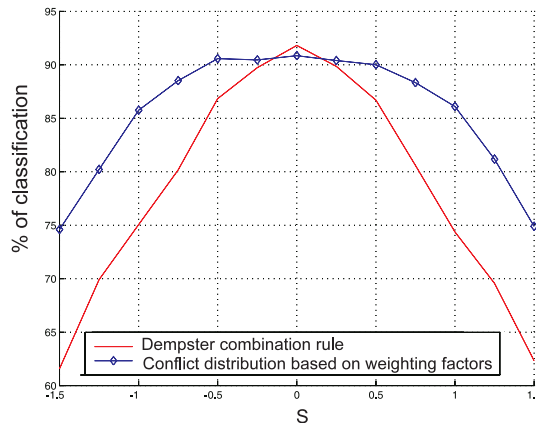


Fig. 2. Evolution of the good classification rate according to S .

to this figure, one can notice that the distribution of the conflicting mass by means of the weighting factor approach improves the good classification rate in comparison with the Dempster's rule of combination. In particular, when

the sensor drift is important ($|S| \in [0.5, 1.5]$), the good classification percentage obtained with our approach is 10% upper than the good classification rate obtained with the Dempster's rule of combination. When the drift is weak ($S \approx 0$), the results given by the two methods are quite similar.

5 Conclusion

In this paper, we have presented a generic framework for the fusion of information sources modeled by means of belief mass functions which allows to define a family of combination operators. Indeed, it is possible to derive different operators based on the definition of a set \mathcal{P} collecting the propositions where the conflicting mass will be distributed and weighting factors associated to each subset $A \in \mathcal{P}$. Several methods have been proposed to obtain the weighting factors. One of this is based on the integration of an additional knowledge an expert could provide in order to solve the conflict. The second one lies on the weighting factor determination by minimizing a mean square error. The tests show the contribution of our approach as well for the resulting mass interpretation by keeping the conflict information.

References

- DEMPSTER, A. (1967): Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics, AMS-38*, 325–339.
- DENOEU, T. (2000): Neural network classifier based on Dempster-Shafer theory. *IEEE Transactions Systems Man and Cybernetics Part A*, 30(2), 131–150.
- DUBOIS, D. and PRADE, H. (1986): On the uncitivity of Dempster rule combination. *International Journal of Intelligent Systems*, 1, 133–142.
- DUBOIS, D. and PRADE, H. (1988): Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.*, 4, 244–264.
- SHAFER, G. (1976): *A Mathematical Theory of Evidence*. Princeton Univerty Press, 1976.
- SMETS, P. (1990): The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 447–458.
- SMETS, P. and KENNES, R. (1994): The transferable belief model. *Artificial Intelligence*, 66(2), 191–234.
- VOORBRAAK, F. (1991): On the justification of Dempster's rule of combination. *Artificial Intelligence*, 48, 171–197.
- YAGER, R.R. (1987): On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41, 93–138.
- ZADEH, L. A. (1979): *On the Validity of Dempster's Rule of Combination of Evidence*, Univerty of California, Berkeley, 1979.
- ZOUHAL, L.M. and DENOEU, T. (1998): An evidence-theoretic k-NN rule with parameter optimization. *IEEE Transactions Systems Man and Cybernetics Part C*, 28(2), 263–271.