A Distance-Based Decision in the Credal Level

Amira Essaid, Arnaud Martin, Grégory Smits, Boutheina Ben Yaghlane

A distance-based decision in the credal level

Amira Essaid[1,2], Arnaud Martin[2], Grégory Smits[2], and Boutheina Ben Yaghlane[3]

[1] LARODEC, University of Tunis, ISG Tunis, Tunisia
[2]IRISA, University of Rennes1, Lannion, France
[3] LARODEC, University of Carthage, IHEC Carthage, Tunisia

Abstract. Belief function theory provides a flexible way to combine information provided by different sources. This combination is usually followed by a decision making which can be handled by a range of decision rules. Some rules help to choose the most likely hypothesis. Others allow that a decision is made on a set of hypotheses. In [6], we proposed a decision rule based on a distance measure. First, in this paper, we aim to demonstrate that our proposed decision rule is a particular case of the rule proposed in [4]. Second, we give experiments showing that our rule is able to decide on a set of hypotheses. Some experiments are handled on a set of mass functions generated randomly, others on real databases.

Keywords: belief function theory, imprecise decision, distance

1 Introduction

Belief function theory [2, 9] allows us to represent all kinds of ignorance and offers rules for combining several imperfect information provided by different sources in order to get a more coherent one. The combination process helps to make decisions later. Decision making consists in selecting, for a given problem, the most suitable actions to take. Today, we are often confronted with the challenge of making decisions in cases where information is imprecise or even not available. In [12], Smets proposed the transferable belief model (TBM) as an interpretation of the theory of belief functions. The TBM emphasizes a distinction between knowledge modeling and decision making. Accordingly, we distinguish the credal level and the pignistic level. In the credal level, knowledge is represented as belief functions and then combined. The pignistic level corresponds to decision making, a stage in which belief functions are transformed into probability functions.

The pignistic probability, the maximum of credibility and the maximum of plausibility are rules that allow a decision on a singleton of the frame of discernment. Sometimes and depending on application domains, it seems to be more convenient to decide on composite hypotheses rather than a simple one. In the literature, there are few works that propose a rule or an approach for making decision on a union of hypotheses [4, 1, 8]. Recently, we proposed a decision rule based on a distance measure [6]. This rule calculates the distance between a combined mass function and a categorical one. The most likely hypothesis to

choose is the hypothesis whose categorical mass function is the nearest to the combined one.

The main topic of this paper is to demonstrate that our proposed decision rule is a particular case of that detailed in [4] and to extend our rule so that it becomes able to give decisions even with no categorical mass functions. We present also our experiments on mass functions generated randomly as well as on real databases.

The remainder of this paper is organized as follows: in section 2 we recall the basic concepts of belief function theory. Section 3 presents our decision rule based on a distance measure proposed in [6]. In section 4, we demonstrate that our proposed rule is a particular case of that proposed in [4]. Section 5 presents experiments and the main results. Section 6 concludes the paper.

2   The theory of belief functions

The theory of belief functions [2, 9] is a general mathematical framework for representing beliefs and reasoning under uncertainty. In this section, we recall some concepts of this theory.

The *frame of discernment* $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ is a set of n elementary hypotheses related to a given problem. These hypotheses are exhaustive and mutually exclusive. The power set of Θ, denoted by 2^Θ is the set containing singleton hypotheses of Θ, all the disjunctions of these hypotheses as well as the empty set.

The Basic belief assignment *(bba)*, denoted by m is a mass function defined on 2^Θ. It affects a value from $[0, 1]$ to each subset. It is defined as:

$$\sum_{A \subseteq 2^\Theta} m(A) = 1. \tag{1}$$

A focal element A is an element of 2^Θ such that $m(A) > 0$. A categorical bba is a bba with a unique focal element such that $m(A) = 1$. When this focal element is a disjunction of hypotheses then the bba models imprecision.

Based on the basic belief assignment, other belief functions (credibility function ad plausibility function) can be deduced.

– Credibility function *bel(A)* expresses the total belief that one allocates to A. It is a mapping from elements of 2^Θ to $[0, 1]$ such that:

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B). \tag{2}$$

– Plausibility function *pl(A)* is defined as:

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \tag{3}$$

The plausibility function measures the maximum amount of belief that supports the proposition A by taking into account all the elements that do not contradict. The value $pl(A)$ quantifies the maximum amount of belief that might support a subset A of Θ.

The theory of belief function is a useful tool for data fusion. In fact, for a given problem and for the same frame of discernment, it is possible to get a mass function synthesizing knowledge from separate and independent sources of information through applying a combination rule. Mainly, there exists three modes of combination:

– Conjunctive combination is used when two sources are distinct and fully reliable. In [10], the author proposed the conjunctive combination rule which is defined as:

$$m_{1\bigcirc\!\!\!\!\cap 2}(A) = \sum_{B \cap C = A} m_1(B) \times m_2(C).\tag{4}$$

The Dempster's rule of combination [2] is a normalized form of the rule described previously and is defined as:

$$m_{1\oplus 2}(A) = \begin{cases} \dfrac{\displaystyle\sum_{B \cap C = A} m_1(B) \times m_2(C)}{1 - \displaystyle\sum_{B \cap C = \emptyset} m_1(B) \times m_2(C)} & \forall A \subseteq \Theta,\ A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases}\tag{5}$$

This rule is normalized through $1 - \displaystyle\sum_{B \cap C = \emptyset} m_1(B) \times m_2(C)$ and it works under the closed world assumption where all the possible hypotheses of the studied problem are supposed to be enumerated on Θ.

– Disjunctive combination: In [11], Smets introduced the disjunctive combination rule which combines mass functions when an unknown source is unreliable. This rule is defined as:

$$m_{1\bigcirc\!\!\!\!\cup 2}(A) = \sum_{B \cup C = A} m_1(B) \times m_2(C)\tag{6}$$

– Mixed combination: In [5], the authors proposed a compromise in order to consider the benefits of the two combination modes previously described. This combination is given for every $A \in 2^{\Theta}$ by the following formula:

$$\begin{cases} m_{DP}(A) = m_{1\bigcirc\!\!\!\!\cap}(A) + \displaystyle\sum_{B \cap C = \emptyset, B \cup C = A} m_1(B) m_2(C)\ \forall A \in 2^{\Theta},\ A \neq \emptyset \\ m_{DP}(\emptyset) = 0 \end{cases}\tag{7}$$

3   Decision Making in the theory of belief functions

In the transferable belief model, decision is made on the pignistic level where the belief functions are transformed into a probability function, named *pignistic probability*. This latter, noted as $BetP$ is defined for each $X \in 2^{\Theta}$, $X \neq 0$ as:

$$betP(X) = \sum_{Y \in 2^{\Theta}, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m(Y)}{1 - m(\emptyset)} \tag{8}$$

where $|Y|$ represents the cardinality of Y.

Based on the obtained pignistic probability, we select the most suitable hypothesis with the maximum $BetP$. This decision results from applying tools of decision theory [4]. In fact, if we consider an entity represented by a feature vector x. A is a finite set of possible actions $A = \{a_1, \ldots, a_N\}$ and Θ a finite set of hypotheses, $\Theta = \{\theta_1, \ldots, \theta_M\}$. An action a_j corresponds to the action of choosing the hypothesis θ_j. But, if we select a_i as an action whereas the hypothesis to be considered is rather θ_j then the loss occurred is $\lambda(a_i|\theta_j)$. The expected loss associated with the choice of the action a_i is defined as:

$$R_{betP}(a_i|x) = \sum_{\theta_j \in \Theta} \lambda(a_i|\theta_j) BetP(\theta_j). \tag{9}$$

Then, the decision consists in selecting the action which minimizes the expected loss. In addition to minimizing pignistic expected loss, other risks are presented in [4].

Decision can be made on composite hypotheses [1, 8]. We present in this paper the Appriou's rule [1] which helps to choose a solution of a given problem by considering all the elements contained in 2^{Θ}. This approach weights the decision functions (maximum of credibility, maximum of plausibility and maximum of pignistic probability) by an utility function depending on the cardinality of the elements. $A \in 2^{\Theta}$ is chosen if:

$$A = \underset{X \in 2^{\Theta}}{argmax}(m_d(X)pl(X)) \tag{10}$$

where m_d is a mass defined by:

$$m_d(X) = K_d \lambda_X \left(\frac{1}{|X|^r} \right) \tag{11}$$

The value r is a parameter in $[0, 1]$ helping to choose a decision which varies from a total indecision when r is equal to 0 and a decision based on a singleton when r is equal 1. λ_X helps to integrate the lack of knowledge about one of the elements of 2^{Θ}. K_d is a normalization factor and $pl(X)$ is a plausibility function.

In the following, we present our decision rule based on a distance measure.

4   Decision rule based on a distance measure

In [6], we proposed a decision rule based on a distance measure. It is defined as:

$$A = argmin(d(m_{comb}, m_A)) \tag{12}$$

This rule aims at deciding on a union of singletons. It is based on the use of categorical bba which helps to adjust the degree of imprecision that has to be kept when deciding. Depending on cases, we can decide on unions of two elements or three elements, etc. The rule calculates the distance between a combined bba m_{comb} and a categorical one m_A . The minimum distance is kept and the decision corresponds to the categorical bba's element having the lowest distance with the combined bba. The rule is applied as follows:

– We consider the elements of 2^{Θ}. In some applications, 2^{Θ} can be of a large cardinality. For this reason, we may choose some elements to work on. For example, we can keep the elements of 2^{Θ} whose cardinality is less or equal to 2.
– For each selected element, we construct its corresponding categorical bba.
– Finally, we apply Jousselme distance [7] to calculate the distance between the combined bba and a categorical bba. The distance with the minimum value is kept. The most likely hypothesis to select is the hypothesis whose categorical bba is the nearest to the combined bba.

Jousselme distance is defined for two bbas m_1 and m_2 as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^t \underline{\underline{D}}(m_1 - m_2)} \tag{13}$$

where $\underline{\underline{D}}$ is a matrix based on Jaccard distance as a similarity measure between focal elements. This matrix is defined as:

$$D(A, B) = \begin{cases} 1 & \text{if A=B=}\emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \forall A, B \in 2^{\Theta} \end{cases} \tag{14}$$

In this paper, we propose to apply the rule through two different manners:

– *Distance type 1* is calculated with categorical bbas ($m(A) = 1$) for all elements of 2^{Θ} except Θ to have an imprecise result rather than a total ignorance.
– *Distance type 2* is calculated with simple bbas such as $m(A) = \alpha$, $m(\Theta) = 1 - \alpha$.

In the following, we show that our proposed rule can be seen as a particular case of that proposed in section 3.

Jousselme distance can be written as:

$$d(m_1, m_2) = \frac{1}{2} \sum_{Y \subseteq \Theta} \sum_{X \subseteq \Theta} \frac{|X \cap Y|}{|X \cup Y|} m(X)m(Y) \tag{15}$$

If we consider the expected loss of choosing a_i, then it can be written as:

$$R_{betP}(a_i|x) = \sum_{Y \in \Theta} \lambda(a_i|Y)BetP(Y).$$

$$R_{betP}(a_i|x) = \sum_{Y \in \Theta} \lambda(a_i|Y) \sum_{X \in \Theta} \frac{|X \cap Y|}{|X|} \frac{m(X)}{1 - m(\emptyset)}. \qquad (16)$$

$$R_{betP}(a_i|x) = \sum_{Y \in \Theta} \sum_{X \in \Theta} \lambda(a_i|Y)\frac{|X \cap Y|}{|X|} \frac{m(X)}{1 - m(\emptyset)}.$$

The equation relative to decision is equal to that for the risk for a value of λ that has to be equal to:

$$\lambda(a_i|Y) = \frac{|X|(1 - m(\emptyset))}{|X \cup Y|}m(X) \qquad (17)$$

In this section, we showed that for a particular value of λ, our proposed decision rule can be considered as a particular case of that proposed in [4]. In the following section, we give experiments and present comparisons between our decision rule based on a distance measure and that presented in [1].

5   Experiments

5.1   Experiments on generated mass functions

We tested the proposed rule [6] on a set of mass functions generated randomly. To generate the bbas, one needs to specify the cardinality of the frame of discernment, the number of mass functions to be generated as well as the number of focal elements. The generated bbas are then combined. We use the Dempster's rule of combination, the disjunctive rule and the mixed rule. Suppose we have a frame of discernment represented as $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and three different sources for which we generate their corresponding bbas as given in Table 1.

Table 1. Three sources with their bbas

	S_1	S_2	S_3
θ_1	0.410	0.223	0.034
θ_2	0.006	0.108	0.300
$\theta_1 \cup \theta_2$	0.039	0.027	0.057
θ_3	0.026	0.093	0.128
$\theta_1 \cup \theta_3$	0.094	0.062	0.04
$\theta_2 \cup \theta_3$	0.199	0.153	0.004
$\theta_1 \cup \theta_2 \cup \theta_3$	0.226	0.334	0.437

We apply combination rules and we get the results illustrated in Table 2.

Table 2. Combination results

	Dempster rule	Disjunctive rule	Mixed rule
θ_1	0.369	0.003	0.208
θ_2	0.227	0	0.128
$\theta_1 \cup \theta_2$	0.025	0.061	0.075
θ_3	0.168	0	0.094
$\theta_1 \cup \theta_3$	0.049	0.037	0.064
$\theta_2 \cup \theta_3$	0.103	0.035	0.093
$\theta_1 \cup \theta_2 \cup \theta_3$	0.059	0.864	0.338

Table 3. Decision results

	Pignistic Probability	Appriou rule	Rule based on distance measure
Dempster rule	θ_1	$\theta_1 \cup \theta_2$	θ_1
Disjunctive rule	θ_1	θ_1	$\theta_1 \cup \theta_2$
Mixed rule	θ_1	θ_1	$\theta_1 \cup \theta_2$

Once the combination is performed, we can make decision. In Table 3, we compare between the results of three decision rules, namely the pignistic probability, the Appriou's rule with r equal to 0.5 as well as our proposed decision rule based on distance measure.

Table 3 shows the decision results obtained after applying some combination rules. We depict from this table that not all the time the rule proposed by Appriou gives a decision on a composite hypotheses. In fact, as shown in Table 3, the application of disjunctive rule as well as the mixed rule lead to a decision on a singleton which is θ_1. This is completely different from what we obtain when we apply our proposed rule which promotes a decision on union of singletons when combining bbas. The obtained results seems to be convenient especially that the disjunctive and the mixed rules help to get results on unions of singletons.

5.2 Experiments on real databases

To test our proposed decision rule, we do some experiments on real databases (IRIS[1] and HaberMan's survival[2]). Iris is a dataset contaning 150 instances, 4 attributes and 3 classes where each class refers to a type of iris plant. HaberMan is a dataset containing results study conducted at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer. This dataset contains 306 instances, 3 attributes and 2 classes

---

[1] http://archive.ics.uci.edu/ml/datasets/Iris
[2] http://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival

(1: patient survived 5 years or longer, 2: patient died within 5 years). For the classification, our experiments are handled in two different manners.

– First, we apply the k-NN classifier [3]. The results are illustrated in a confusion matrix as shown in Table 4 (left side).
– Second, we modify the k-NN classifier's algorithm based on the use of Dempster rule of combination, to make it able to combine belief functions through the mixed rule. Then, Appriou's rule and our proposed decision rule are applied to make decision. Results are illustrated in Table 4.

Table 4. Confusion Matrices for Iris

k-NN classifier			
	θ_1	θ_2	θ_3
θ_1	11	0	0
θ_2	0	11	2
θ_3	0	0	16

Appriou's rule						
	θ_1	θ_2	$\theta_1 \cup \theta_2$	θ_3	$\theta_1 \cup \theta_3$	$\theta_2 \cup \theta_3$
θ_1	11	0	0	0	0	0
θ_2	0	15	0	0	0	0
θ_3	0	1	0	13	0	0

Our decision rule						
	θ_1	θ_2	$\theta_1 \cup \theta_2$	θ_3	$\theta_1 \cup \theta_3$	$\theta_2 \cup \theta_3$
θ_1	10	0	0	0	0	0
θ_2	0	12	0	2	0	1
θ_3	0	0	0	13	0	2

The same tests are done for HaberMan's survival dataset. The results of applying k-NN classifier, Appriou's rule and our decision rule are given respectively in Table 5. For the classification of 40 sets chosen randomly from Iris, we remark that with the k-NN classifier, all the sets having θ_1 and θ_3 as corresponding classes are well classified and only two originally belonging to class θ_2 were classified as θ_3. Appriou's rule gives a good classification for sets originally belonging to classes θ_1 and θ_2 and thus promoting a result on singletons rather than on a union of singletons.

Considering the results obtained when applying our decision rule based on a distance type 1, we note that only 2 sets are not well classified and that 3 have $\theta_2 \cup \theta_3$ as a class. The obtained results are good because our method is based on an imprecise decision which is underlined by the fact of obtaining $\theta_2 \cup \theta_3$ as a class.

Table 5. Confusion Matrices for HaberMan's survival

k-NN classifier		
	θ_1	θ_2
θ_1	34	4
θ_2	12	6

Appriou's rule			
	θ_1	θ_2	Θ
θ_1	34	4	0
θ_2	12	6	0

Our decision rule			
	θ_1	θ_2	Θ
θ_1	34	4	0
θ_2	12	6	0

Considering HaberMan's survival dataset, we note that the k-NN classifier, Appriou's rule as well as our decision rule give the same results where among the sets originally belonging to θ_1, 34 are well classified and among the 18 belonging to θ_2, only 6 are well classified. We obtain the same results as the other rules

because the HaberMan's survival dataset has only two classes and our method is based on getting imprecise decisions and excluding the ignorance.

All the experiments given previously are based on the use of distance type 1. The results shown below are based on distance type 2. In fact, we consider a simple bba and each time, we assign a value α to an element of 2^{Θ}. The tested rule on Iris as illustrated in Table 6 (left side) gives better results with an $\alpha < 0.8$. In addition to that, we obtained decisions on a union of singletons. The tests done on HaberMan's survival as given in Table 6 (right side) shows that with $\alpha > 0.5$, we obtain a better rate of good classification although we did not obtain a good classification for the class θ_2 and no set belongs to Θ. We aim in the future to make experiments on other datasets because HaberMan's survival, for example, does only have 2 classes, so we do not have enough imprecise elements.

Table 6. Rates of good classification

	$\alpha < 0.8$	$\alpha >= 0.8$
Iris	0.95	0.675

	$\alpha <= 0.2$	$\alpha \in [0.3, 0.5]$	$\alpha > 0.5$
HaberMan's survival	0.786	0.803	0.821

6   Conclusion

In this paper, we presented a rule based on a distance measure. This decision rule helps to choose the most likely hypothesis based on the calculation of the distance between a combined bba and a categorical bba. The aim of the proposed decision rule is to give results on composite hypotheses. In this paper, we demonstrated that our proposed rule can be seen as a particular case of that proposed in [4]. We presented also the different experiments handled on generated mass functions as well as on real databases.

References

1. Appriou, A.: Approche générique de la gestion de l'incertain dans les processus de fusion multisenseur. Traitement du Signal 22, pp. 307–319, (2005)
2. Dempster, A.P.: Upper and Lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics, volume 38, pp. 325–339 (1967)
3. Denoeux, T.: A k-nearest neighbor classification rule based on Dempster-Shafer Theory, IEEE Transactions on Systems, Man, and Cybernetics, 25 (5), pp. 804–813, (1995)
4. Denoeux, T.: Analysis of evidence-theoric decision rules for pattern classification, Pattern Recognition 30 (7), pp. 1095–1107, (1997)
5. Dubois, D., Prade, H.: Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence, 4, pp. 244–264 (1988)

6. Essaid, A., Martin, A., Smits, G., Ben Yaghlane, B.: Uncertainty in ontology matching: a decision rule based approach. In proceeding of the International Conference on Information Processing and Mangement Uncertainty, pp. 46–55 (2014)
7. Jousselme, A.L., Grenier, D., Bossé, E.: A New Distance Between Two Bodies of Evidence, Information Fusion, 2, pp. 91–101 (2001)
8. Martin, A., Quidu, I.: Decision support with belief functions theory for seabed characterization. In proceeding of the International Conference on Information Fusion, pp.1–8 (2008)
9. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
10. Smets, P.: The Combination of Evidence in the Transferable Belief Model. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(5), pp. 447–458 (1990)
11. Smets, P.: Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate Reasoning, 9(1), pp. 1–35 (1993)
12. Smets, P. and Kennes, R.: The Transferable Belief Model. Artificial Intelligent 66, pp. 191–234 (1994)
13. Yager, R.R.: On the Dempster-Shafer framework and new combination rules. Information Sciences, 41, pp. 93–137 (1987)