About sources dependence in the theory of belief functions

Mouna Chebbah, Arnaud Martin and Boutheina Ben Yaghlane

Abstract In the theory of belief functions many combination rules are proposed in the purpose of merging and confronting several sources opinions. Some combination rules are used when sources are cognitively independent whereas others are specific to dependent sources. In this paper, we suggest a method to quantify sources degrees of dependence in order to choose the more appropriate combination rule. We used generated mass functions to test the proposed method.

1 Introduction

Decision making is more and more difficult when using imperfect data, however information can be imprecise, uncertain and even not available. Usually decision is made using precise and certain data but available information are not always so. Many theories manage uncertainty such as the *theory of probabilities*, the *theory of fuzzy sets*, the *theory of possibilities* and the *theory of belief functions*. Within imperfect environment, combining several imperfect information helps users and decision makers to reduce the degree of uncertainty by confronting several opinions. The theory of belief functions presents a strong framework for combination.

To combine uncertain information many combination rules can be used. Some of these combination rules are used when sources are cognitively independent like [6, 7, 9, 10, 13] but the cautious rule [5] is applied when sources are dependent.

Mouna Chebbah
LARODEC Laboratory, ISG Tunis, 41 Rue de la liberté, Cité Bouchoucha 2000 Le Bardo, Tunisia
IRISA, University of Rennes 1, rue E. Branly, 22300 Lannion, e-mail: Mouna.Chebbah@gnet.tn

Arnaud Martin
IRISA, University of Rennes 1, rue E. Branly, Lannion, e-mail: Arnaud.Martin@univ-rennes1.fr

Boutheina Ben Yaghlane
LARODEC Laboratory, IHEC Carthage, Carthage Présidence 2016, Tunisia, e-mail: boutheina.yaghlane@ihec.rnu.tn

A source is assumed to be cognitively independent towards another one when the knowledge of the belief of that source does not affect the belief of the first one. In some cases, like when a source is completely dependent on another source, the user can decide to discard the dependent source and its mass functions from the combination.

Some researches are focused on the sources statistical dependence such as [1, 2] and others [12, 11] tackled the cognitive dependence between variables. This paper is focused on sources dependence measuring. Thus, we suggest a method to estimate the dependence between sources.

In the following, we introduce preliminaries of the theory of belief functions in the second section. In the third section, the independence measure is presented. This independence is estimated in three steps, in the first step a clustering technique is applied then similar clusters are matched in the second step and finally a weight is affected to matched clusters. This method is tested on random mass functions in the fourth section. Finally, conclusions are drawn.

2 Theory of belief functions

The theory of belief functions was introduced by [4] and [12] and so called *Dempster-Shafer theory* to model imperfect information held by a source (an expert, a belief holder, etc.). In this section, we will remind some basic notions of this theory as seen in the transferable belief model [10].

The *frame of discernment* $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ is a set of n elementary and mutually exclusive and exhaustive hypotheses. These hypotheses are all the possible and eventual solutions of the problem under study. The *power set* 2^Ω is the set of all subsets made up of hypotheses and union of hypotheses from Ω. The *basic belief assignment (bba)* also called *mass function* is a function defined on the power set 2^Ω and affects a value from $[0,1]$ such that: $\sum\limits_{A \subseteq \Omega} m(A) = 1$. We can also assume that: $m(\emptyset) = 0$. A subset A having a strictly positive mass is called *focal element*. The mass allocated to this focal element A is the source's degree of belief that the solution of the problem under study is in A. In the theory of belief functions, a great number of combination rules [6, 7, 9, 10, 13] are used to summarize all combined mass functions into only one mass function reflecting all the sources beliefs. The first combination rule was proposed by Dempster in [4] and is defined for two distinct mass functions m_1 and m_2:

$$m_{1 \textcircled{\scriptsize$\cap$} 2}(A) = (m_1 \textcircled{\scriptsize$\cap$} m_2)(A) = \begin{cases} \dfrac{\sum\limits_{B \cap C = A} m_1(B) \times m_2(C)}{1 - \sum\limits_{B \cap C = \emptyset} m_1(B) \times m_2(C)} & \forall A \subseteq \Omega,\ A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases} \quad (1)$$

Dempster's rule of combination together with other rules [6, 7, 9, 10, 13] are used to combine independent mass functions. In the case of dependent sources, the cautious rule [5] can be applied. After the combination, the pignistic probability $\mathrm{BetP}(A)$ is generally used to decide.

3 Independence

Independence concept was first introduced in probability theory in the purpose of studying dependent statistical variables. In the probability theory, two variables A and B are assumed to be independent if one of these equivalent conditions is satisfied: $P(A \cap B) = P(A) * P(B)$ or $P(A|B) = P(A)$. *Statistical independence* is generalized from probability theory to the theory of belief functions [1, 2]. Mass functions can be seen as subjective probabilities held by sources (experts, belief holders, ...) who can communicate, thus *cognitive independence* is specially defined in the theory of belief functions. A definition of cognitive independence was first proposed by Shafer ([12], page 149) as "two frames of discernment may be called cognitively independent with respect to the evidence if new evidence that bears on only one of them will not change the degree of support for propositions discerned by the other". Smets [11] claims that two variables are independent when the knowledge of the value taken by one of them does not affect our belief about the other. This paper is not focused on variables independence but on sources independence.

Definition 1. Two sources are independent when the knowledge of the belief provided by one source does not affect the belief of the other source, otherwise these sources are dependent.
Not only communicating sources are considered to be dependent but also sources having the same background of knowledge since their beliefs are similar. In this paper, mass functions provided by two sources are studied in order to reveal any dependence between them. Therefore, the aim is to find dependence between sources if it exists. In the following, we define an independence measure $\mathrm{I_d}$, ($\mathrm{I_d}(s_1, s_2)$ is the independence of s_1 towards s_2) verifying the following axioms:

1. Non-negativity: The independence of a source s_1 on an another source s_2, $\mathrm{I_d}(s_1, s_2)$ cannot be negative, it is a positive or null degree.
2. Normalization: Source independence $\mathrm{I_d}$ is a degree on $[0, 1]$, it is null when the source is dependent from the other one, equal to 1 when it is completely independent and a degree in $[0, 1]$ otherwise.
3. Non-symmetry: If a source s_1 is dependent on a source s_2, s_2 is not necessarily dependent on s_1. Even if s_1 and s_2 are mutually dependent, degrees of dependence are not the same.
4. Identity: $\mathrm{I_d}(s_1, s_1) = 0$. A source is completely dependent from it self.

If two sources s_1 and s_2 are dependent, there will be a relation between their belief functions. The main idea of this paper is to classify mass functions provided by each source, then a study of the similarities between cluster repartitions can reveal any dependence between sources. Once clustering is performed, the idea is to study the sources overall behavior. The proposed method is in three steps, in the first step mass functions of each source are classified then in the second step similar clusters are matched and finally the weights of the linked clusters are quantified in the third step.

3.1 Clustering

In this paper, we use a modified C-means algorithm with the distance on belief functions given by [8] such as in [3] to classify mass functions of one source. The number of clusters C has to be also known, a set T contains n objects $o_i : 1 \leq i \leq n$ which values m_i are belief functions defined on a frame of discernment Ω. For example, a doctor is diagnosing the disease of n patients and giving each time a mass function as an uncertain diagnostic. In that case, patients are considered as these objects o_i to be classified, the frame of discernment Ω contains all the possible diseases and m_i is the mass function provided by the doctor when diagnosing each patient o_i. In this section a clustering technique is performed on mass functions m_i provided by the same source in order to study the overall behavior of a source.

This clustering technique is based on a dissimilarity measure which is used to quantify the dissimilarity of an object o_i towards a cluster Cl_k. The dissimilarity D of the object o_i towards the cluster Cl_k is the mean of distances between m_i the mass function value of the object o_i and all the n_k mass functions classified into the cluster Cl_k as follows:

$$D(o_i, Cl_k) = \frac{1}{n_k} \sum_{j=1}^{n_k} d(m_i^{\Omega}, m_j^{\Omega}) \tag{2}$$

$$d(m_1^{\Omega}, m_2^{\Omega}) = \sqrt{\frac{1}{2}(m_1^{\Omega} - m_2^{\Omega})^t \underline{\underline{D}}(m_1^{\Omega} - m_2^{\Omega})}, \underline{\underline{D}}(A,B) = \begin{cases} 1 & \text{if } A = B = \emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \forall A, B \in 2^{\Omega} \end{cases} \tag{3}$$

Each object is affected to the most similar cluster in an iterative way until reaching an unchanged cluster partition. It is obvious that the number of clusters C has to be fixed. In this paper, we suppose that C is the cardinality of the frame of discernment. In a classification problem, the cardinality of the frame of discernment is the number of classes that is why we choose $C = |\Omega|$ in this paper.

3.2 Cluster matching

Clustering technique, given in section 3.1, is used to classify mass functions provided by both sources s_1 and s_2, the number of clusters is assumed to be the cardinality of the frame of discernment. After the classification, both mass functions provided by s_1 and s_2 are distributed on C clusters. Once clustering performed the most similar clusters have to be linked, a cluster matching is performed for both clusters of s_1 and that of s_2. The dissimilarity between two clusters Cl_{k_1} of s_1 and Cl_{k_2} of s_2 is the mean of distances between objects $o_l \in Cl_{k_1}$ and $o_w \in Cl_{k_2}$:

$$\delta^1(Cl_{k_1}, Cl_{k_2}) = \frac{1}{n_{k_1}} \sum_{l=1}^{n_{k_1}} \frac{1}{n_{k_2}} \sum_{w=1}^{n_{k_2}} d(o_l, o_w) \tag{4}$$

We note that n_{k_1} is the number of objects on the cluster Cl_{k_1}, δ^1 is the dissimilarity towards the source s_1 and d is the distance defined by equation (3). It is obvious that $d(o_l, o_w) \in [0,1]$. $\delta^1(Cl_{k_1}, Cl_{k_2})$ is the mean of pairwise distances between objects of Cl_{k_1} and Cl_{k_2}, thus $\delta^1(Cl_{k_1}, Cl_{k_2}) \in [0,1]$.

A dissimilarity matrix M_1 containing dissimilarities of clusters of s_1 according to clusters of s_2, and M_2 the dissimilarity matrix between clusters of s_2 and clusters of s_1 are defined as follows:

$$M_1 = \begin{pmatrix} \delta^1_{11} & \delta^1_{12} & \dots & \delta^1_{1C} \\ \dots & \dots & \dots & \dots \\ \delta^1_{k1} & \delta^1_{k2} & \dots & \delta^1_{kC} \\ \dots & \dots & \dots & \dots \\ \delta^1_{C1} & \delta^1_{C2} & \dots & \delta^1_{CC} \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} \delta^2_{11} & \delta^2_{12} & \dots & \delta^2_{1C} \\ \dots & \dots & \dots & \dots \\ \delta^2_{k1} & \delta^2_{k2} & \dots & \delta^2_{kC} \\ \dots & \dots & \dots & \dots \\ \delta^2_{C1} & \delta^2_{C2} & \dots & \delta^2_{CC} \end{pmatrix} \tag{5}$$

We note that $\delta^1_{k_1k_2}$ is the dissimilarity between Cl_{k_1} of s_1 and Cl_{k_2} of s_2 and $\delta^2_{k_1k_2}$ is the dissimilarity between Cl_{k_2} of s_2 and Cl_{k_1} of s_1 and $\delta^1_{k_1k_2} = \delta^2_{k_2k_1}$. The dissimilarity matrix M_2 of s_2 is the transpose of the dissimilarity matrix of s_1 noted M_1. Therefore, a unique matrix M_1 can be used to store dissimilarities between all clusters of s_1 and that of s_2. Clusters of s_1 are matched to the nearest clusters of s_2, a cluster Cl_{k_1} of s_1 is matched to the cluster having the minimal dissimilarity $\delta^1_{k_1.}$ and a cluster Cl_{k_2} of s_2 is matched to the cluster having the minimal dissimilarity $\delta^2_{k_2.} = \delta^1_{.k_2}$.Two clusters of s_1 can be linked to the same cluster of s_2. The output are C cluster matchings of s_1, C different cluster matchings of s_2 and $2 \times C$ dissimilarity values of each matched clusters.

3.3 Cluster independence

Once cluster matching is obtained, the degree of independence and dependence between sources are quantified in this step. A set of matched clusters is obtained for both sources and a mass function can be used to quantify the independence between each couple of clusters. Suppose that the cluster Cl_{k_1} of s_1 is matched to Cl_{k_2} of s_2, a mass function m defined on the frame of discernment $\Omega_I = \{Dependent\ Dep, Independent\ Ind\}$ describes how much this couple of clusters is independent or dependent as follows:

$$\begin{cases} m_{k_1k_2}^{\Omega_I}(Dep) = \alpha\,(1-\delta_{k_1k_2}^1) \\ m_{k_1k_2}^{\Omega_I}(Ind) = \alpha\,\delta_{k_1k_2}^1 \\ m_{k_1k_2}^{\Omega_I}(Dep \cup Ind) = 1-\alpha \end{cases} \tag{6}$$

The coefficient α is used to take into account the number of mass functions in each cluster. Mass functions defining sources dependence are not provided by any source whereas they are estimations of the sources dependence. The coefficient α is not the reliability of any source but it can be seen as the reliability of the estimation. Therefore, the more a cluster contains mass functions the more our dependence measure estimation of that cluster is reliable. For example, let us take two clusters the first one containing only one mass function and the second one containing 100 mass functions, it is obvious that the dependency estimation of the second cluster is more precise and significant than the dependency estimation of the first one.

The obtained mass functions quantify the independence of each matched clusters according to each source. Therefore, C mass functions are obtained for each source such that each mass function quantifies the independence of each couple of matched clusters. The combination of C mass functions for each source using Dempster's rule of combination defined by equation (1) is a mass function m^{Ω_I} defining the whole dependence of one source towards the other one: $m^{\Omega_I} = \textcircled{\cap}\, m_{k_1k_2}^{\Omega_I}$.

Two different mass functions $m_{s_1}^{\Omega_I}$ and $m_{s_2}^{\Omega_I}$ are obtained for s_1 and s_2 respectively. We note that $m_{s_1}^{\Omega_I}$ is the combination of C mass functions representing the dependence of matched clusters defined using equation (6). These mass functions are different since cluster matchings are different which verifies the axiom of non-symmetry. $\delta_{k_1k_2}^1, \delta_{k_2k_1}^2 \in [0,1]$ which verifies the non-negativity and the normalization axioms. Finally, pignistic probabilities are computed from these mass functions in order to decide about these sources independence $\mathrm{I_d}$ such that $\mathrm{I_d}(s_1,s_2) = \mathrm{BetP}(Ind)$ and $\overline{\mathrm{I_d}}(s_1,s_2) = \mathrm{BetP}(Dep)$, if $\mathrm{BetP}(Ind) > 0.5$ we can claim that the corresponding source is independent from the other one otherwise it is dependent.

4 Examples on generated mass functions

To test this method we used generated mass functions. Thus, two sets of mass functions are generated for two sources s_1 and s_2. We note that the number of sources is always two (s_1 and s_2) because the dependence is a binary relationship. Thus a source is dependent or independent according to another one. For the sake of simplicity, we take here the discounting factor $\alpha = 1$, thus mass functions are not discounted. To generate bbas, some information are needed: the cardinality of the frame of discernment $|\Omega|$, the number of mass functions. Mass functions are generated as follows:

1. The number of focal elements F is chosen randomly from $[1, |2^{\Omega}|]$. The F focal elements are also chosen randomly from the power set.

2. The interval $[0,1]$ is divided randomly into F continuous sub intervals.
3. A random mass from each sub interval is attributed to focal elements. Masses are attributed to focal elements chosen in the first step. The complement to 1 of the attributed masses sum is affected to the total ignorance $m(\Omega)$.

This method is used to generate a random mass function, thus the number of focal elements and masses are attributed randomly. Using the pignistic transformation, the decided class is not known from the beginning. In some cases generated mass functions are corrected in order to correct the classification result as follows:

1. Generate a mass function as described above,
2. to change the classification result of the generated mass function, masses affected to each focal element are transfered to its union with the decided class.

Dependent sources: When sources are dependent, they are either providing similar belief functions with the same decided class (using the pignistic transformation) or one of the sources is saying the opposite of what says the other one. In the case of sources deciding the same class, the decided class of one source is directly affected by that of the other one. To test this case, we generated 100 mass functions on a frame of discernment of cardinality 5. Both sources are classifying objects in the same way. Applying the method described above, the obtained mass function defined on the frame $\Omega_I = \{Ind,\ \ Dep\}$ and describing the independence of s_1 towards s_2 is $m(Ind) = 0.0217$, $m(Dep) = 0.9783$ meaning that $\mathrm{I_d}(s_1,s_2) = 0.0217$ and $\overline{\mathrm{I_d}}(s_1,s_2) = 0.9783$. Thus s_1 is highly dependent on s_2.
The mass function of the independence of s_2 according to s_1 is $m(Ind) = 0.022$, $m(Dep) = 0.978$. It proves that s_2 is also dependent on s_1 because $\overline{\mathrm{I_d}}(s_2,s_1) = 0.978$.

When sources are indirectly dependent, one of them is saying the opposite of the other one. In other words, when the decision class of the first source is a class A, the second source may classify this object in any other class but not A. In that case, the obtained mass function for the dependence of s_1 according to s_2 is $m(Ind) = 0.0777$, $m(Dep) = 0.9223$ meaning that s_1 is dependent on s_2 because $\overline{\mathrm{I_d}}(s_1,s_2) = 0.9223$.
The mass function of the independence of s_2 according to s_1 is $m(Ind) = 0.0805$, $m(Dep) = 0.9195$, thus s_2 is also highly dependent on s_1 and $\overline{\mathrm{I_d}}(s_2,s_1) = 0.9195$.
Thus s_1 is dependent towards s_2 with a degree 0.978 and s_2 is dependent towards s_1 with a degree 0.9195. s_1 and s_2 are mutually dependent.

Independent sources: We generated randomly 100 mass functions for both sources s_1 and s_2. The number of focal elements is randomly chosen on the interval $[1,\frac{2^{\Omega}}{4}]$ rather than $[1,2^{\Omega}]$ to reduce the number of focal elements. The obtained mass function of the independence of s_1 according to s_2 is $m(Ind) = 0.7211$, $m(Dep) = 0.2789$. The mass function of the independence of s_2 according to s_1 is $m(Ind) = 0.6375$, $m(Dep) = 0.3625$. Thus s_1 is independent towards s_2 because $\mathrm{I_d}(s_1,s_2) = 0.7211$ and s_2 is independent towards s_1 because $\mathrm{I_d}(s_2,s_1) = 0.6375$. s_1 and s_2 are mutually independent.

5 Conclusion

Combining mass functions provided by different sources is helpful when making decision. The choice of the combination rule is conditioned on the sources dependence, thus the cautious rule is especially used when sources are dependent but other rules can be applied with independent sources. In this paper, we suggested a method estimating the dependence degree of one source towards another one. As a future work, we may try to estimate the dependence of one source according to many other sources and not only one source. When one source is dependent on another one, this dependence can be direct (positive dependence) or indirect (negative dependence). Thus, we will also quantify the positive and negative dependence in the case of dependent sources. We will also define the discounting factor which will be a function of the number of mass functions. Finally, we will use the discounting operator in order to take into account the number of provided mass functions because we cannot decide on the sources independence if they do not provide a sufficient number of mass functions.

References

1. Ben Yaghlane, B., Smets, P., Mellouli, K. : Belief function independence: I. The marginal case. International Journal Approximate Reasoning 29, vol. 1, pp. 47–70 (2002).
2. Ben Yaghlane, B., Smets, P., Mellouli, K. : Belief function independence: II. The conditional case. International Journal Approximate Reasoning 31, vol. 1-2, pp. 31–75 (2002).
3. Ben Hariz, S., Elouedi, Z. and Mellouli, K.: Clustering Approach Using Belief Function Theory. In: Euzenat, J., Domingue, J. (eds.) AIMSA'2006, LNCS (LNAI), vol. 4183, pp. 162–171. Springer, Heidelberg (2006).
4. Dempster, A. P.: Upper and Lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics, vol. 38, pp. 325–339 (1967).
5. Denoeux, T.: The cautious rule of combination for belief functions and some extensions. In Proceedings of the Ninth International Conference on Information Fusion (FUSION 2006). Florence, Italy (2006).
6. Dubois, D., Prade, H.: Representation and combination of uncertainty with belief functions and possibility measures. Computational Intelligence, vol. 4, pp. 244-264 (1988).
7. Martin, A., Osswald, C.: Toward a combination rule to deal with partial conflict and specificity in belief functions theory. In Int. Conf. on Information Fusion. Québec, Canada (2007).
8. A.-L. Jousselme, D. Grenier and E. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, pp. 91–101, 2001.
9. Murphy, C.K.: Combining belief functions when evidence conflicts. Decision Support Systems 29, 1–9 (2000).
10. Smets, P., Kennes, R.: The Transferable Belief Model. Artificial Intelligent, vol. 66, pp. 191–234 (1994).
11. Smets, P.: Belief Functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. Classic Works of the Dempster-Shafer Theory of Belief Functions 2008: 633-664.
12. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976).
13. Yager, R. R.: On the Dempster-Shafer Framework and New Combination Rules. Information Sciences, vol. 41, pp. 93–137 (1987).