Idempotent Conjunctive Combination of Belief Functions by Distance Minimization

John Klein, Sébastien Destercke, Olivier Colot

Idempotent conjunctive combination of belief functions by distance minimization

John Klein, Sebastien Destercke, and Olivier Colot

Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France
Technologic University of Compiegne, CNRS, UMR 7253 - Heudiasyc, Centre de Recherche de Royallieu, Compigne, France
{john.klein,olivier.colot}@univ-lille1.fr
sebastien.destercke@hds.utc.fr

Abstract. When combining multiple belief functions, designing a combination rule that selects the least informative belief function among those more informative than each of the combined ones is a difficult task. Such rules, commonly depicted as "cautious", are typically required to be idempotent, since when one is cautious, combining identical information should not lead to the reinforcement of some hypothesis. However, applying the least commitment principle using partial orders is in general not straightforward, mainly due to the non-uniqueness of solutions. Building upon previous work, this paper investigates the use of distances compatible with such partial orders to determine a unique solution to the combination problem. The obtained operators are conjunctive, idempotent and commutative, but lack associativity. They are, however, quasi-associative allowing sequential combinations at no extra complexity.

Keywords: conjunctive combination, idempotence, belief function, distance, optimization

1 Introduction

Combining pieces of evidence coming from different sources of information is one of the most frequently studied problem in the belief function theory. In particular, there exist a rich literature proposing alternatives to Dempster's rule when this latter does not apply, that is when sources of information are either unreliable or non-independent, or both. This paper deals with the second issue, that is the one concerning source independence, and more particularly with the case where this dependence is ill-known and hard to assess.

Under such an assumption, it is common to adopt a cautious approach, also known under least-commitment principle [7] (LCP). A natural consequence of this principle is that if all the sources provide the same mass function, then the result of the combination should be this very mass function, or in other words the combination should be idempotent. However, if idempotence is a consequence of the LCP, satisfying idempotence does not imply satisfying the LCP.

As shown by Dubois and Yager [8], there is virtually an infinity of ways to derive idempotent combination rules, not all of them necessarily following a least-commitment principle. For instance, Cattaneo [2] provides an idempotent rule following a conflict-minimization approach, which may lead to non-least committed results [4].

To satisfy the LCP, we therefore must add additional constraints on the combination results. One such natural constraint is to consider a partial order over informative content of mass functions, and to require the combination result to be one of the maximal element of this partial order within the subset of possible combination results. Unfortunately, such an approach will very often lead to multiple solutions corresponding to all possible maximal elements [6]. Denœux [3] shows that using the canonical decomposition and the associated partial order leads to a unique LCP, idempotent solution, yet this solution has two limitations: the set of possible combination results is much reduced, leading to a not so conservative behavior (as we will see on a simple example in Section 4, and as already pointed out in [4]), and the combination only apply to specific (i.e., non-dogmatic) mass functions.

In this paper, we take inspiration from some of our previous work [9] studying the consistency of distances with partial orders comparing informative contents to propose a new way to derive cautious combination rules. Our approach departs from previous ones, as it is formulated as an optimization problem that naturally satisfies the LCP principle (similarly to what is done by Cattaneo [2] for conflict minimization). The interest of this approach is that if the distance is chosen so as to minimize a strictly convex objective function, we are guaranteed to have a unique solution that satisfies the LCP and is easy to compute. The bulk of the proposal is contained in Section 3, where we present the combination approach and study its properties. Sections 2 and 4 respectively recalls the basics needed in this paper and (briefly) compares our proposal with respect to existing ones.

2 Preliminaries and problem statement

This section briefly recalls the basics of evidence theory (due to space limitations, we will provide references for details).

2.1 Basic concepts

A body of evidence $\mathcal{E}_i$ defined on the space $\Omega = \{\omega_1, \dots, \omega_n\}$ will be modeled by a mass function $m_i : 2^\Omega \to [0,1]$ that sums up to one, *i.e.*, $\sum_{E \subseteq \Omega} m(E) = 1$. In evidence theory, this basic tool models our uncertainty about the true value of some quantity (parameter, variable) lying in Ω. The cardinality of 2^Ω is denoted by $N = 2^n$. A set A is a **focal element** of m_i iff $m_i(A) > 0$. A mass function assigning a unit mass to a single focal element A is called **categorical** and denoted by m_A: $m_A(A) = 1$. If $A \neq \Omega$, the mass function m_A is equivalent to providing the set A as information, while the **vacuous** mass function m_Ω represents ignorance.

Several alternative set functions are commonly used in the theory of belief functions and encode the same information as a given mass function m_i. The **belief**, **plausibility** and **commonality** functions of a set A are defined as

$$bel_i(A) = \sum_{E \subseteq A, E \neq \emptyset} m_i(E), \;\; pl_i(A) = \sum_{E \cap A \neq \emptyset} m_i(E), \;\; q_i(A) = \sum_{E \supseteq A} m_i(E)$$

and respectively represent how much A is implied, consistent and common by the actual evidence.

In this paper, we will also use the **conjunctive weight function** denoted by w_i introduced by Smets [10]. It is only defined for mass functions with $m_i(\Omega) \neq 0$ (*i.e.* non-dogmatic mass functions). We refer to Denœux [3] for a thorough presentation of the conjunctive weight function.

2.2 Comparing mass functions with respect to informative content

When considering two mass functions m_1 and m_2 providing information about the same quantity, a natural question is to wonder if one of these two is more informative than the other one. This question can be answered if the **mass space** $\mathcal{M}$, *i.e.* the set of mass functions over Ω, is endowed with a relevant partial order $\sqsubseteq$ with $m_1 \sqsubseteq m_2$ when m_1 is more informative than m_2. Informative content related partial orders should extend set inclusion, since when $A \subseteq B$, A is more informative than B. Such partial orders[1] can be directly obtained by considering inequality between the set functions $f \in \{pl, q, w\}$, by stating that m_1 is **f-included** in m_2, denoted $m_1 \sqsubseteq_f m_2$, if $f_1 \leq f_2$ where $\leq$ is the element-wise inequality.

Each of these orders is partial in the sense that in general there are some incomparable pairs (m_1, m_2), *i.e.* $m_1 \not\sqsubseteq m_2$ and $m_2 \not\sqsubseteq m_1$. There exist implications between them, as we have

$$m_1 \sqsubseteq_w m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2 \end{cases}. \qquad (1)$$

2.3 Distances and partial orders compatibility

Another way to compare mass functions is by measuring how distant they are. An **evidential distance** is a function $d : \mathcal{M} \times \mathcal{M} \to [0, \infty]$ that satisfies the symmetry, definiteness and triangle inequality properties. In [9], we have formalized the idea of compatibility between a distance and a partial order in the following way:

Definition 1 *Given a partial order $\sqsubseteq_f$ defined over $\mathcal{M}$, an evidential distance d is said to be* ***$\sqsubset_f$-compatible (in the strict sense)*** *if for any mass functions m_1, m_2 and m_3 such that $m_1 \sqsubset_f m_2 \sqsubset_f m_3$, we have:*

$$\max\{d(m_1, m_2); d(m_2, m_3)\} < d(m_1, m_3), \qquad (2)$$

[1] There are others, but due to limited space, we will only deal with these ones.

For some family of set-functions f that are in bijective correspondence with mass functions, an interesting distance $d_{f,k}$ is defined as

$$d_{f,k}(m_1, m_2) = \left(\sum_{A \subseteq \Omega} |f_1(A) - f_2(A)|^k \right)^{\frac{1}{k}}.$$

In particular, we showed that for any $k \in \mathbb{N}^* \setminus \{\infty\}$, $d_{pl,k}$ are $\sqsubset_{pl}$-compatible and $d_{q,k}$ is $\sqsubset_q$-compatible (in the strict sense for all of them).

3 A distance-based cautious conjunctive aggregation

In this section, we introduce the main idea of our new combination operator, relying on distances compatible with the partial orders comparing informative content.

3.1 Conjunctive combination using partial orders

In this paper, rather than seeing a conjunctive combination $\mathcal{E}_1, \ldots, \mathcal{E}_\ell$ as a particular operator defined either on the mass functions $m_1, \ldots, m_\ell$ or on the weight functions $w_1, \ldots, w_\ell$, we simply consider that a mass function m^* resulting from a conjunction should be (1) more informative (in the sense of some partial order $\sqsubseteq_f$) than any $m_1, \ldots, m_\ell$ and (2) should be among the least committed elements (in terms of information) among those, in accordance with the LCP. Formally speaking, if we denote by

$$\mathcal{S}_f(m_i) := \{m \in \mathcal{M} \mid m \sqsubseteq_f m_i\} \tag{3}$$

the set of all mass functions more informative than m_i, then m^* should be such that:

1. $m^* \in \mathcal{S}_f(m_1) \cap \ldots \cap \mathcal{S}_f(m_\ell)$,
2. $\nexists m' \in \mathcal{S}_f(m_1) \cap \ldots \cap \mathcal{S}_f(m_\ell)$ such that $m^* \sqsubset_f m'$.

The first constraint expresses the conjunctive behavior of such an approach. The second constraint says that m^* is a maximal element (*i.e.* a least committed solution) for admissible solutions subject to the first constraint.
The interest of such a solution is that it is rather generic, and does not require any explicit model of dependence. However, it should be noted that the choice of the partial order to consider is not without consequence. Considering those mentioned in section 2.2, Equation (1) tells us that for a same mass function m, $\mathcal{S}_w(m) \subseteq \mathcal{S}_{pl}(m)$, hence the space of solutions will be potentially much smaller when choosing $\sqsubseteq_w$ rather than $\sqsubseteq_{pl}$. In practice and in accordance with the LCP, it seems safer to choose the most conservative partial orders, *i.e.*, in our case $\sqsubseteq_{pl}$ or $\sqsubseteq_q$. We will see in Section 4 that it can have an important impact on the combination results, even for simple examples.

While our definition of the cautious result of a conjunctive combination appears natural, it still faces the problem that many different solutions m^* could actually fit the two constraints, as $\sqsubseteq$ is a partial order. One idea to solve this problem is to use distances that are compatible with $\sqsubseteq$.

3.2 New conjunctive operators from soft LCP

To derive new conjunctive operators, we consider a weaker form of least commitment principle which we call **soft LCP**. This principle states that when there are several candidate mass functions compliant with a set of constraints, the one with minimal distance value from the vacuous mass function should be chosen for some $\sqsubseteq$-compatible distance. According to corollary 4 in [9], we know that the problem induced by the soft LCP is a convex optimization problem with a unique solution if the chosen distance $d_{f,k}$ is $\sqsubset_f$-compatible and if $2 \leq k < \infty$. Let $\star_{k,f}$ denote this operator, for any set of ℓ functions $\{m_1, .., m_\ell\}$, we have

$$m_1 \star_{f,k} .. \star_{f,k} m_l = \underset{m \in \mathcal{S}_f(m_1) \cap .. \cap \mathcal{S}_f(m_\ell)}{\arg\min} d_{f,k}(m, m_\Omega). \qquad (4)$$

The commutativity of the set-intersection and the symmetry property of distance give that $\star_{f,k}$ is commutative. Each operator $\star_{f,k}$ is also idempotent: for any possible solution $m \in \mathcal{S}_f(m_1) \setminus \{m_1\}$, we have $d_{f,k}(m_1, m_\Omega) < d_{f,k}(m, m_\Omega)$ because $d_{f,k}$ is $\sqsubseteq_f$-compatible and $m \sqsubseteq_f m_1 \sqsubseteq_f m_\Omega$, hence $m_1 \star_{f,k} m_1 = m_1$. Each of these operators are also conjunctive by construction, in the sense that the output mass function is more informative than any of the initial mass functions. Indeed if m_i states that ω is not a possible value of the unknown quantity ($pl_i(\omega) = 0$), then any function in $\mathcal{S}(m_i)$ also states so. Since the combination result belongs to $\mathcal{S}(m_i)$, then this piece of information is propagated by $\star_{f,k}$.

This operator is, however, not associative because we can have

$$\mathcal{S}_f(m_1 \star_{f,k} m_2) \subsetneq \mathcal{S}_f(m_1) \cap \mathcal{S}_f(m_2).$$

Consequently, the optimization constraints are not deducible from an output mass function $m_1 \star_{f,k} m_2$. Fortunately, these constraints can be stored and updated iteratively, meaning that the complexity of the combination does not increase with ℓ. In practice, one needs to be able to compute combinations iteratively without storing the whole set of mass functions $\{m_1, \ldots, m_\ell\}$ and restart the combination from scratch when a new function $m_{\ell+1}$ arrives. This property is often referred to as **quasi-associativity**. Let c denote a set function from 2^Ω to $[0;1]$ which is meant to store the problem constraints. Algorithm 1 allows to compute combinations using $\star_{q,k}$ sequentially. The same algorithm works for $\star_{pl,k}$. In practice, what we simply do is storing, for each set A, the lowest commonality (resp. plausibility) value encountered in $\{m_1, \ldots, m_\ell\}$.

4 A brief comparison with related works

As said earlier, there are many works that have dealt with the problem of either cautious conjunctive rules or of conjunctive rules not relying on independence.

Algorithm 1 Sequential combination using $\star_{q,k}$

entries : $\{m_1, .., m_\ell\}$, $k \geq 2$.
$c \leftarrow \min\{q_1; q_2\}$ (entrywise minimum).
$m \leftarrow m_1 \star_{q,k} m_2$.
for i from 3 to ℓ **do**
  $c \leftarrow \min\{c; q_i\}$ (entrywise minimum).
  $m \leftarrow \arg\min_{m'} d_{q,k}(m', m_\Omega)$ subject to $q' \leq c$.
end for
return m.

They depart from the classical conjunctive rule $\textcircled{\cap}$ that assume independence of the sources, and whose formula for a pair (m_1, m_2) of mass functions is

$$m_{1\textcircled{\cap}2}(A) = \sum_{\substack{A_1, A_2 \in 2^{\mathcal{X}} \\ \text{s.t. } A_1 \cap A_2 = A}} m_1(A_1)\, m_2(A_2)\,, \text{ for all } A \subseteq \Omega. \quad (5)$$

Dempster's rule $\oplus$ corresponds to the normalized version of this rule where the mass of the empty set is forced to zero. Choosing an alternative to them is however not so easy. A principled and common approach is to rely on a set of axiomatic properties [5] or to adapt existing rules from other frameworks [4]. In practice, such axioms seldom lead to a unique solution, and it is then necessary to advocate more practical solutions. Our rule can be seen as an instance of such an approach, where the axiom consists in using the LCP over sets of f-included mass functions, and the practical solution is to use a distance compliant with such an axiom. Cattaneo's solution [1] as well as Denoeux [3] cautious rules can also be seen as instances of the same principle. The former defends the fact of reducing the conflict rather than minimizing the informative content, while the latter focuses on using the set $\mathcal{S}_w(m_1) \cap \ldots \cap \mathcal{S}_w(m_\ell)$ and the order $\sqsubseteq_w$, and demonstrates that in this case there is a unique LCP solution known in closed form. This cautious rule is usually denoted by $\textcircled{\wedge}$. Due to lack of space, we will focus on comparing our approach with the most well-known, that is with rules $\textcircled{\cap}$, $\oplus$ and $\textcircled{\wedge}$.

Table 1 summarizes some basic theoretical properties satisfied by operators $\textcircled{\cap}$, $\oplus$, $\textcircled{\wedge}$ and $\star_{f,k}$. From a practical point of view, let us stress that combinations

Table 1. Basic properties of operators $\textcircled{\cap}$, $\oplus$, $\textcircled{\wedge}$ and $\star_{f,k}$.

operator	condition for use	commutativity	associativity	idempotence
$\textcircled{\cap}$	none	yes	yes	no
$\oplus$	$m_{1\textcircled{\cap}2}(\emptyset) < 1$	yes	yes	no
$\textcircled{\wedge}$	$m_1(\Omega) > 0$ and $m_2(\Omega) > 0$	yes	yes	yes
$\star_{f,k}$	none	yes	quasi	yes

using $\star_{f,k}$ for $f \in \{pl, q\}$ and $k = 2$ are really easy to compute. Indeed, quadratic programming techniques can solve equation (4) in a very few iterations. The function $m_\emptyset$ can be used to initialize the minimization as we are sure that it belongs to $\mathcal{S}_f(m_1) \cap .. \cap \mathcal{S}_f(m_\ell)$.

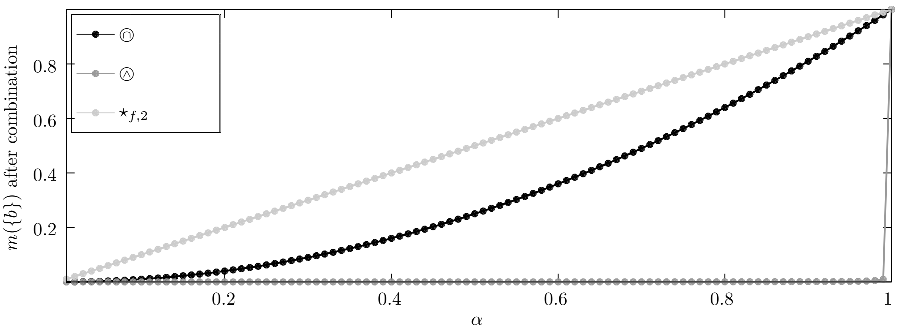

Fig. 1. Mass assigned to $\{b\}$ after combination of $m_1 = \alpha m_{\{b\}} + (1-\alpha)\, m_{\{a\}}$ and $m_2 = \alpha m_{\{b\}} + (1-\alpha) m_{\{c\}}$ with ⓐ, ⓦ and $\star_{f,2}$.

Let us illustrate the operator discrepancies on a simple situation inspired from Zadeh's counter-example [11]. Suppose $m_1 = \alpha m_{\{b\}} + (1-\alpha)\, m_{\{a\}}$ and $m_2 = \alpha m_{\{b\}} + (1-\alpha)\, m_{\{c\}}$ are two mass functions on a frame $\Omega = \{a, b, c\}$. Figure 1 shows the mass assigned to $\{b\}$ after combination by $\textcircled{\cap}$, $\textcircled{\wedge}$ and $\star_{f,2}$. The same masses are obtained for $f \in \{pl, q\}$. A very small mass was assigned to Ω when using $\textcircled{\wedge}$ to circumvent the non-dogmatic constraint.
As could be expected, our rule tries to maintain as much evidence on $\{b\}$ as possible. A striking fact is that we have obviously $m_1 \star_{f,2} m_2 (\{b\}) = \alpha$. More precisely, we have $m_1 \star_{f,2} m_2 = (1-\alpha)\, m_\emptyset + \alpha m_{\{b\}}$.
This result can be proved for any finite $k \geq 2$ when $f = q$. Let $q_{1\wedge 2}$ denote the entrywise minimum of functions q_1 and q_2. In this particular setting, $q_{1\wedge 2}$ happens to be a valid commonality function. Consequently, $m_{1\wedge 2} \in \mathcal{S}_q(m_1) \cap \mathcal{S}_q(m_2)$. By definition of the partial order $\sqsubseteq_q$, for any function $m \in \mathcal{S}_q(m_1) \cap \mathcal{S}_q(m_2)$, we have $m \sqsubseteq_q m_{1\wedge 2}$. Since we also have $m_{1\wedge 2} \sqsubseteq_q m_\Omega$ and $d_{q,k}$ is $\sqsubseteq_q$-compatible, then $m_1 \star_{q,k} m_2 = m_{1\wedge 2}$. In other words, our approach coincides with the minimum rule applied to commonalities in this case.
When $f = pl$, the result can also be proved. For any $m \in \mathcal{S}_{pl}(m_1) \cap \mathcal{S}_{pl}(m_2)$, the constraints $pl(\{a\}) = pl(\{c\}) = 0$ imply that only $\{b\}$ and $\emptyset$ are possible focal sets for m. More precisely, this actually implies that $\mathcal{S}_{pl}(m_1) \cap \mathcal{S}_{pl}(m_2)$ is a segment $(1-\beta)\, m_\emptyset + \beta m_{\{b\}}$ in $\mathcal{M}$ parametrized by $\beta \in [0; \alpha]$. $\sqsubseteq_{pl}$ is a total order for this segment and obviously $m_1 \star_{pl,k} m_2 = (1-\alpha)\, m_\emptyset + \alpha m_{\{b\}}$.
In this example, the behavior of Denœux's cautious rule $\textcircled{\wedge}$ is more questionable, as it keeps almost no mass on $\{b\}$ except when $\alpha = 1$. This is an unconservative behavior, due partly to the fact that $\mathcal{S}_w$ induces stronger constraints than $\mathcal{S}_{pl}$ or $\mathcal{S}_q$. Finally, the conjunctive rule appears to have an intermediate behavior as compared to the two others.

5 Conclusion

This paper introduces an idea allowing cautious conjunctive combinations of mass functions by relying on constraints inducing a more informative mass function than the combined ones on one hand, and on the minimization of distances to total ignorance on the other hand. The metrics used in the minimization procedure must be compatible with partial orders comparing informative contents. This idea can generate several commutative, idempotent and quasi-associative combination operators that are in line with the LCP principle. This procedure allows these operators to be easily interpretable and to rely on sound justifications. Preliminary experimental results show that they have very regular behavior as compared to standard approaches, and comply with some user's expectations.

This study is a start, but the interesting results we obtained call for several possible extensions, for instance by adapting the approach to other combination types (starting with disjunction), and by fully investigating its connection with other rules trying to solve the same problem. Moreover, it would be also interesting to check if our distance-based approach is to some extent compliant with other operations such as conditioning or refining. Finally, these new operators rely on L_k norms ($k \geq 2$) and the influence of parameter k must be studied.

References

1. Cattaneo, M.E.G.V.: Combining belief functions issued from dependent sources. In: Bernad, J., Seidenfeld, T., (Eds.), M.Z. (eds.) Third International Symposium on Imprecise Probabilities and Their Applications (SIPTA'03), pp. 133–147. Carleton Scientific, Lugano (Switzerland) (2003)
2. Cattaneo, M.E.: Belief functions combination without the assumption of independence of the information sources. International Journal of Approximate Reasoning 52(3), 299 – 315 (2011), dependence Issues in Knowledge-Based Systems
3. Denœux, T.: Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial Intelligence 172, 234 – 264 (2008)
4. Destercke, S., Dubois, D.: Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. Information Sciences 181(18), 3925–3945 (2011)
5. Dubois, D., Liu, W., Ma, J., Prade, H.: The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. Information Fusion 32, 12–39 (2016)
6. Dubois, D., Prade, H.: Consonant approximations of belief functions. International Journal of Approximate Reasoning 4(56), 419 – 449 (1990)
7. Dubois, D., Prade, H., Smets, P.: A definition of subjective possibility. International Journal of Approximate Reasoning 48(2), 352 – 364 (2008), in Memory of Philippe Smets (19382005)
8. Dubois, D., Yager, R.R.: Fuzzy set connectives as combinations of belief structures. Information Sciences 66(3), 245–276 (1992)
9. Klein, J., Destercke, S., Colot, O.: Interpreting evidential distances by connecting them to partial orders: Application to belief function approximation. International Journal of Approximate Reasoning 71, 15 – 33 (2016)

10. Smets, P.: The canonical decomposition of a weighted belief. 14th international joint conference on Artificial intelligence 2, 1896–1901 (1995)
11. Zadeh, L.: A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. the Artificial Intelligence Magazine 7, 85–90 (1986)