Aggregating Non-Independent Dempster-Shafer Belief Structures

Ronald R. Yager
Machine Intelligence Institute,
Iona College
New Rochelle, NY 10801
yager@panix.com

Abstract

We suggest an approach to the aggregation of non-independent belief structures that makes use of a weighted aggregation of the belief structures where the weights are related to the degree of dependence. It is shown that this aggregation is non-commutative, the fused value depends on the sequencing of the evidences. We then consider the problem of how best to sequence the evidence. We investigate using the measure of information content of the fused value as a method for selecting the appropriate way to sequence the belief structures.

Keywords: Multi-source data fusion, Dempster-Shafer theory, aggregation, information measures, non-independence.

1. Introduction

The Dempster-Shafer theory of evidence [1] is an important tool in granular computing and particularly useful in the task of multi-source information fusion. Central to its application in information fusion is the use of Dempster's rule for combining belief structures. Implicit in the use of Dempster's rule is the assumption that the belief structures are independent. In many cases this assumption does not necessarily hold. Our objective here is to look at the problem of applying Dempster's rule in the case where there might be some non-independence between the pieces of evidence.

2. Aggregation of Non-Independent Belief Structures

The basic Dempster rule for aggregating belief structures assumes independence between the belief structures being aggregated. If m_1 and m_2 are two belief structures when we calculate $m = m_1 \oplus m_2$ we are assuming that m_1 and m_2 are independent. The concept of independence used here is somewhat a vague idea. Intuitively what is meant by independence is that idea that the pieces of evidence m_1 and m_2 have been determined in some sense by different means. A clear example of independence is when two people separately observe an individual and then each provide an estimate of the person's age. Independence is less clear if the two people are in the same room with each other while observing the person. Here the possibility of one affecting the other exists.

One reason for the concern for independence is the lack of idempotency of the Dempster's rule, $m \oplus m \neq m$. To get some intuition consider a piece of evidence m_1 which is a simple support function $m_1(\{x_1\}) = \alpha$ and $m_1(X) = \overline{\alpha}$. Here the evidence is pointing to the value x_1 with support α. The value $\overline{\alpha}$ can be seen as our uncertainty in this belief. Consider an additional piece of evidence m_2 which formally is the same as m_1 that is $m_2(\{x_1\}) = \alpha$ and $m_2(X) = \overline{\alpha}$.

Combining these two pieces of evidence using Dempster's rule gives $m = m_1 \oplus m_2$ where $m(\{x_1\}) = \alpha^2 + 2\alpha(1-\alpha) = \alpha(2-\alpha)$ and $m(X) = (1-\alpha)^2$. Since $2 - \alpha > 1$ then $\alpha(2-\alpha) > \alpha$ by combining these two pieces of evidence we have increased the support for x_1. If the two pieces of evidence have been supplied by two independent sources then the combining of these to get more support for x_1 appears to be reasonable. On the other hand assume that the evidence m_2 is supplied by the same person who supplied m_1 then combining of these to obtain more support for x_1 would appear to be inappropriate. Here we have a case of complete non-independence.

In the following we shall suggest a formal extension of Dempster's rule to allow for the aggregation of non-independent evidence. More generally we will allow for a degree of

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 289–297
Torremolinos (Málaga), June 22–27, 2008

independence $\lambda \in [0, 1]$ where $\lambda = 1$ is complete independence and $\lambda = 0$ is complete dependence. We shall see that the introduction of non-independent evidence will result in a situation in which the aggregation is no longer commutative. That is the order in which we aggregate the evidence will affect the results. Before proceeding we introduce a weighting operator on belief structures that we will use in the aggregation of non-independent belief structures.

Let m be a belief structure on X with focal elements B_i, i = 1 to q and $m(B_i)$ as the associated weights. Let *a* be a value in the unit interval we define $a \otimes m$ as a belief structure m on X that has as its focal elements all the focal elements of m, B_i = 1 to q, plus X as a focal element X. The associated weights are as follows:

$$m(B_i) = a\, m(B_i) \qquad i = 1 \text{ to } q$$

$$m(X) = (1 - a)$$

Note: If X is one of the focal elements of m, $B_q = X$ then

$$m(B_i) = a\, m(B_i) \qquad i = 1 \text{ to } q - 1$$

$$m(B_q) = a\, m(B_q) + (1 - a).$$

We shall refer to the operation $a \otimes m$ as significance weighting. We note that it is closely related to what Shafer [2] called discounting.

We now describe our approach for the case of two pieces of evidence. Assume we have a piece of evidence m_1 and we now get an additional piece of evidence m_2. Let λ indicate the analyst's perception of the degree to which m_2 is independent at m_1. Using this degree of independence we get that our combined belief structure is

$$m = m_1 \oplus (\lambda \otimes m_2).$$

If $\lambda = 1$ we get that $1 \otimes m_2 = m_2$ that $m = m_1 \oplus m_2$ which is the usual result for Dempster's rule.

On the other hand if $\lambda = 0$ then $0 \otimes m_2 = m_v$ the vacuous belief structure, it has one focal element X. In this case $m = m_1 \oplus m_v = m_1$. Here we get the aggregated value to be m_1.

In the preceding we have implicitly assumed a sequencing of the pieces of evidence. Essentially we have assumed m_1 first and then m_2. While the value of λ is indifferent to the sequencing, the actual result is dependent on the sequencing. If we consider m_2 first in the sequence then our aggregate will be $m = m_2 \oplus (\lambda \otimes m_1)$. If $\lambda = 1$ then the sequencing doesn't matter, we get the same result. At the other extreme if $\lambda = 0$ in the case where m_2 is first in the sequence we get

$$m = m_2 \oplus (0 \otimes m_1) = m_2 \oplus m_v = m_2.$$

Thus we see that the sequencing of the evidence matters. Here then we see the essential non–commutativity of the aggregation in the face of non-independence.

Before we address the issue of determining the sequencing let us introduce a general framework for the aggregation of possibly non-independent belief structures.

Assume $m_1, m_2, \ldots, m_q$ are the collection of belief structures. Let *Seq* be a sequencing of the belief structures. Let Seq(j) be the index of the j^{th} belief structure in the sequencing. Thus our sequence is

$$m_{seq(1)} \to m_{seq(2)} \to \cdots\cdots \to m_{seq(q)}.$$

We formulate the sequential aggregation of these belief structures as

$$m = m_{seq(1)} \oplus (\delta_{seq(2)} \otimes m_{seq(2)}) \oplus \ldots \oplus (\delta_{seq(q)} \otimes m_{seq(q)})$$

$$m = m_{seq(1)} \oplus \bigoplus_{j=2}^{q} (\delta_{seq(j)} \otimes m_{seq(j)}).$$

Here $\delta_{seq(j)}$ is the degree of independence of the evidence $m_{seq(j)}$ from the previous aggregated values. That is $\delta_{seq(k)}$ is the degree of independence of $m_{seq(k)}$ from

$$m_{seq(1)} \oplus \bigoplus_{j=2}^{k-1} (\delta_{seq(j)} \otimes m_{seq(j)}).$$

If by convention we denote $\delta_{seq(1)} = 1$ then we can succinctly express this aggregation as

$$m = \bigoplus_{j=1}^{q} (\delta_{seq(j)} \otimes m_{seq(j)}).$$

We note here that if all the pieces of evidence are independent, for all j $\delta_{seq(j)} = 1$, then we have $m = m_{seq(1)} \oplus m_{seq(2)} + \ldots \oplus m_{seq(q)}$, this is Dempster's rule.

As already noted once when move away from the implicit assumption of independence and allow considerations of dependence we encounter non-commutativity. This introduces considerable complexity. A by-product of this allowance for non-independence will be the requirement that the analyst makes some subjective choices affecting the aggregation process.

In this environment two questions naturally arise. The first is how do we estimate the degree of independence, the values for $\delta_{seq(j)}$ used in the preceding formula. The second question is how do we sequence the evidence in the aggregation.

The first question will require a much deeper discussion of what is independence than we are prepared to undertake at this time. The calculation of degree of independence will clearly be context dependent. It would also appear that non-independent evidence should manifest some degree of similarity.

We can make one formal contribution to the problem of determining the values of $\delta_{seq(j)}$. This is the situation where the calculation of $\delta_{seq(j)}$ is what we shall call decomposable.

Assume $\boldsymbol{\lambda}$ is a $q \times q$ matrix whose components λ_{ij} indicate the degree of independence between m_i and m_j. For this matrix we assume $\lambda_{ii} = 0$ and $\lambda_{ik} = \lambda_{ki}$, it is symmetric.

If we have such a matrix and assume decomposability we can very effectively obtain the values of $\delta_{seq(j)}$. In particular we can obtain for $j > 1$

$$\delta_{seq(j)} = \underset{k=1}{\overset{j-1}{\mathrm{Min}}}[\bar{\delta}_{seq(k)} \vee \lambda_{seq(k)seq(j)}].$$

Here $\bar{\delta}_{seq(k)} = 1 - \delta_{seq(k)}$ and $\vee$ is the max operator. We of course assume $\delta_{seq(1)} = 1$.

In the special case where $\lambda_{ij} = 1$ for all $j \neq i$ then we see that $\delta_{seq(j)} = 1$ for all j. In this case we have essentially assumed independence and we get Dempster's rule.

At the other extreme is the case where $\lambda_{ij} = 0$ for all pairs. In this case since $\delta_{seq(1)} = 1$ we have $\delta_{seq(j)} = 0$ for $j \neq 1$. This is the case of complete dependence. Here we will get as the aggregated value $m = m_{seq(1)}$.

3. On the Issue of Sequencing

We now turn to the issue of deciding on the sequencing of the evidence to be used in our aggregation. We note that if we have q pieces of evidence there are q! different ways to sequence the pieces of evidence. Each of these sequencing can lead to a different aggregated value.

At the most fundamental level the task of deciding the sequencing is going to involve some subjective choices by the agent who is ultimately responsible for the result of the fusion. As there is no absolute predetermined rule for deciding how to sequence the evidence, the choice of how to sequence the evidence must be made by the responsible agent in consultation with their information analyst. In the following we look at some features that can be used as a basis to decide on a sequence.

Clearly any features distinguishing the evidences to be aggregated may be useful. Temporal differences between the pieces of evidence may be useful; this may be particularly useful in a dynamic environment where things are changing. Here we may put the more recent evidence earlier in the sequence. Another feature that may be useful is some distinction between the credibility of the sources. Here we may sequence the evidence by perceived credibility. Here the more credible the earlier in the sequence.

Here I distinguish between what I call external and internal features of a piece of evidence. By internal features of a piece of evidence I mean properties associated with the actual evidence itself, essentially the function m and the associated focal elements. By external features I properties related to who supplied the evidence, when it was supplied, the credibility of the supplier. Possible synonyms for external and internal features could be pedigree and content of the evidence. In situations in which we have no external features distinguishing the evidence or we don't believe the ones we have are useful we must turn to internal properties of the evidence to help decide on the sequencing.

One possible way to sequence the evidence is such that the resulting aggregation provides the most information, the greatest certainty regarding the value of the variable. That is using this approach we fuse the information using all the possible q! sequences and then we select the sequencing that leads to the fused value providing the most information. Central to this type of approach is the ability to compare the information content of the belief structures that result from each sequence. We now turn to this issue.

4. Information Content for Comparing Sequencing

We start the task of developing tools for comparing the information contained in belief structures by looking at the case of the aggregation of two pieces of evidence.

Consider the case where m_1 and m_2 are simple support functions focused on the same value y and their degree of independence is λ. Here $m_1(\{y\}) = a$ and $m_1(X) = 1 - a$ while $m_2(\{y\}) = b$ and $m_2(X) = 1 - b$.
Here we assume a > b. If we sequence them as $m_1 \rightarrow m_2$ then we get as our aggregated value

$$m_{1/2} = m_1 \oplus (\lambda \otimes m_2) = m_1 \oplus m_2$$

where $m_2(\{y\}) = \lambda b$ and $m_2(X) = 1 - \lambda b$. Combining m_1 with m_2 we get

$$m_{1/2}(\{y\}) = 1 - (1 - a)(1 - \lambda b)$$
$$m_{1/2}(X) = (1 - a)(1 - \lambda b)$$

Thus $m_{1/2}(\{y\}) = a + \lambda b \bar{a}$. The bigger $m_{1/2}(\{y\})$ the more information, we are more certain about the value of the variable.

If we sequence them the other way, $m_2 \rightarrow m_1$, we get $m_{2/1}(\{y\}) = b + \lambda a \bar{b}$. Taking the difference we have

$$m_{1/2}(\{y\}) - m_{2/1}(\{y\}) = (a - b)(1 - \lambda).$$

We see that with a > b then $m_{1/2}(\{y\}) > m_{2/1}(\{y\})$. Here to get the most information we sequence them $m_1 \rightarrow m_2$, we first take the evidence with the largest support for y.

Consider another case of m_1 and m_2 still with λ degree of independence

$m_1(\{y\}) = a$	$m_2(\{y, z\}) = a$
$m_1(X) = 1 - a$	$m_2(X) = 1 - a$

Here it is clear that m_1 has more information. Combining these in the sequence $m_1 \rightarrow m_2$ we have $m_{1/2} = m_1 \oplus (\lambda \otimes m_2)$ and we get

$$m_{1/2}(\{y\}) = a \lambda b + a (1 - \lambda b) = a$$
$$m_{1/2}(\{y, z\}) = a (\bar{a} \lambda)$$
$$m_{1/2}(X) = (1 - a)(1 - \lambda a).$$

Combining them in the sequence $m_2 \rightarrow m_1$ we have $m_{2/1} = m_2 \oplus (\lambda \otimes m_1)$ this gives us

$$m_{2/1}(\{y\}) = \lambda a^2 + (\lambda a)(1 - a) = \lambda a$$
$$m_{2/1}(\{y, z\}) = a (1 - \lambda a)$$
$$m_{2/1}(X) = (1 - a)(1 - \lambda a).$$

Since $m_{2/1}(X) = m_{1/2}(X)$ the difference between $m_{1/2}$ and $m_{2/1}$ is determined by the weights on {y} and {y, z}. Since $m_{1/2}(\{y\}) \geq m_{2/1}(\{y\})$ then we see that $m_{1/2}$ is more informative.

Consider another case again with degree of independence λ where

$m_1(\{y, z\}) = a$	$m_2(\{y\}) = 0.5\ a$
$m_1(X) = \bar{a}$	$m_2(\{z\}) = 0.5\ a$
	$m_2(X) = \bar{a}$

While in this case the determination as to which of m_1 and m_2 is more information is not as obvious as in the preceding we can, however, see that m_2 is more informative as with this piece of evidence the source is sure as how the weight of *a* is divided between y and z.

For this example we first calculate $m_{1/2} = m_1 \oplus (\lambda \otimes m_2)$. In this case

$$m_{1/2}(\{y\}) = 0.5\ a \lambda,\ m_{1/2}(\{z\}) = 0.5\ a \lambda$$
$$m_{1/2}(\{y, z\}) = a (1 - \lambda a),\ m_{1/2}(X) = \bar{a} (1 - \lambda a).$$

Calculating $m_{2/1} = m_2 \oplus (\lambda \otimes m_1)$ then we get:

$$m_{2/1}(\{y\}) = 0.5\ a,\ m_{2/1}(\{z\}) = 0.5\ a$$
$$m_{2/1}(\{y, z\}) = \lambda a \bar{a},\ m_{2/1}(X) = \bar{a} (1 - \lambda a).$$

Since $m_{2/1}(X) = m_{1/2}(X)$ the difference between the two will be determined by the other focal elements. Since $m_{2/1}(\{y\}) > m_{1/2}(\{y\})$ and $m_{2/1}(\{z\}) > m_{1/2}(\{z\})$ then we conclude that $m_{2/1}$ has more information.

Since there are q! ways we can sequence q pieces we must eventually compare these q! aggregations and decide on the best. In the special case when all the evidences are completely dependent then as we showed $m = m_{seq(1)}$. It is the value of the first element in the sequence. In this case there are only q possible first values. Furthermore using the information content of the fused value as our determining factor we would select as the fused value the piece of evidence with the most information. This seems reasonable.

5. Using Range Containment

In the preceding simple examples we were able to indicate which belief structures had more information, less uncertainty, for more complex belief structure we need more sophisticated tools to compare belief structure regarding their information content. Our objective here is to find some algorithm that enables us to determine whether the information content of belief structure m_1, $IC(m_1)$ is greater than the information content of belief structure m_2, $IC(m_2)$. In the following we begin this task.

One relationship [3] between two belief structures which definitely characterizes a situation in which we clearly know that $IC(m_1) > IC(m_2)$ is described in the following. Assume m_1 and m_2 are two belief structures on X. For any subset of A of X we define $Range_1(A) = [Bel_1(A), Pl_1(A)]$ and $Range_2(A) = [Bel_2(A), Pl_2(A)]$. We say that m_1 is more informative than m_2 if for all A we have $Range_1(A) \subseteq Range_2(A)$ and there exists at least one A such the $Range_1(A) \subset Range_2(A)$. We shall denote this as $m_1 \Rightarrow m_2$. If $Range_1(A) = Range_2(A)$ for all A we say m_1 and m_2 are equally informative and denote this as $m_1 \Leftrightarrow m_2$.

While the preceding definition correctly captures the idea of m_1 being more informative than m_2 it has two problems. One pragmatic problem associated with this definition is the amount of work needed to calculate Range(A) for all subsets of X. The second problem is more formal; the definition is not complete. That is there are belief structures for which neither $m_1 \Rightarrow m_2$ or $m_2 \Rightarrow m_1$ or $m_1 \Leftrightarrow m_2$ is true. Nevertheless, we shall use this as a starting point for the difficult problem of comparing the information content of Dempster-Shafer belief structures.

Yager [3] provided some tools that can help in the pragmatic issue of comparing belief structures by reducing in some cases the need to calculate Range(A) for all A.

Definition: Assume m_1 is a belief structure with focal elements A_i for i = 1 to p where $m(A_i) = a_i$. Let m_2 be another belief structure with focal element

$$B_{11}, B_{12}, ..., B_{1n_1}, B_{21}, ..., B_{2n_2}, ..., B_{pn_p}$$

where $m(B_{ij}) = b_{ij}$. Furthermore we assume that $A_i \subseteq B_{ij}$ for all j = 1 to n_i and $\sum_{j=1}^{n_i} m(B_{ij}) = a_i$ for i = 1 to p. In this case we say that m_1 entails m_2, we shall denote this as $m_1 \subseteq m_2$.

Yager [3] showed that if $m_1 \subseteq m_2$ then for all subsets A it is the case that

$$[Bel_1(A), Pl_1(A)] \subseteq [Bel_2(A), Pl_2(A)].$$

Thus we see that if $m_1 \subseteq m_2$ then $m_1 \Rightarrow m_2$. We note that $m_1 \not\subset m_2$ doesn't mean that m_1 is not more informative than m_2.

In the above situation we turn the problem of determining whether m_1 is more informative then m_2 into just comparing the focal elements rather than having to compare the range for all the subsets of A.

A somewhat simplified version of the above condition can be formulated. Assume m is a belief function of X with focal elements A_1, ... A_p with weights $m(A_i)$. We can **expand** m by replacing any of its focal elements A_j with two focal elements D_1 and D_2 such that $D_1 = A_j$ and $D_2 = A_j$ where $m(D_1) + m(D_2) = m(A_j)$.

Let m_1 and m_2 be two belief structures which can be expanded such that they have the same number of focal elements A_1, ..., A_q and B_1, ..., B_q such that $A_j \subseteq B_j$ and $m_1(A_j) = m_2(B_j)$ for all j = 1 to q. In this we can show that for any subset A it is the case that

$$[Bel_1(A), Pl_1(A)] \subseteq [Bel_2(A), Pl_2(A)].$$

6. Indifference to Indexing

Our paradigm of using range containment for comparing information content of belief structures can be further enhanced. In particular it can be made indifferent to the indexing. Consider two belief structures m_1 and m_2 defined on X where

$$m_1(\{x_1\}) = a \qquad m_2(\{x_2\}) = a$$
$$m_1(X) = 1 - a \qquad m_2(X) = 1 - a.$$

What should be clear is that these two belief structures are equally informative.

Perhaps a more intuitive manifestation is there would be the case where m_1 and m_2 are Bayesian. For example m_1 is such that $m_1(x_1) = 0.1$, $m_1(x_2) = 0.2$, $m_1(x_3) = 0.3$, $m_1(x_4) = 0.4$ and m_2 such that $m_2(x_4) = 0.1$, $m_2(x_3) = 0.2$, $m(x_2) = 0.3$, $m(x_1) = 0.4$. It is clear the uncertainty in both cases is the same.

Essentially we observe any procedure for comparing the information content of belief structures should be indifferent to the indexing. In the following we generalize this observation. First we introduce the idea of a **replica.**

Let R: X → X be a bijective mapping, it is one to one and onto. Essentially R re-indexes the elements in X, R is sometimes called a permutation. If A is a subset of X by R(A) we shall mean a subset of X in which the elements of A have been re-indexed according to R. Thus if $X = \{x_1, x_2, x_3, x_4\}$ and R is such that: $R(x_1) = x_3$, $R(x_2) = x_4$, $R(x_3) = x_1$ and $R(x_4) = x_2$ then if $A = \{x_1, x_2\}$ we have $R(A) = \{x_3, x_4\}$. We note that the cardinality of R(A) is always the same as A.

Let m be a belief structure on X with focal element B_1, B_2, ….., B_q and weights $m(B_j)$. Let R be a re-indexing function on X. By R(m) we shall mean a new belief structure $\tilde{m}$ with q focal elements $A_j = R(B_j)$ and where $\tilde{m}(A_j) = m(B_j)$. Thus here we have just re-indexed everything. We shall call $\tilde{m}$ a **replica** of m.

We note that a special replica of m is the identity, here $R(x_i) = x_i$.

In the following we use the idea to replica to provide a general characterization of information content of a belief structure.

Let **PIC** be some procedure or rule for comparing the information content of two belief structures such that its application to any two belief structures m_1 and m_2, **PIC**(m_1, m_2), returns one of four states:

a) **PIC**$(m_1, m_2) \Rightarrow IC(m_1) > IC(m_2)$ (m_1 is more informative)

b) **PIC**$(m_1, m_2) \Rightarrow IC(m_2) > IC(m_1)$ (m_2 is more informative)

c) **PIC**$(m_1, m_2) \Rightarrow IC(m_1) = IC(m_2)$ (m_1 and m_2 equally informative)

d) **PIC**$(m_1, m_2) \Rightarrow IC(m_1) <> IC(m_2)$ (m_1 and m_2 are incomparable)

A required property for **PIC** to be a valid procedure for comparing belief structures is that it be *replica indifferent* a property we define in the following. Let $\mathcal{R}$ be the set of all re-indexing procedure on X. Let (R_1, R_2) be any arbitrary pair of re-indexing procedures, R_1 and $R_2 \in \mathcal{R}$. Then replica indifference requires all **PIC**$(R_1(m_1), R_2(m_2))$ that do not evaluate to incomparable, < >, must evaluate to the same (>, <, or =). Thus if there exists a pair R_1 and R_2 such that **PIC**$(R_1(m_1), R_2(m_2))$ evaluates to $IC(R_2(m_2)) > IC(R_1(m_1))$ then for any other pair R_3 and R_4 it must be the case that **PIC**$(R_3(m_1), R_4(m_2))$ evaluates to $IC(R_4(m_2)) > IC(R_3(m_1))$ or $IC(R_4(m_2)) <> IC(R_3(m_1))$. Thus all pairs of replicas that are not incomparable evaluate to the same value in the set {>, <,=}.

The procedure suggested earlier for comparing the informativeness of m_1 and m_2 using the containment of ranges of subsets, $Range_1(A) \subseteq Range_2(A)$, is replica indifferent. Replica indifference enhances the usefulness of this procedure for comparing belief structures. In particular if m_1 and m_2 are two belief structures and there exists a replica of m_2, $m_3 = R(m_2)$ such that for all subset A we have $Range_1(A) \subseteq Range_3(A)$ then m_1 is more informative then m_2.

7. Protoforms for Determining IC Relationship

The combination of replica indifference with the entailment rule provides some very basic protoforms for determining the IC relationship. We first consider the comparison of common classes of belief structure.

Let m_1 be a belief structure so that $m_1(B) = 1$ where B is some subset of X of cardinality k. For simplicity let $B = \{x_1, x_2, ..., x_k\}$. Let m_2 be a Bayesian belief structure with k focal elements

$A_1 = \{x_1\}, A_2 = \{x_2\}, ..., A_k = \{x_k\}$

where $m_2(A_j) = p_j$. We note that for $x_j \notin B$ we have $p_i = 0$.

We now introduce an expanded version of m_1, $\tilde{m}_1$ with k focal elements $B_j = B$, j = 1 to k, where $\tilde{m}_1(B_j) = m_2(A_j) = p_j$. What we observe now is that the relationship between $\tilde{m}_1$ and m_2 is such that they both have k focal elements, A_1, ..., A_k and B_1, ..., B_k with the same weights but $A_j \subseteq B_j$ Thus using our previous result about this situation we can conclude that m_2 has more information that m_1, $IC(m_2) > IC(m_1)$.

Some interesting special cases of the following are worth noting. One special case is where $p_j = \frac{1}{k}$ for $x_j \in B$. Thus we see that a belief structure m_1 with $m_1(B) = 1$ is less informative then a Bayesian one in which $p_j = \frac{1}{k}$ for j = 1 to k. Consider the case where B =

X, m_1 is the vacuous belief structure, m_V. We have shown that the belief structure where $p_i = \frac{1}{n}$ for all $x_i \in X$ has more information. This makes sense, since in the vacuous case we have no information.

In the preceding we have shown that if m_1 has $m_1(B) = 1$ then m_2 with focal elements $A_j = \{x_j\}$ for $x_j \in B$ with $m_2(A_j) = p_j$ has more information than m_1. There is no reasoning why some of p_j can't be zero. From this we can conclude the following. Let m_1 be a belief structure with $m_1(B) = 1$. Let m_2 be a belief structure with $r \leq |B|$ singleton focal elements, $A_i = \{x_i\}$ and $x_i \in B$. Essentially m_2 is a kind of Bayesian structure. In this case m_2 must be more informative than m_1.

Our requirement that the information comparison measure should be indifferent to indexing, the use of replicas, allows us to state the following theorem:

Theorem: Assume m_1 is a belief structure with $m_1(B) = 1$. Let m_2 be a belief structure with q singleton focal elements if $q \leq |B|$ then $IC(m_2) > IC(m_1)$.

We emphasize here that it is not necessary for the elements in B to be the same as those in the focal elements of m_2.

Actually we can provide an even more general result using the idea replica.

Theorem: Let m_1 be a belief with $m(B) = 1$. Let m_2 be a belief structure with q focal elements, A_j for j = 1 to q and $m_2(A_j)$ the associated weights. Let $A = \bigcup_{i=1}^{q} A_i$, all the elements that appear in the focal sets of m_2. Then if $|A| \leq |B|$ we have $IC(m_2) > IC(m_1)$.

Proof: Using the idea of replicas we can determine a replica $\tilde{m}$ of m_1 using a mapping R such that for each $x_j \in A$ there exists a $x_i \in B$ such that $R(x_i) = x_j$. Once having this replica $\tilde{m}$ with $R(B) = \tilde{B}$ then since $A_i \subset \tilde{B}$ the result follows.

Another interesting special situation is the following. Assume m_1 and m_2 are simple support functions where

$$m_1(A) = \alpha \qquad m_2(B) = \beta$$
$$m_1(X) = (1 - \alpha) \qquad m_2(X) = (1 - \beta)$$

Assume $\alpha \geq \beta$ and $|A| \leq |B|$ then $IC(m_1) > IC(m_2)$.

We see this as follows. Assume $\tilde{m}_2$ is a replica of m_2 based on a mapping R such that for each $x_i \in A$ there exists a $x_j \in B$ for where $R(x_j) = x_i$. In this case $R(B) = \tilde{B} \supset A$. In this case we have $\tilde{m}_2(\tilde{B}) = \beta$ and $\tilde{m}_2(X) = (1 - \beta)$. We can now expand m_1 an $\tilde{m}_2$

$$m_1(A_1) = \beta \qquad \tilde{m}_2(B_1) = \beta$$
$$m_1(A_2) = \alpha - \beta \qquad \tilde{m}_2(B_2) = \alpha - \beta$$
$$m_1(X) = (1 - \alpha) \qquad \tilde{m}_2(X) = (1 - \alpha)$$

where $A_1 = A_2 = A$ and $B_1 = \tilde{B}$ and $B_2 = X$. It is clear in this case that $A_j \subseteq B_j$ and hence $IC(m_1) > IC(\tilde{m}_2)$ and hence $IC(m_1) > IC(m_2)$.

Another interesting special case is the following. Assume m_1 and m_2 are such that

$$m_1(A) = \alpha \qquad m_2(B) = \alpha$$
$$m_2(\overline{A}) = 1 - \alpha \qquad m(X) = 1 - \alpha$$

where $|A| \leq |B|$. In this case $IC(m_1) \geq IC(m_2)$.

8. Entropy and Specificity to Compare Belief Structures

While the use of the containment procedure for comparing the information contents of belief structures has been greatly extended it still can often result in incomparability between belief structures. A very basic example of this occurs when we have Bayesian belief structures as shown below. Consider the case where $X = \{x_1, x_2\}$ and m_1 and m_2 are as described below

$$m_1(\{x_1\}) = 0.9 \qquad m_2(\{x_1\}) = 0.5$$
$$m_1(\{x_2\}) = 0.1 \qquad m_2(\{x_2\}) = 0.5.$$

In this case $Range_1(\{x_1\}) = [0.9, 0.9]$ and $Range_1(\{x_2\}) = [0.1, 0.1]$ while $Range_2(\{x_1\}) = [0.5, 0.5]$ and $Range_2(\{x_2\}) = [0.5, 0.5]$. While these intervals are incomparable it is clear that m_1 has less uncertainty than m_2. Thus we need further tools and procedures to help with the comparison of the information content of belief structures.

Earlier we introduced two measures to calculate the uncertainty associated with a Dempster–Shafer belief structure [4]. We recall if m is a belief structure with focal elements B_1,

..., B_q then the specificity of m is defined as $Sp(m) = 1 - \frac{1}{n-1}\sum_{j=1}^{q} m(B_j)(|B_j|-1)$ and the entropy of m is defined as

$$H(m) = -\sum_{j=1}^{q} m(B_j)\ln[Pl(B_j)] .$$

We now explore using these to help formulate a procedure for comparing the information content of belief structures. We first observe that these measures are replica indifferent. Thus if $\tilde{m} = R(m)$ is a replica of m then $Sp(m) = Sp(\tilde{m})$ and $H(m) = H(\tilde{m})$. This feature makes these two measures desirable for building procedures for comparing information content.

We further observe that Sp(m) = [0, 1] where the larger Sp(m) is associated with more information, less uncertainty. For H(m) we have shown that $H(m) \in [0, \ln(n)]$, however, here the larger H(m) the less information, the more uncertainty. In the following we shall find it more convenient to use a normalized version of H(m) which is $G(m) = 1 - \frac{H(m)}{\ln(n)}$. In this case $G(m) \in [0, 1]$ and the larger G(m) the more information, the less uncertainty.

We now shall suggest an additional procedure for comparing the information content of belief structures based on the use of the two measures Sp(m) and G(m). We denote this PIC_2 and described in the following

Assume m_1 and m_2 are two belief structures we shall say that $IC(m_1) > IC(m_2)$ if

$Sp(m_1) \geq Sp(m_2)$ and $G(m_1) \geq G(m_2)$ and at least one of the $\geq$ is an $>$.

Let us look at this procedure for comparing information content.

Consider the case where m_1 and m_2 are Bayesian belief structures. As we have previously show for any Bayesian belief structure the specificity is one, thus $Sp(m_1) = Sp(m_2) = 1$. Thus the only difference between Bayesian belief structures is their G(m) value. Since the larger G(m) the smaller the entropy this appropriately distinguishes between Bayesian belief structures based on the comparison of the entropy. The smaller the entropy the larger the information conflict, thus if $G(m_1) > G(m_2)$ then $IC(m_1) > IC(m_2)$.

Consider now the belief structure where m(B) = 1. In this case G(m) = 1 independent of B while $Sp(m) = 1 - .\frac{|B|-1}{n-1} = \frac{n-|B|}{n-1}$ Thus the smaller B the larger the specificity. Here again the monotonicity will allow us to point to the belief structure with the smaller B as the more informative.

We now consider a situation we introduced earlier in which m_1 and m_2 are two belief structures with the same number of focal sets, A_1, ..., A_q and B_1, ..., and B_q where $m_1(A_j) = m_2(B_j)$ and $A_j \subseteq B_j$ for all j and $A_j \subset B_j$ for at least one j. In the preceding we showed that here we conclude that m_1 is more informative than m_2. Let us look to see if our approach using Sp(m) and G(m) can capture this. First we note that

$$Sp(m_1) = \sum_{j=1}^{q} m(A_j)Sp(A_j) > \sum_{j=1}^{q} m(B_j)Sp(B_j)$$

$Sp(m_1) > Sp(m_2)$.

Thus, m_1 is more specific. Consider the calculation of H(m). Here we have

$$H(m_1) = -\sum_{j=1}^{q} m(A_j) \ln(Pl(A_j))$$

$$H(m_1) = \sum_{j=1}^{q} m(A_j) \ln(1 - \sum_{\substack{i \\ A_i \cap A_j = \varnothing}} m(A_i))$$

$$H(m_2) = -\sum_{j=1}^{q} m(B_j) \ln(Pl(B_j))$$

$$H(m_2) = \sum_{j=1}^{q} m(B_j) \ln(1 - \sum_{\substack{i \\ B_i \cap B_j = \varnothing}} m(B_i)) .$$

Since $m(B_j) = m(A_j)$ the difference will depend only on the terms $\sum_{\substack{i \\ A_i \cap A_j = \varnothing}} m(A_i)$ and $\sum_{\substack{i \\ B_i \cap B_j = \varnothing}} m(B_i)$. Since $A_i \subseteq B_i$ for all i then it follows $H(m_1) \leq H(m_2)$ and hence $G(m_1) \geq G(m_2)$. Thus here since $Sp(m_1) > Sp(m_2)$ and $G(m_1) \geq G(m_2)$ we see that m_1 is more informative than m_2.

9. On the Use of Information Content

In the preceding what became clear was that when we allow for the aggregation of non–independent belief structures the aggregation is non-commutative. The resulting fused value depends on the sequencing (ordering) of the beliefs structures in the aggregation. We suggested one way to address this problem is to use the sequence that results in a fused value that provides the most information, least uncertainty.

In order to implement this approach we needed an algorithm to be able to compare two belief structures m_1 and m_2 and tell which is more informative. In the preceding we suggested two algorithms. The first PIC_1 was based on containment of ranges. The second PIC_2 is based on based on comparing the specificity and entropy of belief structures.

If neither of these two methods allows us to choose one of the belief structures, they are incomparable under PIC_1 and PIC_2, we may try to use a less absolute method for comparing the information in belief structures. Essentially in this case we may only try to find out if $IC(m_1)$ is "better than" $IC(m_2)$. Here the concept "better then" is somewhat subjective. At this point we shall not pursue this only to make some comments. One possibility here is obtain some monotonic function $F(Sp(m), G(m))$ and say m_1 is better than m_2, if $F(Sp(m_1), G(m_1) > F(Sp(m_2), G(m_2))$. We note that Klir [5] and Abellan and Moral [6, 7] investigated functions of this nature..

Our focus in determining the sequencing of non-independent belief structures has been to use the sequence that results in the most informative fused value. In following this approach we are faced with some pragmatic problems. In addition to the possibility of incomparability discussed above this approach can at times involve considerable computational complexity. In particular q pieces evidence require the investigation of q! sequences, a task that can lead to a lot of work. Another option, which avoids this complexity, is to sequence the belief structures directly in terms of their information contents. Thus if we have q belief structures we would use as our sequencing Seq such that Seq(j) is the belief structure with if j^{th} largest amount of information. Here we still have the possibility of incomparability. This problem, however, may somewhat be reduced by the fact that the atomic belief structures may be simple and hence easy to compare.

10. Conclusion

We reviewed some aspects of the Dempster-Shafer theory of evidence. We suggested an approach to the aggregation of non-independent belief structures. This approach made use of a weighted aggregation of the belief structures where the weights are related to the degree of dependence. It was shown that this aggregation is non-commutative, the fused value depends on the sequencing of the evidences. We then considered the problem of how best to sequence the evidence. We investigated using the measure of information content of the fused value as a method for selecting the appropriate way to sequence the belief structures.

References

[1] Yager, R. R. and Liu, L., (A. P. Dempster and G.Shafer, Advisory Editors), (2008) *Classic Works of the Dempster-Shafer Theory of Belief Functions* (Springer: Heidelberg).

[2] Shafer, G., 1976, *A Mathematical Theory of Evidence* (Princeton University Press: Princeton, N.J.).

[3] Yager, R. R., 1986b, The entailment principle for Dempster-Shafer granules. I*nt. J. of Intelligent Systems* 1, pp. 247-262.

[4] Yager, R. R., 1983, Entropy and specificity in a mathematical theory of evidence. *Int. J. of General Systems* 9, pp. 249-260.

[5] Klir, G. J., 2006, *Uncertainty and Information*. (John Wiley & Sons: New York).

[6] Abellan, J. and Moral, S., 1999, Completing a total uncertainty measure in the Dempster-Shafer theory. *International Journal of General Systems* 28, pp. 299-314.

[7] Abellan, J. and Moral, S., 2003, Maximum entropy for credal sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 11, pp. 587-597.

.