The principle of minimum specificity as a basis for evidential reasoning*

Didier Dubois, Henri Prade

Université Paul Sabatier
Laboratoire Langages et Systèmes Informatiques
118, route de Narbonne
31062 TOULOUSE CEDEX

Extended abstract

Recently, the idea of measure of information stemming from Shannon theory has been enlarged in the framework of Shafer's evidence theory. Measures of imprecision, dissonance and others (see Klir [9], Dubois-Prade [7] for complementary surveys) have been attached to a body of evidence, viewed as an allocation of probability m to subsets of a given set Ω called a frame of discernment. Namely m is such that $m(A) \geq 0, \forall A \subset \Omega$ and

$$\sum_{\emptyset \neq A \subseteq \Omega} m(A) = 1 \qquad (1)$$

The pair (F,m) where $F = \{A \mid m(A) > 0\}$ is the set of focal elements, is called a body of evidence describing the possible location of a variable x ranging on Ω. m is called the basic (probability) assignment.

This way of describing uncertain information seems to be rather general since including standard probabilistic descriptions (F contains singletons) and those based on possibility theory [13, 3] (focal elements are nested).

----

* Prepared for presentation at the Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris, June 30 - July 4, 1986.

This paper investigates the possible role of measures of specificity in automated reasoning processes involving granular knowledge pervaded with uncertainty.

1 - Measures of imprecision and specificity

A particular case of a body of evidence is when only one focal element A is known for sure (m(A) = 1). It corresponds to the information granule "x is A", i.e. the value of x lies in A. A granule of information "x is A" is said to be precise if and only if $A = \{\omega\}$ for some $\omega \in \Omega$. Otherwise the granule is said to be imprecise. Precision is of course defined with respect to a frame Ω. The most imprecise granule is "x is Ω" which is a vacuous statement. Clearly, imprecision of A is reflected by its cardinality $|A|$ (supposedly finite), and more generally by $f(|A|)$ where f is an increasing mapping from $\mathbb{N}$ to $[0,+\infty)$. $f(|A|)$ will be called a measure of imprecision of $|A|$. A granule "x is A" is said to be more specific than another granule "x is B" if and only if $f(|A|) \leq f(|B|)$. If we use a decreasing function $f : \mathbb{N} \to [0,+\infty)$, $f(|A|)$ is called a measure of specificity of A.

These measures are readily extended to a body of evidence (F,m) defining

$$f(F,m) = \sum_{A \subseteq \Omega} m(A).f(|A|) \qquad (2)$$

(2) encompasses several noticeable information measures :

- f = identity : $|F,m| = \sum_{A \subseteq \Omega} m(A).|A|$ can be viewed as the (extended) cardinality of (F,m). If m is a usual probability measure $|F,m| = 1$; if (F,m) defines a possibility measure $|F,m|$ is the cardinality of the underlying fuzzy set [4].
- $f(x) = \mathrm{Log}_2 x$. We get the extension of Higashi-Klir measure of possibilistic information [8, 2] $U(F,m)$. It is zero on probability measures and maximal for the vacuous granule "x is Ω".
- $f(x) = 1/x$. We get Yager's measure of specificity [11] $S(F,m)$. It is minimal for "x is Ω", and it takes the value 1 for probability measures.

Properties of these measures of uncertainty are studied in [6, 7], such as monotonicity, additivity, etc ... according to the tradition of information theory.

2 - The principle of minimum specificity

Consider a partially characterized body of evidence under the form of a set of granules {"x is A_i", i = 1,n}, each granule being valued by a grade of credibility α_i. The pair (A_i,α_i) is interpreted as an uncertain item of information whose meaning is "My belief that x is A_i is α_i". The following conventions are adopted to assign a precise meaning to the number $\alpha_i \in [0,1]$:

→ $\alpha_i = 1$ means that "x is A_i" is certain.
→ $\alpha_i = 0$ means total ignorance about "x is A_i".
→ α_i is the grade of credibility $Cr(A_i)$ deriving from some body of evidence $(\mathcal{F},m)$ as defined by Shafer [12] :

$$Cr(A_i) = \sum_{\emptyset \neq B \subseteq A_i} m(B) \qquad (3)$$

that is Cr is a belief function. Especially, if $\bar{A}_i$ is the complement of A_i, $Cr(A_i)+Cr(\bar{A}_i) \leq 1$. The quantity $1-Cr(\bar{A}_i) = Pl(A_i)$ is called the plausibility of A_i ; it is 0 when the fact "x is A_i" is completely false.

A set of uncertain statements $\{(A_i,\alpha_i)|i = 1,n\}$ thus translates into the following proposition : there is a body of evidence $(\mathcal{F},m)$ such that $Cr(A_i) = \alpha_i$, i = 1,n, where Cr is defined by (3). Let $\mathcal{B}_n$ be the set of bodies of evidence $\{(\mathcal{F},m)|Cr(A_i) = \alpha_i, i = 1,n\}$. Generally this set contains more than one element, and we need some criterion to discriminate among them. A natural approach is to select the less specific bodies of evidence in $\mathcal{B}_n$, in the sense of index $f(\mathcal{F},m)$. This is the principle of minimum specificity.

Hence the calculation of a "best" body of evidence representing the contents $\{(A_i,\alpha_i)|i = 1,n\}$ of an uncertain knowledge base comes down to a constrained optimization problem with objective function $f(\mathcal{F},m)$. Let $x_A = m(A)$, $A \subseteq \Omega$ denote the decision variables ; we get the following linear programming problem

$$\left.\begin{aligned}
&\text{maximize } f(F,m) = \sum_{A \subseteq \Omega} x_A \cdot f(|A|) \\
&\text{under the constraints} \\
&\sum_{\emptyset \neq A \subseteq A_i} x_A = \alpha_i \quad , \quad i = 1,n \\
&\sum_{\emptyset \neq A \subseteq \Omega} x_A = 1 \\
&x_A \geq 0, \forall A.
\end{aligned}\right\} \quad (P)$$

where f is a measure of imprecision. This problem can be readily solved via a simplex algorithm. The result may depend upon the choice of f. Some light is shed on this question in the next section. It may happen that (P) has no solution. For instance if $\exists A_j \neq A_i, A_j = \bar{A}_i$ and $\alpha_i + \alpha_j > 1$. Such an occurrence indicates that the knowledge base is inconsistent.

3 - Combination of two items of information

In order to get a better feeling of what the result of (P) may look like, several simple examples are now studied. First note that if the knowledge base is empty, (P) reduces to max $\{f(F,m) \mid \sum_{\emptyset \neq A \subseteq \Omega} x_A = 1\}$, whose solution is simply $x_\Omega = 1$, i.e. (F,m) expresses total ignorance, which is satisfactory. When only one granule (A,α) is available, it is easy to check that the optimal solution of (P) does not depend on f and is defined by

$$x_A = \alpha \qquad ; \qquad x_\Omega = 1-\alpha \tag{4}$$

i.e. it is a simple support belief function focusing on A. [12].
Now consider the case of two granules of information (A,α) and (B,β), such that $A \cap B \neq \emptyset$. It can be proved that only 4 variables x_A, x_B, $x_{AB} = m(A \cap B)$ and x_Ω need be considered and (P) reduces to

maximize $x_\Omega \omega + x_A a + x_B b + x_{AB} c$

under the constraints
$$x_{AB} + x_A = \alpha \qquad (5)$$
$$x_{AB} + x_B = \beta \qquad (6)$$
$$x_\Omega + x_{AB} + x_A + x_B = 1 \qquad (7)$$

where $\omega = f(|\Omega|)$, $a = f(|A|)$, $b = f(|B|)$, $c = f(|A \cap B|)$.

Viewing x_A, x_B and x_Ω as slack variables, this problem has clearly only one degree of freedom, expressed by x_{AB}, and any value of x_{AB} such that, and only such that

$$\max(0,\alpha+\beta-1) \leq x_{AB} \leq \min(\alpha,\beta) \qquad (8)$$

yields a feasible solution. Eliminating slack variables in the objective function, (P) can be reformulated as

$$\underset{x_{AB}}{\text{maximize}} \quad \omega(1-\alpha-\beta) + a\alpha + b\beta + x_{AB}(\omega-a-b+c)$$

under constraints (8).

The solution of (P) clearly depends upon the sign of $(\omega-a-b+c) \triangleq k$ and is only one of the two following ones :

. <u>upper solution</u> $\quad x_{AB} = \min(\alpha,\beta)$; $x_\Omega = 1-\max(\alpha,\beta)$
$\quad x_A = \max(0,\alpha-\beta)$; $x_B = \max(0,\beta-\alpha)$

. <u>Lower solution</u> $\quad x_{AB} = \max(0,\alpha+\beta-1)$; $x_\Omega = \max(0,1-\alpha-\beta)$
$\quad x_A = \min(\alpha,1-\beta)$; $x_B = \min(\beta,1-\alpha)$.

<u>For instance</u> :

. choosing $f(F,m)$ = cardinality = $|F,m|$:
$k = |\Omega| - |A| - |B| + |A \cap C| \geq 0$, hence we obtain the upper solution if $\Omega \neq A \cup B$.

. choosing $f(F,m) = U(F,m)$, $f(x) = \log_2(x)$, then

$$2^k = \frac{|\Omega|.|A\cap B|}{|A|.|B|}$$

Hence if Ω is small enough (e.g. $\Omega = A\cup B$) $2^k < 1$ and the lower solution is obtained. But if Ω is large enough, $2^k > 1$ and the upper solution is recovered.

Remarks

. If $A\cap B = \emptyset$, we must cancel variable x_{AB} in (5-7) and the unique feasible body of evidence is defined by $x_A = \alpha$, $x_B = \beta$, $x_\Omega = 1-\alpha-\beta \geq 0$. If $\alpha+\beta > 1$ the two granules are inconsistent. A way to evaluate a degree of inconsistency is to keep x_{AB} as a slack variable in (5-7), where $x_{AB} = m(\emptyset)$, the weight committed to the empty set ; then the degree of inconsistency can be defined as $\min\{x_{AB} | (5\text{-}7)\}$. This degree is clearly $\alpha+\beta-1$ due to (8), an intuitively satisfactory result.

. Another way of processing the pair of granules (A,α) and (B,β) is to interpret each of them separately as independent granules

$m(A) = \alpha$; $m(\Omega) = 1-\alpha$

$m'(B) = \beta$; $m'(\Omega) = 1-\beta$

and then combine them via Dempster rule [12], i.e. if $A\cap B \neq \emptyset$
$m''(C'') = \sum\{m(C).m(C') | C\cap C' = C''\}$. This process leads to

$m''(A\cap B) = \alpha.\beta$ $\quad m''(\Omega) = (1-\alpha).(1-\beta)$

$m''(A) = \alpha(1-\beta)$ $\quad m''(B) = (1-\alpha)\beta$.

i.e. a feasible solution of (P) lying in between the upper and lower solution ! Hence our approach is a more general way of combining information than Dempster rule. Especially, no independence axiom is needed.

4 - Application to approximate reasoning : the modus ponens

Consider two uncertain granules of information expressed as formulas in propositional logic, say p and $p\rightarrow q$, such that $Cr(p) = \alpha$, $Cr(p\rightarrow q) = \beta$. Note that $p\cap(p\rightarrow q) = p\cap q$ is not a contradiction. Using the results of section 2, where A represents p and B represents $p\rightarrow q$, we can characterize the set of uncertainty measures over the set of propositions generated by p and q as a set of linear constraints, and choose the proper uncertainty measure using the principle of minimum specificity, under the form of a basic assignment m. The credibility of proposition q can be obtained using (3), applied on the upper or the lower solution. We thus obtain the two patterns of inference

<u>upper solution</u>

$$\frac{Cr(p) = \alpha \qquad Cr(p \to q) = \beta}{Cr(q) = \min(\alpha,\beta).} \qquad (I)$$

<u>lower solution</u>

$$\frac{Cr(p) = \alpha \qquad Cr(p \to q) = \beta}{Cr(q) = \max(0,\alpha+\beta-1)} \qquad (II)$$

(I) is always valid if Cr is a necessity measure (see [10]) while (II) is always valid (as a lower bound) if Cr is a probability measure. The actual solution lies in between, i.e. $\max(0,\alpha+\beta-1) \leq Cr(q) \leq \min(\alpha,\beta)$, a result which is already known [5] under the form

$$\min(Cr(A), Cr(B)) \geq Cr(A \wedge B) \geq \max(0,Cr(A)+Cr(B)-1)$$

Contrastedly, using Dempster rule of combination we get the following pattern of inference [1]

$$\frac{Cr(p) = \alpha \qquad Cr(p \to q) = \beta}{Cr(q) = \alpha.\beta} \qquad (III)$$

which lies in between (hence is consistent) with (I) and (III). When $\alpha = \beta = 1$, (I-II-III) collapse into the usual modus ponens. It is noticeable that (I), (II) and (III) use the three basic semi-groups of [0,1] with identity 1, i.e. the triangular norms minimum, product and $\max(0,x+y-1)$. These operations are thus confirmed as proper ones for the propagation of uncertainty, regardless of the uncertainty measure which is used.

<u>Remark</u> : In all patterns (I) (II) (III), we always get $Pl(q) = 1$.

5 - Perspectives

This approach can be extended to deal with any knowledge base containing a set of statements represented, for instance, by means of formulas in propositional logic $\{p_i | i = 1,n\}$ taken as axioms, each p_i being weighted by a grade of credibility α_i. The approximate reasoning procedure can be outlined as follows :

1. Write the constraints of (P) for each axiom p_i, i.e. $Cr(p_i) = \alpha_i$
2. Solve (P) using the minimum specificity principle
3. If a feasible solution exists compute $Cr(q_j)$, $Pl(q_j)$ where the q_j's are the propositions the truth of which must be established.

Step 2 is solved by means of standard linear programming. In other words the inference engine is based on a simplex algorithm. Adding or retrieving granules of knowledge comes down to adding or deleting a constraint. Especially some analysis of why the knowledge base is inconsistent can be performed by checking for violated constraints.

This approach constrasts with another one based on Dempster rule of combination [1].

References

[1] Chatalic, P., Dubois, D., Prade, H. An approach to approximate reasoning based on Dempster rule of combination. Submitted to the 8th IASTED Inter. Symp. on Robotics and Artificial Intelligence, Toulouse, France, June 18-20, 1986.

[2] Dubois, D., Prade, H. A note on measures of specificity for fuzzy sets. Int. J. General Systems, 10, 279-283, 1985. Preprint in BUSEFAL n° 19 (L.S.I., Univ. P. Sabatier, Toulouse, 1984, 83-89.

[3] Dubois, D., Prade, H. Théorie des Possibilités. Applications à la Représentation des Connaissances en Informatique. Masson, Paris, 1985. English translation to be published by Plenum Press.

[4] Dubois, D., Prade, H. Fuzzy cardinality and the modelling of imprecise quantification. Fuzzy Sets & Systems, 16, 199-230, 1985.

[5] Dubois, D., Prade, H. Combination and propagation of uncertainty with belief functions. A reexamination. Proc. 9th Inter. Joint Conf. Artificial Intelligent, Los Angeles, 111-113, 1985.

[6] Dubois, D., Prade, H. Additivity and monotonicity of measures of information defined in the setting of Shafer's evidence theory. BUSEFAL n° 24, (L.S.I., Univ. P. Sabatier, Toulouse), 64-76, 1985.

[7] Dubois, D., Prade, H. Properties of measures of information in evidence and possibility theories. Submitted to Fuzzy Sets & Systems, 1985.

[8] Higashi, M., Klir, G.J. Measures of uncertainty and information based on possibility distributions. Int. J. General Systems, 9, 43-58, 1983.

[9] Klir, G.J. Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like ? Fuzzy Sets & Systems, to appear.

[10] Prade, H. A computational approach to approximate and plausible reasoning, with application to expert systems. IEEE Trans. on Pattern Analysis and Machine Intelligence, 7, 260-283, 1985. Corrections : 7, 747-748, 1985.

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

[12] Shafer, G. A Mathematical Theory of Evidence. Princeton Univ. Press. 1976.

[13] Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets & Systems, 1, 3-28, 1978.