A note on the validation of possibilistic knowledge bases

Didier DUBOIS - Henri PRADE

I.R.I.T.-C.N.R.S.
Université Paul Sabatier, 31062 Toulouse Cedex, France

Roughly speaking, a knowledge base $\mathcal{K}$ will be said to be "coherent" (Ayel and Rousset, 1990) if there does not exist any piece of input data which respects integrity constraints $\mathcal{C}$ and which leads to inconsistency when added to $\mathcal{K}$. Preliminary studies on the characterization of potential inconsistencies in uncertain knowledge bases (represented in the framework of possibility theory) can be found in Yager and Larsen (1990) ; see also Larsen and Nonfjall (1989).

Each piece of information is represented by a possibility distribution π_i in the possibilistic framework. The information is expressed in terms of (fuzzy) restrictions on the possible values of tuples of variables e.g. X, Y, Z, in "X is A", "if X is B and Y is C then Z is D" where A, B, C and D denote subsets which may be fuzzy. Then a piece of information can be represented as a possibility distribution on the Cartesian product of the domain of the variables ; see (Dubois and Prade, 1991). The whole knowledge base $\mathcal{K}$ is associated with the possibility distribution $\pi_{\mathcal{K}}$ obtained as the conjunction of the π_i's representing the pieces of information in $\mathcal{K}$, i.e.

$$\pi_{\mathcal{K}} = \min_i \pi_i.$$

The set of integrity constraints induces a possibility distribution $\pi_{\mathcal{C}}$ which represents to what extent each possible assignment of values to the *input* variables is permitted.

Let us characterize what the coherence of $\mathcal{K}$ with respect to $\mathcal{C}$ means. If $\mathcal{K}$ and $\mathcal{C}$ are stated in terms of restrictions on the possible values of variables, let $X_1, \ldots, X_n$ be the input variables, i.e. the variables which are used to express the input data. It is a subset of all the variables involved in $\mathcal{K}$. Then the coherence of $\mathcal{K}$ with respect to $\mathcal{C}$ means that we have the following inequality

$$\pi_{\mathcal{C}} \leq \text{proj}(\pi_{\mathcal{K}} ; X_1 \times \ldots \times X_n)$$

where $\text{proj}(\pi_{\mathcal{K}} ; X_1 \times \ldots \times X_n)$ is the projection of $\pi_{\mathcal{K}}$ on the Cartesian product of the domains of the X_i. It guarantees that each assignment of values to $(X_1, \ldots, X_n)$ is at least as possible according to $\mathcal{K}$ as according to $\mathcal{C}$. When there is no integrity constraint, $\pi_{\mathcal{C}} = 1$ everywhere, and the above condition writes

$$\text{proj}(\pi_{\mathcal{K}} ; X_1 \times \ldots \times X_n) = \text{dom}(X_1) \times \ldots \times \text{dom}(X_n)$$

where $dom(X_i)$ is the domain of variable X_i (Yager and Larsen, 1990). It entails that $proj(\pi_{\mathcal{K}} ; X_i) = dom(X_i)$, $\forall\ i = 1,n$, but these conditions are not equivalent to the above condition.

In case of possibilistic knowledge bases made of weighted classical logical formulas, the coherence problem corresponds to the research of "no goods" in a possibilistic Assumption-based Truth Maintenance Systems, where assumptions are the input literals (Dubois, Lang and Prade, 1991).

Impossibility qualification

One related problem is then to express the conjunction of integrity constraints under the form of the possibility distribution $\pi_{\mathcal{C}}$. Usually integrity constraints are rather stated in terms of impossibility than in terms of possibility. For instance, e.g. "X is A" is forbidden, is impossible, or more generally "X is A is *almost* impossible".

"X is A is impossible to the degree α" can be translated as the following constraint on the induced possibility distribution $\pi_{\mathcal{C}}$

- if A is an ordinary set

$$\Pi(A) = \sup_{x \in A} \pi_{\mathcal{C}}(x) \leq 1 - \alpha \Leftrightarrow \forall\ x \in A,\ \pi_{\mathcal{C}}(x) \leq 1 - \alpha$$

  i.e. (when $\alpha = 0$ there is no constraint on $\pi_{\mathcal{C}}$ and when $\alpha = 1$, all the values in A are excluded as expected)

- if A is a fuzzy set

  i) if we understand "X is A is almost impossible" as "X is $\bar{A}$ is almost certain" (where $\bar{A}$ denotes the (fuzzy) complement of A), we can apply results about certainty qualification (Dubois and Prade, 1991) which yield

$$\pi_{\mathcal{C}}(x) = \max(\mu_{\bar{A}}(x), 1 - \alpha) = \max(1 - \mu_A(x), 1 - \alpha) = 1 - \min(\mu_A(x), \alpha).$$

  The result is pictured on Figure 1 ;

  ii) starting with the usual definition of the possibility of a fuzzy event leads to write

$$\Pi(A) = \sup \min(\mu_A(x), \pi'_{\mathcal{C}}(x)) \leq 1 - \alpha$$

$$\text{i.e. } \forall x,\ \min(\mu_A(x), \pi'_{\mathcal{C}}(x)) \leq 1 - \alpha$$

$$\text{i.e. } \forall x,\ \pi'_{\mathcal{C}}(x) \leq [\mu_A(x) \rightarrow (1 - \alpha)] \text{ with } a \rightarrow b = \begin{cases} 1 \text{ if } a \leq b \\ b \text{ if } a > b \end{cases}$$

Allowing the greatest possibility degrees compatible with this constraint yields the possibility distribution π'_{ζ} pictured in Figure 1, i.e.

$$\pi'_{\zeta} = \mu_A \rightarrow (1 - \alpha).$$

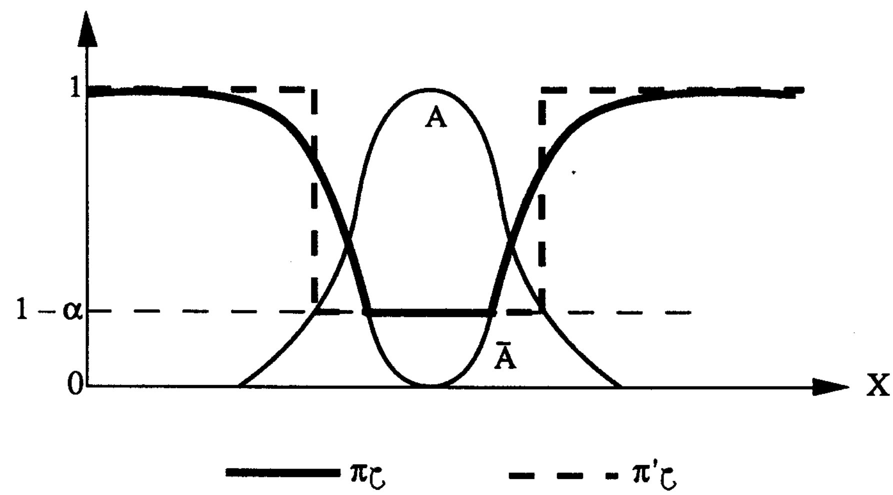

Figure 1

Figure 1 exhibits the difference between $\pi_{\zeta} = 1 - \min(\mu_A, \alpha)$ and $\pi'_{\zeta} = \mu_A \rightarrow (1 - \alpha)$. In the translation of "X is A almost impossible" by π'_{ζ}, the gradual nature of A is lost ; π'_{ζ} corresponds to a rather drastic solution which relates the (ordinary) subset concerned by the impossibility to the level of impossibility, via μ_A. The choice between π_{ζ} and π'_{ζ} is a matter of context.

References

Ayel M., Rousset M.C. (1990) La Cohérence dans les Bases de Connaissances. Cepadues-Editions, Toulouse.

Dubois D., Lang J., Prade H. (1991) A possibilistic assumption-based truth maintenance system with uncertain justifications, and its application to belief revision. Proc. ECAI Workshop on Truth-Maintenance Systems, Stockholm, Aug. 6, 1990 (J.P. Martins, M. Reinfrank, eds.), Lecture Notes in Computer Sciences, n° 515, Springer Verlag.

Dubois D., Prade H. (1991) Fuzzy sets in approximate reasoning – Vol. 1 : Inference with possibility distributions. Fuzzy Sets and Systems, 40, 143-202.

Larsen H.L., Nonfjall H. (1989) Modeling in the design of a KBS validation system. Proc. of the 3rd Inter. Fuzzy Systems Assoc. (IFSA) Congress, Seattle, Wash., Aug. 6-11, 341-344.

Yager R.R., Larsen H.L. (1990) On discovering potential inconsistencies in validating uncertain knowledge bases by reflecting on the input. Tech. Report #MII-1001, Machine Intelligence Institute, Iona College, New Rochelle, NY.