BUSEFAL n° 32 (1987)

A note on the evaluation of conditions involving vague quantifiers in presence of imprecise or uncertain information

Henri PRADE

Laboratoire "Langages et Systèmes Informatiques"
Université Paul Sabatier, 118 route de Narbonne
31062 TOULOUSE Cédex (France)

Abstract

In this note, we are interested in the evaluation of conditions of the form "The value of attribute **a** for Q items of X is in F" or more shortly "Q items of X are F" (and of some other related forms), where X is a set of items, Q denotes a possibly vague proportion (which may be linguistically expressed, e.g. "most"), F is a (possibly fuzzy) subset of the attribute domain of **a**, and where the available knowledge about the value **a**(x) of the attribute **a** for any item x may be imprecise or even vague. The evaluation is based on a fuzzy pattern matching procedure repeated two times. Such conditions may be encountered in queries addressed to an incomplete information data base or in the "if-part" of expert rules.

1. Introduction and background

Let $X = \{x_1, \ldots, x_n\}$ be a finite set of items. Let $A(x_i)$ denote the (possibly fuzzy) subset restricting the (more or less) possible values of the attribute **a** for x_i;$A(x_i)$ represents the available knowledge ; the information is not precise as soon as $A(x_i)$ is not a singleton. Let F be a (possibly fuzzy) subset of the attribute domain D_a of **a**. The extent to which it is possible (resp. necessary or certain) that the value **a**(x) of the attribute **a** for the item x, known to be restricted by A(x), is compatible with F is given by the number $\Pi(F;A(x))$ (resp. $N(F;A(x))$) which is defined by

$$\Pi(F;A(x)) = \sup_{d \in D_a} \min(\mu_F(d), \mu_{A(x)}(d)) \qquad (1)$$

(resp. $N(F;A(x)) = 1 - \Pi(F^c;A(x))$ (2)

$$= \inf_{d \in D_a} \max(\mu_F(d), 1 - \mu_{A(x)}(d)) \quad (3)$$

where the membership function of the complement F^c of F is defined by

$$\mu_{F^c}(d) = 1 - \mu_F(d) \quad (4)$$

The two quantities $\Pi(F;A(x))$ and $N(F;A(x))$ play a central rôle in the fuzzy pattern matching technique developed in (Cayrol, Farreny and Prade, 1980 and 1982).

2. Fuzzy evaluation of the number of items whose attribute value belongs to F

In this section, F is assumed to be an ordinary subset of D_a.

For each $x_i \in X$ we have the pair ($\Pi(F;A(x_i))$, $N(F;A(x_i))$) computed from (1) and (3). The $\Pi(F;A(x_i))$'s, for i = 1,n, define the fuzzy set of items whose attribute value possibly belongs to F. A fuzzy-valued cardinality of this fuzzy set is easily obtained using the following procedure (Prade, 1984), (Dubois and Prade, 1985), (Zadeh, 1983)

1) rank the $\Pi(F;A(x_i))$'s in decreasing order. Let u_i be the value of rank i in this ordering, for i = 1,n and let $u_0 = 1$.

2) the fuzzy-valued cardinality we use is then the fuzzy subset of $\mathbb{N}$ defined by

$$u_0/0 + u_1/1 + \ldots + u_n/n$$

where the membership grade is before the '/' and the corresponding element of $\mathbb{N}$ after ; here + denotes the union of singletons.

Example

$x = \{x_1, x_2, \ldots, x_6\}$ and we have the following values for the $\Pi(F;A(x_i))$'s :

	x_1	x_2	x_3	x_4	x_5	x_6
$\Pi(F;A(x_i))$	0.2	1	0.8	1	0.5	0.5

Then the fuzzy cardinality is given by

$$1/0 + 1/1 + 1/2 + 0.8/3 + 0.5/4 + 0.5/5 + 0.2/6$$

Note that this fuzzy integer is always normalized since we always have $u_0 = 1$. This fuzzy set represents the number of elements which may be found in the fuzzy set defined by the $\Pi(F;A(x_i))$'s. However, note that in case of an ordinary set with $p(\leq n)$ elements, it reduces to the subset of $\mathbb{N}$ $\{0,1,\ldots,p\}$ and not to $\{p\}$. See (Dubois and Prade, 1985) for a discussion.

Similarly, by ordering the $N(F;A(x_i))$'s in a decreasing order we can compute a fuzzy subset of $\mathbb{N}$ which represents the number of items whose attribute value more or less certainly belongs to F. Let v_i be the value of rank i in the ordering of the $N(F;A(x_i))$'s. The fuzzy-valued cardinality is then given by

$$\sum_{i=0}^{n} \min(u_i, 1 - v_{i+1}) / i \qquad (5)$$

where $\sum$ stands for the repeated use of '+' and $v_{n+1} = 0$. (5) must be understood in the following way. We are certain at the degree v_i that there are at least i items which satisfy the condition, and then it is possible at the degree $1 - v_{i+1}$ that there are at most i items (due to (2) and since the negation of "at least i" is "at most i - 1"). Moreover, it is possible at the degree u_i that there are at least i items which satisfy the condition. Finally, $\min(u_i, 1 - v_{i+1})$ is the possibility that there are at least i and at most i items, in other words it is the possibility that there are exactly i items which satisfy the condition (Prade, 1984).

When F is an ordinary subset, we always have

$$\forall i, \Pi(F;A(x_i)) < 1 \Rightarrow N(F;A(x_i)) = 0 \qquad (6)$$

Then the fuzzy subset of $\mathbb{N}$ defined by (5), is always normalized (i.e. there is at least one value with a membership grade equal to 1). In case of complete information (i.e. all the $A(x_i)$'s are singletons), this fuzzy set reduces to a singleton corresponding to a precise integer, which is the number (precisely known in that case) of items which satisfy the condition. Let K be the fuzzy proportion of items of X which satisfy the condition ; we have

$$\forall i, 0 \leq i \leq n, \ \mu_K\left(\frac{i}{n}\right) = \min(u_i, 1 - v_{i+1}) \qquad (7)$$

We are now in position for evaluating (in terms of possibility and necessity) the condition "The value of attribute a for Q items of X is in F" by computing (using (1) and (3) again) the quantities $\Pi(Q^*;K)$ and $N(Q^*;K)$ where

μ_{Q^*} denotes the restriction of the membership function μ_Q (defined from [0,1] to [0,1], since it represents a vaguely specified proportion in the most general case) to the subset of rational number $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$. We have

$$\Pi(Q^*;K) = \max_{i=0,n} \min(\mu_Q(\frac{i}{n}), u_i, 1 - v_{i+1}) \qquad (8)$$

$$N(Q^*;K) = \min_{i=0,n} \max(\mu_Q(\frac{i}{n}), 1 - u_i, v_{i+1}) \qquad (9)$$

$\Pi(Q^*;K)$ (resp. $N(Q^*;K)$) estimates to what extent it is possible (resp. necessary or certain) that the condition "The value of attribute **a** for Q items of X is in F" is satisfied when the available information is represented by the $A(x_i)$'s. Thus, by repeating the fuzzy pattern matching procedure two times, first in order to estimate the compatibility of each item with the requirement, second in order to compare the (fuzzily-known) relative number of items which satisfy this requirement with the requested proportion.

N.B. : In case Q would restrict the possible value of an absolute number (rather than a relative number, i.e. a proportion), the above procedure applies as well changing $\mu_K(\frac{i}{n})$ into $\mu_K(i)$ in (7) and $\mu_Q(\frac{i}{n})$ into $\mu_Q(i)$ in (8)-(9).

3. Case when F is a fuzzy set

When F is a fuzzy set, (6) no longer holds and we only have $\Pi(F;A(x_i)) \geq N(F;A(x_i))$ (provided that $A(x_i)$ is normalized). Then, the fuzzy set of $\mathbb{N}$ defined by (5) and the fuzzy set K may be subnormalized. This is due to the fact that with a fuzzy set we cannot definitely say if a precise attribute value satisfy or not the corresponding requirement. This uncertainty is echoed by the subnormalization of K. However this situation is undesirable here since this subnormalization would blur the computation of $\Pi(Q^*;K)$ and $N(Q^*;K)$.

One way of escaping this problem would be to renormalize K by dividing each membership grade by the height of K. A perhaps less ad hoc way to cope with this difficulty is to modify the definition of N in order to preserve (6). This may be done in the following way.

3.1. A new definition of the necessity of a fuzzy event based on a para-consistent complementation

N is defined from Π by (2) using the fuzzy set complementation (4). There are two other (extreme) ways of defining the complement A^c of a subset A of D by extending the following identities for classical sets

$$A^c = \cap \{S, A \cup S = D\} \qquad (10)$$

or

$$A^c = \cup \{S, A \cap S = \emptyset\} \qquad (11)$$

This yields (see (Dubois and Prade, 1983) for instance), using min and max respectively for defining the intersection and the union, to the following definitions in case of fuzzy sets :

• *para-consistent complementation (from (10))*

$$\mu_{F^c}(d) = \inf\{s \in [0,1], \max(\mu_F(d), s) = 1\} = \begin{cases} 1 \text{ if } \mu_F(d) < 1 \\ 0 \text{ if } \mu_F(d) = 1 \end{cases} \qquad (12)$$

Note that $F^c = [\text{core}(F)]^c$ with $\text{core}(F) = \{d \in D, \mu_F(d) = 1\}$; $F^{cc} \subseteq F$ and $F \cup F^c = D$ (but we have not $F \cap F^c = \emptyset$).

• *intuitionist complementation (from (11))*

$$\mu_{F^c}(d) = \sup\{s \in [0,1], \min(\mu_F(d), s) = 0\} = \begin{cases} 1 \text{ if } \mu_F(d) = 0 \\ 0 \text{ if } \mu_F(d) > 0 \end{cases} \qquad (13)$$

Note that $F^c = [\text{support}(F)]^c$ with $\text{support}(F) = \{d \in D, \mu_F(d) > 0\}$; $F^{cc} \supseteq F$ and $F \cap F^c = \emptyset$ (but we have not $F \cup F^c = D$).

Using the para-consistent complementation for defining $N(F;A(x_i)) = 1 - \Pi(F^c;A(x_i))$ guarantees that $\Pi(F;A(x_i)) < 1 \Rightarrow \Pi(F^c;A(x_i)) = 1$ (provided that $A(x_i)$ is normalized) since $F \cup F^c = D$, and consequently $N(F;A(x_i)) = 0$ if $\Pi(F;A(x_i)) < 1$; then (6) is preserved.

N.B. : Keeping the definition (3) of N and defining Π from N by $\Pi(F;A(x_i)) = 1 - N(F^c;A(x_i))$ using the intuitionist complementation would preserve (6) also.

3.2. Proposed treatment when F is a fuzzy set

When F is a fuzzy set, we compute the $\Pi(F;A(x_i))$'s by (1) and the $N(F;A(x_i))$'s by (2) using the para-consistent complementation (12). Then (6) is preserved, K remains normalized and (8)-(9) still apply without any particular problem.

Let us examine the particular case where F is a fuzzy set and the pieces of information $A(x_j)$ are precise ; i.e. $\forall j$, $A(x_j) = \{d_j\}$, $d_j \in D$. Then $\Pi(F;A(x_j)) = \mu_F(d_j)$ and $N(F;A(x_j)) = 1 - \Pi([\text{core}(F)]^c;A(x_j)) = \mu_{\text{core}(F)}(d_j)$ (which contrasts with the situation where the complementation (4) is used, leading to $\Pi(F;A(x_j)) = N(F;A(x_j)) = \mu_F(d_j)$ in case of precise information). Then the v_i's introduced in section 2. are such that $\exists k,\ v_i = 1$ for $i = 0,k$ and $v_i = 0$ for $i = k + 1,n$. Moreover, $u_K = 1$ (due to (6)). Then the subset of $\mathbb{N}$ defined by (5) can now be written

$$\sum_{i=k}^{n} u_i / i \qquad (14)$$

since $\forall i,\ k \le i \le n - 1,\ v_{i+1} = 0$ and $\forall i,\ 0 \le i \le k - 1,\ v_{i+1} = 1$. Besides, since $\mu_{core(F)}(d_j) = 0 \Rightarrow \mu_F(d_j) < 1$, then $v_j = 0 \Rightarrow u_j < 1$ and in particular $u_{k+1} < 1$. Thus in case F would be an ordinary set, k would be the precise value of the number of items satisfying the requirement.

Yager (1984) has proposed the following estimation of the extent to which the condition "Q items of X are F" is satisfied, where Q and/or F are fuzzy, Q being a relative quantifier, in case of precise information

$$\max_{C \subseteq X} \min(\mu_Q(\frac{|C|}{n}), \min_{j\, d_j \in C} \mu_F(d_j)) \qquad (15)$$

where |C| denotes the cardinality of C. Introducing the rank k defined above, (15) is still equal to

$$\max[\max_{|C| \le k} \min(\mu_Q(\frac{|C|}{n}), \min_{j\, d_j \in C} \mu_F(d_j)), \max(\min_{|C| \ge k+1} (\mu_Q(\frac{|C|}{n}), \min_{j\, d_j \in C} \mu_F(d_j))]$$

$$= \max[\max_{j \le k} \mu_Q(\frac{j}{n}), \max_{j \ge k+1} \min(\mu_Q(\frac{j}{n}), u_j)]$$

$$= \max_{0 \le j \le n} \min(\mu_Q(\frac{j}{n}), u_j) \qquad (16)$$

The expression (15) estimates to what extent it is possible to find a relative number, compatible with Q, of items which are F. In other words it corresponds to the possibility that there is a proportion "at least Q" of items which are F. Indeed, (16) is equal to the expression of $\Pi(Q^*;K)$ taking into account (14), i.e. to

$$\Pi(Q^*;K) = \max_{k \le i \le n} \min(\mu_Q(\frac{i}{n}), u_i)$$

provided that "Q" and "at least Q" are identical, i.e. μ_Q is non-decreasing (which implies that $\max_{0 \le j \le k} \min(\mu_Q(\frac{j}{n}), u_j) = \mu_Q(\frac{k}{n})$). Thus the approach here and Yager's proposal are in agreement in the particular situations where both apply. See also (Dubois, Prade and Testemale, 1986) for the relation with the evaluation of conditions of the kind "at least Q criteria (among n) are satisfied" ; (here, the n criteria are "the value of attribute a for x_i is in F" for i = 1,n) viewed as a special case of weighted pattern matching ; an expression similar to (16) is then obtained.

4. Concluding remarks

The proposed approach which repeatedly make use of the pattern matching procedure first described in (Cayrol, Farreny and Prade, 1980) can be easily implemented and is computationally simple using trapezoidal possibility distributions. Moreover the condition "the value of attribute a for the item x is in F" can be replaced more generally in the approach by a compound condition involving several attributes.

References

Cayrol M., Farreny H., Prade H. (1980) Possibility and necessity in a pattern matching process. Proc. IXth. Inter. Cong. on Cybernetics, Namur, Belgium, 8-13 Sept., 53-65

Cayrol M., Farreny H., Prade H. (1982) Fuzzy pattern matching. Kybernetes, 11, 103-116

Dubois D., Prade H. (1983) Fuzzy set-theoretic differences and inclusions and their use in fuzzy arithmetics and analysis. Presented at the 5th Inter. Seminar on Fuzzy Set Theory, J. Kepler Univ., Linz (Austria), 5-9 Sept., 1983. In : Tech. Rep. "Ensembles Flous-83", n° 191 (L.S.I., Univ. P. Sabatier, Toulouse), 107-129, April 1984

Dubois D., Prade H. (1985) Fuzzy cardinality and the modeling of imprecise quantification. Fuzzy Sets and Sytems, 16, 199-230

Dubois D., Prade H., Testemale C. (1986) Weighted fuzzy pattern matching. Proc. Jour. Nation. sur les Ensembles Flous, la Théorie des Possibilités et leurs Applications, Toulouse, 2-7 June, 1986, 115-145. Revised version in Fuzzy Sets and Systems, to appear

Prade H. (1984) Lipski's approach to incomplete information data bases restated and generalized in the setting of Zadeh's possibility theory. Information Systems, 9(1), 27-42

Yager R.R. (1984) General multiple objective decision functions and linguistically quantified statements. Inter. J. Man-Machine Studies, 21, 389-400

Zadeh L.A. (1983) A computational approach to fuzzy quantifiers in natural languages. Computer and Mathematics with Applications, 9(1), 149-184