RESEMBLANCE IN DATABASE UTILIZATION

Bernadette Bouchon-Meunier, Maria Rifqi

RESEMBLANCE IN DATABASE UTILIZATION

Bernadette BOUCHON-MEUNIER *, Maria RIFQI *

* LAFORIA-IBP, UPMC, Case 169, 4 place Jussieu, 75252 Paris Cedex 05, France.

e-mail : bouchon, rifqi@laforia.ibp.fr

Abstract There exist many quantities evaluating the resemblance between two objects sharing the same features. These various measures differ from one another in properties because of their different utilizations. In this paper, we focus on a class of such quantities and we give a formalization of measures of resemblance.

1 Introduction

Similarity relations have been extensively studied in fuzzy set theory, generally with applications in decision-making or artificial intelligence in mind. In the case of fuzzy subsets, resemblances are generally seen through similarity relations [8].

Extensions of this concept have also been introduced to describe degrees of resemblance, proximity, inclusion, compatibility or satisfiability between two fuzzy sets [4], [10]. The properties satisfied by these degrees differ from one another, according to the purpose of their introduction. In this paper, we focus on a general measure for the degrees corresponding to the idea of resemblance of an object with a reference one in a fuzzy framework. This approach does not cover the idea of inclusion between fuzzy sets, which will be addressed in a forthcoming paper.

2 Resemblance between fuzzy sets

2.1 Presentation of the problem

We consider the following situation : an object O is supposed to be given as a reference, described by means of attributes with fuzzy or crisp possible values. For instance, let us consider an apartment, attributes such as the size, the prize, the location, and their possible values "big", "$250m^2$", "approximately 100000 \$", "downtown" can be taken into account. Let us study a new object O', characterized by means of the same attributes. We want to determine the degree of resemblance of O' with regard to O.

Several purposes could underlie this study. In a database querying process, O is the query and O' a datum in the database. In an analogical reasoning system, O is a well-known phenomenon and O' a new phenomenon for which the value of a few attributes must be found. In case-based reasoning, O is an already solved case and O' is a new case which must be associated with a decision.

2.2 Measure of resemblance

For any set Ω of elements, let us denote by $F(\Omega)$ the set of fuzzy subsets of Ω and by f_A the membership function of any A in $F(\Omega)$. If not specified differently, we use the classical definition of intersection : $f_{A\cap B} = \min(f_A, f_B)$.

We suppose that we are given a *fuzzy measure* M, which is a mapping defined on $F(\Omega)$ and lying in $\mathbb{R}^+$ such that, for every A and B in $F(\Omega)$:

MI1 :$M(\emptyset) = 0$

MI2 : if $B \subseteq A$, then $M(B) \leq M(A)$.

If the values of M are restricted to $[0, 1]$, M is a fuzzy measure introduced by M. Sugeno [5].

Examples of fuzzy measures can be defined as :

$$\begin{aligned} M_1(A) &= \int_\Omega f_A(x)dx & (1)\\ M_2(A) &= \sup_{x\in\Omega} f_A(x) & (2)\\ M_3(A) &= \textstyle\sum_{count}(A) & \\ &= \textstyle\sum_{x\in\Omega} f_A(x) \quad \text{if } \Omega \text{ is finite [3]} & (3) \end{aligned}$$

Definition 1 *An operation on* $F(\Omega)$ *is called a* difference *and denoted by* $-$*, if it satisfies for every* A *and* B *in* $F(\Omega)$ *:*

D1 *: if* $A \subseteq B$*, then* $A - B = \emptyset$*.*

D2 *:* $A - B \subseteq A - (A \cap B)$

D3 *:* $B - A$ *is monotonous with regard to* B *:* $B \subseteq B'$*, entails* $B - A \subseteq B' - A$

In the case of crisp subsets of Ω, $A - B$ can be defined as the complement of $A \cap B$ in A. Examples of differences can be defined as :

$$f^1_{A-B}(x) = \max(0, f_A(x) - f_B(x)) \quad [9] \quad (4)$$

$$f^2_{A-B}(x) = \begin{cases} f_A(x) & \text{if } f_B(x) = 0 \\ 0 & \text{if } f_B(x) > 0 \end{cases} \quad (5)$$

Definition 2 *Let Ω be a set of elements. A* measure of resemblance *on Ω is a mapping $R : F(\Omega) \times F(\Omega) \to [0,1]$ satisfying the following requirements :*

MR1 - *containment : if $B \subseteq A$, then $R(A,B) = 1$*

MR2 - *exclusiveness : if $A \cap B = \emptyset$, then $R(A,B) = R(B,A) = 0$*

MR3 - *external information : if $M(A \cap B) = M(A \cap B')$, then $(M(B - A) \geq M(B' - A)$ entails $R(A,B) \leq R(A,B'))$*

MR4 - *internal information: if $M(B - A) = M(B' - A)$, then $(M(A \cap B) \geq M(A \cap B')$ entails $R(A,B) \geq R(A,B'))$*

These properties can be interpreted as follows : if B is a particular case of A, then B resembles perfectly A; if B has nothing to do with A, then it does not resemble A at all. For a given amount of information common to A and B, then, the smaller the information contained in B and not in A, the more B resembles A. For a given amount of information contained in B and not in A, the bigger the information common to A and B, the more B resembles A.

3 Properties of a measure of resemblance

We study properties of a general measure of resemblance.

Definition 3 *A measure of resemblance is called a F-*measure of resemblance *if it is defined as :*

$$R(A,B) = F(M(A \cap B), M(B - A)) \quad (6)$$

for a function $F : I\!R^+ \times I\!R^+ \to [0,1]$.

Property 1 *Let us consider a function $F : I\!R^+ \times I\!R^+ \to [0,1]$. Then the mapping $R : F(\Omega) \times F(\Omega) \to [0,1]$ defined as : $R(A,B) = F(M(A \cap B), M(B - A))$ is a measure of resemblance if and only if :*

F1 *: $F(u,v)$ is non decreasing in u and non increasing in v*

F2 *: $F(u,0) = 1$*

F3 *: $F(0,v) = 0$*

Property 2 *A measure of resemblance satisfies the following properties :*

MR5 - *reflexivity : $R(A,A) = 1$*

MR6 - *examplarity : if $M(B - A) \geq M(A - B)$, then $R(A,B) \leq R(B,A)$.*

MR7 - *elementary internal information : if $A \cap B = A \cap B'$, then $B - A \supseteq B' - A$ entails $R(A,B) \leq R(A,B')$*

MR8 - *elementary external information : if $B - A = B' - A$, then $A \cap B \supseteq A \cap B'$ entails $R(A,B) \geq R(A,B')$*

Remark : Properties MR7 and MR8 are immediate consequences of properties MR3 and MR4 because of MI2 (monotonicity of M).

In the case of crisp subsets, Tversky has given an axiomatic theory of similarity [6] on a set of features Ω, which is compatible with our definition of a measure of resemblance. He defines the similarity of B to A as : $R(A,B) = f(A \cap B, B - A, A - B)$. In the case where $f(u,v,w)$ is independent of w, we find definition 3. He requires R to satisfy additional properties, called monotonicity, independence and solvability, which are also satisfied by our measures with the definitions given in the following property :

Property 3 *An F-measure of resemblance satisfies the following additional properties :*

MR9 - *monotonicity : if $A \cap B \supseteq A \cap B'$ and $B - A \subseteq B' - A$, then $R(A,B) \geq R(A,B')$*

MR10 - *independence : if $A \cap B = C \cap D$ and $A' \cap B' = C' \cap D', B - A = B' - A'$ and $D - C = D' - C'$, then $R(A,B) \geq R(A',B')$ if and only if $R(C,D) \geq R(C',D')$*

MR11 - *relative independence : if $A \cap B = A \cap D$ and $A \cap B' = A \cap D', B - A = B' - A$ and $D - A = D' - A$, then $R(A,B) \geq R(A,B')$ if and only if $R(A,D) \geq R(A,D')$*

MR12 - *solvability : for any A, B, B' such that $M(A \cap B) > M(A \cap B')$ and $M(B - A) < M(B' - A)$, there exists a pair (P,Q) such that $M(P \cap Q) = M(A \cap B), M(Q - P) = M(B' - A)$ and $R(A,B) > R(P,Q) > R(A,B')$ if $F(u,v)$ is strictly increasing in u and strictly decreasing in v.*

The independance property can be interpreted as follows : it is equivalent to compare the resemblance of a first pair (A,B) to the resemblance of a second pair (A',B') with the same difference, or to compare the resemblance of any other first pair (C,D) to the resemblance of any other second one (C', D'), also with the same difference, as soon as the information common to the first pair is preserved, the information common to the second pair is preserved.

The relative independance is a particular case of the independance, in the case where C, A', C' are replaced by A. It means that, for a given reference object A, it is equivalent to compare the resemblance of any object B with it to the resemblance of any object B' having the same difference with A, for a given information common to A and B, to A and B'.

The solvability property means that, if B resembles A more than B', there exist two objects P and Q, with the same common information as A and B, with the same difference as B' and A, providing a measure of resemblance in the interval $]R(A, B'), R(A, B)[$.

4 Examples of measures of resemblance

The following quantities provide examples of measures of resemblance :

- the possibility of B with respect to A [4] : $S(A, B) = \sup_x \min(f_A(x), f_B(x))$ is a measure of resemblance for normalized fuzzy sets, with the measure of information M_2, whatever the difference may be.

- $R(A, B) = 1 - \sup_{f_A(x)=0} f_B(x)$ [1], is a measure of resemblance for normalized fuzzy sets, with M_2 and the difference defined by f^1.

- $R(A, B) = \inf_x \min(1 - f_B(x) + f_A(x), 1)$ [2] is a measure of resemblance for normalized fuzzy sets, with M_2 and the difference defined by f^1.

- $R(A, B) = M(A \cap B)/M(B)$ is a measure of resemblance, with the measure of information M_1 or M_3 and the difference defined by f^1.

5 Aggregation of resemblances

Let Ω be a set of objects described by means of attributes defined on sets Ω_1, Ω_2... Ω_n. An object O is associated with values $A_1, A_2, ...A_n$ of the attributes, respectively defined as fuzzy subset of Ω_1, Ω_2... Ω_n. Let us consider another object O' associated with $B_1, B_2, ...B_n$. We say that an object O is included in an object O' if and only if A_i is included in B_i for every i, the intersection of O and O' is defined as the object $O \cap O'$ with values of attributes $A_i \cap B_i$, and the difference between objects O and O' is defined as the object $O - O'$ with values of attributes $A_i - B_i$.

We need to give a general degree of resemblance S defined on the set of objects Ω, satisfying properties analogous to those required from each R_i, since they are natural characterizations for the evaluation of resemblances. We prove the following :

Property 4 *Let R_i be measures of resemblance on Ω_i, for $1 \leq i \leq n$. For any triangular norm T, the measure defined for any pair (O, O') of objects of Ω by $S(O, O') = T(R(A_1, B_1), ..., R(A_n, B_n)))$ satisfies the following properties : MR1, MR2, MR7 and MR8.*

Let us now consider the OWA operators [7] defined by : $W = \{w_1, ..., w_n\}$ such that $w_i \in [0, 1]$ $\forall i, \sum_{i=1}^{n} w_i = 1$ and :

$$\begin{aligned} h_W : [0,1]^n &\rightarrow [0,1] \\ h_W(x_1, ..., x_n) &= \sum_{i=1}^{i=n} w_i b_i \end{aligned} \tag{7}$$

where b_i is the i-th biggest value among $x_1, ..., x_n$. They can be used to aggregate measures of resemblance.

Property 5 *Let R_i be measures of resemblance on Ω_i, for $1 \leq i \leq n$. For any OWA-operator h_W, the measure defined for any pair (O, O') of objects of Ω by :*

$$S(O, O') = \sum_{i=1}^{i=n} w_i b_i \tag{8}$$

with b_i the i-th biggest value of $R(A_i, B_i)$, satisfies the following property : MR1, MR2, MR7, MR8.

Proof

MR1 and MR2 are obviously satisfied by the aggregated measure of resemblance by the way of OWA operators.

For the third property MR7, let us consider 3 objects O, O', O'' respectively associated with values of the i-th attribute A_i, B_i, B'_i, $1 \leq i \leq n$. We have to prove that : if $O \cap O' = O \cap O''$ then $O - O' = O - O''$ entails $S(O, O') = \sum_{i=1}^{i=n} w_i b_i \leq S(O, O'') = \sum_{i=1}^{i=n} w_i b'_i$ where b_i is the i-th biggest value of $R(A_i, B_i)$ and b'_i is the i-th biggest value of $R(A_i, B'_i)$, $i = 1, ..., n$. Then, because of our definitions of intersection, difference and inclusion between two objects O and O', we have that : $\forall i$, $A_i \cap B_i = A_i \cap B'_i$, and that $B_i - A_i \supseteq B'_i - A_i$. Thus, because of MR7, we have : $\forall i, R(A_i, B_i) \leq R(A_i, B'_i)$. In order to prove that : $S(O, O') \leq S(O, O'')$, let us prove that $b_i \leq b'_i$, $\forall i$.

Let us consider the ordered list of $\{x_i\}$. Suppose $x_k > x'_p$, where x'_p and x_k are the k-th biggest value. As $x'_i \geq x_i$ $\forall i$, it is necessary that :

$x'_p \leq \min_{i=1,...,k} x'_i$. As x'_p is the k-th biggest value, $x'_p \in x'_i, i = 1, ..., k$, then there exists i such that : $x'_p \geq x'_i \geq x_k$ which is in contradiction with the hypothesis. Hence, it is not possible to have : $w_k.x_k > w_k.x'_p \quad \forall k$.

The proof of MR8 is analogous to the proof of MR7.

6 Conclusion

In the proposed class of measures of resemblances, we do not take into consideration any degree of inclusion: there is no distinction between two elements B and B' included in A because we do not take into account features shared by A and not by B (involved in $A - B$). We shall consider this in a forthcoming paper in order to present a more general framework.

Transitivity and symetry are not required by our axioms but these properties are not excluded. For instance, the third example given in section 4 is T-transitive with regard to the Lukasiewicz triangular norm T. The first example satisfies the property of symetry.

Measures of resemblance satisfying the proposed properties will allow to study problems of typicality and to construct a prototype for a class of objects. The resemblance between a prototype and other elements is evaluated by means of such a measure and a class is built with elements resembling the prototype.

References

[1] B. Bouchon-Meunier. Fuzzy similitude and approximate reasoning. In P. P. Wang, editor, *Advances in Fuzzy Theory and Technology*, pages 161–166. Bookwrights Press, 1993.

[2] B. Bouchon-Meunier and L. Valverde. Analogy relations and inference. In *Proceedings of* 2^{nd} *IEEE International Conference on Fuzzy Systems*, pages 1140–1144, San Fransisco, 1993.

[3] A. De Luca and S. Termini. A definition of a non-probability entropy in the setting of fuzzy sets theory. *Information and Control*, 20:301–312, 1972.

[4] D. Dubois and H. Prade. Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences*, 30:183–224, 1983.

[5] M. Sugeno. *Theory of Fuzzy Integral and its Application.* PhD thesis, Tokyo Institute of Technology, Japan, 1974.

[6] A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.

[7] R.R Yager. On ordered weighted averaging operators on multicriteria decision-making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1988.

[8] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[9] L A. Zadeh. Calculus of fuzzy restrictions. In L. A. Zadeh, K. Fu, K. Tanaka, and M. Shimura, editors, *Fuzzy sets and their applications to cognitive and decision processes.* Academic press, New-York, 1975.

[10] R. Zwick, E. Carlstein, and D. V. Budescu. Measures of similarity among fuzzy concepts : a comparative analysis. *International Journal of Approximate Reasoning*, 1(2):221–242, 1987.