

Constructing prototypes from large databases

Maria Rifqi

LAFORIA-IBP

Université P. et M. Curie

Case 169, 4 place Jussieu

75252 Paris Cedex 05, France.

E-mail : rifqi@laforia.ibp.fr

Abstract

This paper illustrates the use of the model, established in previous papers, enabling to construct different kinds of measures of comparison according to wished properties, as well as to classify the existing measures according to their properties. So, this model helps to choose (or to construct) an appropriate measure for the considered problem. This paper especially looks into the list of measures of comparison in the construction of fuzzy prototypes and then in prototype-based reasoning.

INTRODUCTION

The comparison of descriptions of objects is a usual operation in many domains: psychology, analogy, physical sciences, image processing, clustering, deductive reasoning, case-based reasoning... Comparisons are accomplished because it is thought that if a new problem is sufficiently similar to a known problem then a knowledge transfer from the second one to the first one is possible in order to solve the new problem.

It comes naturally to human agents to reason by analogy, by using prototypes or by reference to past experiences... All these approaches are based on the comparisons of objects. In this paper, we choose, to solve new cases by constructing a prototype - a case which synthesizes all the known cases - and by comparing new cases to prototypes.

We consider that the descriptions of cases are imprecise and inaccurate. We choose to use framework of fuzzy sets to represent them. Our cases are described by means of pairs (attribute, value), when the value is a fuzzy set of the universe of the attribute.

One of the purpose of this paper is to show that it is of great use to work with a frame which takes into consideration the differences among measures of comparison. It was shown [2], [3] that these measures differ from one another because of the application it serves. This frame enables to choose the appropriate type of measures of comparison: their properties reflect the requirements of the application.

Oncoming case-based reasoning, prototype-based reasoning needs to compare objects. The creation of prototypes is analyzed in detail. Before that, we explain our choice of this kind of reasoning and we recall the framework of measures of comparison.

1 REASONING FROM FUZZY PROTOTYPES

Systems using prototype-based reasoning, as well as case-based reasoning, have the advantage not to need to express knowledge under rule form. The difficulty to extract rules from the expert is well known. This problem does not occur in case-based or prototype-based reasoning: knowledge is represented by means of known cases.

The principle of prototype-based reasoning resembles case-based reasoning. As one can see in the description done in introduction, the comparison

in prototype-based reasoning is realized between a case (or an example, or an observation...) and a prototype constructed from old cases (known cases), whereas the comparison is direct in case-based reasoning: the step of creation of prototype does not exist.

One of the interests of prototype-based reasoning is to work with a ‘super case’ which contains the essential information available from cases of a same ‘class’ (or solution...). This involves a gain in the number of comparisons to realize. This number is equal to the number of prototypes instead of the number of cases. Moreover, the addition of a new case is not expensive: the new information it contains is integrated in the prototype and the number of comparisons is not modified.

Such reasoning is relevant when it is easy to have a large set of cases. For example, a bank needs to know if such client is likely to be interested by some of its services. This bank has a large database concerning its clients classified by services that they pay for. In this situation, it is easier and more natural to compare a profile of client with a kind of prototype than with all the clients concerned by the service.

2 THE FRAMEWORK OF MEASURES OF COMPARISON

The framework we used to deal with measures of comparison was described in detail in [3]. It proposes a classification of measures enabling to compare fuzzy characterizations of objects, according to their properties and the purpose of their utilization. It establishes the difference between measures of satisfiability, resemblance, inclusion and dissimilarity. The study is based on concepts analogous to those developed by A. Tversky [18] for his general work on similarities.

2.1 DEFINITIONS

For any set Ω of elements, let $F(\Omega)$ denote the set of fuzzy subsets of Ω , f_A the membership function of any description A in $F(\Omega)$. We use the classical definition of intersection: $f_{A \cap B} = \min(f_A, f_B)$ to describe the elements belonging to A and B . We suppose that we are given a means of evaluat-

ing the weight of the elements of the universe [6] characterized by a fuzzy set through a fuzzy set measure M .

Definition 1 *A fuzzy set measure M is a mapping $: F(\Omega) \rightarrow \mathbb{R}^+$ such that, for every A and B in $F(\Omega)$:*

- $M(\emptyset) = 0$
- if $B \subseteq A$, then $M(B) \leq M(A)$.

First of all, a measure of comparison between A and B takes into account common features, $A \cap B$, and distinctive features, $B - A$ and $A - B$. An M -measure of comparison, more generally, depends on: $M(A \cap B)$, $M(B - A)$ and $M(A - B)$. M is a fuzzy set measure, as for instance the sigma-count of a fuzzy subset.

Definition 2 *An M -measure of comparison on Ω is a mapping $S : F(\Omega) \times F(\Omega) \rightarrow [0, 1]$ such that $S(A, B) = F_S(M(A \cap B), M(B - A), M(A - B))$, for a given mapping $F_S : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow [0, 1]$ and a fuzzy set measure M on Ω .*

An M -measure of comparison can either evaluate the likeliness of two descriptions, it is called an M -measure of similitude, or their differences, it is then called an M -measure of dissimilarity.

Definition 3 *An M -measure of similitude S on Ω is an M -measure of comparison S such that $F_S(u, v, w)$ is non decreasing in u , non increasing in v and w .*

The measure of dissimilarity is not defined as the ‘complementary’ of the measure of similitude but it has specific properties.

Definition 4 *An M -measure of dissimilarity S on Ω is an M -measure of comparison such that $F_S(u, v, w)$ is independent of u and increasing in v and w and satisfying the minimality property: $F_S(u, 0, 0) = 0$ or equivalently $S(A, A) = 0$.*

As it was already noticed, various measures of comparisons exist because requirements differ from an application to another one. This is the reason to distinguish three types of M -measures of similitude: satisfiability, inclusion and resemblance.

1. The *resemblance* is used for a comparison between the descriptions of two objects, of the same level of generality, to decide if they have many common characteristics.

Definition 5 An M -measure of resemblance on Ω is an M -measure of similitude S which satisfies properties of reflexivity: $F_S(u, 0, 0) = 1$ or equivalently $S(A, A) = 1$ and symmetry: $F_S(u, v, w) = F_S(u, w, v)$ or equivalently $S(A, B) = S(B, A)$.

2. The *satisfiability* corresponds to a situation in which we consider a reference object or class and we decide if a new object is compatible with it or satisfies the reference.

Definition 6 An M -measure of satisfiability S on Ω is an M -measure of similitude S such that $F_S(u, v, w)$ is independent of w and satisfies the containment property: $F_S(u, 0, w) = 1$ whatever $u \neq 0$ and w may be or equivalently $S(A, B) = 1$ if $B \subseteq A$ and the exclusiveness property: $F_S(0, v, w) = 0$ whatever v and w may be or equivalently $S(A, B) = 0$ as soon as $A \cap B = \emptyset$.

2.2 INTERPRETATIONS AND EXAMPLES

The generalized and normalized distances [9], [12], [11], [7] for fuzzy subsets are examples of M -measure of dissimilarity. Indeed, asymmetrical M -measure of dissimilarity satisfying the triangular inequality is a distance.

The resemblance measure is the basis of similarity logic [15], [5]. It is also relevant when all the objects under study can be described in a same way. This situation occurs naturally in a case-based reasoning system at the step when it compares cases with another one in order to infer the solution of the most similar to the new one; or in prototype-based reasoning system, at the step when it also compares cases but in order to construct a prototype. M -measures of resemblance S which satisfy an additional property of T -transitivity are indistinguishability relations [17]. In the case where T is the minimum, we obtain measures of similarity.

Lastly, the satisfiability takes place in a situation where a particular object is a leading object. This situation is typical in prototype-based reasoning, where the references are prototypes and a new object must be associated with one of them. Analogy relations [4] and fuzzy similitudes [1] are particular M -measures of satisfiability.

To sum up, it is important to know if all the data available have the same level of generality, if there is some objects that are considered as references, in order to decide what kind of measures of comparison is the most relevant for the application among satisfiability, resemblance or dissimilarity measures.

3 CONSTRUCTION OF FUZZY PROTOTYPES

3.1 PROTOTYPICALITY

Especially in psychology, Eleanor Rosch's study [14] on categorization and prototyping is a reference. It establishes that categories are structured according to a *resemblance family*. A. Tversky [18] has analyzed typicality relations in compatible terms with his "contrast model" and with E. Rosch's study.

Let $P(a, C_i)$ denote the degree of prototypicality of object a , represented by a set of attributes A , with respect to class C_i , with cardinality n . He defines P by:

$$P(a, C_i) = p_n (\lambda \Sigma f(A \cap B) - \Sigma (f(A - B) + f(B - A)))$$

where the summations are over all attributes of the object b in the class C_i , p_n reflects the effect of category size on prototypicality and λ determines the relative weights of the common and distinctive features. An element a of C_i is a prototype if it maximizes $P(a, C_i)$. He also called our attention to the fact that a class may have more than one prototype. With some particular conditions, $P(a, C_i) = \frac{1}{n} \Sigma S(A, B)$, where S is a particular similarity relation of his contrast model: $S(A, B) = \lambda f(A \cap B) - f(A - B) - f(B - A)$.

L. A. Zadeh [20] has also looked into the prototype theory in integrating into it what he consid-

ers as intrinsic to the concept of prototype: the notion of fuzziness and *opacity*. According to his interpretation:

1. a prototype is not unique.
2. the typicality is a question of degree
3. the concept of prototype is opaque : we are not able to define exactly how we compute a degree of typicality.

Zadeh considers a population \mathcal{C} of objects not necessarily distinct a_1, \dots, a_n . In this way, \mathcal{C} is represented as a fuzzy multi-set:

$$\mathcal{C} = \mu_1/(m_1 \times a_1) + \dots + \mu_n/(m_n \times a_n) \quad (1)$$

where μ_i represents the membership degree of a_i in \mathcal{C} and m_i its multiplicity.

The principle to obtain a fuzzy prototype of a population is to stratify \mathcal{C} in grouping objects sharing nearing the same membership degree. For instance:

$$\mathcal{C} = \text{High}/\mathcal{C}_{\text{good}} + \text{Medium}/\mathcal{C}_{\text{borderline}} + \text{Low}/\mathcal{C}_{\text{poor}}$$

where $\mathcal{C}_{\text{good}}$, $\mathcal{C}_{\text{borderline}}$ et $\mathcal{C}_{\text{poor}}$ are multi-sets of good, borderline and poor elements respectively and High, Medium and Poor are fuzzy numbers which represent the corresponding membership degrees. Each level of stratification of \mathcal{C} has its fuzzy prototype obtained thanks to the algorithm of compactification [19] inspired from the Quine-McCluskey algorithm [10]. In iterating this process, we obtain for each level of stratification an object maximally summarized which can be viewed as a fuzzy prototype. For example, $PT(\mathcal{C}_{\text{good}})$ is the fuzzy prototype of $\mathcal{C}_{\text{good}}$, etc. Finally, we obtain the fuzzy prototype of \mathcal{C} :

$$\begin{aligned} PT(\mathcal{C}) &= \text{High}/PT(\mathcal{C}_{\text{good}}) \\ &+ \text{Medium}/PT(\mathcal{C}_{\text{borderline}}) \\ &+ \text{Low}/PT(\mathcal{C}_{\text{poor}}) \end{aligned}$$

3.2 EXTENSION OF THE METHOD

Let X be a set of objects. We suppose that there exists a partition given on X composed by crisp classes \mathcal{C}_j . In the example of a bank's client

database, a class is the set of clients who have subscribed or bought a particular service or product offered by this bank. Let us denote E_k a case of the database and A_i an attribute.

Our aim is to use the general framework of measures of comparison in a system based on prototypicality. For each comparison stepping in the reasoning, an appropriate measure has to be chosen among satisfiability, resemblance or dissimilarity measures.

Our approach consists in improving the principle proposed by Zadeh. In particular, one of the points which has not been explained enough is that of the degrees μ_i . The author himself proposed to interpret them as degrees of compatibility more than membership degrees.

Indeed, G. Kleiber [8], taking up C. Schmidt [16], noticed that the degree of representativeness is too often assimilated to the membership degree. If the class considered is the class of birds, the cock as the penguin belongs totally to this class. On the other hand, the degree of representativeness (or prototypicality) of the cock is certainly larger than the penguin's degree.

In agreement with this remark, we interpret the μ_i degrees as the expression of prototypicality of the object i for its class. A fuzzy prototype is then constructed on the basis of each degree of prototypicality of each example of the class.

We consider that the degree of prototypicality of an object depends positively on its total resemblance to others objects of its class (internal resemblance) and on its total dissimilarity to objects of other classes (external dissimilarity).

Theoretically, examples are all on the same level of generality. *The measure of resemblance is then the appropriate measure for comparing an example to the other members of its class.*

3.3 ALGORITHM

The degree of prototypicality of an example E_k , described by attributes A_i , $1 \leq i \leq N$, according to its class $\mathcal{C}_{j(k)}$ can be computed in following this

algorithm:

- For each example $E_l \neq E_k$ of $C_{j(k)}$:
For A_i , $i = 1$ to N :
 Compute the resemblance
 between E_k and E_l
 according to A_i
 Aggregate the resemblance
 between E_k and E_l
- Aggregate resemblances obtained
between E_k and all the E_l with $l \neq k$
(internal resemblance)
- Repeat the same operations for each
example E_l of $C_{j(l)}$ with $C_{j(l)} \neq C_{j(k)}$ in
replacing the resemblance by the
dissimilarity (external dissimilarity)
- Compute the degree of prototypicality
as a function of the internal resemblance
and the external dissimilarity.

Each step of aggregation is important. An aggregation operator has to be chosen in such a way that the measure obtained after aggregation respects axioms required by the chosen measure of comparison. Aggregation deserves a particular attention because it is rich in consequences [13].

4 INFERENCE OF A SOLUTION

Once prototypes are obtained for each class, a new object can be compared to them. It belongs to a class C_i if it is enough close to the prototypes of this class. We have to choose, once again, an appropriate type of measures. In this particular situation, the comparison is established between an object of high level of generality (the fuzzy prototype) and an object of low level of generality (the new case). *Measures meant to this situation are the measures of satisfiability.* It is clear that the comparison is directional in this case (non symmetrical).

The prototype which resembles the most the new case infers its class.

CONCLUSION

We propose to define a prototype-based reasoning from the general framework of measures of comparison. We use different types of measures of

comparison to create prototypes. These are constructed according to two methods. The choice of the method depends on the application: either the notion of generalization of the prototype is more relevant or it is the notion of representativeness which is more appropriate. Whatever method is carried, the principle is to begin with computing the degree of prototypicality which depends on internal similitude and external dissimilarity.

In order to infer the class of the most similar prototype, we show, once again, that it is useful to distinguish different types of measures of comparison. The measures of comparison used between the new case and a prototype differ according to the chosen method.

References

- [1] B. Bouchon-Meunier (1993). Fuzzy similitude and approximate reasoning, in *Advances in Fuzzy Theory and Technology* P. P. Wang, editor. (Bookwrights Press), pp. 161–166.
- [2] B. Bouchon-Meunier & M. Rifqi (1995). Resemblance in database utilization, in *Proceedings of 6th IFSA World Congress*. (Sao Paulo).
- [3] B. Bouchon-Meunier, M. Rifqi, & S. Bothorel (1996). Towards general measures of comparison of objects. *Fuzzy Sets and Systems, to appear*.
- [4] B. Bouchon-Meunier & L. Valverde (1993). Analogy relations and inference, in *Proceedings of 2nd IEEE International Conference on Fuzzy Systems*. (San Francisco), pp. 1140–1144.
- [5] F. Esteva, P. García, L. Godó, E. Ruspini, & L. Valverde (1994). On similarity logic and the generalized modus ponens, in *Proceedings of FUZZ-IEEE'94*. (Orlando).
- [6] J. Gasos & A. L. Ralescu (1995). The direct and inverse problem of the matching of fuzzy sets depending on fuzziness. *Kybernetes* **24**:9–17.
- [7] A. Kaufmann (1973). *Introduction à la théorie des sous-ensembles flous*. (Masson) Vol. 3.

- [8] G. Kleiber (1991). Prototype et prototypes, in *Sémantique et cognition*. (Editions du C.N.R.S., Paris).
- [9] L. T. Kóczy & K. Hirota (1993). Ordering, distance and closeness of fuzzy sets. *Fuzzy Sets and Systems* **59**:281–293.
- [10] E. J. McCluskey (1965). *An introduction to the theory of switching circuits*. (McGraw-Hill, New-York).
- [11] M. Nowakowska (1977). Methodological problems of measurement of fuzzy concepts in the social sciences. *Behavioral Science* **22**:107–115.
- [12] A. L. Ralescu & D. A. Ralescu (1984). Probability and fuzziness. *Information Sciences* **34**:85–92.
- [13] M. Rifqi (1995). Mesures de similitude et leur agrégation, in *Rencontres sur la Logique Floue et ses Applications, LFA'95*. (Paris), pp. 80–87.
- [14] E. Rosch (1978). Principles of categorization, in *Cognition and categorization* E. Rosch & B. B. Lloyd, editors. (Hillsdale, N. J. : Lawrence Erlbaum Associates), pp. 27–48.
- [15] E. H. Ruspini (1991). On the semantics of fuzzy logic. *International Journal of Approximate Reasoning* **5**:45–88.
- [16] C. Schmidt (1974). The relevance to semantic theory of a study of vagueness, in *10th regional meeting of the Chicago Linguistic Society*. pp. 617–630.
- [17] E. Trillas & L. Valverde (1984). On implication and indistinguishability in the setting of fuzzy logic, in *Management Decision Support Systems Using Fuzzy Sets and Possibility Theory* J. Kacprzyk & R. R. Yager, editors. (Verlag TUV, Rheinland).
- [18] A. Tversky (1977). Features of similarity. *Psychological Review* **84**:327–352.
- [19] L. A. Zadeh (1976). A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man-Machine Studies*.
- [20] L. A. Zadeh (1982). A note on prototype theory and fuzzy sets. *Cognition* **12**:291–297.