Case-based reasonning: a fuzzy approach

Didier Dubois, Francesc Esteva, Pere Garcia, Lluís Godo, Ramon López de Màntaras, Henri Prade

Case-Based Reasoning: A Fuzzy Approach

Didier DUBOIS*, Francesc ESTEVA**, Pere GARCIA**, Lluís GODO**,
Ramon LÓPEZ DE MÀNTARAS** and Henri PRADE*

* Institut de Recherche en Informatique de Toulouse (IRIT), Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cédex 4, France

** Institut d'Investigacio en Intel.ligencia Artificial (IIIA), Consell Superior d'Investigacions Cientifiques (CSIC), Campus Universitat Autonoma de Barcelona, 08193 Bellaterra, Spain

Abstract. This paper is an attempt at providing a fuzzy set-based formalization of case-based reasoning. The proposed approach assumes a principle stating that "the more similar are the problem description attributes, the more similar are the outcome attributes". If this principle is accepted it induces constraints on the fuzzy similarity relations which are acceptable with respect to the cases stored in the memory. The idea of having cases in the memory with different levels of typicality is also discussed. A weaker form of this principle concluding only on the graded possibility of the similarity of the outcome attributes, is also considered. These two forms of the case-based reasoning principle are modelled in terms of fuzzy rules. Then an approximate reasoning machinery taking advantage of this principle enables us to apply the information stored in the memory of previous cases to the current problem. Extensions of the proposed approach in order to handle incomplete or fuzzy descriptions is also considered and studied. The paper does not take into account the learning aspects of case-based reasoning.

1. Introduction

Case-based reasoning (CBR) systems (Kolodner, 1993; Aamodt and Plaza, 1994) are a particular type of analogical reasoning systems which nowadays have an increasing number of applications in different fields and specialized software products. As it is known, the goal of CBR is to infer a solution for a current case from solutions of a family of previously solved problems, the memory of precedent cases. Theoretical and empirical works have focused among others on the definition and elicitation of similarity measures, on retrieving the relevant cases, on extrapolating pieces of knowledge from cases in the memory, on the logical modelling of the inference mechanism and on the management of incomplete, imprecise or uncertain description of cases. Although these different steps apparently require some graded notion of similarity and some approximate reasoning capabilities, there have been rather few attempts for introducing fuzzy set-based tools in analogical reasoning until recently, up to some exceptions (Farreny and Prade, 1982; Bouchon-Meunier and Valverde, 1993). However, in case-based reasoning, some works have focused on the handling of fuzzy descriptions in the retrieval step (Salotti, 1989, 1992; Jaczynski and Trousse, 1994), on the learning of fuzzy concepts from fuzzy examples (Plaza and López de Màntaras, 1990), on the integration with rule-based reasoning (Dutta and Bonissone, 1993), and very recently on the logical modelling of the inference mechanisms based on similarity measures (Plaza et al., 1996a), in the use of fuzzy predicates for expressing preferences when computing similarities (Bonissone and Cheetman, 1997) and in the use of fuzzy rules for guiding case-based reasoning (Dubois et al., 1997b). See (Dubois and Prade, 1994) for a general overview on similarity-based approximate reasoning and for an investigation of the potentials of fuzzy logic for modeling the different steps mentioned above.

Following a paper by the same authors (Dubois et al.1997b), this work only deals with problems where cases can be given as n-tuples of completely, incompletely or fuzzily described attribute values, this set of attributes being divided in two non-empty disjoint subsets: the subset of problem description attributes and the subset of solution or outcome attributes, denoted by $\mathcal{S}$ and $\mathcal{T}$ respectively. These subsets are taken according to the problem we deal with. A *case* will be denoted as a tuple (s, t) where s and t stand for complete sets of precise attribute values of $\mathcal{S}$ and $\mathcal{T}$ respectively. In order to perform a case-based reasoning we assume that we have a finite set M of known cases or precedents, called *case base* or *memory* (M is thus a set of pairs (s,t)), and a current problem description, denoted by s_0, for which the precise values of all attributes belonging to $\mathcal{S}$ are known. Then case-based reasoning aims at extrapolating or estimating the value t_0 of the attributes in $\mathcal{T}$, for the current problem.

Case-based reasoning, in general, assumes the following implicit principle: "*similar situations give (or may give) similar outcomes*". Thus, a similarity relation S between problem descriptions and a similarity measure T between outcomes are needed. In this paper we briefly discuss how to get them. In terms of the relations S and T this implicit CBR-principle can be expressed as,

"*the more similar are the problem description attributes in the sense of S,*
the more similar are the outcome attributes in the sense of T"

and will be modelled throughout this paper in the framework of fuzzy rules. A key idea is that each case in M, together with the CBR-principle, induces a fuzzy gradual rule. That is, for each (s,t) in M and a current case s_0, we have the following basic reasoning pattern. From the gradual rule "the closer a problem description to s, the closer its solution to t", and from the observation s_0, we conclude that a value t_0 is possible for the current case if it is at least as close to t as s_0 is close to s.

Throughout this paper we will refer to what we call deterministic case-based problems when the above principle is applicable, otherwise we will refer to non-deterministic problems (where only a weaker form of the principle, concluding only on the *possibility* that the outcome attributes are similar, can be used).

Section 2 provides a refresher on similarities. Section 3 and 4 describe the fuzzy set modelling of deterministic and non-deterministic CBR problems respectively for completely described cases. Section 5 generalizes the two above models to deal with incompletely described cases. We use a running example for illustrative purposes.

2. Background on fuzzy similarity relations

In CBR, the evaluation of the similarity between cases is a crucial matter. In this paper we model the notion of similarity in a general sense by means of fuzzy relations. Given a universe U of possible values for a feature, a fuzzy relation S on U is a mapping S: UxU $\to [0,1]$[1] and the properties usually required for our purpose of similarity modelling are the following (Zadeh,1971):

(i) $\forall$ u $\in$ U, S(u,u) = 1, *(reflexivity)*

(ii) $\forall$ u,v $\in$ U, S(u,v) = S(v,u), *(symmetry)*

(iii) $\forall$ u,v,w $\in$ U, S(u,v) $\otimes$ S(v,w) $\leq$ S(u,w), *($\otimes$-transitivity)*

where $\otimes$ is usually a t-norm operation[2]. While *reflexivity* and *symmetry* are minimum properties that are clearly required when evaluating the closeness of cases, *transitivity*, even graded, does not always seem compulsory. Indeed, in case-based reasoning, given a current situation s_0, in order to retrieve the most similar cases from the memory M, we estimate the similarity of s_0 with each of the situations s_i in M pairwisely, and we do not compute the similarity between s_0 and s_j by transitivity from the similarity of s_0 and s_i and that of s_i and s_j. Transitivity may even be thought to be an undesirable property in some settings, since s_0 may be somewhat intermediary between two situations which are not close themselves. Moreover we may also require the *separating* property ((i') $\forall$ u,v $\in$ U, S(u,v) = 1 if and only if u = v). Besides, on ordered universes we shall also require that the similarity be *convex*, namely for all x,y,u,v $\in$ U such that [x,y] $\supseteq$ [u,v], then S(x,y) $\leq$ S(u,v).

Example 1. Many similarity measures which are often used in CBR make a rather neat distinction between those elements which are considered to be similar from those which are considered dissimilar by means of some threshold. As an extreme case consider, on the set of real numbers, the non-fuzzy similarity relation S_ε defined as

$$S_\varepsilon(x, y) = \begin{cases} 1, & \text{if } |x-y| \leq \varepsilon \\ 0, & \text{otherwise} \end{cases}.$$

This type of relation is clearly not transitive. But even if a relation is $\otimes$-transitive for some operation $\otimes$, the corresponding notion of extended transitivity can be very different. For instance, in case $\otimes$ is the Lukasiewicz t-norm, $\otimes$-transitivity does not restrict at all the value of S(x,z) as soon as S(x,y) + S(y,z) $\leq$ 1 for all y. The situation in that sense would be very different if the relation is min-transitive, which is much stronger.

The more general relations we shall consider for CBR-problems are fuzzy relations satisfying (i) and (ii), which are called *proximity* relations. If a proximity relation further satisfies (iii) w.r.t. a t-norm $\otimes$, it is called a $\otimes$-*similarity* relation. In both cases we say that

---

[1]In this paper, we use the notations $S(s_1,s_2)$ and $T(t_1,t_2)$ for denoting the degrees of similarity for simplicity, rather than using the notations $\mu_S(s_1,s_2)$ and $\mu_T(t_1,t_2)$ commonly used in the fuzzy set literature where they distinguish between a fuzzy set F and its membership function μ_F. Thus, in this paper we shall write F(u) instead of $\mu_F(u)$.

[2]A t-norm $\otimes$ is a non-decreasing binary operation on [0,1] satisfying associativity, commutativity, 1 being the neutral element and 0 being an absorvent element. Noticeable t-norms are min, product and Lukasiewicz operation (a $\otimes$ b = max(0, a + b – 1)).

the relation is *separating* if it satisfies (i'). Nevertheless, for the sake of simplicity, from now on we will generally use the term *similarity* to denote both proximity and $\otimes$-similarity relations whenever no further precision is required. See (Ovchinnikov, 1991) for a good overview on similarity relations.

Usually, the global evaluation of the similarity between two multiple-feature descriptions is obtained by aggregating degrees of similarities for each feature. The aggregation has to be done in such a way that the resulting similarity relation should preserve properties, like reflexivity, symmetry (and possibly $\otimes$-transitivity), of the individual similarities. In general min-combination preserves reflexivity and symmetry. With respect to transitivity, it is worth noticing that if S_1 and S_2 are $\otimes$-transitive then both $S_\otimes((x_1, x_2), (y_1, y_2)) = S_1(x_1, y_1) \otimes S_2(x_2, y_2)$ and $S_{min}((x_1, x_2), (y_1, y_2)) = \min(S_1(x_1, y_1), S_2(x_2, y_2))$ are still $\otimes$-transitive. It is clear that the min-combination gives the minimum value to the similarity, and thus keeps the most discriminating value as a global value of the similarity, while the max-combination would give as a value the least discriminating one. Other well-known combination operations like averages, take values between min and max (but do not usually preserve transitivity).

Moreover we may think of a weighted aggregation if we consider that we are dealing with a fuzzy set of features having different levels of importance. For instance, if we aggregate the similarity degrees by means of the min operation, a weighted version (e.g., Dubois and Prade, 1988) can be defined by

$$S(x, y) = \min_{i=1,n} \max(S_i(x_i, y_i), 1 - \lambda_i)$$

with $\max_{i=1,n} \lambda_i = 1$, where case x (resp. y) is described by the vector of feature values $(x_1, \ldots, x_n)$ (resp. $(y_1, \ldots, y_n)$) and λ_i is the level of importance of the i^{th} feature. Clearly $\lambda_i = 1$ means that the feature is fully important for the assessment of the global similarity, while if $\lambda_i = 0$, the feature is not taken into account. An easy computation shows that reflexivity, symmetry and $\otimes$-transitivity are preserved by this weighted aggregation, as pointed out by Fodor and Roubens(1994) who also suggest other weighted aggregations of fuzzy relations. Moreover separatingness is also preserved if $\lambda_i \neq 0$ for all $i = 1, 2, \ldots, n$.

In the following example, and in the next sections, different proximity and similarity relations are presented and used.

Example 2. Suppose we have a data base about second-hand cars. Suppose also that every car has exactly 6 attributes in our data base and suppose that the problem description attributes are the first five, i.e., $\mathcal{S}$ = {*year, power, mileage, equipment, shape*}, and the outcome attribute is the last one, i.e. $\mathcal{T}$= {*price*}. The ranges of *year, power, mileage* and *price* are numerical and the range of the *equipment* and *shape* attributes consists of 4 qualitative levels, linearly ordered: "bad"< "poor"< "good" < "excellent". The global similarity S will be defined by aggregation of similarities S^i (i = 1, ..., 5) for each description attribute. S^1 and S^3, corresponding to *year* and *mileage*, are defined by,

$$S^i(u,v) = \min(u,v) / \max(u,v).$$

This seems to be a reasonable similarity measure for these attributes because it is not uniform, that is, the greater are two values having the same difference, the more similar they are. Here the similarity relative to the power (S^2) is assumed to be linear w.r.t. the difference of power. For S^2 we use

$$S^2(u,v) = 1 - (|u - v| / 1000).$$

Finally, the similarities S^4 and S^5, between the elements of the qualitative range are taken as,

$$S^i(u,v) = \begin{cases} 1, & \text{if } u = v \\ 2/3, & \text{if u and v are consecutive w.r.t. the attribute order} \\ 1/3, & \text{if there is exactly one element between u and v} \\ 0, & \text{otherwise} \end{cases}$$

Obviously all these S^i are reflexive and symmetric and thus they are proximity relations. Just for the sake of illustration, in Table 1 we present an small memory of cases (second hand cars), together with a car (case $C_0 = (s_0, t_0)$, with $s_0 = (s_0{}^1, s_0{}^2, s_0{}^3, s_0{}^4, s_0{}^5)$) whose market value is to be estimated.

cases (s, t)	years old (s^1)	power (s^2)	mileage (s^3)	equipement (s^4)	shape (s^5)	price (t)
C_1	1	1.300	20.000	poor	good	8.000
C_2	2	1.600	30.000	excellent	poor	7.000
C_3	2	1.600	40.000	good	good	5.000
C_4	3	1.500	60.000	excellent	poor	5.000
C_0	2	1.600	50.000	poor	good	?

Table 1. Cases of the second hand cars example.

An important comment is in order here. The aggregation of the S^i 's presupposes that they are commensurate. In practice it means that an expert should define meaningful fuzzy proximity relation, using the same scale, namely [0, 1]. Otherwise, different numerical encodings of the proximity relations on each attribute domain (maintaining the same total orderings) can lead, whatever the chosen aggregation function is, to different global proximity relations. Once the S^i 's are defined we can use different conjunctive aggregation functions, such as min and product, for obtaining the global similarity. For simplicity in this example we are not introducing any importance weighting in the aggregation although it may be natural to do it in such an example.

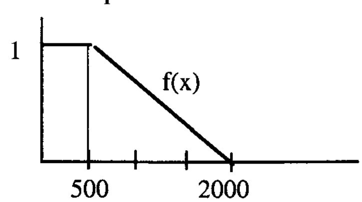

Fig. 1

Besides, only for illustrative purposes, a possible definition for T is $T(u, v) = f(|u-v|)$, f(x) being the function depicted in Figure 1. In the next section, in the deterministic case, we will discuss the problem of coherence of the memory and of the gradual rules obtained from cases through the CBR-principle. We will also discuss which constraints T has to satisfy if we require that coherence be assured when M and S are given.

3. Deterministic CBR Problems

In the deterministic setting, for each case (s_i,t_i) of the memory M, the principle assumed to hold is that "the more similar is s_i to the input s_0, the more similar is t_i to the outcome". We also assume that a similarity S on the set of problem description attribute values U (domain of $\mathcal{S}$), and a similarity T on the set of solution attribute values V (domain of $\mathcal{T}$) are available. In terms of the similarities S and T, the principle can be interpreted as the fuzzy gradual rule:

the more X is $S(s_i)$, the more Y is $T(t_i)$,

where X and Y are variables ranging on $\mathcal{S}$ and $\mathcal{T}$ respectively, $S(s_i)$: $U \to [0, 1]$ and $T(t_i)$: $V \to [0, 1]$ are fuzzy sets defined as $S(s_i)(s) = S(s_i, s)$ and $T(t_i)(t) = T(t_i, t)$ respectively. The semantics of such rules is given by the following constraint on the joint possibility distribution $\pi_{X,Y}$ (Dubois and Prade, 1992):

$$\pi_{X,Y}(s, t) = S(s_i, s) \to T(t_i, t)$$

where $\to$ denotes the residuated implication such as $x \to y = 1$ if $x \leq y$, and $x \to y = y$ otherwise (Gödel implication).

It is worth noticing that the gradual rule "the more X is $S(s_i)$, the more Y is $T(t_i)$" means that the similarity of any s with s_i constrains the similarity of t_i with any t associated with s at a minimum level, i.e., $S(s_i,s)$ is a lower bound of $T(t_i,t)$. Thus, $\forall\ \alpha \in [0,1]$, we have

$$s \in S(s_i)_\alpha \Rightarrow t \in T(t_i)_\alpha \tag{1}$$

where $S(s_i)_\alpha = \{(s_i,s') \mid S(s_i,s') \geq \alpha\}$ is the α-cut of $S(s_i)$ and $T(t_i)_\alpha$ is similarly defined. In particular, if $S(s_i,s) = 1$, we should have $T(t_i,t) = 1$. Moreover, if T is such that $T(t_i,t_j) = 1 \Leftrightarrow t_i = t_j$ (separating property of T), then the classical functional dependency

$$s_i = s \Rightarrow t_i = t, \tag{2}$$

is a consequence of (1) using the reflexivity of S. Constraint (1) is then clearly stronger than (2). Obviously, when $S(s_i,s) = 0$, $T(t_i,t)$ is no longer constrained.

Given an input value $s = s_0$, the solution is a fuzzy set F_i with membership grades $F_i(s_0)(t) = S(s_i,s_0) \to T(t_i,t)$. In the following we only use the core of this fuzzy set, i.e., the elements t with membership 1. Thus the solution set for Y is:

$$E_i(s_0) = \{t \mid T(t_i,t) \geq S(s_i,s_0)\},$$

i.e. the set of t's such that the implication $S(s_i,s_0) \to T(t_i,t)$ takes the value 1. This amounts to take the "crisp" implication $x \to^* y = 1$ if $x \leq y$, $x \to^* y = 0$ otherwise, instead of $\to$. If the memory M contains n cases, then we have n fuzzy rules and the core of the fuzzy set solution for an input value s_0 is the intersection of sets $E_i(s_0)$, that is,

$$E(s_0) = \bigcap\nolimits_{i=1,\ldots,n} \{t \mid T(t_i,t) \geq S(s_i,s_0)\}. \tag{3}$$

Notice that $E(s_0)$ can be empty if T is not permissive enough. This set is not empty for any input s_0 if the family of gradual rules "the more X is $S(s_i)$, the more Y is $T(t_i)$ " for (s_i,t_i) in M is coherent in the sense of (Dubois et al., 1996). This coherence means that the family of crisp rules of the form (1) for i=1,n and for $\alpha \in [0, 1]$ is coherent, which is equivalent to say that if the condition parts of a subset of rules in this family have a non-empty intersection, then their conclusion parts should also intersect. Formally speaking, it means that for any I contained in $\{1,2,\ldots,n\}$ and any $\alpha_i \in (0,1]$ for $i \in I$, it holds

$$\bigcap\nolimits_{i\in I} S(s_i)_{\alpha_i} \neq \emptyset \Rightarrow \bigcap\nolimits_{i\in I} T(t_i)_{\alpha_i} \neq \emptyset\ . \tag{4}$$

In summary, $E(s_0)$ is not empty for any s_0 if and only if condition (4) is satisfied.

Besides, if we want to be coherent with respect to the memory M, the solution $E(s_j)$ for each s_j appearing in M has to contain t_j. This condition is equivalently expressed by the constraint

$$\forall (s_i,t_i),(s_j,t_j) \in M,\ S(s_i,s_j) \leq T(t_i,t_j). \tag{5}$$

So (5) expresses that when s_i and s_j are close, t_i and t_j should be at least as close.

Only if T is separating does (4) imply (5). Indeed if T is separating and (4) is satisfied, $E_i(s_i) = \{t_i\}$ and then from (3) condition (5) follows. But in general (4) does not imply (5) as it is shown in the following example. Let $M = \{(s_1,t_1), (s_2,t_2)\}$ and define two proximity relations S and T as follows: $S(s, s') = 3/4$ for any $s \neq s'$ and $T(t_1,t'') = T(t_2,t'') = 1$ for some t" different from t_1 and t_2, and $T(t, t') = 1/2$ otherwise. Obviously condition (5) fails but (4) holds (t" is in all the intersections of $T(t_i)_{\alpha_i}$ for any α_i).

On the other hand (5) does not imply (4) as it is shown in the continuation of example 2 below. Thus, once the "deterministic model" is adopted, and the similarity S is defined, the CBR principle induces constraints on the similarity relation T: both conditions (4) and (5) should be enforced when building T.

It is worth noticing that the constraints (5) are easy to handle. The interesting T is the minimal one (in the sense of fuzzy set inclusion ($T \supseteq T'$ iff $T(t,t') \geq T'(t,t')$) yielding the most informative gradual rule. Formally speaking, given M and S, the minimal solution for convex T satisfying the constraints (5) is defined by

$$T(t, t') = 1 \text{ if } t = t',\ T(t, t') = \max_{(s_i,\, t_i),(s_j,\, t_j)\in M} \{S(s_i, s_j) \mid t,t' \in [t_i, t_j]\} \text{ otherwise.} \tag{6}$$

Even if S is separating, T may not be so; it happens when M contains two cases (s,t) and (s,t') with $t \neq t'$. Moreover the transitivity of S does not entail the transitivity of T. Transitivity of T can be obtained, if necessary, taking the transitive closure.

Of course the relation T defined by (6) provides only a lower bound of the minimal T satisfying both (5) and (4). In this paper we shall not further discuss the computation of the minimal solution of this optimisation problem.

Example 2 (continued). Let us compute T_{min} and T_{prod} given by (6) with S being obtained using first the minimum and then the product aggregations:

(i) $T_{min}(t, t') = 1$ if $t = t'$, $T_{min}(t, t') = 1/2$ if $t,t' \in [5.000, 8.000]$ and $T_{min}(t, t') = 0$ otherwise.

(ii) $T_{prod}(t, t') = 1$ if $t=t'$, $T_{prod}(t, t') = 3/10$ if $t,t' \in [5.000, 7.000]$, $T_{prod}(t, t') = 7/60$ if $t,t' \in [5.000, 8.000]$ and one of t or t' does not belong to [5.000, 7.000] and $T_{prod}(t, t') = 0$ otherwise.

For an input s_0, it can be checked that the (core of the fuzzy set) solution in case of the minimum aggregation is {5.000} and the empty set in case of the product. The last result shows that (5) can hold while (4) is not satisfied. This is the reason why it is important in practice to use a pair S, T such that both (4) and (5) are satisfied.

Remark.- We may think of the following generalisation of the deterministic model. The idea is to introduce some weighting α_i associated to the cases (s_i,t_i) of M and thus, to weaken the constraints (5) in the following way:

$$\forall(s_1,t_1),(s_2,t_2) \in M,\ \min(\alpha_1, \alpha_2, S(s_1,s_2)) \le T(t_1,t_2). \qquad (7)$$

Moreover the gradual rule $S(s_i) \to T(t_i)$ is changed accordingly into $\min(\alpha_i, S(s_i)) \to T(t_i)$, which is also equivalent to $S(s_i) \to (\alpha_i \to T(t_i))$ if the implication is residuated with respect to the minimum (Gödel implication). This means that the case (s_i,t_i) is changed into the imprecise case $(s_i,T(t_i)_{\alpha_i})$. Indeed $\alpha_i \to T(t_i)(t) = 1$ if $\alpha_i \le T(t_i)(t)$. Thus the solution is,

$$E(s_0) = \bigcap_{(s_i,t_i)\in M} E_i^*(s_0),$$

where $E_i^*(s_0) = \{t' \mid \min(\alpha_i, S(s_i,s_0)) \le T(t_i,t')\}$. Notice that when $\alpha_i = 1$ we have $E_i^*(s_0) = E_i(s_0)$ and when $\alpha_i = 0$, $E_i^*(s_0)$ becomes the whole set of possible values for t and it amounts to delete the case (s_i,t_i) from the memory M.

This extension based on the introduction of degrees α_i is related to the so-called proximity entailment in (Dubois et al., 1997a; Dubois et al., 1997b). From an interpretation point of view we may think of α_i as a kind of typicality weight and the modified gradual rule can be read as

"*the more similar is s_0 to a typical case s, the more similar is t_0 to t*".

Generally speaking, memories of cases are not usually well structured, and they may contain cases that can be considered as atypical or not coherent with the rest of the cases stored in the memory. These cases could make the finding of a meaningful proximity T satisfying (4) and (5) impossible. Thus the introduction of a measure of *typicality* of the cases of the memory M could be interesting for assessing the relevance of cases.

4. Non-Deterministic Problems

The deterministic principle-based model may be felt too strong in some practical applications where M may for instance simultaneously include cases like (s,t) and (s,t') with $t \neq t'$. Indeed in such a case the enforcing the constraints (4) and (5) may lead to use a similarity relation T which is too permisive. An alternative solution is to use a weaker version of the CBR principle stating that

"the more similar s_1 and s_2, the more *possible* t_1 and t_2 are similar".

In our context, this weak principle becomes a non-deterministic (fuzzy) dependency rule of the form "the more similar is s to s_0 (in the sense of S), the more *possible* is the similarity of t to t_0 (in the sense of T)". It should be pointed out that this rule only concludes on the *possibility* of t_0 being similar to t. This acknowledges the fact that, often in practice, a database may contain cases which are rather similar with respect to the problem description attributes, but which are sensibly distinct with respect to outcome attribute(s). This emphasizes that case-based reasoning can only lead to cautious conclusions.

The formal expression of the above principle requires to clarify the intended meaning of "possible" in it. Rules of the form "the more X is A, the more possible Y is B" correspond to a particular kind of fuzzy rules called "possibility rules" (Dubois and Prade, 1996). They express that "the more X is A, the more possible B is a range for Y", which can be understood as "$\forall u$, if X = u, it is possible at least at the degree A(u) that Y lies in B". When B is an ordinary subset, it clearly expresses that i) if $v \in B$, v is possible for Y at least at the level A(u) if X = u, and ii) if $v \notin B$, nothing is said about the minimum possibility level of value v for Y. It leads to the following constraint on the joint possibility distribution $\pi_{Y|X}$ representing the rule (where $\pi_{Y|X}(v,u)$ estimates to what extent Y = v is possible when X = u), namely

$$\begin{cases} \pi_{Y|X}(v) \geq A(u) \text{ if } v \in B \\ \pi_{Y|X}(v) \geq 0 \text{ if } v \notin B \end{cases}.$$

When both A and B are fuzzy sets it generalizes into

$$\forall\, u \in U,\ \forall\, v \in V,\ \min(A(u), B(v)) \leq \pi_{Y|X}(v,u). \tag{8}$$

This clearly gives back the above expression when $B(v) \in \{0,1\}$. This model of fuzzy rule is close to Mamdani's (1977) original proposal in fuzzy logic-based control.

Coming back to our CBR problem, since we apply the principle "the more similar are s and s_0 (in the sense of S), the more possible is that t and t_0 are similar (in the sense of T)" then, according to (8), the fuzzy set of possible values t' for t_0 with respect to case (s, t) is given by

$$\pi_{t_0}(t') = \min(S(s, s_0), T(t, t')). \tag{9}$$

As it can be seen, what is obtained is the fuzzy set T(t) of values t' T-similar to t, "truncated" by the global degree $S(s,s_0)$ of similarity of s and s_0. The inequality in (8), which leads to a max-based aggregation of the various contributions obtained from the comparison with each case (s,t) in the memory M of cases, acknowledges the fact that each new comparison may suggest new possible values for t_0. Since (9) applies to all the pairs $(s,t) \in M$, we obtain the following fuzzy set E_{s_0} of possible values t' for t_0 :

$$E_{s_0}(t') = \max_{(s,t)\in M} \min(S(s,s_0), T(t,t')). \tag{10}$$

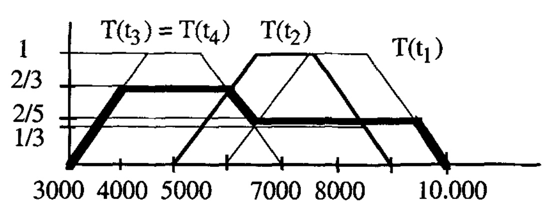

Fig. 2

The representation of this set in the case of our second-hand cars example is given in Figure 2 when S is the min-aggregation of the S^i's and T is the proximity relation given at the end of Example 2. Notice that the resulting set E may be quite uninformative if rather similar cars have very different prices in M, as it is expected. However the maximal height part of the fuzzy set in Fig. 2 is informative enough since it suggest with strength 2/3, a price in the interval [4000, 6000].

5. Dealing with incomplete cases

In this section we extend the fuzzy approach to CBR in order to deal with cases which may be incompletely described. To do this, we propose that each (possibly incompletely described) case $(\underline{s}_i, \underline{t}_i) \in M$ be understood as the generic fuzzy rule:

If X is $\underline{s}_i$-*similar*, then Y is $\underline{t}_i$-*similar*

where $\underline{s}_i$ and $\underline{t}_i$ are fuzzy sets, and s_i-*similar* and t_i-*similar* are also fuzzy sets with membership functions:

$\underline{s}_i\text{-}similar(s) = \max_{s'} \min(S(s', s), \underline{s}_i(s'))$ and

$\underline{t}_i\text{-}similar(t) = \max_{t'} \min(T(t', t), \underline{t}_i(t'))$,

that is, the convex hulls of $\underline{s}_i$ and $\underline{t}_i$ with respect to S and T respectively. Then, taking a particular current (possibly incompletely described) problem as the fuzzy statement "X is $\underline{s}_0$", we apply the traditional fuzzy logic machinery, namely the Generalized Modus Ponens (Zadeh,1979) in order to infer a fuzzy statement "Y is $\underline{t}^i_0$", where $\underline{t}^i_0$ is the fuzzy solution set for $\underline{s}_0$ w.r.t. case $(\underline{s}_i, \underline{t}_i) \in M$. $\underline{t}^i_0$ can be obtained as a sup-min composition of $\underline{s}_0$ with the fuzzy relation $R_i(s, t) = \underline{s}_i\text{-}similar(s) \to \underline{t}_i\text{-}similar(t)$ (resp. $R_i(s, t) = \min(\underline{s}_i\text{-}similar(s), \underline{t}_i\text{-}similar(t))$) in the deterministic (resp. non-deterministic) approach, namely:

$$\underline{t}^i_0(t) = \sup_s \min(\underline{s}_0(s), R_i(s, t)) \tag{11}$$

Next we briefly show how to proceed for obtaining the general (fuzzy) solution set $\underline{t}_0$ in the two models.

• **Deterministic model.** In the deterministic model, the fuzzy rules are interpreted as gradual rules, and thus the solution from the i-th case is given by

$$\underline{t}^i_0(t) = \sup_s \min(\underline{s}_0(s),\ \underline{s}_i\text{-}similar(s) \to \underline{t}_i\text{-}similar(t)),$$

but the general solution is not given by $\bigcap_i \underline{t}^i_0$ since the $\sup_s$ and the $\min_i$ do no commute in the expression (12) below corresponding to the application of the fuzzy set machinery to the whole set of cases:

$$E(\underline{s}_0)(t) = \sup_s \min(\underline{s}_0(s), \min_i R_i(s, t)) \tag{12}$$

$\bigcap_i \underline{t}^i_0$ is only an upper bound of $E(s_0)$. It is easy to check that, when s_0, s_i, t_i, are precisely known, we recover the approach of Section 3.

Example 2 (continued). Suppose that we have a case $(\underline{s}, \underline{t})$ that is only described for the year (2), power (1.600), mileage (30.000), shape (poor) and price (7.000) attributes, and the value for equipment attribute is missing. Then $\underline{s} = \{(2, 1.600, 30.000, x, \text{poor}) \mid x \in \{\text{bad, poor, good, excellent}\}\}$ and $\underline{t} = \{7.000\}$. If the current case is imprecisely defined by $\underline{s}_0 = \{(2, x, 40.000, \text{good, good}) \mid x \in [1.300, 2.000]\}$, the core of the solution[3] for $\underline{s}_0$ with respect to the case $(\underline{s}, \underline{t})$ is $\text{core}(E(\underline{s}_0)) = \{t' \mid \max\{T(t, t') \mid t \in \underline{t}\} \geq 1/3\} = [5.500, 8.500]$, i.e., the elements of $E(\underline{s}_0)$ with membership 1.

• **Non-deterministic model.** In the non-deterministic model, the fuzzy rules are interpreted as possibility rules, that is, the relation R_i describing the rule is defined as $R_i(s,t) = \min(\underline{s}_i\text{-}similar(s), \underline{t}_i\text{-}similar(t))$,

[3]when again S is the min-aggregation of the S^i's and T is the proximity relation given at the end of Example 2

$$\underline{t}^i{}_0(t) = \sup_s \min(\underline{s}_0(s), \min(\underline{s}_i\text{-}similar(s), \underline{t}_i\text{-}similar(t))).$$

In this setting, the general solution $E(\underline{s}_0)$ is the disjunctive aggregation of the $\underline{t}^i{}_0$'s, that is:

$$E(\underline{s}_0)(t) = \max_i \underline{t}^i{}_0(t).$$

Again, it is easy to check that we recover the approach of Section 4 when $\underline{s}_0$, $\underline{s}_i$, $\underline{t}_i$, are precisely known.

6 Conclusion and Further Work

In this paper we have been concerned with the modelling of some aspects of CBR using fuzzy set-based techniques, as well as their applicability. The basic tool is the use of fuzzy similarity relations both between problem descriptions and between outcomes of the cases. Each case in the memory is interpreted either as a gradual fuzzy rule (deterministic model) or as a possibility rule (non-deterministic model).

This fuzzy framework gives us the means of defining to what extent a memory M is coherent w.r.t. the use of a pair (S, T) of fuzzy similarity relations in the deterministic approach. Given a CBR-framework (M, S, T) its coherence Coh(M,S,T) is defined as a measure (between 0 and 1), which estimates how much the cases given in the memory are in agreement with our theoretical model, and in some sense the possibility to find a solution for a current case. We have

$$Coh(M,S,T)= \mathrm{Min}\{S(s_i , s_j) \to T(t_i , t_j) \,/\, (s_i , t_i), (s_j , t_j) \in M\}.$$

(M,S,T) is then said to be coherent iff Coh(M, S, T) = 1. In particular, $Coh(\{(s_1, t_1), (s_2, t_2)\}) = 1$ iff $S(s_1, s_2) \leq T(t_1, t_2)$. Moreover note that $Coh(M \cup \{(s, t)\}) \leq Coh(M)$, i.e. the introduction of a new case cannot increase the coherence. Then, given a CBR problem (M, S, T, s_0), it can be checked that a value t is a solution for s_0 in the deterministic approach iff $Coh(M \cup \{(s_0,t)\}) = 1$. A more detailed investigation of this notion of coherence is left for further research.

Acknowledgements. The authors acknowledge partial support of the bilateral CSIC - CNRS grant called"Similarity based-reasoning and its applications". The IIIA authors also acknowledge partial support from the CYCIT project SMASH (TIC96-1138-C04-01).

References

Aamodt A., Plaza E. (1994) Case-based reasoning: Foundational issues, methodological varoations and system approaches. Artificial Intelligence Communications, 7(1), 39-59.

Bonissone P., Cheetman W. (1997) Financial Applications of Fuzzy Case-Based Reasoning to Residential Property Valuation. Proc. of the 6th IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'97), Barcelona, Spain, 37-44.

Bouchon-Meunier B., Valverde L. (1993) Analogy relations and inference. Proc. of the 2nd IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'93), San Francisco, CA, March 28-April 1st, 1140-1144.

Dubois D., Esteva F., Garcia P., Godo L., Prade H. (1997a) A logical approach to interpolation based on similarity relations. Int. J. of Approximate Reasoning, 17 n°1, pp.1-36

Dubois D., Esteva F., Garcia P., Godo L., López de Mantaras R., Prade H. (1997b) Fuzzy Modelling of Case-based Reasoning and Decision, Proc. Int. Conf. Case-based Reasoning ICCBR'97 (D.Leake and E. Plaza, eds.), Springer Verlag, 599-611.

Dubois D., Prade H. (1988) Possibility Theory, Plenum Press, New York.

Dubois D., Prade H. (1992) Gradual inference rules in approximate reasoning. Information Sciences, 61, 103-122.

Dubois D., Prade H. (1994) Similarity-based approximate reasoning. In: Computational Intelligence Imitating Life (Proc. of the IEEE Symp., Orlando, FL, June 27-July 1st, 1994) (J.M. Zurada, R.J. Marks II, X.C.J. Robinson, eds.), IEEE Press, New York, 69-80.

Dubois D., Prade H. (1996) What are fuzzy rules and how to use them. Fuzzy Sets&Systems, 84, 169-185.

Dubois D., Prade H., Ughetto L. (1996) Coherence of fuzzy knowledge bases. Proc. of the 5th IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'96), New Orleans, LO, Sept. 8-11, 1996, IEEE Press, 1858-1864.

Dutta S., Bonissone P.P. (1993) Integrating case- and rule-based reasoning. Int. J. of Approximate Reasoning, 8, 163-203.

Farreny H., Prade H. (1982) About flexible matching and its use in analogical reasoning. Proc. of the 1982 Europ. Conf. on Artificial Intelligence (ECAI'82), Orsay, France, July 12-14, 43-47.

Fodor J., Roubens, M. (1994) Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Academic, Dordrecht, The Netherlands.

Jaczynski M., Trousse B. (1994) Fuzzy logic for the retrieval step of a case-based reasoner. Proc. of the EWCBR'94, 313-321.

Kolodner J. (1993) Case-Based Reasoning. Morgan Kaufmann, San Mateo, CA.

Mamdani E.H. (1977) Application of fuzzy logic to approximate reasoning using linguistic systems. IEEE Trans. Comput., 26, 1182-1191.

Ovchinnikov S.V. (1991) Similarity relations, fuzzy partitions, and fuzzy orderings. Fuzzy Sets and Systems, 40, 107-126.

Plaza E., Esteva F., Garcia P., Godo L., López de Màntaras R. (1996a) A logical approach to case-based reasoning using fuzzy similarity relations. To appear in Information Sciences.

Plaza E., López de Màntaras R. (1990) A case-based apprentice that learns from fuzzy examples. In: Methodologies for Intelligent Systems, Vol. 5 (Z.W. Ras, M. Zemankova, M.L. Emrich, eds.), Elsevier, 420-427.

Salotti S. (1989) Représentation centrée objet et filtrage flou pour raisonner par analogie: Le système FLORAN. Actes du 7ème Congrés "Reconnaissance des Formes et Intelligence Artificielle" (RFIA'89), Paris, 29 nov.-1er dec., 1695-1707.

Salotti S. (1992) Filtrage flou et représentation centrée objet pour raisonner par analogie: Le système FLORAN. (In French) PhD thesis, University of Paris XI, Orsay, France.

Zadeh L.A. (1971) Similarity relations and fuzzy orderings. Information Sciences, 177-200.

Zadeh L.A. (1979) A theory of approximate reasoning. In: Machine Intelligence, Vol. 9 (J.E. Hayes, D. Michie, L.I. Mikulich, eds.), Elsevier, New York, 149-194.