Fuzzy set-based models in case-based reasoning

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 Cedex 4, France

** Institut d'Investigacio en Intel.ligencia Artificial (IIIA), Conseil 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, which does not take into account the learning aspects of case-based reasoning, assumes a principle stating that "the more similar are the problem description attributes, the more similar are the outcome attributes". 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. A particular instance of case-based reasoning, named case-based decision is especially investigated. A logical formalization of the basic case-based reasoning inference is also proposed. Extensions of the proposed approach in order to handle imprecise or fuzzy descriptions or to manage more general forms of the principle underlying case-based reasoning are briefly discussed in the conclusion.

Key Words: Case-based reasoning; similarity; fuzzy rules; functional dependency; decision; interpolation.

1 - Introduction

Case-based reasoning amounts to inferring from known case(s) which are similar enough to a newly encountered situation, that what is true in the known case(s) might still be true, up to some suitable adaptation, in this new situation. Case-based reasoning may be viewed as a particular form of analogical reasoning, since an analogy is a relational statement of the form, *D is in the same relation w.r.t. C as B is w.r.t. A*, which is usually seen as a way of "guessing" D, knowing A, B and C. Analogical reasoning has been investigated for a long time in Artificial Intelligence (Kling, 1971; Carbonell ,1981; Gentner, 1983) and the interest in this research has been considerably renewed by the development of case-based reasoning (Kolodner, 1993; Aamodt and Plaza, 1994); see (Py, 1994) for a survey on different models of analogical reasoning. Any attempt to provide a general formalization of this common way of reasoning raises several critical issues, especially on such questions as

– the definition and practical elicitation of similarity measures,
– the retrieving of relevant cases,
– the logical modelling of the inference mechanism,
– the adaptation of the retrieved cases in order to extrapolate plausible values for the variable(s) of interest in the current situation,
– the incomplete, imprecise, or even uncertain, description of available cases, or of the current problem.

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; Prade, 1984, 1987; 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). See (Dubois and Prade, 1994) for a general overview on similarity-based approximate reasoning.

This paper provides a more systematic investigation of the potentials of fuzzy logic-based approaches for the different issues mentioned above. It is also an attempt at formalizing the basic patterns of case-based reasoning and decision. Let us first specify how case-based reasoning is viewed in this paper.

A case is viewed as a n-tuple of precise attribute values, this set of attributes being divided in two non-empty disjoint subsets: the problem description attributes subset and the soultion or outcome attributes subset 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 the estimating the value t_0 of the attributes in $\mathcal{T}$, for the current problem. It is therefore assumed everywhere in this paper that the memory only includes completely informed cases (with the exception of the conclusion section).

In the case-based reasoning setting it is also assumed that the attributes belonging to the outcome set can be related, in some way, to the problem description attributes. The goal of case-based reasoning is to estimate the value of the outcome attribute(s) of the current case taking into account the case base of precedents M. We may think of several ways for doing this. A first approach consists in looking for similarities and regularities inside the memory of cases by, for instance, clustering the cases into classes, and then building rules by induction which can be applied to the current problem s_0. This learning-based approach (whose learning step may be developed independently of any particular s_0, or on the contrary conditioned by the consideration of a particular s_0) is not investigated in this paper. In the following, it is rather assumed that a basic *principle* stating that "*similar situations give (or may give) similar outcomes*" holds. Therefore, a similarity measure between problem descriptions and a similarity measure between outcomes are needed and supposed to be given (or obtained in a preliminary step not investigated here).

As already said, case-based reasoning amounts to inferring, from a known case which is *similar enough* to the encountered situation, that what is true in the known case *might* still be true in this situation. In order to guarantee the validity of the result of the inference, the above mentioned principle relating the similarities of situations to the similarities of outcomes is supposed to hold. This idea has been emphasized for a long time in analogical reasoning, see, e.g., (Bourrelly et al., 1983). In particular, (Davies and Russell, 1987) have proposed to control the inference process by means of explicitly stated functional dependencies expressing that if two descriptions are identical with respect to the values of a set $\mathcal{S}$ of attributes, they should be also identical with respect to (an)other attribute(s) $\mathcal{T}$. However, this form of meta-rule or principle is too restrictive for two kinds of reasons: i) perfect identity is required rather than (graded) similarity, and nothing is said in case of approximate identity, ii) the functional dependencies forbid to have identical cases w.r.t. to $\mathcal{S}$ which differ w.r.t. attributes in $\mathcal{T}$, which is not very realistic in practice. Indeed two very similar second hand cars might be sold at different prices, for instance. In (Arrazola et al., 1989) and (Léa Sombé, 1990) some remedies to these two limitations have been suggested in terms of fuzzy functional dependencies.

In this paper, a more systematic discussion of the modelling and use of the case-based reasoning principle is provided, based on the assumption that the similarity measure between cases is given by two fuzzy similarity relations S and T defined on the set of problem description attributes $\mathcal{S}$ and on the set of outcome attributes $\mathcal{T}$ respectively. We suppose that these fuzzy relations are primitive notions and known in advance (although we might think of

learning them from the set of precedents stored in the memory). Methods for obtaining fuzzy relations, or checking their adequacy are out of the scope of this paper. Nevertheless, see Appendix 1 on how to build global similarity measures from elementary ones pertaining to single attributes, and for technical definitions.

Expressed in terms of the fuzzy relations S and T, the implicit Case-Based Reasoning Principle can be expressed by the following rule:

"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"

A problem in the framework of our case-based reasoning model will be denoted by a 4-tuple (M, S, T, s_0) where M stands for a case-base or memory, S and T stand for the similarity relations and s_0 stands for the current case. The goal of case-based reasoning is to estimate the outcome t_0 corresponding to the current problem s_0.

Throughout this paper we will refer to what we call *deterministic* case-based problems (M, S, T, s_0) 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). We use the term "deterministic" since, as it will become clear in next section, when the principle holds, a problem description s indeed determines the outcome t (in the sense that two identical problem descriptions should be associated with the same outcome).

Section 2 describes the fuzzy set modelling of deterministic and non-deterministic case-based reasoning problems. Section 3 discusses case-based decision models. Section 4 provides a logical formalization of the deterministic case. Concluding remarks briefly discuss various extensions of the proposed approach. Finally, in Appendix 1 we overview the notion of fuzzy similarity relations and related issues about the elicitation and aggregation of similarity relations, while Appendix 2 is about the relationship of Fuzzy Gradual Rules and Sugeno and Nishida's Interpolation Method.

2 - A Fuzzy Set Framework for Case-Based Reasoning and Decision

Fuzzy rules (see Dubois and Prade (1996) for an introduction) provide a tool for modelling the above expressions of the implicit Case-Based Reasoning Principle. We first consider the deterministic situation.

2.1 - Deterministic Problems

In the deterministic setting, the principle is supposed to express that "the more similar s_1

and s_2, the more similar t_1 and t_2" where (s_1,t_1) and (s_2,t_2) are cases of M. This can be modelled by the constraint

$$\forall (s_1,t_1),(s_2,t_2) \in M,\ S(s_1,s_2) \leq T(t_1,t_2) \tag{1}$$

where S and T are fuzzy proximity relations[1] (they range on [0,1]). S and T are supposed to be symmetric ($\forall s_1, s_2$, $S(s_1,s_2) = S(s_2,s_1)$, $T(t_1,t_2) = T(t_2,t_1)$), and reflexive ($\forall s_1$, $S(s_1,s_1) = 1$, $T(t_1,t_1) = 1$). It is worth noticing that (1) should be understood in the following way: the similarity of s_1 and s_2 constrains the similarity of t_1 and t_2 at a minimum level, i.e., $S(s_1,s_2)$ is a lower bound of $T(t_1,t_2)$. In particular if $S(s_1,s_2) = 1$ then $T(t_1,t_2)$ should be 1 also. Expression (1) corresponds to the representation of a gradual rule in the sense of (Dubois and Prade, 1992a); see also (Prade, 1988). In particular, $\forall\ \alpha \in (0,1]$, we have

$$(s_1,s_2) \in S_\alpha \Rightarrow (t_1,t_2) \in T_\alpha \tag{2}$$

where $S_\alpha = \{(s,s'),\ S(s,s') \geq \alpha\}$ is the α-cut of S and T_α is similarly defined. So (2) expresses that when s_1 and s_2 are close, t_1 and t_2 should be at least as close. Clearly, when $S(s_1,s_2) = 0$, $T(t_1,t_2)$ is no longer constrained. Moreover, if T is such that $T(t_1,t_2) = 1 \Leftrightarrow t_1 = t_2$ (separating property of T), then the classical functional dependency

$$s_1 = s_2 \Rightarrow t_1 = t_2, \tag{3}$$

is a consequence of (1) or (2) using the reflexivity of S. Constraint (1) is then clearly stronger than (3). In the paper, we always assume that T is separating *in deterministic problems* (i.e., if M agrees with the strong version of the Case-Based Reasoning Principle). Constraint (1) can be indeed viewed as a particular type of fuzzy functional dependency (see Raju and Majumdar, 1988; Dubois and Prade, 1991; Bosc et al., 1993), where the separating property of T is often hypothesized.

Let us know examine how (1) is used in the case-based inference process. Let $(s,t) \in M$, we have for the current situation s_0

$$S(s,s_0) \leq T(t,t_0)$$

where t_0 is unknown. Thus, the constraint defines a set of possible values for t_0, namely $\{t_0,\ S(s,s_0) \leq T(t,t_0)\}$. Since, it applies for any (s,t) in M, we obtain the following set E of possible values for t_0

$$E = \bigcap_{(s,t)\in M} \{t' \in \mathcal{T} \mid S(s,s_0) \leq T(t,t')\}. \tag{4}$$

---

[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 all the paper, we write F(u) instead of $\mu_F(u)$.

Note that E may be empty if T is not permissive enough. For instance, assume that $\mathcal{S} = \mathcal{T} = \mathbb{R}$, S and T are defined by $S_L(s,s') = L(|s - s'|)$, $T_K(t,t') = K(|t - t'|)$ with L and K given in Figures 1(a)-(b), M = {(7,10), (3,12)} and $s_0 = 5$. Then $S_L(s_0,7) = S_L(s_0,3) = 1/3$. Thus, t_0 according to case (7,10) is such that $t_0 \in [10 - 2/3, 10 + 2/3]$ while according to case (3,12) we get $t_0 \in [12 - 2/3, 12 + 2/3]$, and thus E = Ø.

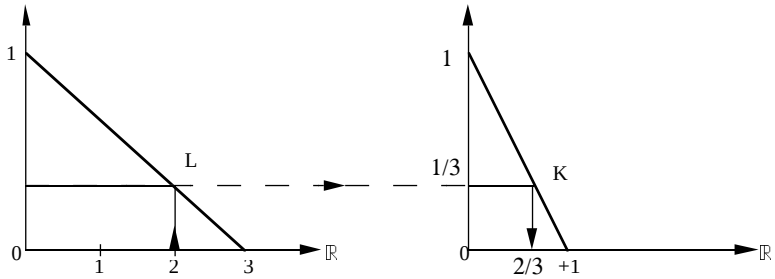

Figure 1

When E ≠ Ø, an interpolation mechanism is embedded in (2), see Appendix 2. The non-emptiness of E can be guaranteed by the coherence of the set of fuzzy gradual rules which can be built from M, S and T. Namely, these fuzzy rules are obtained by stating that for each (s,t) ∈ M, the more similar to s (in the sense of S) a value in $\mathcal{S}$ is, the more similar to t (in the sense of T) the associated value in $\mathcal{T}$ should be, or more formally, ∀ (s,t) ∈ M, ∀s', ∀t', S(s,s') ≤ T(t,t'). The coherence of this set of gradual rules is equivalent to the condition

$$\forall\, \alpha \in (0,1],\ \bigcap_{(s,t)\in M} (S(s))_\alpha \neq \emptyset \Rightarrow \bigcap_{(s,t)\in M} (T(t))_\alpha \neq \emptyset \qquad (5)$$

with S(s) (resp. T(t)) being the set of values in $\mathcal{S}$ (resp. is $\mathcal{T}$) close to s (resp. t). (5) clearly guarantees the non-emptiness of E. See (Dubois, Prade and Ughetto, 1996) for the formal definition of the coherence of a set of fuzzy rules, its equivalence with the condition (5) in the case of gradual rules, and practical checking procedures.

2.2 - Non-Deterministic Problems

Clearly, if we want to apply the principle expressed by (1), the case-base M should satisfy (1) for any pair of cases (s_1,t_1) and (s_2,t_2), and even satisfy (5) if we want to have always E non-empty. This requirement may be felt to be too strong in some practical applications where M may for instance simultaneously include cases like (s,t) and (s,t') with t ≠ t', which violates (3), and thus (1). In such a case, we suggest to use a weaker version of the Case-Based Reasoning Principle stating that "the more similar s_1 and s_2, the more *possible* t_1 and t_2 are similar". The formal expression of this principle requires to clarify the intended meaning of 'possible' in this meta-rule; this will be done in the next paragraph. A non-deterministic (fuzzy) dependency rule is thus 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.

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 conditional 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

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

When both A and B are fuzzy sets it generalizes into

$$\forall\, u \in U, \forall\, v \in V, \min(A(u), B(v)) \leq \pi_{Y|X}(v,u). \quad (6)$$

This clearly gives back the above expression when $B(v) \in \{0,1\}$. See Dubois and Prade (1992b) for a full justification of (6) by decomposing B into non-fuzzy level cuts $B_\alpha = \{v, B(v) \geq \alpha\}$. This model of fuzzy rule is close to Mamdani (1977)'s original proposal in fuzzy logic-based control.

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)", the fuzzy set of possible values t' for t_0 is given by

$$\pi_{t_0}(t') \geq \min(S(s,s_0), T(t,t')). \quad (7)$$

As it can be seen, what is obtained is the fuzzy set $\{t\} \circ T = T(t,\cdot)$ 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 (6), which leads to a max-based aggregation of the 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 (7) applies to all the pairs $(s,t) \in M$, we obtain the following fuzzy set E of possible values t' for t_0 ($E(t') = \pi_{t_0}(t')$)

$$E(t') = \max_{(s,t)\in M} \min(S(s,s_0), T(t,t')). \quad (8)$$

Note that if T reduces to the ordinary equality, (7) gives

$$E(t) = \max_{(s,t)\in M} S(s,s_0) \quad (9)$$

i.e., t is all the more possible as a value for t_0, when there exists a case (s,t) in the memory such that s and s_0 are more similar (in the sense of S).

Remark

Besides, much stronger dependency knowledge of the form "the more similar the values of s and s_0, the more *certain* the similarity of t_0 and t" can be also modelled. This can be modelled by certainty rules (Dubois and Prade, 1996). It corresponds in our problem to the constraint

$$E(t') \leq \max(T(t,t'),\ 1 - S(s,s_0)) \quad (10)$$

w.r.t. one case (s,t) which expresses, in possibility theory, that it is certain at least at the degree $S(s,s_0)$ that $t' \in T(t)$. Considering all the cases in M we get

$$E(t') \leq \min_{(s,t)\in M} \max(T(t,t'),\ 1 - S(s,s_0)). \quad (11)$$

We might think of using certainty rules in case-based reasoning. But, they are not appropriate since they do not enable us to modulate the width of the neigborhood around t in terms of the similarity between s and s_0, they only attach a level of uncertainty to $\{t\} \circ T = T(t)$, depending on $S(s_0,s)$, as expressed by (10). When $S(s,s_0) = 0$, the case (s,t) has no influence on E given by (11); when $S(s,s_0) = 1$, t' is a possible value for t_0 insofar as t' is similar to t where $(s,t) \in M$, as expressed by (11). But $S(s,s_0)$ has no influence here on the neigborhood of values around t which is considered; this contrasts with (4) where $S(s,s_0) \leq T(t,t_0)$, $\forall\ (s,t) \in M$, constrains the similarity to t of the possible values of t_0.

3 - Case-Based Decision

3.1 - Gilboa and Schmeidler's Approach

Computing the result of a case-based inference is not only a matter of retrieving similar relevant cases, even if no adaptation is to be performed, as shown by the case-based decision paradigm. Recently, Gilboa and Schmeidler (1995) have advocated a similarity-based approach to decision where a case is described as a triple (problem, act, result) and where a decision-maker's non-negative utility function u assigns a numerical value u(r) to a result r. When faced with a new situation s_0, the decision-maker is supposed to choose an act *a* which maximizes a counterpart of classical expected utility used in decision under uncertainty, namely

$$U_{s_0,M}(a) = \sum_{(s,a,r)\in M} S(s_0,s) \cdot u(r) \quad (12)$$

where S is a non-negative function which estimates the similarity of situations, here the similarity of the current situation s_0 against already encountered ones stored in the memory M.

Moreover it is assumed that $\forall s, \forall a, \exists! r$ such that $(s, a, r) \in M$ (i.e., results are uniquely determined by the act applied to the context of a given problem), and $\forall s, \exists! a$ such that $(s, a, r) \in M$ and $u(r) \neq 0$ (i.e., it means that only the best act in context s is stored in the memory). Gilboa and Schmeidler (1995) give an axiomatic derivation of this U-maximization, within a formal model.

In order to relate the case-based reasoning framework of the previous section to the above approach to decision problems, we consider the triples (s, a, r) as cases ((s,a), r) and we assume $S((s,a), (s_0,a)) = S(s,s_0)$. Under a rather special hypothesis on the contents of M, it is then possible to retrieve (12) by applying the approach of Section 2.1 in the deterministic case. Indeed let us try to estimate the utility $u(r_0)$ attached to the act *a* applied to the current problem description s_0. Then (4) yields

$$E_a = \bigcap_{(s,a,r)\in M} \{u(r') \mid S(s,s_0) \leq T(u(r), u(r'))\}. \tag{13}$$

Using results in Appendix 2, it can be checked that under the hypothesis that M and S are such that for s_0, $\exists!(s_1,s_2)$ such that $(s_1, a, r_1) \in M$, $(s_2, a, r_2) \in M$, $S(s_0,s_1) > 0$, $S(s_0,s_2) > 0$ and moreover that $S(s_0,s_1) + S(s_0,s_2) = 1$, then E_a contains only one element e_a equal to

$$e_a = \frac{\sum_{(s,a,r)\in M} S(s,s_0) \cdot u(r)}{\sum_{(s,a,r)\in M} S(s,s_0)} \tag{14}$$

where we recognize (12) upto a normalization factor depending on the act *a*. An appropriate choice of T in (13) is also required in order to have (13) reducing to the value (14). See Appendix 2. This shows that for particular S and T, (13) embeds a linear interpolation mechanism. One way of satisfying the requirement of the above hypothesis when the problems s in M can be linearly ordered, is in fact to assume that a different S, say S_s, is associated with the neighborhood of each problem s, in such a way that S(s,s') decreases to 0 when s' goes away from s and coincide with one of the two closest neighbors of s in M. Then it is easy to make $S_{s_1}(s_0,s_1) + S_{s_2}(s_0,s_2) = 1$; see Appendix 2.

Note that (12) is also akin to the fuzzy vote procedure used by Bensana et al. (1988) for selecting a (scheduling) decision b in a situation s_0 from a set of rules if s is $S^{(i)}$ then the recommended act is $a^{(i)}$ with weight $w^{(i)}$ (where $w^{(i)}$ can be viewed as the utility of the result of the act $a^{(i)}$) as the one maximizing an index of the form

$$U(a) = \sum_{i:a=a^{(i)}} S^{(i)}(s_0) \cdot w^{(i)} \tag{15}$$

where $S^{(i)}$ is a fuzzy set describing a fuzzy class of situations.

At the reasoning level, the above decision procedures make use of similarity notions which can be viewed as the amount of "stretching" of the typical situation described in the condition part of the rule (i.e., the tolerance expressed by $S^{(i)}$ for extending the range of use of

the rule) to encompass the current problem s_0. This idea of stretching a set of interpretations (compatible here with the condition part of a rule) is also explicitly at work in the logical approach presented in Section 4 (which encodes the approach of Section 2.1).

Although (12) looks like an expected utility expression where probabilities are replaced by similarity degrees, its intuitive interpretation is quite different from the decision under uncertainty situation. First, note that there is no constraint on the sum $\sum_s S(s,s_0)$, in particular it has not to sum to 1 (in (14) the normalizing factor depends on act *a*!). The idea is rather to look for acts that in several similar situations had results with a high utility. However, a drawback of (12) is it unability to distinguish between the two following types of "extreme" situations

(i) For act a_1, $\exists!$ s* such that $S(s_0,s^*)$ is high, $u(r^*)$ is *high* and $(s^*, a_1, r^*) \in M$ and $S(s_0,s)$ is zero for any other s such that $(s, a, r) \in M$;

(ii) For act a_2, $\exists$ $s_1, \ldots, s_n$, with n sufficiently large, such that $S(s_0,s_i)$ is high for i = 1,n and $u(r_i)$ is *low* but non-zero where $(s_i, a_2, r_i) \in M$, while $S(s_0,s)$ is zero for any other s such that $(s, a, r) \in M$

and we have $U_{s_0,M}(a_1) < U_{s_0,M}(a_2)$.

Thus, it may look strange to prefer act a_2, which always gave rather poor results in situations similar to s_0, to act a_1, which gave a very good result in a situation quite similar to s_0 (even if this situation is unique). Moreover (12) somewhat compensates between good results r and bad results r' attached to the same act *a* for distinct problems s and s' which are both similar to s_0 (if we have both (s, a, r) and (s', a, r') in M).

3.2 - Alternative Approach

Another idea is to look for acts which for similar problems always gave good results. Then, for a given act *a*, we are interested in computing a degree of inclusion of the fuzzy set of problems which are similar to s_0 and where act *a* was experienced, into the fuzzy set of situations where act *a* led to good results.

The function u, like S, is now supposed to range on the real interval [0,1], with the following interpretations: S(s,s') = 1 means perfect similarity of s and s', S(s,s') = 0 means that s and s' are not at all similar as previously, while u(r) = 1 means that the result r is among the best ones, while u(r) = 0 means that r is among the worst ones. Thus, the fuzzy set of situations similar to s_0 is represented by $S(.,s_0) = \{s_0\} \circ S$, and u is the membership function of the fuzzy set of good results. The following degree of inclusion enables us to select the act(s) *a*, if any, which for problems similar to s_0 always gave good results:

$$U*_{s_0,M}(a) = \min_{(s,r):\ (s,a,r)\in M} S(s,s_0) \rightarrow u(r) \qquad (16)$$

where $\rightarrow$ is a multiple-valued implication connective (i.e., $x \rightarrow y$ increases with y, decreases with x, and coincides with material implication on {0,1}). Assuming

- $U*_{s_0,M}(a) = 1$ only if $\{s, (s, a, r) \in M, S(s,s_0) > 0\} \subseteq \{s, (s, a, r) \in M \text{ and } u(r) = 1\}$
- $U*_{s_0,M}(a) = 0$ as soon as $\exists s, S(s,s_0) = 1, (s, a, r) \in M$ and $u(r) = 0$.

leads to choose an implication $x \rightarrow y$ of the form $x \rightarrow y = n(x) \perp y$ where n is an involutive negation function and $\perp$ a disjunction operation. In case of a purely ordinal interpretation of [0,1] where only the ordering of the levels is meaningful, we are led to use $x \rightarrow y = \max(1 - x, y)$,[2] i.e.,

$$U*_{s_0,M}(a) = \min_{(s,r):\ (s,a,r)\in M} \max(u(r), 1 - S(s,s_0)) \qquad (17)$$

which expresses that the existence of a case (s, a, r) in M does not penalize act *a* w.r.t. s_0, if r is a good result, or if s is not similar to s_0.

Remark: Expressions (16) and (17) make sense for case-based decision only when they take into account acts "a" that have been experienced in at least some situation very similar to s_0. Otherwise, it could happen that $U*_{s_0,M}(a)$ would be very high while $S(s, s_0)$ very low for all triples $(s, a, r) \in M$. For example, if $S(s, s_0) = 0$ for all $(s, a, r) \in M$, then $U*_{s_0,M}(a) = 1$. To overcome this problem, a possible alternative is, for each act a, to renormalize[3] the fuzzy set $S(., s_0)$ over the situations s such that $(s,a,r) \in M$, or perhaps better, to take into account only those situations whose similarity with s_0 is above a certain level α. This amounts to compute the parametrized inclusion degrees of the α-cut sets $M(a,\alpha) = \{(s, a, r) \in M \mid S(s, s_0) \geq \alpha\}$ into the fuzzy set of good results, i.e.:

$$U*_{s_0,M,\alpha}(a) = \min_{(s,r):\ (s,a,r)\in M(a,\alpha)} u(r).$$

Another option of keeping track the absence of situations very similar to s_0 would be a soft discount of (16) or (17) (once $S(.,s_0)$ was already renormalized) by performing the minimum of this result with the value $\max\{S(s, s_0) \mid (s, a, r) \in M\}$.

$U*_{s_0,M}$ is a rather drastic criterium since it requires that in *all* the problems similar to s_0, act *a* led in good results. A more "optimistic" behaviour can consist in selecting all the acts which led to a good result for at least one problem similar to s_0, i.e., the dual criteria

$$U^*_{s_0,M}(a) = \max_{(s,r):\ (s,a,r)\in M} \min(u(r), S(s,s_0)). \qquad (18)$$

---

2 Since $\max(1 - x, y) = 1$ if and only if $y = 1$ whenever $x > 0$. On an ordinal scale, $1 - (\cdot)$ is replaced by the order-reversing function of the scale.

3 There are several ways of renormalizing a fuzzy set μ. If the scale is numerical the usual option is define $\mu'(u) = \mu(u) / \max\{\mu(v) \mid v \in U\}$, while if the scale is qualitative we may define $\mu''(u) = \mu(u)$ for $u \neq u_0$ and $\mu''(u_0) = 1$, where $\mu(u_0) = \max\{\mu(v) \mid v \in U\}$.

Thus $U^*_{s_0,M}(a)$ is maximum as soon as it exists a case corresponding to a problem completely similar to s_0 where the act *a* led to an excellent result. Note that

$$U^*_{s_0,M}(a) \geq U_{*s_0,M}(a). \quad (19)$$

Interestingly enough, (16) and (17) can be viewed as the counterparts of qualitative expectations recently introduced in decision making under uncertainty (Dubois and Prade, 1995), just as (12) may be seen as a counterpart of the classical expected utility.

This approach can be also related to the approach proposed in Section 2.2 for the non-deterministic case. This point of view leads to estimate the utility of act *a* for problem s_0 by the fuzzy set E_a of possible values t' defined as

$$E_a(t') = \max_{(s,a,r)\in M} \min(S(s,s_0), T(u(r), t')) \quad (20)$$

or, as

$$E_a(u(r)) = \max_{(s,a,r)\in M} S(s,s_0) \quad (21)$$

when taking T to be the classical identity relation.

Thus, (17) and (18) can be also viewed as scalar summaries, which are pessimistic and optimistic respectively, of the fuzzy set of values whose membership function is given by (21). Indeed (17) and (18) are weighted median operations; see (Dubois and Prade, 1986).

Besides, it is possible in practice to make $U_{*s_0,M}$ less drastic by only requiring that for *most* cases (rather than all) where *a* is present, the obtained results are good when the situation is similar to s_0, thus allowing for a few exceptions, see, e.g., Dubois and Prade (1996b).

4 - A Similarity Logic Setting for Case-Based Reasoning

In this section we propose an alternative setting, closer to logical formalisms than the one described in Section 2.1, for the inference processes that take place in case-based reasoning systems when solving "deterministic" problems. The approach is inspired in the logical formulation of case-based inference proposed in Plaza et al. (1996), which in turn relies on two kinds of graded consequence relations that play a major role in similarity-based reasoning as shown in Dubois et al. (1995, 1996). For another attempt at formalizing case-based reasoning inference in a logical setting see (Lieber and Marquis, 1996).

The underlying idea is that when a current problem description s_0 is compared with a precedent case (s,t) of the memory M, there are two basic steps:

(i) comparing s_0 with s, by means of the similarity relation S, and

(ii) extrapolating the problem solution t for s to s_0, according to their similarity degree

computed in (i).

Our aim is to provide a logical account for the whole inference process. But before that we introduce the notions of approximate and proximity entailments from Dubois et al. (1996) which pertain to the inference steps (i) and (ii) respectively.

4.1 - Background

Let $\mathcal{L}$ be a finite Boolean language, Ω its set of interpretations and let S be a similarity relation on Ω. Ruspini (1991) defines the degree $I_S(p \mid \omega)$ to which an interpretation ω is close to some model ω' of a proposition p ($\omega' \models p$), i.e.,

$$I_S(p \mid \omega) = \sup_{\omega' \models p} S(\omega,\omega'). \tag{22}$$

This defines, for each proposition p, the fuzzy set [p*] of interpretations which are close to p, by just considering

$$[p^*](\omega) = I_S(p \mid \omega) \tag{23}$$

The fuzzy set [p*] is such that its α-level cuts $[p^*]_\alpha = \{\omega \mid [p^*](\omega) \geq \alpha\}$, for $\alpha \in (0,1]$, define a family of nested approximations of the set [p] of models of p, i.e., at each level α, $[p^*]_\alpha$ is the set of interpretations which are α-similar to p. Based on this idea, in Dubois et al. (1996) the following graded *approximate entailment* is introduced:

$$\omega \models^\alpha p \text{ iff } \omega \in [p^*]_\alpha \tag{24}$$

and more generally,

$$q \models^\alpha p \text{ iff } \omega \in [p^*]_\alpha \text{ for all models } \omega \text{ of q, i.e., iff } [q] \subseteq [p^*]_\alpha \tag{25}$$

which can be equivalently expressed as $q \models^\alpha p$ iff $I_S(p \mid q) \geq \alpha$, where $I_S(p \mid q) = \inf_{\omega \models q} I_S(p \mid \omega)$ is known as the implication degree of p given q. For a context given under the form of some proposition K, it will be also useful to introduce the following notation

$$q \models_K^\alpha p \text{ iff } q \wedge K \models^\alpha p \tag{26}$$

For modelling step (ii) above, we are viewing a case (p, q), where p and q are taken as any pair of propositions of $\mathcal{L}$, as a kind of extrapolative entailment relation between p and q, with the following intended meaning: given a certain context K, p implies q in the classical sense, but moreover, if p is "close" to being true then q is also "close" to being true. In other words, in that context, the neighborhood of models of p should lie in the neighborhood of the models of q. This is in accordance with the view in Section 2.1 of a case as a gradual fuzzy rule "the more ω in [p*], the more ω in [q*]", which is therefore expressed as

$$[p^*](\omega) \leq [q^*](\omega), \text{ for all models } \omega \text{ of K.} \tag{27}$$

Furthermore this notion of entailment, called *proximity entailment*, can be easily graded as well, just by noticing that $[p^*](\omega) \leq [q^*](\omega)$ iff $[p^*](\omega) *\!\!\rightarrow [q^*](\omega) = 1$, $*\!\!\rightarrow$ being the implication function defined by residuation from any t-norm $*$.[4] Given such an implication $*\!\!\rightarrow$, it is natural to define

$$p \mid\!\equiv_K^{\alpha} q \text{ iff } [p^*](\omega) *\!\!\rightarrow [q^*](\omega) \geq \alpha, \text{ for all models } \omega \text{ of K.} \tag{28}$$

It is worth mentioning that $\models_K^{\alpha}$ and $\mid\!\equiv_K^{\alpha}$ coincide only when K is uninformative, i.e., when $[K] = \Omega$. However it is easy to show the following relationship between approximate and proximity entailments:

$$\text{if } p' \models^{\alpha} p \text{ and } p \mid\!\equiv_K^{\beta} q \text{ then } p' \models_K^{\alpha * \beta} q \tag{29}$$

This inference pattern can be read as follows: if p' makes p α-true and, in a given context K, p extrapolatively entails q to the degree β then, in the same context K, p' makes q $\alpha * \beta$-true.

4.2 - Logical Modelling of Case-Based Inference

Now we make use of the logical framework just introduced in the previous subsection to reformulate the inference process involved in the case-based reasoning problems presented in Section 2.1. The main ideas are the following ones:

- step (i) is formulated as computing the approximate entailment degree of s by s_0,
- each case (s, t) is modelled as a proximity entailment, given a suitable context K,
- step (ii) is then, due to the inference pattern (29), formulated as computing the approximate entailment degree of t given s_0 in the context of K.

First of all, we need to introduce some notations. Given the sets of attribute predicate symbols $\mathcal{S}$ and $\mathcal{T}$, corresponding to problem description attributes and to the outcome attributes respectively, $L_{\mathcal{S}}$ (resp. $L_{\mathcal{T}}$) will denote the set of atomic formulas of the form A(c), where $A \in \mathcal{S}$ (resp. $A \in \mathcal{T}$) and c is an attribute value. Actually $L_{\mathcal{S}}$ and $L_{\mathcal{T}}$ are nothing but a set of propositional variables since we will not deal with object variables. This simplifies the semantical notion of interpretation. An interpretation will be just a value assignment for each attribute. We will denote by Ω_L and Ω_T the set of interpretations for $L_{\mathcal{S}}$ and $L_{\mathcal{T}}$ respectively, and we define $\Omega = \Omega_L \times \Omega_T$. Assuming that the problem descriptions and outcomes are complete sets of precise attribute values for attributes in S and T, we can identify problem descriptions with interpretations of Ω_L and outcome descriptions with interpretations of Ω_T.

---

4 The use of residuated implication ($x *\!\!\rightarrow y = \sup\{z \in [0,1], x * z \leq y\}$ is also in accordance with the semantics proposed in Dubois and Prade (1996) for the so-called gradual fuzzy rules.

According to the previous sections, we assume to have a similarity S on Ω_L and a similarity T on Ω_T. Then one can define the product similarity $S \times T$ on Ω by

$$S \times T((s_1,t_1), (s_2,t_2)) = \min(S(s_1,s_2), T(t_1,t_2))$$

and it is easy to check that the corresponding marginals[5] of $S \times T$ on Ω_L and Ω_T are S and T respectively.

Given a case base $M = \{(s_i,t_i)\}_{i \in I}$, and taking into account that problem and solution descriptions are indeed interpretations of $\Omega_L$and Ω_T respectively, step (i) amounts to compute the similarity degree $S(s_i,s_0)$ between the current description and the precedent cases, while the context K needed to perform the extrapolative inference (i.e., the logical counterpart of our "deterministic principle") is assumed to be of the form $K = \{(s,t) \mid [s_i^*](s) \ast\!\!\rightarrow [t_i^*](t) \geq \beta_i\} = \{(s,t) \mid S(s_i,s) \ast\!\!\rightarrow T(t_i,t) \geq \beta_i\}$, where the parameters β_i are usually taken to be 1 (this is the situation that exactely corresponds to the expression (1) of our deterministic principle[6]), but in general they could be assigned a lower value, denoting a weaker form of graduality between s_i and t_i. Then , the whole inference process described by (i) and (ii) can be formulated as follows:

$$\begin{array}{ll} \text{from} & s_0 \models^{\alpha_i} s_i \\ & s_i \mid\equiv_K^{\beta_i} t_i \\ \\ \text{infer} & s_0 \models_K^{\alpha_i * \beta_i} t_i \end{array}$$

where $\alpha_i = S(s_0,s_i)$. In terms of the level-cuts, the inferred result is that $[s_0 \wedge K] \subseteq [t_i^*]_{\alpha_i * \beta_i}$, for all $i \in I$. Therefore, the estimated solution t_0 for s_0 should verify

$$t_0 \in \bigcap_{(s_i,t_i) \in M} \{t \mid T(t_i,t) \geq \alpha_i * \beta_i\}$$

which reduces to (5) when $\beta_i = 1$ for all $i \in I$.

5 - Conclusions and Further Work

In this paper we have been concerned with the modelling of some aspects of case-based reasoning and decision 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. These similarity relations allow us to give a precise meaning to the implicit principle underlying "deterministic" case-based reasoning, i.e., "the more similar

---

5 Given a similarity R on $U \times V$, its marginal R_U on U is defined as $R_U(u_1,u_2) = \inf\{R((u_1,v), (u_2,v)) \mid v \in V\}$, for all $u_1, u_2 \in U$.

6 Notice that the condition $S(s_i,s) \ast\!\!\rightarrow T(t_i,t) \geq \beta_i$ is equivalent to $S(s_i,s) * \beta_i \leq T(t_i,t)$, which is a more general constraint than expression (1).

the problem descriptions, the more similar their outcomes", by interpreting each case in the memory as a gradual fuzzy rule. This interpretation has been also formalized in a logical setting by means of similarity-based graded consequence relations. Furthermore, a weaker form of the above principle has been proposed in order to cope with "non-deterministic" problems. Fuzzy possibility rules have then been used instead of gradual rules to model these non deterministic cases.

The ideas contained in this paper are a starting point regarding the potential of fuzzy techniques and models for case-based reasoning. We envisage to further develop these ideas in at least four different aspects.

(i) It may be useful to extend the inference power by taking into account not only similarities but also *dissimilarities* between the cases and the current problem by means of different types of gradual rules. These rules may allow, for instance, to express generalized functional dependencies of the form "the less similar s and s_0, the less similar t and t_0", etc. More generally, we may take advantage of dependencies, like "the more similar s_0 and s (according to a given subset of attributes), the more possible for t_0 the value(s) obtained from t by taking into account, in a specified way, the amount of dissimilarity between s and s_0 (according to another set of attributes)", as already suggested in Léa Sombé (1990). An example of such a generalized dependency is "if two second-hand cars are similar but their mileages differ from some amount x, then the difference between their prices is about f(x)".

(ii) *Dealing with cases whose attributes value are imprecise or unknown*: when the description of cases is pervaded by imprecision or uncertainty, we can no longer be sure about the identity or even the approximate equality of two feature values. Let us assume that our information about feature values is represented by means of possibility distributions. Let π and π' be two possibility distributions pertaining to the same feature x for two different cases. $\pi(u) = 1$ means that the value u is completely possible for x (note that several distinct values u and u' may be completely possible for x), while $\pi(u) = 0$ means that value u is totally impossible for x. Then we can compute to what extent it is possible and it is certain respectively that the two ill-known feature values are similar in the sense of the fuzzy relation S by

$$\prod_{\pi\times\pi'}(S) = \sup_{u,v} \min(S(u,v), \pi\times\pi'(u,v))$$

and

$$N_{\pi\times\pi'}(S) = \inf_{u,v} \max(S(u,v), 1 - \pi\times\pi'(u,v))$$

respectively, where $\pi\times\pi'(u,v) = \min(\pi(u), \pi'(v))$. When $\exists! u_0$, $\pi(u_0) = 1$, $\exists! v_0$, $\pi(v_0) = 1$ and $\forall\ u \neq u_0$, $\pi(u) = 0$, $\forall\ v \neq v_0$, $\pi'(v) = 0$, then $\prod_{\pi\times\pi'}(S) = N_{\pi\times\pi'}(S) = S(u_0,v_0)$, i.e., we recover the completely informed situation case. Note also that when S reduces to the perfect identity, i.e., $S(u,v) = 1 \Leftrightarrow u = v$ and $S(u,v) = 0$ otherwise, N(S) is always zero as soon as the information on the two cases is not precise and certain. The possibility and necessity degrees $\prod_{\pi\times\pi'}(S)$ and $N_{\pi\times\pi'}(S)$ are easy to compute for trapezoidal possibility distributions, and can be aggregated by min operation (or its weighted version; see Appendix 1), still keeping their possibility and necessity semantics. See (Dubois and

Prade, 1987) for details.

(iii) *Incorporating learning capabilities*: as we have already mentioned in the introduction, a possible approach could consist in looking for similarities and regularities inside the memory of cases by, for instance, clustering the cases into classes, and then building rules by induction which can be applied to the current problem s_0. Another learning aspect would be that of assessing the relevance of attributes along the lines proposed in (Plaza et al., 1996b). We may also think of building the similarity relations S and T from the memory M in such a way that the set of cases leads to a coherent set of fuzzy rules (in the sense of (5)); this may include a step where very similar cases in M are in fact fused into one fuzzy case.

(iv) *Using interpolation for case adaptation*: the fuzzy set-based model presented in Section 2.1, as well as the logical model proposed in Section 4, incorporate a natural interpolation mechanism (see Dubois et al., 1995, 1996) which may be very helpful for the case adaptation step, specially when the outcome attribute domains are of numerical nature.

Acknowledgements: The authors are grateful to Enric Plaza and Adriana Zapico for valuable comments and discussions on preliminary drafts of this paper. The authors acknowledge partial support of the CSIC/CNRS grant called"Similarity based-reasoning and its applications". Francesc Esteva, Pere Garcia and Lluís Godo have been also partially supported by the E.U. Copernicus project MUM (No.10053).

References

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

Arrazola I., Plainfossé A., Prade H., Testemale C. (1989) Extrapolation of fuzzy values from incomplete data bases. Information Systems, 14, 487-492.

Bensana E., Bel G., Dubois D. (1988) OPAL: A multi-knowledge-based system for industrial job-shop scheduling. Int. J. of Prod. Res., 26, 795-819.

Bosc P., Dubois D., Prade H. (1994) Fuzzy functional dependencies —An overview and a critical discussion—. Proc. of the 3rd IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'94), Orlando, June 26-29, 325-330.

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.

Bourrelly L., Chouraqui E., Ricard M. (1983) Formalisation of an approximate reasoning: The analogical reasoning. Proc. of the IFAC Symp. on Fuzzy Information, Knowledge Representation and Decision Analysis, Marseille, 19-21.

Carbonell J. (1981) A computational model of problem solving by analogy. Proc. of the 7th IJCAI, Vancouver, 147-152.

Davies T.R., Russell S.J. (1987) A logical approach to reasoning by analogy. Proc. of the 10th Inter. Joint Conf. on Artificial Intelligence (IJCAI'87), Milano, Italy, Aug. 23-28, 264-270.

Dubois D., Esteva F., Garcia P., Godo L., Prade H. (1995) Similarity-based consequence relations. In: Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Proc. of the Europ. Conf. ECSQARU'95, Fribourg, Switzerland, July 1995) (C. Froidevaux, J. Kohlas, eds.), Lecture Notes in Artificial Intelligence, Vol. 946, Springer Verlag, Berlin, 171-179.

Dubois D., Esteva F., Garcia P., Godo L., Prade H. (1996) A logical approach to interpolation based on similarity relations. Int. J. of Approximate Reasoning, to appear.

Dubois D., Grabisch M., Prade H. (1994) Gradual rules and the approximation of control laws. In: Theoretical Aspects of Fuzzy Control (H.T. Nguyen, M. Sugeno, R. Tong, R.R. Yager, eds.) Wiley, New York, 147-181.

Dubois D., Prade H. (1986) Weighted minimum and maximum operations in fuzzy set theory. Information Sciences, 39, 205-210.

Dubois D., Prade H. (1987) Théorie des Possibilités — Applications à la Représentation des Connaissances en Informatique. (2nd edition) Masson, Paris.

Dubois D., Prade H. (1991) Certainty and uncertainty of (vague) knowledge and generalized dependencies in fuzzy data bases. Proc. of the Inter. Fuzzy Engineering Symp. (IFES'91), Yokohama, Japan, Nov. 13-15, 239-249.

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

Dubois D., Prade H. (1992b) Fuzzy rules in knowledge-based systems —Modelling gradedness, uncertainty and preference—. In: An Introduction to Fuzzy Logic Applications in Intelligent Systems (R.R. Yager, L.A. Zadeh, eds.), Kluwer Academic Publ., Dordrecht, 45-68.

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. (1995) Possibility theory as a basis for qualitative decision theory. Proc. of the 14th Inter. Joint Conf. on Artificial Intelligence (IJCAI'95), Montréal, Canada, Aug. 20-25, 1924-1930.

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

Dubois D., Prade H. (1996b) Semantics of quotient operators in fuzzy relational databases. Fuzzy Sets and Systems, 78, 89-93.

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.

Gilboa I., Schmeidler D. (1995) Case-based decision theory. The Quarterly J. of Economics, August, 607-639.

Gentner D. (1983) Structure mapping: A theoretical framework for Analogy. Cognitive Science, 7, 155-170.

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

Kling R.E. (1971) A paradigm for reasoning by analogy. Artificial Intelligence, 2, 147-178.

Klawonn F., Kruse R. (1993) Equality relations as a basis for fuzzy control. Fuzzy Sets and Systems, 54, 147-156.

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

Léa Sombé (Group) (1990) Reasoning by analogy. In: Reasoning Under Incomplete Information in Artificial Intelligence (Léa Sombé, eds.), Wiley, New York, 418-424.

Lieber J., Marquis P. (1996) Domain-independent similarity relations for case-based reasoning in a logical framework. Proc. of the Poster Session of the 9th Inter. Symp. on Methodologies for Intelligent Systems (ISMIS'96), Zakopane, Poland, June 9-13, 230-241.

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

Niiniluoto I. (1988) Analogy and similarity in scientific reasoning. In: Analogical Reasoning (D.H. Helman, ed.), Kluwer Academic Publ., 271-298.

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.

Plaza E., López de Màntaras R., Armengol E. (1996b) On the importance of similitude: An entropy-based assessment. Report IIIA 96/14, IIIA-CSIC, University of Barcelona, Bellaterra, Spain.

Prade H. (1984) Analogie et flou. BUSEFAL (IRIT, Univ. P. Sabatier, Toulouse, France), 18, 83-91.

Prade H. (1987) Analogie, flou et systèmes experts. BUSEFAL (IRIT, Univ. P. Sabatier, Toulouse, France), 32, 108-114.

Prade H. (1988) Raisonner avec des règles d'inférence graduelle — Une approche basée sur les ensembles flous. Revue d'Intelligence Artificielle, 2(2) (Hermès, Paris), 29-44.

Py M. (1994) Un modèle conceptuel de raisonnement par analogie. Revue d'Intelligence Artificielle (Hermès, Paris), 8(1), 63-99.

Raju K.V.S.V.N., Majumdar A.K. (1988) Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems. ACM Trans. on Database Systems, 13(2), 129-166.

Rosch E., Mervis C.B. (1975) Family ressemblances studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Ruspini E.H. (1991) On the semantics of fuzzy logic. Int. J. of Approximate Reasoning, 5, 45-88.

Russell S.J. (1989) The Use of Knowledge in Analogy and Induction. Pitman, London.

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.

Sugeno M., Nishida M. (1985) Fuzzy control of model car. Fuzzy Sets and Systems, 16, 103-113.

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.

Appendix 1 - Similarity Representation

The evaluation of the similarity between two multiple-feature descriptions of cases is crucial in case-based reasoning. Usually a global similarity degree between two cases given by a multiple-feature description is obtained by aggregating degrees of similarities pertaining to one feature.

In this appendix we introduce the notion of fuzzy similarity relations and we study some aspects of the elicitation of a similarity relation and the aggregation of a family of similarity relations.

Fuzzy Similarity Relations

"The concept of similarity relation is essentially a generalization of the concept of an equivalence relation" (Zadeh, 1971). A similarity relation S is a binary fuzzy relation defined on a set U satisfying the three following properties

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

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

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

where $*$ is a binary operation defined on [0,1] such that $0 * 0 = 0 * 1 = 1 * 0 = 0$, $1 * 1 = 1$. Fuzzy relations satisfying (i) and (ii) are often called proximity relations. The operation $*$ is usually a t-norm operation, that is, $*$ is a non-decreasing operation on [0,1] satisfying associativity, commutativity, 1 being the neutral element and 0 being an absorvent element. Noticeable t-noms are min, product and Lukasiewicz operation, i.e., $\max(0, a + b - 1)$. See (Ovchinnikov, 1991) for a recent overview on similarity relations.

Sometimes a stronger property than reflexivity is required for similarity relations:

(i') $\forall\, u \in U,\ S(u,v) = 1$ if and only if $u = v$.

If (i') holds, the similarity relation is said to be *separating*. Interestingly enough, if a similarity relation is separating then it is always $*$-transitive for $*$ being the "smallest" t-norm operation, i.e.,

$$a * b = \begin{cases} \min(a,b) \text{ if } \max(a,b) = 1 \\ 0 \text{ otherwise} \end{cases}$$

Similarity relations are closely related to the idea of distance. In particular if S is based on the Lukasiewicz t-norm, 1 – S is a pseudometric, and $*$-transitivity is equivalent to the triangle inequality; the min-transitivity, which is stronger, corresponds to ultrametrics.

$*$-transitivity can be also easily interpreted in terms of composition of fuzzy relations, namely (iii) can be written

$$S \circ_* S \subseteq S \tag{1}$$

where $\circ_*$ composition is defined as $S \circ_* R(u,w) = \sup_v S(u,v) * R(v,w)$ and fuzzy set inclusion is pointwisely defined by an equality between membership functions. We wrote $\circ$ simply if $*$ is the min t-norm. As we can see (iii) (or (1)) extends transitivity to degrees of similarity which are intermediary between 0 and 1 (transitivity of non-fuzzy relations is retrieved as a particular case). In the following we shall encounter *weaker* forms of transitivity-like properties where the similarity or approximate equality slightly deteriorates when applying transitivity. While reflexivity and symmetry are minimum properties that are clearly required when evaluating the closeness of cases (proximity relations), extended transitivity does not seem always compulsory in case-based reasoning. For instance, we can think of the following reason why transitivity is not only a non required property but an undesirable property in some settings. In case-based reasoning, given a current situation s_0, to retrieve the most similar cases from the repertory M, we measure pairwise the similarity of s_0 with each of the situations s_i in M, but 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.

Another reason refers specifically to the fact that many similarity measures which are often used in case-based reasoning make a quite 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 defined as

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

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

Elicitation of Similarity Relations

A key issue is then the problem of the practical elicitation of these elementary similarity measures. This has been discussed by many authors from cognitive and formal philosophy points of view, see, e.g., (Rosch and Mervis, 1975; Niiniluoto, 1988). In the following we

only consider the easiest, but often-encountered, situation where we deal with a numerically-valued feature, in particular we consider that this feature is valued on the real numbers. (It can be extended when the domain of the feature values is discrete but linearly ordered.) In this subsection we consider two types of elicitation:

- The first one given by means of a class of fuzzy numbers centered on 0. These fuzzy numbers are used to measure how much x is close to y, assuming that x is approximately y if and only if (x – y) is approximately 0 and this is measured by a fuzzy number L centered on 0.
- The second one given by means of a class of fuzzy numbers centered on 1. These fuzzy numbers are used to measure how much x is close to y, assuming that x is approximately y if and only if (x / y) is approximately 1.

Following the first suggestion given before, an approximate equality can be modelled by a fuzzy relation S of the form

$$S_L(x,y) = L(|x - y|) \tag{2}$$

which depends on the difference between feature values x and y. For instance,

$$\forall x,\ \forall y,\ S(x,y) = \max\left(0, \min\left(1, \frac{\delta + \varepsilon - |x - y|}{\varepsilon}\right)\right) = \begin{cases} 1 \text{ if } |x - y| \le \delta \\ 0 \text{ if } |x - y| > \delta + \varepsilon \\ (\delta + \varepsilon - |x - y|) / \varepsilon \text{ otherwise} \end{cases} \tag{3}$$

where δ and ε are respectively positive and strictly positive parameters which modulate the approximate equality, that is, we consider the family of fuzzy sets whose membership functions are defined in Figure A1.1. These fuzzy sets L are centered in 0 and L(d) = L(–d) in order to ensure the symmetry of the approximate equality relation.

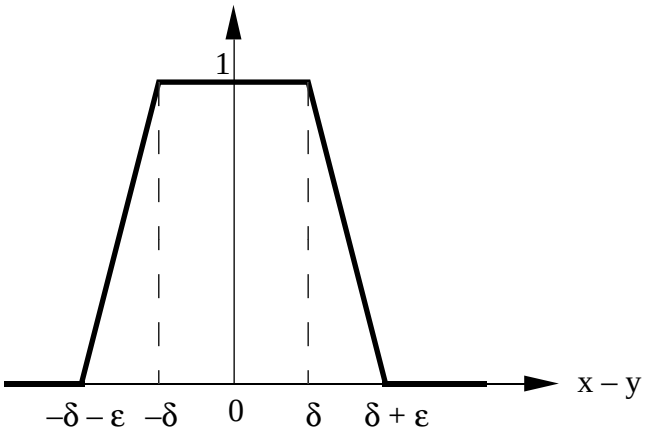

Figure A1.1

Then the elicitation only requires the identification of two thresholds: i) δ which determines the upper limit of the difference between x and y to which the user is completely

indifferent, and ii) $\delta + \varepsilon$ which corresponds to the smaller difference for which the user considers that the two values are not at all similar. Notice that, in general the resulting approximate equality relation S_L is not $*$-transitive (for the usual t-norm operations $*$) because of the (fuzzy) threshold determined by $\delta \neq 0$ and ε.

Observe that the equality relation S_L, defined by L, is reflexive and symmetric, and is separating if and only if $\delta = 0$.

The $*$-transitivity property for S_L is expressed by the following inequality

$$L(|x - y|) \geq L(|x - z|) * L(|z - y|)$$

or equivalently

$$L(a + b) \geq L(a) * L(b) \text{ for any } a \geq 0 \text{ and } b \geq 0.$$

Observe that the trivial fuzzy relations defined by the fuzzy sets L_1 and L_2, where $L_1(a) = 1$ for any a, and $L_2(0) = 1$, $L_2(a) = 0$ otherwise, satisfy the $*$-transitivity. It is easy to prove that when $*$ is min-conjunction or product-conjunction only these trivial fuzzy relations fulfil the $*$-transitivity. However, if $*$ is the Lukasiewicz-conjunction, we have:

- when $\delta > 0$, $L(a + b) \geq \max(0, L(a) + L(b) - 1)$ if and only if $L(a) = 1$ for any a,
- when $\delta = 0$, it is easy to check that all the fuzzy relations defined by the fuzzy sets of the family described in Figure A1.1 satisfy Lukasiewicz-transitivity. See (Klawonn and Kruse, 1993).

Following the second suggestion the relationship "x is close to y" (between real values) may be also understood as "the quotient x/y is close to 1", in the sense that the possible values of x/y are restricted by a fuzzy set P modelling "around 1". In other words, we define a proximity relation S_P such that

$$S_P(u,v) = P(u/v) \tag{4}$$

where P is a fuzzy number which satisfies $P(1) = 1$, $P(t) = P(1 / t)$ (to ensure the symmetry of S_P). Then, given two fuzzy relations S_P and $S_{P'}$, their composition results in a simple fuzzy arithmetic calculus on P and P'. Namely

$$\begin{aligned} S_P \circ S_{P'}(u,w) &= \sup_v \min(P(u / v), P'(v / w)) \\ &= \sup_{(s,t):s\cdot t=u/w} \min(P(s), P'(t)) = P \otimes P'(u / w) \end{aligned}$$

where $\otimes$ denotes the extension of product to fuzzy numbers. Thus,

$$S_P \circ S_{P'} = S_{P \otimes P'}. \tag{5}$$

This expresses that x is close to y in the sense of $S_{P \otimes P'}$. It accounts for the degradation of the transitivity since $P \otimes P \supset P$ and thus $S_{P \otimes P} \supset S_P$. Similarly, the composition of two fuzzy relations S_L and $S_{L'}$ of the form (2) amounts to the addition of two fuzzy numbers L and L', where $L \oplus L'$ is defined by

$$(L \oplus L')(|x - y|) = \sup_{s,t,u,v:\ |x-y|=|s-t|+|u-v|} \min(L(|s - t|), L'(|u - v|))$$

and can be easily computed (e.g., Dubois and Prade, 1987) since if L is defined by the two parameters δ and ε, and L' similarly by δ' and ε', then $L \oplus L'$ is defined by $\delta + \delta'$ and $\varepsilon + \varepsilon'$. Then the degradation of the transitivity is clearly expressed by the enlargement of L and L' into $L \oplus L'$ (expressed by the increase of the parameters).

Aggregation of Similarity Relations

If we assume that each feature is associated with an attribute domain equipped with a similarity relation modelling approximate equality on this domain, the problem is then to aggregate the degrees of similarity between the objects pertaining to each feature into a global similarity measure. This means that the resulting measures should still have properties like reflexivity, symmetry and $*$-transitivity. In general min-combination preserve reflexivity and symmetry properties. With respect to the transitivity, it is worth noticing that if S_1 and S_2 are $*$-transitive then $S_*((x_1,x_2), (y_1,y_2)) = S_1(x_1,y_1) * S_2(x_2,y_2)$ and $S_{min}((x_1,x_2), (y_1,y_2)) = \min(S_1(x_1,y_1), S_2(x_2,y_2))$ (for 2 features but it can be easily generalized to n features) are still $*$-transitive.

It is clear that 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 gives as a value the least discriminating one. Other well-known combination operations like averages, take values between min and max.

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 min operation, a weighted version can be given by

$$S(x,y) = \min_{i=1,n} \max(S_i(x_i,y_i), 1 - \lambda_i) \qquad (6)$$

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 *-transitivity are preserved by this weighted aggregation. Moreover separation is also preserved if $\lambda_i \neq 0$ for all i = 1, 2, …, n. See (Dubois and Prade, 1987) for a justification

of (6) and other weighted conjunctions. More generally, one could think of weights λ_i being a function of the particular values that some attributes can take, in order to capture situations in which one attribute may be more important than another one.

Appendix 2 - Fuzzy Gradual Rules and Sugeno and Nishida's Interpolation Method

Sugeno and Nishida (1985)'s method starts from rules with non-fuzzy conclusion parts of the form "if X is $A^{(i)}$ then $Y = b^{(i)}(x)$ (we do not deal here with compound conditions for simplicity), and computes the output y as the weighted sum

$$y = \sum_i \mu_{A^{(i)}}(x) \cdot b^{(i)}(x) / \sum_i \mu_{A^{(i)}}(x) \qquad (I)$$

which indeed looks like an interpolation. When $b^{(i)}(x)$ does not depend on x, this result can be obtained by applying Zadeh (1979)'s approximate reasoning combination and projection approach viewing the rules as *gradual rules* expressing that "the closer X is to $a^{(i)}$, the closer Y is to $b^{(i)}$", i.e., modelling them as inequality constraints of the form $A^{(i)}(u) \leq B^{(i)}(v)$. Then the subset of V obtained by combining the results of the rules for the input $X = x_0$ is given by $\min_{i=1,n} A^{(i)}(x_0) \rightarrow B^{(i)}(v)$ where the implication defined by $a \rightarrow b = 1$ if $a \leq b$ and $a \rightarrow b = 0$ if $a > b$ encodes the above interpretation of the rules. When the $A^{(i)}$'s and $B^{(i)}$'s make suitable fuzzy partitions of U and V respectively (it guarantees that $\forall u, \sum_i A^{(i)}(u) = 1$), which is in particular the case with the usual partitions made of triangular membership functions as in Figure A2.1, it can be shown that the fuzzy subset of V which is thus obtained is nothing but the singleton {y} computed by Sugeno's method (I). Let us explain the situation in more details.

Let us consider a collection of gradual rules of the form "the closer X is to a_i, the closer Y is to b_i" where (a_i,b_i), i = 1,n are pairs of *scalar* values. The first problem is to represent "close to a_i", by means of a fuzzy set A_i. It seems natural to assume that $A_i(a_{i-1}) = A_i(a_{i+1}) = 0$ since there are special rules adapted to the cases $X = a_{i-1}$, $X = a_{i+1}$. Moreover if $u \neq a_i$, $A_i(u) < 1$ for $u \in (a_{i-1},a_{i+1})$, since information is only available for $u = a_i$. Hence A_i should be a fuzzy interval with support (a_{i-1},a_{i+1}) and core $\{a_i\}$. Besides, by symmetry, since the closer x is to a_{i-1}, the farther it is from a_i, A_{i-1} should decrease when A_i increases, and $A_i\left(\frac{a_i + a_{i+1}}{2}\right) = A_{i-1}\left(\frac{a_{i-1} + a_i}{2}\right) = 0.5$. The simplest way of achieving this is to let

$$\forall\, u \in [a_{i-1},a_i],\ A_{i-1}(u) + A_i(u) = 1$$

an example of which are triangular-shaped fuzzy sets as in Figure A2.1.

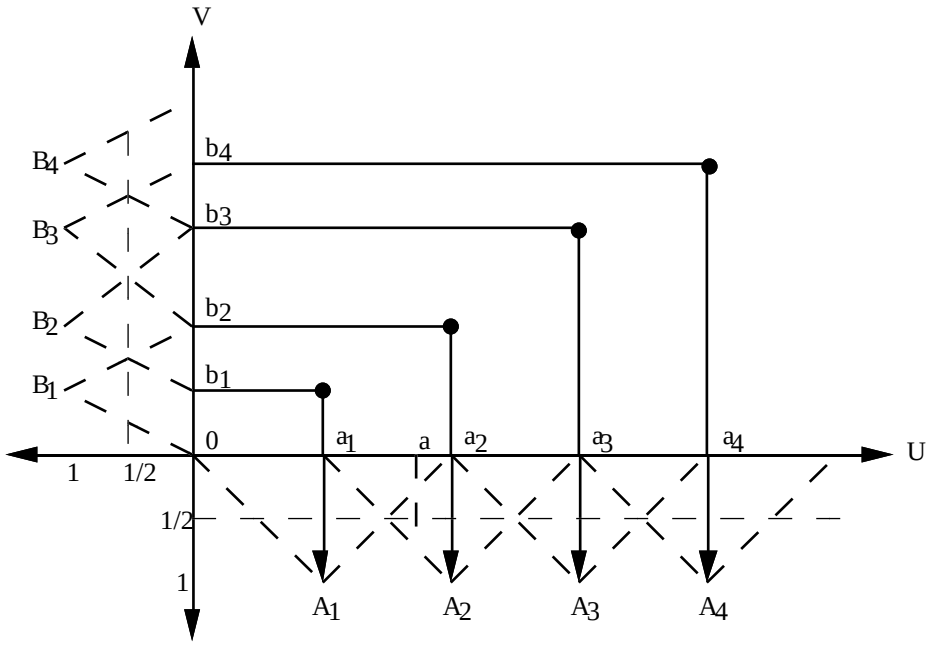

Figure A2.1

Clearly the conclusion parts of the rules should involve fuzzy sets B_i whose meaning is "close to b_i", with similar convention. In other words, each rule is understood as "the more X is A_i, the more Y is B_i", i.e., here "the closer X is to a_i, the closer Y is to b_i". In that case the output associated with the precise input X = a, where $a_{i-1} < a < a_i$ is

$$B = (\alpha_{i-1} \rightarrow B_{i-1}) \cap (\alpha_i \rightarrow B_i) = (B_{i-1})_{\alpha_{i-1}} \cap (B_i)_{\alpha_i}$$

where $\alpha \rightarrow B$ corresponds to the level cut B_α, where $\rightarrow$ is the implication defined above and $\alpha_{i-1} = A_{i-1}(a)$, $\alpha_i = A_i(a)$ with $\alpha_{i-1} + \alpha_i = 1$. Then it can be easily proved (without the assumption of a triangular shape) that there exists a unique value y = b such that B(b) = 1, which exactly corresponds to the result of a linear interpolation, i.e., we have $b = \alpha_{i-1} \cdot b_{i-1} + \alpha_i \cdot b_i$. See Figure A2.2.

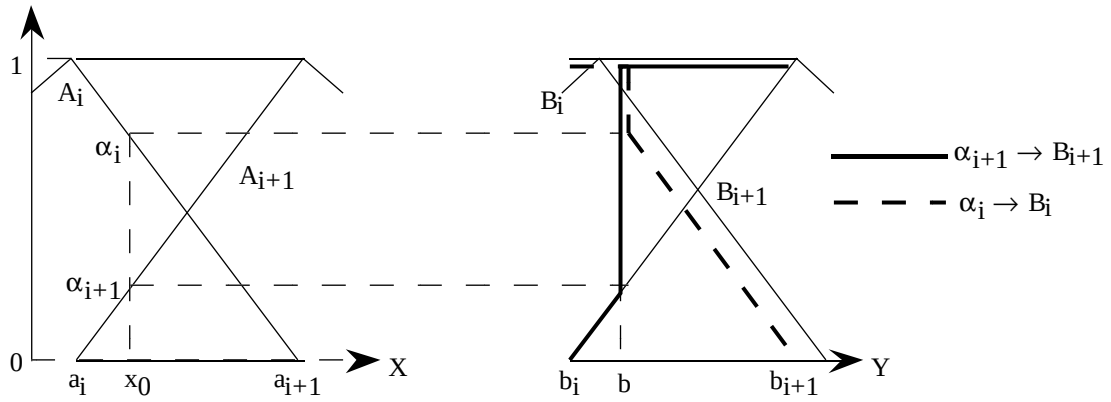

Figure A2.2: The core of $(\alpha_i \rightarrow B_i) \cap (\alpha_{i+1} \rightarrow B_{i+1})$ is the result of a linear interpolation

It is a theoretical justification for Sugeno and Nishida (1985)'s inference method where the

conclusion parts of the rules are precise values b_i and where a linear interpolation is performed on the basis of the degrees of matching $\alpha_i = A_i(a)$, in the particular case of single input rules. Hence reasoning with gradual rules does model interpolation, linear interpolation being retrieved as a particular case. The more complicated case of gradual rules with compound conditions, i.e., rules of the form "the more X_1 is A_1, …, and the more X_n is A_n, the more Y is B" is studied in details in (Dubois, Grabisch and Prade, 1994); in this more general case, where the "and(s)" in the compound condition are modelled by a fixed operation like 'product' or 'min', the gradual rule approach only gives a bracketting of the result obtained by Sugeno and Nishida' method.