



HAL
open science

Fuzzy set modelling in case-based reasoning

Didier Dubois, Henri Prade, Francesc Esteva, Pere Garcia

► **To cite this version:**

Didier Dubois, Henri Prade, Francesc Esteva, Pere Garcia. Fuzzy set modelling in case-based reasoning. *International Journal of Intelligent Systems*, 1998, 13 (4), pp.345–373. 10.1002/(SICI)1098-111X(199804)13:43.0.CO;2-N . hal-04028797

HAL Id: hal-04028797

<https://hal.science/hal-04028797>

Submitted on 17 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fuzzy Set Modelling in Case-Based Reasoning*

Didier Dubois and Henri Pradet, ‡

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

Francesc Esteva, Pere Garcia, Lluís Godo, and Ramon López de Màntaras
Institut d'Investigació en Intel·ligència Artificial (IIIA), Consell Superior d'Investigacions Científiques (CSIC), Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

This paper is an attempt at providing a fuzzy set formalization of case-based reasoning and decision. Learning aspects are not considered here. The proposed approach 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. © 1998 John Wiley & Sons, Inc.

I. 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 ,

*This paper is a revised and expanded version of a conference paper.¹

†Author to whom correspondence should be addressed.

‡The authors should be cited alphabetically, in spite of this address-based presentation.

knowing A , B , and C . Analogical reasoning was investigated for a long time in *Artificial Intelligence*^{2,3,4} and the interest in this research was considerably renewed by the development of case-based reasoning and analogical reasoning; see Kolodner,⁵ Aamodt and Plaza,⁶ Py⁷ for surveys. 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 were rather few attempts for introducing fuzzy set-based tools in analogical reasoning until recently, up to some exceptions.^{8,9,10,11,12} However, in case-based reasoning, some works focused on the handling of fuzzy descriptions in the retrieval step,^{13,14,15} on the learning of fuzzy concepts from fuzzy examples,¹⁶ on the integration with rule-based reasoning,¹⁷ and very recently on the logical modelling of the inference mechanisms based on similarity measures.¹⁸ See Dubois and Prade¹⁹ 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 nonempty disjoint subsets: the problem description attributes subset, and the solution or outcome attributes subset, denoted by \mathcal{S} and \mathcal{T} , respectively. These subsets are defined 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. 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 by induction building rules 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 holds, stating that “similar situations give (or may give) similar outcomes.” 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 previously stated, 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 was emphasized for a long time in analogical reasoning, see, e.g., Bourrelly et al.²⁰ In particular, Davies and Russell,²¹ and Russell²² 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 also be identical with respect to (an)other attribute(s) \mathcal{T} . However, this form of metarule 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.

Prade,⁹ Arrazola et al.,²³ and Léa Sombé²⁴ suggested some remedies to these two limitations 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.

In this work, the similarity measure between cases is given by two fuzzy 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 A 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 *nondeterministic* 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 strong 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 II describes the fuzzy set modelling of deterministic and nondeterministic case-based reasoning problems. Section III discusses case-based decision making. Section IV provides a logical formalization of the deterministic case. Concluding remarks briefly discuss various extensions of the proposed approach.

II. A FUZZY SET FRAMEWORK FOR CASE-BASED REASONING AND DECISION

Fuzzy rules (see Dubois and Prade²⁵ for an introduction) provide a tool for modelling the above expressions of the implicit case-based reasoning principle. We first consider the deterministic situation.

A. 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

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

where S and T are fuzzy relations§ (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

§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 $\mu_F(u)$.

$[\forall s_1, S(s_1, s_1) = 1, T(t_1, t_1) = 1]$. Technically speaking, such a relation is named a *proximity relation* in the fuzzy set literature. In the following, we shall often refer to it as a *similarity relation*, although in its technical sense a fuzzy similarity relation is supposed to enjoy some kind of generalized transitivity (an assumption which is not required here for S and T).

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 Prade²⁶ and Prade.²⁷ In particular, $\forall \alpha \in (0, 1]$, we have

$$(s_1, s_2) \in S_\alpha \Rightarrow (t_1, t_2) \in T_\alpha \quad (2)$$

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

$$s_1 = s_2 \Rightarrow t_1 = t_2 \quad (3)$$

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,²⁸ Dubois and Prade,²⁹ Bosc et al.³⁰), where the separating property of T is often hypothesized.

Let us now 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

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

Note that E may be empty if T is not permissive enough. For instance, assume that $\mathcal{S} = \mathbb{R}$, 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 Figure 1, $M = \{(7, 10), (3, 12)\}$ and $s_0 = 5$. Then $S(s_0, 7) = S(s_0, 3) = \frac{2}{3}$. Thus, t_0 according to case (7, 10) is such that $t_0 = 10 \pm 0.66$ while according to case (3, 12) we get $t_0 = 12 \pm 0.66$, and thus $E = \emptyset$. When $E \neq \emptyset$, an interpolation mechanism is embedded in (2); see Appendix B where it is shown how this mechanism can be captured by fuzzy gradual rules. The nonemptiness of E can be guaranteed by the coherence of the set of fuzzy

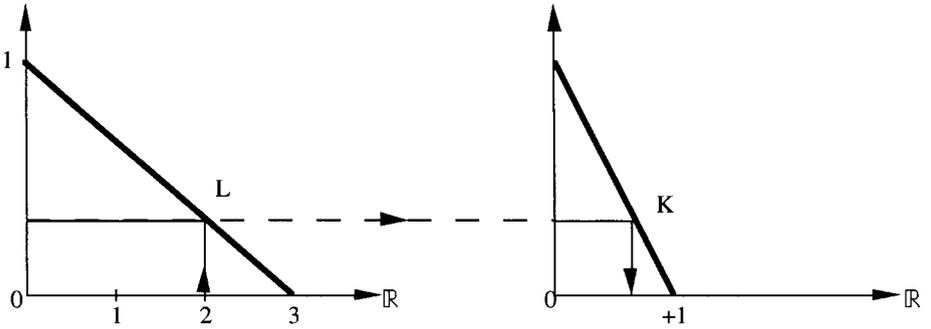


Figure 1.

gradual rules which can be built from M , S , and T . Namely, these fuzzy rules are obtained by stating that for each $(s, t) \in M$, 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, $\forall (s, t) \in M, \forall s', \forall t', S(s, s') \leq T(t, t')$. The coherence of this set of gradual rules is equivalent to the following condition:

$$\forall N \subseteq M, \text{ and for all families of positive values } \{\alpha(s, t), (s, t) \in N\} \text{ in the unit interval, } \bigcap_{(s, t) \in N} (S(s))_{\alpha(s, t)} \neq \emptyset \text{ implies } \bigcap_{(s, t) \in N} \times (T(t))_{\alpha(s, t)} \neq \emptyset \quad (5)$$

[with $S(s)$ [resp. $T(t)$] being the fuzzy set of values in \mathcal{S} (resp. in \mathcal{T}) close to s (resp. t)]. Equation (5) clearly guarantees the nonemptiness of E . Dubois, Prade, and Ughetto³¹ address 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 supply practical checking procedures.

The condition (5) has a major influence on the building of a memory. Two options can be envisaged: the similarity relations S and T are prescribed and augmenting M with a new case c is possible only if $M \cup \{c\}$ verifies (5). Alternatively, the similarity relations are updated each time a case c is entered into M , so as to maintain the coherence requirement (5). This topic is out of the scope of the paper.

B. Nondeterministic 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 the coherence condition (5) in order to ensure that E is not 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 $t \neq t'$, 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.” A proper understanding of this principle requires that the intended meaning of “possible” in this metarule be clarified; this will be done in the next paragraph. A nondeterministic (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 somewhat distinct with respect to outcome attribute(s). This emphasizes that, in this type of situations, 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.”³² 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

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

When both A and B are fuzzy sets it generalizes into

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

This clearly gives back the above expression when $B(v) \in \{0, 1\}$. Dubois and Prade²⁵ provide a full justification of (6) by decomposing B into nonfuzzy level cuts $B_\alpha = \{v, B(v) \geq \alpha\}$. This model of fuzzy rule is close to Mamdani’s³³ 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

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

Clearly, what is obtained is the fuzzy set $\{t\} \circ T = T(t, \cdot)$ of values t' which are T similar to t , “truncated” from above 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, a much stronger kind of 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 by so-called certainty rules.³² 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 neighborhood 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). However, $S(s, s_0)$ has no influence here on the neighborhood of values around t which is considered; this contrasts with (4) where the inequality $S(s, s_0) \leq T(t, t_0)$, $\forall (s, t) \in M$, constrains the similarity to t of the possible values of t_0 .

III. CASE-BASED DECISION

A. 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 Schmeidler³⁴ advocated a similarity-based approach to decision where a case is described as a triple (*problem, act, result*) and where a decision-maker's nonnegative 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 nonnegative 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 if two cases (s, a, r) and (s', a', r') belong to M , then if $s = s'$ and $a = a'$ then $r = r'$ holds (i.e., results are uniquely determined by the act applied to the context of a given problem), and $\forall (s, a, r) \in M, \exists ! a$ such that $u(r) \neq 0$ (i.e., it means that only the best act in context s is stored in the memory). Gilboa and Schmeidler³⁴ give an axiomatic derivation of this U maximization, within a formal model.

The case-based reasoning framework of the previous section and the above approach to decision problems agree in a particular case. Let us regard the triples (s, a, r) as cases $((s, a), u(r))$ and assume $S((s, a), (s_0, a)) = S(s, s_0)$. It is then possible to relate (12) to the approach of Section II A 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

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

Then assume that the set of situations s in M where a given act a was experienced, can be linearly ordered into $s_1, \dots, s_{i-1}, s_i, s_{i+1}, \dots, s_n$. We suppose that there exists a fuzzy proximity relation S such that for s_0 , there exists only one pair of situations (s_i, s_{i+1}) such that $(s_i, a, r_i) \in M, (s_{i+1}, a, r_{i+1}) \in M, S(s_0, s_i) > 0, S(s_0, s_{i+1}) > 0$ and moreover that $S(s_0, s_i) + S(s_0, s_{i+1}) = 1$. Note that it entails $S(s_i, s_{i+1}) = 0$ (since S is reflexive and s_0 can be anywhere in between s_i and s_{i+1}). However, the ordering of the situations s_i in M (associated with an act a) somewhat accounts for an idea of proximity between them (in the sense that s_i can be considered as closer to s_{i-1} and s_{i+1} than to s_{i-k} or s_{i+k} for $k > 1$). We can then associate with S a fuzzy relation T such that $T(t, u(r_i)) + T(t, u(r_{i+1})) = 1, \forall t$. Then the set E_a defined by (13) contains only one element e_a equal to

$$e_a = \sum_{(s, a, r) \in M} S(s, s_0) \cdot u(r) = S(s_i, s_0) \cdot u(r_i) + S(s_{i+1}, s_0) \cdot u(r_{i+1})$$

Thus, a particular case of (12) can be obtained by the fuzzy functional dependency method of Section II A. See Appendix B for a detailed justification. This is due to the fact that for particular S and T , (13) embeds a linear interpola-

tion mechanism. Note that one way of satisfying the requirement $S(s_0, s_i) + S(s_0, s_{i+1}) = 1$ for S when the problems s in M can be linearly ordered, is in fact to assume that S is locally defined in the neighborhood of each problem s by S_s , in such a way that $S_s(s, s_0)$ decreases to 0 when s_0 goes away from s and coincides with one of the two closest neighbors of s in M . Then it is easy to make $S_{s_i}(s_0, s_i) + S_{s_{i+1}}(s_0, s_{i+1}) = 1$. Equation (13) could be more generally used instead of (12), for estimating the worth of act a in situation s_0 . However, a set-valued utility is obtained.

Note that (12) is also akin to the fuzzy vote procedure used by Bensana et al.³⁵ 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

$$U(a) = \sum_{i: a=a^{(i)}} S^{(i)}(s_0) \cdot w^{(i)} \quad (14)$$

where $S^{(i)}$ is a fuzzy set describing a fuzzy class of situations. At the reasoning level, the above decision procedures (12) and (14) 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 IV (which encodes the approach of Section II A).

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, maximizing (12) is not equivalent to maximizing

$$U'_{s_0, M}(a) = \frac{\sum_{(s, a, r) \in M} S(s, s_0) \cdot u(r)}{\sum_{(s, a, r) \in M} S(s, s_0)} \quad (15)$$

since the normalizing factor depends on act a .

The underlying idea in (12) is to look for acts that in several similar situations had results with a high utility. However, a drawback of (12) is its inability to distinguish between an act for which there is only one case in M very similar to s_0 which led to a good result, and an act for which there are many cases in M , all of them with a high degree of similarity to s_0 but which led to rather poor results: this may then yield estimates $U_{s_0, M}(a)$ which favor the second act. Namely, we cannot distinguish between the two following types of “extreme” situations where it is possible to have $U_{s_0, M}(a_1) < U_{s_0, M}(a_2)$:

- (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 $(s, a_1, r) \in M$;

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

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).

B. Alternative Approach

Another approach to case-based decision 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 utility function u , like the similarity 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 excellent ones (the best one we may imagine), 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 G of good results. Under this scaling hypothesis, note that (12) is nothing but the scalar fuzzy cardinality³⁶ of the intersection of the fuzzy set $S_a(s_0)$ of situations similar to s_0 where a was applied and of the fuzzy set G_a of situations where good results were obtained for act a (using the product for defining intersection pointwisely). In case we would normalize (12) into (15), as previously suggested, we would compute a genuine inclusion degree of $S_a(s_0)$ into G_a , of the form $|S_a(s_0) \cap G_a|/|S_a(s_0)|$ (where $|A|$ denotes the cardinality of A). However, more qualitative inclusion degrees are also worth considering.

Let us introduce the following type of degree of inclusion which also enables us to select the act(s) a , if any, which for problems similar to s_0 *always* gave good results,

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

where (16) is a multiple-valued implication connective (i.e., $x \rightarrow y$ increases with y , decreases with x , and coincides with material implication on $\{0, 1\}$). It seems natural to require that

- $U_{*s_0, M}(a) = 1$ 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 $u(r) = 0$.

Thus, act a receives the greatest rate only if it led to results with the maximum satisfaction degree (i.e., 1) in all the situations which have a nonzero similarity degree with s_0 (and where a was applied). The act a receives the smallest rate if there exists a situation s fully similar to s_0 [i.e., $S(s_0, s) = 1$] where a completely fails [i.e., $u(r) = 0$]. This suggests 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. It can be checked that it is enough for having the above requirements satisfied. In case of a purely ordinal use of $[0, 1]$, where only the ordering of the levels is meaningful, we are led to use $x \rightarrow y = \max(1 - x, y)$, i.e.,

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

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 .

$U_{*s_0, M}$ is a rather drastic criterium since it requires that in *all* the problems similar to s_0 , act a led to good results. A more “optimistic” behavior 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

$$U_{s_0, M}^*(a) = \max_{(s, r): (s, a, r) \in M} \min(u(r), S(s, s_0)) \quad (18)$$

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.

Interestingly enough, (17) and (18) can be viewed as the counterparts of qualitative expectations recently introduced in decision making under uncertainty,³⁷ just as (12) may be seen as a kind of counterpart of the classical expected utility.

Expressions (17) and (18) make sense for case-based decision only when they take into account acts a that are experienced in at least some situation s such that $S(s, s_0) = 1$. Otherwise, it could happen that $U_{*s_0, M}(a)$ is very high while $S(s, s_0)$ is very low for all triples $(s, a, r) \in M$. To overcome this problem, a possible way is, for each act a , to renormalize[¶] the fuzzy set $S_a(s_0)$ over the situations s such that $(s, a, r) \in M$. But to keep track of the absence of situations very similar to s_0 , a soft discounting of (17) and (18) [once $S(\cdot, s_0)$ is already renormalized] should be also performed. Namely, letting $h_s(a, s_0) =$

[¶]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.

[¶]There are several ways of renormalizing a fuzzy set F . If the scale is numerical the usual option is to define $F'(u) = F(u) / [\max\{F(u) | u \in U\}]$, while if the scale is qualitative we may define $F''(u) = F(u)$ for $u \neq u_0$ and $F''(u_0) = 1$, where $F(u_0) = \max\{F(u) | u \in U\}$.

$\max\{S(s, s_0) | (s, a, r) \in M\}$, we can define

$$\begin{aligned}\overline{U}_{*s_0, M}(a) &= \min(h_S(a, s_0), U_{*s_0, M}^{\text{norm}}(a)) \\ \overline{U}_{s_0, M}^{\#}(a) &= \max(1 - h_S(a, s_0), U_{s_0, M}^{*\text{norm}}(a))\end{aligned}\quad (19)$$

where $U_{*s_0, M}^{\text{norm}}$ and $U_{s_0, M}^{*\text{norm}}$ are computed by applying (17) and (18) once $S(\cdot, s_0)$ is renormalized. This preserves the inequality $U_{s_0, M}^* \geq U_{*s_0, M}$ between the pessimistic and the optimistic utilities [which only holds if $S(\cdot, s_0)$ is normalized], i.e., we have $\overline{U}_{s_0, M}^{\#} \geq \overline{U}_{*s_0, M}$. Thus $\overline{U}_{s_0, M}^{\#}(a)$ is maximal as soon as

- (i) There exists a case corresponding to a problem completely similar to s_0 where the act a led to an excellent result; or
- (ii) Act a has never been applied to a situation somewhat similar to s_0 (so if we are optimistic we should consider it).

On the other hand, $\overline{U}_{*s_0, M}(a)$ is 0 when there is no situation similar to s_0 in M . Dubois and Prade give some more details on this proposal elsewhere.³⁸

This approach can be also related to the approach proposed in Section II B for the nondeterministic 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 letting T be the classical identity relation. Then, (17) and (18) can be also viewed as scalar summaries, which are pessimistic and optimistic, respectively, of the fuzzy set of utility values whose membership function is given by (21). Indeed (17) and (18) are weighted median operations.³⁹

Besides, it would be 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, e.g., Dubois and Prade⁴⁰ describe a way of introducing such a quantified procedure for a different, but formally analogous problem.

IV. A SIMILARITY LOGIC SETTING FOR CASE-BASED REASONING

In this section we propose a setting, closer to logical formalisms than the one described in Section II A, for the inference processes that take place in case-based reasoning systems when solving deterministic problems. The approach is inspired by the logical formulation of case-based inference proposed by Plaza et al.,¹⁸ which in turn relies on two kinds of graded consequence

relations that play a major role in similarity-based reasoning as shown by Dubois et al.^{41,42} For another attempt at formalizing case-based reasoning inference in a logical setting, see the paper by Lieber and Marquis.⁴³

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. However, before that we introduce the notions of approximate and proximity entailments due to Dubois et al.⁴² which pertain to the inference steps (i) and (ii), respectively.

A. Background

Let \mathcal{L} be a finitely generated propositional language, Ω its set of Boolean interpretations, and let S be a similarity relation on Ω . Ruspini⁴⁴ defines the degree $I_S(p|\omega)$ to which an interpretation ω is close to some model ω' of a proposition p ($\omega' \models p$), i.e.,

$$I_S(p|\omega) = \sup_{\omega' \models p} S(\omega, \omega') \quad (22)$$

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

$$[p^*](\omega) = I_S(p|\omega) \quad (23)$$

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, the following graded *approximate entailment*^{41,42} is defined by

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

and more generally,

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

which can be equivalently expressed as $q \models^\alpha p$ iff $I_S(p|q) \geq \alpha$, where I_S is generalized for pairs of propositions by $I_S(p|q) = \inf_{\omega \models q} I_S(p|\omega)$. It 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

$$q \models_K^\alpha p \quad \text{iff } q \wedge K \models^\alpha p \quad (26)$$

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 II A of a case as a gradual fuzzy rule “the more ω in $[p^*]$, the more ω in $[q^*]$,” which is therefore expressed as

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

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

$$p \models_K^\alpha q \quad \text{iff } [p^*](\omega) \otimes [q^*](\omega) \geq \alpha, \quad \text{for all models } \omega \text{ of } K \quad (28)$$

It is worth mentioning that \models_K^α and \models_K^α 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,

$$\text{if } p' \models^\alpha p \text{ and } p \models_K^\beta q \text{ then } p' \models_K^{\alpha \otimes \beta} q \quad (29)$$

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 $\alpha \otimes \beta$ -true.

B. Logical Modelling of Case-Based Inference

Now we exploit the logical framework just introduced in the previous subsection to reformulate the inference process involved in the case-based reasoning problems presented in Section II A. 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

#The use of a residuated implication ($x \otimes \rightarrow y = \sup\{z \in [0, 1], x \otimes z \leq y\}$) is also in accordance with the semantics for the so-called gradual fuzzy rules.^{26,32} See Appendix A for some background on t -norms \otimes .

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_S and Ω_T the set of interpretations for $L_{\mathcal{S}}$ and $L_{\mathcal{T}}$, respectively, and we define $\Omega = \Omega_S \times \Omega_T$. 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_S and outcome descriptions with interpretations of Ω_T .

According to the previous sections, we assume to have a similarity S on Ω_S 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** of $S \times T$ on Ω_S 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 Ω_S and Ω_T , respectively, step (i) amounts to computing 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 $K = \{(s, t)[s_i^*](s) \otimes \rightarrow [t_i^*](t) \geq \beta_i\} = \{(s, t) | S(s_i, s) \otimes \rightarrow T(t_i, t) \geq \beta_i\}$, where the parameters β_i are usually taken to be 1 (this is the situation that exactly corresponds to the expression (1) of our deterministic principle††), but in general they could be assigned a lower value, denoting a weaker form of gradualness between s_i and t_i . Then, the whole inference process described by (i) and (ii) can be formulated as follows,

$$\begin{aligned} \text{from } s_0 &\models^{\alpha_i} s_i \\ & s_i \Vdash_K^{\beta_i} t_i \\ \text{infer } s_0 &\models_K^{\alpha_i \otimes \beta_i} t_i \end{aligned}$$

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

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

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

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

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

V. CONCLUSIONS AND FURTHER WORK

In this paper we are 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 proximity relations both between problem descriptions and between outcomes of the cases. These proximity relations allow us to give a precise meaning to the implicit principle underlying deterministic case-based reasoning, i.e., “the more similar the problem descriptions, the more similar their outcomes,” by interpreting each case in the memory as a gradual fuzzy rule. This interpretation was also formalized in a logical setting by means of similarity-based graded consequence relations. Furthermore, a weaker form of the above principle was proposed in order to cope with “nondeterministic” problems. Fuzzy possibility rules were then used instead of gradual rules to model these nondeterministic 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.²⁴ 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

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

and

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

respectively. 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 $\prod(S) = N(S) = S(u_0, v_0)$, 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 $\Pi(S)$ and $N(S)$ are easy to compute for trapezoidal possibility distributions, and can be aggregated by min operation (or its weighted version; see Appendix A), still keeping their possibility and necessity semantics. See Dubois and Prade⁴⁵ for details.

- (iii) *Incorporating learning capabilities*: as we 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 by Plaza et al.⁴⁶ 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 II A, as well as the logical model proposed in Section IV, incorporate a natural interpolation mechanism (see Dubois et al.^{41,42}) which may be very helpful for the case adaptation step, especially when the outcome attribute domains are of numerical nature.

The authors are grateful to Enric Plaza 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 were also partially supported by the E. U. Copernicus project MUM (No. 10053).

References

1. D. Dubois, F. Esteva, P. Garcia, L. Godo, R. López de Màntaras, and H. Prade, "Fuzzy modelling of case-based reasoning and decision," *Proceedings of the Second International Conference on Case-Based Reasoning (ICCB'97)*, Providence, July 1997, Lecture Notes in Computer Sciences 1266, Springer-Verlag, Berlin, pp. 599–610.
2. R.R. Kling, "A paradigm for reasoning by analogy," *Artificial Intelligence*, **2**, 147–178 (1971).
3. J. Carbonell, "A computational model of problem solving by analogy," *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI'81)*, Vancouver, 1981, pp. 147–152.
4. D. Gentner, "Structure mapping: A theoretical framework for analogy," *Cognitive Science*, **7**, 155–170 (1983).
5. J. Kolodner, *Case-Based Reasoning*, Morgan Kaufmann, San Mateo, CA, 1993.
6. A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations and system approaches," *Artificial Intelligence Communications*, **7**, 39–59 (1994).
7. M. Py, "Un modèle conceptuel de raisonnement par analogie," *Revue d'Intelligence Artificielle* (Hermès, Paris), **8**, 63–99 (1994).
8. H. Farreny and H. Prade, "About flexible matching and its use in analogical reasoning," *Proceedings of the 1982 European Conference on Artificial Intelligence (ECAI'82)*, Orsay, France, July 12–14, 1982, pp. 43–47.
9. H. Prade, "Analogie et flou" and "Analogie, flou et systèmes experts," *BUSEFAL* (IRIT, Univ. P. Sabatier, Toulouse, France), **18**, 83–91 (1984); **32**, 108–114 (1987).
10. I.B. Turksen and Z. Zhong, "An approximate analogical reasoning approach based on similarity measures," *IEEE Transactions on Systems, Man and Cybernetics*, **18**, 1049–1056 (1988).

11. B. Bouchon-Meunier and L. Valverde, "Analogy relations and inference," *Proceedings of the Second IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'93)*, San Francisco, CA, March 28–April 1, 1993, pp. 1140–1144.
12. R.R. Yager, "A unified view of case based reasoning and fuzzy modelling," in *Fuzzy Logic Foundations and Industrial Applications*, Da Ruan, Ed., Kluwer Academic, Boston, 1996, pp. 5–26.
13. S. Salotti, "Représentation centrée objet et filtrage flou pour raisonner par analogie: Le système FLORAN," *Actes du 7ème Congrès "Reconnaissance des Formes et Intelligence Artificielle" (RFIA'89)*, Paris, 29 nov.–1er déc., 1989, pp. 1695–1707.
14. S. Salotti, "Filtrage flou et représentation centrée objet pour raisonner par analogie: Le système FLORAN," (in French), Ph.D. thesis, University of Paris XI, Orsay, France, 1992.
15. M. Jaczynski and B. Trousse, "Fuzzy logic for the retrieval step of a case-based reasoner," *Proceedings of the European Workshop on Case-Based Reasoning'94*, 1994, pp. 313–321.
16. E. Plaza and R. López de Màntaras, "A case-based apprentice that learns from fuzzy examples," in *Methodologies for Intelligent Systems*, Z.W. Ras, M. Zemankova, and M.L. Emrich, Eds., Elsevier, 1990, Vol. 5, pp. 420–427.
17. S. Dutta and P.P. Bonissone, "Integrating case- and rule-based reasoning," *International Journal of Approximate Reasoning*, **8**, 163–203 (1993).
18. E. Plaza, F. Esteva, P. Garcia, L. Godo, and R. López de Màntaras, "A logical approach to case-based reasoning using fuzzy similarity relations," *Information Sciences*, to appear.
19. D. Dubois and H. Prade, "Similarity-based approximate reasoning," in *Computational Intelligence Imitating Life (Proceedings of the IEEE Symposium, Orlando, FL, June 27–July 1, 1994)*, J.M. Zurada, R.J. Marks II, and X.C.J. Robinson, Eds., IEEE Press, New York, 1994, pp. 69–80.
20. L. Bourrelly, E. Chouraqui, and M. Ricard, "Formalisation of an approximate reasoning: The analogical reasoning," *Proceedings of the International Federation of Automatic Control Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis*, Marseille, 1983, pp. 19–21.
21. T.R. Davies and S.J. Russell, "A logical approach to reasoning by analogy," *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI'87)*, Milano, Italy, Aug. 23–28, 1987, pp. 264–270.
22. S.J. Russell, *The Use of Knowledge in Analogy and Induction*, Pitman, London, 1989.
23. I. Arrazola, A. Plainfossé, H. Prade, and C. Testemale, "Extrapolation of fuzzy values from incomplete data bases," *Information Systems*, **14**, 487–492 (1989).
24. Léa Sombé (Group), "Reasoning by analogy," in *Reasoning Under Incomplete Information in Artificial Intelligence*, Léa Sombé, Ed., Wiley, New York, 1990, pp. 418–424.
25. D. Dubois and H. Prade, "Fuzzy rules in knowledge-based systems—Modelling gradedness, uncertainty and preference—," in *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, R.R. Yager and L.A. Zadeh, Eds., Kluwer Academic, Dordrecht, 1992, pp. 45–68.
26. D. Dubois and H. Prade, "Gradual inference rules in approximate reasoning," *Information Sciences*, **61**, 103–122 (1992).
27. H. Prade, "Raisonner avec des règles d'inférence graduelle—Une approche basée sur les ensembles flous," *Revue d'Intelligence Artificielle*, **2** (Hermès, Paris), 29–44 (1988).
28. K.V.S.V.N. Raju and A.K. Majumdar, "Fuzzy functional dependencies and lossless joint decomposition of fuzzy relational database systems," *ACM Transactions on Database Systems*, **13**, 129–166 (1988).
29. D. Dubois and H. Prade, "Certainty and uncertainty of (vague) knowledge and generalized dependencies in fuzzy data bases," *Proceedings of the International Fuzzy Engineering Symposium (IFES'91)*, Yokohama, Japan, Nov. 13–15, 1991, pp. 239–249.

30. P. Bosc, D. Dubois, and H. Prade, "Fuzzy functional dependencies—An overview and a critical discussion—," *Proceedings of the Third IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'94)*, Orlando, June 26–29, 1994, pp. 325–330.
31. D. Dubois, H. Prade, and L. Ughetto, "Coherence of fuzzy knowledge bases," *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'96)*, New Orleans, LA, Sept. 8–11, 1996, IEEE Press, Englewood Cliffs, 1996, pp. 1858–1864.
32. D. Dubois and H. Prade, "What are fuzzy rules and how to use them," *Fuzzy Sets and Systems*, **84**, 169–185 (1996).
33. E.H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic systems," *IEEE Transactions on Computers*, **26**, 1182–1191 (1977).
34. I. Gilboa and D. Schmeidler, "Case-based decision theory," *The Quarterly Journal of Economics*, **Aug.**, 607–639 (1995).
35. E. Bensana, G. Bel, and D. Dubois, "OPAL: A multi-knowledge-based system for industrial job-shop scheduling," *International Journal of Production Research*, **26**, 795–819 (1988).
36. A. De Luca and S. Termini, "A definition of a non-probabilistic entropy in the setting of fuzzy sets theory," *Information Control*, **20**, 301–312 (1972).
37. D. Dubois and H. Prade, "Possibility theory as a basis for qualitative decision theory," *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, Montréal, Canada, Aug. 20–25, 1995, pp. 1924–1930.
38. D. Dubois and H. Prade, "A fuzzy set approach to case-based decision," *Proceedings of the Second European Workshop on Fuzzy Decision Analysis and Neural Networks for Management, Planning and Optimization (EFDAN'97)*, R. Felix, Ed., Dortmund, Germany, June 10–11, 1997, Fuzzy Demonstration-Zentrum, Joseph-von-Fraunhofer Str. 20, D-44227, 1997, pp. 1–9.
39. D. Dubois and H. Prade, "Weighted minimum and maximum operations in fuzzy set theory," *Information Sciences*, **39**, 205–210 (1986).
40. D. Dubois and H. Prade, "Semantics of quotient operators in fuzzy relational databases," *Fuzzy Sets and Systems*, **78**, 89–93 (1996).
41. D. Dubois, F. Esteva, P. Garcia, L. Godo, and H. Prade, "Similarity-based consequence relations," in *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Proceedings of the European Conference ECSQARU'95, Fribourg, Switzerland, July 1995)*, C. Froidevaux, and J. Kohlas, Eds., Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin, 1995, Vol. 946, pp. 171–179.
42. D. Dubois, F. Esteva, P. Garcia, L. Godo, and H. Prade, "A logical approach to interpolation based on similarity relations," *International Journal of Approximate Reasoning*, **17**, 1–36 (1997).
43. J. Lieber and P. Marquis, "Domain-independent similarity relations for case-based reasoning in a logical framework," *Proceedings of the Poster Session of the Ninth International Symposium on Methodologies for Intelligent Systems (ISMIS'96)*, Zakopane, Poland, June 9–13, 1996, pp. 230–241.
44. E.H. Ruspini, "On the semantics of fuzzy logic," *International Journal of Approximate Reasoning*, **5**, 45–88 (1991).
45. D. Dubois and H. Prade, *Possibility Theory—An Approach to Computerized Processing of Uncertainty*, Plenum, New York, 1988.
46. E. Plaza, R. López de Mántaras, and E. Armengol, "On the importance of similitude: An entropy-based assessment," *Advances in Case-Based Reasoning (Proceedings of the Third European Workshop EWCBR'96, Lausanne, Switzerland, Nov. 1996)*, Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin, 1996, Vol. 1168, pp. 324–338.
47. L.A. Zadeh, "Similarity relations and fuzzy orderings," *Information Sciences*, 177–200 (1971).
48. S.V. Ovchinnikov, "Similarity relations, fuzzy partitions, and fuzzy orderings," *Fuzzy Sets and Systems*, **40**, 107–126 (1991).

49. E. Rosch and C.B. Mervis, "Family resemblances studies in the internal structure of categories," *Cognitive Psychology*, **7**, 573–605 (1975).
50. I. Niiniluoto, "Analogy and similarity in scientific reasoning," in *Analogical Reasoning*, D.H. Helman, Ed., Kluwer Academic, 1988, pp. 271–298.
51. F. Klawonn and R. Kruse, "Equality relations as a basis for fuzzy control," *Fuzzy Sets and Systems*, **54**, 147–156 (1993).
52. D. Dubois and H. Prade, "Additions of interactive fuzzy numbers," *IEEE Transactions on Automatic Control*, **26**, 926–936 (1981).
53. J. Fodor and M. Roubens, *Fuzzy Preference Modelling and Multicriteria Decision Support*, Kluwer Academic, Dordrecht, 1994.
54. M. Sugeno and M. Nishida, "Fuzzy control of model car," *Fuzzy Sets and Systems*, **16**, 103–113 (1985).
55. L.A. Zadeh, "A theory of approximate reasoning," in *Machine Intelligence*, J.E. Hayes, D. Michie, and L.I. Mikulich, Eds., Elsevier, New York, 1979, Vol. 9, pp. 149–194.
56. D. Dubois, M. Grabisch, and H. Prade, "Gradual rules and the approximation of control laws," in *Theoretical Aspects of Fuzzy Control*, H.T. Nguyen, M. Sugeno, R. Tong, and R.R. Yager, Eds., Wiley, New York, 1994, pp. 147–181.

APPENDIX A: SIMILARITY MODELING BASED ON FUZZY RELATIONS

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 each feature.

In this appendix we introduce the notion of fuzzy proximity and similarity relations, and we study some aspects of the representation of a proximity relation (which pertain to their elicitation), and the aggregation of a family of similarity relations.

While reflexivity and symmetry are minimal properties that are clearly required when evaluating the closeness of cases (thus, by proximity relations), (extended) transitivity is not always compulsory in case based reasoning. Indeed, in case-based reasoning, given a current situation s_0 , in order to retrieve the most similar cases from the repertory M , we estimate the similarity of s_0 with each of the situations s_i in M pairwise, but we do not compute the similarity between s_0 and s_i 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.

However, for the sake of completeness, our refresher on fuzzy proximity and similarity relations contains a discussion of the modelling of various forms of transitivity (including a softening thereof) in the fuzzy setting.

Basic Definitions

The concept of similarity relation is essentially a generalization of the concept of an equivalence relation.⁴⁷ A similarity relation S , in the sense of fuzzy set literature, is a binary fuzzy relation defined on a set U satisfying the three

following properties

- (i) $\forall u \in U, S(u, u) = 1$.
- (ii) $\forall u \in U, \forall v \in U, S(u, v) = S(v, u)$, (symmetry).
- (iii) $\forall u \in U, \forall v \in U, \forall w \in U, S(u, v) \otimes S(v, w) \leq S(u, w)$, (\otimes -transitivity).

where \otimes is a binary operation defined on $[0, 1]$ such that $0 \otimes 0 = 0 \otimes 1 = 1 \otimes 0 = 0$, $1 \otimes 1 = 1$. Fuzzy relations satisfying (i) and (ii) are called proximity relations. The operation \otimes is usually a t -norm operation, that is, \otimes is a nondecreasing operation on $[0, 1]$ satisfying associativity, commutativity, 1 being the neutral element and 0 being an absorbent element. Noticeable t norms are min, product, and Lukasiewicz operation, i.e., $\max(0, a + b - 1)$. See Ovchinnikov⁴⁸ for a recent overview on similarity relations.

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

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

If (i') holds, the proximity relation is said to be *separating*. Interestingly enough, if a proximity relation is separating then it is always \otimes transitive for \otimes being the “smallest” t norm operation, i.e.,

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

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 \otimes -transitivity is equivalent to the triangle inequality; the min-transitivity, which is stronger, corresponds to ultrametrics.

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

$$S \circ_{\otimes} S \subseteq S \tag{A1}$$

where \circ_{\otimes} composition is defined as $S \circ_{\otimes} R(u, w) = \sup_v S(u, v) \otimes R(v, w)$ and fuzzy set inclusion is pointwisely defined by an inequality between membership functions. We write \circ simply if \otimes is the min t -norm. Clearly, (iii) (or (A1)) extends transitivity to degrees of similarity which are intermediary between 0 and 1 (transitivity of nonfuzzy relations is retrieved as a particular case). In the following we shall encounter *weaker* forms of transitivitylike properties where the similarity or approximate equality slightly deteriorates when applying transitivity.

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. Just as an extreme case consider, on the set of real numbers, the

nonfuzzy 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. However, even if a relation is \otimes -transitive for some operation \otimes , the corresponding notion of extended transitivity can be very weak. For instance, in the case of \otimes being 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.

Representation of Proximity Relations

A key issue is then the problem of the practical definition of elementary proximity or similarity relations. This was discussed by many authors from cognitive and formal philosophy points of view, see, e.g., Rosch and Mervis,⁴⁹ Niiniluoto.⁵⁰ 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 representation:

- 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

$$S_L(x, y) = L(|x - y|) \tag{A2}$$

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

$$\begin{aligned} \forall x, \forall y, S_L(x, y) &= \max\left(0, \min\left(1, \frac{\delta + \varepsilon - |x - y|}{\varepsilon}\right)\right) \\ &= \begin{cases} 1 & \text{if } |x - y| \leq \delta \\ 0 & \text{if } |x - y| > \delta + \varepsilon \\ (\delta + \varepsilon - |x - y|)/\varepsilon & \text{otherwise} \end{cases} \end{aligned} \tag{A3}$$

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 A.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.

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 \otimes -transitive (for the usual t -norm operations \otimes) 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. It is separating if and only if $\delta = 0$.

The \otimes -transitivity property for S_L is expressed by the following inequality

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

or equivalently

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

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 \otimes -transitivity. It is easy to prove that when \otimes is min conjunction or product-conjunction only these trivial fuzzy relations fulfil the \otimes -transitivity. However, if \otimes 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 A.1 satisfy Lukasiewicz-transitivity. See Klawonn and Kruse.⁵¹

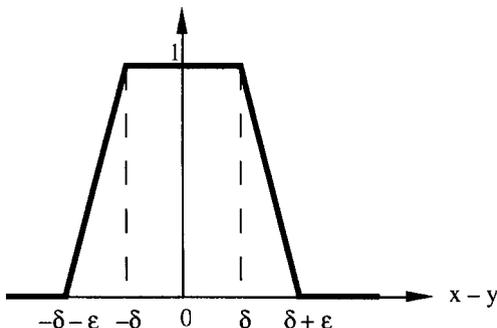


Figure A.1.

The min composition of two fuzzy relations S_L and $S_{L'}$ of the form (A2) 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 Prade⁴⁵) 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 into $L \oplus L$ (expressed by the increase of the parameters), that contains L . Note that if we use sup- \otimes convolution for computing $L \oplus L$, when \otimes is Lukasiewicz t -norm, then $L \oplus L = L$ if $\delta = 0$ (Dubois and Prade⁵²) and if the shape of L is linear.

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

$$S_P(u, v) = P(u/v) \quad (\text{A4})$$

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,

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

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

$$S_P \circ S_{P'} = S_{P \odot P'} \quad (\text{A5})$$

This expresses that x is close to y in the sense of $S_{P \odot P'}$. It again accounts for the degradation of the transitivity since $P \odot P \supset P$ and thus $S_{P \odot P} \supset S_P$.

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 \otimes -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 \otimes transitive then $S_{\otimes}((x_1, x_2), (y_1, y_2)) = S_1(x_1, y_1) \otimes S_2(x_2, y_2)$ and $S_{\min}((x_1, x_2), (y_1, y_2)) = \min(S_1(x_1, y_1), S_2(x_2, y_2))$ (for 2 features but it can be easily generalized to n features) are still \otimes 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 but will not 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 min operation, a weighted version can be given by

$$S(x, y) = \min_{i=1, n} \max(S_i(x_i, y_i), 1 - \lambda_i) \quad (\text{A6})$$

with $\max_{i=1, n} \lambda_i = 1$, where case x (resp. y) is described by the vector of feature values (x_1, \dots, x_n) (resp. (y_1, \dots, 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. Moreover, separation is also preserved if $\lambda_i \neq 0$ for all $i = 1, 2, \dots, n$. See Dubois and Prade^{39,45} for a justification of (A6) 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. For a full account of aggregation of fuzzy relations, see the book by Fodor and Roubens.⁵³

APPENDIX B: FUZZY GRADUAL RULES AND SUGENO AND NISHIDA'S INTERPOLATION METHOD

The purpose of this appendix is to provide the reader with a brief account of the relation between fuzzy gradual rules (which were used in the paper for discussing deterministic case-based reasoning problems) and interpolation mechanisms, since Gilboa and Schmeidler's³⁴ approach to case-based decision, recalled in Section III A, reduces to classical interpolation in some particular cases.

Sugeno and Nishida's⁵⁴ fuzzy control 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

$$y = \frac{\sum_i A_i(x) \cdot b_i(x)}{\sum_i A_i(x)} \quad (\text{B1})$$

where the A_i s are fuzzy sets defined in U , the $b_i(x)$ s are scalar values in V , and y ranges on V . In practice, in fuzzy control, U and V are subparts of the real line. Equation (B1) indeed looks like interpolation. When $b_i(x)$ does not depend on x , this result can be obtained by applying Zadeh's⁵⁵ 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, [in the sense 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 B.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 (B1). 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 further it is from a_i , A_{i-1} should decrease when A_i increases, and $A_i((a_i + a_{i+1})/2) = A_{i-1}((a_{i-1} + a_i)/2) = 0.5$. The

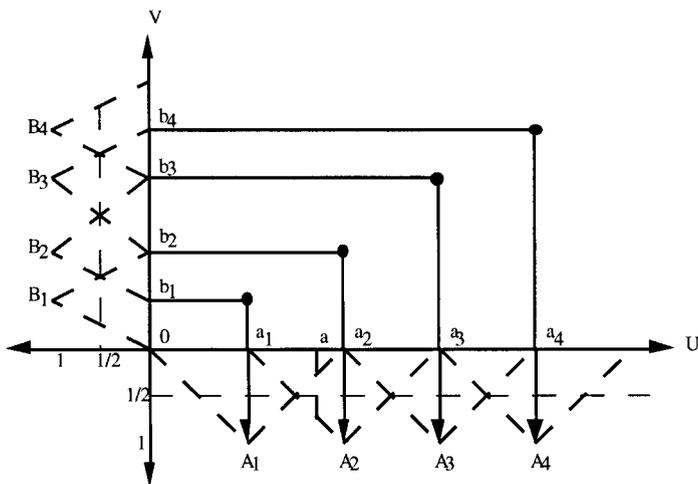


Figure B.1.

simplest way of achieving this is to let

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

an example of which are triangular-shaped fuzzy sets as in Figure B.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 \rightarrow is the implication defined above in this appendix, and so $\alpha \rightarrow B$ corresponds to the level cut B_α , with $\alpha_{i-1} = A_{i-1}(a)$, $\alpha_i = A_i(a)$, and $\alpha_{i-1} + \alpha_i = 1$. 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 B.2.

It is a theoretical justification for Sugeno and Nishida’s⁵⁴ 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, \dots , and the more X_n is A_n , the more Y is B ” is studied in detail in (Dubois, Grabisch, and Prade⁵⁶); in this more general case, where the “and(s)” in the compound condition are modelled

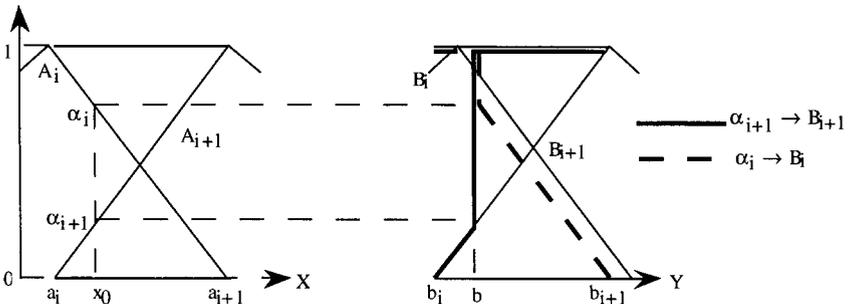


Figure B.2. The core of $(\alpha_i \rightarrow B_i) \cap (\alpha_{i+1} \rightarrow B_{i+1})$ is the result of a linear interpolation.

by a fixed operation like “product” or “min,” the gradual rule approach only gives an approximation of the result obtained by Sugeno and Nishida’s method.

Let us now apply the above results to the case-based decision problem considered in Section III A. Take the a_i s as situations s_i s stored in a memory M to which a given act was applied, and take the b_i s as the utilities of the corresponding results. See Figure B.1. Moreover, take A_i as the fuzzy set $S(\cdot, s_i)$ and B_i as the fuzzy set $T(\cdot, u(r_i))$. It is then easy to check that the conditions required above for the a_i s, b_i s, A_i s, and B_i s, completely agree with those stated in Section III A. They allow to retrieve Gilboa and Schmeidler’s³⁴ utility expression (12) by applying the fuzzy gradual rules approach to deterministic case-based problems.