A fuzzy set approach to case-based decision

Didier DUBOIS and Henri PRADE*

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

Abstract: This paper is an attempt at providing a fuzzy set-based approach to case-based decision. Case-based decision consists in selecting an action to be applied to a current problem on the basis of a set of cases storing the results of various actions applied to similar, previously encountered, problems. Recently, Gilboa and Schmeidler have presented an axiomatic justification of a counterpart of the expected utility used in decision under uncertainty, where similarity degrees play a role somewhat analogous to probability, and have proposed to apply it to case-based decision. This proposal resembles Sugeno's approach to fuzzy control. The relation between the two approaches is investigated. Besides, another approach, based on possibility and necessity measures, is presented and discussed. The idea is to favor actions which have never given bad results in problems similar to the current problem. A much more permissive view considers all the actions which have given good results (at least one time) in a similar problem.

Key Words: Case-based decision; similarity; gradual rule; possibility theory.

1 - Introduction

Analogical reasoning, and more particularly, case-based reasoning is pervasive in the way humans handle practical problems. It is based on the idea that if two problems are similar with respect to a set X of attributes describing them, they should, or at least may, be still similar with respect to other attribute(s) Y pertaining to them. Then, we can take advantage of past experience about problems which are similar (with respect to X) to a current problem in order to infer a plausible value for Y in the current problem. Case-based reasoning is now a very active research area in Artificial Intelligence (Kolodner [11]). Nevertheless, there has been rather few attempts at formalizing case-based reasoning. A proposal was made by Davies and Russell [2] to use acknowledged functional dependencies from attributes X to Y for controlling the inference process. However, this framework is very restrictive since inference is possible only if there is a perfect identity between the X-values describing the two problems, and if it is known that when the X-values are equal, the Y-values should be equal (i.e., that the functional dependency "X determines Y" holds). Then, it may be tempting to allow for a graded assessment of similarity between problems and to take advantage of "fuzzy" functional dependencies, as suggested in (Léa Sombé [12]). This idea will be briefly investigated in Section 2, where "strong" dependencies expressing that the value of X (fuzzily) determines the value of Y, are contrasted with "weak" dependencies stating only that if the X-values are similar, it is *possible* that the Y-values are similar.

Besides, Gilboa and Schmeidler [10] have recently proposed an axiomatized approach to case-based decision, where a similarity-based counterpart of expected utility is justified. Interestingly enough, the decision procedure they advocate looks somewhat similar to Sugeno and Nishida [15] approach to fuzzy rule-based control, where a control value is computed by interpolation between values recommended by fuzzy rules, on the basis of the degrees of matching of the current situation with these rules.

In Section 3, Gilboa and Schmeidler's proposal is recalled and its relationship with fuzzy control techniques is discussed in relation with the strong modelling of fuzzy functional dependencies. Then, Section 4 proposes another approach to case-based decision more in agreement with the weak dependency view.

2 - Fuzzy Functional Dependencies in Case-Based Reasoning

Let M denote a memory of experienced cases represented by pairs (s_i,t_i) for i = 1,n, where s_i denotes a problem and t_i the associated outcome (solution). The current problem will be denoted by s_0 and its intended solution by t_0.

A fuzzy functional dependency (e.g., Raju and Majumdar [14]) of the form "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, can be modelled by the constraint

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

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 ($\forall s_1$, $S(s_1,s_1) = 1$, $T(t_1,t_1) = 1$). According to (1), 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 [5]. In particular, $\forall\ \alpha \in (0,1]$, we have

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

where $S_\alpha = \{(s,s'), S(s,s') \geq \alpha\}$ is the α-cut of S and T_α is similarly defined. So it expresses that when s_1 and s_2 are close, t_1 and t_2 should be at least as close. Moreover, if T is such that $T(t_1,t_2) = 1 \Leftrightarrow t_1 = t_2$, then the classical functional dependency $s_1 = s_2 \Rightarrow t_1 = t_2$ is a consequence of (1) using the reflexivity of S. Constraint (1) is then clearly stronger than a classical functional dependency in this case, since the above implication holds for any α-cut.

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

$$E = \bigcap_{(s,t)\in M} \{t' \in \mathcal{T} \mid S(s,s_0) \leq T(t,t')\}. \quad (2)$$

Note that E may be empty if T is not permissive enough. It can be shown that the non-emptiness of E can be guaranteed by the coherence (in the sense of (Dubois, Prade and Ughetto [9])) of the set of fuzzy gradual rules "the more s_0 is S-similat to s_i the more t_0 should be T-similar to t_i for each case $(s_i,t_i) \in M$.

The requirement that the value of s uniquely determines the value of t, or at least that when s_i and s_j are close t_i and t_j should be close also, 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 much different from t'. Indeed two very similar second hand cars might be sold at different

---

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 $\mu_F(u)$.

prices, for instance. In such a case, we suggest to use a weaker version of the principle underlying case-based reasoning, 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 the clarification of the intended meaning of 'possible' in the above rule.

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 [8]). 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

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

See Dubois and Prade [6] for a full justification of its semantics when both A and B are fuzzy. This model of fuzzy rule is close to Mamdani [13]'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

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

As it can be seen, what is obtained is the fuzzy set of values t' T-similar to t, "truncated" by the global degree $S(s,s_0)$ of similarity of s and s_0. Since it applies to all the pairs $(s,t) \in M$, we obtain the following fuzzy set E of possible values t' for t_0 ($E(t') = \pi_{t_0}(t')$)

$$E(t') = \max_{(s,t)\in M} \min(S(s,s_0),\ T(t,t')). \tag{3}$$

3 - 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 [10] 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

$$U_{s_0,M}(a) = \Sigma_{(s,a,r)\in M}\ S(s_0,s) \cdot u(r) \tag{4}$$

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 [10] 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 (4) by applying the approach of Section 2 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 (2) yields

$$E_a = \bigcap_{(s,a,r)\in M} \{u(r') \mid S(s,s_0) \leq T(u(r), u(r'))\}. \tag{5}$$

Using results in the Appendix, 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

$$e_a = \frac{\sum_{(s,a,r)\in M} S(s,s_0) \cdot u(r)}{\sum_{(s,a,r)\in M} S(s,s_0)} \tag{6}$$

where we recognize (4) upto a normalization factor depending on the act *a*. An appropriate choice of T in (5) is also required in order to have (5) reducing to the value (6). See Appendix. This shows that for particular S and T, (5) 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. Moreover we have to make $S_{s_1}(s_0,s_1) + S_{s_2}(s_0,s_2) = 1$.

Note that (4) is also akin to the fuzzy vote procedure used by Bensana et al. [1] 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

$$U(a) = \sum_{i:a=a^{(i)}} S^{(i)}(s_0) \cdot w^{(i)} \tag{7}$$

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.

Although (4) 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 (6) 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 (4) 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 (4) 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).

4 - 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:

$$U_{*s_0,M}(a) = \min_{(s,r):\ (s,a,r)\in M} S(s,s_0) \to u(r) \tag{8}$$

where (16) is a multiple-valued implication connective (i.e., $x \to y$ increases with y, decreases with x, and coincides with material implication on {0,1}). Assuming 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 $u(r) = 0$.

This leads to choose an implication $x \to y$ of the form $x \to y = n(x) \perp y$ 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 $x \to y = \max(1 - x, y)$,[2] i.e.,

$$U_{*s_0,M}(a) = \min_{(s,r):\ (s,a,r)\in M} \max(u(r), 1 - S(s,s_0)) \tag{9}$$

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 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

$$U^*_{s_0,M}(a) = \max_{(s,r):\ (s,a,r)\in M} \min(u(r), S(s,s_0)). \tag{10}$$

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

$$U^*_{s_0,M}(a) \geq U_{*s_0,M}(a), \tag{11}$$

provided that the fuzzy set $S(\cdot, s_0)$ of situations s which are similar to s_0 and where act *a* was

---

[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.

already experienced (i.e., such that $\exists$ r and (s, a, r) $\in$ M for the considered act *a*), is a normalized fuzzy set. Indeed when

$$h_S(s_0) = \max_{s:\ (s,a,r)\in M} S(s,s_0) < 1,$$

it means that act *a* was never experienced on a situation completely similar to s_0. In particular when {s, (s, a, r) $\in$ M, $S(s,s_0) > 0\} = \emptyset$, $U_{*s_0,M}(a) = 1$ using (9) which is not satisfactory. It suggests to modify (9) into

$$U_{*s_0,M}(a) = \min(h_S(s_0), \min_{(s,r):\ (s,a,r)\in M} \max(u(r), 1 - S^*(s,s_0)) \qquad (12)$$

where $S^*(\cdot, s_0)$ is a renormalized version of $S(\cdot, s_0)$ (e.g., $S^*(s,s_0) = S(s,s_0) / h_S(s_0)$ with $h_S(s_0) \neq 0$). The rationale behind (12) is that our willingness to apply act *a* to s_0 is upper bounded by the existence of situations completely similar to s_0 where act *a* was experienced. Moreover $S(\cdot, s_0)$ is renormalized in order to obtain a meaningful degree of inclusion. Thus, (12) corresponds to the expression of the compound condition "there exists situations similar to s_0 where act *a* was used and the situations *which are the most similar to s_0* are among the situations where act *a* led to good results". Note that the similarity is no longer estimated in an absolute manner, but in a relative way, hence the normalization. However with (12), the inequality (11) is not guaranteed, except if we change $S(\cdot, s_0)$ into $S^*(\cdot, s_0)$ in (10) also. Besides, it would be natural to have the optimistic evaluation all the greater as the act *a* was never applied to situations similar to s_0 in the past (indeed in this case, the lack of information should benefit to act *a*). This leads to modify (10) into

$$U^*_{s_0,M}(a) = \max(1 - h_S(s_0), \max_{(s,r):\ (s,a,r)\in M} \min(S(s,s_0), u(r))). \qquad (13)$$

Using (12) and (13), the inequality (11) is preserved. Moreover when $h_S(s_0) = 1$ (i.e., the set of situations similar to s_0 is normalized), (9) and (10) are retrieved. Note that we have $U^*_{s_0,M}(a) = 1$ and $U_{*s_0,M}(a) = 0$, in two different situations: i) act *a* was experienced in situations completely similar to s_0 and both excellent and very bad results were obtained according to the considered cases; ii) act *a* has never been applied to a situation somewhat similar to s_0. Note that, this contrasts with $U_{s_0,M}$ defined by (4) which gives a non-zero result in the first case and is zero in the second case. However, $U_{(s_0,M)}$ cannot always distinguish between acts which always gave good results and acts which fail sometimes.

Interestingly enough, (9) can be viewed as the counterpart of qualitative expectations recently introduced in decision making under uncertainty (Dubois and Prade [7]), just as (4) may be seen as a counterpart of the classical expected utility.

We might also think of applying the approach proposed in Section 2 based on weak functional dependencies. This point of view leads to estimate the utility of act *a* for problem s_0 as the fuzzy set E_a of possible values t' defined as

$$E_a(t') = \max_{(s,a,r)\in M} \min(S(s,s_0), T(u(r), t')), \qquad (14)$$

by application of a weak dependency rule stating that an act *a* applied to similar situations possibly yields results with similar values (which requires the knowledge of the relation T). When T is the identity (which expresses that the value of the result of an act remains unchanged when applied to situations which are somewhat similar), (14) can be simplified into

$$E_a(t') = \max_{s:\ (s,a,r)\in M \text{ and } u(r) = t'} S(s,s_0). \qquad (15)$$

Thus, (9) and (10) can be also viewed as pessimistic and optimistic scalar summaries (in a sense which is to be clarified), of the fuzzy set of values whose membership function is given by (15). Indeed (9) and (10) are weighted median operations; see (Dubois and Prade [4]).

5. Concluding remarks

This is short note is clearly preliminary. A deeper study is necessary for understanding when the different case-based decision methods which have been suggested can be applied. Besides, it is possible in practice to make $U*_{s_0,M}$ as defined in Section 4 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. This is a topic for further research.

References

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

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

[3] Dubois D., Grabisch M. and Prade H., Gradual rules and the approximation of control laws, in: Theoretical Aspects of Fuzzy Control, Edited H.T. Nguyen, M. Sugeno, R. Tong and R.R. Yager, Wiley, New York, 1995, pp. 147-181.

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

[5] Dubois D. and Prade H., Gradual inference rules in approximate reasoning, Information Sciences, Vol. 61, 1992, pp. 103-122.

[6] Dubois D. and Prade H., Fuzzy rules in knowledge-based systems —Modelling gradedness, uncertainty and preference—, in: An Introduction to Fuzzy Logic Applications in Intelligent Systems, Edited by R.R. Yager and L.A. Zadeh, Kluwer Academic Publ., Dordrecht, 1992, pp. 45-68.

[7] Dubois D. and Prade H., 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, 1995, pp. 1924-1930.

[8] Dubois D. and Prade H., What are fuzzy rules and how to use them, Fuzzy Sets and Systems, Vol. 84, 1996, pp. 169-185.

[9] Dubois D., Prade H. and Ughetto L., 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, 1996, pp. 1858-1864.

[10] Gilboa I. and Schmeidler D., Case-based decision theory, The Quarterly J. of Economics, August 1995, pp. 607-639.

[11] Kolodner J., Case-Based Reasoning, Morgan Kaufmann, San Mateo, CA, 1993.

[12] Léa Sombé (Group), Reasoning by analogy, in: Reasoning Under Incomplete Information in Artificial Intelligence, Edited by Léa Sombé, Wiley, New York, 1990, pp. 418-424.

[13] Mamdani E.H., Application of fuzzy logic to approximate reasoning using linguistic systems, IEEE Trans. Comput., Vol. 26, 1977, pp. 1182-1191.

[14] Raju K.V.S.V.N. and Majumdar A.K., Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems, ACM Trans. on Database Systems, Vol. 13, No. 2, 1988, pp. 129-166.

[15] Sugeno M. and Nishida M., Fuzzy control of model car, Fuzzy Sets and Systems, Vol.

16, 1985, pp. 103-113.

[16] Zadeh L.A., Similarity relations and fuzzy orderings, Information Sciences, 1971, pp. 177-200.

[17] Zadeh L.A., A theory of approximate reasoning, in: Machine Intelligence, Vol. 9, Edited by J.E. Hayes, D. Michie and L.I. Mikulich, Elsevier, New York, 1979, pp. 149-194.

Appendix: Fuzzy Gradual Rules and Sugeno and Nishida's Interpolation Method

Sugeno and Nishida [15]'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

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

which indeed looks like an interpolation. When $b^{(i)}(x)$ does not depend on x, this result can be obtained by applying Zadeh [17]'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 $\forall u$, $\Sigma_i A^{(i)}(u) = 1$), which is in particular the case with the usual partitions made of triangular membership functions as in Figure A.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.

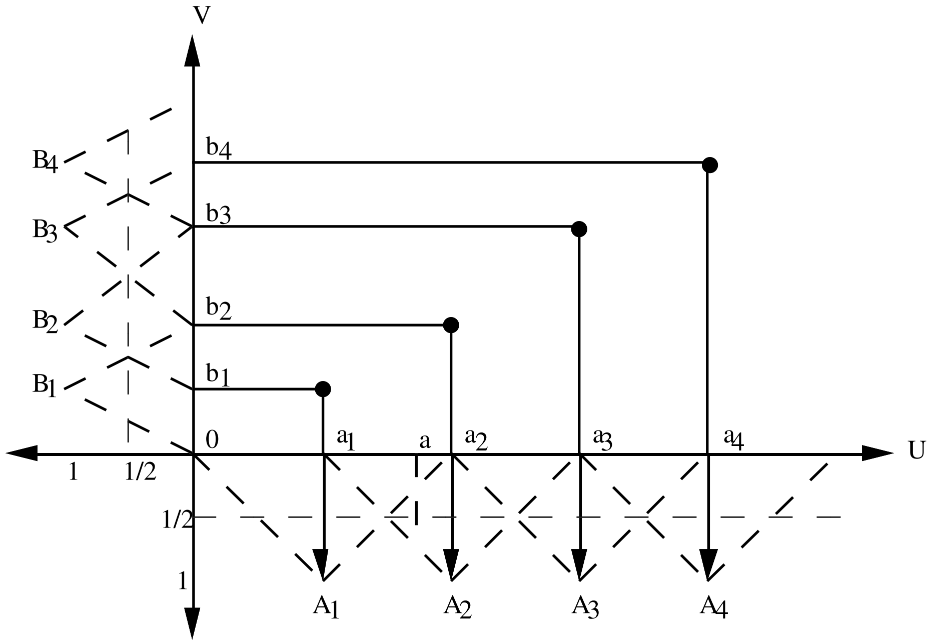

Figure A.1

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 A.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

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

where $\alpha \to B$ corresponds to the level cut B_α, where $\to$ 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 A.2. It is a theoretical justification for Sugeno and Nishida [15]'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 $\alpha_i = A_i(a)$. 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 [3]).

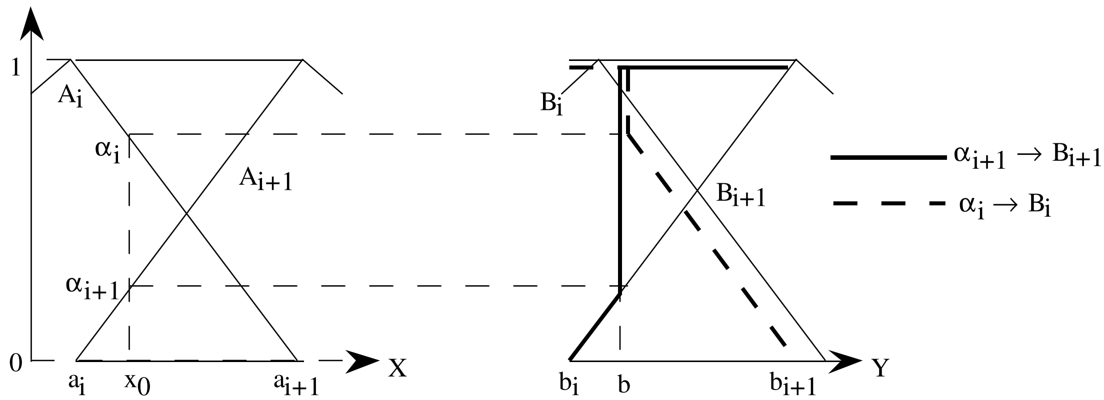

Figure A.2: The core of $(\alpha_i \to B_i) \cap (\alpha_{i+1} \to B_{i+1})$ is the result of a linear interpolation