A New Perspective on Reasoning with Fuzzy Rules

D. DUBOIS[1], H. PRADE[1], and L. UGHETTO[2]

[1] IRIT – CNRS, Université Paul Sabatier, 31062 TOULOUSE Cedex 4, FRANCE
[2] IRIN, Université de Nantes, BP 92208, 44322 NANTES Cedex 3, FRANCE
Didier.Dubois@irit.fr Henri.Prade@irit.fr
Laurent.Ughetto@irin.univ-nantes.fr

Abstract. This paper puts forward the idea that information encoded on a computer may have a negative or a positive emphasis. Negative information corresponds to the statement that some situations are impossible. It is often the case for pieces of background knowledge expressed in a logical format. Positive information corresponds to observed cases. It is often encountered in data-driven mathematical models, learning, etc. The notion of an "if…, then…" rule is examined in the context of positive and negative information. It is shown that it leads to the three-valued representation of a rule, after De Finetti, according to which a given state of the world is an example of the rule, a counterexample to the rule, or is irrelevant for the rule. This view also sheds light on the typology of fuzzy rules. It explains the difference between a fuzzy rule modeled by a many-valued implication, and expressing negative information, and a fuzzy rule modeled by a conjunction (a la Mamdani) and expressing positive information. A new compositional rule of inference adapted to conjunctive rules, specific to positive information, is proposed. Consequences of this framework on interpolation between sparse rules are also presented.

1 Introduction

Fuzzy rules are often considered to be a basic concept in fuzzy logic [1]. Originally, fuzzy rules were meant to represent human knowledge in the design of control systems, when mathematical models were lacking [2]. However, in the last ten years or so, fuzzy rules have been mainly used as a tool for approximating functions from data. This trend gets away from the original motivations of fuzzy logic [3]. It seems to be due to the formal analogy between the mathematical models of fuzzy rules and artificial neurons, and the popularity of the "soft computing" paradigm. However restricting fuzzy rules to such a role puts them in direct competition with many other and older methods for approximating functions.

In parallel, the last ten years have witnessed a relative decline of basic research in fuzzy expert systems and approximate reasoning based on fuzzy rules (surveyed in [4]). One reason may be the lack of articulated expert knowledge in many industrial applications of fuzzy systems, and the popularity drop of symbolic Artificial Intelligence methods in industry. More recently several researchers have insisted on the fact that the originality of the fuzzy rule-based approach to modeling lies in the interpretability of the models. This point is all the more valuable as topics like data mining, knowledge discovery, and various forms of learning techniques have become an important challenge in information technology, due to the huge amount of data stored in nowadays information systems. Beyond the quest for numerical performance, looking for meaning in data has become again a relevant issue. The extraction of fuzzy association rules from data has become a topic of interest [5], [6]. One may expect that this new trend will revive the concerns on approximate reasoning, with a view to develop smart information systems. This revival may also benefit from the popularity of Bayesian belief networks whose counterparts in possibility theory have been recently studied (see for instance Borgelt and Kruse [7,8] or Gebhardt [9]).

In order to extract fuzzy rules from data and exploit them in inference processes, it is necessary to have a clear understanding of the existing typology of fuzzy rules, their formal models and their various purposes. However, the state of the art on fuzzy rules, as inherited from the fuzzy modeling and control literature is somewhat confusing. There are basically three methodologies for fuzzy inference, and they disagree with each other [10].

- First, the Mamdani approach [11] is the oldest one and the first to be used on real control systems. Its peculiarity is the use of a conjunction for modeling rules, instead of the material implication, and the use of a disjunction for aggregating the rules.
  This approach has been ill-understood from a logical point of view, and is actually not in agreement with classical logic. However it has been used for the most part under precise inputs, and its anomalous logical behavior has been largely ignored because it is hidden by the defuzzification schemes. It has been strongly advocated by systems engineering scholars, like Mendel who considers that results obtained with fuzzy versions of material implications "violate engineering commonsense" [12]. However, formal difficulties with Mamdani's reasoning method from the logical point of view, once we try to use it beyond what control engineers do with it, have been pointed out quite early ; in 1980 by Baldwin and Guild [13] and more recently, in 1989 by Di Nola et al. [14].
- Another important approach to formalizing approximate reasoning is the one proposed by Zadeh himself [15]. In his view, each piece of information (be it a rule, a natural language sentence, etc.) can be described as a fuzzy restriction on a set of possible worlds (the universe of discourse). Reasoning with a knowledge base leads to the conjunctive combination of the fuzzy restrictions induced by each piece of information, followed by the projection of this global fuzzy restriction on the domain of some variable of interest.
  An example of pattern of approximate reasoning is the generalized modus ponens, which exploits a fuzzy fact and a fuzzy rule, and computes a fuzzy conclusion. In this approach, a rule is modeled by the fuzzy extension of a material implication, and the inference process is clearly a generalization of classical logic inference. Hence Zadeh's approach is at odds with Mamdani's methods and other fuzzy systems techniques.
- The third fuzzy inference methodology is based on fuzzy rules with precise, numerical conclusion parts. This trend, now known as TSK approach, has been initiated by Takagi and Sugeno in 1985 [16] and includes neuro-fuzzy methods using radial-basis functions and so on.
  This methodology has the merit of simplifying Mamdani's systems by retaining their useful part only. In particular, it lays bare their real purpose, which is interpolation between local models having fuzzy validity ranges, rather than inference in the usual sense. There is no (logical) reasoning task left in TSK systems. Neuro-fuzzy techniques do not involve logic either.

In fact, there is a major misunderstanding between logicians from Artificial Intelligence, and systems engineers. The former are interested by the representation of knowledge, including human knowledge, and the extraction of articulated knowledge from data. The latter are interested by the derivation of a useful and accurate model of a system from observed data. There is a clash of intuitions between the knowledge-driven tradition of logical approaches and the data-driven framework of systems engineering [17], and it becomes particularly acute in the framework of fuzzy systems when observing the critical views of the engineering camp on fuzzy implications and the logical approach (as developed in Mendel's paper [12]) by scientists who appear not to be familiar with logic and automated reasoning.

In fact, if Mamdani rules are logically anomalous, it is not because they are wrong models. If implicative rules do not always meet the engineers expectations, it is not because the logical approach is flawed for modeling purposes. The reason is that Mamdani's fuzzy rules and implicative rules model radically different kinds of information, what is called here *positive* and *negative* information respectively.

The aim of this paper is:

1. to show that the typology of fuzzy rules, previously proposed by the authors [18,19] manages to reconcile the knowledge-driven logical tradition and data-driven engineering tradition;
2. to give a logical account of Mamdani's fuzzy rules in terms of case-based reasoning and to show that inference in this setting should not rely on a sup-min composition;
3. to propose a reasoned discussion of the problem of interpolating with sparse rules in the framework of positive and negative information.

The ultimate aim of this research is to try to rebuild some bridge between fuzzy systems and knowledge representation.

The paper is organized as follows. Section 2 advocates the existence of two types of information that are involved in problem-solving tasks, and shows that they can be captured in the setting of possibility theory. It corresponds to two notions of possibility, one pertaining to logical consistency and the other to the idea of guaranteed or observed possibility. Section 3 tries to formally define the notion of "if. . . then. . . " rules in the light of this distinction, and shows that both an implicative and a conjunctive views are necessary to get the complete picture. These results shed light on the nature of so-called association rules in data mining. Section 4 recalls the typology of fuzzy rules previously proposed by the authors. Section 5 shows that the existence of two kinds of information calls for two forms of fuzzy inference: one based on the sup-min composition, that is a form of constraint propagation, and the other based on a dual principle that derive conclusions that are guaranteed possible, despite the incompleteness of the input information. Lastly, Section 6 deals with the interpolation problem with sparse rules and shows that, again, two approaches exist which address it.

2 Negative vs. Positive Information

When an agent states a proposition p, it may convey one of two meanings. First, the agent may mean that all the situations where p is false are impossible, ruled out. For instance, if one is told that "museums are closed during night-time", it clearly means that no museum is open in the night. However, by stating p, the agent may sometimes mean that all the situations where p is true are guaranteed possible (because, for instance, actually observed, or formally permitted, etc.). This view is very different from the first one. For instance, if somebody tells you that "between 3 and 4 p.m., museums are open", this information gives an example of time points when the museums are normally open. It does not mean that museums are necessarily closed at 2 or 5 p.m. In the first case, a piece of information discards situations as being impossible. Hence it can be called negative information. In the second case the emphasis is put on showing examples where the information is valid and such kind of information can be termed positive. Note that if the information expressed by a proposition p is negative, a situation where p is true is not necessarily possible (there are some hours in daytime when museums are closed too). Similarly, if the information expressed by a proposition p is positive, a situation where p is false is not necessarily impossible (it is unknown actually).

In many fields, positive and negative information is found conjointly. In relational database technology, negative information consists of integrity constraints, while positive information is stored as a set of tuples. In legal reasoning, laws express what is forbidden (negative information) while past cases used for jurisprudence are positive information. In logical approaches to Artificial Intelligence, logical sentences describe constraints on possible worlds (negative information) and contrast with observed data. In mathematics, necessary conditions are a form of negative information, while sufficient conditions have a positive touch.

In general it can be said that negative information is modeled by constraints discarding impossible situations, while positive information consists of observations or examples.

2.1 Modeling Positive and Negative Information in Possibility Theory

In possibility theory, the available information is represented by means of possibility distributions which rank-order the possible states of affairs in a given referential set or attribute domain. A possibility distribution assigns to each value $u \in U$ of the variable X a degree of possibility $\pi_X(u)$.

Very early, in 1978, Zadeh proposed to represent pieces of information by means of possibility distributions [20]. Elementary propositions such as "John is tall" are of the form "X is A_i" where A_i is a fuzzy predicate and X is the variable (ranging on a domain U) underlying the statement (e.g. the height of John). The main role of possibility distributions is to discard states of affairs inconsistent with the available knowledge. Indeed, $\pi_X(u) = 0$ means that the assignment $X = u$ is totally excluded if the statement "X is A_i" is taken for granted. If μ_{A_i} denotes the membership function of the fuzzy set A_i on U, and π_X denotes a possibility distribution on U, Zadeh proposed the equality $\pi_X = \mu_{A_i}$ expressing that the statement "X is A_i" induces a possibility distribution that can be equated with μ_{A_i}. In fact, it should be an inequality since the presence of one piece of information should not forbid the existence of other ones which further restrict the set of possible values. Then, the statement "X is A_i" is represented by the inequality:

$$\forall u \in U,\ \pi_X(u) \leq \mu_{A_i}(u)\ \ . \tag{1}$$

Such a statement actually means "it *must be* the case that X is (in) A_i". It represents negative information.

Given an event $A \subseteq U$ and a possibility distribution π_X on U expressing the available knowledge "X is A_i", one can define a possibility measure Π, given by:

$$\forall A \subseteq U,\ \Pi_X(A) = \sup{}_{u \in A}\ \pi(u)\ \ .$$

A possibility measure $\Pi_X(A)$ evaluates to what extent the event A is not inconsistent with the information "X is A_i".

A dual degree of necessity N_X is given by:

$$\forall A \subseteq U,\ \mathrm{N}_X(A) = \inf{}_{u \notin A}\ (1 - \pi_X(u))\ \ .$$

A necessity measure $\mathrm{N}_X(A)$ evaluates to what extent the event A is implied by the information "X is A_i". When $\mathrm{N}_X(A) = 1$, it means the following: given that "X is A_i", A is certain (i.e., true in all non-impossible situations). When $\mathrm{N}_X(A) > 0$, it means the following: given that "X is A_i", A is normally true (true in all the most plausible situations where $\pi_X(u) = 1$). These evaluations are in full agreement with the forms of deduction from a belief set encountered in classical logic.

Positive information corresponds to converse inequalities w.r.t. the above ones [21]. The statement "X is A_i" now designates a subset of values testified as possible for X[1], because all the values in A_i have been observed as possible by an information source. A membership function represents a *guaranteed* possibility distribution δ_X. The main role of guaranteed possibility distributions is to point out states of affairs that have been or can be actually observed (if $\delta_X(u) = 1$). A value $\delta_X(u) = 0$ means that it is not known if the assignment $X = u$ is possible or not, when the statement "X is A_i" is taken for granted. Then, the guaranteed possible values for X can be described by the inequality:

$$\forall u \in U,\ \delta_X(u) \geq \mu_{A_i}(u) \ . \tag{2}$$

Such a statement actually means "it *can be* the case that X is (in) A_i". This dual view of possibilistic information is also advocated by Weisbrod [22] who calls $\delta_X(u)$ *degree of support* of the assumption "$X = u$".

A guaranteed possibility measure is then given by:

$$\forall A \subseteq U,\ \Delta_X(A) = \inf{}_{u \in A}\ \delta_X(u) \ .$$

The guaranteed possibility measure $\Delta_X(A)$ evaluates to what extent all situations where A is true are indeed possible as induced by the positive information that "X is A_i".

When both positive and negative information is available, there should be a coherence between them, according to which the observed information should not be considered impossible. It leads to the inequality $\delta_X \leq \pi_X$. In other words, negative information acts as an integrity constraint for positive information. This coherence requirement is discussed in more details by Weisbrod [22], Dubois, Hajek and Prade [17] and Ughetto et al. [23].

This basic distinction between positive and negative information is important, since human knowledge is both made of restrictions or constraints on the possible values of tuples of variables (as in the logic-based view of Artificial Intelligence), often induced by general laws, and made of examples of values known for sure as being possible, generally obtained in the form of observations, or as reported facts (as in database practice). This distinction plays a central role for the representation as well as the handling of fuzzy information.

2.2 Combining Pieces of Information

Obviously, adding observations accumulate possible worlds while adding constraints delete possible worlds. So as a basic principle, it is clear that negative information should be combined conjunctively, while positive information should be combined disjunctively. This claim is also made in [22]. In Artificial Intelligence, logic-based knowledge representations model negative information and aims at restricting a set of possible states of the world (the models of the propositions which constitute the knowledge base). Each interpretation which satisfies all the propositions in the belief base (representing the available knowledge) is then considered as possible, as not being forbidden. Thus, the addition of new pieces of negative information will just reduce the set of possible states, since the set of models of the belief base can be reduced only by the addition of new constraints. The information is said to be complete if only one possible state for the represented world remains. A statement is true (resp. false) for sure if the set of its models contains (resp. rules out) the possible states of the world; in case

[1] The set A_i may be fuzzy if values can be guaranteed possible to a degree.

of incomplete information, one may ignore if a statement is true or false. Several pieces of negative information A_i, $i = 1, \ldots, n$ are naturally aggregated conjunctively into:

$$\forall u \in U,\ \pi_X(u) \leq \min_{i=1,\ldots,n} \mu_{A_i}(u)\ \ . \tag{3}$$

If there is no information as to whether the pieces of information A_i are independent or not, there will be no reinforcement effect between them. Therefore, they are aggregated using the $\min$ operation, due to the inequalities:

$$\forall i \in \{1, \ldots, n\},\ \pi_X \leq \mu_{A_i}\ \ .$$

Then, once all the constraints are taken into account, the minimal specificity principle is applied, which allocates to each value (or state of the world) the greatest possibility degree in agreement with the constraints. This principle is one of least commitment. It states that anything not declared impossible is possible. It is characteristic of negative information and it leads to enforcing an equality in (3).

By contrast, a collection of pieces of positive information forms a database, or a memory of cases, and the more it contains the more numerous are the guaranteed possible situations. When several pieces of positive information A_j, $j = 1, \ldots, m$ are available, the inequality (2) becomes:

$$\forall u \in U,\ \delta_X(u) \geq \max_{j=1,\ldots,m} \mu_{A_j}(u)\ \ . \tag{4}$$

Here again the aggregation using the $\max$ operation is due to the inequalities (2), and the lack of information about the independence of the pieces of information A_j.

As opposed to the logical tradition, databases generally use the closed world assumption (CWA) which is the opposite of the minimal specificity principle. It says that anything not observed as actually possible is considered to be false. Only what is explicitly known is represented, and then the CWA allows for the default exclusion of what is regarded as false: a statement is either true (present in the database) or considered as false, because it is not known to be true. When new pieces of data are available, they are just added to the database, and the corresponding information, which was considered as false so far, is then considered as true.

There is no explicit representation of ignorance, only the storage of accepted statements. In the possibilistic setting, applying the closed world assumption comes down to allocating to each value (or state of the world) the least possibility degree in agreement with the observed data. The CWA is thus characteristic of positive information and leads to enforcing an equality in (4).

Note that if a logical language is used for representation purposes, the conjunctive normal form is the most natural one for representing negative information. Indeed, a conjunction of clauses properly reflects the combination of negative information (rules, constraints, regulations). Typically, it encodes a collection of "if…then…" rules in a knowledge base, as for instance in the Prolog language. On the contrary, the disjunctive normal form is the most natural one for representing positive information. Indeed, a disjunction of phrases properly reflects the combination of positive information (a memory of cases, a collection of t-uples in a relational database, …). See [17] for details.

3 What is a Rule ?

"If…then…" rules have been widely used for knowledge representation purposes in Artificial Intelligence since the late sixties, with the emergence of expert systems, and in Fuzzy Logic, with the widespread popularity of fuzzy control and fuzzy systems, since the mid seventies. To-date, rules, also called association rules, are found to be a convenient format for describing interpretable knowledge extracted from data. However, there has been little consensus on the exact nature of rules and their correct mathematical model. In the tradition of logic, as inherited by logic programming, a rule is modeled by a material implication. However, the fact that, when doing so, a rule takes the truth-value "true" when the condition part is false has led to paradoxes in logic, and to computational and representational difficulties in logic programming. In the latter area, researchers have felt the need to change the semantics of negation, so as to circumvent these difficulties (negation by failure, which is a form of CWA). In the tradition of expert systems, a rule is understood as a production rule, associated to a modus-ponens-like deduction process. A "rule" is thus a kind of inference rule, however without a clear mathematical status. In more recent probabilistic expert systems, rules are encoded as conditional probabilities in a belief network. This view of a weighted rule, if mathematically sound, is at odds with the logical tradition, since the probability of a material implication describing a rule clearly differs from the corresponding conditional probability. This observation [24] has led to a vivid debate in philosophical circles since the late seventies [25] without fully settling the case.

More recently some scholars like Calabrese [26], Goodman et al. [27], Dubois and Prade [28] have suggested that a rule is not a two-valued entity, but a three valued one. To see it, consider a database containing descriptions of items in a set S. If a rule "if X is A, then Y is B" is to be evaluated in the face of this database, it clearly creates a 3-partition of S, namely:

- the set of examples of the rule: $A \cap B$ [2],
- the set of counter-examples to the rule: $A \cap \overline{B}$,
- the set of items for which the rule is irrelevant: $\overline{A}$,

as shown on Fig. 1.

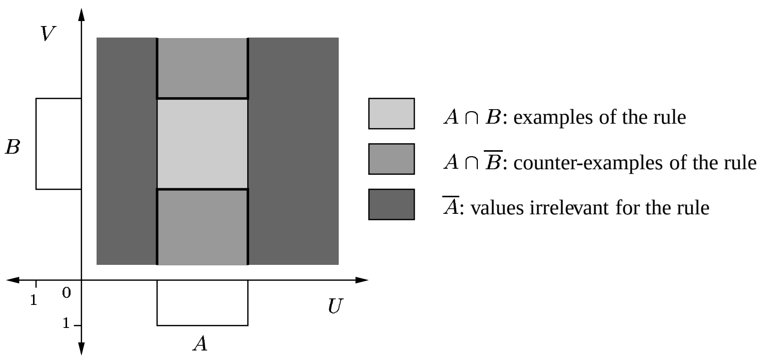

Fig. 1. Partition on $U \times V$ induced by the rule "if X is A, then Y is B".

Each case should be encoded by means of a different truth-value. The two first cases only corresponding to the usual truth-values "true" and "false" respectively. The third case corresponds to a third truth-value which, according to the context, can be interpreted as "unknown", undetermined, irrelevant, etc. This idea of a rule as a "tri-event" actually goes back to De

[2] When A and B are not defined on the same universe, the cylindrical extensions are assumed. Then, for the sake of simplicity, $A \cap B$ is given for $(A \times V) \cap (U \times B)$, and $A \cup B$ is given for $(A \times V) \cup (U \times B)$.

Finetti (1936) [29] who even proposed it as a genuine example of three-valued proposition which could be captured by Lukasiewicz logic. It is also the backbone of De Finetti's approach to conditional probability [30]. Indeed it is obvious to see that the probability $P(B|A)$ is entirely defined by $P(A \cap B)$ and $P(A \cap \overline{B})$. More recently it was shown by Dubois and Prade [28] that the three-valued semantics of rules was a possible semantics for the calculus of conditional assertions of Kraus, Lehmann and Magidor [31], which is the basis of one approach to nonmotonic reasoning. It is also in full accordance with Adams conditional logics [32] based on infinitesimal probabilities.

This framework for modeling a rule produces a precise mathematical model: a rule is modeled as a pair of disjoint sets representing the examples and the counter-examples of a rule, namely $(A \cap B,\ A \cap \overline{B})$. This definition has several consequences. First, it justifies the claim made by De Finetti that a conditional probability is the probability of a particular entity which can be called a conditional event.

Moreover it precisely shows that material implication only partially captures the intended meaning of an "if…then…" rule. It is obvious that the set of items where the material implication $\overline{A} \cup B$ is true is the complement of the set of counter-examples of a rule. Hence the usual logical view does not single out the examples of the rule, only its counter-examples. This is clearly in agreement with the fact that propositions in classical logic represent negative information. On the other hand, the set of examples of a rule is $A \cap B$ and clearly represents positive information.

Note that in the data mining literature, the merit of an association rule $A \implies B$ extracted from a database is evaluated by two indices: the support degree and the confidence degree, respectively corresponding to the probability of the conjunction and the conditional probability. This proposal may sound ad hoc. However the deep reason why two indices are necessary to evaluate the quality of a rule is because the rule generates a 3-partition of the database, and two evaluations are needed to picture their relative importance. In fact the primitive quality indices of an association rule are the proportion of its examples and the proportion of its counter-examples. All other indices derive from these basic evaluations.

The three-valued representation of an "if…then…" rule also strongly suggests that a rule contains both positive and negative information embedded in it. It also sheds light on the debate between Mamdani's rules in fuzzy logic versus implicative rules: they are the two sides of the notion of a rule. It makes no sense to discard one of them. However each side may serve a specific purpose.

For instance, the three-valued representation of a rule may solve some well-known paradoxes of confirmation discussed by scholars like Hempel [33]. Consider the statement "all ravens are black" modeled in first order logic as:

$$\forall x,\ \text{Raven}(x) \to \text{Black}(x)\ , \tag{5}$$

whose validity is to be experimentally tested. Observing black ravens indeed confirms this statement while the observation of a single non-black one would refute it. However since this sentence also writes $\forall x,\ \neg\text{Raven}(x) \vee \text{Black}(x)$, it is logically equivalent to its contraposed form:

$$\forall x, \neg\text{Black}(x) \to \neg\text{Raven}(x)\ . \tag{6}$$

So observing a white swan, which does not violate the sentence, should also confirm the claim that all ravens are black. This is hardly defendable. Using a three-valued representation of the rule, it is clear that while black ravens are examples of the rule "all ravens are black", white swans are irrelevant items for this rule. Conversely, white swans are examples of the rule "all non-black items are non-ravens", and black ravens are in turn irrelevant. Both rules have the same set of counterexamples (all ravens that are not black, if any) but their examples differ, and only examples should meaningfully confirm rules. Hence the three-valued representation enables a rule to be distinguished from its contraposed form.

This representation also tolerates non-monotonicity, as claimed above. It is intuitively satisfying to consider that that a rule R_1 = "if A then B" is safer than a rule R_2 = "if C then D", if R_1 has more examples and less counterexamples than R_2 (in the sense of inclusion). Then R_1 is said to be entailed by R_2 (see Goodman et al. [27]). In particular, the two inclusions $C \cap D \ \subseteq \ A \cap B$ and $A \cap \overline{B} \ \subseteq \ C \cap \overline{D}$ generally ensure that $P(B \mid A) \ \geq \ P(D \mid C)$.

When rules are represented by material implications, it is easy to see that R_2 classically entails R_1 as soon as R_1 has less counterexamples than R_2, since it means $\overline{C} \cup D \ \subseteq \ \overline{A} \cup B$. Examples of the rules are never involved in classical entailment. As a consequence "if A then B" always implies "if A and C then B", classically. However using the three-valued representation, this is no longer the case. Indeed while the rule "if A and C then B" has generally less counterexamples than the rule "if A then B", it also has less examples, since $A \cap B \cap C \ \subseteq \ A \cap B$. So the rule "if A and C then B" cannot be a consequence of the rule "if A then B" any more in the three-valued setting.

Now, consider a rule "if X is A then Y is B", where X and Y are two attributes with domains U and V respectively. The sets A and B are now viewed as subsets of attribute values in U and V respectively. This rule implicitly defines a relation R between these domains. However the outline of this relation, as prescribed by the rule, depends on the way it is interpreted:

1. **Negative view.** The rule is viewed as a constraint of the form "if X is A, then Y *must* be B". In other words, if $x \in A$ and $y \notin B$, then $(x, y) \notin R$, or equivalently $R \subseteq \overline{A} \cup B$. This view emphasizes only the counter-examples to the rule. It is the implicative form of the rule. Pairs of attribute values in $A \cap \overline{B}$ are deemed impossible. Other ones remain possible, as shown on Fig. 2.

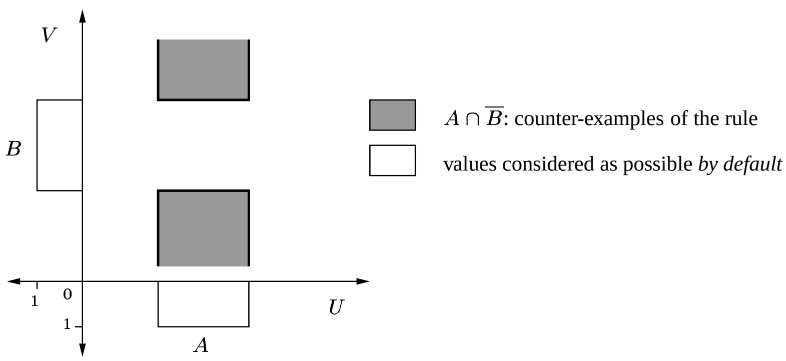

Fig. 2. Negative view of the rule "if X is A, then Y is B".

2. **Positive view.** The rule is viewed as a case of the form "if X is A, then Y *can* be B". In other words, if $x \in A$ and $y \in B$, then $(x, y) \in R$, or equivalently $A \cap B \subseteq R$. This view emphasizes only the examples of the rule. It is the conjunctive form of the rule. Pairs of attribute values in $A \cap B$ are guaranteed possible. It is not known if other ones are possible or not, as shown on Fig. 3.

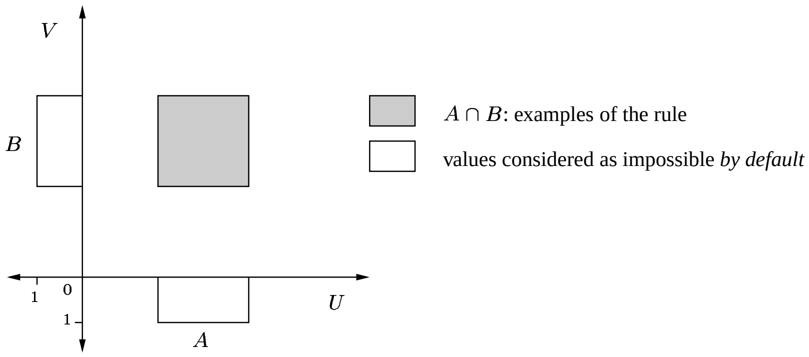

Fig. 3. Positive view of the rule "if X is A, then Y is B".

The full picture of a rule of the form "if X is A then Y is B" involves these two sides and corresponds to a double inclusion:

$$A \cap B \subseteq R \subseteq \overline{A} \cup B \ . \tag{7}$$

Namely the rule specifies a subset of the Cartesian Product $U \times V$ which includes examples and excludes counter-examples.

This dual view of fuzzy rules, and more particularly the advantages of using conjointly implicative rules (encoding negative information) and conjunctive rules (encoding positive information) in the same rule-based system, has been emphasized in [23].

This basic distinction between positive and negative information has never been emphasized in classical rule-based systems (e.g. expert systems). However, this is not really surprising. Generally, in these systems, when the rules encode a mapping from a numerical domain U to another one V, the condition parts of the rules are defined from a classical partition of U. As they do not overlap, a precise input triggers only one rule, which gives only one conclusion on V. Then it does not matter whether the conclusion encodes positive or negative information. By contrast, when dealing with uncertainty (imprecision, fuzziness, etc.) the condition parts of rules may overlap, and the inputs may be imprecise or fuzzy. Then, several rules may be triggered, each giving a (different) conclusion on the same domain. Hence these conclusions have to be aggregated, and their nature (positive or negative) determines the way they have to be aggregated, as shown in the following. It can explain why this distinction between positive and negative information first appeared in fuzzy rule-based systems. However, it is also relevant in some classical systems, where it could clarify the aggregation processes. For instance some data mining algorithms produce rules whose condition and conclusion parts are not subject to any a priori constraint. As the condition parts of these rules may overlap, a precise input can trigger several rules, whose conclusions have to be aggregated. Usually, this aggregation is ad-hoc, or based on some evaluation function of the produced outputs. Sometimes, only one rule is chosen as the most "representative" one. As a consequence, these methods do not take into account the semantics nor the potential incoherence of the produced rules.

4 Retrieving the Typology of Fuzzy Rules

Let $A \circ R$ denote the upper image of A through R:

$$A \circ R \ = \ \{v \,/\, \exists u \in A,\ (u, v) \in R\} \ .$$

It is interesting to notice the following equivalences.

- For implicative rules, $R \subseteq \overline{A} \cup B$ is equivalent to $A \circ R \subseteq B$. The latter inclusion is a formal translation of the linguistic expression corresponding to the rule: "if X is A, then Y *must* be B".
- For conjunctive rules, $A \cap B \subseteq R$ is equivalent to $A \circ \overline{R} \subseteq \overline{B}$. The latter inclusion is a bit less obvious. It means that if a value v for Y is not reachable from a value u for X in A, then (u, v) cannot be an example of the rule.

These four inclusions lead to four approaches to the modeling of fuzzy rules by fuzzy relations. Indeed, when considering fuzzy rules, these equivalences no longer hold. It means that, regardless of the choice of particular fuzzy connectives for conjunction or disjunction, four kinds of fuzzy rules fundamentally exist, as pointed out in [19].

4.1 Implicative Rules

Consider a knowledge base $\mathcal{K} = \{A_i \to B_i,\ i = 1, \ldots, n\}$, made of n parallel fuzzy rules (i.e., rules with the same input space U and the same output space V). These rules are interpreted as constraints (negative information). The fuzzy connective representing the arrow must thus be a generalized material implication. Each rule "if X is A_i, then Y *must* be B_i" (denoted $A_i \to B_i$) is represented by a conditional possibility distribution $\pi^i_{Y|X} = \mu_{A_i \to B_i}$ (the membership function of $A_i \to B_i$), which is determined according to the semantics of the rule. X is the tuple of input variables (on which information can be obtained) and Y is the tuple of output variables (about which we try to deduce information). If the implication is such that $1 \to a = a$, for $X = u \in A$ where A is not fuzzy, the fuzzy conclusion of a fuzzy rule is a fuzzy constraint on the values of Y. Each output value v is at most $\mu_B(v)$-possible:

$$\pi(u, v) \leq \mu_B(v), \quad \text{for all } \ v \in V \quad .$$

According to (3), the possibility distribution $\pi^{\mathcal{K}}$ representing the base $\mathcal{K}$ is obtained as the (min-based) conjunction of the $\pi^i_{Y|X}$'s:

$$\pi^{\mathcal{K}} = \min_{i=1,\ldots,n} \pi^i_{Y|X} \quad . \tag{8}$$

This equation shows that such rules are viewed as (fuzzy) constraints since the more rules, the more constraints, the smaller the number of values that satisfy them, and the smaller the levels of possibility. The possibility distribution $\pi^{\mathcal{K}}$ is then an upper bound of possibility for pairs of values in $U \times V$. Moreover, this conjunctive combination implies that some output values, which are possible according to some rules, can be forbidden by other ones. Then, the possibility degree $\pi^{\mathcal{K}}(u, v) = 0$ means that if $X = u$, then v is an impossible value for Y, i.e., (u, v) is an impossible pair of input/output values. By contrast, $\pi^{\mathcal{K}}(u, v) = 1$ denotes ignorance only. It means that for the input value $X = u$, no rule in $\mathcal{K}$ forbids the value v for the output variable Y. However, the addition of a new rule to $\mathcal{K}$ (expressing a new piece of knowledge) may lead to forbid this value. A possibility degree $\pi^{\mathcal{K}}(u, v) > 0$ means that the pair (u, v) is not known as totally impossible, with respect to the current knowledge.

Note that since, with the negative information view, rules express constraints on possible situations, it may happen that a knowledge base is inconsistent with a given input value $X = u$. In the fuzzy case, it means that the fuzzy set of possible values $\{(v, \pi(u, v)),\ v \in V\}$ is sub-normalized, if not empty. Then the rule base is partially useless as inference cannot be carried out for some input values. This feature has been considered an impediment when using implicative rules for fuzzy modeling. Indeed using the Takagi-Sugeno approach or the Mamdani approach, this situation never occurs. However, in the logic of negative information, it can be viewed as an expected and valuable feature. It means that this approach to fuzzy rule-based systems enables them to be tested for logical consistency [34], which can be used to force the human expert providing the rules (or the learning system) to propose meaningful sets of rules (forbidding rule bases containing statements like "if X is A, then Y is B" and "if X is A, then Y is *not* B" simultaneously, for instance). Generally, a set of implicative rules is logically coherent if and only if the fuzzy set of possible values $\{(v, \pi(u,v)), v \in V\}$ is normalized for all $u \in U$.

According to the typology of fuzzy rules proposed in [19], there are two main kinds of implicative rules, whose prototypes are *certainty* and *gradual* rules.

Certainty rules are of the form "The more X is A, the more certainly Y lies in B", as in "The younger a man, the more certainly he is single". They are based on the fuzzy translation of the inclusion $R \subseteq \overline{A} \cup B$. They correspond to the following conditional possibility distribution modeling the rule:

$$\forall (u,v),\ \pi_{Y|X}(v,u) \leq \max(1 - \mu_A(u), \mu_B(v)) \ . \tag{9}$$

In (9), A and B are combined with Kleene-Dienes implication: $a \to b = \max(1-a, b)$. Given such a rule, and an observation A' for the input variable X, the (fuzzy) set B' denotes the induced restriction on the possible values for the output variable Y. For a precise input $A' = \{u_0\}$, the conclusion B' is such that $\forall v \in V,\ \mu_{B'}(v) \geq 1 - \mu_A(u_0)$, i.e., a uniform level of uncertainty $1 - \mu_A(u_0)$ appears in B'. Then, "Y is B" is certain only to the degree $\mu_A(u_0)$, since values outside B are possible to the complementary degree. A similar behavior is obtained with the implication $a \to b = 1 - a \star (1-b)$, where $\star$ is the product instead of min. Clearly, certainty rules extend propositions with certainty levels by making the certainty level depend on the level of satisfaction of fuzzy properties.

Gradual rules are of the form "The more X is A, the more Y is B", as in "The redder the tomato, the riper it is". They are based on the fuzzy translation of the inclusion $A \circ R \subseteq B$. They correspond to the constraint:

$$\forall u \in U,\ \mu_A(u) \star \pi_{Y|X}(v,u) \leq \mu_B(v) \ , \tag{10}$$

where $\star$ is a conjunction operation. The greatest solution for $\pi_{Y|X}(v,u)$ in (10) (according to the minimal specificity principle which calls for the greatest permitted degrees of possibility) corresponds to the residuated implication:

$$\mu_{A \to B}(u,v) = \sup\{\beta \in [0,1],\ \mu_A(u) \star \beta \leq \mu_B(v)\} \ . \tag{11}$$

When $\star$ is min, (11) corresponds to Gödel implication: $a \to b = 1$ if $a \leq b$, and b if $a > b$. If only a crisp relation between X and Y is supposed to underlie the rule, it can be modeled by Rescher-Gaines implication: $a \to b = 1$ if $a \leq b$, and 0 if $a > b$.

For an input u_0, (11) provides an enlargement of the core of B, i.e., the less X satisfies A, the larger the set of values near the core of B which are completely possible for Y. This expresses a similarity-based tolerance: if the value of X is close to the core of A, then Y is close to the core of B and, more precisely, the closer X is to the core of A, the closer Y must be to the core of B. This is the basis for an interpolation mechanism when dealing with several overlapping rules ([35], [10]).

4.2 Conjunctive Rules

In the fuzzy control tradition, rule-based systems are often made of conjunction-based rules, as Mamdani-rules for instance. From the previous section it is clear that this representation considers fuzzy rules as cases, pieces of positive information. These cases, denoted $A_i \wedge B_i$ represent clusters of imprecise observations, i.e., couples of jointly possible input/output values. Each conjunctive rule is then represented by a joint "guaranteed possibility" distribution $\delta^i_{X,Y} = \mu_{A_i \wedge B_i}$.

If the conjunction $\wedge$ is such that $1 \wedge a = a$, for a precise input $X = u \in A$ where A is not fuzzy, the fuzzy conclusion of a fuzzy rule $A \wedge B$ is a collection of examples of feasible or reachable values for Y. Namely, each value $v \in V$ is at least $\mu_B(v)$-possible. If the rule does not apply (i.e., if $X = u \notin A$), then the inference mechanism leads to the conclusion $Y = \emptyset$ which, in the context of positive information means that Y is unknown, in the sense that no output is guaranteed possible.

Since conjunctive rules model imprecise information, the guaranteed possibility distribution representing the database is obtained by a disjunctive combination of the conjunctive rules, which appropriately corresponds to the accumulation of observations and leads to those pairs of values (u, v) whose possibility/feasibility is guaranteed to some minimal degree. Then, the counterpart to (8) is:

$$\delta_{\mathcal{K}} = \max_{i=1,\ldots,n} \delta^i_{X,Y} \quad . \tag{12}$$

The distribution $\delta_{\mathcal{K}}$ is then a lower bound of possibility for pairs of values in $U \times V$.

Thus, a possibility degree $\delta_{\mathcal{K}}(u, v) = 1$ means that if $X = u$, then v is a totally possible value for Y. This is a guaranteed possibility degree. By contrast, $\delta_{\mathcal{K}}(u, v) = 0$ only means that if $X = u$, no rule can guarantee that v is a possible value for Y. By default, v is considered as not possible (since possibility cannot be guaranteed). A membership degree 0 to B' represents ignorance, while a degree 1 means a guaranteed possibility. Thus, a conclusion $B' = \emptyset$ should not be understood here as "all the output values are impossible", but as "no output value can be guaranteed".

Among objections to fuzzy rule aggregation a la Mamdani coming from the logical camp is the observation that adding an empty rule "if anything, then anything", to a rule base destroys the available information (because $(\bigcup_i A_i \times B_i) \cup (U \times V) = U \times V$). This objection would be valid in the negative information view where the rule "if anything, then anything" has indeed empty contents (it is a tautology), and the disjunctive combination of information is absurd. However, in the logic of positive information, the rule "if X is U then Y is V" is not a tautology but has very extreme contents. It says that whatever the (input) value u for X, any value v for the output variable Y is guaranteed possible. In other words, there is absolutely no relationship between X and Y. No surprise if this type of positive information should destroy all previous observations. With this model, the empty rule is obviously "If X is U, then Y is $\emptyset$" which means that whatever $u \in U$, no value can be guaranteed possible for Y. Then, suitably, $(\bigcup_i A_i \times B_i) \cup (U \times \emptyset) = \bigcup_i A_i \times B_i$.

As for implicative rules, there are two main kinds of conjunctive rules, called *possibility* and *antigradual* rules [19].

Possibility rules are of the form "the more X is A, the more possible Y lies in B", as in "the more cloudy the sky, the more possible it will rain soon". They are based on the inclusion $A \cap B \subseteq R$. They correspond to the following possibility distribution modeling the rule:

$$\forall (u,v) \in U \times V,\ \delta_{X,Y}(u,v) \geq \min(\mu_A(u), \mu_B(v)) \ . \tag{13}$$

These rules, modeled with the conjunction *min*, correspond to the ones introduced by Mamdani and Assilian in 1975 [36]. For an input value u_0 such that $\mu_A(u_0) = \alpha$, a possibility rule expresses that, when $\alpha = 1$, B is a set of possible values for Y (to different degrees if B is fuzzy). When $\alpha < 1$, values in B are still possible, but they are guaranteed possible only up to the degree α. To obtain B', the output set associated with u_0, the set B is then truncated from above. Finally, if $\alpha = 0$, the rule does not apply, and $B' = \emptyset$ as already said.

Antigradual rules, which are another type of conjunctive rule (see [19]), express that "the more X is A, the larger the set of guaranteed possible values for Y is, around the core of B", as in "the more experienced a manager, the wider the set of situations he can manage". They are based on the inclusion $A \circ \overline{R} \subseteq \overline{B}$. For an input $A' = \{u_0\}$, if $\mu_A(u_0) = \alpha < 1$, the values in B such that $\mu_B(v) < \alpha$, cannot be guaranteed. Such a rule expresses how values which are guaranteed possible can be extrapolated on a closeness basis.

5 Inference with Different Types of Rules

This section is concerned with the inference from a fuzzy system under a fuzzy input, in the light of the distinction between positive and negative information. Given a fuzzy system described by a fuzzy relation, it is widely agreed that given a fuzzy input, the fuzzy output should be computed by means of a sup-min composition. It yields the upper image of a fuzzy set through the fuzzy relation. The second part of this section shows that this approach should be restricted to the handling of negative information represented by implicative rules, and that the inference from Mamdani systems under fuzzy inputs (or more generally from systems made of conjunctive rules, encoding positive information) should be based on another principle which comes down to finding the lower image of a fuzzy set through a fuzzy relation ([37], [38]).

5.1 Inference with Implicative Rules

In this section, a set of implicative rules is considered. It is modeled by a joint possibility distribution $\pi^{\mathcal{K}}$ after (8).

In order to compute the restriction induced on the values of the output variable Y, given a possibility distribution π'_X restricting the values of the input variable X, π'_X is then combined conjunctively with $\pi^{\mathcal{K}}$ and projected on V, the domain of Y (in agreement with (3)):

$$\pi_Y(v) = \sup_{u \in U} \min(\pi^{\mathcal{K}}(u,v), \pi'_X(u)) \ . \tag{14}$$

This combination-projection is known as *sup-min* composition (or Compositional Rule of Inference) and often denoted $\circ$. Then, given a set of rules $\mathcal{K}$ and an input set A', which means that the ill-known, real value for X lies in A', one can deduce the output B' given by:

$$B' \ = \ A' \circ \bigcap_{i=1}^{n} A_i \to B_i \ = \ A' \circ R^{\mathcal{K}} \ , \tag{15}$$

with $\mu_{R^{\mathcal{K}}} = \pi^{\mathcal{K}}$. The obtained fuzzy set B' is then an upper bound of the possible values for the output variable Y.

Several features of this kind of fuzzy inference deserve to be recalled [4]:

1. The rule-by-rule approach to inference (also called FITA: first infer then aggregate) is not a complete procedure in the presence of fuzzy inputs. Namely, letting $B'_i = A' \circ (A_i \to B_i)$ then $A' \circ (\bigcap_i \ Ai \ \to \ B_i) \ \subseteq \ \bigcap_i B'_i$. The rule-by-rule approach is less precise. For instance, for the input $A_i \cup A_j$, it may yield V instead of $B_i \cup B_j$, in the general case. Indeed it is often the case that the core of A_j is not included in the support of A_i, so that $(A_i \cup A_j) \circ (A_i \to B_i) = V$. Only the FATI approach (first aggregate then infer) obtains the most precise results, in agreement with (15).
2. The more rules are triggered, the more precise the conclusion derived from a given input. This conclusion is always a normalized fuzzy set when the fuzzy input is normalized, unless the rule base is not coherent in the sense depicted in Section 4.1.
3. The suitable property $A_j \circ (\bigcap_i \ A_i \ \to \ B_i) \ = \ B_j$, called inferential independence does not hold in the general case. However, it has been shown (see [39]) that, if the used fuzzy implication satisfies the modus ponens $A \circ (A \to B) = B$ (typically with Gödel implication), then the rules are inferentially independent, in particular when the A_i's form a fuzzy partition in the usual sense [40], property which is expected from a fuzzy system where each rule covers a specific area of the input space.
   The failure of modus ponens for certainty rules, for non-crisp sets, agrees with the semantics of these rules. Indeed, certainty rules express that Y is in B_i only for input values X in the *core* of A_i, i.e., input values X which are *typical* for A_i.

5.2 Inference with Conjunctive Rules

The usual approach to reasoning with Mamdani systems under fuzzy inputs is based on the sup-min inference:

$$\begin{aligned} B' &= A' \circ (\textstyle\bigcup_i A_i \times B_i) \\ &= \textstyle\bigcup_i A' \circ (A_i \times B_i) \qquad \text{(rule by rule)} \\ &= \textstyle\bigcup_i \ \min\left(\mathrm{Cons}(A', A_i), B_i\right) \qquad \text{(truncation)} \ , \end{aligned}$$

where the consistency is given by: $\mathrm{Cons}(A', A_i) = \sup_{u \in U} \min(\mu_{A'}(u), \mu_{A_i}(u))$.

At first glance, the situation is better because the FITA method is complete and is equivalent to the more cumbersome FATI method. However the approach is very questionable for inference with conjunctive rules, and it leads to serious anomalies.

- First, the use of the sup-min composition considers the fuzzy relation $R = \bigcup_i \ A_i \times B_i$ as a constraint combined with the input. So R is interpreted as negative information while each granule of it ad been considered as positive information.
- Then, when a fuzzy partition is used, there is no way of achieving $A_j \circ (\bigcup_i \ A_i \times B_i) = B_j$, except if the A_i's are disjoint [14].
- Lastly, the more rules are triggered the more imprecise the obtained conclusion, since a disjunction of partial conclusions is performed. Adding rules leads to more imprecise conclusions while one would expect the fuzzy system to be more informative with this kind of inference.

In order to overcome these difficulties, it is useful to consider the fuzzy relation obtained from a set of conjunctive rules for what it really is, namely positive information. A conjunctive rule base is actually a memory of fuzzy cases. Then what appear to be anomalies under the negative information view becomes natural. It is clear that adding a new conjunctive rule to a fuzzy case memory should expand the possibilities, not reduce them.

The fuzzy input still consists in a restriction on the values of the input variable, and is thus of a different nature. It is in some sense negative information. So the question is: “how to exploit a set of fuzzy cases, which for each input value describes the fuzzy set of guaranteed possible output values, on the basis of negative imprecise information on the input ?”

In fact what has to be computed, via an appropriate projection, are the output values which are guaranteed possible for Y, for *all* values of X compatible with the restriction on the input value.

Then, considering a set of conjunctive rules $\mathcal{K} = \{A_i \wedge B_i,\ i = 1, \ldots, n\}$, the computation of B' can no longer be achieved via the sup-min composition applied to the disjunctive aggregation of the rules. Indeed, the sup-min composition applied to too imprecise an input A', such that $A_i \cap A_{i+1} \subseteq A' \subseteq A_i \cup A_{i+1}$, leads to too large a conclusion $B_i \cup B_{i+1}$, since $A' \times (B_i \cup B_{i+1})$ contains values which are not guaranteed possible, as shown on Fig. 4, on the two hatched zones.

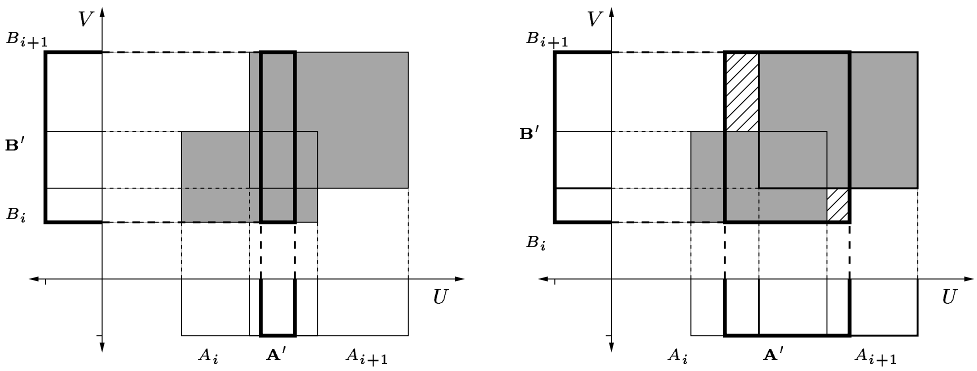

Fig. 4. Sup-min composition with Mamdani rules.

The expected conclusion, in terms of guaranteed possible values, is given for a non fuzzy input A' by:

$$\mu_{B'_\wedge}(v) = \inf_{u \in A'} \max_i \delta^i(u, v) \ . \tag{16}$$

What is computed is the intersection of the sets of images of precise inputs compatible with A'. Any value $Y = v$ in this intersection is guaranteed possible, by any input value compatible with A'. $B'_\wedge$ is the lower image [37] of A' via the relation R aggregating the fuzzy cases.

When A' is fuzzy, this equation can be generalized by (see [41]):

$$\mu_{B'_\wedge}(v) = \inf_{u \in U} \left(\mu_{A'}(u) \to \max_i \delta^i(u, v) \right) \ , \tag{17}$$

where $\to$ is Gödel implication: $a \to b = 1$ if $a \leq b$ and b otherwise. It can be checked that for usual fuzzy partitions, if $A' = A_i$ in (17), then $B' = B_i$, a result that cannot be obtained using the sup-min composition (see [14]).

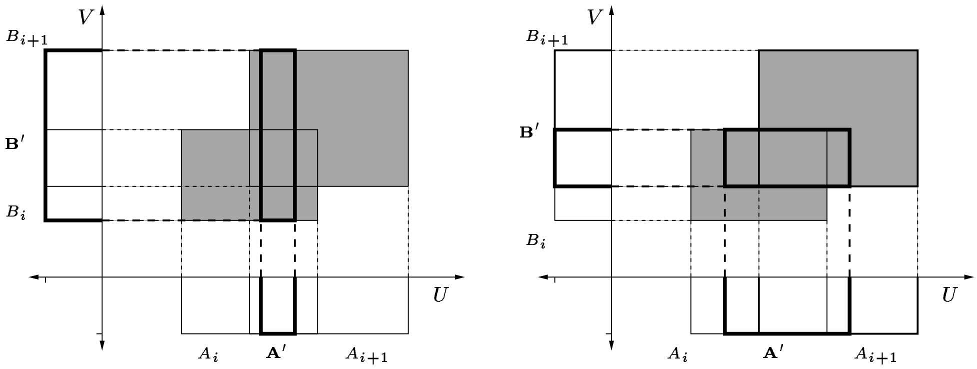

Fig. 5. B' gives the guaranteed possible values for Y, $\forall u \in A'$, for two different A'.

In the following, $\mathcal{K}_\rightarrow$ denotes a set of implicative rules, encoding negative information, and $\mathcal{K}_\wedge$ a set of conjunctive rules, encoding positive information.

Generally speaking, the logical consequence mechanism works in an opposite way for $\pi^{\mathcal{K}_\rightarrow}$ and $\delta^{\mathcal{K}_\wedge}$ (see [17]). Indeed, a fuzzy set C is a consequence of the knowledge base $\mathcal{K}_\rightarrow$ iff $\pi^{\mathcal{K}_\rightarrow} \leq \mu_C$, since $\min(\pi^{\mathcal{K}_\rightarrow}, \mu_C) = \pi^{\mathcal{K}_\rightarrow}$. By contrast, C is a consequence of the base of examples $\mathcal{K}_\wedge$ iff $\delta^{\mathcal{K}_\wedge} \geq \mu_C$, since $\max(\delta^{\mathcal{K}_\wedge}, \mu_C) = \delta^{\mathcal{K}_\wedge}$. By the way, this is in agreement with the idea of defuzzification, which consists in choosing a value in the set $\{t,\ \delta^{\mathcal{K}_\wedge}(t) > 0\}$.

One can also notice that, while the sup-min composition is monotonic ($A \subseteq A'$ implies $B \subseteq B'$), the focusing operation described by (17) is anti-monotonic ($A \subseteq A'$ implies $B \supseteq B'$), as shown on Figs. 4 and 5 respectively. The reason is that the more imprecise the input, the more restricted the set of output values attainable from all realizations of the imprecise input.

The $\inf_{u \in A'}$ in equation (16) (and its fuzzy counter-part in equation (17)) expresses that the output values in B' have to be guaranteed *for all* possible value in the input set A', while the $\sup_{u \in A'}$ in the sup-min composition would suggest that a value in B' can be guaranteed possible if it has been observed for one value in A', which is obviously wrong.

The focusing operation expressed by (16) and (17) (where A' is not combined disjunctively with the rest of the information corresponding to the rules), should not be confused with a Modus-Ponens-like counterpart for bases of examples, where the conjunctive rules and the input A' would play the same role, unlike in (17). Indeed, let us consider the two premisses: i) the values in $A \times B$ are guaranteed possible at least at the degree α, i.e., $\left(\inf_{(u,v) \in A \times B}\ \delta(u,v)\right) \geq \alpha$, and ii) the values in $\overline{A} \times V$ (where $\overline{A}$ is the complement of A on U) are guaranteed possible at least at the degree β, i.e., $\left(\inf_{(u,v) \in \overline{A} \times V}\ \delta(u,v)\right) \geq \beta$. Then one can deduce that the values for Y which lie in B are guaranteed possible at the degree $\min(\alpha, \beta)$, whatever $u \in U$, since $\inf_{(u,v) \in U \times B}\ \delta(u,v) \geq \min(\alpha, \beta)$, as it is clear on Fig. 6.

Formally, it corresponds to the following inference scheme: if $\Delta(A \times B) \geq \alpha$ and $\Delta(\overline{A}) \geq \beta$, then $\Delta(B) \geq \min(\alpha, \beta)$.

Note that the reasoning theory for positive information developed by Weisbrod [22] strongly differs from the proposal of this section. Weisbrod avocates the universality of the sup-min composition for inference, including with positive information. However the proposed rationale is debatable.

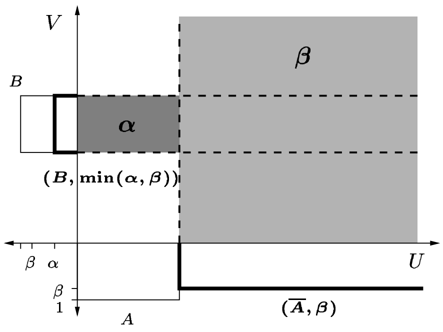

Fig. 6. Computation of guaranteed possible values.

6 Interpolation by Completion between Sparse Rules

When a rule base is made of incomplete knowledge, it may occur that for a given input $A^\star$, no rule applies. Then, the obtained conclusion is either $B^\star = V$ with implicative rules, or $B^\star = \emptyset$ with conjunctive rules, both situations representing total ignorance, as expected. However, when the rules are sparse, and the domains of the variables are real intervals, it can be natural to interpolate between values given by (at least) two neighboring rules, especially when a gradual and smooth variation of Y in terms of X is assumed.

The interpolation problem can be summarized as follows. Given a set of rules "if X is A_j, then Y is B_j", where the A_j and the B_j are ordered on their domains, and given an imprecise input A' between A_j and A_{j+1}, find an informative conclusion $B^\star$ between B_j and B_{j+1} (assuming a non decreasing monotonic function between (A_j, B_j) and (A_{j+1}, B_{j+1})).

The input $A^\star$ is said to be *between* A_j and A_{j+1} if and only if:

$$\inf\Big(\mathrm{Supp}(A_j)\Big) \leq \inf\Big(\mathrm{Supp}(A^\star)\Big) \leq \sup\Big(\mathrm{Supp}(A^\star)\Big) \leq \sup\Big(\mathrm{Supp}(A_{j+1})\Big) \ ,$$

where $\mathrm{Supp}(A_j)$ is the support of A_j. The betweenness property is denoted $A_j \leq A^\star \leq A_{j+1}$.

Several interpolation methods from sparse fuzzy rules, based on different principles have been proposed (e.g. [42] for some references). Hirota and Kóczy [43,44] use level-cuts to find set-valued outputs and rebuild the fuzzy output from the interval-valued outputs. The latter are obtained by interpolating separately from the left-hand sides and the right-hand sides of the level cuts of the conditions and conclusions of the fuzzy rules and of the input. These authors notice that sometimes the bounds of the results are not in the expected order, because the interpolation weights for left-hand sides and for the right-hand sides are not related to each other. Since then many papers have appeared which try to cope with this anomaly. In contrast Dubois and Prade [45] have suggested a view of fuzzy interpolation based on the extension principle applied to the fuzzy points defined by the Cartesian products of the left-hand side and the right-hand side of each rule. Other interpolation principles have been proposed, based on shape interpolation (Bouchon-Meunier et al. [46], [47] for instance). Jenei [48] tried to state axioms a well-behaved fuzzy interpolation method should satisfy.

In this section, only the two first mentioned interpolation methods are discussed, and re-interpreted in the framework of this paper. It is shown that both schemes derive from standard deductive inference performed on the rule base consisting of the two rules, completed with linearly interpolated rules [49,50]. The method of Kóczy and Hirota is shown to rely on implicative rules, viewed as constraints, while the other method uses conjunctive rules, viewed as imprecise data. For the sake of simplicity, this section considers that rules have only one input (the input space has one dimension), and that the conditions and conclusions of rules are non-fuzzy closed intervals (see [49] for the extension to fuzzy sets and to multi-input rules). Moreover, as a convenient notation the two considered sparse rules are denoted (A_0, B_0) and (A_1, B_1).

In the case of conjunction-based rules, the two rules between which the interpolation is performed represent two points in $A_0 \times B_0$ and in $A_1 \times B_1$ respectively, which are guaranteed to be possible. The scheme consists in linearly interpolating between these imprecise points. Formally, given $(x_0, y_0) \in A_0 \times B_0$ and $(x_1, y_1) \in A_1 \times B_1$, and assuming $x \in A^\star$, the problem is to find the output $B^\star$ as the range of:

$$y_1 \cdot \frac{x - x_0}{x_1 - x_0} + y_0 \cdot \frac{x_1 - x}{x_1 - x_0} \quad ,$$

using sensitivity analysis. It comes down to computing the area scanned by all the straight lines between the points $(x_0, y_0) \in A_0 \times B_0$ and $(x_1, y_1) \in A_1 \times B_1$. It leads to the relation Δ depicted on Fig. 7. Obviously, points (x, y) in Δ can be outside $A_0 \times B_0$, and outside $A_1 \times B_1$, even when $x \in A_0$ or $x \in A_1$. This is compatible with the semantics of conjunctive rules only.

From this construction, Δ is an imprecise function obtained by linear interpolation between imprecise points. As a consequence, the output of the interpolation problem with input $A^\star$, namely $B^\star$, should be computed considering the extension principle. Then, the output $B^\star$ is given by the sup-min composition: $B^\star = A^\star \circ \Delta$, and can be computed via interval arithmetics (see [45]). However, given $A^\star = A_i$, for $i = 0, 1$, it leads to $B^\star \supseteq B_i$ (the equality is obtained when $B_0 = B_1$ only).

In fact this method of linear interpolation from imprecise points can yield very imprecise results even if the input is precise, say $X = x^\star$. Indeed it is clear that the result is the interval on the Y-axis obtained by cutting the shaded area on Fig. 7 by the straight line $X = x^\star$. With imprecise inputs, the output is all the more imprecise. The result could be made more precise using the inf-max composition, yielding the set of output values that can be attained from all linear functions in Δ using any value in the input set $A^\star$. Using the inf-max composition instead of the sup-min composition is not a ad-hoc proposition. It can be justified by the semantics of the conjunctive rules, as shown later. Moreover, it leads to the expected result $B^\star = B_i$ when $A^\star = A_i$.

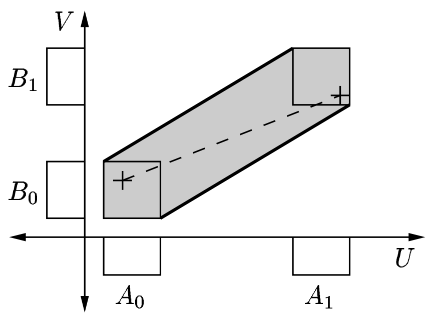

Fig. 7. Interpolation with imprecise points.

The point of view of implicative rules is to consider that the two rules no longer represent imprecisely located points, but are constraints on the range of Y, for $X \in A_0 \cup A_1$. They forbid linear functions $Y = c \cdot X + d$ between X and Y whereby $c \cdot X + d$ lies outside B_0 if $X \in A_0$ or outside B_1 if $X \in A_1$. Let $A_i = [a_i, a'_i]$, $B_i = [b_i, b'_i]$ for $i = 0, 1$. Formally, the output $B^\star$ is the range of $c \cdot X + d$ where $x \in A^\star$, under the following constraints on coefficients c and d:

$$
\begin{aligned}
c \cdot a_0 + d &\in B_0 \ , \\
c \cdot a'_0 + d &\in B_0 \ , \\
c \cdot a_1 + d &\in B_1 \ , \\
c \cdot a'_1 + d &\in B_1 \ .
\end{aligned}
\tag{18}
$$

The obtained relation between X and Y is the area R scanned by all the straight lines that do not cross the areas forbidden by the two rules (that is, containing no counterexamples of any of the rules, as shown on Fig. 8). The two straight lines that delimit R are defined by the pair of points $\{(a_0, b_0), (a_1, b_1)\}$ and $\{(a'_0, b'_0), (a'_1, b'_1)\}$. Thus, the proposed interpolation principle used with implicative rules corresponds to the case where the linear model has to be the same between A_0 and A_1 as on $A_0 \cup A_1$. The output of the interpolation problem with input $A^\star$ is actually $B^\star = A^\star \circ R$ (a sup-min composition), and can be computed as follows: for any $x \in [a_0, a'_1]$, define λ_i such that $x = a_i \cdot \lambda_i + a'_i \cdot (1 - \lambda_i)$ for $i = 0, 1$. It can be easily seen that $B^\star = [u, v]$ with $u = b_0 \cdot \lambda_0 + b'_0 \cdot (1 - \lambda_0)$ and $v = b_1 \cdot \lambda_1 + b'_1 \cdot (1 - \lambda_1)$. This is Kóczy and Hirota's method. However one may have $u > v$ since there is no relationship between λ_0 and λ_1. It can be called constrained rule-based linear interpolation.

Sometimes the two straight lines delimiting R may cross so that there is no value y corresponding to a given input x. This is when the set of constraints defining the interpolation problem is inconsistent: there may be no point $(x, c \cdot x + d)$ satisfying the constraints (18). This is true in particular if the conclusion parts of the rules are precise and are different while condition parts are imprecise, since the only possible linear functions compatible with such rules are constant functions. When R exists, the output is obviously more precise than with the former method.

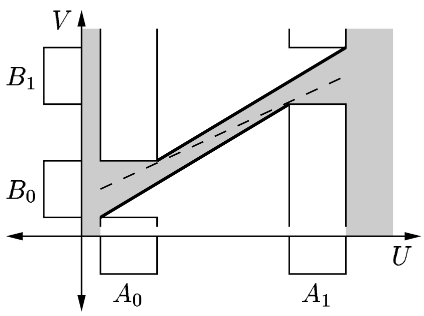

Fig. 8. Interpolation between and on $A_0 \cup A_1$.

The two linear interpolation schemes introduced above can be obtained by a unique principle which consists in completing the set of two rules $\mathcal{K}$, either by implicative or conjunctive-based rules. This interpolation principle simply consists in adding to $\mathcal{K}$ all the rules of the form (A_λ, B_λ), for $\lambda \in [0, 1]$, where A_λ and B_λ are obtained by linear interpolation, by:

$A_\lambda = \lambda \cdot A_1 + (1-\lambda) \cdot A_0$ and $B_\lambda = \lambda \cdot B_1 + (1-\lambda) \cdot B_0$. An infinite set of rules is obtained, namely $\mathcal{L} = \{(A_\lambda,\ B_\lambda),\ \lambda \in [0,1]\}$. Now, it can be shown that the relations corresponding to $\mathcal{L}$ are the ones depicted Figs. 7 and 8.

Conjunctive rules, representing ill-known points correspond to the relation $A_i \times B_i$, where $\times$ is the Cartesian product. As these rules encode possible values, they are aggregated disjunctively (possible values add). Thus, the relation $R_\mathcal{L}$ corresponding to the set of rules $\mathcal{L}$ is $R_\mathcal{L} = \bigcup_{\lambda \in [0,1]} A_\lambda \times B_\lambda$, which correspond to the grey part of Fig. 7.

Implicative rules, representing parts of a fuzzy rules correspond to the relation $A \to B$ and are aggregated conjunctively. In this case, the relation $R_\mathcal{L}$ becomes $R_\mathcal{L} = \bigcap_{\lambda \in [0,1]} A_\lambda \to B_\lambda$, which corresponds to the grey part of Fig. 8.

Since this interpolation method only consists in completing the initial set of two rules by linearly interpolated rules, the inference method should be the usual one, applied to the completed set of rules: the sup-min composition for implicative rules, and the inf-max composition for conjunctive rules. Then, when $A^\star = A_\lambda$, whatever $\lambda \in [0,1]$, the output is suitably $B^\star = B_\lambda$. This again advocates for the inf-max composition with conjunctive rules.

Since the two interpolation methods can be described by means of regular (completed) rule bases, it is clear to see why the extension principle approach always provides an output: conjunctive rules never conflict. On the other hand it is clear that the case when the Kóczy-Hirota method gives an empty result corresponds to an incoherent completed knowledge base of implicative rules. Hence rather that trying to mend the method, it is advisable either to change the rules (relax the constraints) or to change the interpolation method (non-linear interpolation, or interpolation in-between the rules only [49]). Also note that the interpolation method by completion of a set of conjunctive rules and the inf-max composition can also lead to empty results. It occurs when the input $A^\star$ is too imprecise, and then no value for Y can be guaranteed possible whatever the value of X in $A^\star$. However, this is not a case of incoherence!

The extension of the conjunctive rule-based interpretation to fuzzy rules is simple [45]. The extension of the implicative rule-based interpretation to fuzzy rules is clear in theoretical terms (using a fuzzy completed knowledge based, and applying the sup-min inference). However it is tricky in practice as subject first to a coherence test, and if coherent the computation of the fuzzy output inherits all difficulties of inference from a set of implicative rules, but now it is an infinite set (the completed set of all interpolated rules). First results appear in [49]. This is a topic for further research.

7 Conclusion

This paper has emphasized two complementary types of information called negative and positive information. Negative information acts as constraints that exclude possible worlds while positive information models observations that enable new possible worlds.

It has been shown that "if…then…" rules convey both kinds of information, through their counter-examples and examples respectively. The existence of different types of fuzzy rules, whose representation is based either on implications or on conjunctions can be explained by the existence of these two antagonistic views of information. These two points of view on rules lead to two specific and distinct inference modes. A general method by completion of the set of rules can be applied in both situations for interpolating between sparse rules.

Fuzzy rules and especially conjunctive rules have been shown recently to be useful in case-based prediction [51] for modeling the principle that similar inputs possibly have similar outputs; implicative rules are then used for interpreting the repertory of cases as constraints on similarity relations.

References

1. Zadeh, L.: The calculus of fuzzy if-then rules. AI Expert **7** (1992) 23–27
2. Zadeh, L.: Outline of a new approach to the analysis of complex system and decision processes. IEEE Transactions on Systems, Man and Cybernetics **3** (1973) 28–44
3. Dubois, D., Prade, H.: Soft computing, fuzzy logic and artificial intelligence. Soft Computing **2** (1998) 7–11
4. Bouchon-Meunier, B., Dubois, D., Godo, L., Prade, H.: Fuzzy sets and possibility theory in approximate and plausible reasoning. In Bezdek, J., Dubois, D., Prade, H., eds.: Fuzzy Sets in Approximate Reasoning and Information Systems. The Handbooks of Fuzzy Sets. Kluwer, Boston (1999) 15–190
5. Gyenesei, A.: A fuzzy approach for mining quantitative association rules. Acta Cybernetica **15** (2001) 306–320
6. Huellermeier, E.: Fuzzy association rules semantic issues and quality measures. In Reusch, B., ed.: Computational Intelligence Theory and Applications. LNCS 2206. Springer Verlag, Berlin, Germany (2001) 380–391
7. Borgelt, C., Kruse, R.: Graphical Models — Methods for Data Analysis and Mining. J. Wiley & Sons, Chichester, United Kingdom (2001)
8. Borgelt, C., Kruse, R.: Learning from imprecise data: Possibilistic graphical models. Computational Statistics and Data Analysis **38** (2002) 449–463
9. Gebhardt, J.: Learning from data: possibilistic graphical models. In: Abductive Reasoning and Uncertainty Management Systems. Volume 4 of Handbook of Defeasible reasoning and Uncertainty Management Systems (DRUMS). Kluwer Academic Publishers (2000) 315–389
10. Dubois, D., Prade, H., Ughetto, L.: Fuzzy logic, control engineering and artificial intelligence. In Verbruggen, H., Zimmermann, H.J., Babuska, R., eds.: Fuzzy Algorithms for Control Engineering. Kluwer Academic Publishers (1999) 17–57
11. Mamdani, E.: Application of fuzzy logic to approximate reasoning using linguistic systems. IEEE Transactions on Computers **26** (1977) 1182–1191
12. Mendel, J.: Fuzzy logic systems for engineering: A tutorial. Proc. IEEE – Special issue on fuzzy logic with engineering applications **83** (1995) 345–377
13. Baldwin, J., Guild, N.: Modelling controllers using fuzzy relations. Kybernetes **9** (1980) 223–229
14. Di Nola, A., Pedrycz, W., Sessa, S.: An aspect of discrepancy in the implementation of modus ponens in the presence of fuzzy quantities. International Journal of Approximate Reasoning **3** (1989) 259–265
15. Zadeh, L.: A theory of approximate reasoning. In J.E. Hayes, D.M., Mikulich, L., eds.: Machine Intellignece. Elsevier, NY, USA (1979) 149–194
16. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man and Cybernetics **15** (1985) 116–132
17. Dubois, D., Hajek, P., Prade, H.: Knowledge driven vs. data driven logics. Journal of Logic, Language and Information **9** (2000) 65–89
18. Dubois, D., Prade, H.: A typology of fuzzy "if..., then..." rules. In: Proceedings of the 3rd International Fuzzy Systems Association Congress (IFSA'89), Seattle, WA (1989) 782–785
19. Dubois, D., Prade, H.: What are fuzzy rules and how to use them. Fuzzy Sets and Systems **84** (1996) 169–186 Special issue in memory of Prof A. Kaufmann.
20. Zadeh, L.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems **1** (1978) 3–28
21. Dubois, D., Prade, H.: Possibility theory as a basis for preference propagation in automated reasoning. In: Proceedings of the 1st IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'92), San Diego, Ca, USA (1992) 821–832
22. Weisbrod, J.: A new approach to fuzzy reasoning. Soft Computing **2** (1998) 89–99
23. Ughetto, L., Dubois, D., Prade, H.: Implicative and conjunctive fuzzy rules - A tool for reasoning from knowledge and examples. In: Proceedings of the 16th American National Conference on Artificial Intelligence (AAAI'99), Orlando, Fl, USA (1999) 214–219
24. Lewis, D.: Probabilities of conditionals and conditional probabilities. Journal of Philosophical Logic **3** (1973) Reprinted in Harper et al. [25].

25. Harper, W., Stalnaker, R., Pearce, G., eds.: Ifs. D. Reidel, Dordrecht, The Netherlands (1981)
26. Calabrese, P.: An algebraic synthesis of the foundations of logic and probability. Information Sciences **42** (1987) 187–237
27. Goodman, I., Nguyen, H., Walker, E.: Conditional Inference and Logic for Intelligent Systems: a theory of measure-free conditionals. North-Holland, Amsterdam, The Netherlands (1991)
28. Dubois, D., Prade, H.: Conditional objects as non-monotonic consequence relationships. IEEE Transactions on Systems, Man and Cybernetics **24** (1994) 1724–1739
29. De Finetti, B.: La logique des probabilités. In: Actes du Congrès International de Philosophie Scientifique. Volume IV., Paris, France, Hermann et Cie, Paris, France (1936) 1–9
30. De Finetti, B.: Theory of Probability — A Critical Introductory Treatment. Volume 1 & 2. John Wiley & Sons, Chichester, United Kingdom (1974)
31. Kraus, K., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44** (1990) 167–207
32. Adams, E.: The Logic of Conditionals. D. Reidel, Dordrecht, The Netherlands (1975)
33. Hempel, C.: A purely syntactic definition of confirmation. Journal of Symbolic Logic **8** (1943) 121–143
34. Dubois, D., Prade, H., Ughetto, L.: Checking the coherence and redundancy of fuzzy knowledge bases. IEEE Transactions on Fuzzy Systems **5** (1997) 398–417
35. Dubois, D., Grabisch, M., Prade, H.: Gradual rules and the approximation of control laws. In Nguyen, H., Sugeno, M., Tong, R., Yager, R., eds.: Theoretical Aspects of Fuzzy Control. Wiley (1994) 147–181
36. Mamdani, E., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. International Journal on Man-Machine Studies **7** (1975) 1–13
37. Dubois, D., Prade, H.: Upper and lower images of a fuzzy set induced by a fuzzy relation — Applications to fuzzy inference and diagnosis. Information Sciences **64** (1992) 109–120
38. De Baets, B., Kerre, E.: Fuzzy relational compositions. Fuzzy Sets and Systems **60** (1993) 109–120
39. Ughetto, L.: Inferential independence of fuzzy rules. In: Proceedings of the 13th European Conference on Artificial Intelligence (ECAI'98), Brighton, United Kingdom, John Wiley & Sons (1998) 605–609
40. Dubois, D., Martin-Clouaire, R., Prade, H.: Practical computing in fuzzy logic. In Gupta, M., Yamakawa, T., eds.: Fuzzy Computing. North-Holland, Amsterdam, Holland (1988) 11–34
41. Dubois, D., Prade, H.: Fuzzy rules in knowledge-based systems – modeling gradedness, uncertainty and preference. In Yager, R., Zadeh, L., eds.: An Introduction to Fuzzy Logic Applications in Intelligent Systems. Kluwer Academic Publishers, Boston (1992) 45–68
42. Bouchon-Meunier, B., Dubois, D., Marsala, C., Prade, H., Ughetto, L.: A comparative view of interpolation methods between sparse fuzzy rules. In: Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference. Volume 5., Vancouver, Canada, IEEE Press (2001) 2499–2504
43. Kóczy, L., Hirota, K.: Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases. Information Sciences **71** (1993) 169–201
44. Kóczy, L., Hirota, K.: Approximate reasoning by linear rule interpolation and general approximation. International Journal of Approximate Reasoning **9** (1993) 197–225
45. Dubois, D., Prade, H.: On fuzzy interpolation. International Journal of General Systems **28** (1999) 103–114
46. Bouchon-Meunier, B., Delechamp, J., Marsala, C., Mellouli, N., Rifqi, M., Zerrouki, L.: Analogy and fuzzy interpolation in case of sparse rules. In: Proceedings of the EUROFUSE-SIC Joint Conference, Budapest, Hungary (1999) 132–136
47. Bouchon-Meunier, B., Marsala, C., Rifqi, M.: Interpolative reasoning based on graduality. In: Proceedings of 9th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2000), San Antonio, Tx, USA (2000) 483–487
48. Jenei, S.: A new approach for interpolation and extrapolation of compact fuzzy quantities. In: Proceedings of the 21th Linz Seminar on Fuzzy Set Theory, Linz, Austria (2000) 13–18

49. Ughetto, L., Dubois, D., Prade, H.: Fuzzy interpolation by convex completion of sparse rule bases. In: Proceedings of 9th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2000), San Antonio, Tx, USA (2000) 465–470
50. Ughetto, L., Dubois, D., Prade, H.: Interpolation linéaire par ajout de règles dans une base incomplète. In: Proceedings des Rencontres Francophones sur la Logique Floue et ses Applications (LFA'00), La Rochelle, France, Cépaduès, Toulouse (2000) 71–78
51. Dubois, D., Hüllermeier, E., Prade, H.: Flexible control of case based prediction in the framework of possibility theory. In: Proceedings of the 5th European Workshop on Case-Based Reasoning (EWCBR'00), LNCS 1898 – Springer Verlag (2000) 61–73