Implicative and conjunctive fuzzy rules – A tool for reasoning from knowledge and examples

Laurent UGHETTO*, Didier DUBOIS and Henri PRADE****

* IRISA, IUT de Lannion, rue Edouard Branly, BP 150, 22302 LANNION Cedex, FRANCE
e-mail: Laurent.Ughetto@iut-lannion.fr
** IRIT – CNRS, UPS, 118 route de Narbonne, 31062 TOULOUSE Cedex 4, FRANCE
e-mail: {dubois,prade}@irit.fr

Abstract

Fuzzy rule-based systems have been mainly used as a convenient tool for synthesizing control laws from data. Recently, in a knowledge representation-oriented perspective, a typology of fuzzy rules has been laid bare, by emphasizing the distinction between implicative and conjunctive fuzzy rules. The former describe pieces of generic knowledge either tainted with uncertainty or tolerant to similarity, while the latter encode examples-originated information expressing either mere possibilities or how typical situations can be extrapolated.

The different types of fuzzy rules are first contrasted, and their representation discussed in the framework of possibility theory. Then, the paper studies the conjoint use of fuzzy rules expressing knowledge (as fuzzy constraints which restrict the possible states of the world), or gathering examples (which testify the possibility of appearance of some states). Coherence and inference issues are briefly addressed.

Introduction

Fuzzy rules of the form "if X is A, then Y is B", where A and/or B are fuzzy sets, are often advocated as the basic unit used in fuzzy logic-based systems for expressing pieces of knowledge (Zadeh 1992), or modeling data. Although expressiveness is increased by the introduction of fuzzy sets in if-then rules, and by the existence of a wide panoply of possible operators for connecting the membership functions of A and B in the representation of the rules (e.g., (Dubois & Prade 1996)), little attention has been paid to the possible intended semantics of fuzzy rules. Indeed, researchers involved in fuzzy modeling use sets of fuzzy rules as black box tools for the approximation of control laws. In this type of works, the intended meaning of the fuzzy rules as a summary of data meaningful for a human operator is not a major concern. Besides, works more oriented towards knowledge engineering have mainly focused on the study of the properties of the generalized modus ponens, introduced by (Zadeh 1979), which extends inference to fuzzy rules.

However, a formal study (Dubois & Prade 1996) has pointed out that there exist different types of fuzzy rules with very different intended semantics. A first dichotomy must be made between implicative and conjunctive rules. The former, whose representation is of the form $\mu_A \to \mu_B$ (where $\to$ is a multiple-valued implication operator), express a more or less strict *constraint* on the values allowed for Y, conditioned by the value taken by X. The latter, whose representation is of the form $\mu_A \wedge \mu_B$ (where $\wedge$ denotes a, maybe non-symmetric, conjunction), gather sets of pairs of values which are *known as* (more or less) *feasible* for (X, Y). Thus, given the value for X, implicative (resp. conjunctive) rules *forbid* (resp. *guarantee possible*) values for Y.

This basic distinction is important since expert knowledge can be composed of both restrictions or constraints on the possible values on the one hand (e.g., induced by general laws), and of examples of possible values on the other hand (e.g., induced by observations). In this case, using simultaneously implicative and conjunctive rules allows to represent these two kinds of knowledge in the same rule base. These two types of information may also reveal some incoherence, when a constraint forbids values which are assessed as possible by an example.

Moreover, reasoning in AI is usually driven either from generic knowledge, expressed by, maybe fuzzy, expert rules (e.g., (Ruspini, Bonissone, & Pedrycz 1998), (Ayoun & Grabisch 1997)), or from data or examples, as in Case-Based Reasoning (e.g., (Bonissone & Cheetham 1997) in the fuzzy case), or in KDD which aims at extracting rules from data. In this perspective, distinguishing between the two kinds of rules or, even better, using them simultaneously is also of interest when rules are induced from both positive and negative examples of a concept. Indeed, these examples can lead to conjunctive and implicative rules respectively.

Besides, the choice between several types of fuzzy implication or conjunction operators leads to a more accurate representation of knowledge, where we can further distinguish between rules involving uncertainty in their conclusions, and rules which take benefit of fuzzy sets for expressing tolerance to similarity (without genuine uncertainty).

First, the semantics of the four main kinds of fuzzy rules is presented, emphasizing the difference between implication-based and conjunction-based rules. Then, the conjoint use of these two kinds of rules is studied. Knowledge representation, inference and coherence issues are addressed.

Different fuzzy rules for different information

In possibility theory, the available information is represented by means of possibility distributions which rank-order the possible values in a given referential set or attribute domain. A piece of information "X is (in) A_i", where X is a variable ranging on a domain U, and A_i is a subset of U (maybe fuzzy), is represented by the constraint:

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

where π_X is a possibility distribution restricting the values of X. Several such pieces of information are naturally aggregated conjunctively into:

$$\forall u \in U,\ \pi_X(u) \leq \min_i \mu_{A_i}(u). \tag{2}$$

Then, once all the constraints are taken into account, a minimal specificity principle is applied, which allocates to each value (or state of the world) the greatest possibility degree in agreement with the constraints. It leads to the equality:

$$\forall u \in U,\ \pi_X(u) = \min_i \mu_{A_i}(u). \tag{3}$$

Observation-based information corresponds to the converse inequalities. Let A_i be a subset of values testified as possible for X since all the values in A_i have been observed as possible for X by a source i (A_i may be a fuzzy set if some values are less guaranteed as possible for X). Then, the feasible values for X are restricted by the constraint:

$$\forall u \in U,\ \delta_X(u) \geq \mu_{A_i}(u). \tag{4}$$

If several sources provide examples of possible values for X, all this information is aggregated disjunctively into:

$$\forall u \in U,\ \delta_X(u) \geq \max_i \mu_{A_i}(u). \tag{5}$$

A converse principle, of maximal specificity, expressing that nothing can be guaranteed if it has not been observed, leads to limit the set of feasible values for X to:

$$\forall u \in U,\ \delta_X(u) = \max_i \mu_{A_i}(u). \tag{6}$$

This two-sided approach is applied to possibility distributions representing fuzzy rules in the following.

The semantics of the four main kinds of fuzzy rules, of the form "if X is A_i, then Y is B_i" is now detailed. The difference between implication-based and conjunction-based models is particularly addressed, emphasizing ideas first introduced in (Dubois & Prade 1996) or (Weisbrod 1996).

Implicative rules: restrictions of possible values

In the possibilistic framework (e.g.,(Dubois & Prade 1996)), each piece of knowledge is represented by a possibility distribution π^i on the Cartesian product of the domains of the involved variables, which expresses a (fuzzy) restriction on the possible values of these variables. Thus, considering 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 output space V), each rule "if X is A_i, then Y is 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 the tuple of non-input variables (on which we try to deduce information). According to (2), 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:

$$\pi^{\mathcal{K}} = \min_{i=1,\ldots,n} \pi^i_{Y|X}. \tag{7}$$

This equation shows that 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. $\pi^{\mathcal{K}}$ is then an upper bound of possible values.

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

$$\pi_Y(v) = \sup_{u \in U} \min(\pi^{\mathcal{K}}(u,v), \pi'_X(u)). \tag{8}$$

This combination-projection is known as *sup-min* composition and often denoted $\circ$. Then, given a set of rules $\mathcal{K}$ and an input A', one can deduce the output B' given by:

$$B' = A' \circ \bigcap_{i=1}^{n} A_i \to B_i = A' \circ R^{\mathcal{K}}, \tag{9}$$

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.

If, for a given precise input $A' = \{u^0\}$, the rule $A_i \to B_i$ does not apply, i.e., $\mu_{A_i}(u^0) = 0$, the *sup-min* composition yields the conclusion $B' = V$, the entire output space. This conclusion is in accordance with the conjunctive combination of the rules. Indeed, when a rule does not apply, it is not supposed to modify the conclusion B' given by the other rules. Thus V plays the role of the neutral element for the aggregation operator. This is why implicative rules are combined conjunctively.

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; (u,v) is an impossible pair of input/output values. By contrast, $\pi^{\mathcal{K}}(u,v) = 1$ denotes ignorance. 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) can 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.

As a consequence, the conclusion $B' = V$, obtained for a given precise input $A' = \{u^0\}$, should not be understood as "each output value is possible (for sure)" but rather as "the knowledge base gives no information, then it leads to no restriction on the values of the output variable", i.e., this case of total ignorance leads to an uncertainty level uniformly equal to 1. In conclusion, a membership degree 0 to B' means impossibility, while a degree 1 represents ignorance.

According to the typology of fuzzy rules proposed in (Dubois & Prade 1996), 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", or "The more crowded is the cafeteria in the morning, the more certainly it is about ten o'clock". This statement corresponds to the following conditional possibility distribution modeling the rule:

$$\forall (u,v),\ \pi_{Y|X}(v,u) \leq \max(\mu_B(v), 1 - \mu_A(u)). \tag{10}$$

Clearly, A and B are combined with Kleene-Dienes implication: $a \to b = \max(1-a, b)$. For a precise input $A' = \{u^0\}$, $\forall v \in V,\ \mu_{B'}(v) \geq 1 - \mu_A(u^0)$ holds, i.e., a uniform level of uncertainty $1 - \mu_A(u^0)$ appears in B' (see Figure 1.a). 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.

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". This statement corresponds to the constraint:

$$\forall u \in U,\ \mu_A(u) \star \pi_{Y|X}(v,u) \leq \mu_B(v), \tag{11}$$

where $\star$ is a conjunction operation. The greatest solution for $\pi_{Y|X}(v,u)$ in (11) (according to the minimal specificity principle which calls for the greatest permitted degrees of possibility) corresponds to the residuated implication:

$$\mu_{A \to B}(u,v) = \sup\{\beta \in [0,1],\ \mu_A(u) \star \beta \leq \mu_B(v)\} \tag{12}$$

When $\star$ is min, equation (12) 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 underly the rule, it can be modeled by Rescher-Gaines implication: $a \to b = 1$ if $a \leq b$, and 0 if $a > b$.

Applying (9) with one rule, the core of B' is enlarged w.r.t. B (see Figure 1.c), i.e., the less X satisfies A, the larger the set of values in the support of B which are completely possible for Y. This embodies a tolerance to similarity: if the value of X is close to the core of A, then Y is close to the core of B.

Conjunctive rules: guaranteed possible values

In the fuzzy control tradition, rule-based systems are often made of conjunction-based rules, as Mamdani-rules for instance. These rules, denoted $A_i \wedge B_i$, can no more be viewed as constraints, but rather as pieces of data, i.e., as couples of conjointly possible input/output (fuzzy) values. Each rule is then represented by a conjoint possibility distribution:

$$\delta^i_{X,Y} = \mu_{A_i \wedge B_i}.$$

A first justification of this interpretation comes directly from the semantics of the conjunction. Moreover, given a precise input $A' = \{u^0\}$, and a conjunctive rule $A_i \wedge B_i$, if the rule does not apply (i.e., $\mu_{A_i}(u^0) = 0$), then the *sup-min* composition leads to the conclusion $B' = \emptyset$. This implies a disjunctive combination of the conjunctive rules, which appropriately corresponds to an accumulation of data and leads to a set of values whose possibility/feasibility is guaranteed to some minimal degree. Equation (7) is then turned into:

$$\delta_{\mathcal{K}} = \max_{i=1,\dots,n} \delta^i_{X,Y}. \tag{13}$$

The distribution $\delta_{\mathcal{K}}$ is then a lower bound of possible values.

From a set of conjunctive rules $\mathcal{K}$ and an input A', the *sup-min* composition leads to an output B' given by:

$$B' = A' \circ \bigcup_{i=1}^{n} A_i \wedge B_i = A' \circ R_{\mathcal{K}}. \tag{14}$$

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 as "all the output values are impossible", but as "no output value can be guaranteed".

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

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". It corresponds to the following possibility distribution modeling the rule:

$$\forall (u,v),\ \min(\mu_A(u), \mu_B(v)) \leq \delta_{X,Y}(u,v). \tag{15}$$

These rules, modeled with the conjunction *min*, correspond to the ones introduced by Mamdani and Assilian in 1975.

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 set B is then truncated as shown on Figure 1.b. Finally, if $\alpha = 0$, the rule does not apply, and $B' = \emptyset$ as already said.

Antigradual rules have been obtained by symmetry relations between (10), (11) and (15) (see (Dubois & Prade 1996) for details). They correspond to the inequality:

$$\forall (u,v),\ \ \mu_A(u) \star (1 - \delta_{X,Y}(u,v)) \geq 1 - \mu_B(v), \tag{16}$$

which can also be written, when the conjunction $\star$ is min:

$$\forall (u,v),\ \ \mu_A(u) \wedge \mu_B(v) \leq \delta_{X,Y}(u,v), \tag{17}$$

where $\wedge$ is the non-commutative multiple-valued conjunction: $a \wedge b = b$ if $a + b > 1$ and 0 otherwise.

Such a rule expresses that "the more X is A, the larger the set of 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". For a given input value $A' = \{u^0\}$, this rule means that if $\mu_A(u^0) = 1$, all the values in B are possible for Y ($B' = B$). If $\mu_A(u^0) = \alpha < 1$, the values in B such that $\mu_B(v) < \alpha$, cannot be guaranteed, as shown on Figure 1.d. Such a rule expresses how values which are guaranteed possible can be extrapolated on a closeness basis.

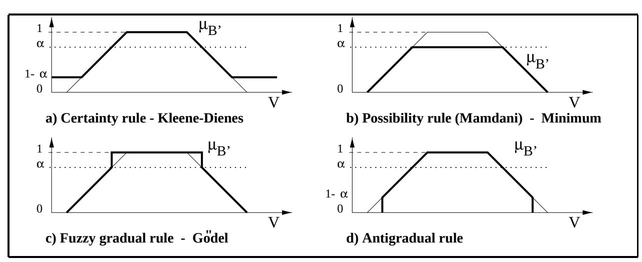

Figure 1: Inference with different kinds of fuzzy rules, and a precise input value $A' = \{u^0\}$ such that $\mu_A(u^0) = \alpha$.

Choosing between different types of rules

Given a fuzzy rule and a precise input $A' = \{u^0\}$, the conclusions B' obtained for each of the four kinds of fuzzy rules presented in the previous sections are depicted on Figure 1. Clearly, the B''s are obtained from B by applying four different thresholding functions: horizontal ones for certainty and possibility rules (low degrees of possibility are increased, and high degrees are decreased respectively), and vertical ones for gradual and antigradual rules (enlargement of the core, and squeezing of the support respectively). Thus, these four kinds of fuzzy rules can be considered as the basic ones, in particular if the considered triple (conjunction, disjunction, negation) is (min, max, 1-.).

For the valuation scale, only a complete order on the membership degrees is assumed when using *min*, *max*, and the order reversing operation $1 - (.)$. They can belong to a discrete, linearly ordered scale of membership degrees. Then, $1 - (.)$ is just the order reversing map of the scale. When the continuous $[0, 1]$ scale is used as a ratio scale, the *product* can be used instead of *min* for the conjunction. In this case, the behavior of the four kinds of rules can be made continuous, with a similar semantics. Thus, the triple of operators (conjunction, disjunction, negation), has to be chosen depending on the scale of the membership degrees. Besides, the particular case of Lukasiewicz implication $a \to b = \min(1, 1 - a + b)$ is worth noticing. Indeed, it combines the effects of both certainty and gradual rules (addition of an uncertainty level, and enlargement of the core of the output). Similarly, the bounded sum $a \wedge b = \min(a+b, 1)$ combines the effects of the two conjunctive rules.

It is important to use the right kind of rule for the representation of each piece of knowledge. First, the choice has to be made between implicative and conjunctive models.

For if-then rules expressing constraints, as for instance "if the speed of a car is highly too fast, then the driver must brake strongly", or "if a vegetable is big and orange, then it is a pumpkin", the problem is to determine if it corresponds to certainty or gradual rules. Consider an input $A' = \{u^0\}$ such that $\mu_A(u^0) = \alpha$, with a partial membership degree to the condition part of the rule ($0 < \alpha < 1$). The answer depends on the effect of α on B', as shown on Figure 1.

If the rule expresses typicality, a partial membership degree leads to an uncertainty level on the conclusion, and it is a *certainty rule*. It is the case with the second rule: being big and orange is typical of a pumpkin, if the vegetable is not really big, it can be a pumpkin, but it is not certain. Then the rule expresses "The more a vegetable is big and orange, the more certainly it is a pumpkin". A certainty rule is a rule which holds in normal cases; counter-examples should correspond to rather exceptional situations.

If by contrast the rule expresses a closeness relation, or a gradual evolution of a variable with respect to another, a partial membership degree leads to a less precise conclusion, and it is a *gradual rule*. It is the case for the first rule which expresses that the strength of the braking must be proportional to the speed of the car, and which writes "the higher the over-speed of a car, the stronger the driver must brake".

For rules expressing examples of possible values, as for instance "if someone is very rich, then this person can access to numerous means of transport, including the least common (and most expensive) ones", or "if a city is big, then its shops are open in the evening", the problem is to determine if they correspond to possibility or to antigradual rules. As for implicative rules, the answer depends on the effect of α on B'.

If the rule expresses that the whole conclusion is more or less possible, α leads to a bounding of the possibility degrees of the values in B. It is then a *possibility rule*, as in the second example which expresses the level of possibility that shops are open at "evening" time. It then writes "the bigger a city, the more possible its shops are open in the evening".

By contrast, if the rule only gives a set of more or less possible values, and α leads to the deletion of the less possible values, it is then an *antigradual rule*. It is the case in the first example, which means that the less common means of transport (supposed here to be the most expensive ones) cannot be guaranteed as possible ones for not very rich people. The rule is understood as "the richer someone, the more numerous the means of transport this person can access to".

Sometimes, certainty and possibility rules can be contrasted according to counter-examples. Indeed, the rule "the younger someone, the more it is certain that s/he is single" should be a certainty rule since counter-examples are rather exceptional. By contrast, the rule "the older someone, the more it is possible s/he has been married" is a possibility rule since even if non-married old persons are less numerous than married ones, they are not exceptional at all.

Joint use of implicative and conjunctive rules

Usually, fuzzy rule-based systems are made of parallel rules of the same kind. This section shows the interest of using several kinds of rules in the same rule base, and in particular one kind of implicative with one kind of conjunctive rules.

Thus, the considered knowledge base $\mathcal{K} = \mathcal{K}_\to \cup \mathcal{K}_\wedge$ is composed of a set $\mathcal{K}_\to = \{A_i \to B_i,\ i = 1, \ldots, n\}$ of implicative rules and a set $\mathcal{K}_\wedge = \{A_j \wedge B_j,\ j = 1, \ldots, m\}$ of conjunctive rules, where $\to$ and $\wedge$ are multiple-valued implication and conjunction operators.

Dealing with both fuzzy constraints and examples

Information pertaining to a domain can be composed of both examples of possible values, and of constraints expressing sets of impossible values. To accurately represent this information, examples and constraints must be distinguished, using conjunctive and implicative rules together.

For instance, consider an expert system for assessing the buying price of a one-roomed flat in a big city. The considered input variables are the surface (`Size`, in m^2), the proximity to the university (`Puni`) and to the town center (`Pcen`, in minutes). The output variable is the price (`Pr`$\times 1000$ dollars). An expert salesman can give the following rules (which are very sketchy, and then not very realistic, for the sake of simplicity).

- the more `Puni` is (12,15,20,23), the more certainly `Pr` is (30,35,60,65),
- the more `Puni` is (12,15,20,23) and the more `Pcen` is (7,10,20,23), the more certainly `Pr` is (40,45,55,60),
- the more `Puni` is (12,15,20,23) and the more `Pcen` is (20,23,30,33), the more certainly `Pr` is (30,35,50,55),

where (a,b,c,d) is a trapezoïdal shaped fuzzy set whose support is $]a, d[$, and core $[b, c]$. For instance, (12,15,20,23) means “approximately $[15, 20]$” (the interval $]12, 23[$ expressing what is meant by approximately) and could be given a linguistic label, as “not too far”.

Certainty rules have been chosen here since the given prices are considered as boundaries. The upper bound is the maximal price a buyer would pay for the flat, and the lower bound the minimal price the seller would accept. However, gradual rules, which encode the notion of proximity or resemblance, are also of interest in this context, and particularly if knowledge about “reference flats” is available, as it seems natural to assess: “the more similar two flats, the more similar their prices should be” (see (Dubois *et al.* 1998)).

In this kind of application, another important source of knowledge is the database of recently sold flats. This is also the case in many engineering sciences which are data-driven rather than knowledge-driven. The available information is often under the form of data, each piece of data corresponding to an actually observed situation. By contrast, each model of a knowledge base expressing constraints represents a potentially observable situation only. For instance, consider the following entries of a database:

`Size` (m^2)	`Puni` (mn)	`Pr` (dollars)
30	18	43,000
29	15	47,000
35	13	52,000
32	20	45,000

These data can be summarized by a conjunctive rule, like:

- the more `Size` is (28,30,32,36) and the more `Puni` is (12,15,20,23), the more possibly `Pr` is (40,43,50,55).

This extraction can be done either by a human expert, or an adequate KDD process. It is not discussed here. The membership grades should reflect the typicality of the examples.

Usually, fuzzy rules extracted from rough data are conjunctive ones, since they seem more natural to produce from a (generally incomplete) set of examples. However, data can be composed of both positive and negative examples, and the latter could lead to implicative rules, since they express impossible values. Depending on the applications, negative examples may be rather difficult to find, as in the flat pricing problem, where impossible values are more naturally assessed through constraints provided by experts.

This example shows that both implicative and conjunctive rules are required in order to accurately represent all the available knowledge.

Inference mechanisms

The considered rule base $\mathcal{K}$ contains two kinds of knowledge, represented in $\mathcal{K}_{\rightarrow}$ and $\mathcal{K}_{\wedge}$, whose representations $\pi^{\mathcal{K}_{\rightarrow}}$ and $\delta^{\mathcal{K}_{\wedge}}$ express an upper and a lower bound of possible values respectively. This is why the fuzzy inference mechanism (the Generalized Modus Ponens) should not be applied directly on $\mathcal{K}$, but separately on $\mathcal{K}_{\rightarrow}$ and $\mathcal{K}_{\wedge}$. Then, with only one kind of implicative and one kind of conjunctive rules, no special inference mechanisms is required. The methods consists in partitioning $\mathcal{K}$ into $\mathcal{K}_{\rightarrow}$ and $\mathcal{K}_{\wedge}$, and applying the usual algorithms.

With conjunctive rules, it is possible to apply the usual rule by rule inference technique of classical expert systems. Indeed, in equation (14), $\bigcup$ and $\circ$ commute, and then:

$$B'_{\wedge} = A' \circ (\bigcup_{j=1}^{m} A_j \wedge B_j) = \bigcup_{j=1}^{m} (A' \circ A_j \wedge B_j).$$

With implicative rules, this approach (called FITA, for first infer then aggregate) should not be applied as soon as the input A' is fuzzy. Only a global inference (called FATI for first aggregate then infer) has to be used, since only the following inclusion generally holds:

$$B'_{\rightarrow} = A' \circ (\bigcap_{i=1}^{n} A_i \rightarrow B_i) \subseteq \bigcap_{i=1}^{n} (A' \circ A_i \rightarrow B_i).$$

However, for certainty rules, the addition of well-chosen redundant rules allow to design a rule by rule inference method (Ughetto, Dubois, & Prade 1997) and, for gradual rules, specific inference techniques have been proposed.

For a (maybe fuzzy) input A', the rule base $\mathcal{K}$ leads to a double information: an upper bound $B'_{\rightarrow}$ and a lower bound $B'_{\wedge}$ of the possible values for the variable Y. With these two bounds, usual *defuzzification* methods are no longer appropriate when a precise output is required. An intuitive method consists in choosing one of the values v which maximize both $\mu_{B'_{\rightarrow}}(v)$ and $\mu_{B'_{\wedge}}(v)$. Otherwise, the two bounds provide an accurate view of the possible range of values for Y.

If this choice involves an optimization criterion, the output value can be chosen according to a notion of higher order uncertainty. In our example, if a flat sizing 31 m^2 and at 18 mn from university is to be sold, the previous rules give the range of prices depicted on Figure 2. It means that we are sure to sell it between 43 and 50. It is also possible, but not certain, to sell it between 55 and 60, and one cannot expect more than 65. In order to sell the flat very rapidly, the price can range in $[35, 40]$. If only money (and not time) is involved, the price can be then around 60.

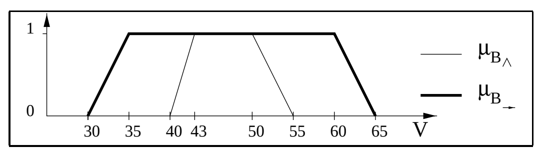

Figure 2: Possible prices for a flat (thousand dollars)

The inference mechanism becomes less simple when the input A' is also ill-known, and only bounded by $A'_{\rightarrow}$ and $A'_{\wedge}$. This is the case in particular when rules have to be chained. Here again, the natural approach consists in using $A'_{\rightarrow}$ with $\mathcal{K}_{\rightarrow}$ for obtaining $B'_{\rightarrow}$, and $A'_{\wedge}$ with $\mathcal{K}_{\wedge}$ for obtaining $B'_{\wedge}$.

Accurate representation of what is known

If the boundaries of the conclusion, namely $\{B'_{\rightarrow}, B'_{\wedge}\}$ are more difficult to handle than a usual fuzzy set B', they allow for an accurate representation of what is known about the possible values of Y. With only one kind of fuzzy rules, the membership degree of an output value v to the conclusion B' can be interpreted as follows. For implicative rules:

- $\mu_{B'}(v) = 0$ means that v is impossible,

- $\mu_{B'}(v) = 1$ means ignorance, as no rule forbids v. By default, v is considered possible.

For conjunctive rules:

- $\mu_{B'}(v) = 1$ means that v is guaranteed to be possible,
- $\mu_{B'}(v) = 0$ means ignorance, as no rule can guarantee v. By default, v is considered not possible.

Now, for sets containing both implicative and conjunctive rules, the case of ignorance is no more ambiguous since:

- $\mu_{B'_{\rightarrow}}(v) = 1$ and $\mu_{B'_{\wedge}}(v) = 1$ means that v is guaranteed to be completely possible (certainly possible),
- $\mu_{B'_{\rightarrow}}(v) = 1$ and $\mu_{B'_{\wedge}}(v) = 0$ means ignorance on v which is neither guaranteed, nor forbidden,
- $\mu_{B'_{\rightarrow}}(v) = 0$ and $\mu_{B'_{\wedge}}(v) = 0$ means that v is certainly impossible.

Absence of information is no more interpreted *by default* as possibility or impossibility, but expresses ignorance only.

Coherence checking

Validation is an important issue for rule-based systems, in order to avoid inconsistent conclusions especially. In the possibilistic framework, a set of rules is said to be coherent if for all (allowed) input variable, there is at least one output value totally compatible with the input value and the rules:

The rule base $\mathcal{K} = \{A_i \rightarrow B_i,\ i = 1, \ldots, n\}$ *is coherent if* $\forall u \in U,\ \sup_{v \in V} \pi^{\mathcal{K}}(u, v) = 1$.

According to this definition, it is easy to show that only implicative rules can be incoherent. Indeed, conjunctive rules represent only a lower bound of the possibility distribution $\delta^{\mathcal{K}}$. Thus, it is not possible to prove that $\delta^{\mathcal{K}}(u, v) < 1$, and the rule base $\mathcal{K}_{\wedge}$ is always coherent. Examples cannot be incoherent, while constraints can be. Far from being a drawback, it can be considered as a good property of implicative rules. Since coherence checking algorithms have been designed (see for instance (Dubois, Prade, & Ughetto 1997)), potential incoherence can be detected and removed.

However, checking the coherence of $\mathcal{K}_{\rightarrow}$ is not sufficient to ensure the coherence of $\mathcal{K}$. Indeed, for a given precise input, an output value can be guaranteed possible by a conjunctive rule and forbidden by an implicative rule. It is then necessary to check the coherence of $\mathcal{K}_{\wedge}$ with respect to $\mathcal{K}_{\rightarrow}$.

A set of rules $\mathcal{K} = \mathcal{K}_{\rightarrow} \cup \mathcal{K}_{\wedge}$ *is said to be coherence if* $\mathcal{K}_{\rightarrow}$ *is coherent and if* $\mathcal{K}_{\wedge}$ *is coherent w.r.t.* $\mathcal{K}_{\rightarrow}$*, i.e., if:*

$$\left\{ \begin{array}{ll} & \forall u \in U,\ \exists v \in V \textit{ such that } \pi^{\mathcal{K}_{\rightarrow}}(u, v) = 1, \\ \textit{and} & \forall (u, v) \in U \times V,\ \pi^{\mathcal{K}_{\rightarrow}}(u, v) \geq \delta^{\mathcal{K}_{\wedge}}(u, v). \end{array} \right.$$

Efficient coherence checking algorithms for sets of parallel certainty or gradual rules have been proposed in (Dubois, Prade, & Ughetto 1997). They can be used to validate $\mathcal{K}_{\rightarrow}$.

The coherence of $\mathcal{K}_{\wedge}$ w.r.t. $\mathcal{K}_{\rightarrow}$ is rather simple to check. Indeed, according to the following propositions, it comes down to check the coherence of each rule $A_j \wedge B_j$ in $\mathcal{K}_{\wedge}$ w.r.t. each rule $A_i \rightarrow B_i$ in $\mathcal{K}_{\rightarrow}$ such that $A_i \cap A_j \neq \emptyset$:

• *A set of conjunctive rules* $\mathcal{K}_{\wedge} = \{A_j \wedge B_j,\ j = 1, \ldots, m\}$ *is coherent with respect to a set of implicative rules* $\mathcal{K}_{\rightarrow} = \{A_i \rightarrow B_i,\ i = 1, \ldots, n\}$ *if and only if each rule in* $\mathcal{K}_{\wedge}$ *is coherent w.r.t. each rule in* $\mathcal{K}_{\rightarrow}$.

• *A conjunctive rule* $A_j \wedge B_j$ *and an implicative rule* $A_i \rightarrow B_i$ *are always coherent if* $A_i \cap A_j = \emptyset$.

Coherence conditions can be defined for the different pairs of rules (see (Ughetto 1997)).

Conclusion

This paper has advocated the interest of distinguishing between different kinds of rules for representing data and knowledge, which can be appropriately modeled in the fuzzy sets and possibility theory framework. It has been also shown how to check the coherence of sets of different types of rules and how to use them in inference. The differences between the various kinds of fuzzy rules are meaningful from a cognitive modeling point of view. Each kind, either constraint-based or example-based, requires a separate processing, leading to two conclusions which can be then fused, and whose coherence can be discussed. The typology of fuzzy rules should be also relevant when trying to extract rules from data in learning. The distinction between data and knowledge is discussed in a more general logical setting by (Dubois, Hajek, & Prade 1997).

References

Ayoun, A., and Grabisch, M. 1997. Tracks real-time classification based on fuzzy rules. *Int J of Int Syst* 12:865–876.

Bonissone, P., and Cheetham, W. 1997. Financial application of fuzzy case-based reasoning to residential property valuation. In *Proc. of the 6th IEEE Int. Conf. on Fuzzy Syst. (FUZZ-IEEE'97)*, 37–44.

Dubois, D., and Prade, H. 1996. What are fuzzy rules and how to use them. *Fuzzy Sets and Syst.* 84(2):169–186.

Dubois, D.; Esteva, F.; Garcia, P.; Godo, L.; de Mantaras, R. L.; and Prade, H. 1998. Fuzzy sets in case-based reasoning. *Int. J. of Int. Syst.* 13:345–373.

Dubois, D.; Hajek, P.; and Prade, H. 1997. Knowledge-driven versus data-driven logics. Technical Report 97-46R, IRIT. To appear in *J. of Logic, Language and Information*.

Dubois, D.; Prade, H.; and Ughetto, L. 1997. Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Trans. on Fuzzy Syst.* 5(3):398–417.

Ruspini, E.; Bonissone, P.; and Pedrycz, W., eds. 1998. *Hanbook of fuzzy computation*. Institute of Physics Publ.

Ughetto, L.; Dubois, D.; and Prade, H. 1997. Efficient inference algorithms with fuzzy inputs. In *Proc. of the 6th IEEE Int. Conf. on Fuzzy Syst. (FUZZ-IEEE'97)*, 567–572.

Ughetto, L. 1997. Utilisations conjointes de différents types de règles floues. In *Proc. 7⁰ Renc. Franc. sur la Log. Floue et Appl. (LFA'97)*, 269–276. Cépaduès, Toulouse.

Weisbrod, J. 1996. A combined approach to fuzzy reasoning. In *Proc. of the 4th Europ. Conf. on Intel. Tech. and Soft Comp. (EUFIT'96)*, 554–557.

Zadeh, L. 1979. A theory of approximate reasoning. In J.E. Hayes, D. M., and Mikulich, L., eds., *Machine Intell.* NY, USA: Elsevier. 149–194.

Zadeh, L. 1992. The calculus of fuzzy if-then rules. *AI Expert* 7(3):23–27.