Efficient Approximate Reasoning with Positive and Negative Information

Chris Cornelis, Martine De Cock, and Etienne Kerre

Fuzziness and Uncertainty Modelling Research Unit,
Department of Applied Mathematics and Computer Science,
Ghent University,
Krijgslaan 281 (S9), B–9000 Gent, Belgium
{chris.cornelis,martine.decock,etienne.kerre}@UGent.be
http://fuzzy.UGent.be

Abstract. Starting from the generic pattern of the Generalized Modus Ponens, we develop an efficient yet expressive quantitative model of approximate reasoning that tries to combine "the best of different worlds"; following a recent trend, we make a distinction between *positive* or observed ("guaranteed") fuzzy rules on one hand, and *negative* or restricting ones on the other hand, which allows to mend some persistent misunderstandings about classical inference methods. To reduce algorithm complexity, we propose inclusion–based reasoning, which at the same time offers an efficient way to approximate "exact" reasoning methods, as well as an attractive implementation to the concept of reasoning by analogy.

Keywords: approximate reasoning, positive and negative information, possibility theory, inclusion measures

1 Introduction and Motivation

Reasoning with imprecise information expressed as fuzzy sets (possibility distributions) has received much attention over the past 30 years. More specifically, researchers have undertaken various attempts to model the following reasoning scheme (an extension of the modus ponens logical deduction rule), known as Generalized Modus Ponens (GMP):

$$\begin{array}{l} \textbf{IF } X \text{ is } A \textbf{ THEN } Y \text{ is } B \\ \quad X \text{ is } A' \\ \hline \qquad\qquad\qquad\qquad Y \text{ is } B' \end{array}$$

where X and Y are assumed to be variables taking values in the respective universes U and V; furthermore $A, A' \in \mathcal{F}(U)$ and $B, B' \in \mathcal{F}(V)$[1].

[1] By $\mathcal{F}(U)$ we denote all fuzzy sets in a universe U, i.e. mappings from U to $[0, 1]$.

M.Gh. Negoita et al. (Eds.): KES 2004, LNAI 3214, pp. 779–785, 2004.

Traditionally, the if–then rule is represented by a fuzzy relation R (a fuzzy set in $U \times V$), and to obtain an inference B' about Y, the direct image $R' \uparrow_{\mathcal{T}} A$ of A' under R by means of a t–norm[2] $\mathcal{T}$ is computed[3], i.e. for v in V,

$$B'(v) = R \uparrow_{\mathcal{T}} A'(v) = \sup_{u \in U} \mathcal{T}(A'(u), R(u,v)) \tag{1}$$

R is typically modelled by either a t–norm $\mathcal{T}$ or an implicator[4] $\mathcal{I}$, such that for all u in U and v in V

$$R(u,v) = \mathcal{T}(A(u), B(v)) \tag{2}$$

$$\text{or,}\quad R(u,v) = \mathcal{I}(A(u), B(v)) \tag{3}$$

This choice gives rise to the *conjunction–based*, resp. *implication–based* model of approximate reasoning (see e.g. [1]). Also (1) can be easily generalized to a batch of parallel fuzzy rules (as in a fuzzy expert system); in this paper we do not consider this extended setting.

Two important points should be made w.r.t. this “de facto” procedure:

1. Regarding *semantics*, Dubois et al. [4] recently pointed out that when R is modelled by a t–norm as in (2), the application of (1) invokes undesirable behaviour of the reasoning mechanism.
2. Regarding *complexity*, the calculation of the supremum in (1) is a time–consuming process. When $|U| = m$ and $|V| = n$, the complexity of a single inference amounts to $\mathcal{O}(mn)$.

We are convinced that these arguments can be identified as the main causes why the application of approximate reasoning has been restricted so far to simple control tasks, and why only crisp numbers are used as input values to the GMP (as in Mamdani controllers). In this paper, starting from the distinction between positive and negative information in the light of possibility theory (Section 2), in Section 3 we present a unified reasoning mechanism that takes into account a rule’s intrinsic nature. Section 4 tackles the efficiency issue: we show that inclusion–based approximate reasoning, as a natural tool for reasoning by analogy, may reduce complexity to $\mathcal{O}(m + n)$ without harming the underlying rule semantics.

2 Positive and Negative Information in Possibility Theory

Possibility theory is a formalism that tries to capture in mathematical terms imprecise (typically, linguistic) information about the more or less plausible values

---

[2] A t–norm $\mathcal{T}$ is an increasing, commutative, associative $[0,1]^2 \to [0,1]$ mapping that satisfies $\mathcal{T}(x,1) = x$ for all x in $[0,1]$.

[3] This procedure is also known as Compositional Rule of Inference (CRI).

[4] An implicator $\mathcal{I}$ is a $[0,1]^2 \to [0,1]$ mapping with decreasing first and increasing second partial mappings that satisfies $\mathcal{I}(0,0) = 1$ and $\mathcal{I}(1,x) = x$ for all x in $[0,1]$.

that a variable may assume. For instance, a statement like “decent houses in Gent do not come cheap” flexibly restricts plausible buying prices by pointing out that a low budget will not get you very far in this town. Compare this to “I found a nice place in Gent for about 100.000 EUR”, which gives a guarantee (by explicit example) that properties in a given price range can be found. The examples we quoted are instances of what is called negative or constraint–based and positive or observation–based information respectively in the literature [4]: the first kind rules out certain values for the buying price X, while the second kind only designates certain observed values as “definitely possible” while saying nothing about the rest.

To mathematically represent both types of information, possibility distributions and guaranteed possibility distributions were introduced. Formally, a possibility distribution π_X on a variable X (e.g. buying price) in a universe U is a $U \rightarrow [0,1]$ mapping such that $\pi_X(u) = p$ means that it is (at most) possible to degree p that X can take the value u. Possibility degrees typically emerge by evaluating a fuzzy set like “expensive” and subsequently imposing

$$\pi_X(u) \leq \mathit{expensive}(u) \tag{4}$$

The inequality allows for the existence of other information items restricting X (specifically, new constraints can make the possibility degrees decrease). On the other hand, a guaranteed possibility distribution δ_X on U is another $U \rightarrow [0,1]$ mapping such that $\delta_X(u) = d$ means that it is (at least, or guaranteed) possible to degree d that X can take the value u. In our example, δ_X would be obtained by setting

$$\delta_X(u) \geq \text{about-100.000-EUR}(u) \tag{5}$$

Analogously to (4), the inequality is meant to imply that later on new observations can make the guaranteed possibility degrees increase. Clearly, constraint–based and observation–based information induce dual types of inequalities; it is generally assumed that $\delta_X(u) \leq \pi_X(u)$ for $u \in U$, an integrity constraint expressing that impossible values cannot be guaranteed.

3 A Unified Framework for Approximate Reasoning with Positive and Negative Information

Intuitively, a rule reflects some pattern or regularity from real life. It gains strength when a lot of instances in which the regularity holds are observed, and when only few counterexamples occur. Formally, in a crisp setting, given A in $\mathcal{P}(U)$, B in $\mathcal{P}(V)$, the couple (u, v) in $U \times V$ is called

- a *positive example* if $u \in A$ and $v \in B$
- a *negative example* if $u \in A$ and $v \notin B$

w.r.t. the crisp rule“IF X is A THEN Y is B”. It is clear that all positive examples are given by $A \times B$ and all negative ones by $A \times coB$. Remark also that the couples (u,v) in $coA \times V$ are neither positive nor negative; in e.g. [4] they are called *irrelevant examples.* To represent a rule as a crisp relation R between U and V there are two ways to proceed:

- *Negative Approach.* $R = co(A \times coB)$. This means that negative examples are explicitly excluded (do not belong to R) while other couples in $U \times V$ are considered possible by default for X and Y. In other words: "if X takes a value in A, then Y must certainly be in B (and can impossibly be in $co(B)$)". The rule is thus treated as a constraint, i.e. as a piece of negative information in the light of possibility theory.
- *Positive Approach.* $R = A \times B$. The rule's representation coincides with its positive examples. Positive examples are explicitly supported, while for other couples in $U \times V$, due to lack of evidence, we put $R(u,v) = 0$. In other words: "If X is in A, it's perfectly possible (but not necessary) that Y is in B". This means that R carries positive information[5].

The above can be straightforwardly extended to fuzzy sets, leading to these formulas for the representation of a fuzzy rule in the implication– and conjunction–based model; for an implicator $\mathcal{I}$ and a t–norm $\mathcal{T}$, (u,v) in $U \times V$, we distinguish between

$$R_{\mathcal{I}}(u,v) = \mathcal{I}(A(u), B(v)) \tag{6}$$

$$R_{\mathcal{T}}(u,v) = \mathcal{T}(A(u), B(v)) \tag{7}$$

The above analysis of the anatomy of a fuzzy rule makes it possible to imagine the rule base of a fuzzy expert system (e.g. to determine a suitable price for a house) being built up of both negative rules expressing restrictions (typically obtained from experts) and positive rules expressing observed relationships (emerging e.g. from a suitable data mining process). It also reveals that both kinds of rules should be processed in a different way, as was noted in [4]. To see this, we revert to the crisp case. Assume that X's values are restricted to the crisp subset A' of U. If $R = co(A \times coB)$, then v in V cannot be *excluded* as a possible value for Y provided there exists a u from A' such that $(u,v) \in R$, i.e.

$$B' = \{v \mid v \in V \text{ and } (\exists u \in U)(u \in A' \text{ and } (u,v) \in R)\} \tag{8}$$

So $B' = R \uparrow A'$, the direct image of A' under R. On the other hand, if $R = A \times B$, then v in V can be *guaranteed* as a possible value for Y only insofar as each of the (u,v), with $u \in A'$, can be guaranteed, hence

$$B' = \{v \mid v \in V \text{ and } (\forall u \in U)(u \in A' \Rightarrow (u,v) \in R)\} \tag{9}$$

B' is also known as the subdirect image $R \lhd A'$ of A' under R. Straightforward fuzzification of these formulas using a t–norm $\mathcal{T}$ and an implicator $\mathcal{I}$ gives rise to, for v in V,

[5] Remark that R cannot be seen as a constraint, since it would mean that also irrelevant examples are excluded (do not belong to R), something the rule definitely does *not* imply.

$$B'_1(v) = R_{\mathcal{I}} \uparrow_{\mathcal{T}} A'(v) = \sup_{u \in U} \mathcal{T}(A'(u), \mathcal{I}(A(u), B(v))) \quad (10)$$

$$B'_2(v) = R_{\mathcal{T}} \lhd_{\mathcal{I}} A'(v) = \inf_{u \in U} \mathcal{I}(A'(u), \mathcal{T}(A(u), B(v))) \quad (11)$$

Notice in particular the nice dual symmetry of the inference results in the implication– and conjunction–based model; also remark that if A' is a crisp singleton of U, then (11) coincides with the application of CRI using formula (1) where $R = R_{\mathcal{T}}$. Incidentally, this explains why in Mamdani controllers the processing of positive information by means of the direct image does not cause problems[6].

4 Inclusion–Based Approximate Reasoning

Parallel to the mainstream approach to approximate reasoning based on the CRI, an extensive body of literature (see e.g. [1]) is concerned with so–called reasoning by analogy, the rationale of which can be summed up as: "Similar inputs A' should lead to similar outputs B'". This paradigm has inspired several authors to propose various kinds of similarity–based reasoning. Generically, given an if–then rule "IF X is A THEN Y is B", and an observation A' about X, this procedure can be summarized as

1. **Comparison.** A' is compared to A by means of a similarity measure $\mathcal{SIM}$, i.e. $\alpha = \mathcal{SIM}(A', A)$ is determined, with $\alpha \in [0, 1]$. The higher α, the more A' and A are considered similar.
2. **Modification.** The consequent B is modified into the conclusion B' by means of $B'(v) = f(\alpha, B)(v)$, where f is called a modification mapping. Normally, $f(1, B) = B$, and $f(\alpha, B) \neq B$ when $\alpha < 1$.

The above procedure is attractive from an efficiency point of view: the calculation of the similarity of A' and A usually takes $\mathcal{O}(m)$ time and needs to be performed only once, so overall complexity is $\mathcal{O}(m + n)$. Its semantics however fail to meet intuition. Indeed, the predominant characteristics that most people look for in a similarity measure are reflexivity and symmetry. Reflexivity assures that when $A' = A$, then $\mathcal{SIM}(A', A) = 1$, so $B' = B$ as well, a very natural integrity condition. Symmetry, on the other hand, may be harmful to our purposes, as the following example shows.

Example 1. Consider a crisp rule "IF X is in [0,100], THEN Y is in [10,20]". Now suppose that $A' = \{50\}$. Regardless of whether we take the negative or the positive view (cfr. Section 3) of this rule, we expect the conclusion to be $[10, 20]$. Similarity–based reasoning cannot obtain this: if $SIM(\{50\}, [0, 100])$ were to be equal to 1, then *by symmetry*, for the rule "IF X is 50, THEN Y is in [10,20]"

[6] One of those problems being e.g., if $A' = U$ and $R = R_T$ in (1), then $B' = V$, i.e. while one is fully uncertain about the value of X, all values for Y would be explicitly guaranteed.

and the observation $A' = [0, 100]$, the conclusion would be $[10, 20]$ as well, which is meaningless.

As we argued in [3], this problem can be mended by replacing the degree of similarity of A' and A by a degree of fulfilment or satisfaction of A by A'. In the example, being equal to 50 satisfies the constraint of being between 0 and 100, so the rule should apply. Vice versa, knowing only that X is between 0 and 100 certainly does not meet the criterion of being exactly 50, so the converse rule should not apply. A handy tool to evaluate fulfilment is an inclusion measure $\mathcal{INC}$, i.e. a $\mathcal{F}(U)^2 \to [0, 1]$ mapping that determines to what extent a fuzzy set is a subset of another one. In particular, we can use $\mathcal{INC}_\mathcal{I}$, defined by, for an implicator $\mathcal{I}$,

$$\mathcal{INC}_\mathcal{I}(A', A) = \inf_{u \in U} \mathcal{I}(A'(u), A(u)) \tag{12}$$

This measure is not symmetrical, and one may check that $\mathcal{INC}_\mathcal{I}(A, B) = 1 \iff A' \subseteq A$ when $\mathcal{I}$ is equal to the residual implicator $\mathcal{I}_\mathcal{T}$ of a t–norm $\mathcal{T}$, i.e. $\mathcal{I}_\mathcal{T}(x, y) = \sup\{\gamma \mid \gamma \in [0, 1] \text{ and } \mathcal{T}(x, \gamma) \leq y\}$ for x, y in $[0, 1]$.

The issue that remains to be settled is how to choose the modification mapping f. The following two theorems show that, for a particular choice of the connectives, a nice relationship with the inference results for fuzzy rules treated as negative, resp. positive information, can be established when we put $f = \mathcal{I}_\mathcal{T}$, resp. $f = \mathcal{T}$.

Theorem 1. *Let $\mathcal{T}$ be a left–continuous t–norm, then for v in V*

$$\sup_{u \in U} \mathcal{T}(A'(u), \mathcal{I}_\mathcal{T}(A(u), B(v))) \leq \mathcal{I}_\mathcal{T}\left(\mathcal{INC}_\mathcal{I}(A', A), B(v)\right) \tag{13}$$

Theorem 2. *Let $\mathcal{T}$ be a left–continuous t–norm, then for v in V*

$$\inf_{u \in U} \mathcal{I}_\mathcal{T}(A'(u), \mathcal{T}(A(u), B(v))) \geq \mathcal{T}\left(\mathcal{INC}_\mathcal{I}(A', A), B(v)\right) \tag{14}$$

In other words, the "exact" results (10) and (11) can be approximated by their inclusion–based counterparts. Since the infimum over U needs to be calculated only once, savings are made w.r.t. complexity ($\mathcal{O}(m + n)$ instead of $\mathcal{O}(mn)$). Moreover, the approximation is conservative in a sense that e.g. $\mathcal{I}_\mathcal{T}\left(\mathcal{INC}_\mathcal{I}(A', A), B(v)\right)$ does not impose more restrictions than the result obtained with CRI, which warrants soundness of the method. Initial experimental results in [2] indicate the strength of the approach.

5 Conclusion

The distinction between rules expressing positive and negative information opens up new directions that allow a deeper insight into the nature of approximate reasoning. As a trade–off between expressivity and efficiency, in this paper we have developed a method based on an inclusion measure, motivated in terms of fulfilment, for processing both kinds of information.

Acknowledgements

Chris Cornelis and Martine De Cock want to thank the Fund for Scientific Research–Flanders for funding the research reported on in this paper.

References

1. Bouchon–Meunier B., Dubois, D., Godo, L., Prade, H.: Fuzzy Sets and Possibility Theory in Approximate and Plausible Reasoning. In: Fuzzy sets in approximate reasoning and information systems, Kluwer Academic Publishers (1999) 15–190
2. Cornelis, C.: Two–sidedness in the Representation and Processing of Imprecise Information (in Dutch), Ph. D. thesis, Ghent University
3. Cornelis, C., Kerre, E.E.: Inclusion–Based Approximate Reasoning. In: Lecture Notes in Computer Science **2074** (2001) 221–230
4. Dubois, D., Prade, H., Ughetto, L.: A New Perspective on Reasoning with Fuzzy Rules. In: International Journal of Intelligent Systems **18(5)** (2003) 541–563