**Medical Fuzzy Expert Systems and Reasoning about Beliefs**

Hájek, Petr
1994

# INSTITUTE OF COMPUTER SCIENCE

## ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

# Medical fuzzy expert systems and reasoning about beliefs

Petr Hájek Dagmar Harmancová

Technical report No. 632

December 1994

**Institute of Computer Science, Academy of Sciences of the Czech Republic**
**Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic**
**phone: (+422) 6641 3418, (+422) 6641 4244 fax: (+422) 8585789**
**e-mail: hajek@uivt.cas.cz, dasa@uivt.cas.cz**

**INSTITUTE OF COMPUTER SCIENCE**

**ACADEMY OF SCIENCES OF THE CZECH REPUBLIC**

# Medical fuzzy expert systems and reasoning about beliefs

Petr Hájek Dagmar Harmancová

Technical report No. 632
December 1994

**Abstract**

Reasoning on belief using fuzzy logic (as present in some prominent medical fuzzy expert systems) is examined from the point of view of formal logic.

**Keywords**

many-valued logic, fuzzy logic, probabilities, nedical expert systems

# 1 Introduction

This is a working paper describing a research program in a more or less initial stage of development.
Its aim is

- to stress the difference between vagueness (fuzziness) on the one hand and uncertainty as degree of belief on the other hand (this difference seems to be disregarded in some medical expert systems)

- to call the reader's attention to a very elegant and simple formal logical system of fuzzy logic

- to present some observations and results on possibilities of handling uncertainty (in particular, probability) in fuzzy logic.

Let us start with the distinction (made by Professor Zadeh) between fuzzy logic on broad and narrow sense, the broad sense being "everything concerning fuzziness" and the narrow being "the underlying formal logical calculi" (including many-valued systems, possibly with non-standard quantifiers as "usually"etc.). Indeed, many-valued logic has proved to be a very comfortable formal home for fuzzy logic. A very crucial feature of many valued (propositional) logic is its *truth-functionality*: the truth degree of a compound formula is determined by the truth-degrees of its component. Various systems differ in *how* the truth degrees of components determine the truth degree of the compound formula (recall the theory of t-norms and conorms as possible semantics of conjunction and disjunction); but they agree in the truth-functionality. In contradiction with this, degrees of belief are *not* truth functional; e.g. it is clear that we cannot compute the probability of $A$ & $B$ ($A, B$ being crisp propositions) from probabilities of $A$ and of $B$; since $P(A \ \& \ B)$ is *not a function* of $P(A)$, $P(B)$ (if no additional assumptions of independence etc. are made). This distinction has been observed by several authors, see e.g. [2, 3]. Recall the attempts of probabilistic justification of MYCIN-like systems that turned out to be pseudoprobabilistic, (cf. [7, 6]).

Disregarding the difference between truth degree of fuzzy propositions (admitting a truth-functional calculus) and degrees of belief (e.g. probabilities) of crisp notions (which are inherently not truth-functional) brings the danger of wrong conclusions and conceptual illness, if not justified by some theoretical foundations. This is why some well-known and successful medical expert systems appear to need careful theoretical analysis which either encovers some deeper ways of interpretation of the calculus used or will result into some recommendations of re-consideration. Such analysis is our future plan; here we present some preliminary considerations that are hoped to be of independent interest.

Let us close this introduction with two examples of rules in existing systems.

Our first example is from PNEUMON-IA, a medical expert system based on the shell MILORD [4]. The rule reads:

R08004 IF 1) Community acquired pneumonia is *almost sure*
2) Bacterial disease is *possible*
3) (No aspiration) is *very possible*
THEN [Possible]
Enterobacteria is *quite possible*

There are nine "linguistic certainty values" like *quite possible, almost sure* etc.; formally, they are first handled as truth-modifiers (hedges) but later they seem to be treated as truth-degrees in a nine-valued fuzzy logic. In the former case, the question may be posed if a terminology, referring to truth rather than certainty (hence quite true, almost absolutely true) would not be more adequate; in the latter, our discussion on a truth-functional treatment of degrees of certainty may apply. Is, for example, "bacterial disease" a crisp notion, either present or absent and we have some belief ("possible") on its presence or is it a fuzzy notion which may be more or less true so that it is modifiable by a hedge "possible"? (The authors plan to analyze such questions in a joint work with authors of MILORD.)

Our second example is CADIAG-2 [8]. The rule reads:

Example 8
IF (the patient shows low back pain, and
a limitation of motion of the lumbar spine, and
a diminished chest expansions, and
the patient is male, and
is between 20 and 40 years of age)
THEN (the diagnosis may be ankylosing spondylitis)
WITH (the frequency of occurence value of the above combination with ankylosing spondylitis is *very often* [.90] and the strength of confirmation value of the combination for ankylosing spondylitis is *very strong* [.80]).

Here the members of the antecedent (like "the patient shows low back pain") and the succedent ("diagnosis ankylosing spondylitis") are understood as fuzzy propositions having some truth degrees and also the rule is understood as defining a fuzzy relation (in fact two relations: occurrence and strength); fuzzy relational calculus is applied *but* relative frequences are taken as truth values of the relations; relative frequences are like probabilities and are, in general, not truth functional. Can this be justified?

Let us be explicit in saying that we do not question the quality of practical results obtained by these systems which may well be very high due to careful "tuning" as well as to the simplicity of deduction patterns used; but the question of a theoretical justification remains.

The main question of this paper reads as follows:

It is tempting to deal with beliefs (probabilities) of crisp propositions as if they were truth degrees of some fuzzy propositions and to deal with conditional probabilities as if they were truth degrees of some implications. Can one do this? If so, how? What can one conclude?

Let us mention two related papers. First [9] discusses the problem of probabilities of fuzzy propositions and fuzzy propositions on probabilities (like "the probability of $p$ is large"). [1] discusses, among other things, inference mechanisms of fuzzy logic that derive upper and lower bounds for degrees of (un)certainty. We shall present a result which is more specific concerning the calculus (just a variant of Łukasiewicz's logic) but more general concerning the proof mechanism (arbitrary graded proof allowed).

The rest of the paper is organized as follows: In Section 2 we survey a simplified version of Pavelka's fuzzy logic, needed in following sections, and introduce some notational conventions. In Section 3 and 4 we offer two approaches to the problem of dealing with degrees of belief (probabilities) of crisp propositions as with truth degrees of some (other) fuzzy propositions. In Section 5 we conclude with some remark.

## 2 Rational Pavelka's logic

Here we survey a simplified version of Pavelka's variant of Łukasiewicz's logic (RPL) as presented in [5].

**Formulas** are built from propositional atoms $p_1, p_2, \ldots$ and truth constants $\overline{r}$ for each *rational* $r \in [0,1]$ using connectives $\rightarrow$ and $\neg$; other connectives are defined thus:

$$\begin{array}{lll} \varphi \mathbin{\&} \psi & \text{stands for} & \neg(\varphi \rightarrow \neg\psi) \\ \varphi \underline{\vee} \psi & \text{stands for} & \neg\varphi \rightarrow \psi \\ \varphi \vee \psi & \text{stands for} & (\varphi \rightarrow \psi) \rightarrow \psi \\ \varphi \wedge \psi & \text{stands for} & \neg(\neg\varphi \vee \neg\psi) \\ \varphi \leftrightarrow \psi & \text{stands for} & (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi) \end{array}$$

Thus e.g. $(\overline{0.7} \rightarrow p)\underline{\vee}((q\&r) \rightarrow \overline{0.6})$ is a formula.

Truth degrees are *reals* from the unit interval [0,1].

The truth functions for $\rightarrow, \neg$ are:

$$\begin{array}{rll} r \rightarrow s & = 1 & \text{if } r \leq s, \\ & = 1 - r + s & \text{otherwise}, \\ \neg r & = 1 - r. & \end{array}$$

This determines the truth functions for other connectives, thus $(r \mathbin{\&} s) = max(0, r+s-1)$, $(r\underline{\vee}s) = min(r+s, 1)$, $(r \vee s) = max(r,s)$, $(r \wedge s) = min(r,s)$.

A *graded formula* is a pair $(\varphi, r)$ where $\varphi$ is a formula and $r \in [0,1]$ is rational.

A *fuzzy theory* is a mapping associating to each formula a rational number - its degree of being an axiom (fuzzy set of formulas, rational-valued, or a certain set of graded formulas). In particular, we have the fuzzy theory of logical axioms.

*Logical axioms* are

(i) Rose-Rosser's axioms (all in degree 1)

$$\varphi \rightarrow (\psi \rightarrow \varphi)$$

$$(\varphi \rightarrow \psi) \rightarrow ((\psi \rightarrow \chi) \rightarrow (\varphi \rightarrow \chi))$$

$$(\neg\varphi \rightarrow \neg\psi) \rightarrow (\psi \rightarrow \varphi)$$

$$((\varphi \rightarrow \psi) \rightarrow \psi) \rightarrow ((\psi \rightarrow \varphi) \rightarrow \varphi)$$

(ii) bookkeeping axioms: (for arbitrary rational $r, s \in [0,1]$):

$$\overline{r} \text{ in degree } r,$$
$$\overline{\neg r} \leftrightarrow \neg\overline{r} \text{ in degree } 1,$$
$$\overline{r \rightarrow s} \leftrightarrow (\overline{r} \rightarrow \overline{s}) \text{ in degree } 1.$$

*Deduction rules* are:
*modus ponens*: from $(\varphi, r)$ and $(\varphi \to \psi, s)$ derive $(\psi, r \mathbin{\&} s)$
*truth constant introduction*: from $(\varphi, s)$ derive $(\overline{r} \to \varphi, r \to s)$.

A *graded proof* from a fuzzy theory $T$ is a sequence of graded formulas

$$(\varphi_1, r_1), \dots (\varphi_n, r_n)$$

such that for each $i$, $(\varphi_i, r_i)$ is a logical axiom (i.e. $\varphi_i$ is a logical axiom in degree $r_i$) or $(\varphi_i, r_i)$ is an axiom of $T$ (i.e. $T(\varphi_i) = r_i$) or $(\varphi_i, r_i)$ follows from some previous member(s) of the sequence by a deduction rule. A formula $\varphi$ is *provable in $T$ in degree $r$* if there is a graded proof from $T$ whose last element is $(\varphi, r)$. The *provability degree* of $\varphi$ is $|\varphi|_T = sup\{r \mid \varphi \text{ provable in } T \text{ in degree } r\}$. Caution: $|\varphi|_T$ may be irrational.

Note that this logic has a natural semantics and a completeness theorem (saying that the truth degree of $\varphi$ in $T$ equals to the provability degree of $\varphi$ in $T$); see [5] for definitions and details. Here we only stress the general notion of a graded proof, fully analogous to the notion of proof in crisp logic.

We shall discuss general many-valued logics but RPL will be our outstanding example.

To discuss our main question, let us distinguish propositional variables and connectives of crisp (two-valued) logic ($p, q, \dots, -, \cap, \cup, \to$) and of fuzzy logic $fp, fq, \dots, \neg, \&$ or $\wedge$, $\underline{\vee}$ or $\vee$ [or both], $\to$). Assume a fuzzy logic to be given, i.e. choice of truth values, connectives and their truth tables. Investigate graded formulas (i.e. pairs $(\varphi, \alpha)$ where $\varphi$ is a formula and $\alpha$ a truth degree) and assume some sound fuzzy logical axiom system (if $(\varphi, \alpha)$ is an axiom then each evaluation $e$ of atoms satisfies $e(\varphi) \geq \alpha$). Assume we have a conjunction & and implication $\to$ such that the fuzzy modus ponens is sound: for each $e$ if $e(\varphi) \geq \alpha$ and $e(\varphi \to \psi) \geq \beta$ then $e(\psi) \geq \alpha \mathbin{\&} \beta$. This is in particular the case of RPL. But a sound deductive system with modus ponens is available for many many-valued logics. Note that each graded formula $(\varphi, \alpha)$ defines some conditions on the values of atoms of $\varphi$ in an evaluation $e$ necessary and sufficient to the fact that $e(\varphi) \geq \alpha$. E.g. $e(p \to q) \geq \alpha$ in Lukasiewicz's logic iff $e(p) \mathbin{\&} \alpha \leq e(q)$.

# 3 Direct translation

One just *identifies* propositional variables of crisp logic with those of fuzzy logic and interprets each formula $\varphi$ of fuzzy logic by its crisp counterpart $c(\varphi)$. (Clearly, various formulas non-equivalent in fuzzy logic may have counterparts equivalent in crisp logic, notably $\varphi \mathbin{\&} \psi$ and $\varphi \wedge \psi$ are not equivalent in fuzzy logic but both go to $\varphi \cap \psi$.)

We work with fuzzy theories $T$ in the above sense; $T$ is a fuzzy set of formulas. We write $(\varphi, \alpha) \in T$ for $T(\varphi) = \alpha$.

Let WPL (weak Pavelka's logic) stand for RPL without truth constants and corresponding axioms and deduction rules, i.e. just Łukasiewicz's logic but with fuzzy theories.

Let $P$ be a probability on crisp formulas; a fuzzy theory $T$ *respects* $P$ if for each $\varphi, P(c(\varphi)) \geq T(\varphi)$.

**Fact.** If $T$ respects $P$ and $T \vdash^w (\varphi, r)$ (i.e. $T$ proves $(\varphi, r)$ in WPL) then $P(c(\varphi)) \geq r$.

**Proof** standard (just check modus ponens: if $P(A) \geq r$ and $P(\neg A \vee B) \geq s$ then $P(B) \geq r \,\&\, s$.)

**Example.** If $T = \{(p, \alpha), (q, \alpha)\}, 1/2 < \alpha < 1$ then $T \vdash^w (p \,\&\, q, \alpha \,\&\, \alpha)$ (where $\alpha \,\&\, \alpha$ is $\max(0, 2\alpha - 1)$), also $T \vdash^w (p \wedge q, \alpha \,\&\, \alpha)$ but *not* $T \vdash^w (p \wedge q)(\alpha)$.

On the other hand, $T \vdash (p \wedge q, \alpha)$ (in RPL) due to the completeness theorem.

This fact relies heavily on Lukasiewicz's implication; in next section we present a more general approach.

# 4 Linguistic (cautious) translation

We associate with each crisp formula $\varphi$ a fuzzy propositional variable $f\varphi$ (read: $\varphi$ *is PROBABLE*, or *PROBABILITY_OF_*$\varphi$ *is HIGH*). This is a fuzzy proposition; given a joint probability $P$, we are free to define $e(f\varphi) = P(\varphi)$, i.e. assign $P(\varphi)$ as the truth-value of $f\varphi$. We have $f(\neg\varphi) \equiv \neg f(\varphi)$ (i.e. the last formula has value 1 under the evaluation $e$ above; but, on the contrary $f(\varphi \to \psi) \equiv (f\varphi \to f\psi)$ need not have value 1 and likely for other binary connectives. Once more: the formulas [$\varphi$ *is PROBABLE* & $\psi$ *is PROBABLE*] and [($\varphi$ & $\psi$) *is PROBABLE*] may well have *different* truth values.

One easy way how to read a "knowledge base" consisting of pairs $(\varphi_i, \alpha_i)$ where $\varphi_i$ is a crisp formula and $\alpha_i \in [0, 1]$ is a fuzzy theory consisting of pairs $(\varphi_i^*, \alpha_i)$ where $\varphi_i^*$ results from $\varphi_i$ by replacing each atom $p$ by $fp(PROBABLE\_p)$. Thus for example, a "rule" $(p \to q, \alpha)$ will be read $(PROBABLE\_p \to PROBABLE\_q, \alpha)$ and express the fact that $P(p) \,\&\, \alpha \leq P(q)$.

Thus we may investigate knowledge bases $K$ (fuzzy axiom systems) consisting of graded formulas $(\varphi, \alpha)$ where $\varphi$ is a formula built from atoms $fp_1, \ldots, fp_n$. (Typically, $K$ will consist of rules of the form e.g. $(fp_1 \,\&\, \neg fp_2$
$\ldots \,\&\, fp_k \to fq, \alpha))$. Data $D$ may have form $(fp_i, \alpha_i)$ or $(\neg fp_i, \alpha_i)$ for some $i$; inference is proving in the fuzzy theory $K \cup D = T$. If both the data and the knowledge base is consistent with probability $P$ (i.e. $(\varphi(fp_1, \ldots fp_n), \alpha) \in T$ implies that the expression $\varphi(P(p_1), \ldots, P(p_n))$ is $\geq \alpha$ [caution: we do NOT mean $P(\varphi(p_1 \ldots, p_n)) \geq \alpha$ !!] and $T$ proves $(\psi(fp_1, \ldots fp_n), \beta)$ then we know $\psi(P(p_1), \ldots P(p_n)) \geq \beta$; in particular, if $T$ proves $(fq, \beta)$ for an atom $q$, then we know $P(q) \geq \beta$.

Observe that this is true for *any* fuzzy logic satisfying our minimal conditions, *not just for Lukasiewicz!*

Here we get in fact only information on probabilities of *atoms*; one has to remember that e.g. the rule

$$((fp \,\&\, fq) \to fr, 0.9)$$

is understood as saying $P(p) \,\&\, P(q) \,\&\, 0.9 \leq P(r)$, *not* anything about conditional probabilities.

# 5 Concluding remarks

(1) *Fuzzy probabilistic logic.*
If we allow all atoms $PROBABLE\_\varphi$ (for each $\varphi$, briefly $PBLE\_\varphi$) we may formulate axioms sound for each probability, e.g.

$$(PBLE\_\varphi \to PBLE\_(\varphi \cup \psi), 1)$$

$$(PBLE\_(\varphi \cap \psi) \to (PBLE\_\varphi \wedge PBLE\_\psi), 1)$$

$$(PBLE\_\varphi \,\&\, PBLE\_\psi) \to PBLE\_(\varphi \cap \psi), 1)$$

etc.

Can we have an axiomatization probabilistically complete in some sense?

For a fuzzy theory $T$ whose atoms are $PBLE\_p$ for $p$ *atomic* we have some trivial observations: assume $T$ is consistent, then it has a model - evaluation $e$ respecting all the axioms (now we work in the logic RPL). But then we may assume that $e(PBLE\_p_i) = P(p_i)$ for some probability $P$ since we may freely choose probabilities of propositional atoms; the axioms say *nothing* on the probabilities of compound formulas.

(2) *Possibilities of use in expert systems*
If we think in probabilistic terms in expert systems we are mainly interested in *conditional* probabilities: we have a piece $E$ of knowledge (information) and are interested in $P(H|E)$, $H$ being e.g. a hypothesis. $E$ may for example imply $p_1$; then $P(p_1|E) = 1$.On the contrary, we have stressed that our use of truth degrees as expressing inequalities concerning expressions built from (unconditional) probabilities of atomic formulas does not allow us to express general conditional probabilities. Is there any help?

We suggest the following: Do not think of a knowledge base as of *one* unconditional probability on atoms but on a *system of conditional probabilities* $P_i(\ldots) = P(\ldots|E_i)$, $i = 1, \ldots, n$, where $E_i$ are some possible pieces of evidence. Assume further that each of the probabilities $P_i$ satisfies the conditions expressed by the knowledge base; but a single $P_i$ may of course satisfy more, e.g. it may be the case that $P_1(p) = 1$. (This may be given by the available data.) Caution: we do not assume that the $E_i$ run over all possible events by which we may conditionalize; these are just evidences expected or admitted or considered in uses of the knowledge base. For example if our knowledge base consists of the single formula $p \to q$ with the truth value 0.8 then we assume that for all $i$ we have $P_i(p) \,\&\, 0.8 \leq P_i(q)$, thus an $E$ implying $p \,\&\, \neg q$ and of positive probability cannot be one of $E_i$'s. We do not claim that this understanding of a fuzzy knowledge base is to be recommended; we only say that *if* one wants the truth values of atoms to be their probabilities then ours is one possible way how to do that.

(3) Our last remark: we repeatedly stress the working character of the paper. Nevertheless, we hope that it might contribute to better understanding of the problem of dealing with beliefs in fuzzy logic and serve as a starting point of theoretical (logical) analysis of various particular approaches, not necessarily only those mentioned above.

**Acknowledgements**

Thanks are due to Lluis Godo for valuable comments. The work on this paper was partially suported by the COPERNICUS grant 10053 - MUM (Managing uncertainty in medicine).

# Bibliography


[1] Bonissone, P. P. Sumarizing and propagating uncertain information with triangular norms. *Int.Journal of Approximate Reasoning 1* (1987), 71–101.

[2] Dubois, D., and Prade, H. Non-standard theories of uncertainty in knowledge representation and reasoning. *Knowledge Engeneering Review 9*, 4 (1994), 399–416.

[3] Godo, L., and López de Mántaras, R. Fuzzy logic. In *Encyclopaedia of Computer Science* (1993).

[4] Godo, L., López de Mántaras, R., Sierra, C., and Verdaguer, A. MILORD: The architecture and the management of linguistically expressed uncertainty. *International Journal of Intelligent Systems 4*, 4 (1989), 471–501.

[5] Hájek, P. Fuzzy logic and arithmetical hierarchy. *Fuzzy Sets and Systems* (to appear).

[6] Hájek, P., Havránek, T., and Jiroušek, R. *Uncertain Information Processing in Expert Systems.* CRC Press, Boca Raton, Florida, 1992.

[7] Johnson, R. W. Independence as Bayesian updating methods. In *Uncertainty in Artificial Intelligence* (1986), Kanal and Lemmer, Eds., Elsevier Science Publishers B.V. North Holland, pp. 197–200.

[8] Kolarz, G., and Adlassnig, K.-P. Problems in establishing the medical expert systems CADIAG-1 and CADIAG-2 in rheumatology. *Journal of Medical Systems 10*, 4 (1986), 395–405.

[9] Zadeh, L. Is probability theory sufficient for dealing with uncertainty in AI: A negative view. In *Uncertainty in Artificial Intelligence* (1986), Kanal and Lemmer, Eds., Elsevier Science Publishers B.V. North Holland, pp. 103–115.