A SYNTHETIC VIEW OF APPROXIMATE REASONING TECHNIQUES

Henri PRADE

Laboratoire "Langages et Systemes Informatiques"
Universite Paul Sabatier, 118 route de Narbonne
3 1062 TOULOUSE CEDEX
FRANCE

Abstract: This paper presents a review of different approximate reasoning techniques which have been proposed for dealing with uncertain or imprecise knowledge ,especially in expert systems based on production rule methodology. Theoretical approaches such as Bayesian inference,Shafer's belief theory or Zadeh's possibility theory as well as more empirical proposals such as the ones used in MYCIN or in PROSPECTOR, are considered. The presentation is focused on two basic inference schemes : the deductive inference and the combination of several uncertain or imprecise evidences relative to a same matter. Several kinds of uncertainty are taken into account in the models which are described in the paper : different degrees of certainty or of truth may be associated with the observed or produced facts or with the "if.., then..." rules; moreover the statements of facts or of rules may be imprecise or fuzzy and the values of the degrees of certainty which are used may be only approximately known. An extensive bibliography, to which it is referred in the text, is appended.

Key words: uncertainty; imprecision; inference; combination of evidence; belief function; possibility; necessity; degree of truth; fuzzy set; production systems; expert systems.

1-Introduction: The pieces of information with which we have to deal in decision or reasoning processes, may be uncertain, imprecise or even vague, incomplete, mutually inconsistent, and time-varying. Thus,so-called approximate,inexact, plausible reasoning methods are strongly needed in knowledge engineering. Inference procedures with uncertainty are becoming more important in rule-based expert systems since the knowledge given by human experts is often uncertain or imprecisely stated.

For a long time, the Bayesian model had been the only numerical approach to inference with uncertainty,since no quantification was introduced in the patterns of plausible reasoning analyzed by Polya[33]. Several mathematical models of uncertainty, which depart from the usual probability approach, have been recently proposed, particularly Shafer's belief theory[40]and Zadeh's possibility theory [53]. In the sane time, many researchers in Artificial Intelligence have felt a need for alternatives of the standard Bayesian approach (see[45] for a discussion) and have proposed and used (generally with success) more empirical models, particularly in expert systems such as *MYCIN* [41], *PROSPECTOR* [11], *CASNET* [481, *SPERIL* [20],(see also [27],[13],[27],[39]).

In the following, we try to present a synthetic view of most of these proposals within the compass of a small number of pages. The paper is organized in three main parts. The first part is devoted to an unified presentation of the different mathematical approaches of uncertainty, including Shafer's belief theory, Zadeh's possibility and fuzzy set theories and probability theory. Then, the deductive inference scheme with various kinds of uncertainty is dealt with in the framework of probability and/or possibility theory. Lastly, the combination of several uncertain or imprecise evidences, relative to a same matter and possibly partially inconsistent, is discussed using Dempster-Shafer's approach.

II-Different mathematical approaches of uncertainty embedded in a common framework

a)- representaing an uncertain body of evidence and measuring the uncertainty of events :

We start with an elementary presentation of Shafer's belief theory [48](see also[15] or[5] for an A.I.-oriented introduction),whose probability theory and Zadeh's possibility theory are particular cases,as it will be seen later.

Let U be the exhaustive set of the possible (and mutually exclusive) values of a quantity X. X may be a numerical quantity or just a "parameter" that takes non-numerical values. U will be supposed finite for simplicity, except if the contrary is explicitly stated. U is sometimes called the universe of discourse or the frame of discernement. In the following, we are interested in propositions of the form "the true value of X is in A",or more briefly "X is in A",where $A \subseteq U$. Let $\mathcal{P}(U)$ denote the set of subsets of U; $\mathcal{P}(U)$ may be viewed as the set of the events,each event corresponding to a proposition.

Let m be a set function from $\mathcal{P}(U)$ to[0,1],such that

$$m(\emptyset) = 0; \sum_{A \subseteq U} m(A) = 1 \qquad (1)$$

m is called a basic probability assignment and is supposed to represent an uncertain body of evidence regarding the value of some quantity X. A certain body of evidence would be represented by a statement 'X is in F, exactly and certainly' (i.e,m(F)=1 and $\forall A \neq F$, m(A)=0); note that a certain body of evidence does not necessarily lead to a precise specification of the value of X(the value of X would be known with precision only if F is a singleton of U). When the body of evidence is uncertain, the subset F is randomnized (indeed mathematically speaking,(1)expresses that m represents a random set). m is a probability distribution on $\mathcal{P}(U)$,not onU. Thus,m(A) is not the probability of the proposition 'X is in A' in general. The subsets $A \subseteq U$ such that m(A)>0 are called the focal elements,they correspond to the possible contents of the uncertain evidence under consideration.

From m a so-called belief (or credibility) function Cr,from $\mathcal{P}(U)$ to[0,1], is defined as

$$\forall A \subseteq U,\ Cr(A) = \sum_{B \subseteq A} m(B) \qquad (2)$$

Thus the credibility of the proposition "X is in A" is computed from the values of m attached to the propositions 'X is in B' which entail 'X is in A'.
By duality a so-called plausibility function Pl, from $\mathcal{P}(U)$ to [0,1], is defined as, the overbar denoting the complement, $\forall A \subseteq U, Pl(A) = 1 - Cr(\bar{A})$ (3)

$$= \sum_{A \cap B \neq \emptyset} m(B) \quad (4)$$

Thus the plausibility of the proposition 'X is in A' is computed from the values of m attached to the propositions which do not entail "X is not in A"; the plausibility of an event corresponds to the non-credibility of the opposite event, which is natural. It can be checked that

. $Cr(\emptyset)=0$; $Pl(\emptyset)=0$; $Cr(U)=1$; $Pl(U)=1$ (5)
. if $A \subseteq B$, then $Cr(A) \leq Cr(B)$, $Pl(A) \leq Pl(B)$ (6)
. $\forall A \subseteq U, Pl(A) \geq Cr(A)$ (7)
. $\forall A \subseteq U, Cr(A)+Cr(\bar{A}) \leq 1, Pl(A)+Pl(\bar{A}) \geq 1$ (8)

All these properties are in agreement with the intuitive meaning generally attached to the concepts of plausibility and credibility. For instance (8) shows that two opposite events may seen both plausible while it is allowed to find an event not credible and the opposite event too. In this framework the situation of total ignorance can be easily modelled, while it is not possible to deal with ignorance in an effective manner in the framework of probability theory. With this apporach it is possible to distinguish between the lack of belief and disbelief since $Cr(A)=0$ does not entails $Cr(\bar{A})=1 \Leftrightarrow Pl(A)=0$, while $Cr(\bar{A})=1$ entails $Cr(A) = 0$.

This model does not seem in disagreement with the basic ideas used by Colby and Smith[5] in their empirical but quantitative approach of belief.

Two important particular cases of credibility and plausibility functions are got for special structures of the set of focal elements.

- When the only possible focal elements are singletons of U (i.e. m(A)=0 as soon as A has more than one element), we clearly recover a probability measure P=Cr=Pl and (3) reduces to $P(A)+P(\bar{A})=1$. Thus probability corresponds, in terms of m, to an evidence which is precise but "inconsistent" (the focal elements are mutually exclusive since they are singletons).
- When the set of focal elements can be ordered in a nested sequence $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n$, it can be shown (see for instance[40] that

$\forall A \subseteq U, \forall B \subseteq U, Cr(A \cap B) = \min(Cr(A), Cr(B))$ (9)
$\forall A \subseteq U, \forall B \subseteq U, Pl(A \cup B) = \max(Pl(A), Pl(B))$ (10)

(9) and (10) contrast with $\forall A \subseteq U, \forall B \subseteq U$, if $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$ (11)

A plausibility function that satisfies (10) is nothing but a possibility measure in the sense of Zadeh[54]; a credibility function which satisfies (9) is called a necessity measure (see Dubois Prade[8]) since the duality expressed by (3) mirrors the fact that the necessity of an event corresponds to the impossibility of the opposite event in modal semantics. In the following, a possibility measure and the dual measure of necessity will be respectively denoted by Π and N, the relation $\forall A \subseteq U, N(A) = 1 - \Pi(\bar{A})$ (12)
contrasts with the probabilistic situation where $P(A) = 1 - P(\bar{A})$.

Possibility and necessity correspond, in terms of m, to an evidence which is consistent (all the focal elements are nested; in that case Shafer[40] speaks of consonant plausibility and belief functions) but which is not precise in general with respect to the specification of the value of the quantity X under consideration (at most one focal element is a singleton).

(5), (9) and (10) yield as immediate consequences

. $\forall A \subseteq U, \min(N(A), N(\bar{A}))=0, \max(\Pi(A), \Pi(\bar{A}))=1$ (13)
. $\forall A \subseteq U, N(A)>0 \rightarrow \Pi(A)=1, \Pi(A)<1 \rightarrow N(A)=0$ (14)

(14) expresses that an event must be completely possible before being somewhat necessary. An event whose probality or necessity is equal to 1 can be regarded as certain, while it is not the case for an event whose possibility is 1 since the opposite event may also have a possibility equal to 1.

An important consequence of (10) is the fact that a possibility measure (and then the dual necessity measure) can be expressed in terms of a so-called possibility distribution π, from U to [0,1], defined by $\forall u \in U, \pi(u) = \Pi(\{u\})$ and then

$$\forall A \subseteq U, \ \Pi(A) = \max_{u \in A} \pi(u) \quad (15)$$

$$\forall A \subseteq U, \ N(A) = \min_{u \in \bar{A}} (1 - \pi(u)) \quad (16)$$

The extension of (15) and (16) to non-finite sets such that $\mathbb{R}^n$, is obvious. Since $\Pi(U) = 1$, π is normalized in the sense that $\max_{u \in U} \pi(u)=1$. When π is the characteristic function of a crisp subset Q of U (then m(Q)=1 and m(A)=0 if A≠Q), (15) and (16) reduce respectively to $\Pi(A) = \begin{cases} 1 & \text{if } A \cap Q \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$ and $N(A) = \begin{cases} 1 & \text{if } A \supseteq Q \\ 0 & \text{otherwise} \end{cases}$, which shows the links between possibility and intersection and between necessity and inclusion. When evaluating the possibility of an event, only the most favorable case is taken into account as it is shown by (15), which departs from probability where the evaluation is cumulative ($P(A) = \sum_{u \in A} P(\{u\})$).

From a practical point of view, it is important that probability, possibility and necessity measures can be directly expressed in terms of a distribution which require only |U|-1 numerical values to be defined, while $2^{|U|}-2$ values are needed to define a function m.

N.B.: For a discussion of the axiomatics of plausibility, credibility functions, possibility, probability and necessity measures in a common framework, see Dubois Prade[9]; see also[9] for a "possibilistic" interpretation of histograms concurrently with the usual probabilistic interpretation; this double interpretation enables to build possibility, probability and necessity measures based on a same body of evidence such that $\forall A, N(A) \leq P(A) \leq \Pi(A)$, which is satisfactory for the intuition.

b) - Fuzzy events and vague statements : From an uncertain body of evidence, we may want to evaluate the probability, the possibility, or the necessity of a proposition which is itself vaguely stated. Zadeh[51] has introduced the concept of a fuzzy set in order to represent vague predicates such as "tall", "large", ... More recently, in PRUF, this author has proposed to represent a vague proposition by equating the possibility distribution of the variable whose value is restricted by the proposition, with the membership function of the fuzzy set representing the meaning of the vague predicate (possibly compound) present in the proposition; this approach enables also to deal with fuzzy quantifiers such as 'most', 'several', 'a few', ... See[53].

Let $\tilde{\mathcal{P}}(U)$, the set of fuzzy subsets of U. A fuzzy set is defined by its membership function μ_A from U to [0,1]. Usual set-operations are extended to fuzzy sets in the following way[51](see[8]for alternative definitions).

. complementation $\forall u\in U, \mu_{\bar{A}}(u)=1-\mu_A(u)$ (17)

. intersection $\forall u\in U, \mu_{A\cap B}(u)=\min(\mu_A(u),\mu_B(u))$ (18)

. union $\forall u\in U, \mu_{A\cup B}(u)=\max(\mu_A(u),\mu_B(u))$ (19)

The usual properties of set-operations are preserved except the non-contradiction and excluded-middle laws which are weakened in $\forall u\in U, \min(\mu_A(u),\mu_{\bar{A}}(u))\leq\frac{1}{2}$ and in $\forall u\in U, \max(\mu_A(u),\mu_{\bar{A}}(u))\geq\frac{1}{2}$; it is not surprising that an ill-defined set somewhat overlaps its complement which is itself not-accurately defined. The fuzzy set inclusion is defined by $A\subseteq B \Leftrightarrow \forall u, \mu_A(u)\leq\mu_B(u)$ (20). Note that viewing μ_A and μ_B as two possibility distributions representing two uncertain bodies of evidence, the possibility and the necessity measures defined from μ_B are respectively greater and smaller than the possibility and the necessity measures defined from μ_A, which is quite natural since the body of evidence represented by μ_B is less informative than the one represented by μ_A. The standard definitions of the probability, of the possibility and of the necessity of a fuzzy event $A\in\tilde{\mathfrak{P}}(U)$ are respectively given by

$P(A)=\sum_{u\in U}\mu_A(u).p(u)$ with $p(u)=P(\{u\})$ (Zadeh[52]) (21)

$\Pi(A)=\max_{u\in U}\min(\mu_A(u),\pi(u))$ with $\pi(u)=\Pi(\{u\})$ (Zadeh[53]) (22)

$N(A) = \min_{u\in U}\max(\mu_A(u), 1-\pi(u))$ (23)

When U is not finite, '$\sum$' is replaced by a Lebesgue-Stielges integral, 'max' and 'min' by 'sup' and 'inf'. (21) is the expression of an expectation and (11) still holds using (18) and (19). It can be easily checked that (22) and (23) generalizes (15) and (16) and that (9), (10), (12) still holds using (17)-(19) as well as $\forall A\in\tilde{\mathfrak{P}}(U), \Pi(A)\geq N(A)$; however, only weaker versions of (13) and (14) hold, due to the lack of non-contradiction and excluded-middle laws for fuzzy sets. From a multi-valued logic point of view, (22) still evaluates the non-emptiness of a generalized intersection while (23) is a degree of inclusion. The fact that probability, possibility and necessity measures are increasing with respect to set inclusion as defined by (20), is compatible with the entailment principle (see[54]) "if X is in A, then X is in B" as soon as $A\subseteq B$ in the sense of (20). It may seems surprising that a measure of a fuzzy event is a scalar rather than a fuzzy number; indeed an alternative approach which yields fuzzy numbers exists (see[8] for instance), but scalar evaluations are generally sufficient for practical purposes. Smets [43] has defined the plausibility and the credibility of fuzzy event in a natural way as upper and lower expectations (see[6]). Yager[50] has introduced generalized plausibility and credibility functions defined from fuzzy focal elements (see also[21]).

c) non-interactivity : The notion of non-interactivity in possibility theory plays a role analogous to that of independence in probability theory. Two quantities or variables, X and Y, which take their values respectively on U and V are said to be non-interactive if their join possibility distribution $\pi_{X,Y}$, from UxV to [0,1] supposed to be normalized, is such that

$\forall(u,v)\in U\times V, \pi_{X,Y}(u,v)=\min(\pi_X(u), \pi_Y(v))$ (24)

where π_X and π_Y are the marginal possibility distributions obtained from $\pi_{X,Y}$ by projection :

$\pi_X(u)=\sup_{v\in V}\pi_{X,Y}(u,v)$ and $\pi_Y(v)=\sup_{u\in U}\pi_{X,Y}(u,v)$ (25)

(24) expresses the fact that the fuzzy set of the possible values of X does not depend on the value of Y and reciprocally. Defining the cartesian product of two fuzzy sets A and B of U and V respectively, by

$\forall u\in U, \forall v\in V, \mu_{A\times B}(u,v)=\min(\mu_A(u),\mu_B(v))$ (26)

it can be shown, if X and Y are non-interactive variables (see[35]), that

$\forall A\in\tilde{\mathfrak{P}}(U), \forall B\in\tilde{\mathfrak{P}}(V), \Pi_{X,Y}(A\times B)=\min(\Pi_X(A),\Pi_Y(B))$ (27)

$\Pi_{X,Y}(A+B)=\max(\Pi_X(A),\Pi_Y(B))$ (28)

$N_{X,Y}(A\times B)=\min(N_X(A),N_Y(B))$ (29)

$N_{X,Y}(A+B)=\max(N_X(A),N_Y(B))$ (30)

where $A+B=\overline{\bar{A}\times\bar{B}}$, and $\Pi_{X,Y}, N_{X,Y}, \Pi_X, N_X, \Pi_Y, N_Y$ are the possibility and necessity measures built from $\pi_{X,Y}$, π_X and π_Y respectively. When X and Y are interactive, the sign '=' is replaced by '$\leq$' in (27) and (28) and by '$\geq$' in (29) and (30).

Remark : In MYCIN[41], a measure of belief and a measure of disbelief in the hypothesis h knowing the evidence e, respectively denoted by MB(h,e) and MD(h,e), are used. From the definitions of MB(h,e) and of MD(h,e) in terms of the probabilities P(h|e) and P(h), it can be easily checked that ('$\neg$' means the negation)

$MB(\neg h,e) = MD(h,e)$ (31)

$MB(h,e)>0 \rightarrow MD(h,e)=0; 1-MD(h,e)<1 \rightarrow MB(h,e)=0$ (32)

(31) and (32) are analogous to (12) and (14) respectively, viewing MB(h,e) as a necessity measure and MD(h,e) as the complement to 1 of a possibility measure. Indeed, the following formulae are used in MYCIN ($\wedge$ and $\vee$ denote the conjunction and the disjunction respectively).

$MD(h_1\wedge h_2,e)=\max(MD(h_1,e),MD(h_2,e))$; $MD(h_1\vee h_2,e)=\min(MD(h_1,e), MD(h_2,e))$ (33)

$MB(h_1\wedge h_2,e)=\min(MB(h_1,e),MB(h_2,e))$; $MB(h_1\vee h_2,e)=\max(MB(h_1,e),MB(h_2,e))$ (34)

which are the exact counterparts of (27)-(30).

d) degree of truth : We close this background on the representation and the measurement of uncertainty with a short discussion of the relation between the notion of a degree of truth and possibility and necessity measures. A degree of truth may be viewed as a measure of the conformity between a representation of the contents of the proposition under consideration and a representation of what is actually known of the reality. Thus the degree of truth of a proposition is relative to our state of knowledge. Let us suppose that the uncertain body of evidence is represented by a possibility distribution π_e and that the contents of the proposition p="X is in A' under consideration is also represented by a possibility distribution π_p (which is crisp if p is a non-vague statement). Then, using (22) and (23) we can compute the possibility that p is true given e:

$\Pi(p,e)=\sup_{u\in U}\min(\pi_p(u), \pi_e(u))$ (35)

and the necessity that p is true given e :

$$N(p,e)=\inf_{u\in U} \max(\pi_p(u), 1-\pi_e(u)) \quad (36)$$

N.B.: Cayrol, Farreny, Prade [4] have designed a procedure of pattern-matching where $\Pi(p,e)$ and $N(p,e)$ are used in order to evaluate the semantic similarities between patterns and data.
It has been shown that either when π_e is the characteristic function of a singleton (i.e. the evidence is precise) or when the proposition p is non-vague, the quantity

$$v(p,e)=\frac{\Pi(p,e)+N(p,e)}{2} \quad (37)$$

can be regarded as a genuine degree of truth in the sense that the truth-functionality is preserved for the negation, the conjunction and the disjunction; see Dubois Prade [10] and also Gaines [14]. For instance, we have

$$v(\neg p,e) = 1 - v(p,e) \quad (38)$$

As pointed out in Prade [34], the certainty factor $CF(h,e)=MB(h,e)-MD(h,e)$, used in MYCIN, can be viewed as a degree of truth up to a scaling effect since

$$v(h,e)=\frac{1+CF(h,e)}{2} \quad (39)$$

coincides with (37) in the analogy presented in the remark of section II.c. Zadeh [54] has introduced a fuzzy degree of truth named 'compatibility'. The compatibility of π_p with respect to π_e is a fuzzy set $CP(p,e)$ of the real interval $[0,1]$ whose membership function is defined by

$$\forall r\in[0,1],\ \mu_{CP(p,e)}(r)= \sup_{r=\pi_p(u)} \pi_e(u) \quad (40)$$
$$= 0 \text{ if } \pi_p^{-1}(r) = \emptyset$$

$CP(p,e)$ is nothing but the fuzzy set of the possible values of the variable $\pi_p(u)$ when the possible values of u are restricted by the possibility distribution π_e. It can be shown [35] that

$$\Pi(p,e) = \sup_{r\in[0,1]} \min(r,\mu_{CP(p,e)}(r)) \quad (41)$$

$$N(p,e) = \inf_{r\in[0,1]} \max(r,1 - \mu_{CP(p,e)}(r)) \quad (42)$$

Thus $CP(p,e)$ contains the informations $\Pi(p,e)$ and $N(p,e)$. Using extended operations on fuzzy numbers [8], $CP(\neg p,e)$ can be easily expressed in terms of $CP(p,e)$, and $CP(p\wedge q,e)$ or $CP(p\vee q,e)$ in terms of $CP(p,e)$ and of $CP(q,e)$, See [35], [7].
Conversely, given the possibility distribution π_p of the proposition under consideration, and its fuzzy truth-value τ, it is possible to get the greatest solution (in the sense of fuzzy set inclusion) π_e^+ of the equation $\tau=CP(p,e)$; we have

$$\forall u\in U, \pi_e^+(u) = \mu_\tau(\pi_p(u)) \quad (43)$$

π_e^+ represents what we may conclude about the reality knowing that the proposition p is τ-true. See [54], [35]. Note that $\pi_e^+ = \pi_p$ for $\mu_\tau(r)=r$, $\forall r$; if τ is a crisp subset of $[0,1]$ (particularly if τ is a scalar value) π_e^+ is non fuzzy.

III - Deductive inference with uncertainty :

a) - The general problem under consideration : Let us consider a causal link between a variable Y taking its values in V and a variable X taking its values in U, expressed by a collection of rules,

if X is in Ai, then Y is in Bi with gi (44)

where $\forall_i$, $A_i\subseteq U$, $B_i\subseteq V$ and g_i is a degree of uncertainty of the rule such as a probability or a possibility degree or is a truth-value; the subsets A_i and B_i may be fuzzy; moreover, whatever its nature g_i may be only approximately known and the set of its possible values represented by a fuzzy number. Given some uncertain body of evidence concerning the value of X, the problem consists in deducing what can be said about the value of Y taking into account the uncertainty of the rules and the fact that they may be only vaguely stated. Using representations in terms of probability or in terms of possibility, this general problem can receive at least approximate solutions in general. In the following, we review the main results.

b) - Complete description of the causal link : Let us suppose we know a possibility distribution π_X of the variable X and a conditional possibility distribution $\pi_{Y|X}$ defined on VxU which restricts the possible values of Y when the value of X is given. Then the possibility distribution $\pi_{X,Y}$ attached to the pair (X,Y) is obtained as

$$\forall u\in U, \forall v\in V, \pi_{X,Y}(u,v)=\min(\pi_{Y|X}(v,u), \pi_X(u)) \quad (45)$$

By projection we get the possibility distribution of Y

$$\forall v\in V, \pi_Y(v)=\sup_{u\in U} \min(\pi_{Y|X}(v,u), \pi_X(u)) \quad (46)$$

The analogous of (46) in terms of probability is clearly

$$\forall v\in V\ p_Y(v)=\sum_{u\in U} p_{Y|X}(v,u).p_X(u) \quad (47)$$

The problem in possibility theory of deriving $\pi_{X|Y}$ or $\pi_{X|Y}$ from $\pi_{X,Y}$ remains open, however see the proposals of Nguyen [30], Hisdal [18], Zadeh [53]. We may think of using operators other than min in (46); other noticeable eligible operators are the product or the operator T defined by $T(a,b)=\max(0,a+b-1)$; however, min is the greatest of all the possible eligible operators (see [35]). In general $\pi_{Y|X}$ or $\pi_{Y|X}$ are not available since the A_i's and B_i's of (44) are not just singletons; then we have to deal with an incomplete description of the causal link.

c) - Incomplete description of the causal link :

α) - in terms of probability: Let us consider the n rules "if X is in A_i, then Y is in B_i" with $P(B_i|A_i)$ (and Y is in $\overline{B_i}$ with $P(\overline{B_i}|A_i)=1-P(B_i|A_i)$ together with the rule "if X is in Ao, then Y is in V ($P(V|Ao)=1$) where $(Ao,A_1,\dots,A_n)$ forms a partition of U (i.e. the A_i's are exhaustive and mutually exclusive), then from p_X and these collection of rules, we can deduce

$$p_Y(v)=\sum_{v\in B_j} \frac{P(Bj|Aj)}{|Bj|}.P(Aj)+\sum_{v\notin B_k} \frac{P(\overline{B_k}|Ak)}{|\overline{B_k}|}.P(Ak)+\frac{P(Ao)}{|V|}$$

$$\text{where } P(Ai)=\sum_{u\in Ai} p_X(u) \quad (48)$$

with the underlaying assumption that the probability is distributed uniformly in the designated subset because no information is available. See [21]. Ishizuka, Fu and Yao [21] have proposed an extension of this approach when the Ai's are fuzzy sets.

β) - In terms of possibility : Again we consider a collection of rules

if X is in Ai, then Y is in Bi is τ_i true (49)

where the Ai's and the Bi's may be normalized fuzzy sets (which can be viewed as possibility distributions)

and where τi is a fuzzy truth-value (i.e. a fuzzy set of [0,1] which may represent linguistic truth-qualifications, see [54], [7]). Note that there is no hypothesis of mutual exclusiveness or of exhaustiveness concerning the Ai's.
Each rule (49) expresses a partial knowledge about the unknown possibility distribution $\pi_{Y|X}$. More precisely "if X is in Ai, then Y is in Bi" can be understood as

$$\forall v\in V, \mu_{Bi}(v) \geq \sup_{u\in U} \min(\pi_{Y|X}(v,u), \mu_{Ai}(u)) \quad (50)$$

(50) expresses that the possibility distribution of Y, computed from (46) with $\pi_X = \mu_{Ai}$ is included in Bi.
The greatest solution of (50) with respect to fuzzy set inclusion is given by (see [38], [35])

$$\forall u\in U, \forall v\in V, \pi_{Y|X}(u,v) = \mu^1_{Ai\to Bi}(u,v) = \begin{cases} 1 \text{ if } \mu_{Ai}(u) \leq \mu_{Bi}(v) \\ \mu_{Bi}(v) \text{ if } \mu_{Ai}(u) > \mu_{Bi}(v) \end{cases} \quad (51)$$

$\mu^1_{Ai\to Bi}$ defined by (51) corresponds to a multivalent implication which is sometimes known as Gödel-Brouwer implication (see [14]); when we use the product instead of min in (50), we get

$$\forall u\in U, \forall v\in V, \mu^2_{Ai\to Bi}(u,v) = \min(1, \frac{\mu_{Bi}(v)}{\mu_{Ai}(u)}) \text{ if } \mu_{Ai}(u) \neq 0 \quad (52)$$
$$= 1 \text{ if } \mu_{Ai}(u) = 0$$

which corresponds to a multivalent implication already considered by Goguen [17] and Gaines [14]; lastly, when we use $T(a,b) = \max(0, a+b-1)$, we get

$$\forall u\in U, \forall v\in V, \mu^3_{Ai\to Bi}(u,v) = \min(1, 1-\mu_{Ai}(u)+\mu_{Bi}(v)) \quad (53)$$

which corresponds to Łukasiewicz implication. It can be shown that

$$\mu^3_{Ai\to Bi} \geq \mu^2_{Ai\to Bi} \geq \mu^1_{Ai\to Bi} \quad (54)$$

Using (43), (46) and (51), from the rule i and an uncertain body of evidence concerning the value of X represented by π_X, supposed to be normalized, we deduce

$$\forall v\in V, \pi_Y(v) = \sup_{u\in U} \min(\mu_{\tau i}(\mu^1_{Ai\to Bi}(u,v)), \pi_X(u)) \quad (55)$$

It can be noticed that $\mu_{Ai\times Bi}$ (defined by (26)) is also a solution of (50), however it is not the greatest solution generally and its use in (55) leads to a smaller π_Y (in the sense of fuzzy set inclusion) which thus would be arbitrarily precise. When $\mu_{\tau i}(r) = r$, $\forall r$, if $\pi_X = \mu_{Ai}$ (and more generally if $\pi_X \leq \mu_{Ai}$), we get $\pi_Y = \mu_{Bi}$ by (55) which is in conformity with the usual modus ponens; if $\pi_X = \mu_{\overline{A}i}$, $\pi_Y(v) = 1$, $\forall v\in V$ (i.e. Y remains indeterminate) which is natural; as soon as π_X is not completely included in μ_{Ai}, some level of indetermination appears (more precisely it can be shown that $\forall v$, $\pi_Y(v) \geq \sup_{\mu_{Ai}(u)=0} \pi_X(u)$ (see [35]).
It must be pointed out that (55) does not allow to infer from "if X is in Ai, then Y is in Bi" and from "X is almost in Ai" that "Y is almost in Bi" except if we introduce fuzzy tolerance relations R_u and R_v which enable to enlarge Ai and Bi respectively in $\sup_{u'\in U} \min(\mu_{Ai}(u), \mu_{R_U}(u,u')) \geq \mu_{Ai}(u)$ and in $\sup_{v'\in V} \min(\mu_{Bi}(v), \mu_{R_V}(v,v')) \geq \mu_{Bi}(v)$, μ_{R_U} and μ_{R_V} modeling approximate equalities; see [35].
Provided that the collection of rules is consistent with the existence of a unique $\pi_{Y|X}$, the global result is obtained by aggregating the different results given by (55) for each rule by means of the idempotent operator 'min'.
N.B. : The approximate reasoning scheme which leads to (55) can also be discussed equivalently in terms of fuzzy truth-values introducing the compatibility of μ_{Ai} with respect to π_X and the compatibility of μ_{Bi} with respect to π_Y, see Baldwin [2]; it can also be shown that (55) generalizes the rule of detachment in multivalent logics, see [7] and [35]. For other discussions see [25], [28], [42], [46], [47].
Remark : If we suppose that a possibility measure Π and its dual necessity measure N are defined on a set of propositions p, q, ..., then it can be shown that (Prade [35])

$$.N(p\to q) \geq a \text{ and } N(p) \geq b \to N(q) \geq \min(a,b) \quad (56)$$
$$.N(p\to q) = 1 \text{ and } \Pi(p) \geq b \to \Pi(q) \geq b \quad (57)$$
$$.\Pi(p\to q) \geq a \text{ and } N(p) = 1 \to \Pi(q) \geq a \quad (58)$$

compared to

$$.P(p\to q) \geq a \text{ and } P(p) \geq b \to P(q) \geq \max(0, a+b-1) \quad (59)$$
$$.P(q|p) \geq a \text{ and } P(p) \geq b \to P(q) \geq a.b \quad (60)$$

where P is a probability measure.

γ) - Mixed problem : Lastly, we consider a situation where probability and possibility are mixed together. We have the n + 1 rules
if X is in Ai, then Y is in Bi   i=1,n
if X is in Ao, then Y is in V where the $A_o, A_1, \ldots, A_n$ are non-fuzzy and form a partition of U and the Bi's are normalized fuzzy sets. A distribution of probability $p_i = P(A_i)$ $(\sum_{i=o}^{n} p_i = 1)$ is known. The problem is to know what can be said of the proposition "Y is in B" in terms of probability. As it has been recognized by Zadeh [56], the information consists of a probability distribution and of a conditional possibility distribution $\pi_{Y|X}(v,u) = \mu_{Bi}(v)$ if $u\in Ai$. This problem has been considered by Dempster [6] when the Bi are non-fuzzy. When they are, using the degree of intersection $\Pi(B,Bi)$ and the degree on inclusion of Bi in B obtained from (22) and (23), i.e.

$$\Pi(B,Bi) = \sup_{v\in V} \min(\mu_B(v), \mu_{Bi}(v));$$
$$N(B,Bi) = \inf_{v\in V} \max(\mu_B(v), 1-\mu_{Bi}(v))$$

we are in position to compute the plausibility

$$Pl_Y(B) = \sum_{i=1}^{n} p_i . \Pi(B|Bi) + p_o \quad (61)$$

and the credibility $Cr_Y(B) = \sum_{i=1}^{n} p_i . N(B|Bi)$ (62)

where (61) and (62) are straightforward generalizations of (4) and (2) using degrees of intersection and of inclusion respectively. Lastly, when the values of the pi's are approximatively known and represented by fuzzy numbers, expressions (61) and (62) can be generalized and remain computationally tractable using recent results in fuzzy arithmetics by Dubois and Prade (see [35]).

IV- Combining : In this section we successively deal with the combination of uncertain bodies of evidence and with the combination of uncertain informations

a)- combining uncertain bodies of evidence : Let m_1 and m_2 be two basic probability assignments (in the sense of (1)) representing two uncertain bodies of evidence relative to a same matter. Dempster's rule of combination [6] enables to combine them in order to get a new basic probability assignment m defined by

$$m(\emptyset)=0; \forall C\neq\emptyset, m(C)= \frac{\sum_{A_i \cap B_j = C} i,j\ m_1(A_i)\cdot m_2(B_j)}{\sum_{A_i\cap B_j \neq \emptyset} i,j\ m_1(A_i)\cdot m_2(B_j)} \quad (63)$$

Note that m does not exist when there is no common part between the focal elements of m_1 and those of m_2.

Dempster's rule of combination is associative. Zadeh [55] has discussed the normalization in (63) which may be unsuitable since when the denominator of (63) is not 1, we have $\sum_{A_i\cap B_j=\emptyset} i,j\ m_1(A_i).m_2(B_j)>0$ which evaluates the degree to which the two evidences are dissonant; the normalization conceals the existence of this dissonance. Some particular cases of (63) are noticeable. When m_1 and m_2 reduce to two probability distributions p_1 and p_2, (63) gives back the combined probability distribution p : $\forall u\in U, p(u)= \frac{p_1(u).p_2(u)}{\sum_{u\in U} p_1(u).p_2(u)}$ (64)

If U has only two elements u and $\bar{u}$ (i.e. there are only two possibilities), (64) gives $p(u) = \frac{p_1(u).p_2(u)}{1-p_1(u)-p_2(u)+2p_1(u).p_2(u)}$ (65)

This aggregation operation has been extensively used by Kayser [22]. When m_1 reduces to p_1 and $m_2(A)=1$ (which entails $m_2(B)=0, \forall B\neq A$), (63) gives the usual conditioning $p(u)= \frac{p_1(u)}{p_1(A)}$ if $u\in A$ and $p(u)=0$ if $u\notin A$ with $p_1(A)=\sum_{u\in A} p_1(u)$.

If m_1 and m_2 are such that, $\exists A, m_1(A)=1-m_1(U)$ and $m_2(A)=1-m_2(U)$, i.e. A is the unique and common focal element (apart U) of m_1 and m_2. Then (63) gives for m, whose focal elements are also A and U

$$\begin{cases} m(A)=m_1(A)+m_2(A)-m_1(A).m_2(A) \\ m(U)=(1-m_1(A)).(1-m_2(A)) \end{cases} \quad (66)$$

When m_1 and m_2 reduce to two possibility distributions π_1 and π_2, the result given by (63) is not a possibility distribution except in some noticeable particular cases (when apart of U, m_1 and m_2 have only one focal element each, but not necessarily the same). If we want to obtain a possibility distribution π, several proposals can be made with some justifications (see [35]), among them, we have the analogous of (64)

$$\forall u\in U, \pi(u)=\frac{\min(\pi_1(u),\pi_2(u))}{\sup_{u\in U} \min(\pi_1(u),\pi_2(u))} \quad (67)$$

In (67), the use of the normalization may be discussed as in the case of Dempster's rule and we may think of using other operations than 'min', see [35].

b) - combining uncertain informations : First we briefly recall the Bayesian model. Knowing the conditional probabilities $P(h|e_1)$ and $P(h|e_2)$ of the hypothesis h when the evidences e_1 and e_2 are respectively observed the probability of h when we observe e_1 and e_2 is equal to (see [21] or [45])

$$P(h|e_1,e_2)=\frac{P(h|e_1).P(h|e_2)}{P(h)} \cdot \frac{P(e_1).P(e_2)}{P(e_1,e_2)} \quad (68)$$

where P(h) is the a priori probability of the hypothesi . When the two sources of evidence are independent, $P(e_1,e_2)=P(e_1).P(e_2)$ and (68) can be simplified.

In PROSPECTOR [11], [12], a rule of combination, which is derived of the Bayesian model with some additional assumptions for "subjective" probability has been proposed; some limitations of this rule has been pointed out in [32]. See also [31]. In MYCIN [40], the belief or disbelief measures are combined in the following way

. If $MB(h,e_1)>0$ and $MB(h,e_2)>0$ (which entails $MD(h,e_1)=MD(h,e_2)=0$) (69)

$$MB(h,e_1\wedge e_2)=MB(h,e_1)+MB(h,e_2)-MB(h,e_1).MB(h,e_2)$$

. If $MD(h,e_1)>0$ and $MD(h,e_2)>0$ (which entails $MB(h,e_1)=MB(h,e_2)=0$) (70)

$$MD(h,e_1\wedge e_2)=MD(h,e_1)+MD(h,e_2)-MD(h,e_1).MD(h,e_2)$$

In case of a positive measure of belief and a positive measure of disbelief there is a conflict. Note that (69) and (70) are similar to (66). Adams ['11 has shown that (69) or (70) are consistent with Bayesian model provided that the two sources of evidence are independent (remember that MB(h,e) and MD(h,e) are defined from P(h|e) and P(h)). Lastly, Ishizuka, Fu, Yao [19] have proposed a slightly different rule of combination for the certainty factors CF(h,e) in MYCIN.

V - Concluding remarks : The intented purpose of this paper is to give an unified view of the mathematical models we have at our disposal in order to deal with uncertainty. The presentation has particularly emphasized two recent theories which are still ill-known specially in A.I.: Shafer's belief theory and Zadeh's possibility theory (see [23], [49], [26], [44] for examples of use of this latter theory in diagnosis and expert systems). The main applications of these models to two inference schemes, particularly important for applications to expert systems, the deductive inference and the combination, have been surveyed and the link with more empirical approaches, such as the one used in MYCIN, has been stressed. Unfortunately it was not possible to present each proposal in great details within the compass of a small number of pages. Lastly, important classes of approximate or plausible reasoning techniques have not been considered here, especially analogical reasoning and default reasoning (see [36]); although these approaches do not use numerical quantification in general it would be interesting to consider them in an enlarged synthesis in the future.

REFERENCES :

[1] Adams, J.B. (1976) A probability model of medical reasoning and the MYCIN model. Math. Biosci. 32, 177-186.

[2] Baldwin, J.F. (1979) A new approach to approximate reasoning using a fuzzy logic. Fuzzy Sets & Syst 2, 309-325.

[3] Barnett, J.A. (1981) Computational methods for a mathematical theory of evidence. Proc. 7th I.J.C.A.I., Vancouver, 868-875.

[4] Cayrol, M., Farreny, H., Prade, H. (1982) Fuzzy patter matching. Kybernetes, 11, 103-116.

[5] Colby, K.M., Smith, D.C. (1969) Dialogues between hu mans and an artificial belief systems. Proc. 1st, I.J.C.A.I., Washington, D.C., 319-324.

[6] Dempster, A.P. (1967) Upper and lower probabilitie induces by a multivalent mapping. Ann. Math. Stat. 38, 325-339.

[7]Dubois D., Prade H.(1979)Operations in a fuzzy-valued logic,Infor.& Cont.,43,n° 2,224-240.
[8]Dubois D.,Prade,H.(1980) Fuzzy Sets & Systems:Theory and Applications.Academic Press,New York.
[9]Dubois D.,Prade H.(1982)On several representations of an uncertain body of evidence.In:Fuzzy Information and Decision Processes(N.M.Gupta,E.Sanchez,eds.),North-Holland,167-181).
[10]Dubois D.,Prade H.(1982)Degree of truth and truth-functionality.Proc.2nd World Conf.on Maths.at the Service of Man,Las Palmas, 262-265.
[11]Duda R.,Gachnig J.,Hart P.(1981)Model design in the Prospector consultant system for mineral exploration.In:Expert Systems in the Micro-Electronic Age (D.Michie,ed),Edinburgh,Univ.Press,153-167.
[12]Duda R.O.,Hart P.E. Nilsson N.J.(1976)Subjective bayesian methods for rule-based inference systems.Tech. Note n° 124,SRI-Inter,Menlo Park.
[13]Friedman L.(1981)Extended plausible inference. Proc.7th I.J.C.A.I., Vancouver,487-495.
[14]Gaines B.R.(1976)Foundations of fuzzy reasoning. Int.J.Man-Machine Stu.,8, 623-668.
[15]Garvey T.D.,Lowrance J.D.,Fischler MA. (1981) An inference technique for integrating knowledge from disparate sources.Proc.7th I.J.C.A.I.,Vancouver,319-325.
[16]Giles R.(1980)A computer program for fuzzy reasoning.Fuzzy Sets & Syst. 4, 221-234.
[17]Goguen J.A.(1969)The logic of inexact concepts. Synthese, 19, 325-373.
[18]Hisdal E.(1978)Conditional possibilities.Independence and noninteraction.Fuzzy Sets & Syst.1,283-297
[19]Ishizuka M.,Fu K.S.,Yao J.T.P.(1981)A theoretical treatment of certainty factor in production systems. Memo CE-STR-81-6,Purdue Univ.,In.
[20]Ishizuka M.,Fu K.S.,Yao J.T.P.(1981)Inexact inference for rule-based damage assessment of existing structures.Proc.7th I.J.C.A.I.,Vancouver,837-842.
[21]Ishizuka M.,Fu K.S.,Yao J.T.P.(1981)Inference procedure with uncertainty for problem reduction method.Memo TR-EE-81-33,Purdue Univ.,In.
[22]Kayser D.(1979) Vers une modélisation du raisonnement"approximatif".Proc.Colloq."Représentation des Connaissances et Raisonnement dans les Sciences de l'Homme"(M.Borillo,ed.),Saint-Maximin,Publ.by INRIA, 440-457.
[23]Lesmo L.,Saitta L.,Torasso P.(1983)Fuzzy production rules:A learning methodology.To appear in:Advanced in Fuzzy Set Theory and Applications(P.P.Wang,ed.), Plenum Press.
[24]Lesser V.R.,Reed S.,Pavlin J.(1980)Quantifying and simulating the behavior of knowledge-based interpretation systems.Proc.1st Annual N.Conf.Art.Intellig. Stanford,111-115.
[25]Mamdani E.H.(1977)Application of fuzzy logic to approximate reasoning using linguistic systems.IEEE Trans.on Comp. C-26, 1182-1191.
[26]Martin-Clouaire R.(1982)A computer-aided medical diagnosis method based on a fuzzy set theoretical approach.M.S.Thesis,Univ.of Saskatchewan.
[27]Michalski R.S.,Chilausky R.L.(1980)Knowledge acquisition by encoding expert rules versus computer induction from examples:a case study involving soybean pathology.Int.J.Man-Machine Stu.12, 63-88.
[28]Mizumoto M.,Fukami S.,Tanaka K.(1979)Fuzzy conditional inferences and fuzzy inferences with fuzzy quantifiers.Proc.6th I.J.C.A.I.,Tokyo,589-591.
[29]Ng S.W.,Walker A.(1980)Max-min chaining of weighted causal assertions is loop free.Proc.1st Annual Nat. Conf.Art.Intellig.,Stanford,105-107.
[30]Nguyen H.T.(1978)On conditional posssibility distributions Fuzzy Sets & Syst.,1, 299-309.
[31]Pearl J.(1982)Reverend Bayes in inference engines:A distributed hierarchical approach.Proc.AAAI.
[32]Pednault E.P.D.,Zucker S.W.,Muresan L.V.(1981) On the independence assumption underlying subjective bayesian updating.Art.Intellig.16, 213-222.
[33]Pólya G.(1974)Mathematics and plausible reasoning.Vol.II:Patterns of plausible inference.Princeton Univ.Press,2nd Ed.,1968.
[34]Prade H.(1980)Plausible reasoning for artificial intelligence.Proc.of Selected Papers of the 1st Int. Conf.on Art.Intellig. & Information-Control Syst.of Robots.Bratislava,57-64.
[35] Prade H.(1982)Modèles mathématiques de l'imprécis et de l'incertain en vue d'applications au raisonnement naturel.Thèse d'Etat,Toulouse III.
[36]Reiter R.(1980)A logic for default reasoning.Art. Intellig.,13, 81-132.
[37]Rescher N.(1976)Plausible Reasoning.Van Gorcum, Amsterdam.
[38]Sanchez E.(1977)Solutions in composite fuzzy relations equations:Application to medical diagnosis in Brouwerian logic.In:Fuzzy Automata and Decision Processes(M.M.Gupta,G.N.Sari-is,B.R.Gaines,ed.),North-Holland,221-234.
[39]Schefe P.(1980)On foundations of reasoning with uncertain facts and vague concepts.Int.J.of Man-Machine Stu.,12, 35-62.
[40] Shafer G.(1976)A mathematical Theory of Evidence. Princeton Univ.Press.
[41]Shortliffe E.H., Buchanan B.G.(1975)A model of inexact reasoning in medicine.Math.Biosci.,23,351-379.
[42]SilvaO.de O., Souza C.de R.(1982)Changing the fuzzy rule of detachment.Proc.12th IEEE Int.Symp.Multiple-Valued Logic,Paris,267-271.
[43] Smets P.(1981)Medical diagnosis and degrees of belief.Fuzzy Sets & Syst.5, 259-266.
[44] Soula G.,Sanchez E.(1982)Soft deduction rules in medical diagnosis processes.In:Approximate Reasoning in Decision Analysis(MM.Gupta,E.Sanchez,eds.)North-Holland,77-88.
[45]Szolovits P.Pauker S.G.(1978)Categorical and probabilistic reasoning in medical diagnosis.Art.Intellig. 11, 115-144.
[46] Tsukamoto Y.(1979)An approach to fuzzy reasoning method.In:Advances in Fuzzy Set Theory and Applications (M.M.Gupta,R.K.Ragade,Yager,eds.) North-Holland,137-149.
[47] Umano M.,Mizumoto M.,Tanaka K.(1979)A system for fuzzy reasoning.Proc.6th I.J.C.A.I.Tokyo,917-919.
[48] Weiss S.M.,Kulikowski C.A.,Amarel S.,Safir A.(1978) A model-based method for computer-aided medical decision making.Art.Intellig.,11, 145-172.
[49] Whalen T.,Schott B.(1981)Fuzzy production systems for decision support.Porc.Int.Conf.Cyber.& Society, Atlanta,649-653.
[50] Yager R.R.(1982)Generalized probabilities of fuzzy events from fuzzy belief structures.Tech.Rep.MII-203, Iona College, New Rochelle,N.Y.
[51] Zadeh L.A.(1965)Fuzzy Sets.Infor.& Cont.8,338-353.
[52] Zadeh L.A.(1968)Probability measures of fuzzy events J. Math.Anal.& Appl. 23, 421-427.
[53] Zadeh L.A.(1978)Fuzzy sets as a basis for a theory of possibility.Fuzzy Sets & Syst.1, 3-28.
[54] Zadeh L.A.(1979)A theory of approximate reasoning. Machine Intelligence, 9, Elsevier, 149-194.
[55] Zadeh L.A.(1979) On the validity of Dempster's rule of combination of evidence.Memo UCB/ERL/M79/24,Berkeley.
[56] Zadeh LA.(1979)Fuzzy sets and information granularity.In:Advances in Fuzzy Set Theory and Applications,3-18.