Fuzzy relation equations and causal reasoning

Didier DUBOIS – Henri PRADE

Institut de Recherche en Informatique de Toulouse – C.N.R.S.
Université Paul Sabatier, 118 route de Narbonne
31062 TOULOUSE Cedex – FRANCE

Abstract : Fuzzy relation-based models for handling uncertainty (in a non-probabilistic way) in diagnosis problems have been introduced by Sanchez about twenty years ago, and since then applied and developed by several researchers. The paper first reviews the existing proposals, also including the association-based abductive model proposed by Reggia and his co-workers, which uses non-fuzzy relations. A new model is presented in order to have a more expressive representation capable of handling uncertainty, and also of distinguishing, i) between manifestations which are certainly absent from those which are not (yet) observed, and ii) between manifestations which cannot be caused by a given disorder and manifestations for which we do not know if they can or cannot be caused by this disorder. This new model is based on possibility theory and the so-called twofold fuzzy sets, previously introduced by the authors.

Key words : Fuzzy relation equation ; diagnosis; abduction ; possibility ; certainty.

Contribution to a special issue of Fuzzy Sets and Systems entitled
"Equations and Relations on Ordered Structures : Mathematical Aspects and Applications"
(A. Di Nola, W. Pedrycz, S. Sessa, eds.)

1 - Introduction

Among various schemes of fuzzy reasoning which have been proposed at the end of the seventies in order to manage uncertainty in knowledge-based systems, fuzzy relation-based models for diagnosis problems stand apart from the mainstream of research on fuzzy rule-based expert systems. Fuzzy relation-based models describe the association of causes (disorders, deseases, …) and effects (manifestations, symptoms, …) by means of a fuzzy relation ; then possible causes of a set of observed effects are identified by solving a fuzzy relation equation (Sanchez[23]). A fuzzy relation enables the modeller to weight the association between a disorder (or a set of simultaneous disorders, see Sanchez [24]) and a manifestation (or a set of manifestations), and a fuzzy set can represent the levels of presence (intensity, or certainty of presence according to the cases) of manifestations. Although different applications have been presented, e.g. Tsukamoto and Terano [26], Shahinpoor and Wells[25], Pappis and Sugeno [17], Adlassnig et al. [1], Asse et al. [2], Kitowski and Bargiel [16], Gmytrasiewicz et al. [15], this approach to abductive reasoning has not encountered the same interest as the deductive approach to approximate reasoning based on combination and projection of possibility distributions, introduced by Zadeh [30]. However, fuzzy relation equations are also at the core of explanation systems designed for expert systems handling uncertainty in the framework of fuzzy set and possibility theory (Yager [27], Farreny and Prade[13,14]). The limited success of fuzzy relation-based models for diagnosis may be partially due to a lack of precise interpretation of the degrees of association between disorders and manifestations, and of the degrees of membership in the fuzzy set of observations, in many works. From a knowledge representation point of view, the fuzzy sets used in abductive reasoning models usually have a conjunctive interpretation; indeed there is no idea of mutual-exclusiveness between the symptoms which are more or less certainly produced by a given disease for instance. This contrasts with the disjunctive interpretation of fuzzy sets restricting the more or less possible, but mutually exclusive, values of single-valued attributes in approximate reasoning based on possibility distributions.

The next section introduces the diagnosis problem as considered in relational models. The completely informed case is first dealt with and then a model is proposed for the case where we only have incomplete information about the manifestations which are present and about the manifestations which are indeed caused by a given disorder. This situation can be interpreted in terms of two-valued possibility and necessity measures. Namely we distinguish between manifestations whose presence is necessarily true (or if we prefer, certain) and those whose presence is only possible. The proposed model is compared to the parsimonious covering theory developed by Reggia et al.[22]. The fuzzy relation approach is surveyed and discussed in Section 3; the problem of the intended meaning of the membership degrees is especially addressed. Section 4 presents a new model based on twofofd fuzzy relations and twofold fuzzy sets [8], which has a greater expressive power. This new model, introduced in [12] is a graded version of the one proposed in the incompletely informed case but remains purely qualitative. Especially manifestations more or less certainly absent as well as those more or less certainly present are taken into account. Similarly the model manipulates the fuzzy set of manifestations which are more or less certainly produced by a disorder and the fuzzy set of manifestations which cannot be, more or less certainly, produced by this disorder. Section 5 discusses the improvements obtained by introducing levels of uncertainty in the expression of partially known causal relationships. Lines for further research are pointed out in the concluding remarks.

2 - Relational Approach : the Crisp Model

2.1 - Reasoning in Diagnosis Problems : The Completely Informed Case

Let $\mathcal{S}$ be a system whose current state is described by means of a n-tuple of *binary* attributes $(a_1, \ldots, a_i, \ldots, a_n)$. When $a_i = 1$ the manifestation m_i is said to be present ; when $a_i = 0$, it means that m_i is absent. When there is no manifestation present, $\mathcal{S}$ is said to be in its normal state and this state is described by the n-tuple $(0, \ldots, 0, \ldots, 0)$. Let $\mathcal{M}$ denote the set of the n possible manifestations $\{m_1, \ldots, m_i, \ldots, m_n\}$. Let $\mathcal{D}$ be a set of possible disorders $\{d_1, \ldots, d_j, \ldots, d_k\}$. A disorder can be present or absent. To each d_i is associated the set $M(d_j)$ of manifestations which are entailed, or preferably, caused, produced, by the presence of d_j *alone*. The completely informed case is first considered, where all the present manifestations are observed and where the set of manifestations which appear when a disorder is present is perfectly known. Thus if $m_i \notin M(d_j)$ it means that m_i is not caused by d_j. A relation R on $\mathcal{D} \times \mathcal{M}$ is thus defined by $(d_j,m_i) \in R \Leftrightarrow m_i \in M(d_j)$, which associates manifestations and disorders.

Given a set M^+ of present manifestations which are observed, the problem is to find what disorder(s) *may* have produced the manifestations in M^+. It is supposed that the set $M^- = \mathcal{M} - M^+ = \overline{M^+}$ is the set of manifestations which are absent, i.e. all manifestations which are present are observed. While deductive reasoning enables us to predict the presence of manifestation(s) from the presence of disorder(s), abductive reasoning looks for possible cause(s) of observed effects. In other words, one looks for plausible explanations (in terms of disorders) of an observed situation. Clearly while it is at least theoretically possible to find out all the possible causes which may have led to a given state of the system $\mathcal{S}$, the ordering of the possible solutions according to some levels of plausibility is out of the scope of logical reasoning, strictly speaking. However one may for instance prefer the solutions which involve a small number of disorders, and especially the ones, if any, which rely on only one disorder. This is called the principle of parsimony. In case several disorders may be jointly present, the set of manifestations produced by the presence of a pair of disorders $\{d_i,d_j\}$ alone has to be defined, and more generally by a tuple of disorders. In the hypothesis that effects can be added and do not interfere, it holds that:

$$M(\{d_i,d_j\}) = M(d_i) \cup M(d_j) \tag{1}$$

and consequently

$$\overline{M(\{d_i,d_j\})} = \overline{M(d_i)} \cap \overline{M(d_j)}$$

i.e. the manifestations which are absent are those which are not produced by d_i or d_j separately. If this hypothesis is not acceptable, a subset M(D) of entailed manifestations should be prescribed for each subset $D \subseteq \mathcal{D}$ of disorders which can be jointly present. Under this new hypothesis, situations where disorder d_i followed by d_j has not the same effects in terms of manifestations as d_j followed by d_i are excluded, since the set $\{d_i, d_j\}$is not ordered. In other words, a relation on $2^{\mathcal{D}} \times \mathcal{M}$ is used, rather than on $\mathcal{D} \times \mathcal{M}$. If some disorders can never be jointly present, $2^{\mathcal{D}}$ should be replaced by the appropriate set $\mathcal{A}$ of associations of disorders which indeed make sense.

In the completely informed case described above, the following properties hold:

i) $M^+ = \overline{M^-}$, i.e. all the present manifestations are observed, and equivalently all the manifestations which are not observed are indeed absent;

ii) $\forall d$, $M(d) = M(d)^+ = \overline{M(d)^-}$, where $M(d)^+$ (resp. $M(d)^-$) is the set of manifestations which are certainly present (resp. certainly absent) when disorder d alone is present.When $M(D) \neq \bigcup_{d\in D} M(d)$, this condition is supposed to hold $\forall\ D \in 2^{\mathcal{D}}$ in the completely informed case (and not only for $D = \{d\}$).

The potential set $\widehat{D}$ of all disorders which can individually be responsible for M^+ is given by

$$\widehat{D} = \{d \in \mathcal{D}, M(d) = M^+\}. \tag{2}$$

Note that $M(d) = M^+ \Leftrightarrow \overline{M(d)} = M^-$. Namely no disorder outside $\widehat{D}$ can cause (M^+, M^-). Clearly, if $\widehat{D} = \emptyset$, one must check for the set $\widehat{DD}$ of groups of disorders D, which together may have caused M^+,

$$\widehat{DD} = \{D \in \mathcal{A} \subseteq 2^{\mathcal{D}}, M(D) = M^+\}. \tag{3}$$

Using the principle of parsimony, one might consider that the smaller the cardinality of D the more plausible it is. If M(D) can be obtained as $\bigcup_{d\in D} M(d)$, then the set D_0 of disorders which alone *partially* explains M^+

$$D_0 = \{d \in \mathcal{D}, M(d) \subseteq M^+\} \tag{4}$$

may be of interest for building elements of $\widehat{DD}$. Clearly $D_0 \supseteq \hat{D}$.

2.2 - The Case of Incomplete Information

Now consider the case where information about the relation disorder-manifestation is still complete ($\forall$ d $\in \mathcal{D}$, $M(d) = M(d)^+ = \overline{M(d)^-}$), but where the present manifestations are not necessarily completely observed. When not all the information is available, the set M^+ of manifestations which are certainly present and the set M^- of manifestations which are certainly absent no longer form a partition of $\mathcal{M}$; indeed $M^+ \cap M^- = \emptyset$ but $M^+ \cup M^- \neq \mathcal{M}$, or equivalently, there exists a non-empty set $M^0 = \mathcal{M} - (M^+ \cup M^-)$ of manifestations, the presence or the absence of which are completely unknown. In other words, one may be unaware of the presence of some manifestations, and perhaps, only a subset of the absent manifestations are known to be absent. Then (2) is changed into

$$\hat{D} = \{d \in \mathcal{D}, M^+ \subseteq M(d) \subseteq \overline{M^-}\} \tag{5}$$

since it is only known that the set of manifestations which are indeed present is lower bounded by the set M^+ and upper bounded by the set $\overline{M^-}$. (5) also writes

$$\hat{D} = \{d \in \mathcal{D}, M(d) \cap M^- = \emptyset \text{ and } \overline{M(d)^-} \cap M^+ = \emptyset\}. \tag{6}$$

The generalization of (5)-(6) to subsets D of joint disorders is straightforward.

Another particular case of incomplete information is when observations are complete ($M+ = \overline{M^-}$), but some manifestations may sometimes be present or absent with a given disorder; i.e. for some d, it is sometimes not known if a manifestation m follows or not from d; in that case $m \notin M(d)^+$ and $m \notin M(d)^-$. In other words, the union of the set $M(d)^+$ of manifestations which are certainly produced by d alone and the set $M(d)^-$ of manifestations which certainly cannot be caused by d alone, no longer covers $\mathcal{M}$, i.e. $\exists d$, $M(d)^+ \cup M(d)^- \neq \mathcal{M}$, but, $M(d)^+ \cap M(d)^- = \emptyset$ always holds. Denoting $M(d)^0 = \mathcal{M} - (M^+(d) \cup M^-(d))$, $m \in M(d)^0$ means that m is only a *possible* manifestation of d. In particular m may be absent or present when d is present. An unknown disorder can always be introduced, i.e. d_0 such that $M(d_0)^0 = \mathcal{M}$, i.e. whose manifestations are unknown. Hence it is not a closed-world model. Then, the set $\hat{D}$ of potential disorders which alone can explain $M+ = \overline{M^-}$ is given by

$$\hat{D} = \{d \in \mathcal{D}, M(d)^+ \subseteq M^+ \text{ and } M(d)^- \subseteq M^-\}. \tag{7}$$

Obviously, when $M(d)^+ = \overline{M(d)^-} = M(d)$, (2) is recovered. Besides, (7) can be easily generalized over to a subset of joint disorders.

In the general case, both the information pertaining to the manifestations and the information relative to the association between disorders and manifestations is incomplete. Then d belongs to the set $\widehat{D}$ of potential disorders each of which can alone explain both M^+ and M^- if and only if d does not produce with certainty any manifestation which is certainly absent in the evidence, and no observed manifestations is ruled out by d. Formally:

$$\widehat{D} = \{d \in \mathcal{D},\ M(d)^+ \subseteq \overline{M^-} \text{ and } M(d)^- \subseteq \overline{M^+}\}. \tag{8}$$

This also writes

$$\widehat{D} = \{d \in \mathcal{D},\ M(d)^+ \cap M^- = \emptyset \text{ and } M(d)^- \cap M^+ = \emptyset\}. \tag{9}$$

Clearly, (8) reduces to (2) in the completely informed case, since then $\overline{M^-} = M^+$ and $M(d)^- = \overline{M(d)^+}$. More generally, it is worth-noticing that $\widehat{D}$ gathers all the manifestations in $\mathcal{D}$ which cannot be ruled out by the observations. If $M^- = \emptyset$ and $M(d)^- = \emptyset$, i.e. no information is available on the manifestations certainly absent, it can be verified that $\widehat{D} = \mathcal{D}$ and the whole set of possible disorders is obtained.

The unknown disorder d_0 whose manifestations are unknown is such that $M(d_0)^+ = \emptyset = M(d)^-$. Hence $d_0 \in \widehat{D}$, i.e. d_0 can be always considered as potentially responsible for observed manifestations. The membership test (8) is thus very permissive, i.e. $\widehat{D}$ can be very large, and contains what may look like irrelevant causes, since they cannot be ruled out. Namely a disorder d may belong to $\widehat{D}$ defined by (8)-(9) even if $M(d)^+ \cap M^+ = \emptyset$ and $M(d)^- \cap M^- = \emptyset$. Indeed, in this case, nothing forbids $M(d)^+ \cup M(d)^- \subseteq M^0$. This means that $\widehat{D}$ includes disorders, no sure manifestations of which are observed, and no forbidden manifestations are for sure absent. Such disorders may still be present when M^+ and M^- are observed, since all the sure information we have about them pertains to manifestations in M^0 about which no observation is available (this is true for the unknown disorder d_0).

In other words, (9) achieves a consistency-based diagnosis only, that is, only the disorders which are incompatible with manifestations observed as present or absent are rejected. In order to be more selective, one may turn to abductive reasoning. First, among the disorders in $\widehat{D}$, consider those, if any, which have at least a weak relevance to the observations, namely the subset $\widehat{D}^*$ of $\widehat{D}$, defined by

$$\widehat{D}^* = \{d \in \widehat{D},\ M(d)^+ \cap M^+ \neq \emptyset \text{ or } M(d)^- \cap M^- \neq \emptyset\}. \tag{10}$$

This eliminates the disorders such that $M(d)^+ \cup M(d)^- \subseteq M^0$, i.e., the disorders which are not suggested by the observations while not being ruled out by them either (and d_0 particularly). It is worth-noticing that the refinement of $\widehat{D}$ by $\widehat{D}^*$ makes sense only in the general case of incomplete information. Indeed, $\widehat{D} = \widehat{D}^*$, when $\widehat{D}$ is defined by (5) (if $M^+ \neq \emptyset$ or $M^- \neq \emptyset$), or (7) (if $M(d)^+ \neq \emptyset$ or $M(d)^- \neq \emptyset$). Obviously, one may still refine $\widehat{D}^*$ by strengthening the requirements on d, by asking for a conjunction rather than a disjunction in (10) or by replacing the conditions by more demanding ones like $M(d)^+ \subseteq M^+$ (all sure manifestations of d are observed) or $M(d)^+ \supseteq M^+$ (all observed manifestations are among the ones which certainly accompany d), and similar conditions for the subsets pertaining to the absence of manifestations. In particular, the subset $\widehat{D}^{**}$ of $\widehat{D}$ defined by

$$\widehat{D}^{**} = \{d \in \widehat{D}, M(d)^+ \supseteq M^+ \text{ and } M(d)^- \supseteq M^-\} \tag{11}$$

gathers the disorders, if any, which offer a complete coverage of the observations, but which may also have some sure effects or sure absence of effects which have remained unobserved (i.e., which are in M^0). A genuine abductive task is performed by (11).

When $\widehat{D} = \emptyset$, one can look for explanations in terms of subsets of disorders which are not singletons. (3) is then extended by

$$\widehat{DD} = \{D \in \mathscr{A} \subseteq 2^{\mathscr{D}}, M(D)^+ \subseteq \overline{M^-} \text{ and } M(D)^- \subseteq \overline{M^+}\} \tag{12}$$

for the subsets of disorders which jointly may explain M^+ *and* M^-. $\widehat{DD}$ can obviously be refined by extending the counterparts of (10) or (11) for defining "relevant" subsets of disorders; for instance (10) is generalized by

$$\widehat{DD}^* = \{D \in \widehat{DD}, M(D)^+ \cap M^+ \neq \emptyset \text{ or } M(D)^- \cap M^- \neq \emptyset\}. \tag{13}$$

As expected, what is present and what is absent play symmetrical roles, exchanging + and – in (8)-(12).

Note that if $M^- = \emptyset$, i.e. if only manifestations which are certainly present are known, (9) (or (12)) may yield a result $\widehat{D} \neq \mathscr{D}$ (or $\widehat{DD} \neq \mathscr{A}$) provided that $\overline{M(d)^-} \neq \mathscr{M}$, i.e. we have non-trivial information on the set of manifestations $\overline{M(d)^-}$ (or $\overline{M(D)^-}$) which *may* be produced by a disorder d (or a subset of disorders D) alone; indeed $\overline{M(d)^-} \supseteq M(d)^+$ (resp. $\overline{M(D)^-} \supseteq M(D)^+$), and thus $\overline{M(d)^-}$ (resp. $\overline{M(D)^-}$) gathers the manifestations in $M(d)^+$ (resp. $M(D)^+$) which are certainly produced by d (resp. D) and the manifestations for which we do not know if they can or cannot follow from d (resp. D). For $M^- = \emptyset$, (9) and (12) write

$$\widehat{D} = \{d \in \mathscr{D}, \overline{M(d)^-} \supseteq M^+\}. \tag{14}$$

$$\widehat{DD} = \{D \in \mathscr{A} \subseteq 2^{\mathscr{D}}, \overline{M(D)^-} \supseteq M^+\} \tag{15}$$

In the non-completely informed case, the hypothesis (1) that effects can be added and do not interfere writes (for two disorders)

$$\begin{cases} M(\{d_i,d_j\})^+ = M(d_i)^+ \cup M(d_j)^+ \\ \text{and} \\ M(\{d_i,d_j\})^- = M(d_i)^- \cap M(d_j)^-. \end{cases} \tag{16}$$

Clearly, (16) reduces to (1) in the completely informed case. Note that the second equality of (16) still writes

$$\overline{M(\{d_i,d_j\})^-} = \overline{M(d_i)^-} \cup \overline{M(d_j)^-}$$

which says that the possible manifestations of two simultaneous disorders gather the manifestations possibly produced by each disorder, as for certain manifestations.

2.3 - Link with the Reggia-Peng Approach

Reggia et al.[22] (see also Peng and Reggia [20,21]) have extensively studied a relation-based formulation of diagnosis problems. In their model they assume the knowledge of a relation R between disorders and manifestations, such that the fact that the pair (d_j,m_i) satisfies this relation "means d_j may directly cause m_i. Note that this does *not* mean that d_j *necessarily* causes m_i, but only that it *might*", as stated in (Peng and Reggia[21]). But these authors do not explain why they make this choice for interpreting the "causal" relation between $\mathcal{M}$ and $\mathcal{D}$. Moreover the set M^+ of "manifestation known to be present" is supposed to be available.

Thus, in their model, what is known is the set $R(d) = \overline{M(d)^-}$ of manifestations possibly attached to a disorder, for each disorder d, while $M(d)^+ = \emptyset$; it is also assumed that (16) holds for computing $\overline{M(D)^-}$ for $D \subseteq \mathcal{D}$. Since in this model $M^- = \emptyset$, (14)-(15) applies, and indeed a singleton {d} such that $\overline{M(d)^-} \supseteq M^+$ and more generally, a set of disorders D such that $\overline{M(D)^-} \supseteq M^+$(applying (16), i.e. $\overline{M(D)^-} = \bigcup_{d \in D} \overline{M(d)^-}$) is called a "*cover*" of M^+ by Peng and Reggia[21].

These authors more particularly look for so-called "*parsimonious*" covers, especially relevant, irredundant and minimum covers. D is a *relevant* cover if it is a cover and if $\forall\, d \in D$, $\exists\, m \in M^+$, $(d,m) \in R$; D is *irredundant* if none of its proper subsets is also a cover of M^+; D is *minimum* if its cardinality is smallest among all covers of M^+. Clearly a one-disorder cover is a minimal cover, a minimal cover is an irredundant cover, an irredundant cover is a relevant cover, a relevant cover is obviously a cover. The set of relevant covers is defined by

$$DD^* = \{D \in \widehat{DD},\ \forall\, d \in D,\ \overline{M(d)^-} \cap M^+ \neq \emptyset\} \tag{17}$$

since in Peng and Reggia[21] the relation R is defined by $(d,m) \in R \Leftrightarrow m \in \overline{M(d)^-}$, using our notations. Note that Peng and Reggia's notion of relevance does not coincide with the one defined by (10) and (13) (or by (11)) for two reasons. First, with definition (17) a one-disorder cover is necessarily relevant, or in other words this notion of relevance is non-trivial only with covers which are not singletons, which is not the case with (10)-(13), and second, $\widehat{DD}^*$, defined by (13), is empty in Peng and Reggia's particular case since $M(D)^+ = \emptyset$ and $M^- = \emptyset$. Peng and Reggia's definitions could easily be extended to a framework where possible and sure manifestations are told apart. For instance the set of relevant covers, in their sense, could be defined by the subset of disorders $\{D \in \widehat{DD}, \forall\ d \in D, \overline{M(d)^-} \cap M^+ \neq \emptyset$ or $\overline{M(d)^+} \cap M^- \neq \emptyset\}$, which extends (17). Consequently, our relational diagnosis model is more general than Peng and Reggia', in the incompletely informed case.

3 - The Fuzzy Approach

3.1 - The Meaning of the Grades of Membership

In the late seventies Sanchez [23,24], Tsukamoto and Terano[26] already developed diagnosis methods based on a fuzzy relational model. Several slightly different proposals have been made. We consider here the simplest version which corresponds to the type of diagnosis problem presented in the preceding section, where, i) R is a fuzzy relation defined on $\mathcal{D} \times \mathcal{M}$, and ii) M^+ is a fuzzy set. Indeed, it seems desirable in practice to be able to model a more gradual association between manifestations and disorders, and to take into account the uncertainty or vagueness pervading the observation of manifestations. Although this suggests the possible interest of some fuzzy set-based methods, it is not right away clear what should be the precise interpretation of the degrees of association $\mu_R(d_j,m_i)$, of the degrees of membership $\mu_{M^+}(m_i)$ when the relation R or the subset M^+ is fuzzy, or of the degrees attached to disorders which are then proposed for the explanation of M^+, knowing R. For instance Kitowski and Bargiel [16] spoke of the partial occurrence of a disorder, or of the uncertain observation of a manifestation, but other interpretations can be thought of.

There are basically two types of interpretations which can be contemplated, one pertaining to the level of fulfilment of a gradual manifestation, the other to the uncertainty of statements pertaining to observations. Namely, we may be uncertain about the presence of a manifestation because its observation is, difficult, costly, or even untimely, or manifestations may be a matter of degree (e.g. in medical diagnosis, the fever of a patient may be more or less high). In the first interpretation the uncertainty pervading our knowledge of actual manifestations stems from the use of easy, or quick observation methods whose results are imprecise. This may occur even if the presence or absence of a manifestation is an all or nothing matter : the presence of a binary manifestation can be more or less certain ($\mu_{M^+}(m_i)$) or more or less possible ($1 - \mu_{M^-}(m_i) \geq \mu_{M^+}(m_i)$) due to unreliable or imprecise measurement. This case is studied in Section 4. On the contrary, $\mu_{M^+}(m_i)$ can model the *intensity* of a manifestation m_i. This second interpretation is only possible if the underlying attribute a_i is no longer binary. For instance, instead of considering that m_i means "fever" or "not fever", it looks more natural to consider "high fever" as a manifestation. The vagueness pertaining to the word "high" reflects the fact that fever is a matter of degree and can be measured on a numerical scale S (ranging from 35 to 42 degrees, say). "High fever" refers to a fuzzy set F of S (that means "close to 40", say) in a given context. If $x \in S$ is the temperature of the patient, then $\mu_{M^+}(\text{high fever}) = \mu_F(x) = 1 - \mu_{M^-}(\text{high fever})$, since M^- refers to the absent manifestations and $\mu_{M^-}(\text{high fever}) = 1$ means "not high fever".

Similarly the degree of association $\mu_R(d_j,m_i)$ may account for the uncertainty that m_i follows from the presence of d_j, or for the intensity of manifestation m_i when d_j is present (which again supposes that the severeness of the effect can be graded) ; if the level of disorder d_j may itself be more or less severe $\mu_R(d_j,m_i)$ may even correspond to a gradual rule, like the more severe the disorder, the stronger the manifestation. The intensity of the disorder then simply reflects the level of matching between (numerical) measurements and the most acute forms of the manifestations caused by the disorder (like "high fever"). On the contrary if we interpret $\mu_R(d_j,m_i)$ as a degree of uncertainty it may either mean to what extent m_i *necessarily* follows from the presence of d_j or only to what extent it is *possible* that m_i is present when d_j is present. In particular if the fuzzy set $M(d_j)$ ($\mu_{M(d_j)}(m_i) = \mu_R(d_j,m_i)$) gathers the manifestations which more or less *certainly* or necessarily follow from d_j, the fact that m_i is such that $\mu_R(d_j,m_i) = 0$ means that either m_i does not follow from d_j, or that we do not know (i.e. we have absolutely no certainty that m_i is caused by d_j) ; by contrast if $M(d_j)$ is the fuzzy set of manifestations which more or less *possibly* follows from d_j, $\mu_R(d_j,m_i) = 0$ means that m_i cannot be caused by d_j, but $\mu_R(d_j,m_i) = 1$ expresses no certainty that m_i *should* accompany d_j, it only expresses that it is fully possible. If $\mu_R(d_j,m_i)$ does reflect the uncertainty of presence of manifestation m_i due to disorder d_j, but the end-points of the unit interval are interpreted as impossible (0) and certain (1), then it is difficult to imagine that $\mu_R(d_j,m_i)$ represents something else than the probability that d_j causes m_i when d_j is present, as used by Peng and Reggia [21], this probability being subjective or objective.

According to the interpretation we have in mind, it will lead to different models with different interpretations of the results. To some extent the literature on fuzzy relational equations for diagnosis suffers from a lack of concern for these representational issues.

3.2. - Fuzzy Sets of Abduced Disorders Considered in the Literature

In the following we survey some proposals made in the fuzzy set literature for handling the diagnosis problem on the basis of fuzzy relation (in)equations. We shall try to recognize fuzzy counterparts of various sets of disorders we have introduced in Section 2 for the crisp case, keeping in mind the problem of the interpretation of the membership grades.

Let M^+ be the fuzzy set of observations and R be a fuzzy relation on $\mathcal{D} \times \mathcal{M}$ modelling the association between disorders and manifestations. As already said R induces for each $d \in \mathcal{D}$, the fuzzy set M(d) of manifestations fuzzily associated with d, i.e., $\forall\, d \in \mathcal{D}$, $\forall\, m \in \mathcal{M}$, $\mu_{M(d)}(m) = \mu_R(d,m)$ whatever it means. Given M^+ and R, the diagnosis problem is then stated as the determination of some subset $D \subseteq \mathcal{D}$ of disorders such that the composition of D with R yields M^+, which can be written symbolically $D \circ R = M^+$, and which in the simplest models translates into the fuzzy relation equation

$$\forall\, i = 1,n,\ \max_{j=1,k} \min(\mu_D(d_j), \mu_R(d_j,m_i)) = \mu_{M^+}(m_i). \tag{18}$$

This equation can be straightforwardly generalized to the multiple disorder case where R is defined on $2^{\mathcal{D}} \times \mathcal{M}$ [24]. The equation (18) may be also written with a triangular norm different from min, such as the product or the operation $\max(0, \cdot + \cdot - 1)$ (see Pedrycz, [19] for instance). It corresponds to different manners of aggregating the grade of presence of a disorder with its degree of association with a manifestation. In the following we only consider the use of the triangular norm min in (18). The fuzzy set D of solutions d_j of equation (18) is the fuzzy counterpart of $\widehat{D}$ in equation (2), where one assumes complete information on the manifestations which are present and on those which are for sure associated with a given disease. Thus the understanding of (18) should be in terms of intensity of presence for a manifestation rather than uncertainty of presence. Namely, the more intense a disorder d_j assumed to be present, and the greater its degree of association with a manifestation, the greater the intensity of this manifestation, and conversely the greater this intensity, the greater the intensity of the disorders associated with this manifestation. Such an interpretation presupposes that the attributes which are used to describe the manifestations are no longer binary, but gradual, like "fever" in "high fever".

The equation (18) where D is unknown may have no solution. If several solutions exist, the solution(s) which is(are) minimal (in the sense of set inclusion, defined by $F \subseteq G \Leftrightarrow \mu_F \leq \mu_G$ for fuzzy set(s)), is(are) worth-considering ; see Pappis and Sugeno [17]. They correspond to the idea of irredundant covers in Reggia et al.'s model.

If the equation $D \circ R = M^+$ has no solution, we can relax it into $D \circ R \subseteq M^+$ or $D \circ R \supseteq M^+$. In the first case we look for an explanation of a subset of manifestations, while in the second case we look for a superset thereof, i.e. the set of potential causes D should explain all the observations but is allowed to have some unobserved effects also. This leads to consider various (fuzzy) sets of interest in $\mathcal{D}$. Namely, there are three kinds of set of causes which are particularly worth-considering, as pointed out in [11] :

i) the set of disorders whose manifestations lie in M^+ is defined in the non-fuzzy case by $D_0 = \{d_j \in \mathcal{D} \mid M(d_j) \subseteq M^+\}$ where $M(d_j) = \{d_j, (d_j,m_i) \in R\}$. D_0 is still obtained in the fuzzy case as the largest solution of the equation $D \circ R \subseteq M^+$. Note that if we have the fuzzy set inclusions $M(d_i) \subseteq M^+$ and $M(d_j) \subseteq M^+$ we also have $M(\{d_i,d_j\}) = M(d_i) \cup M(d_j) \subseteq M^+$, where the fuzzy union is defined by the max operation on the membership degrees (this is the extension of (1) to the fuzzy setting). D_0 is also the largest solution of $D \circ R = M^+$ when a solution exists. The membership function of D_0 is then given by

$$\mu_{D_0}(d_j) = \min_i \mu_R(d_j,m_i) \rightarrow \mu_{M^+}(m_i) \tag{19}$$

where $\rightarrow$ denotes the so-called Gödel implication, i.e. $\alpha \rightarrow \beta = 1$ if $\alpha \leq \beta$ and $\alpha \rightarrow \beta = \beta$ if $\alpha > \beta$. This is a fuzzy counterpart of (4). Since D_0 may gather several distinct explanations of a subset of manifestations, it is then interesting to look for minimal solution(s) of the equation $D \circ R = D_0 \circ R$ where D is unknown ; see Di Nola et al. [7] for algorithms for computing these solutions (there may be several ones which cannot be compared with respect to set inclusion). Each such a minimal solution is made of a set of disorders which altogether are sufficient to explain the greatest subset of observations in M^+ which is explainable according to R, namely $D_0 \circ R$, and which has no effect outside M^+ ;

ii) the set of disorders such that each of them *alone* explains at least all the manifestations in M^+ (and may produce some others), i.e. with crisp sets, the set $\widehat{D} = \{d_j \in \mathcal{D} \mid R(d_j) \supseteq M^+\}$. $\widehat{D}$ is defined in the fuzzy case by

$$\mu_{\widehat{D}}(d_j) = \min_i \mu_{M^+}(m_i) \rightarrow \mu_R(d_j,m_i). \tag{20}$$

(20) appears to be the fuzzy counterpart of (7), which suggests that $\mu_R(d_j,m_i)$ should then be interpreted as the extent to which it is possible that d_j causes m_i. Obviously $\widehat{D}$, like D_0, may be empty. If $\widehat{D}$ is empty, we may look for 2-element sets $\{d_i,d_j\}$ such that $\{d_i,d_j\} \circ R \supseteq M^+$ (with $\{d_i,d_j\} \circ R = R(d_i) \cup R(d_j)$), and more generally for any subsets of disorders with a given number of elements which *altogether* explain M^+ (and maybe more). In the case of 2-element sets, it is extended to the fuzzy case by the following expression

$$\mu_{\widehat{DD}}(\{d_i,d_j\}) = \min_k [\mu_{M^+}(m_k) \rightarrow \max(\mu_R(d_i,m_k), \mu_R(d_j,m_k))] \tag{21}$$

where $\mu_{\widehat{DD}}(\{d_i,d_j\})$ estimates to what extent the pair (d_i,d_j) provides an explanation of all the manifestations in M^+. Yager [27, 28] has proposed a slightly different index based on Lukasiewicz implication rather than Gödel's which, in a fuzzy set framework, provides an estimate of the extent to which a (fuzzy) set of causes D explains all the observations in M^+. Its expression is

$$\text{Explain}(M^+ / D) = \min_i [\mu_{M^+}(m_i) \rightarrow \mu_{D \circ R}(m_i)] \tag{22}$$

where the fuzzy set $D \circ R$ is defined by the max-min composition. It can be easily checked that $\text{Explain}(M^+ / \{d_i,d_j\})$ is nothing but $\mu_{\widehat{DD}}(\{d_i,d_j\})$ defined above. The quantity $\text{Explain}(M^+ / D)$ estimates to what extent D may be considered as a solution of the equation $D \circ R \supseteq M^+$, i.e. equation (7), where $D \circ R = \overline{M(D)^-}$;

iii) the set of disorders whose manifestations are partly in M^+ which is defined in the crisp case by $D^* = \{d_j \in \mathcal{D} \mid R(m_i) \cap M^+ \neq \emptyset\}$, and in the fuzzy case by

$$\mu_{D^*}(d_j) = \max_i \min(\mu_R(d_j,m_i), \mu_{M^+}(m_i)). \tag{23}$$

It corresponds to the set of relevant disorders in the sense of (9). Technically speaking, D^* is the upper image of the fuzzy set M^+ induced by R, while $\widehat{D}$ and D_0 defined above are respectively the lower and the lower inverse image of M^+ induced by R [11]. Clearly, D^* is usually much larger than D_0 or $\widehat{D}$, and gathers all the causes of potential interest with respect to the set of observations M^+ ; however the set of manifestations of the disorders in D^* does not necessarily cover M^+, nor is included in M^+.

Other worth-considering (fuzzy) subsets of disorders can be defined in the framework of possibility theory, in particular the subset of disorders which alone explain the most *important* manifestations in M^+, or *most* of the manifestations in M^+ [11]. The fuzzy subset of causes which alone explain the most important manifestations in M^+ is obtained by weighting μ_{M^+} in a way which is in agreement with possibility theory. Let $w_i \in [0,1]$ be the importance rate of observation m_i ; if $w_i = 1$, the manifestation m_i is definitely important (and thus has to be explained) ; if $w_i = 0$ the manifestation has no importance (and we have not to look for its cause). Then the fuzzy subset of disorders $\widehat{D}^w$, which alone explains the most important manifestations in M^+, is defined by substituting $\mu_{M^{+w}} = \min(w_i, \mu_{M^+}(m_i))$ to $\mu_{M^+}(m_i)$ in the above expressions of $\mu_{\widehat{D}}$. Indeed M^{+w} represents the fuzzy set of important observations. Note that in case M^+ is an ordinary subset, i.e. $\mu_{M^+}(m_i) = 0$ or 1, weighting the importance of the observations is nothing but substituting w_i to $\mu_{M^+}(m_i)$ when m_i is observed. This suggests yet another possible interpretation of the membership grades, namely the importance given to the observation rather than the intensity, or the certainty degree of the manifestation. Clearly $\widehat{D}^W$ is larger than $\widehat{D}$, in the sense of fuzzy set inclusion.

In this framework, modelling a quantifier like "most" comes down to assigning a high degree of importance to k manifestations, arbitrarily chosen in M^+, with k "close to" the number of observed effects in M^+ (this can be defined as a fuzzy set defined on the set of integers). In order to estimate to what extent the disorder d_j explains "most" of the effects in M^+ we compute the maximum of $\mu_{\widehat{D}^w}(d_j)$ on all the permutations σ of $[\![1,n]\!]$, i.e. we compute

$$\max_\sigma \min_i [\min(w_{\sigma(i)}, \mu_{M^+}(m_i)) \to \mu_R(d_j,m_i)] \tag{24}$$

where $w_{\sigma(i)} = 1$ if $\sigma(i) \in [\![1,k]\!]$ and $w_\sigma(i) = 0$ otherwise ; we get the set of causes which alone explain at least k manifestations in M^+. This readily extends to subsets of causes of a given cardinality. In the general case, the weights $w_{\sigma(i)}$ may lie in [0,1] and capture the idea of taking k as a fuzzy number. The analogy between this quantification problem and the one addressed by Yager [29]'s OWA is worth noticing.

4 - Possibilistic diagnosis under graded uncertainty

This section presents a graded counterpart of the incomplete information model of Section 2.2. Namely M^+ and M^- are now fuzzy sets of manifestations which are more or less certainly present, and more or less certainly absent respectively. However the requirement $M^+ \cap M^- = \emptyset$ is kept (where the intersection is defined by the min operation), i.e. we cannot be somewhat certain both of the presence and of the absence of the same manifestation simultaneously. Similarly, $M(d)^+$ (more generally $M(D)^+$) and $M(d)^-$ (more generally $M(D)^-$) will denote the fuzzy sets of manifestations which are respectively more or less certainly present and more or less certainly absent when disorder d alone is present (more generally when the subset D of disorders is present). Obviously, $\forall d$, $M(d)^+ \cap M(d)^- = \emptyset$ (and $\forall D$, $M(D)^+ \cap M(D)^- = \emptyset$) are also assumed.

By complementation (defined by $\mu_{\overline{F}} = 1 - \mu_F$), we obtain the fuzzy sets $\overline{M^-}$, $\overline{M(d)^-}$ (and more generally $\overline{M(D)^-}$) of manifestations which are more or less possibly present, respectively, in the considered situation, when d is present, or when disorders in D are altogether present. This corresponds to the usual duality between what is (more or less) certain, i.e. necessarily true, and what is (more or less) possibly true. Indeed a pair of dual possibility and necessity measures ∏ and N are related by the relation $\Pi(A) = 1 - N(\overline{A})$, for any event A (here A represents the presence of a manifestation) ; see Dubois and Prade [9], for instance.

Note that $M^+ \subseteq \overline{M^-}$, $M(d)^+ \subseteq \overline{M(d)^-}$, and $M(D)^+ \subseteq \overline{M(D)^-}$ in the sense of fuzzy set inclusion. An even stronger inclusion holds. Since $M^+ \cap M^- = \emptyset$, we have

$$\{m_i \in \mathcal{M}, \mu_{M^+}(m_i) > 0\} \subseteq \{m_i \in \mathcal{M}, \mu_{M^-}(m_i) = 0\} = \{m_i \in \mathcal{M}, \mu_{\overline{M^-}}(m_i) = 1\} \quad (25)$$

i.e. the support of M^+ is included in the core of $\overline{M^-}$; the same holds for $M(d)^+$, $\overline{M(d)^-}$, or $M(D)^+$, $\overline{M(D)^-}$. This is in agreement with the fact that for crisp events A, we have $N(A) > 0 \Rightarrow \Pi(\bar{A}) = 0 \Rightarrow \Pi(A) = 1$ since then one of A or $\bar{A}$, at least, should be completely possible [9].

A pair of fuzzy sets (F,G) such that $F \cap \bar{G} = \emptyset$ is called a twofold fuzzy set [8]. Twofold fuzzy sets (F,G) have been introduced for modelling incompletely known sets, i.e. sets for which we know elements gathered in F, which more or less certainly belong to it, as well as other elements, gathered in $\bar{G}$, which more or less certainly do not belong to it. But $F \cup \bar{G}$ may not cover the whole referential. Similarly, the pairs $(M(d)^+, \overline{M(d)^-})$ and $(M(D)^+, \overline{M(D)^-})$ define twofold fuzzy relations on $\mathcal{D} \times \mathcal{M}$, and $2^{\mathcal{D}} \times \mathcal{M}$ respectively.

The extension to fuzzy sets of the equation (5) (and more generally (6)) is not straighforward. This equation gives the set $\widehat{D} = \{d \in \mathcal{D},\ M(d)^+ \subseteq \overline{M^-} \text{ and } M(d)^- \subseteq \overline{M^+}\}$ of potential disorders which alone may explain both M^+ and M^-. Indeed there exist different indices for estimating the inclusion between two fuzzy sets. These indices are of the form

$$\mathrm{Inc}(A,B) = \inf_x \mu_A(x) \rightarrow \mu_B(x) \quad (26)$$

where the implication operation has to be chosen [3]. The choice may be directed by desired properties such as the requirement to have Inc(A,A) = 1 or the satisfying of the contraposition properties $\mathrm{Inc}(A,B) = \mathrm{Inc}(\bar{B},\bar{A})$. If we restrict ourselves to the three following implication operations which can be directly defined from the conjunction minimum, so as to remain in an ordinal setting :

$$a \Rightarrow b = 1 - \min(a, 1 - b) = \max(1 - a, b) \qquad \text{(Dienes-Rescher)}$$
$$a \rightarrow b = \sup\{t \in [0,1], \min(a,t) \leq b\} = \begin{cases} 1 \text{ if } a \leq b \\ b \text{ if } a > b \end{cases} \qquad \text{(Gödel-Brouwer)}$$
$$a \twoheadrightarrow b = (1 - b) \rightarrow (1 - a) = \begin{cases} 1 \text{ if } a \leq b \\ 1 - a \text{ if } a > b \end{cases}$$

we observe that i) only $\Rightarrow$ satisfies the contraposition property, ii) $a \Rightarrow b = 1 \Leftrightarrow (a > 0 \Rightarrow b = 1)$ while $a \rightarrow b = 1 \Leftrightarrow a \leq b \Leftrightarrow a \twoheadrightarrow b = 1$.

These properties lead to choose the implication operation '$\Rightarrow$ ' for estimating the degrees of inclusion of $M(d)^+$ into $\overline{M^-}$, and of $M(d)^-$ into $\overline{M^+}$, since i) the contraposition property is natural here (we should have $\mathrm{Inc}(M(d)^+,\overline{M^-}) = \mathrm{Inc}(M^-,\overline{M(d)^+})$ and $\mathrm{Inc}(M(d)^-, \overline{M^+}) = \mathrm{Inc}(M^+,\overline{M(d)^-})$, and ii) the disorder d is a perfect candidate for explaining M^+ and M^-, i.e. d is such that $\mu_{\hat{D}}(d) = 1$, only if the support of $M(d)^+$ is included in the core of $\overline{M^-}$ and the support of $M(d)^-$ is included in the core of $\overline{M^+}$, i.e., only if all the manifestations more or less certainly caused by d are among manifestations which are not somewhat certainly absent, and if all the manifestations more or less certainly absent when d is present are among the manifestations which are not somewhat certainly present. In the previous sentence, "not somewhat certainly absent" (resp. : present) should be understood as "with a complete possibility of presence" (resp. : absence), referring to manifestations whose presence (resp. : absence) has been actually observed as well as those which have not been measured yet. $\hat{D}$ is thus defined by

$$\begin{aligned}
\forall\, d \in \mathscr{D},\ \mu_{\hat{D}}(d) &= \min(\mathrm{Inc}(M(d)^+,\overline{M^-}),\ \mathrm{Inc}(M(d)^-,\overline{M^+})) \\
&= \min(\min_{i=1,n} \max(1 - \mu_{M(d)^+}(m_i),\ 1 - \mu_{M^-}(m_i)), \\
&\qquad\qquad \min_{i=1,n} \max(1 - \mu_{M(d)^-}(m_i),\ 1 - \mu_{M^+}(m_i))) \\
&= 1 - \max(\max_{i=1,n} \min(\mu_{M(d)^+}(m_i),\ \mu_{M^-}(m_i)), \\
&\qquad\qquad \max_{i=1,n} \min(\mu_{M(d)^-}(m_i),\ \mu_{M^+}(m_i))) \\
&= 1 - \max(\mathrm{cons}(M(d)^+,M^-),\ \mathrm{cons}(M(d)^-,M^+)) \qquad (27)
\end{aligned}$$

where cons(F,G) is the max-min consistency of two fuzzy sets, evaluating to what extent their intersection is not empty (Zadeh [30]). (27) clearly expresses that a disorder d is all the less a candidate explanation as the fuzzy set of its more or less certain effects overlaps the fuzzy set of manifestations more or less certainly absent, or as the fuzzy set of effects which are more or less certainly absent when d is present overlaps the fuzzy set of manifestations which are more or less certainly present. This is an extension of consistency-based diagnosis to graded uncertainty.

In the particular case where $M^- = \varnothing$, i.e. we only have positive observations, $\hat{D}$ is defined by

$$\forall\, d \in \mathscr{D},\ \mu_{\hat{D}}(d) = \min_{i=1,n} \max(1 - \mu_{M^+}(m_i),\ \mu_{\overline{M(d)^-}}(m_i)). \qquad (28)$$

This expression has to be compared with (20). Remembering that $\mu_R(d,m_i) = \mu_{\overline{M(d)^-}}(m_i)$ in (20), the two expressions appear to be similar except that in (20) Gödel's implication is used instead the implication $a \Rightarrow b = \max(1 - a, b)$ used in (28). The presence of Gödel's implication in (20) comes from the fact $\hat{D}$ is defined as the solution of a fuzzy relation equation in analogy with the relation equation $D \circ R = M^+$ in the completely informed case. In this latter case there is an equivalence between $D \circ R = M^+$ and $D = \{d \in \mathscr{D}, \neg(\exists\, m \in M^+, (d,m) \notin R)\}$; this equivalence no longer holds in the fuzzy case. However $\hat{D}$ as given by (28) is still the greatest solution of $R \circ D \subseteq M^+$ if the composition $\circ$ is of the form sup-$*$ where $*$ is the non-commutative conjunction $a * b = b$ if $a > 1 - b$ and 0 otherwise [11].

Clearly (27) and (28) straightforwardly extend to subsets D of disorders which altogether explain both M^+ and M^-, substituting D to the singleton {d} in (27) and (28). Indeed when $\widehat{D} = \emptyset$, we have to look for two-element subsets D which may account for M^+ and M^-, and then for three-element subsets if there is no two-element one, and so on until a plausible explanation is found. The decomposition properties (8) when they hold, easily extend to the fuzzy case under the form

$$\begin{cases} \mu_{M(\{d_i,d_j\})^+} = \max(\mu_{M(d_i)^+}, \mu_{M(d_j)^+}) \\ \mu_{M(\{d_i,d_j\})^-} = \min(\mu_{M(d_i)^-}, \mu_{M(d_j)^-}). \end{cases} \quad (29)$$

Note that (29) is coherent with the definition of the union of twofold fuzzy sets. However, if we look for multiple disorders explaining a given set of manifestations, it is clear that we shall have a trade-off problem between small sets of disorders which are most plausible in the sense of the parsimony principle, and bigger sets of disorders which are more plausible because they ensure a better covering of the observed manifestations. This topic, along with the semantics of (29) for the representation of independent disorders, requests further investigation.

Turning to abduction, the idea of relevance expressed by equation (10) in the crisp case can be easily extended using the consistency index. Namely,

$$\mu_{\widehat{D}^*}(d) = \min(\mu_{\widehat{D}}(d), \max(\mathrm{cons}(M(d)^+, M^+), \mathrm{cons}(M(d)^-, M^-))). \quad (30)$$

and the subset $\widehat{D}^{**}$ of $\widehat{D}$ of the disorders whose sure effects cover the observations can be also extended by

$$\mu_{\widehat{D}^{**}}(d) = \min(\mu_{\widehat{D}}(d), \mathrm{inc}'(M^+, M(d)^+), \mathrm{inc}'(M^-, M(d)^-)) \quad (31)$$

where inc'(F,G) is a less drastic inclusion index than inc(F,G), since we only require that $\mathrm{inc}(F,G) = 1 \Leftrightarrow \mu_F \le \mu_G$ for expressing the coverage property. A lack of coverage by disorder d means that there is a manifestation m accepted as present to some degree($\mu_{M^+}(m) > 0$) which is not forecast by disorder d. It leads to choose

$$\mathrm{inc}'(F,G) = \inf_m \mu_F(m) \to \mu_G(m) \quad (32)$$

with $a \to b = 1$ if $a \le b$, and b otherwise.

5. Discussion and example

The modelling of uncertainty remains qualitative in the above approach. Indeed, we could use a finite completely ordered chain of levels of certainty ranging between 0 and 1, i.e. $\ell_1 = 0 < \ell_2 < \ldots < \ell_n = 1$ instead of [0,1], with $\min(\ell_i, \ell_k) = \ell_i$ and $\max(\ell_i, \ell_k) = \ell_k$ if $i \leq k$, and $1 - \ell_i = \ell_{n+1-i}$, and $\ell_i \rightarrow \ell_k = 1$ if $\ell_i \leq \ell_k$, $\ell_i \rightarrow \ell_k = \ell_k$ if $\ell_i > \ell_k$.

Taking into account the incomplete nature of the information about the presence or absence of manifestations decreases the discrimination power when going from the completely informed case (equation (2)) to the incomplete information case (equations (8) or (9)), since then the number of possible disorders in $\widehat{D}$ increases. This is due to the fact that now there are manifestations which are neither certain nor impossible and consequences of the presence of a given disorder d which are only possible, as pictured in Figure 2, while $M(d)^+ = M(d)$ and $M(d)^- = \overline{M(d)}$ in Figure 1 (similar figures can be drawn for M^+ and M^-).

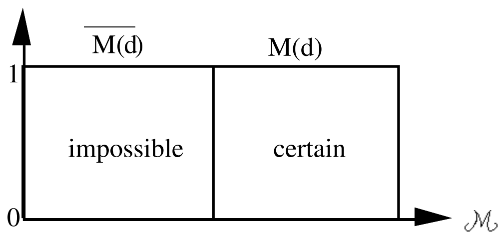


FIGURE 1. Completely informed situation

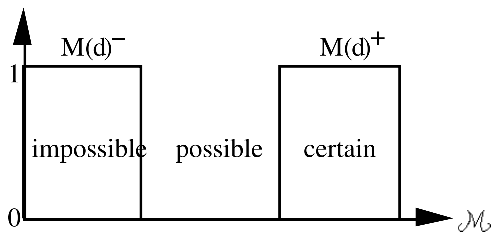


FIGURE 2. Partially informed situation

This suggests that, in order to improve the discrimination power of the model, we have to refine the non-fuzzy model in such a way that consequences (resp. manifestations) previously expressed as certain (resp. certainly present) and impossible (resp. certainly absent) remain classified in the same way and where some possible consequences (resp. possibly present manifestations) are now allowed to be either somewhat certain (resp. somewhat certainly present) or somewhat impossible (resp. somewhat certainly absent). See Figure 3. Then (27) enables us to rank-order the possible disorders which are compatible with the observations. This counterbalances the increase of candidates due to the incompleteness of the information.
It can be easily shown that adding preference levels on top of incompleteness can at the same time enable the discrimination power of the completely informed situation to be recovered, and also enable the extra disorders obtained on the partially informed situation to be ranked in terms of their plausibility levels. Indeed, assume that the dichotomy M(d), $\overline{M(d)}$ of Figure 1 represents a first approximation of the fuzzy sets $\widetilde{M}(d)^+$ and $\widetilde{M}(d)^-$ of a modified Figure 3 in the sense that

$$M(d) = \mathrm{support}(\widetilde{M}(d)^+)\ ;\ \overline{M(d)} = \mathrm{support}(\widetilde{M}(d)^-)$$

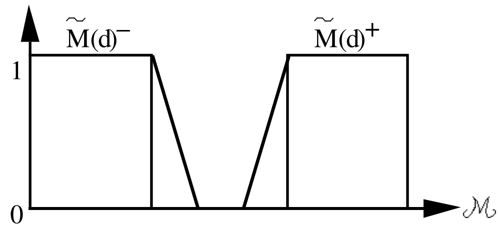


FIGURE 3. Partially informed situation with preferences

and moreover $M(d)^+ = \text{core}(\widetilde{M}(d)^+)$, $M(d)^- = \text{core}(\widetilde{M}(d)^-)$. Then it can be proved that both results of the completely informed and partially informed situations can be recovered by the fuzzy model. Namely, let $\widehat{D}_c$, $\widehat{D}_p$, and $\widehat{D}$ be the sets of disorders obtained via equations (2), (8) and (31) respectively. We shall assume non-fuzzy incomplete observations, in which case (2) becomes $\widehat{D}_c = \{d \in \mathcal{D}, M^+ \subseteq M(d) \subseteq \overline{M^-}\}$. Then we can show the following equalities, $\widehat{D}_c = \text{core}(\widehat{D})$; $\widehat{D}_p = \text{support}(\widehat{D})$ (see [12] for a proof). This result indicates that our approach, although much more qualitative than one based on probability theory, still possesses the ability to rank-order the set of plausible unique disorders explaining an incomplete set of manifestations.

Example: Let us show the benefit of grading uncertainty on an example. Let $\mathcal{D} = \{d_1, d_2, d_3, d_4\}$, $\mathcal{M} = \{m_1, m_2, m_3\}$ and let us first consider the completely informed case

$$M(d_1)^+ = \{m_2\},\ M(d_1)^- = \{m_1, m_3\}$$
$$M(d_2)^+ = \{m_3\},\ M(d_2)^- = \{m_1, m_2\}$$
$$M(d_3)^+ = \{m_1, m_2, m_3\},\ M(d_3)^- = \emptyset$$
$$M(d_4)^+ = \{m_2, m_3\},\ M(d_4)^- = \{m_1\},$$
and $$M^+ = \{m_2, m_3\},\ M^- = \{m_1\}.$$

Then, we get $\widehat{D} = \{d_4\}$.

Let us weaken our example into an incompletely informed case

$$M(d_1)^+ = \emptyset,\ M(d_1)^- = \{m_1\}$$
$$M(d_2)^+ = \{m_3\},\ M(d_2)^- = \emptyset$$
$$M(d_3)^+ = \{m_3\},\ M(d_3)^- = \emptyset$$
$$M(d_4)^+ = \emptyset,\ M(d_4)^- = \{m_1\}.$$

Then $\widehat{D} = \{d_1, d_2, d_3, d_4\}$, i.e. all causes are possible. Let us now consider the following table where the previous incompletely informed case is completed with some gradual uncertainty information that is in accordance with the complete information case.

	$M(d_1)^+$	$M(d_1)^-$	$M(d_2)^+$	$M(d_2)^-$	$M(d_3)^+$	$M(d_3)^-$	$M(d_4)^+$	$M(d_4)^-$	$\mathcal{M}$	
membership m_1	0	1	0	0.9	0.7	0	0	1	0	M^-
m_2	0.8	0	0	0.6	0.5	0	0.7	0	1	M^+
grades m_3	0	0.8	1	0	1	0	0.3	0	1	M^+
$\mathrm{cons}(M(d)^+,M^-)$	0		0		0.7		0			
$\mathrm{cons}(M(d)^-,M^+)$		0.8		0.6		0		0		
eq. (27): $\mu_{\hat{D}}(d)$	0.2		0.4		0.3		1			

Table 1

Thus we obtain that d_4 is much more plausible than d_2, itself more plausible than d_3, which is more plausible than d_1, i.e. a ranking of the possible disorders is obtained. This is in accordance with the complete information case as well as the incomplete information case, but significantly improves their findings. Note that all disorders are relevant in the sense of (30), but only d_4 is fully relevant in the sense of (31).

Clearly, the above approach straightforwardly extends to subsets D of disorders which altogether explain both M^+ and M^-, substituting D to the singleton {d} in (22), (23), (30) and (31). Indeed when $\hat{D} = \emptyset$, we have to look for two-element subsets D which may account for M^+ and M^-, and then for three-element subsets if there is no two-element one, and so on until a plausible explanation is found.

6 - Concluding Remarks

In this paper we have analyzed the fuzzy set approach to diagnosis problems and proposed a new model, more expressive for representing the available information, and where the intended meaning of the membership degrees has been clarified: here, they pertain to uncertainty. The approach presented in this section is currently applied to a fault diagnosis problem in satellites [5,6]. More elaborate uncertainty-handling models also allowing for non-binary attributes for expressing the intensity of manifestations and the intensity of association between disorders and manifestations still have to be developed.

In spite of its greater expressiveness, the present model still has several limitations. Let us mention two of them. The relational model we consider associates directly disorders and manifestations. More generally we may have two relations, between $\mathcal{D}$ and an intermediary set $\mathcal{S}$, and between $\mathcal{S}$ and $\mathcal{M}$; see (Peng and Reggia[20]) on this point. Besides, we are not capable of capturing the most general kind of incomplete information. For instance we cannot express that we are certain that manifestations m_i or m_j are present (but perhaps one of them is absent), or a similar information stating that when disorder d is present it is certain that m_i or m_j are present (and that m_k or m_ℓ are absent) for instance. See [10] for the modelling of such pieces of knowledge in case of graded uncertainty. The treatment of the most general kind of incomplete information would require to work with a (fuzzy) relation R defined on $2^{\mathcal{D}} \times 2^{\mathcal{M}}$.

Another topic for further research is the embedding of the proposed approach into a logical formalism , in order to relate it with methods based on possibilistic assumption-based truth-maintenance systems ; see (Benferhat et al.[4]). More generally it would be interesting to develop a logical framework where it is possible to express both weighted causal rules associating manifestations to disorders and weighted evocation rules (in the sense of Pearl [18]) associating plausible disorders to manifestations, and perform local reasoning tasks with this kind of hybrid knowledge.

References

1. Adlassnig K.P., Scheithauer W., Kolarz G. (1986) Fuzzy medical diagnosis in a hospital. In : Fuzzy Logic in Knowledge Engineering (H. Prade, C.V. Negoita, eds.), Verlag TÜV Rheinland, Köln, 275-294.

2. Asse A., Maizener A., Moreau A., Willaeys D. (1987) Diagnosis based on subjective information in a solar energy plant. In : Approximate Reasoning in Intelligent Systems, Decision and Control (E. Sanchez, L.A. Zadeh, eds.), Pergamon Press, 159-173.

3. Bandler W., Kohout L. (1980) Fuzzy power sets and fuzzy implication operators. Fuzzy Sets and Systems, 4, 13-30.

4. Benferhat S., Dubois D., Lang J., Prade H. (1994) Hypothetical reasoning in possibilistic logic : basic notions and implementation issues. In : Advances in Fuzzy Systems : Applications and Theory – Vol. I (P.Z. Wang, K.F. Loe, eds.), World Scientific Publ., Singapore, 1-29.

5. Cayrac, D., Dubois, D., Haziza, M., and Prade, H., Relational approach to fault diagnosis based on a functional model, in Proc. of the 8th Inter. Conf. on Artificial Intelligence in Engineering (AIENG'93), Rzevski, G., Pastor, J., and Adey, R. A., Eds., Toulouse, June 1993, Computational Mechanics Publ., Ashurst, UK, 358-367

6. Cayrac, D., Dubois, D., Haziza, M., and Prade, H., Possibility theory in fault mode effects analyses - a satellite fault diagnosis application. Proc. 3rd IEEE Inter. Conf. On Fuzzy Systems, Orlando, Fl., 1176-1181.

7. Di Nola A., Sessa S., Pedrycz W., Sanchez E. (1989) Fuzzy Relation Equations and their Applications to Knowledge Engineering. Kluwer Academic Publ., Dordrecht.

8. Dubois D., Prade H. (1987) Twofold fuzzy sets and rough sets — Some issues in knowledge representation. Fuzzy Sets and Systems, 23, 3-18.

9. Dubois D., Prade H. (with the collaboration of Farreny H., Martin-Clouaire R., Testemale C.) (1988a) Possibility Theory — An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (French editions, Masson, 1985 and 1987).

10. Dubois D., Prade H. (1988b) Incomplete conjunctive information. Comput. Math. Applic., 15(10), 797-810.

11. Dubois D., Prade H. (1992) Upper and lower images of a fuzzy set induced by a fuzzy relation: applications to fuzzy inference and diagnosis. Information Sciences, 64, 203-232.

12. Dubois, D., and Prade, H., A fuzzy relation-based extension of Reggia's relational model for diagnosis handling uncertain and incomplete information, Proc. of the Inter. Conf. on Uncertainty in Artificial Intelligence, Heckerman, D., and Mamdani, E. H., Eds., 1993, 106-113.

13. Farreny H., Prade H. (1989) Positive and negative explanations of uncertain reasoning in the framework of possibility theory. In : Fuzzy Logic for the Management of Uncertainty (L.A. Zadeh, J. Kacprzyk, eds.), Wiley, 1992, 319-333.

14. Farreny H., Prade H. (1990) Explications de raisonnements dans l'incertain. Revue d'Intelligence Artificielle, 4(2), 43-75.

15. Gmytrasiewicz P., Hassberger J.A., Lee J.C. (1990) Fault tree based diagnostics using fuzzy logic. IEEE Trans. on Pattern Analysis and Machine Intelligence, 12(11), 1115-1119.

16. Kitowski J., Bargiel M. (1987) Diagnosis of faulty states in complex physical systems using fuzzy relational equations. In : Approximate Reasoning in Intelligent Systems, Decision and Control (E. Sanchez, L.A. Zadeh, eds.), Pergamon Press, 175-194.

17. Pappis C.P., Sugeno M. (1985) Fuzzy relational equations and the inverse problem. Fuzzy Sets and Systems, 15, 79-90.

18. Pearl J. (1988) Embracing causality in default reasoning. Artificial Intelligence, 35, 259-271.

19. Pedrycz W. (1991) Processing in relation structures : fuzzy relational equations. Fuzzy Sets and Systems, 40, 77-106.

20. Peng Y., Reggia J.A. (1987) Diagnostic problem-solving with causal chaining. Int. J. of Intelligent Systems, 2(3), 265-302.

21. Peng Y., Reggia J.A. (1990) Abductive Inference Models for Diagnostic Problem-Solving. Springer Verlag, New York.

22. Reggia J.A., Nau D.S., Wang P.Y., Peng H. (1985) A formal model of diagnostic inference. Information Sciences, 37, 227-285.

23. Sanchez E. (1977) Solutions in composite fuzzy relation equations : application to medical diagnosis in Brouwerian logic. In : Fuzzy Automata and Decision Processes (M.M. Gupta, G.N. Saridis, B.R. Gaines, eds.), North-Holland, Amsterdam, 221-234.

24. Sanchez E. (1979) Inverses of fuzzy relations, application to possibility distributions and medical diagnosis. Fuzzy Sets and Systems, 2, 75-86. Preliminary version in : Proc. of the 1977 IEEE Conf. on Decision & Control, New Orleans, Louisiana, Dec. 7-9, 1977, Vol. 2, 1384-1389.

25. Shahinpoor M., Wells D.J. (1980) Application possibilities for fuzzy failure analysis and diagnosis of reactor plant components and areas. Nucl. Eng. & Design, 61, 93-100.

26. Tsukamoto Y., Terano T. (1977) Failure diagnosis by using fuzzy logic. Proc. of the IEEE Conf. on Decision and Control, New Orleans, 1390-1395.

27. Yager R.R. (1985) Explanatory models in expert systems. Int. J. of Man-Machine Studies, 23, 539-550.

28. Yager R.R. (1988a) Measures of accountability and credibility in knowledge based diagnosis systems. In : Engineering Risk and Hazard Assessment – Vol. III (A. Kandel, E. Avni, eds.), CRC Press, Boca Raton, Fl., 1-15.

29. Yager R.R. (1988b) On ordered weighted averaging operators in multi-criteria decision-making. IEEE Trans. on Systems, Man and Cybernetics, 18, 183-190.

30. Zadeh L.A. (1979) A theory of approximate reasoning. In : Machine Intelligence, Vol. 9 (J.E. Hayes, D. Michies, L.I. Mikulich, eds.), Halstead Press, New York, 149-194.

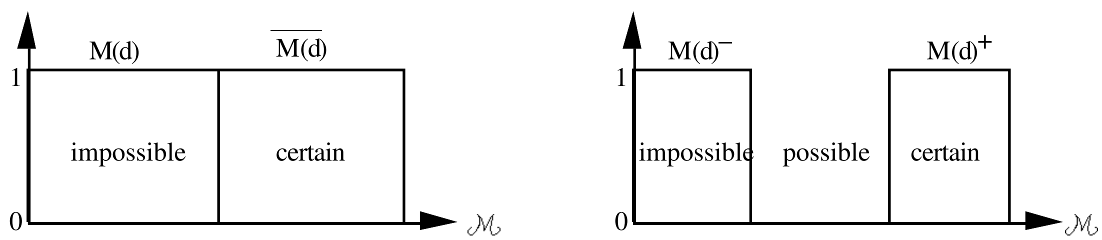


FIGURE 1. Completely informed situation

FIGURE 2. Partially informed situation

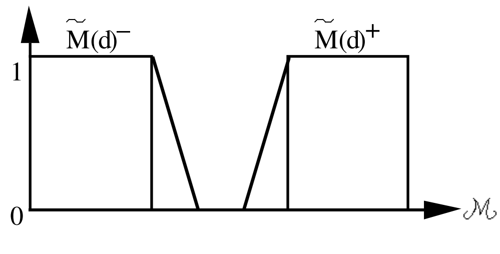
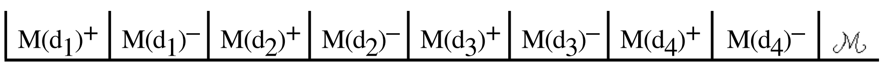


FIGURE 3. Partially informed situation with preferences



membership grades	m_1	0	1	0	0.9	0.7	0	0	1	0	M^-
	m_2	0.8	0	0	0.6	0.5	0	0.7	0	1	M^+
	m_3	0	0.8	1	0	1	0	0.3	0	1	M^+
$\mathrm{cons}(M(d)^+,M^-)$		0		0		0.7		0			
$\mathrm{cons}(M(d)^-,M^+)$			0.8		0.6		0		0		
eq. (27): $\mu_{\widehat{D}}(d)$		0.2		0.4		0.3		1			

Table 1